Volume Title: Essays in the Economics of Crime and Punishment

Volume Author/Editor: Gary S. Becker and William M. Landes, eds.

Volume Publisher: NBER

Volume ISBN: 0-87014-263-1

Volume URL: http://www.nber.org/books/beck74-1

Publication Date: 1974

Chapter Title: An Economic Analysis of the Courts

Chapter Author: William M. Landes

Chapter URL: http://www.nber.org/chapters/c3629

Chapter pages in book: (p. 164 - 214)

An Economic Analysis of the Courts

William M. Landes

University of Chicago and National Bureau of Economic Research

"The object of our study, then, is prediction, . . . The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law." Oliver Wendell Holmes, Jr., The Path of Law (1897).

In the folklore of criminal justice a popular belief is that the accused will have his case decided in a trial. Empirical evidence does not support this belief. Table 1 indicates that most cases are disposed of before trial by either a guilty plea or a dismissal of the charges. What factors determine the choice between a pretrial settlement and a trial? What accounts for the large proportion of settlements compared to trials? How are certain aspects of the criminal justice process such as the bail system and court delay related to the decision to settle or to go to trial? The main

This study has been supported by a grant for the study of law and economics from the National Science Foundation to the National Bureau (Grant Number GS-3314). The views expressed in this essay are not attributable to the National Science Foundation, whose support I gratefully acknowledge. I should like to thank Professors Gary Becker, Solomon Fabricant, Laurence Miller, Sherwin Rosen, Finis Welch and Neil Wallace, and Elisabeth Landes for helpful criticisms. I also received useful comments at seminars at the NBER, Columbia, Rochester, U.C.L.A. and the University of Chicago. Charles H. Berry, Eugene P. Foley, and Eli Goldston provided valuable advice as members of the reading committee of the National Bureau's Board of Directors.

TABLE 1
DISPOSITION OF CRIMINAL CASES

Area (Year)	Number of Defendants	Trials Number	Trials Per Cent	Guilty Pleas Number	Guilty Pleas Per Cent	Dismissed Number	Dismissed Per Cent
132 State County Courts (1962)	7,510[a]	1,394	19	5,293	70	823	11
U.S. District Courts (1967)	31,535	4,208	13	23,131	73	4,196	13

SOURCES.—Lee Silverstein, Defense of the Poor in Criminal Cases in American State Courts, A Field Study and Report (2 v. 1965); Ann. Rep., Admin. Off. of the United States Courts, 1967.

[a] Number of felony defendants in sample.

purpose of this essay is to answer these questions by means of a theoretical and empirical analysis of the criminal justice system using standard tools of economic theory and statistics.

A theoretical model is first developed that identifies the variables relevant to the choice between a settlement and a trial. The basic assumption of the model is that both the prosecutor and the defendant maximize their utility, appropriately defined, subject to a constraint on their resources. It is shown that the decision to settle or to go to trial depends on the probability of conviction by trial, the severity of the crime, the availability and productivity of the prosecutor's and defendant's resources, trial versus settlement costs, and attitudes toward risk. We then analyze the effects of the bail system and court delay on settlements, and consider several proposals for improving the bail system and reducing court delay. These include "preventive detention," monetary compensation to defendants not released on bail, and the imposition of a money price for the use of the courts. The model is further useful in evaluating the frequently made argument that the criminal justice system discriminates against low-income defendants. This proposition is analyzed by relating a defendant's income or wealth to his decision to settle or go to trial, the probability of his conviction, and his sentence if convicted. The interactions of these factors with the bail system and court delay are also examined.

The second part of this study is an empirical analysis from published data on the disposition of cases in state and federal criminal courts.

Multiple regression techniques are used to test the effects on the demand for trials (or conversely, settlements) and on the probability of conviction of the following: (1) pretrial detention; (2) court queues; (3) the size of the potential sentence; (4) judicial expenditures; (5) subsidizing defendants' legal fees; and (6) demographic variables such as population size, region, county income, per cent nonwhite, and urbanization. Finally, in the appendix a theoretical and empirical analysis on the demand for civil cases is presented.

I. THE MODEL

We make the following assumptions.

(1) There are n defendants.

(2) The probability of conviction in a trial for the ith defendant ($i = 1, \ldots, n$) depends on the prosecutor's and defendant's inputs of resources, R_i^* and R_i respectively, into the case. That is,

$$P_i^* = P_i^*(R_i^*, R_i; Z_i)$$

and

$$P_i = P_i(R_i^*, R_i; Z_i), \tag{1}$$

where P_i^* is the prosecutor's and P_i is the defendant's estimates of the probability of conviction by trial. P_i^* can be greater, less than, or equal to P_i. Z_i denotes other factors affecting the level of P_i^* and P_i; for example, the availability of witnesses, the defendant's past record, his alibi, etc. Inputs of R_i^* would tend to raise P_i^* and P_i, while inputs of R_i would tend to lower them so that

$$\begin{aligned} \frac{\partial P_i^*}{\partial R_i^*} &\geqslant 0 \qquad & \frac{\partial P_i^*}{\partial R_i} &\leqslant 0 \\ \frac{\partial P_i}{\partial R_i^*} &\geqslant 0 \qquad & \frac{\partial P_i}{\partial R_i} &\leqslant 0. \end{aligned} \tag{2}$$

(3) The sentence, S_i, the defendant would receive if convicted in a trial is known to the prosecutor and defendant and independent of R_i^* and R_i.[1]

1. There is some justification for this assumption other than mathematical simplicity since most crimes carry statutory penalties which are presumably known to both parties and independent of R_i^* and R_i. However, statutory penalties usually set a minimum and maximum sentence for the defendant convicted in a trial, and within this range the sentence received would partly depend on R_i^* and R_i. To allow the sentence to be a function of R_i^* and R_i would substantially complicate the model at this point (for example, two sen-

(4) Initially, there is no money charge for the use of the courts nor a nonmoney cost in terms of court delay or queues.

PROSECUTOR

Let the prosecutor's decision rule be to maximize the expected number of convictions weighted by their respective S_i's – he prefers longer to shorter sentences – subject to a constraint on the resources or budget available to his office (B).[2] This decision rule coincides with the social optimum in the following sense. If expected sentences are regarded as prices the community charges for various offenses, then the prosecutor's behavior is equivalent to maximizing the community's "profit" for a given level of prosecution expenditures.

The prosecutor maximizes $E(C)$ where

$$E(C) = \sum_{i=1}^{n} P_i^* S_i + \lambda(B - \sum_{i=1}^{n} R_i^*), \tag{3}$$

which yields the equilibrium conditions

$$\frac{\partial P_1^*}{\partial R_1^*} \cdot S_1 = \frac{\partial P_2^*}{\partial R_2^*} \cdot S_2 = \cdot \cdot \cdot , = \frac{\partial P_n^*}{\partial R_n^*} \cdot S_n. \tag{4}$$

Thus, the prosecutor allocates greater resources to cases, ceteris paribus, where the sentence is greater and where P_i^* is more responsive to changes in R_i^*.[3] If all n defendants need not be prosecuted, one would also predict charges would be dismissed when the prosecutor sees little chance of conviction regardless of his resource input into the trial, or given a conviction he expects a negligible sentence. The formulation of (3) is suffi-

tences would have to be included – the defendant's estimate and the prosecutor's estimate) without substantially changing the analysis of the trial versus settlement decision. In a later section on wealth effects, I allow the sentence to be a function of resource inputs.

2. Other decision rules are possible; for example, maximizing the expected number of convictions without weighting by the S_i's. The difficulty here is that the prosecutor, in order to increase his convictions and conserve his resources, would often be willing to drop a murder charge against a suspected murderer if the latter agreed to plead guilty to a minor offense (for example, a traffic violation). A simple way of eliminating this in the model is to weight by the S_i's. (See the analysis of a settlement presented below.) Note that fines could be included in S_i by specifying a rate at which the prosecutor transforms fines into sentences keeping his utility constant.

3. We assume the price of a unit of R_i^* is \$1.00 and $\partial^2 P_i^* / \partial R_i^* < 0$. It should also be noted that (4) does not necessarily have a unique solution unless one assumes that the prosecutor takes as given the defendant's inputs, R_i's. If he readjusts his inputs of R_i^* to changes or anticipated changes in any of the R_i's, then the defendants may in turn readjust their R_i's, and so forth. This process need not converge to a unique solution.

ciently general to give the prosecutor discretion over the type of charge brought against each defendant. The charge selected would be one that maximized $E(C)$. Further, the maximization of $E(C)$ together with the assumption that $\partial P_i^*/\partial R_i^* \geq 0$ imply (possibly unrealistically) that the prosecutor would suppress any evidence that reduces the probability of conviction.

Scarce resources provide an incentive for the prosecutor to avoid a trial and negotiate a pretrial settlement with the defendant. From (3) and (4) it follows that if the prosecutor's transaction costs of a settlement equal his optimal resource expenditure on a trial, he would be willing to offer the suspect a reduction in the sentence below S_i in exchange for a plea of guilty (which makes $P_i^* = P_i = 1$).[4] However, since trial costs probably exceed these transaction costs, he would be willing to offer a further sentence reduction as the savings in resources can be used to increase the conviction probabilities in other cases. If ΔS_i denotes the sentence reduction that is a positive function of the difference between the prosecutor's trial costs and transaction costs of a settlement, then

$$So_i = P_i^* S_i - \Delta S_i, \tag{5}$$

where So_i is the minimum sentence the prosecutor is willing to offer the defendant for a guilty plea.[5] From (5) we note that the terms offered the defendant will be more favorable the lower P_i^* and S_i, and the greater the prosecutor's resource saving from a settlement. Finally, suppose that certain cases bring the prosecutor considerable notoriety only if a trial occurs. If notoriety were desired, the sentence variable, S_i, could be increased by a notoriety factor (for example $S_i(1 + j_i)$ where j_i is a positive function of the amount of notoriety and is ≥ 0). Hence in some cases So_i could be greater than $P_i^* S_i$ and even S_i. Unless otherwise stated, we assume $So_i < P_i^* S_i$.

---

4. The prosecutor's transaction costs of a settlement would equal his time spent explaining the terms of the offer to the suspect and judge, paperwork in his office, etc. These costs will generally be less than his total costs of reaching a settlement since the latter may involve substantial negotiating or bargaining costs in order to arrive at a sentence more preferred than the minimum sentence he is willing to offer in a settlement.

5. A settlement that releases resources from any one case will increase the R_i^*'s in other cases. Thus, the R_i^*'s that initially satisfy (4) are not the final equilibrium values because adjustments take place as cases are settled. Moreover, these adjustments raise the So's in cases not yet settled. I largely ignore these secondary effects in the analysis.

DEFENDANT

If the defendant goes to trial, the outcome is either of two mutually exclusive states: a conviction state with an endowment Wc defined as

$$Wc = W - s \cdot S - r \cdot R \tag{6a}$$

or a nonconviction state with an endowment Wn defined as

$$Wn = W - r \cdot R. \tag{6b}$$

W is his wealth endowment prior to arrest, s equals the present value of the average pecuniary and nonpecuniary losses per unit of jail sentence, r is the average price of a unit of R, and S and R are defined as before.[6] I assume Wc is nonnegative.

Let U be a continuous utility function over the defendant's endowment. His expected utility from going to trial is then

$$E(U) = PU(Wc) + (1 - P)U(Wn). \tag{7}$$

Since inputs of R lower P, Wc and Wn, the defendant would select a level of R to maximize $E(U)$ such that

$$-P'[U(Wn) - U(Wc)] = r[PU'(Wc) + (1 - P)U'(Wn)], \tag{8}$$

where $P' = dP/dR$ and U' denotes the marginal utility (>0) of the endowment in each state.[7] The left-hand side of (8) represents marginal returns of R and the right-hand side, marginal costs of R.[8] An analysis of the determinants of the optimal R is presented later.

6. The subscript i is deleted, since it is explicit that we are now dealing with one defendant.

7. The qualification stated in footnote 3 regarding the prosecutor's equilibrium inputs of R^* also applies to the defendant's equilibrium inputs of R.

8. The second-order condition for the optimum R requires that the rate of change of marginal returns be less than the rate of change of marginal cost. That is,

$$-P''[U(Wn) - U(Wc)] + rP'[U'(Wn) - U'(Wc)] < -rP'[U'(Wn) - U'(Wc)] - r^2[PU''(Wc) + (1 - P)U''(Wn)],$$

where $P'' = d^2P/dR^2$, and U'' = the rate of change of U'. P'' is assumed >0 to indicate diminishing marginal product of R in reducing P. If $U'' = 0$, the last three terms above are zero and hence marginal returns are falling while marginal costs are constant. If $U'' \neq 0$,

TRIAL VERSUS SETTLEMENT

Let $r \cdot \hat{R}$ equal the defendant's transaction costs of a settlement.[9] Note that the defendant's trial costs, $r \cdot R$, are greater than $r \cdot \hat{R}$ because a defendant going to trial will in the process of rejecting a settlement incur most of the costs in $r \cdot \hat{R}$, and in addition he has expenditures on the trial. The defendant would choose between a trial or settlement on the basis of whether his expected utility from the former, $E(U)$, were greater or less than his utility from the latter. Similarly, the prosecutor would choose the alternative that maximizes his conviction function, $E(C)$. Therefore, a necessary condition for a settlement is that both the defendant and prosecutor simultaneously gain from a settlement compared to their expected trial outcomes. This requires that

$$\pi = U(W - s \cdot So - r \cdot \hat{R}) - E(U) > 0, \tag{9}$$

because one can then find a negotiated sentence somewhat greater than So, the minimum offer of the prosecutor, that leaves the defendant with a utility from a settlement greater than $E(U)$ and at the same time increases $E(C)$ for the prosecutor above its value in a trial. Although (9) explicitly allows for the prosecutor's and defendant's transaction costs of a settlement, the attempt to reach mutually acceptable terms may in certain cases involve substantial bargaining costs that are large enough to prevent a settlement even though $\pi > 0$. In spite of this qualification, I will assume that $\pi > 0$ is not only a necessary but also a sufficient condition for a settlement. Alternatively, $\pi < 0$ is a necessary and sufficient condition for a trial. These conditions are Pareto optimal in that if $\pi > 0$, both parties expect to gain from a settlement, and if $\pi < 0$, both parties expect to gain from a trial.

We can derive the following implications from (9) regarding the likelihood of settling and the resulting sentence.

1. Although the precise sentence in a pretrial settlement is indeterminate, it must lie between the extremes defined by (9). Within this range it would depend on the relative bargaining strengths of the parties involved. In general, one would expect a smaller negotiated sentence the

---

marginal costs may be rising, falling or constant with increases in R since the two terms on the right-hand side are of opposite sign. Similarly, when $U'' < 0$, marginal returns may actually rise since $rP'[U'(Wn) - U'(Wc)]$ is positive but when $U'' > 0$, marginal returns must fall.

9. Similar to the definition of the prosecutor's transaction costs (see *supra* note 4), $r\hat{R}$ would be generally less than the defendant's total costs of negotiating a settlement since $r\hat{R}$ excludes bargaining costs.

smaller the probability of conviction in a trial. A smaller P raises $E(U)$ and thus lowers the maximum sentence accepted by the defendant, while a smaller P^* reduces the minimum acceptable to the prosecutor. For identical reasons, a lower sentence if convicted by trial, S, should lead to a lower negotiated sentence.

2. π will be positive and a settlement chosen whenever

$$s \cdot So < r(R - \hat{R}), \tag{10}$$

since this implies $U(W - s \cdot So - r\hat{R}) > U(Wn)$, and by definition $U(Wn) \geq E(U)$. This result is independent of the defendant's attitude toward risk and his estimate of the conviction probability, because regardless of the trial outcome he is always better off with a settlement. (10) implies that a trial is less likely for offenses with small expected sentences (since So depends on S and P^*) relative to the defendant's differential cost of going to trial, $r(R - \hat{R})$.[10] Except when explicitly stated to the contrary, I now assume $s \cdot So > r(R - \hat{R})$ so that Wn is greater than $(W - s \cdot So - r\hat{R})$.

3. If both parties agree on the probability of conviction by trial ($P^* = P$), a settlement will take place for defendants who are risk averse ($U'' < 0$) or risk neutral ($U'' = 0$).[11] When $P^* = P$, one can show that a trial is equivalent to an unfair gamble (that is, the expected trial endowment is less than the settlement endowment).[12] Risk neutral suspects maximize their expected endowment and, therefore, refuse the trial "gamble," and a fortiori risk averse suspects also refuse the "gamble." On the other hand, a trial can still occur for a risk preferrer ($U'' > 0$) even though $P^* = P$.[13]

---

10. This provides an explanation of why many persons plead guilty to traffic violations instead of spending considerable time in traffic court disputing them.

11. U'' denotes the rate of change of U' with respect to one's endowment.

12. A trial is an unfair gamble if

$$(W - s \cdot So - r\hat{R}) - [P \cdot Wc + (1 - P)Wn] > 0, \tag{i}$$

which can be rewritten as

$$s \cdot \Delta S + r(R - \hat{R}) > 0, \tag{ii}$$

using (5), (6a, 6b) and the assumption $P^* = P$. Since we have assumed $s \cdot \Delta S$ and $r(R - \hat{R})$ are both positive, (ii) holds and the gamble is unfair.

13. Given risk preference, a negative π, which leads to a trial, would be more likely the greater the preference for risk, the larger $r\hat{R}$, and the smaller $s \cdot \Delta S$ and rR. To prove this differentiate (9) partially with respect to these variables. The partial derivatives are negative for $r\hat{R}$, and positive for $s \cdot \Delta S$ and rR, indicating that π falls with respect to increases in $r\hat{R}$ and decreases in $s \cdot \Delta S$ and rR.

4. Suppose the prosecutor and defendant differ in their estimates of the trial conviction probability. If $P^* < P$, a trial becomes an even less favorable gamble in comparison to $P^* = P$, and hence risk averse and risk neutral suspects would continue to settle.[14] Risk preferrers are also more likely to settle since π in (9) rises. $P^* > P$ is the more interesting case because this provides an explanation in addition to risk preference of why trials occur. When $P^* > P$ a trial becomes a more favorable gamble compared to $P^* = P$, and hence π falls, increasing the chances of a trial. Moreover, if $P^* > P$, one can show that the likelihood of a trial is generally greater for defendants accused of crimes that carry stronger penalties.[15]

Several additional points are worth noting in regard to the settlement versus trial decision:

5. The greater the savings in costs from a settlement, other things the same, the smaller So and $r\hat{R}$, and the more likely a settlement. This suggests that policy measures designed to eliminate or subsidize the defendant's legal fees, which in turn reduce the cost differential between a trial and a settlement, will increase the proportion of trials.

6. Suppose a not-guilty verdict in a trial produces pecuniary and nonpecuniary returns to the defendant. This would raise $E(U)$ and make a trial more likely. Similarly, publicity gains to the prosecutor from a trial would raise So, as previously noted, and also make a trial more likely.

7. The question of whether the defendant did in fact commit the crime he is charged with does not explicitly enter the analysis. The prosecutor and defendant have been assumed to react to the probability of

14. As P^* falls, So falls, which in turn increases $(W - s \cdot So - r\hat{R})$. Similarly, the increase in P lowers $[PWc + (1 - P)Wn]$. Thus, the value of (i) in footnote 12 rises relative to the case where $P^* = P$. Since (i) is already >0 when $P^* = P$, it is obviously >0 when $P^* < P$.

15. Differentiating π with respect to S and noting that $So = P^* \cdot S - \Delta S$ (see (5)) yields $\partial\pi/\partial S \lesseqgtr 0$ according as

$$\frac{P}{P^*} \lesseqgtr \frac{U'(W - s \cdot So - r\hat{R})}{U'(Wc)}. \tag{i}$$

$\partial\pi/\partial S < 0$ when $U'' \geq 0$ since $U'(W - s \cdot So - r\hat{R}) \geq U'(Wc)$ and $P < P^*$. Thus, risk preferring and risk neutral defendants are more likely to go to trial as S rises given $P^* > P$. When $U'' < 0$ risk aversion), both sides of (i) are <1, and the sign of $\partial\pi/\partial S$ is indeterminate. However, if the degree of risk aversion is weak (the right-hand side of (i) is close to one), risk averters are also more likely to go to trial as S rises.

In another sense, the likelihood of a trial is *always* greater for large than for small sentences. We have already shown in (10) that a trial will not occur when $s \cdot So$ is less than the difference in costs between a trial and a settlement, $r(R - \hat{R})$. Thus, for very small sentences $r(R - \hat{R})$ is likely to dominate and a settlement will take place.

conviction and other variables in choosing between settling and going to trial, while their behavior has not been directly influenced by the actual guilt or innocence of the defendant. However, this factor may enter in two ways. First, the amount and quality of the evidence against the defendant seems likely to diminish in the case of an innocent person. This would reduce the probability of conviction in a trial or even lead the prosecutor to dismiss charges more readily since P^* may be close to zero. Second, an innocent person may have an aversion to lying so that he would have a greater reluctance to plead guilty to an offense than a guilty person. This can be interpreted as imposing psychic losses on a guilty plea for an innocent suspect which would reduce $U(W - s \cdot So - r\hat{R})$ in (9) and hence increase the likelihood of a trial.

8. We observed in the introduction that a large fraction of cases are settled before trial. Our analysis predicts this if in most cases the prosecutor and suspect agree on the expected outcome of a trial, the costs of a trial to both parties exceed their settlement costs, and suspects are generally risk averse in their trial versus settlement choice.

WEALTH AND SENTENCE EFFECTS

In this section two further questions are considered. (1) Do the resources (R) invested by the defendant in a trial rise as the sentence increases? (2) Do the resources invested increase with the level of the defendant's initial endowment or wealth? The latter question is directly related to the widespread claim that the criminal justice system works less favorably for low income suspects than for affluent ones,[16] because if the defendant's investment of resources rises with wealth, then both the probability of conviction in a trial and a negotiated sentence would tend to be lower for wealthier defendants.

To determine the effect of an increase in the sentence, we take the total differential of the first-order condition in (8) with respect to S and R.[17] This yields $dR/dS \gtreqless 0$ according as

16. See Patricia M. Wald, Poverty and Criminal Justice, in U.S. Pres. Comm'n on Law Enforcement and Admin. of Justice, Task Force on the Admin. of Justice, Task Force Report: The Courts, at 139, app. C (1967).

17. The differential is

$$\frac{dR}{ds} = \frac{s[P'U'(Wc) - rPU''(Wc)]}{-P''[U(Wn) - U(Wc)] + 2rP'[U'(Wn) - U'(Wc)] + r^2[PU''(Wc) + (1 - P)U''(Wn)]}.$$

The second-order condition for $E(U)$ to be a maximum requires that the denominator be <0. Hence, $dR/dS \gtreqless 0$ as $P'U'(Wc) - rPU''(Wc) \lesseqgtr 0$.

$$\frac{-P'}{r \cdot P} \gtrless -\frac{U''(Wc)}{U'(Wc)}, \tag{11}$$

where $-U''(Wc)/U'(Wc)$ is a measure of absolute risk aversion. From (11) it follows that $dR/dS > 0$ for defendants who are risk preferrers or risk neutral. If defendants are risk averse, the sign of dR/dS is uncertain. It is more likely to be positive the more responsive P to increases in R, the lower r, and the smaller the level of absolute risk aversion.[18] In sum, for a group of defendants differing in their attitudes toward risk, we might expect to find a greater investment of resources on average for defendants charged with crimes carrying longer sentences. Note that this need not lead to an observed negative relation between the probability of conviction and the severity of the crime since we have previously shown that an increase in the potential sentence also induces the prosecutor to allocate more resources to the case.

The value of one's time is generally related positively to one's income and wealth. In consequence, an increase in the defendant's wealth will lead to an increase both in r and s, the prices per unit of R and S respectively. To show this for r, let R be produced by both inputs of market goods such as the services of lawyers, expert witnesses, etc., and inputs of one's time. The optimal input combination is where the marginal products of the inputs over their respective marginal factor costs are equal. Since defendants with greater wealth attach higher prices to their time input, they would not only substitute more market intensive methods of producing R, but would also have a higher r.[19] Moreover, it follows from the equilibrium condition in (8) that a rise in r will lead to fewer inputs of R. In contrast, the increase in s as wealth rises will usually result in an increase in R.[20] Thus, to predict the net effect of an increase in wealth,

18. (11) may also be rewritten as

$$\frac{e}{rR/Wc} \gtrless -Wc\frac{U''(Wc)}{U'(Wc)},$$

where e is the elasticity of P with respect to R, rR/Wc is the share of R in the suspect's conviction wealth, and $-WcU''(Wc)/U'(Wc)$ is a measure of relative risk aversion. The value of the latter is often argued to hover around 1. (See Kenneth Arrow, Aspects of the Theory of Risk-Bearing, 33–37 (1965).) Thus, if rR/Wc were small, one would expect $dR/dS > 0$ for risk averse suspects.

19. If higher income or wealth defendants are more productive in their use of time to produce R, then the marginal product of time would be positively related to income. This would work to offset the substitution of market inputs for time as income and wealth rose. Further, r need not increase with wealth.

20. The condition under which $dR/ds > 0$ is identical to that for $dR/dS > 0$ in (11).

which increases both r and s, one would have to determine the relative magnitudes of these two offsetting forces. In addition, a change in wealth even if s and r were to remain constant may change the equilibrium input of R if tastes for risk depend on wealth. We analyze below the case of risk neutrality and in the mathematical appendix we consider nonneutral tastes for risk.

The total differential of (8) with respect to W and R, assuming risk neutrality, gives $dR/dW \gtrless 0$ according as

$$-P'E_s \gtrless E_r \frac{r}{s \cdot S}, \tag{12}$$

where $E_s = \partial s/\partial W(W/s)$ and $E_r = \partial r/\partial W(W/r)$. E_r will be <1 since the price of market inputs is unaffected and some substitution of market inputs for time takes place as wealth rises. E_s can be assumed equal to 1 because as a first approximation the per unit value of time in jail is proportional to wealth. The optimality condition for R (see (8)) becomes with risk neutrality $-P' = r/s \cdot S$, and, therefore, dR/dW will be positive when $E_s > E_r$.[21] Thus, the amount of resources invested in a trial would tend to rise and the probability of conviction would tend to fall with increases in wealth. Note that this result also implies a lower negotiated sentence for wealthier defendants.[22]

Suppose that the penalty for conviction is not a jail sentence but instead a money fine. E_s would equal zero with a fine since changes in the value of time do not alter the dollar value of a fine, and dR/dW would be negative. Therefore, the effect of wealth on R reverses when penalties

21. An identical result holds when R affects not only P but also S. With risk neutrality, the first-order condition for the optimal R becomes

$$-P'(s \cdot S) - P(s \cdot S') = r, \tag{i}$$

where $S' = \partial S/\partial R \leqslant 0$. The total differential of (i) with respect to W and R yields $dR/dW \gtreqless 0$ as

$$E_s[-P'(s \cdot S) - P(s \cdot S')] \gtreqless E_r r, \tag{ii}$$

which gives $dR/dW \gtreqless 0$ as

$$E_s \gtreqless E_r. \tag{iii}$$

22. An increase in R that is anticipated by the prosecutor would lower P^* and hence his minimum offer, So, while the reduction in P would raise the defendant's expected utility from a trial and lower the maximum sentence he would accept to settle. Other things the same, these forces should work to lower the negotiated sentence.

are in terms of money and not time for risk neutral defendants.[23] Once risk aversion or preference is introduced, the effect of changes in wealth on R cannot in general be specified unless one has explicit knowledge of additional parameters of the defendant's utility function. Nevertheless, one can presume that if the deviation from risk neutrality is small, the effects of wealth on R will follow the effects for risk neutrality.

II. SOME APPLICATIONS

THE BAIL SYSTEM

In the United States the typical procedure for bail is that shortly after the defendant's arrest a bond is set as security for his appearance at trial. If the defendant can post the amount of the bond through a deposit of cash or other assets, or have a professional bondsman do it for him, he is released until trial. The bondsman's fee runs about 10 per cent of the value of the bail bond. If the defendant does not meet the bail requirement, he remains in jail. The bond is generally forfeited should the released defendant fail to appear at trial.

Several implications of the bail system can be derived from our model.

1. Bail costs would be deducted from the defendant's endowment, W, so that both $U(W - s \cdot So - r\hat{R})$ and $E(U)$ in (9) would fall. For defendants released on bail there would be no obvious change in $\hat{\pi}$ (since equal dollar amounts are subtracted from $(W - s \cdot So - r\hat{R})$, Wc and Wn) and hence no reason to expect a change in their use of trials compared to settlements. Bail costs for defendants not released would equal the opportunity cost of their time in prison plus losses from restrictions on their consumption and freedom. These costs would be greater for a trial than a settlement because the delay in reaching trial generally exceeds the time taken to negotiate a settlement.[24] This in turn would lower $E(U)$

23. G. S. Becker, Crime and Punishment: An Economic Approach, this volume, presents a similar argument without presenting a proof. However, he argues that the incentive to use time to reduce the probability of a sentence is unrelated to earnings, and the incentive to use money to reduce the probability of a fine is also unrelated to earnings. These results would follow when in the former case R is produced solely by time and in the latter case R is produced solely by market inputs. However, once R is produced by both time and market inputs there is always an incentive to substitute market inputs for time as earnings rise.

24. Empirically, the time difference appears to be positive. For example, in the 89 United States district courts the median queues in 1967 were as follows: jury trial = 5.7 months; court trial = 3.9 months, and settlement (guilty plea) = 1.9 months. See, 1967 Ann. Rep. Admin. Off. of the United States Courts, 269–71, table D6.

relative to $U(W - s \cdot So - r\hat{R})$ in (9), raise π and make a settlement more likely for defendants not released on bail. Thus, given a positive time differential between a trial and settlement one would predict proportionately more settlements among defendants not released than those released on bail.[25]

2. The defendant in jail is restricted in his use of resources to reduce the probability of conviction. This can be interpreted as either raising the costs or lowering the marginal products of his market and time inputs. For example, in the case of market inputs, detention would hamper consultation with lawyers, and in the case of time inputs, the defendant would have greater (even prohibitive) difficulty in seeking out witnesses and in engaging in other investigatory activities. These factors increase the marginal cost of producing a given R and lower the defendant's input of R.[26] Thus, other things the same, the probability of conviction by trial should be greater for defendants not making bail than for those making bail.[27] As noted earlier, a higher probability of conviction by trial also leads to worse terms in a settlement. One should add that for these reasons the prosecutor always has an incentive to request the judge to set high bail charges.

3. Finally, if making bail is positively correlated with income, then

25. Two additional points should be noted. (a) If the defendant not released on bail were given credit toward his sentence for time in prison prior to disposition of his case, the only bail deduction in (9) would be from Wn, the defendant's endowment if he is not convicted. π would still rise. However, the rise in π would be negatively related to the probability of conviction. In the limit, if the probability of conviction equaled 1, court delay would leave π unchanged. (b) Bail costs of defendants released on bail will generally *not* be greater for a trial than a settlement. The bondsman's fee is independent of whether the defendant goes to trial or accepts a settlement, and a majority of felony defendants who make bail use bondsmen. (See, Lee Silverstein, Bail in the State Courts – A Field Study and Report, 50 Minn. L. Rev. 621, 647–52 (1966).) And the returns from assets (except cash) used as security for bail bonds will continue to be received by the owner.

26. Defendants not making bail may have available added resources for legal services that would have been used to finance bail. These can offset the higher costs of R so that the probability of conviction need not increase. However, it is also possible that their resources will decline should the loss in income (excluding a consumption allowance) exceed the cost of financing bail. In the latter case, capital market difficulties would presumably have prevented their release.

27. Critics of the bail system have recognized this point. For example, see Report of the Att'y Gen. Comm. on Poverty & the Admin. of Fed. Crim. Just. 74–76 (1963). Also note that the increase in cost of R for jailed defendants may be partly offset by the greater availability of "legal" advice from other inmates. However, if this factor were sufficiently important, one would observe defendants who were able to meet bail requirements accepting pretrial detention instead.

the effects of pretrial jailing, cited above, would fall most heavily on low-income defendants.

Proposals for bail reform generally focus on eliminating income as an indirect criterion of pretrial release. The Federal Bail Reform Act of 1966 requires that criminal defendants in federal courts (which cover a small minority of criminal defendants) be released prior to trial unless there is reason to believe they would flee. The "President's Commission" suggests placing greater reliance on release of defendants without bail, accompanied by certain restrictions on their behavior (for example, restrictions on travel, associations), while simultaneously confining suspects whose release would pose a significant threat to the community, regardless of their financial ability to make bail,[28] the latter provision being a form of "preventive detention." If these reforms were to result in the pretrial release of more defendants and more low-income ones we would predict the following: a decline in the negative correlations between income and the effects of pretrial jailing outlined above; a reduction in the fraction of defendants convicted since fewer defendants would be restricted in their use of R; and an increase in the demand for trials as differential bail costs between a trial and settlement go to zero for more defendants.[29] The latter would probably increase court delay.

These reforms leave persons detained in the same position as before and, moreover, their position relative to defendants released may worsen if the latter group does not pay for their release. Suppose those detained were paid a monetary compensation that increased with the length of their detention. We could then eliminate much of the discriminatory aspects of the bail system while still detaining persons believed to be dangerous. A higher marginal cost of R for detained suspects would still be present, but they would have additional resources to mitigate the adverse effects of this on the probability of conviction. Compensation would reduce the defendant's incentive for a settlement as the differential bail costs between a trial and settlement decline and approach zero for full compensation. If compensation were paid out of the prosecutor's budget, the latter's incentive for a settlement would increase given that the payment were greater for a trial than a settlement. This in turn would lower his minimum offer, So, and raise $U(W - So - r\hat{R})$ in (9). Hence, the incentive

28. U.S. Pres. Comm'n on Law Enforcement and Admin. of Justice, Task Force on the Admin. of Justice, Task Force Report: The Courts, at 38–40 (1967). [Hereinafter, The Courts.]

29. Other effects could be added. For instance, a predicted increase in crime from reducing the average probability of conviction, and a savings in resources used for pretrial detention.

for a settlement need not fall with compensation. Note as *So* falls this tends to reduce the positive difference between negotiated sentences for defendants not released compared to defendants released on bail.[30]

COURT DELAY

It is widely recognized that the courts are burdened with a larger volume of cases than they can efficiently handle. The results are often long delays prior to trial, and hasty considerations when cases reach trial.[31] This is not surprising since users pay a nominal money fee, if any, and a queue develops to ration the supply.

To understand the implications of nonmoney and money pricing on the demand for courts, assume initially there is a money price, M, paid by the loser that clears the market.[32] We also assume that the prosecutor's budget is not increased to cover these court costs. M affects both the prosecutor's and defendant's demand for trial. First, it reduces the minimum sentence offered by the prosecutor, *So* in (5), by a positive function of $(1 - P^*) \cdot M$. This, in turn, raises π in (9) and increases the likelihood of a settlement. Second, it lowers the defendant's wealth if convicted by trial, *Wc*, by an amount equal to M, reducing $E(U)$ and raising π in (9). This also increases the chance of a settlement. The larger is M. the greater the increase in π, and the more settlements that take place. Thus, a downward sloping demand curve for the courts is generated. Further, one can venture from the analysis of (9) that as M rises, the reduction in quantity demanded of trials (hereafter, trial demand) will be primarily from cases where there is not a significant disagreement between the prosecutor and defendant over the probability of conviction, and where the sentence if convicted by trial tends to be small.[33] Put differently, cases that still go to trial as M rises are where there are significant disagreements over the probability of conviction and where penalties are severe. Moreover, changing the allocation of the payment of M has little effect on the above results, since whether the loser or winner pays M, or both share M, a money price always increases π in (9) and reduces trial demand.

30. For a detailed analysis of alternative bail systems see my paper, The Bail System: An Economic Approach, this volume.

31. The Courts. *supra* note 28, at 80–90.

32. This does not mean that defendants are immediately brought to trial. Some time is required by both defendant and prosecutor to prepare a case for trial. Current delays are alleged to run considerably in excess of this.

33. Optimal values of P^*, P, R^* and R may change as M rises since the prosecutor and defendant must now allocate some resources to losses from expected court fees.

Compare this pricing scheme with one in which the courts are heavily subsidized, taking the extreme example of a zero money price. As M goes to zero, So and Wc rise, π falls and without an increase in supply, trial demand would exceed supply. Let us assume trial dates are allocated on the basis of waiting time since arraignment and a trial queue develops. The queue will reach an equilibrium size because, as we will show, trial demand is a decreasing function of waiting or queueing time. An increase in the queue imposes losses on the prosecutor as it (a) reduces the number of convictions in the current year from cases commenced in that year by delaying trial convictions,[34] and (b) ties up resources in a case for a longer period of time. These losses increase as the queue lengthens, inducing the prosecutor to offer a lower sentence in exchange for a guilty plea. Although $U(W - s \cdot So - r\hat{R})$ in (9) then rises (as So falls) the incentive for the defendant to settle as the queue grows will depend on whether or not he is released on bail. For defendants not released, the longer the queue the higher the bail costs of a trial and hence the lower their expected utility from a trial.[35] This factor, together with the response of the prosecutor, leads to the prediction that the demand for trials will fall as the queue lengthens for defendants not released on bail. On the other hand, for defendants released on bail the net effect on their expected utility of an increase in the queue is unclear. The discounted loss from a sentence received in a trial would diminish or increase as the penalty is pushed into the future, depending on whether earnings are rising at a slower or faster rate than the defendant's discount rate. In addition, the defendant's earnings may be adversely affected during the period he is free on bail due to his being under indictment. If on balance their expected utility falls or remains constant and the prosecutor's losses rise, one would expect an increase in π and a reduction in trial demand as the queue lengthens for defendants free on bail. However, one would predict that the demand for trials among defendants released would be less responsive to an increase in the queue than the demand among defendants not released, since the

---

34. Even if the prosecutor had no time preference with respect to convictions, an increase in the queue would still impose losses on him. For example, suppose the prosecutor is in office for 5 years. An increase in the queue during his tenure would lead to fewer convictions and a lower weighted conviction function than a constant queue because he will have left to his successor a greater stock of cases than his predecessor had left to him.

35. This would be partially offset by giving credit towards the eventual sentence for time spent in jail awaiting trial. Wc would be unchanged as the queue lengthened, providing the time spent in jail awaiting trial was less than S, but Wn would still fall. Hence, $E(U)$ would continue to fall as the queue increased.

cost of an increase in the queue is greater for the latter than the former group.

These points are illustrated in Figure 1, where Q_t = trial queue, Q_p = pretrial settlement queue, and T = fraction of trials per unit of time. D_1 and D_2 denote the trial demand curves for defendants not released and released on bail, respectively. Assume initially that there is no money charge for trials, the number of defendants not released on bail equals the number released, and credit against one's sentence is not given for pretrial detention. When $(Q_t - Q_p) = 0$, T would be the same for defendants released and not released, since the differential bail costs between a trial and settlement are zero for both groups. As $(Q_t - Q_p)$ rises, due to a reduction in supply of trial services, the differential bail costs rise by a greater amount for the not released than for the released defendants and hence the reduction in trials will tend to be greater for the former than the latter group. Thus, D_1 diverges from D_2 as $(Q_t - Q_p)$ increases. If the equilibrium queue initially equaled $\bar{Q}$, T would equal T_1 for defendants not released and T_2 for defendants released. Suppose a money charge for trials is established that is sufficient to reduce $(Q_t - Q_p)$ to zero, keeping the number of trials constant. As a first approximation, the demand curves for trials of the released and not released defendants would be identical because the differential bail costs between a trial and settlement are now

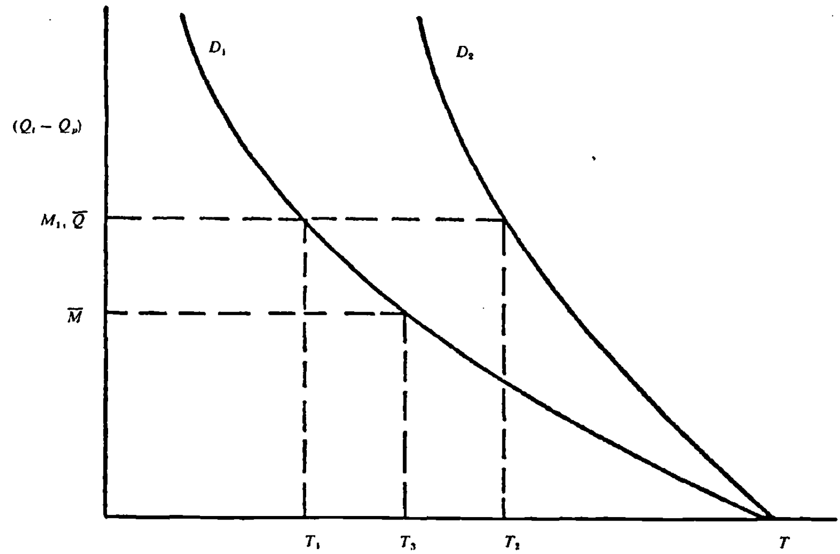

FIGURE 1

zero for both groups.[36] If the defendant's trial fee were set at M_1, a price that equals the maximum amount that defendants who are not released would pay at the margin for the same number of trials in order to reduce $(Q_t - Q_p)$ to zero, the aggregate number of trials demanded would be less than the available supply as $2T_1 < T_1 + T_2$.[37] In order for demand to be equated with supply, the defendant's court fee must be less than M. Let $\bar{M}$ in Figure 1 equal the market clearing court fee. At $\bar{M}$ the fraction of trials for each group would equal T_3 and by assumption, $2T_3 = T_1 + T_2$. Thus, a money charge for the courts that kept constant the number of trials can lead to an increase in the use of the courts on the part of defendants not released on bail, and a reduction in use among defendants released on bail. Moreover, if the supply curve of trials were positively sloped with respect to a money price, one would also expect an increase in the total number of trials.

In sum, we should note that although a zero money price is often advocated as a means of not discouraging low-income defendants from using the courts, its effect can be the opposite. A zero price operating with a bail system that tends to detain in jail low-income defendants will discourage the latter group from going to trial. In contrast, an appropriate money price may reduce the demand for the courts of defendants released on bail, and by reducing the trial queue can increase the use of the courts by defendants who do not make bail.[38] Surprisingly, the literature that criticizes court delay makes no mention of the possibility of charging a money price, which not only reduces delay, but can distribute the use of the courts more equally among defendants independent of their ability to make bail.

---

36. Note that the demand curves may differ. For example, if the average wealth of released defendants exceeds that of jailed defendants and the wealth elasticity of trial demand is not zero, then the demand curve of the former group will be to the right of D_1. However, as long as it is still to the left of D_2, the results that follow will still hold.

37. The prosecutor's court fee as $(Q_t - Q_p)$ falls to zero must be large enough to keep *So* constant. If other methods of allocating court fees (for example, winner or loser pays) were used, we could no longer assume that D_1 is the demand curve when trials are priced. Although the geometry would become more complicated when different pricing schemes are used, the results of the analysis would not be substantially altered.

38. An alternative scheme that would produce similar results is to continue a zero money price for the courts but allow defendants to buy and sell their places on the queue. This would presumably reduce the differential costs between a trial and settlement for defendants not released on bail relative to those released, and hence lead to a shift in court use from the latter to the former group. For example, if $\$X$ = the equilibrium price for a place in the queue that makes $(Q_t - Q_p) = 0$, the differential trial cost would be $\$X$ for both defendants released and not released, and their trial demands would be approximately equal.

III. EMPIRICAL ANALYSIS

In the legal area readily accessible and systematically collected data are quite limited. However, two sources of data were found that make it possible to test a number of the important hypotheses in the theoretical model. The first source is an American Bar Foundation (ABF) study, in which over 11,000 felony defendants in 1962 were sampled from state court dockets in nearly 200 counties.[39] From this sample we can estimate for several counties within most states the number of defendants released on bail and their average bail charge, the number going to trial, and the number dismissed, acquitted and sentenced. The second major source of data is for the 89 U.S. district courts where annually published statistics on civil and criminal cases are available.[40] These data contain information of civil and criminal court queues, the number of cases going to trial, the disposition of cases, and the number of criminal defendants receiving subsidized legal services. It should be added that most criminal defendants have their cases decided in state not in U.S. courts. In 1962 about 300,000 persons were charged with felonies in the state courts, while about 30,000 criminal defendants annually have their cases disposed in the U.S. district courts.[41]

THE DEMAND FOR TRIALS

The theoretical analysis suggests the following demand function for criminal trials:

$$T = f(B, Q_t, Q_p, S, D, U), \tag{13}$$

39. Lee Silverstein, Defense of the Poor in Criminal Cases in American State Courts, A Field Study and Report (2 v. 1965). Note that a felony is generally defined as any crime punishable by imprisonment of more than one year.

40. See various years of the Ann. Rep. Admin. Off. of the United States Courts and Fed. Offenders in the United States District Courts 1967.

41. Lee Silverstein, *supra* note 39, at 7-8 and Ann. Rep., Admin. Off. of the United States Courts, *supra* note 40. The types of offenses also differ in the state and U.S. courts. Offenses in the U.S. courts include forgery, counterfeiting, interstate transportation of stolen goods and vehicles, postal theft, and violation of immigration laws, liquor laws and other federal statutes, while it includes few cases of murder, assault, robbery, and other "violent" crimes. The latter types of offenses are concentrated in state courts. The one exception is the U.S. district court in the District of Columbia which handles all criminal offenses in the area. In the empirical analysis of the U.S. Courts I have excluded the District of Columbia in order to have comparable offenses across districts.

where T is the fraction of defendants going to trial, B is the fraction of defendants released on bail, Q_t and Q_p are the average trial and pretrial settlement queues respectively, S is the average sentence if convicted by trial, D is the average cost differential between a trial and settlement, and U is the combined effect of all other factors.[42] We would predict on the basis of our model that B, Q_p, and S will have positive effects on T, while Q_t and D will have negative effects on T. Unfortunately, data limitations prevent us from estimating the partial effects of these variables in a single equation. The ABF sample of state county courts has no data on queues or cost differentials between trials and settlements, while the data for the U.S. courts contain no information on bail. Therefore, the analysis will use the ABF data to test bail effects, and the U.S. data to test queue effects. At the same time we will point out possible biases and alternative interpretations of the results that arise from leaving out either the bail or queue variables.

STATE COUNTY COURTS

Least-squares multiple regression equations were estimated across state county courts in 1962. These equations were of the following general form:

$$T = \alpha + \beta_1 B + \beta_2 S + \beta_3 Pop + \beta_4 Re + \beta_5 NW + \beta_6 Ur + \beta_7 Y + u. \quad (14)$$

The variables in (14) are defined as follows:

T: the fraction of defendants in a county court whose cases were disposed of by trial in 1962. Cases where a plea of guilty was made at time of trial are not counted as trials.

B: the fraction of defendants in each county released on bail in 1962.

S: the average time served of first-released prisoners in 1964 who had sentences of one year or longer. S is an estimate of the average sentence, if convicted by trial, of felony defendants in 1962. Releases in 1964 are used because the average time served in state prisons of first-released prisoners was about two years, and hence 1964 should be the average release year for defendants sentenced in 1962.

42. U would include factors derived from the π function (equation (9)) such as the distribution of estimates of the probability of conviction by trial, and attitudes toward risk. I have not been able to directly measure these variables and hence they are largely ignored in the empirical analysis. U also includes several demographic variables that will be specified in the statistical estimation of (13).

Pop: county population in 1960.

Re: region dummy variable that equals 1 for counties in South and 0 for non-South counties.

NW: per cent nonwhite population in county in 1960.

Ur: per cent urban population in county in 1960.

Y: median family income in county in 1959.

Weighted regressions on T are presented in Table 2 for counties in the U.S., the non-South and South.[43] In the U.S. and non-South equations (2.1–2.2) the regression coefficients of the bail variable have the predicted positive sign and are always highly significant, while in the South the coefficient is not significantly different from zero. Before discussing these results in greater detail, an interesting interpretation can be given to the bail regression coefficient. T can be written as

$$T = \frac{\lambda_1 N_1 + \lambda_2 N_2}{N}, \tag{15}$$

where N is the number of defendants, N_1 is the number not released on bail, and N_2 is the number released. λ_1 and λ_2 are the average propensities to go to trial of the not released and released group respectively. The theory predicts that $\lambda_2 > \lambda_1$, providing that the trial queue is longer than the settlement queue. Since $N_1/N + N_2/N = 1$ and $N_2/N = B$, (15) can be rewritten as

$$T = \lambda_1 + (\lambda_2 - \lambda_1)B. \tag{16}$$

Therefore, from a set of observations on T and B, the intercept in a simple regression of T on B would be an estimate of λ_1 and the beta coefficient on B would be an estimate of $(\lambda_2 - \lambda_1)$. A positive beta coefficient would be consistent with the prediction that $\lambda_2 > \lambda_1$. The interpretations of these regression coefficients are modified with the addition of other independent variables, X_i, which enter the regression indirectly through their influence on λ_1 and λ_2. For example, let

$$\lambda_1 = c_1 + \sum_{i=2}^{j} \beta_i X_i \tag{17}$$

43. Observations were weighted by the $\sqrt{n}$ where n is the number of defendants sampled in each county. The range of n is from 3 to 349 with a mean of 58, and n generally rises with the size of the county population. Weighted regressions were computed because of the likelihood of larger variances in the residuals as n declined. However, unweighted regressions were also computed, and as it turned out, the weighting made little difference in the results.

TABLE 2

WEIGHTED REGRESSIONS AND *t*-VALUES FOR CRIMINAL TRIALS IN 1962, STATE COUNTY COURTS[a]

Equation Number	Area	Counties	Dependent Variable	Regression Coefficients and *t*-Values[b]								R^2[c]
				α	*B*	*S*	*Pop*	*Re*	*NW*	*Ur*	*Y*	
2.1	U.S.	132	*T*	−.065	.348	−.0002	.037	.098	.014	.062	−.002	.26
				(.624)	(4.071)	(.087)	(3.312)	(2.723)	(.106)	(.709)	(.101)	
2.2	N.S.	100	*T*	.002	.498	.003	.038		−.217	.129	−.043	.45
				(.027)	(6.521)	(1.515)	(4.263)		(1.548)	(1.513)	(2.552)	
2.3	South	32	*T*	.229	−.012	−.009	.463		.109	−.570	.095	.48
				(.833)	(.059)	(1.746)	(3.215)		(.488)	(2.716)	(2.042)	

SOURCES.—*T* and *B* from Lee Silverstein, Defense of the Poor in Criminal Cases in American State Courts, A Field Study and Report (2 v. 1965); *Pop*, *NW*, *Ur* and *Y* from U.S. Bureau of the Census, County and City Data Book 1962; *S* from U.S. Bureau of Prisons, National Prisoner Statistics Detailed Reports: State Admissions and Releases, 1964, table R-5.

[a] Although the ABF sample covered nearly 200 counties, many had to be excluded because there was no reporting on the number of defendants who made bail. I have no reason to believe that the group of excluded counties would have differed systematically from the counties included in the regression equation. Two counties in New Jersey were excluded because no data on *S* were available for New Jersey.

[b] *t*-values in parentheses.

[c] All R^2's are unadjusted in Table 2 and all other tables unless explicitly stated to the contrary.

and

$$\lambda_2 = c_2 + \sum_{i=2}^{j} \alpha_i X_i. \tag{18}$$

By substitution into (16) we have,

$$T = c_1 + (c_2 - c_1)B + \sum_{i=2}^{j} \beta_i X_i + \sum_{i=2}^{j} (\alpha_i - \beta_i)B \cdot X_i. \tag{19}$$

Estimates of equation (19) were not successful because of the large amount of multicollinearity resulting from the inclusion of interaction variables. This tended to eliminate statistical significance from any of the independent variables. However, if we set $\alpha_i = \beta_i$ for all $_i = 2, \text{———}, j$, the interaction variables drop out and (19) reduces to the form of equations estimated in Table 2. It also follows from (17) and (18) that when $\alpha_i = \beta_i$ the regression coefficient on B is a measure of $(\lambda_2 - \lambda_1)$, the difference in trial propensities between defendants released and not released on bail.

Estimates of $(\lambda_2 - \lambda_1)$ from Table 2 for the U.S. and non-South imply, for example, that the release of an additional 20 defendants on bail, other things the same, would lead to a desired increase of about 7 to 10 trials as a result of the reduction in trial costs associated with making bail. One can also get a rough idea of the increased demand for trial if the existing bail system were replaced with a system of preventive detention that released all defendants except a few "hard-core" criminal suspects (for example, 10 per cent). The weighted means of T and B are about .18 and .45 respectively. Therefore, the release of additional defendants to bring the number released to 90 per cent would lead to an increase in the fraction desiring trials from 18 per cent to between 34 and 40 per cent, or roughly a 100 per cent increase in desired trials.[44]

Although no direct measures of trial queues are available in the ABF data, longer trial queues are generally thought to exist in large urban areas. If the county population variable is interpreted as an imperfect proxy for the difference between trial queues and settlement queues, the sign of the regression coefficient on the population variable would depend on the relative strength of two opposing forces. On the one hand, longer queues discourage trials, but on the other hand, longer queues may

44. This is the increase in trials desired with no change in the trial queue. With an increased demand and unchanged court capacity the queue would presumably grow so that the actual increase in T would be less than the desired increase. In fact, if the courts were fully employed, the queue would grow until the costs of waiting were just sufficient to make desired trials equal to the previous level of trials.

be the result of an increased demand for trials. In Table 2, the coefficient on the population variable is positive and significant in all regressions, which suggests that the positive association of trials with queues dominates.

Further evidence on the effects of population size appears in Table 3 where separate regressions are given for counties in the non-South with populations greater than 450,000, between 100,000 and 450,000, and less than 100,000, and in the South with populations greater than 200,000 and less than 108,000. In Table 3 not only does the bail variable continue to have a positive effect on trials in all non-South equations, but its coefficient (or $(\lambda_2 - \lambda_1)$) has a systematically greater value as county size rises (.09 in eq. 3.3, .31 in eq. 3.2, and .75 in eq. 3.1), which is precisely what one would expect if $(Q_t - Q_p)$ was positively correlated with county population.[45] $(\lambda_2 - \lambda_1)$ is statistically significant in the non-South except in counties of less than 100,000. This result could be observed if the difference between $(Q_t - Q_p)$ was negligible in small counties. Moreover, the empirical finding that the coefficient on the bail variable increases as county population size rises is indirect evidence that $(Q_t - Q_p)$ is in fact larger in counties with bigger populations.

Let us briefly consider the empirical results for the South. The three regression coefficients on the bail variable in the South in Tables 2 and 3, were negative and not significantly different than zero. One possible explanation is that $(Q_t - Q_p)$ is negligible for counties sampled in the South so that $(\lambda_2 - \lambda_1)$ would approach zero, and hence the regression coefficients on bail would not be significant.[46] A second explanation is that greater measurement errors in the bail variable may exist in the South compared to the non-South, which would lower the value of regression coefficients on bail in the South relative to the non-South. Along these lines it might be argued that justice is more informally administered in the South, particularly in rural areas, and this would produce poorer records on bail. (A similar argument may be used to rationalize the nonsignificant but positive bail coefficient in non-Southern counties of less than 100,000.) However, it is questionable whether this argument should be given much weight since a nonsignificant bail variable was also ob-

---

45. If we refer to Figure 1, *supra*, we note that at a given value of $(Q_t - Q_p)$ the difference getween D_2 and D_1 equals $\lambda_2 - \lambda_1$ and as $(Q_t - Q_p)$ increases, $\lambda_2 - \lambda_1$ increases due to the increase in bail costs of defendants not released compared to defendants released.

46. Data for the federal courts indicate that queues are somewhat lower in the South. In 1966, the mean civil Q_t's were 22.0 and 15.7 months for the non-South and South, and the mean criminal Q_t's were 6.3 and 5.3 months in the two areas. However, $(Q_t - Q_p)$ for criminal cases was 3.8 and 3.5 months in the non-South and South.

TABLE 3

WEIGHTED REGRESSIONS AND *t*-VALUES FOR CRIMINAL TRIALS IN 1962 BY COUNTY POPULATIONS, STATE COUNTY COURTS

Equation Number	Area	Counties	Dependent Variable	Regression Coefficients and *t*-Values							R^2
				α	*B*	*S*	*Pop*	*NW*	*Ur*	*Y*	
3.1	Non-South >450,000	30	*T*	.312 (1.011)	.750 (4.589)	.004 (.923)	.045 (2.864)	−.479 (1.778)	−.063 (.238)	−.081 (2.468)	.37
3.2	Non-South 100–450,000	27	*T*	−.041 (.200)	.309 (2.619)	.001 (.223)	.101 (.493)	.114 (.157)	−.120 (.616)	.014 (.430)	.31
3.3	Non-South <100,000	43	*T*	.151 (1.777)	.085 (1.126)	−.001 (.895)	.044 (.084)	.010 (.058)	.099 (1.556)	−.026 (1.562)	.17
3.4	South >200,000	17	*T*	−.194 (.417)	−.010 (.036)	−.016 (2.667)	.524 (2.674)	.590 (2.249)	−.642 (1.394)	.188 (3.633)	.81
3.5	South <108,000	15	*T*	.255 (.783)	−.186 (.775)	.0001 (.013)	−4.269 (2.402)	−.252 (.974)	−.067 (.342)	.088 (1.200)	.59

SOURCE.—Table 2, *supra*.

served for the South in counties with populations greater than 200,000.

Although a sufficient explanation is not available for the South, the overall results of Tables 2 and 3 give strong support to the hypothesis that the frequency of trials is greater among defendants released on bail than those not released. A positive and statistically significant relationship between bail and trials was observed for the U.S. and the non-South, and the latter region included more than three-fourths of our observations. This finding is consistent with the prediction of the theoretical model that the costs of going to trial compared to settling are increased by not making bail, which in turn reduces the likelihood of a trial.

Several additional comments on the results in Tables 2 and 3 are in order. (1) It might be argued that the bail variable is a proxy for wealth so that a finding that T increases with B is due to differences in wealth and not to greater trial costs for those not released. First, there is nothing in the theoretical analysis that indicates that wealth directly affects the choice between a trial and a settlement. If wealth were positively correlated with the ability to make bail, one would observe wealthier defendants going to trial, but the theoretical explanation lies with differences in costs not differences in wealth. Second, the empirical analysis of the U.S. courts in the next section contains indirect estimates of a defendant's wealth which show that increases in wealth have no observable positive effect on trials. (2) A second criticism, which if valid would weaken my conclusion that increases in B lead to increases in T, is that spurious correlation exists between B and T. The argument can be made that defendants planning to plead guilty will not be willing to incur the costs of making bail, while those planning to go to trial will incur these costs. If this were true, an increase in T would lead to an observed increase in B and not the reverse. Although this argument has some plausibility, it has a defect. A defendant planning a guilty plea presumably desires the most favorable terms in a settlement. We showed in the model that one effect of not making bail is to raise the probability of conviction in a trial, which in turn results in worse settlement terms. Therefore, it is not obvious that the defendant planning to settle will find it any less desirable to post bail than the defendant planning a trial, since both suffer losses from not being released on bail. (3) The regression coefficients on the sentence variable, S, do not support the hypothesis that the likelihood of a trial is greater for defendants accused of crimes carrying longer sentences. In 4 of 8 equations in Tables 2 and 3 the sign of S was positive, and in only one equation (3.4), where S had a negative effect, was the variable significant. The inconclusive behavior of S may partly be attributed to the data. The theoretical analysis calls for a variable that measures the average severity of offenses for defendants sampled in each county in 1962, while data limitations have forced us to use the average time served by all felons in a state who were first released in 1964. Fortunately, the regressions for the U.S. courts provide us with a stronger test of the sentence hypothesis because data on sentences in the U.S. courts correspond more closely to the theoretical requirements. (4) The NW, Ur and Y variables, which are included in the regressions of Tables 2 and 3, should be viewed as demographic characteristics of counties rather than as indicators of socioeconomic classes of defendants, since the relation of these variables to defendants may be remote. There are no prior expectations on the effects of these variables on trials, and their regression coefficients do not show a consistent pattern. In Table 2 Y and NW have negative effects on T in the non-South and positive effects in the South, while Ur has a positive effect in the non-South and a negative effect in the South. About two-thirds of the NW, Ur and Y regression coefficients are not statistically significant.[47]

U.S. COURTS

Least-squares regression equations were estimated across U.S. district courts in 1967 of the following form:

$$T_1 = \alpha + \beta_1 Q_t + \beta_2 Q_p + \beta_3 S + \beta_4 D + \beta_5 Re + \mu. \qquad (20)$$

All variables are in natural logs except Re and hence the regression coefficients are elasticity measures. The variables in (20) are defined as follows:

T_1: ratio of defendants whose cases were disposed of by trials during 1967 to the total number of defendants disposed of in 1967. Regressions were also fitted on T_2 which equals the ratio of trials to defendants for 1968.

Q_t: weighted average of median time intervals from filing to disposition by court trial and by jury trial in 1967, where weights are the proportion of court and jury trials respectively.

Q_p: median time interval from filing to disposition by a plea of guilty in 1967.

S: weighted average of sentences received by convicted defendants whose

47. Regressions were also estimated without the NW, Ur and Y variables. The regression results for the bail, population, and sentence variables were largely unaffected.

cases were disposed of in 1967, where weights are the proportion of convicted defendants receiving each type of sentence.[48]

D: proportion of criminal defendants disposed of in 1967 who are assigned counsel by the court under provisions of the Federal Criminal Justice Act of 1964. The Act provides for counsel when defendants are unable to pay all or part of their legal fees.[49] Thus, D is a direct measure of the fraction of defendants with subsidized legal counsel. Since the ability to pay for counsel is related to the defendant's wealth, D would also serve as a rough measure of the fraction of defendants with low incomes or wealth.

Re: region dummy variable where 1 is assigned to district courts in the South and 0 to district courts in the non-South.

Data on Q_t are available for only 44 of 89 district courts in 1967 (see note b to Table 4), while data on Q_t and Q_p are not available for other years. Regressions were first estimated for the 44 districts in 1967. However, in order to incorporate observations for the remaining districts, and to work with years other than 1967, a proxy variable for Q_t was used that can be computed for all years and districts. The proxy for Q_t in year m is the ratio of pending cases (Pc) at the end of year m-1 to the average annual number of cases that go to trial ($\bar{T}$) in years m and m-1. One would expect Pc to estimate the backlog and $\bar{T}$ to roughly measure the availability of trial services, and hence $Pc/\bar{T}$ should serve as a measure of Q_t even though not all pending cases eventually go to trial.[50] The accuracy of

48. Obvious problems arise in evaluating a diversity of sentences that include imprisonment, fines, probation with and without supervision, suspended sentence, etc. The Administrative Office of the U.S. Courts has devised a common set of values for these sentences (see Fed. Offenders, *supra* note 40, at 4) that assign 0 to suspended sentences and probation without supervision, 1–4 for fines and various terms of probation with supervision, and 3–50 for imprisonment with sentences that range from 1 to more than 120 months. Although higher values are generally given to more severe sentences, the method is still arbitrary. For example, why all fines and probation with supervision from 1 to 12 months are both assigned the value 1 is never explained. Nevertheless, the use of this variable as an estimate of the average potential sentences of accused defendants in each district seems preferable to using just the mean prison sentence, since the latter group includes only 38 per cent of all defendants disposed of in 1967, while the former group includes 77 per cent. Both measures suffer because they exclude defendants not convicted when the relevant theoretical variable is the average potential sentence faced by all defendants before disposition of their case.

49. See The Courts, *supra* note 28, at 59–61.

50. There were 10,771 pending criminal cases in the beginning of fiscal year 1967, and 3,924 trials in the 1967 fiscal year. Since the average trial queue is about six months, this would suggest as a first approximation that roughly one-half of the trials in 1967 were from pending cases in the beginning of the year. Thus, about 20 per cent of pending cases would go to trial.

TABLE 4

WEIGHTED REGRESSIONS[a] AND t-VALUES FOR CRIMINAL TRIALS IN U.S. DISTRICT COURTS, 1960, 1967 AND 1968

Equation Number	Year	Districts	Dependent Variable	Regression Coefficients and t-Values								R^2
				α	Q_t [b]	Q_p	$Pc/\bar{T}$	S	D	Re	$E(T)$ [c]	
4.1	1967	44	T_1	−2.814 (4.853)	−.246 (1.666)	.629 (4.201)		.610 (2.255)	.392 (2.529)	.159 (1.249)		.61
4.2	1967	44	T_1	−2.646 (6.321)		.608 (6.027)	−.407 (5.698)	.413 (1.976)	.151 (1.199)	−.039 (.379)		.78
4.3	1968	44	T_2	−1.999 (3.533)	−2.50 (1.803)	.591 (3.788)		.296 (1.116)	.461 (3.166)	.095 (.822)		.57
4.4	1968	44	T_2	−1.880 (3.920)		.532 (4.494)	−.311 (3.849)	.143 (.598)	.349 (2.627)	−.047 (.430)		.67
4.5	1967	89	T_1	−2.750 (6.894)		.578 (5.882)	−.463 (6.573)	.495 (2.518)	.194 (1.620)	−.087 (.856)		.66
4.6 [d]	1960	86	T	−.365 (.323)			−.410 (5.142)	.315 (1.246)		.037 (.293)	1.287 (3.954)	.42
4.7 [e]	1960	43	T	1.294 (.996)			−.208 (2.338)	.242 (.805)		.054 (.372)	1.958 (5.751)	.58

[See following page for Notes.]

NOTES TO TABLE 4

SOURCES.—T_1, Q_t, Q_p, $Pc/\bar{T}$ for equations 4.1 through 4.5 are from 1967 Ann. Rep., Admin. Office of the United States Courts, tables D6, D1, C7; T_2 for equations 4.1 through 4.5 is from 1968 Ann. Rep., Admin. Office of the United States Courts, table D6; S and W for equations 4.1 through 4.5 are from Fed. Offenders in the United States District Courts 1967, tables D10, D11; T, $Pc/\bar{T}$, $E(T)$ in equations 4.6 and 4.7 are from 1960 Ann. Rep. Admin. Office of the United States Courts tables C7, D1, D3, D4; S for equations 4.6 and 4.7 is from U.S. Bureau of Prisons, Federal Prisons 1960, table 20.

[a] Weighted by $\sqrt{n}$ where n is the number of defendants disposed (equations 4.1 through 4.5) and the number of cases commenced (equations 4.6 and 4.7). All variables except Re are in natural logarithms.

[b] Q_t is the district average of the median court trial queue and median jury trial queue. Data on either median were available only if there were at least 25 observations for that type of trial. 44 out of 89 districts had figures on at least one of two trial queues. Since most trials are jury trials (about 67 per cent), 29 out of the above 44 districts did not publish any information on court trial queues. The latter were estimated by assuming the ratio of the court trial to jury trial queue in the circuit (the 89 district courts are divided into ten circuits) was equal to the ratio in the district. Information was available on the aggregated circuit level, and hence the median court trial queue in a district could be directly estimated. Estimates were generally required in districts that had a small proportion of court relative to jury trials. Therefore, any errors in estimating the court trial queue would have a small effect on the weighted average Q_t. Finally, note that in 5 districts the queue for court trials but not jury trials was available. The procedure described above was then used to estimate the jury trial queue.

[c] $E(T)$ in the ith district equals $\sum_{j=1}^{16} T_j^* O_{ji}$ where T_j^* is the proportion of defendants in the jth offense category whose cases were disposed of by trial for all districts taken together in 1960, and O_{ji} is the proportion of defendants accused of the jth offense in district i. There were 16 offense categories. Thus, variations in $E(T)$ across districts are due solely to differences in the composition of offenses. $E(T)$ was devised to take account of the possibility that differences in the fraction of trials across districts were the result of $E(T)$ rather than the queue. Data did not permit a similar calculation for 1967–68.

[d] In 1960 there were 86 district courts. By 1967 there were 89 district courts as several were eliminated and new ones were added. There are several small differences between the 1960 and 1967 data. They are: (1) T is the ratio of the number of cases that went to trial over the number of cases commenced in 1960. This differs from 1967 where the trial data are for defendants not cases, and the denominator is disposed defendants not commenced cases. In a given year the number of new defendants is about 25 per cent greater than the number of cases commenced, indicating that the number of cases with more than one defendant exceed the number of defendants involved in more than one case. Since this is reflected in the numerator of the trial variable as well, the correlation in a given year between T for cases and defendants would be very high; (2) S in 1960 is the average prison sentence of convicted defendants, and excludes defendants who were fined or put on probation with supervision. The latter groups are included in the 1967 S variable.

[e] Regressions computed from districts in 1960 that match those districts in 1967 that had data on Q_t.

$Pc/\bar{T}$ as an estimate of trial queues was checked by running simple regressions of Q_t, Q_p, and $(Q_t - Q_p)$ on $Pc/\bar{T}$.[51] These equations indicate that $Pc/\bar{T}$ is positively and significantly related to Q_t and $(Q_t - Q_p)$, accounting for nearly half the variation in these variables. Although $Pc/\bar{T}$ is also positively related to Q_p, it is substantially more important in explaining variations in Q_t and $(Q_t - Q_p)$. Therefore, $Pc/\bar{T}$ is not merely a measure of general delay in the disposition of criminal cases, but, on the contrary, is a measure of differential delay between trials and guilty pleas. This result allows us to estimate regressions for all 89 district courts in 1967, and to check the stability of the model over time by fitting equations to an earlier year, 1960, in which direct data on queues were absent.

If equation (20) estimates a demand curve for trials, the theoretical analysis predicts that the regression coefficient on Q_t (and $Pc/\bar{T}$) will be negative, and the regression coefficient on Q_p will be positive. However, single-equation estimates may identify a supply curve instead, if the demand for trials varied more than the supply across districts. In the latter case, higher observed values for Q_t would have resulted from shifts to the right in demand curves, giving rise to a positive coefficient on Q_t. Similar behavior would produce a negative coefficient on Q_p if a reduction in guilty pleas lowered Q_p. I have attempted to deal with this identification problem in two ways: (1) Equation (20) includes S and D variables that are expected to lead to shifts in the demand for trials. By holding S and D constant, the likelihood of identifying a demand curve is increased. A region variable, Re, also enters equation (20), but it is not obvious that Re operates more on the demand than supply side of trials. (2) Regressions have been estimated with a 1968 trial variable, T_2, against 1967 values of Q_t and Q_p. If defendants and prosecutors form their expectations about current queues on the basis of last year's queue, then Q_t in 1967 could still be inversely related to T_2 even though demand shifts in 1967 had caused a positive correlation between T_1 and Q_t.

---

51. The regression equations for the 44 districts in 1967 were

$$Q_t = \underset{(14.961)}{1.272} + \underset{(5.689)}{.420}(Pc/\bar{T}) \qquad R^2 = .44$$

$$Q_p = \underset{(4.087)}{.489} + \underset{(1.603)}{.167}(Pc/\bar{T}) \qquad R^2 = .06$$

$$(Q_t - Q_p) = \underset{(4.155)}{.516} + \underset{(5.844)}{.629}(Pc/\bar{T}) \qquad R^2 = .45$$

$Pc/\bar{T}$. Q_t, Q_p, and $(Q_t - Q_p)$ are in natural logs, and all observations are weighted by $\sqrt{n}$ where n is the number of defendants disposed of in each district in 1967.

Regression estimates of equation (20) are presented in Table 4 for the years 1960, 1967, and 1968. In districts where Q_t and Q_p are available, the regression results strongly support the hypothesis that increases in Q_t, holding Q_p constant, have significant negative effects on T_1 and T_2, and increases in Q_p, with Q_t constant, have significant positive effects on T_1 and T_2. When $Pc/\bar{T}$ is substituted for Q_t, in equations 4.2 and 4.4, $Pc/\bar{T}$ has the predicted negative sign and is statistically significant, while Q_p has the same effects on T_1 and T_2 as before.[52] Further, when the sample is expanded to include all 89 districts in equation 4.5, the signs and significance of Q_p and $Pc/\bar{T}$ are similar to the results for the 44 districts. This suggests that any biases in estimating the effects of queues on trial demand due to excluding 45 districts in equations 4.1–4.4 are probably of small magnitude. Estimation of regressions for 1960 indicates that for all districts (equation 4.6) the queue, as measured by $Pc/\bar{T}$, had about the same effects as in 1967 and 1968. However, for districts in 1960 that match the districts in which Q_t were available in 1967 (equation 4.7), the regression coefficient of $Pc/\bar{T}$ was still negative but with a smaller absolute value. The latter partially results from the absence in 1960 of a measure of Q_p. Since Q_p and $Pc/\bar{T}$ are positively correlated. part of the positive effect of Q_p on trials would be picked up by $Pc/\bar{T}$ which, in turn, would diminish the negative effect of $Pc/\bar{T}$ on trials. I have tested this for 1967 by reestimating equation 4.2 without Q_p, which reduces the regression coefficient of $Pc/\bar{T}$ from –.407 to –.331.

Although the regressions in Table 4 are consistent with the hypothesis on Q_t and Q_p, these results contain an interesting puzzle. In both equations 4.1 and 4.3, trials are substantially more responsive to changes in Q_p than Q_t. One possible explanation is that errors in measurement are more important in Q_t than Q_p. Q_t is based on a sample of defendants that in each district averages less than 25 per cent of the sample size of Q_p, and in addition, Q_t often had to be estimated because either data on the jury trial queue or the court trial queue were absent (see note b in Table 4).[53]

52. The significance of Q_p improves with this substitution, and the R^2's rise. The former is due to the substantially higher correlation between Q_p and Q_t than between Q_p and $Pc/\bar{T}$ (.54 compared to .24), while the latter is related to some spurious negative correlation since trials are present in the denominator of $Pc/\bar{T}$ and the numerator of T_1 and T_2. This spurious correlation probably explains why the absolute value of the regression coefficient of $Pc/\bar{T}$ is larger than Q_t.

53. Errors in measurement of Q_t would also bias downward the regression coefficient of Q_p since the regression coefficient of Q_t and the partial correlation between Q_t and Q_p are of opposite signs. See G. C. Chow, Demand for Automobiles in the United States, A Study in Consumer Durables app. I (13 Contributions to Economic Analysis 1957).

The effects on trial demand of the remaining variables in Table 4 may be summarized as follows: (1) S has a positive sign in all regressions as predicted by the theoretical analysis, and is statistically significant in the 1967 equations. The lack of significance in equations 4.3 and 4.4 is probably due to the fact that T_2 denotes defendants going to trial in 1968, whereas S refers to defendants sentenced in 1967. The nonsignificance of S in 1960 reflects the less comprehensive measure of S in that year. S is the average prison sentence in 1960, while in 1967, S includes defendants who were fined, placed on probation, and sentenced to prison. (2) D measures the fraction of defendants with subsidized legal counsel. These subsidies reduce the cost of a trial relative to a settlement, providing unsubsidized legal fees are greater for the former than the latter, and this in turn increases the demand for trials. The results of Table 4 support this hypothesis. The coefficient on D is positive in all regressions and significant in 4 out of 5 equations. This finding is relevant to the previous analysis of state courts where it was shown that defendants making bail had higher trial propensities. The latter was explained in terms of cost differentials between a trial and settlement that were greater for defendants not making bail. However, an alternative explanation was that wealthy persons were more likely to go to trial, and hence the observed relation between bail and trials resulted from the positive correlation between wealth and the ability to make bail. The analysis of the U.S. courts does not support this view. If differences in wealth per se were an important determinant of trial demand and wealthier defendants were more likely to go to trial, then the coefficient on D would have had a negative sign, since D should be inversely related to the fraction of wealthy defendants in a district. Thus, the results in Table 4, together with the findings for state courts, indicate that the cost differential between trials and settlements, and not differences in wealth among defendants, is an important factor in trial demand.[54] (3) The South dummy variable, Re, had no systematic effect on trials in Table 3, which contrasts with the strong positive effect of Re in the state data. This is not surprising, since the region effect in the state courts may have picked up the effect of lower trial queues in the South, whereas queues were held constant in the U.S. courts.

54. If one still believed that wealth was an important variable in trial demand, then the observed positive coefficients on D would show that wealthier defendants were *less* likely to go to trial. This contradicts the results of the state data which showed that wealthier defendants were *more* likely to go to trial. One final note is that if wealth were a determinant of the ability to make bail in the U.S. courts, then the coefficient on D could have a negative sign. However, the Bail Reform Act (to the extent it is effective) would have reduced or eliminated the correlation between wealth and the ability to make bail.

THE PROBABILITY OF CONVICTION

STATE COUNTY COURTS

The theoretical analysis predicted that if the defendant were not released on bail, the costs of his resource inputs would rise, leading to a reduction in these inputs and an increase in the probability of conviction. Therefore, a decline in the fraction of defendants making bail should result in an increase in the fraction of defendants convicted. A major difficulty in testing this hypothesis relates to the direction of causation between the bail and conviction variables. At the time bail is set a prima facie case is often made against the accused. If the preliminary evidence points to his guilt, a higher bail bond is likely to be set, which would lower his chance of being released. Hence, a selection process would take place before the final disposition of cases whereby defendants with a higher probability of conviction would be less likely to make bail. I have attempted to deal with this problem by including as independent variables both the fraction of defendants released on bail (B) and the average money bail charge (C) in regressions on the fraction of defendants convicted. Since setting a high money bail is a method of detaining a defendant with a high initial probability of conviction, then including C as an independent variable has the effect of holding constant differences across counties in these probabilities.[55] This in turn would remove from the regression coefficient on B any negative correlation due to higher conviction probabilities reducing the fraction of defendants released on bail.

Weighted regression equations of the form

$$P = \alpha + \beta_1 B + \beta_2 Pop + \beta_3 R_e + \beta_4 NW + \beta_5 Ur + \beta_6 Y + \beta_7 C + \beta_8 T + u \quad (21)$$

are presented in Table 5. B, Pop, Re, NW, Ur, Y and T are defined as before (see pp. 184–85) and P and C are defined as follows:

P: the fraction of felony defendants sentenced to prison in each county. Some convicted felony defendants received only fines so that P understates the total

55. The inclusion of C is only an approximation to holding constant variations in the probabilities because C may reflect other factors as well. For example, the severity of the offense, variations in the fraction of defendants not appearing for trial, attitudes of judges, etc.

TABLE 5

WEIGHTED REGRESSIONS AND *t*-VALUES FOR CRIMINAL CONVICTIONS IN 1962, 70 COUNTY COURTS IN U.S.

Equation Number	Dependent Variable	Regression Coefficients and *t*-Values									R^2
		α	*B*	*Pop*	*Re*	*NW*	*Ur*	*Y*	*C*	*T*	
5.1	*P*	.683	−.438	.010	.063	−.127	.076	−.006			.23
		(4.771)	(3.821)	(.748)	(1.141)	(.809)	(.581)	(.235)			
5.2	*P*	.655	−.369	.007	.062	−.142	.105	−.019	.015		.25
		(4.545)	(2.917)	(.544)	(1.129)	(.909)	(.800)	(.663)	(1.280)		
5.3	*P*	.676	−.247	.020	.051	−.090	.056	−.017	.011	−.289	.30
		(4.808)	(1.822)	(1.447)	(.947)	(.585)	(.431)	(.632)	(.989)	(2.130)	
5.4	*Ds* + *A*	.076	.227	−.019	.033	.010	.103	−.031	.015	−.299	.30
		(.641)	(1.978)	(1.566)	(.719)	(.074)	(.940)	(1.330)	(1.520)	(2.600)	
5.5	*Ds*	.115	.150	−.012	.063	−.042	.101	−.031	.013	.011	.19
		(1.044)	(1.408)	(1.114)	(1.482)	(.348)	(.994)	(1.440)	(1.442)	(.101)	
5.6	*A*	−.039	.077	−.006	−.030	.052	.002	.0002	.002	.288	.67
		(1.019)	(2.089)	(1.653)	(2.039)	(1.236)	(.056)	(.020)	(.565)	(7.789)	
5.7	*P* [a]	.824	−.430	.016	−.073	.091	.014	−.028			.17
		(7.409)	(4.307)	(1.283)	(1.729)	(.631)	(.147)	(1.292)			

SOURCES.—See Table 2 *supra*, for all variables except *C*. *C* is from Lee Silverstein, Bail in the State Courts—A Field Study and Report, 50 Minn. L. Rev. 621, tables 2, 3 & 4 (1966).

[a] Equation 5.7 is for all 132 counties where data on *B* are available whereas equations 5.1 through 5.6 are for 70 counties where data on *C* are available.

number of defendants receiving penalties. However, data available in a few counties indicate that the fraction of defendants receiving only fines was negligible.

C: average dollar amount of bail set for defendants in each county.

The negative and statistically significant effect of *B* in equation 5.1 reflects in part the negative correlation described above that runs from *P* to *B*. *C* has a positive effect on *P* in equation 5.2, suggesting that defendants with greater conviction probabilities had *C* set at higher amounts.[56] As expected both the absolute value of the regression coefficient on *B* and its significance are reduced when *C* is entered. We have previously shown that defendants released on bail have greater propensities to go to trial. One would like to determine to what extent the observed effect of *B* on *P* in 5.2 is due to differences in the method of disposition of cases (that is, trials versus settlements) between defendants released and not released on bail. In equation 5.3 a trial variable (*T*) has been added. *T* further reduces the regression coefficient of *B* and its significance because defendants going to trial are less likely to be sentenced to prison (that is, the regression coefficient on *T* is negative and significant) and more likely to make bail. In sum, the results of equations 5.1–5.3 support the hypothesis that the probability of conviction is increased for a defendant when he is not released on bail. At the mean values of *P* and *B* (both are about .5) the regression coefficient on *B* in 5.3 implies that the frequency of prison convictions is .38 for defendants released on bail and .62 for defendants not released, holding *C* and *T* constant. Observe that the coefficient of *B* is reduced by about 40 per cent when *C* and *T* are held constant—15 per cent due to *C* and 25 per cent due to *T*.

Regressions are also presented in Table 5 on the fraction of defendants dismissed (*Ds*) and the fraction acquitted (*A*). These results confirm the previous findings that defendants released on bail are less likely to be convicted. The regression coefficients on *B* are positive in all three equations where *C* and *T* are held constant, and statistically significant in two. Note that 15 per cent of defendants in the sample were acquitted or dismissed, 50 per cent were sentenced to prison, while the remaining 35 per cent were generally given suspended sentences or placed on probation. The latter type of sentences, where the defendant's costs are small

56. Data on *C* are available in only 70 of the 132 counties used in analysis of trials in county courts. The exclusion of 62 counties does not create any obvious biases since a regression computed for all 132 counties without *C* (equation 5.7) yields similar coefficients to equation 5.1.

in comparison to prison sentences, should probably be viewed as non-convictions. For this reason P would be a better measure of convictions than $1 - (Ds + A)$. The positive though nonsignificant coefficients on C in the Ds and A regressions suggest that increases in defendants sentenced to prison as C rises, which are found in equations 5.2 and 5.3, come from a reduction in probations and suspended sentences rather than from fewer dismissals and acquittals.

Other findings in Table 5 may be summarized as follows. (1) The population variable generally has a positive effect on convictions, indicating that longer trial queues across counties tend to increase the fraction of convictions. One should be cautious with this interpretation because of the uncertain relation between queues and population size and the lack of strong statistical significance of the population variable. (2) The demographic variables, NW, Ur and Y are not statistically significant in any regression. (3) An additional problem relates to the interpretation of the regression coefficient of B. Although a negative effect of B on convictions is found, this could be due to a greater average wealth rather than to a lower cost of resources for defendants released on bail. The relationship between wealth and convictions will be examined in the analysis of the U.S. court data.

Data on judicial expenditures in 1966–67 are available for twenty counties with populations greater than 450,000. A reasonable assumption is that these expenditures are positively correlated with the size of the prosecutor's budget in a county. We would then predict from the theoretical analysis that the proportion of defendants convicted would be greater in counties with larger judicial expenditures per defendant. This hypothesis is consistent with findings in Table 6 where judicial expenditures, denoted by J, have a positive effect on convictions in all regressions.[57] The primary effect of an increase in J is to reduce the proportion of cases dismissed, while there is no significant effect on acquittals. Moreover, the increase in the fraction of defendants going to prison as J rises can be accounted for solely by a reduction in dismissals. At the mean values of J, Ds and P, a 15 per cent rise in J reduces Ds from .13 to .11 and increases P from about .50 to .52. Thus, the major economizing move as judicial expenditures fall is to reduce the number of cases prosecuted – hence an increase in dismissals.

57. J is not divided by the number of defendants in a county because this information is not available. However, population size, which is probably positively correlated with the number of defendants in a county, is held constant in the regression equations.

TABLE 6

WEIGHTED REGRESSIONS AND *t*-VALUES FOR CRIMINAL CONVICTIONS IN 1962, 20 COUNTY COURTS IN U.S.

Equation Number	Dependent Variable	Regression Coefficients and *t*-Values										
		α	*B*	*Pop*	*Re*	*NW*	*Ur*	*Y*	*C*	*T*	*J*	R^2
6.1	*P*	.813	−.539	.028	−.033	.399	−.120	.013	−.015	−.378		.55
		(1.841)	(1.495)	(1.208)	(.262)	(.869)	(.222)	(.172)	(.603)	(1.487)		
6.2	*P*	.396	−.369	−.011	.147	.249	−.554	.123	−.008	−.255	.007	.67
		(.851)	(1.079)	(.350)	(.975)	(.584)	(1.011)	(1.347)	(.365)	(1.066)	(1.796)	
6.3	*Ds* + *A*	.903	.042	.042	−.157	−.124	.904	−.257	.017	.186	−.011	.82
		(2.887)	(.182)	(2.035)	(1.539)	(.433)	(2.453)	(4.191)	(1.155)	(1.145)	(4.052)	
6.4	*Ds*	1.013	−.141	.045	−.125	−.097	.775	−.241	.010	−.088	−.010	.76
		(3.308)	(.626)	(2.237)	(1.255)	(.344)	(2.149)	(4.010)	(.643)	(.555)	(3.862)	
6.5	*A*	−.110	.183	−.003	−.032	−.028	.129	−.016	.008	.275	−.001	.82
		(.755)	(1.704)	(.333)	(.663)	(.206)	(.747)	(.565)	(1.126)	(3.617)	(.579)	

SOURCES.—See Tables 2 and 5, *supra*, for variables except *J*. *J* is from U.S. Bureau of the Census, Criminal Justice Expenditure and Employment for Selected Large Governmental Units 1966–1967, tables 16 & 20 (State and Local Gov't Special Studies No. 51, 1969).

U.S. COURTS

Two conviction variables are used in regressions computed across U.S. district courts in 1967.

P: the fraction of defendants sentenced to prison.
F: the fraction of defendants receiving a fine only.

Prison sentences were more numerous than fines as the weighted means of P and F were .38 and .07 respectively in the 89 districts.

Of considerable interest in Table 7 is the behavior of the variable D, the proportion of defendants assigned counsel by the court. D has a positive and significant effect on P, and a negative though nonsignificant effect on F in all equations. This suggests that increases in wealth of defendants reduce (increase) the frequency of convictions for offenses carrying prison sentences (fines) since D serves as a proxy variable for the fraction of lower income defendants in a district (see p. 192). These findings are consistent with a "wealth" hypothesis developed in the theoretical section which predicted for risk neutral defendants that (a) when penalties were in the form of jail sentences a rise in wealth would lead to an increase in the defendant's resource inputs and a subsequent fall in the probability of conviction and (b) when penalties are in the form of fines an increase in wealth would lower his inputs and raise the probability of conviction. A related interpretation of the increase in P as D rises is that court-assigned lawyers are less effective and less able than privately hired lawyers. This is not at variance with the "wealth" hypothesis because more able lawyers can be counted as more units of the defendant's resource inputs than less able ones. However, the ability explanation would also predict that privately hired lawyers would reduce the conviction rate on fines, and the reverse is found in Table 7.[58]

An increase in delay between a trial and settlement is associated with an increase in the fraction of defendants sentenced to prison. The coefficients are positive for Qt and $Pc/\bar{T}$ and negative for Qp in the P regressions in Table 7. We also observed in the state data that the population

58. A possible reconciliation is that more able lawyers are able to lower the defendant's conviction costs by shifting penalties from prison sentences to fines. This explanation would be consistent with both a positive effect of D on P and a negative effect of D on F. Another possible interpretation of the observed effects of D on P and F is that in districts where wealth is higher (and hence D lower) the types of crimes committed are more likely to be those carrying fines rather than jail sentences.

TABLE 7

WEIGHTED REGRESSIONS AND t-VALUES FOR CRIMINAL CONVICTIONS IN U.S. DISTRICT COURTS,[a] 1967

Equation Number	Districts	Dependent Variable	Regression Coefficients and t-Values							R^2
			α	Q_t	Q_p	$Pc/\bar{T}$	D	Re	T_1	
7.1	89	P	.593		−.096	.022	.086	.032	.046	.34
			(11.730)		(3.587)	(1.069)	(3.226)	(1.319)	(1.858)	
7.2	89	F	−.057		.061	−.012	−.026	−.027	−.040	.27
			(1.711)		(3.469)	(.875)	(1.471)	(1.707)	(2.441)	
7.3	44	P	.612		−.119	.030	.081	.028	.049	.37
			(6.217)		(2.656)	(.955)	(2.049)	(.852)	(.984)	
7.4	44	P	.502	.094	−.154		.086	.021	.042	.46
			(6.825)	(2.717)	(3.815)		(2.378)	(.701)	(1.232)	
7.5	44	F	.015		.050	−.002	−.012	−.019	−.005	.32
			(.292)		(2.227)	(.149)	(.591)	(1.151)	(.204)	
7.6	44	F	.018	.002	.047		−.011	−.019	−.002	.32
			(.449)	(.108)	(2.149)		(.558)	(1.158)	(.104)	

SOURCES.—See Table 4, *supra*, for all variables except P and F. P and F are from Federal Offenders in the United States District Courts 1967, table D10.

[a] All variables are in natural log form, except P, F, and Re.

variable (interpreted as a proxy for trial delay) had a positive effect on convictions. One reason for the positive association between trial delay and prison convictions may be that the prosecutor becomes more selective with respect to the cases he prosecutes as trial delay increases. That is, he selects from an inventory of cases the ones he believes to have the greatest probability of conviction and the highest sentences if convicted in order to maximize his weighted conviction function. Moreover, if the prosecutor views fines as light penalties in comparison to jail sentences,[59] we would expect a negative relation between trial delay and the frequency of defendants fined. The equations on F give some support to this hypothesis although the regression coefficients on the Qt and $Pc/\bar{T}$ variables are not significant.

IV. SUMMARY AND CONCLUSIONS

The model developed in this essay utilizes two behavioral assumptions: the prosecutor maximizes the expected number of convictions weighted by their sentences, subject to a budget constraint, and the defendant maximizes the expected utility of his endowments in various states of the world. Both participants can influence the probability of conviction by their input of resources into the case, and cases are disposed of either by a trial or a voluntary pretrial settlement between the prosecutor and defendant. A settlement results in either a dismissal or a guilty plea. The major implications of the model are the following:

1. A settlement is more likely to take place (a) the smaller the sentence if convicted by trial, (b) the greater the resource costs of a trial compared to a settlement, (c) the greater the defendant's aversion to risk, and (d) the greater the defendant's estimate of the probability of conviction by trial relative to the prosecutor's estimate. We further showed that if the defendant and prosecutor agree on the expected outcome of a trial, a decision to go to trial is analogous to accepting an unfair gamble. In this instance, a settlement would result for risk neutral and risk averse defendants.

2. The defendant's investment of resources into his case is related both to the sentence if convicted by trial and to wealth. Generally, the

59. The Administrative Office of the U.S. Courts assigns the value 1 to fines and 3 to imprisonment of 1 to 6 months in calculating a weighted average of the severity of all sentences (see Fed. Offenders, *supra* note 40, at 35, table 10 and note 48, *supra*). This indirectly suggests that fines are of small magnitude compared to jail sentences of a few months.

resource investment is greater for crimes carrying larger sentences. Under the special assumption of risk neutrality (or presumably where the deviation from risk neutrality is small), increases in the defendant's wealth lead to greater resource investments when penalties are jail sentences and to smaller investments when penalties are money fines.

3. Court delays increase the opportunity costs of a trial compared to a settlement for defendants not released on bail. This leads to a smaller likelihood of going to trial for these defendants than for defendants released on bail. The greater the court delay the greater the difference in trial demand between the two groups. Pretrial detention also raises the marginal costs of the defendant's resources and hence lowers his input. Therefore, defendants not released on bail are likely to have higher conviction probabilities in a trial and receive longer sentences if they settle than defendants released on bail. If making bail is a positive function of wealth, then the effects of pretrial jailing fall primarily on low-income defendants. We argued that paying a defendant not released on bail for time spent in jail prior to disposition of his case, or alternatively, crediting him for this time towards his eventual sentence and paying him only if he is not convicted would eliminate much of the "discriminatory" aspects of the current bail system.

4. In the absence of money pricing for the courts a trial queue arises to ration the limited supply. An equilibrium queue is reached because trial costs increase with the length of the queue. Queues could be reduced by charging a money price for trials, which reduces demand, leading to more settlements. Various methods of allocating the court fee—loser pays, winner pays, defendant and prosecutor share the cost—are consistent with a downward sloping demand curve for trials. Pricing trials will not only reduce delay but can also distribute trials more equally among defendants independent of their ability to make bail.

Available data on criminal defendants in state county courts and in U.S. district courts enabled us to test a number of the hypotheses developed in the theoretical analysis. Multiple regressions were estimated for various cross sections in selected years from 1957 to 1968. The principal findings of the empirical analysis may be summarized as follows.

1. The propensity to go to trial was smaller for defendants not released on bail than defendants released, holding constant the average sentence and several demographic variables. This was observed for state county courts in the United States as a whole, and the non-South. Moreover, results from the U.S. district courts indirectly indicate that increases in wealth do not increase trial demand. Thus, the observed relation between bail and trials in state courts is probably due to cost dif-

ferences as predicted by the model rather than to differences in wealth that are positively correlated with the ability to make bail.

2. The absolute difference in trial propensities between defendants released and not released on bail increased as county population rose. One explanation for this finding is that court delay is greater in counties with larger populations. Note that direct measures of court delay were not available for the state courts.

3. Trial demand was negatively related to trial delay and positively related to settlement delay across U.S. district courts for 1960, 1967, and 1968. Thus, as the queue differential between a trial and settlement increased, the demand for trials fell.

4. Subsidizing defendant's legal fees in the U.S. district courts increased the demand for trials. This is consistent with the hypothesis that as the cost differential between a trial and settlement falls, the demand for trials increases.

5. District courts in which the average sentence was greater had proportionately more trials as predicted by the model. However, the results for the sentence variable in the county courts were inconclusive. The latter may be due to the crudity of the sentence variable used in counties.

6. The probability of conviction as measured by the proportion of defendants sentenced to prison was greater for defendants not released on bail than for defendants released on bail in county courts. This was observed in regressions which held constant, among other factors, both the size of money bail and the method of disposition (that is, trial or settlement). Money bail was included as an independent variable to reduce spurious correlation between the conviction and bail variables since defendants who were more likely to be convicted were also likely to have bail set at higher amounts, reducing their chance of release on bail. Regressions using the proportion of defendants acquitted and dismissed as the dependent variable supported the finding that defendants not released on bail were more likely to be convicted.

7. Convictions leading to prison sentences were lower in districts where estimates of the average wealth were higher, while convictions resulting in monetary fines were greater where average wealth was higher. One interpretation of this result is that the effect of wealth on the defendant's investment of resources into his case depended on whether penalties were jail sentences or fines.

8. Conviction rates were higher in district courts where trial delay was greater, and in county courts where judicial expenditures were larger. The former may result from a greater selectivity on the part of

the prosecutor with respect to cases he prosecutes as the backlog increases. The latter was consistent with the hypothesis that the size of the prosecutor's budget determined the proportion of defendants convicted.

APPENDIX A

CIVIL CASES

We can extend our model to make it applicable to civil cases. The plaintiff replaces the prosecutor. Damages replace sentences. Both the plaintiff and defendant maximize their expected utility. It is assumed that civil trials decide both the question of the defendant's liability and the amount of damages. Only the defendant's guilt was at issue in criminal cases; the sentence if convicted was fixed and known prior to trial. A similar assumption for damages is not justified because statutory penalties generally do not exist for various types of civil suits. This modification requires that the plaintiff and defendant form expectations not only on the probability of the defendant being found liable, but also on the size of damages. With these changes, the analysis of civil cases remains quite similar to the model for criminal cases. To avoid excessive duplication I present only a brief outline of the civil model and its more important results.

In civil suits the plaintiff and defendant each select a level of resource inputs that maximizes his expected utility in the event of a trial. The plaintiff's inputs raise both the estimates of the probability that the defendant will be found liable and the amount of damages awarded in a trial, while the defendant's inputs lower these estimates. The plaintiff will determine a settlement payment ($= X$) that yields him the same utility as his expected utility from a trial. X would be the minimum sum accepted by the plaintiff to settle. If the payment of X by the defendant yields him a higher utility than his expected utility from a trial, a settlement will take place. This follows because one can find a payment somewhat greater than X that gives both parties a higher utility from a settlement than their expected utilities from a trial. It can further be shown that a settlement is likely when the following factors are present: (1) both parties have similar expectations on the probability that the defendant will be found liable in a trial; (2) both parties have similar estimates of the damages, given that the defendant is found liable in a trial; (3) neither party has strong preferences for risk; (4) the costs of a trial including lawyer's fees, time costs of the plaintiff and defendant, court fees, etc., exceed the costs of a settlement. Alternatively, the more dissimilar the plaintiff's and defendant's estimates of liability and damages (providing the plaintiff's estimates are higher), the greater their preference for risk, and the lower court costs relative to settlement costs, the more likely a trial.[60]

60. This result is similar to one derived by R. H. Coase, The Problem of Social Cost, 3 J. Law & Econ. 1 (1960). Coase shows that with well-defined property rights, and in the absence of transaction costs, a private agreement will be reached between individuals that

The analysis of charging a money price for the courts, as opposed to queuing costs, is similar for civil and criminal cases. For example, a money price (*M*) will raise the maximum settlement offered by the defendant in a civil suit and lower the minimum settlement accepted by the plaintiff. *M* narrows the gap between what the defendant offers and what the plaintiff is willing to accept (providing the former was initially less than the latter sum), and increases the likelihood of a settlement. Moreover, the greater *M* the fewer civil cases that go to trial. As *M* falls to zero, the demand for trials will increase and a queue is likely to develop. The queue rations demand in the following way. As the queue lengthens, the discounted value of damages awarded in a trial falls. This would lower both the amount the plaintiff will accept and the amount the defendant will offer in a settlement. However, there are probably some costs to the defendant from trial delay. For example, his ability to dispose of assets (particularly if they are directly involved in the suit) and his ability to obtain funds in the capital market may be adversely affected by his being involved in litigation. If, on the average, the gains and costs of delay to the defendant offset each other, or the costs dominate, then the defendant's settlement offer would remain constant or increase as delay increases. The net effect would be a reduction in desired trials as the queue lengthened, since the defendant's settlement offer remains constant or increases while the plaintiff reduces the amount he is willing to accept.

There are several additional points on court delay that should be noted. (1) The analysis of queueing in civil cases is almost identical to criminal cases where the defendant is released on bail. In the latter, the prosecutor reduces his minimum sentence offer as delay increases, while the defendant's response is affected by two offsetting forces. Delay pushes his potential sentence from a trial further into the future, reducing its present value, while simultaneously his current earnings may be adversely affected by being under indictment. (2) A system analogous to the bail system could be instituted for civil cases. This would require, for example, defendants in civil suits to either pay a sum to the court per unit of time, or forgo the returns from all or part of their assets by depositing them with the court during the period between filing and disposition of the case. One effect of this procedure would be to make trial demand more responsive to a change in the queue as the costs of delay rise to the defendant. This is similar to the greater responsiveness of criminal trial demand for defendants not released on bail relative to those released. (3) A requirement that the defendant pay interest on any sum awarded the plaintiff in a trial would have little effect on trial demand or court queues. Interest payments would raise both the defendant's settlement offer and the minimum sum acceptable to the plaintiff in a settlement. Hence, as a first ap-

---

internalizes externalities. If we interpret the absence of transaction costs in civil cases as the availability of information on damages at zero cost and zero bargaining costs of a settlement, and we generalize Coase's notion of well-defined property rights to include identical expectations over property rights or liability decisions in a trial, then Coase's theorem on private agreements would also include pretrial settlements in the absence of risk preference.

proximation it would not close the gap between the defendant's offer and the plaintiff's acceptance sum and, therefore, would have no effect on the trial versus settlement decision. (4) Differences in the rate at which the plaintiff and defendant discount future damages awarded at a trial can give rise to differences in the response of trial demand to a change in the queue. The higher the plaintiff's discount rate relative to the defendant's, the larger the plaintiff's losses and the smaller the defendant's gains from an increase in the queue. This, in turn, would reduce the sum acceptable to the plaintiff by a greater amount than it reduces the defendant's offer, making a settlement more likely.

We can test the hypothesis that the demand for civil trials is negatively related to the length of the trial queue. The statistical specification of the demand function is

$$T = \alpha + \beta_1 Q_t^* + \beta_2 E(T) + \beta_3 Re + \mu. \tag{22}$$

Data were from the 86 U.S. district courts in 1957–61. The variables are in natural log form except *Re* and are defined as follows:

T: The ratio of the number of trials from cases that commenced in 1957 over the number of cases commenced in 1957.[61]

Q_t^*: Estimate of the expected trial queue in 1957 where Q_t^* is an exponentially declining weighted average of 1957, 1956, and 1955 median trial queues.[62] Q_t, the median trial queue in 1956, was also used as an estimate of the expected trial queue.

---

61. A frequency distribution of civil cases by length of time from filing to disposition by trial is published for each year from 1957 to 1961. (After 1961 only median trial queues are available.) This allows us to trace over time the eventual disposition (that is, trial or settlement) of cases commenced in 1957 assuming all of the latter cases are disposed of within four years from the date of filing. Since civil trial queues average about one-and-one-half years in the U.S. courts, a frequency distribution of trials is an important advantage in estimating *T*. For example, if the number of trials in a given year were used as the numerator of *T*, it would be difficult to choose an appropriate denominator for *T* because the trials were from cases commenced over several different time periods with an average queueing time to trial of one-and-one-half years. Moreover, it would be equally difficult to choose a value for the expected trial queue. Frequency distributions of trials are not available for criminal cases, but the above problems are not as great since criminal queues average about six months.

62. Derived from the assumption that persons form expectations of future queues on the basis of past expectation and adjustment based on ratio of current value to previous expected value. That is,

$$Q_{t_y}^* = Q_{t_{(y-1)}}^* [Q_{t_y} / Q_{t_{(y-1)}}^*]^\gamma \tag{i}$$

where Q_t^*'s are expected and Q_t's are actual median queues in a district, y is the year, and γ is the adjustment coefficient. (i) can be rewritten as the following infinite series:

$E(T)$: Expected fraction of trials in a district are estimated by dividing civil cases commenced in each district in 1957 into five broad groups, and then multiplying each group by the fraction of trials in that group for all U.S. district courts in 1957.[63] Therefore, the inclusion of $E(T)$ allows us to hold constant differences in the distribution of types of cases across districts.

Re: Region dummy variable that equals 1 for district courts in South and 0 for non-South district courts.

Table 8 presents regression estimates of equation (22). Separate regressions were also computed for districts in the non-South and South. All regression coefficients on Q_t^* and Q_t have the predicted negative signs, are highly significant, and are of similar magnitude. In sum, these results support the hypothesis that the demand for trials is negatively related to the size of the trial queue.

A difficulty in interpreting the findings of Table 8 arises from the way in which the trial variable is measured. An unknown number of civil cases that would come under the jurisdiction of the U.S. courts are settled before they are filed. Since these cases are excluded from the denominator of T, the true proportion of civil cases going to trial in a district each year is less than the observed fraction. This measurement error in T will not bias the regression coefficients if the error is uncorrelated with the independent variables. However, we can show that the error

---

$$Q_{t_y}^* = Q_{t_y}^{\gamma} \cdot Q_{t_{(y-1)}}^{\gamma(1-\gamma)}, \ldots \cdot Q_{t_{(y-\infty)}}^{\gamma(1-\gamma)\infty} \qquad \text{(ii)}$$

Q_t^* in 1957 was approximated in the empirical analysis by using three previous values for Q_t. In logs this becomes

$$\log Q_{t_{57}}^* = \gamma \log Q_{t_{57}} + \gamma(1-\gamma) \log Q_{t_{56}} + \gamma(1-\gamma)^2 \log Q_{t_{55}}. \qquad \text{(iii)}$$

γ was initially set equal to .4, but to have the weights sum to 1 all weights were proportionally raised by a factor of 1.2755.

63. Fraction of trials for various categories are as follows:

1. U.S. Plaintiff (excludes land condemnation and forfeiture cases)	.031
2. U.S. Defendant (ex. habeas corpus)	.188
3. Federal Question (ex. habeas corpus)	.123
4. Diversity	.151
5. Admiralty	.081

Note that 53,343 civil cases were commenced in 1957 in 86 U.S. District Courts and the number of cases excluded above were 4,613. These were excluded because data on queues and trials for each district do not include these types of cases. Note that when the U.S. government is involved in a suit as a defendant there is much greater likelihood of a trial than when the U.S. is the plaintiff. One explanation is that the costs to the defendant from delay (that offset the gains from the reduction in the present value of a trial settlement), such as his inability to dispose of assets or to obtain funds in the capital markets, may not be present when the defendant is the U.S. government. Hence, for a given queue one would expect more trials when the U.S. government is the defendant than when it is the plaintiff.

TABLE 8

WEIGHTED REGRESSION EQUATIONS[a] AND t-VALUES FOR CIVIL TRIALS IN U.S. DISTRICT COURTS, 1957–61

Equation Number	Area	Districts	α	Q_t^*	Q_t	$E(T)$	Re	R^2
				Regression Coefficients and t-Values				
8.1	U.S.	86	−.307	−.410		.354	−.337	.28
			(.413)	(5.444)		(1.051)	(4.014)	
8.2	U.S.		−.422		−.369	.352	−.314	.27
			(.563)		(5.188)	(1.032)	(3.757)	
8.3	Non-South	52	.164	−.429		.546		.45
			(.238)	(6.033)		(1.770)		
8.4	Non-South		.196		−.400	.596		.46
			(.285)		(6.079)	(1.940)		
8.5	South	34	−2.005	−.376		−.255		.12
			(1.087)	(2.080)		(.288)		
8.6	South		−2.399		−.323	−.386		.10
			(1.283)		(1.818)	(.420)		

SOURCES.—1956–1962 Ann. Rep., Admin. Off. of the United States Courts, tables C1, C3 and C5.

[a] Each observation weighted by $\sqrt{n}$ where n equals the number of cases commenced in 1957.

in T is likely to be positively correlated with the trial queue, and this in turn will bias downward the absolute value of the queue elasticities.[64]

In Table 8 $E(T)$ has a positive and significant effect on T in the U.S. and non-South, but a negative and nonsignificant effect in the South. Overall, $E(T)$ was less important than trial queues in explaining variations in T across districts. Re which is significant at the .01 level indicates that the fraction of civil trials was about 30 per cent lower in the South holding the queue and $E(T)$ constant. This result is puzzling in view of the finding that Re had no significant effect on the

64. Let t = the number of trials in a district, F = the number of cases filed, and C = the number of cases filed plus those settled before filing. Further, assume that $K \cdot F = C$, where $K > 1$. Suppose the relationship between t/C and Q_t^* is

$$t/C = Q_t^{*\beta} e^{\mu}, \tag{i}$$

while the estimating equation is

$$t/F = Q_t^{*\hat{\beta}} e^{\hat{\mu}} \tag{ii}$$

where μ and $\hat{\mu}$ are error terms. (ii) can be rewritten as

$$\log K + \log (t/C) = \hat{\beta} \log Q_t^* + \hat{\mu}. \tag{iii}$$

Let $E = \log K$, $Y = \log (t/C)$ and $X = \log Q_t^*$, and let e, y, and x denote deviations from demand for criminal trials in the U.S. courts. A possible explanation is that the average size of damages in civil suits in the South is considerably lower than in the non-South. Thus, the negative effect on T of lower damages would be picked up by the Re variable.

APPENDIX B

MATHEMATICAL NOTES: WEALTH EFFECTS

In this section we analyze the effect of changes in W on R when the defendant has nonneutral tastes for risk. Risk aversion is assumed. The case of risk preference can easily be worked out from the example of risk aversion. Inputs of R are assumed to reduce both P and S (that is, $S' = \partial S/\partial R < 0$) in contrast to the assumption in Section I that S was a constant and independent of R.

The first- and second-order conditions for $E(U)$ to be a maximum may be written, respectively, as

$$-P'[U(Wn) - U(Wc)] - sS'PU'(Wc) - r[PU'(Wc) + (1-P)U'(Wn)] = 0 \quad (23)$$

and

$$\begin{aligned} -R''[U(Wn) - U(Wc)] + 2rP'[U'(Wn) - U'(Wc)] + r^2[PU''(Wc) \\ + (1-P)U''(Wn)] - 2sS'P'U'(Wc) - sS''PU'(Wc) \\ + 2rsS'PU''(Wc) + (sS')^2PU''(Wc) < 0. \quad (24) \end{aligned}$$

Relative risk aversion at Wn is defined as follows:

$$A(Wn) = -WnU''(Wn)/U'(Wn). \quad (25)$$

$A(Wc)$ is similarly defined at Wc. Taking the total differential of (23) with respect to W and R, noting that (24) is negative, and substituting $A(Wn)$ and $A(Wc)$ gives $dR/dW \gtreqless 0$ as

$$\begin{aligned} U'(Wn)\left[-P'k - r'(1-P) + \frac{r(1-P)A(Wn)k}{Wn}\right] \\ + U'(Wc)\left[P'm - s'S'P + \frac{sS'PA(Wc)m}{Wc} - r'P + \frac{rPA(Wc)m}{Wc}\right] \gtreqless 0, \quad (26) \end{aligned}$$

---

their respective means. The least-squares estimator of $\hat{\beta}$ is

$$\hat{\beta} = \beta + \frac{\Sigma xe}{\Sigma x^2}, \quad \text{(iv)}$$

which will be an unbiased estimator of β only if Cov $(x, e) = 0$. However, it is more likely that as Q_t^* rises, K will also rise, since the incentives to settle (both before and after filing) increase with the size of Q_t^*. This implies that Cov $(x, e) > 0$. Given that β and $\hat{\beta}$ are negative, this would result in $|\hat{\beta}|$ underestimating $|\beta|$.

where $r' = \partial r/\partial W > 0$, $s' = \partial s/\partial W$, $k = (1 - r'R)$ and $m = (1 - s'S - r'R)$. Note that $0 < k < 1$, $0 < m < 1$ and $k > m$. m and k are both positive because an increase in W must increase both Wn and Wc. Even with further simplifying assumptions the sign of (26) is indeterminate. For example, suppose $A(Wn) = A(Wc) = 1$ and let $Er = r'(W/r)$ and $Es = s'(W/s)$ where $0 \leqslant Er, Es \leqslant 1$. This gives $dR/dW \gtrless 0$.

$$\frac{U'(Wn)}{U'(Wc)} \gtrless Wn \frac{\begin{bmatrix} -P'(W - E_s sS - E_r rR)Wc \\ - sS'P[W - (E_s Wc + E_s sS + E_r rR)] \\ - rP[W - (E_r Wc + E_s sS + E_r rR)] \end{bmatrix}}{Wc[-P'(W - E_r rR)Wn + r(1 - P)(1 - E_r)W]}. \quad (27)$$

The sign of dR/dW cannot be determined from (27) without additional information about the defendant's utility function, the elasticities of s and r with respect to W, and the productivity of R in reducing S. If $E_s = E_r = 1$, (27) becomes

$$\frac{U'(Wn)}{U'(Wc)} \gtrless \frac{Wc}{Wn}. \quad (28)$$

In the special case of a Bernoulli utility function, where the utility of wealth equals its logarithm, then $dR/dW = 0$, since $U'(Wn)/U'(Wc) = Wc/Wn$.

In general, the effects of changes in wealth on the defendant's input of resources are indeterminate once nonneutral tastes for risk are introduced. This conclusion is valid even when the strong assumption is made that relative risk aversion equals one for all levels of the defendant's wealth. Nevertheless, one still presumes that if the deviation from risk neutrality is small, the effects of wealth on R will be similar to those for risk neutrality.