Finding an Optimum Legal Policy Level: The Undesirability of Doing Too Much or Too Little in the Law

Stuart S. Nagel
University of Illinois at Urbana - Champaign

Follow this and additional works at: https://researchrepository.wvu.edu/wvlr
Part of the Criminal Law Commons, Criminal Procedure Commons, Jurisprudence Commons, and the Law Enforcement and Corrections Commons

Recommended Citation
Available at: https://researchrepository.wvu.edu/wvlr/vol86/iss1/7

This Article is brought to you for free and open access by the WVU College of Law at The Research Repository @ WVU. It has been accepted for inclusion in West Virginia Law Review by an authorized editor of The Research Repository @ WVU. For more information, please contact ian.harmon@mail.wvu.edu.
FINDING AN OPTIMUM LEGAL POLICY LEVEL:
THE UNDESIRABILITY OF DOING TOO MUCH
OR TOO LITTLE IN THE LAW

STUART S. NAGEL*

I. INTRODUCTION

The legal process tends to be epitomized by controversial issues that relate to finding an optimum level for a variety of legal policies, whereby doing too much or too little is undesirable. For example, due process or fair procedure involves a constant struggle between providing too much due process and providing too little due process. If too much due process is provided, then many persons who should have been found guilty will go free in criminal cases, and many who should have been held liable will not be so held in civil cases. This may mean high monetary and nonmonetary costs to the legal system and society that provides too much due process. On the other hand, if too little due process is provided, then many innocent persons will be found guilty in criminal cases and in civil cases many persons will be held liable who should not have been so held. That will mean high costs to the system that provides too little due process. Somewhere in the middle of a valley-shaped cost curve, the cost to the system reaches a minimum. At that point, we have an optimum balance or an optimum level of due process.

Similarly, any legal rule can be worded or applied in an overly strict way or an overly lenient way. If environmental protection standards become too strict, we suffer unduly high cleanup costs, but if the standards become too lenient, we suffer unduly high pollution damage costs. Likewise, contract law standards can become too strict and therefore interfere with freedom of contract. Thus, society will possibly incur large costs by the reduction in the free flow of business. If contract law standards become too lenient, however, then we might encourage large societal costs since the party with less expertise or weaker bargaining power will be exploited. In tort law, automobile negligence standards could be so strict as to slow traffic almost to a standstill, or so lenient as to paralyze potential drivers and pedestrians from venturing into the streets. Similar problems of doing too much or too little can occur in criminal law, divorce law, housing law, or any field of law. Somewhere between these extremes is an optimum point where the sum of the total costs to society is at a minimum.

The purpose of this article is to illustrate some basic principles that are used to arrive at an optimum level on a legal policy where doing too much or too little is undesirable. This article will illustrate these principles by determining: (1) the optimum attorney time in handling civil cases, (2) the optimum level of bail bond for various crimes, (3) the optimum percentage of defen-

* Professor of Political Science, University of Illinois at Urbana—Champaign; B.S. 1957, J.D. 1958, Ph.D. Political Science, 1961, Northwestern University; Member, Illinois Bar.
dants to hold in jail prior to trial, (4) the optimum time to allow between ar-
est and trial, (5) the optimum jury size, and (6) the optimum sentence length
for convicted defendants. All of these legal problems are suited for optimum
policy level analysis since they present situations where doing too much or
too little is undesirable.¹

II. Maximizing Benefits Minus Costs

A typical optimum level problem involves a policy that causes an increase
in benefits as the policy is increased, but also causes an increase in costs as
the policy is increased. The object in such situations is to find the policy level
that maximizes the positive difference between the benefits and the costs.
One example is determining the optimum amount of attorney time to devote
to handling a given case, where we would like to maximize the predicted in-
come minus the predicted expenses. Another example is determining the op-
timum amount of bail-bond to charge an arrested defendant, where we would
like to maximize the probability of appearing in court minus the probability
of being held in jail.

A. Finding an Optimum Amount of Attorney Time per Case

If an attorney is being paid on a flat fee basis of say 900 dollars for handl-
ing a case, the optimum amount of time to devote is none at all if one can get
away with it. Doing so will thereby minimize one's expenses, and thus max-
imize income minus expenses. More realistically, the optimum time is
whatever the ethical/efficient minimum is considered to be to complete the
case. If an attorney is being paid on an hourly basis, the optimum amount of
time to devote is an infinite amount of time if one can get away with it. Doing
so will thereby maximize one's income, and thus maximize income minus ex-
penses. This assumes for each hour, the hourly rate is above what the at-
torney considers his or her time to be worth. More realistically, the optimum
time is whatever the ethical/efficient maximum is considered to be. The most
interesting type of case is the contingency fee case where the attorney gets
paid a percentage of the damages collected, such as thirty-three percent.

¹ On finding an optimum policy level in general, see D. MacRae & J. Wilde, Policy Analysis
152-75 (1982); and M. White, Managing Public Systems: Analytic Techniques for Public Ad-
ministration 278-90 (1980). The present article is especially concerned with methods of relating
alternative policies to multiple goals in order to arrive at an optimum policy level. The main
general categories of appropriate methods are inductive observation (Section II and III), deductive
analysis (Section IV), and the use of authority (Section V), with trial-and-error sensitivity analysis
interspersed. For further discussion of ways of establishing relations between policies and goals,
see "Sources of Goals, Policies, and Relations" in Nagel, Public Policy: Goals, Means and
1. Relating damages to hours

To determine the optimum time in a contingency fee case, it would help to know what damages were awarded in a large quantity of cases that have characteristics similar to those of the case we are currently considering. Along with information on the damages awarded, we would like to know the number of billable hours devoted to each case. "Billable hours" as used here means hours spent on a case for which a client could be legitimately billed if an hourly rate were being charged even though the case may be handled on a contingency fee, flat fee, or salaried basis. The relationship between damages awarded and billable hours can be shown in a graph like Figure 1. Each dot represents a case. By seeing where the dots are positioned on the vertical axis, we can tell the amount of damages that were awarded in each case. For many cases there were no damages because liability was not established and therefore the attorney lost the case. By seeing where the dots are positioned on the horizontal axis, we can tell how many billable hours were devoted to each case.

After graphing such a set of dots, it might then be appropriate to try to fit a meaningful curve to the dots in such a way as to minimize the sum of the distances from the dots to the curve. That can be done by a common-sense eyeballing approach or a more sophisticated computerized analysis. One might expect a well-fitting curve to have a kind of S-shape like that shown in Figure 1. Such a curve indicates that prior to a certain number of hours, putting in additional time is not likely to raise the damages substantially above zero. In other words, cases tend to require an initial minimum of hours before additional time can have an effect on the damages received. Likewise, after some number of hours, additional time probably has little effect in increasing the damages awarded. Thus, each type of case probably has a bottom range of hours before hours substantially affect damages, and a top range of hours after which the influence of additional hours seems to plateau out.

With that kind of information, one might logically say give the case about seventy-five hours since that is the number of hours roughly corresponding to the upper turning point on the S-shaped curve. We would not want to give the case just twenty-five hours corresponding to the lower turning point since those twenty-five hours would be wasted by virtue of the fact that twenty-five hours only brings us to the point where we first start taking off from a zero damages level. If we devote a number of hours between twenty-five and seventy-five, then at first glance it looks as if we would be suffering some opportunity costs in the sense of missing the additional damages that could be obtained by putting in some additional hours. It may, however, be wasteful to devote seventy-five hours to such a case because the value of seventy-five hours to us may be greater than the one-third contingency fee corresponding to a seventy-five hour allocation.
2. Relating fees and costs to hours

Figure 2 represents an improvement on Figure 1 for viewing the time allocation problem. First, it shows the relation between hours allocated and cost to the lawyer, not just damages awarded. Like any business firm or individual, a lawyer is interested not in maximizing income, but in maximizing income minus expenses, or benefits minus costs. For the sake of simplicity, we assume here that the only cost is our time and that our time is worth thirty dollars an hour. By thirty dollars an hour, we are not referring to the charge to clients (which may be fifty dollars an hour), but rather that we would be willing to work for as low as thirty dollars an hour.

Second, Figure 2 refers to fees received rather than damages awarded. We calculate fees received by taking one-third of the damages awarded in Figure 1, since one-third is the usual contingency fee. The third and most important difference from Figure 1 is that this new information should allow us to more rationally decide how many hours to allocate to a case of the type shown in Figure 2. The ideal quantity is about sixty-five hours, rather than the seventy-five hours where the fee received approaches the maximum. It makes more sense to allocate only sixty-five hours because that is where a
maximum positive difference is between the fee received curve and the time-cost line. Either below or above sixty-five hours, we would be taking a loss or not making as much profit. Billable hours between points A and C are predicted to yield a profit, but at point B (where an asterisk is shown), the profit is at a maximum. Figure 1 thus emphasizes how the S-shaped damages awarded curve is derived, and Figure 2 emphasizes what to do with such a curve in order to make a profit-maximizing time-allocation decision.

When we say sixty-five hours is the optimum allocation for cases of this type, we do not mean that when a lawyer has put in sixty-five hours on a case like that, he should then stop whatever he is doing regardless of what is happening. Likewise, we do not mean that if a lawyer has put in only fifty hours on such a case that he should throw in an extra fifteen hours regardless of what those fifteen hours involve. The sixty-five hours is just an average optimum that may be meaningfully deviated from in light of aggravating and extenuating circumstances. Relevant variables justifying more or less billable hours might include how experienced the lawyer is, how cautious he is, and how much help he has from assistants, research tools, and counsel for other
parties. Also relevant is how difficult the case is in terms of the quantity of factual and legal issues and the newness of those issues. The classification of cases, however, is likely to include some general notions of difficulty in that airline crashes are more difficult than dog bites. Although lawyers are generally seeking to maximize fees minus costs, rather than maximizing fees, it is logical to expect a lawyer to spend more time on a case where there is the potential of a large fee because he would regret losing such a case more than he would regret losing a small fee case, but the case classification is also likely to consider that. Although aggravating and extenuating circumstances may influence the extent to which one should deviate from the predicted optimum hours, they should not affect where that optimum generally tends to be for a case with a given set of characteristics.²

B. Finding an Optimum Bail Bond to Charge Arrested Defendants

1. The problem, the data, and one solution

An ideal bond for a given crime would be a bond that maximizes the percentage of defendants who appear for trial without committing a crime minus the percentage of defendants who are held in jail as being unable to meet the bond set. That kind of bond is analogous to arriving at a price or quantity for a product that will maximize income minus expenses since a defendant appearing for trial is like income, and a defendant being held is clearly an expense.

To arrive at the kind of theoretically ideal bond, we could obtain data for many prior criminal cases showing (1) the crime for which the defendant was charged, (2) the amount of bond the defendant was asked to pay in order to be released, (3) whether the defendant was held or released, and (4) if released, whether the defendant appeared for trial without committing a crime. With that kind of data for each case, we could plot points like those in Figure 3. The horizontal axis shows ten different bond-level categories from $0 on up to whatever the highest bond may have been for the crime that is being analyzed. The vertical axis shows probabilities or percentages of appearing or being held. The X's indicate for each bond category the percentage of defendants in the category who were held in jail pending trial. The dots indicate for each bond category the percentage of released defendants in the category who appeared for trial without having been arrested.

One would logically expect the X's to exhibit an S-shaped relation between bond category and the probability of being held. That means that with low bond categories, we would expect only a small percentage of defendants to be held, and at high bond categories nearly all the defendants might be held. Among the middle categories, there would be a positive slope like that

² For further detail on finding an optimum amount of attorney time per case, see Nagel, Attorney Time Per Case: Finding an Optimal Level, 32 U. FLA. L. REV. 424 (1980).
shown in Figure 3. Likewise, one would expect the dots to also exhibit an S-shape relation between bond category and the probability of appearing. That means that at the low bonds, defendants would not have much of an incentive to appear in order to retrieve their bond money, especially if they feel the case will be resolved as a simple bond forfeiture if they do not appear. At higher bond levels, the defendants would theoretically have more of an incentive to appear since they have much more of a bond to recover and since they are likely to feel that nonappearance will result in a warrant for their arrest, rather than closing the case as if the bond were a fine. With those assumptions as plotted in Figure 3, one can see that the optimum bond is in approximately Category 4.

FIGURE 3 FINDING AN OPTIMUM BAIL BOND TO CHARGE ARRESTED DEFENDANTS

(Hypothetical data for the crime of disorderly conduct in medium sized cities in the state of Illinois)

2. Other solutions

Preliminary data gathered from New York City tends to support the meaningfulness of this analysis. Defendants, though, were found to have a

---

A. Schaffer, Bail and Parole Jumping in Manhattan in 1967 (Vera Institute of Justice, 1968).
high rate of appearance even at extremely low bonds. Seventy-six percent appeared when the bond level was only one to twenty-five dollars, although the appearance rate is lower in traffic matters. The rate of appearance tends to drop off in the highest bond category, probably because the fear of conviction and imprisonment in that category is great enough to offset the desire to retrieve the bond money and avoid being a fugitive. The rate dropped from ninety-three percent (when the bond was 501 to 1,000 dollars) down to eighty-nine percent (when the bond was 1,001 to 2,500 dollars). There were not enough cases in the sample to analyze the over 2,500 dollars category. This pattern, however, still makes the Figure 3 analysis meaningful because the percent-appearing curve is positively sloped. Therefore, there is no need to resort to an averaging approach to arrive at nondiscretionary bonds if one can do so through a more meaningful benefit-cost analysis approach. The prior averages may also reflect the tendency of bondsetting judges to set bonds higher than necessary in order to avoid personally-embarrassing releasing costs, rather than societally-expensive holding costs.

The major defect in either a benefit-cost approach or an averaging approach to nondiscretionary bond-setting is that the bonds set are likely to place an unnecessary and inequitable burden on low income defendants. This is a good example of where a solution that is efficient in the sense of maximizing benefits minus cost may be inequitable in the sense of distributing the benefits of being released disproportionately to non-indigent defendants, and distributing the costs of being held disproportionately to the indigent defendants. The nonindigent tend to pay higher taxes to support pretrial jailing, but that does not offset the often needless incarceration of indigent defendants who would be good risks to release, but they simply cannot meet the optimum or average bond for the crimes with which they have been charged.

The New York City data showed that in the bond category of $101 to $500 (which seems to be a low bond amount), fifty-seven percent of the defendants with such bonds could not meet them and had to be held in jail pending trial. Once incarcerated these defendants were (1) more vulnerable to needlessly pleading guilty, (2) less capable of preparing their defenses, and (3) more susceptible to embittering criminogenic influences. The system could give judges the discretion to raise or lower the nondiscretionary bond up to twenty-five percent in order to consider special circumstances including ability to pay, but that may not be sufficient. The system could also allow release on recognizance for good risks who cannot meet the nondiscretionary bond in the same way nondiscretionary sentencing systems allow for the possibility of probation.

Perhaps the system should provide that all defendants can either be released on their own recognizance or not released until a speedy trial. To prevent abuses of discretion in making such release/nonrelease decisions, judges could be required to follow a point-setting table similar to the one
OPTIMUM POLICY LEVEL

devised by the Vera Institute of New York City. It allocates points to such
criteria as how long the defendant has lived in the community, and how long
he has held his current job. The judges can be informed how many points
each criterion is worth as based on previous statistical analysis of the relations
between various criteria or predictors and whether or not the released
defendants appeared for trial without being arrested. Such a point system
would amount to a set of quantitative guidelines for releasing defendants
without having to put up bond money, since money bonds inherently
discriminate across economic classes even when the bond figures are low.

The guidelines could also allow for adding or subtracting points for ag-
gravating and mitigating circumstances (which are explicitly stated), while
still preserving the basic objectivity of the system. This analysis illustrates
finding an optimum level that maximizes benefits minus costs. It also il-
lustrates how a problem that begins as an optimum level problem can end as
an either-or problem. In other words, what we really want to know is not how
much bail-bond to set for each defendant, but rather whether or not to
release each defendant on his or her own recognizance.4

III. MINIMIZING THE SUM OF CONFLICTING COSTS

Another typical optimum level problem involves a policy that increases
one set of costs when the policy is increased, but also increases a second set
of costs when the policy is decreased. The object in such situations is to find
the policy level that minimizes the sum of the two sets of costs. Illustrative
examples include the percentage of defendants to hold in jail prior to trial, or
the amount of time to be consumed between arrest and trial.

A. Finding an Optimum Percentage of Defendants to Hold in Jail Prior to
   Trial

1. The problem and the data

One policy problem is finding an optimum percentage of defendants to
hold in jail prior to trial in order to minimize the sum of the holding costs and
releasing costs. An alternative way of stating the problem would be to max-
imize the sum of the holding benefits (i.e., the releasing costs avoided) and
the releasing benefits (i.e., the holding costs avoided). One could also say the
problem is one of maximizing the holding benefits minus the holding costs, or
one of maximizing the releasing benefits minus the releasing costs.

In 1969, questionnaires were mailed to judges, prosecutors, defense at-
torneys, and bail project directors from numerous cities, although here we

4 For further detail on finding an optimum bail bond to charge arrested defendants, see
Nagel, Neef & Schramm, Decision Theory and the Pre-Trial Release Decision in Criminal Cases,
only use the data from the twenty-three cities that provided complete information for all the basic variables. The basic variables are:

1. The costs per month to hold an inmate in jail.
2. The average length of time spent in jail for those defendants who are held in jail pending trial.
3. The number of individual defendants arraigned in 1968, i.e., brought before a magistrate or judge to determine whether the individual should be held or released before the trial.
4. The percentage of arraigned individuals held pending trial.
5. The percentage of arraigned individuals who were released, but who failed to appear in court for their trials.
6. The percentage of arraigned individuals who were released, but who were arrested for committing another crime while released.
7. The percentage of persons held in jail before trial who were found guilty.

From these variables we can calculate the total holding cost (THC), total releasing cost (TRC), and the total over-all cost (TC). The total holding cost is the sum of the jail costs, the loss of the gross national product when a defendant is held in jail, and cost based on a monetary valuation of the bitterness felt by defendants who are subsequently found not guilty. The total releasing cost is the sum of the costs to society when a released defendant commits a crime, and the costs of rearresting such defendants.

---

6 This cost is in dollars. The data was originally expressed as cost per day per inmate. These figures were multiplied by 30 to put them on a per month basis.
7 These costs were calculated as follows:
   Total holding costs (THC) = THC₁ + THC₂ + THC₃, where
   THC₁ = Total jail cost = #H • LJ • $JC per day • 30 days;
   THC₂ = Total GNP cost = #H • LJ • $360 per month;
   THC₃ = Total bitterness cost = #H • LJ • $300 per month • (1.00 - %HG);
   and where,
   JC = Jail costs per day per inmate;
   LJ = Average length of time spent in jail (in months) for those defendants who are held in jail pending trial;
   A = Number of individual defendants arraigned;
   %H = Percentage of arraigned individuals held pending trial;
   %HG = Percentage of persons held in jail before trial who were subsequently found guilty;
   #H = %H • A = Number of individuals held.
8 These costs were calculated as follows:
   Total releasing costs (TRC) = TRC₁ + TRC₂, where,
   TRC₁ = Total rearrest cost = #R • $200 • %R_p;
   TRC₂ = Total crime-committing cost = #R • $1000 • Rc;
   and where,
   %R_p = Percentage of arraigned individuals who were released, but who were arrested for committing another crime while released;
   %R_c = Percentage of arraigned individuals who were released, but who failed to appear in court for their trials.
Constants in these costs include $360 a month as the average gross national product when a defendant was kept in jail for thirty days pending trial. That figure is based on an assumption of two dollars an hour in 1969 wages at forty hours a week and four and one-half weeks a month. In calculating bitterness costs, it is assumed that bitterness is mainly generated among those detained defendants who are not subsequently found guilty (i.e., the complement of the percentage of those defendants who are found guilty). The bitterness costs for these defendants are figured at approximately $300 a month. This means that if these defendants were paid $300 a month, they would generally feel compensated for their bitterness. The figure is not equal to lost wages on the assumption that these defendants recognize that legitimate, nondeliberate mistakes can be made in the criminal justice process. It is assumed that it would cost approximately $200 to rearrest a defendant who fails to show up for trial. It is further assumed, in the absence of any better information, that the average released defendant who commits a crime while released costs the victims and society approximately 1,000 dollars. The effects of changing these tentative figures on the optimum percentage to hold can be determined as part of a sensitivity analysis designed to determine how the changes affect the optimum percentage to hold.

2. Graphing the problem

For each of the twenty-three cities, we can plot a point showing the percent of defendants held and the total holding costs. If we draw a smooth curve through those twenty-three points that minimizes the sum of the distances squared from the points to the line, we get a positive convex curve like that shown in Figure 4.9 Similarly, for each of the twenty-three cities, we can plot a point showing the percent held and the total releasing costs. If we draw a curve through those twenty-three points, we get a negative convex curve like that shown in Figure 4.10

Since total costs are the sum of holding costs and releasing costs, we can show an asymmetrical valley-shaped total cost curve by adding the distance from the horizontal axis up to the total holding costs (THC) curve, and the distance from the horizontal axis up to the total releasing costs (TRC) curve

\[ A = \text{Number of defendants who were arraigned}; \]
\[ \#R = A - \#H = \text{Number of defendants who were released.} \]

9 The equation for this curve is \( THC = 1185 \times (%H)^{1.31} \) where, THC = Total holding costs; %H = percentage of arraigned individuals held pending trial. This means that as %H becomes 100 percent (or 1.00 when expressed as a decimal), then the total holding costs equal $1,185 per defendant. The 1.31 means that as %H goes up 1 percent, the holding costs go up at a faster rate of 1.31 percent. At the left end of the graph, if %H becomes one percent or .01, then the total holding costs per defendant equal $3, although for simplicity here we ignore the fixed holding costs that remain constant regardless of how low the percentage of defendants held becomes.

10 The equation for that curve is \( TRC = 77 \times (%H)^{-1.7} \), where TRC = Total releasing costs; %H = Percentage of arraigned individuals held pending trial.
FIGURE 4 DETERMINING AN OPTIMUM PERCENTAGE OF DEFENDANTS TO HOLD

PERCENTAGE OF DEFENDANTS HELD BEFORE TRIAL (%H)

Formulas: THC = $1185(%H)^{1.31}; TRC = $77(%H)^{0.17}; TC = THC + TRC.

in order to create the total costs (TC) curve.\footnote{The equation for that curve is TC = 1185(%H)^{1.31} + 77(%H)^{0.17}, since TC = THC + TRC. See supra notes 8 and 9. Figure 1 is not drawn exactly to scale in order to enlarge the area of Figure 1 where the total cost curve reaches its bottom point. Also, if the percentage held were allowed to go to zero percent rather than one percent on the figure, then at that point TRC and TC would equal infinity. At one percent, THC = $3, TRC = $188, and TC = $171.} From Figure 4 we can see that the total costs increase as the percentage of defendants held decreases from four percent to zero percent. Similarly, the total costs also increase as the percentage of defendants held increases from four percent to one hundred percent. Therefore, from Figure 4 we can see that the total costs are at a minimum when four percent of the arraigned individuals are held in jail pending trial.\footnote{This is so because from four percent to zero percent, the TRC curve is rising faster than the slowing THC curve is falling; whereas after four percent, the THC curve is rising faster than the slowing TRC curve is falling. At four percent, the slope of the TRC curve is the same as the slope of the THC curve. The easiest way to determine where the total cost curve bottoms out is to}
What we have done here is analogous to what a rational business firm producing one product is supposed to do when it decides on the optimum level of goods to produce in order to maximize its total profits. In the absence of this kind of analysis the decisions or policies reached may not be profitable or productive. This is especially likely to be the case if the individual decision makers are not seeking to maximize societal benefits minus societal costs. This may happen where individual judges are not bothered by the societal holding costs, but are quite sensitive to the embarrassment that comes from releasing a defendant who fails to appear or commits a crime while released. In other words, judges' individual benefit-cost analysis may not necessarily be consistent with a benefit-cost analysis for society in general, and thus they may be slighting societal productivity.  

B. Finding an Optimum Level of Time to Consume from Arrest to Trial

Suppose you are a court administrator whose concern is processing cases. In a world where there are no costs, the ideal delay time for processing cases might be zero delay, so that cases are disposed of as fast as they arrive. In the real world, however, that might be too expensive, and it might be undesirable to rush so quickly. On the other hand, too much delay may mean missed opportunities and substantial harm to those who are kept waiting for a governmental service. Optimum level analysis in the context of optimum time consumption is designed to determine the optimum level of delay in the sense of minimizing the sum of the delay costs (DC) and the speed-up costs (SC). That is the equivalent of saying we want to maximize the delay benefits (which are the speed-up costs-saved) minus the delay costs, or the same as saying we want to maximize the speed-up benefits (which are the delay costs-saved) minus the speed-up costs. It is easier, however, to talk about minimizing the sum of the two kinds of costs rather than talk about costs-saved.

1. Measuring delay costs

Figure 5 shows what is involved in an optimum level analysis for a hypothetical metropolitan court system. To apply the analysis, we need to develop an equation showing the relation between delay costs and time consumed. A survey and an accounting analysis might show that every extra day of time consumed in completing a criminal case is worth about seven dollars per jailed defendant to the system. The seven dollars represents waste in holding those defendants in jail who will receive an acquittal, dismissal, or

---

For further detail on finding an optimum percentage of defendants to hold in jail prior to trial, see S. Nagel & M. Neef, Legal Policy Analysis: Finding an Optimum Level or Mix 1-74 (1977).
probation when their case is tried. Of that seven dollars, about two dollars represents wasted jail maintenance costs, and five dollars represents lost gross national product that could have been earned. The two dollars is calculated by noting that it costs six dollars per day to maintain a defendant in jail, and one-third of them receive nonjail dispositions upon trial, meaning two dollars per day is wasted by delaying the nonjail disposition. The five dollars is calculated by noting that defendants can earn about fifteen dollars a day if they are not in jail, and that approximately one-third of them would not be in jail if their cases came up sooner, meaning an additional five dollars per day is wasted by delay.

There might also be about three dollars per day wasted per released defendant. The three dollars represents waste in releasing those defendants who will be jailed when they are eventually tried and convicted, but who during the delay commit a crime or have to be rearrested for failure to appear in court. That three dollars is determined by calculating (1) the crime-
committing costs or the rearresting costs for the average released defendant, (2) multiplied by the low probability of the occurrence of crime-committing or rearresting, (3) multiplied by the middling probability of being convicted and jailed if the case were to come to disposition, and (4) divided by the number of days released.

If we assume half the arrested defendants are jailed and half are released, then the seven dollar delay cost per day per jailed defendant becomes $3.50, and the three dollar delay cost per day per released defendant becomes $1.50. Thus the total cost per day per case would be five dollars (or $3.50 plus $1.50). If the five dollars per day were a constant figure, we could say delay costs (DC) equal five dollars times T days, or $5(T). Since the likelihood of crime committing and the need for rearresting increase as delay increases, the relation between DC and T might be better expressed by an equation of the form DC = $5(T)^2. This equation tells us that when T is one day, DC is five dollars; but when T is X days, DC is not five dollars times X, but rather DC increases at an interesting rate. More specifically, as T goes up one percent, DC goes up two percent.

2. Measuring speed-up costs

The more time that is consumed, the delay costs increase at a possibly higher rate. However, the faster we dispose of cases, the greater the speed-up costs might be. These costs mainly include the monetary costs of hiring additional personnel or introducing new facilities or procedures. Suppose from past experience our data reveals that with only twenty judges, cases average seventy-five days per case; with forty judges, thirty-eight days; with sixty judges, twenty-five days; with eighty judges, nineteen days. We can meaningfully assume that with zero judges, the number of days would rise to infinity, and in order to get the number of days down to zero, we would have to have an infinite number of judges.

The speed-up costs curve shown in Figure 5 incorporates that data and those assumptions. A curve of that kind can be expressed by the equation $J = a/T$, where $J$ stands for the number of judges, and $T$ stands for time in days per average case. If $J = a/T$, then $T = a/J$. The "a" in the $J = a/T$ equation is the number of judges needed to get time down to one day per case (i.e., $T = 1$), and the "a" in the $T = a/J$ equation is the number of days consumed when there is only one judge (i.e., $J = 1$) if the relationships are carried out to their logical extremes even though at the extremes the empirical data does not apply. From the above data, we can determine that $a = 1500$. This means, according to our data, that $J = 1500/T$ and $T = 1500/J$.\textsuperscript{14}

\textsuperscript{14} The 1500 numerical value of "a" can be determined with a statistical calculator by simply going through steps like the following:

1. Insert information for the first data point. Doing so means inserting a 20 for 20
Instead of talking just in terms of the relation between the number of judges and the number of days consumed, we should be talking in terms of the cost of judges and the number of days consumed. If one judge costs $40,000 a year, that means $110 per day at a 365 day year. Thus the equation \( J = \frac{1500}{T} \) should be changed to \( Y_2 = \frac{165,000}{T} \). The \( Y_2 \) is the speed-up costs or the additional judge costs, and the $165,000 is simply $110 times the previous “a” (also called the scale coefficient) of 1500 to show we have increased the scale by $110 per judge per day. This equation perfectly fits the above data although in real life the equation-fitting might provide a good fit, but not a perfect fit.

3. Minimizing the sum of the costs

Given the relationship between delay costs and time consumed of \( DC = \frac{5(T)^2}{2} \) and the relation between speed-up costs and time consumed of \( SC = \frac{165,000}{T} \), the relation between total costs (TC) and time consumed is logically \( TC = \frac{5(T)^2}{2} + \frac{165,000}{T} \). The easiest way to find the optimum value of \( T \) with that equation is to go through steps like the following:

1. Guess any value for \( T \) between 0 and some reasonable maximum like 200. For example, try fifty days.
2. Insert that value in place of \( T \) in the above equation that relates \( TC \) to \( T \).
3. Solve for \( TC \) given the guessed value of \( T \). Doing so with a \( T \) of fifty days yields a \( TC \) of $15,800.
4. Try another guessed value like forty days. Doing so yields a value of \( TC \) of $12,125. This tells us that reducing the number of days from fifty days will reduce the total costs.
5. Try another even lower guessed value since going lower seems to help. For example, try ten days. It yields a \( TC \) of $17,000 which is higher than $12,125. That tells us the optimum value of \( T \) must be somewhere between fifty and ten days.

judges, and then hitting the X button to show this is the variable from which we are predicting. Doing so also means inserting a 75 for 75 days and then hitting the Y button to show this is the variable to which we are predicting.

2. Do the same thing for the second data point of 40 judges and 38 days, and then the same thing for 60 judges and 25 days, and so on until we run out of data points.

3. Then hit a button labeled “a” and out will come a 1500 on the display. Some calculators will not read out 1500 at this step. Instead, with the above date, they will read out 3.18. One then has to hit the button labeled 10^x and out comes 1500.

4. One should then also hit a button labeled “b” and out will come a -1 on the display showing that the exponent of \( T \) is -1, which is the same as saying 1500 divided by \( T \).

5. Before going through the above four steps, one has to inform the calculator that we want to know the numerical values for a nonlinear statistical equation like the speed-up costs curve. If there is no button or program instruction for doing that, then one has to hit the logarithm button before inserting each of the data items. One, however, does not have to have any awareness of how the calculator arrives at its answers in order to be able to make use of the answers.
6. Therefore we should now try a guessed value for $T$ that is lower than forty but higher than ten, such as thirty. Doing so yields a $TC$ of $10,000 which is the lowest thus far.

7. We may be able to go still lower on the total cost by going slightly below or above thirty days.

8. Twenty-eight days is an improvement. So is twenty-six, but twenty-four is worse than twenty-six. That tells us the optimum must be between twenty-four and twenty-six.

9. If the days have to be in whole numbers, then the optimum must be twenty-five days. At that figure, our total costs are only $9,725.

If twenty-five days is the optimum number of days, then the optimum number of judges is sixty judges, since $J = \frac{1500}{T}$ or $60 = \frac{1500}{25}$. This means that our court system could be minimizing its total costs if we have sixty full-time judges, rather than any number above or below sixty.¹⁵

IV. ARRIVING AT COST CURVES DEDUCTIVELY RATHER THAN STATISTIcALLY

A. Relating Conviction Rates to Jury Size

In all of the above examples, the relations between the policies and the costs or benefits were determined by a statistical analysis that involved inductively eyeballing or algebraically inferring a cost curve. An alternative approach involves deducing an equation for one or more cost curves from premises that are empirically validated or intuitively accepted. A good example of that approach involves attempting to determine the effects on conviction rates of changing twelve-person juries to six-person juries. At first glance one might think an appropriate way to determine that relation would be simply to compare the conviction rates in states that use twelve-person juries with those that use six-person juries. Such an approach is likely to be meaningless, however, because any differences we find in the conviction rates may be determined by other differences, such as the characteristics of the law, the people, or the cases in the two states or two sets of states, rather than by differences in their jury sizes. This is especially true since nearly all the six-person jury systems are in the South.

¹⁵ For further detail on finding an optimum level of time consumed from arrest to trial, see Nagel & Neef, Time-Oriented Models and the Legal Process: Reducing Delay and Forecasting the Future, 1978 WASH. U.L.Q. 467 (1978). The whole process for finding the optimum level of delay which minimizes the sum of the delay costs and the speed-up costs should not take more than a couple of minutes with a $15 calculator after total costs have been related to delay. If one has access to a programmable calculator, the solution can be found in a matter of seconds because one only has to insert the guessed value for each innovation since the other numbers are stored in the calculator. One can therefore quickly observe the value of $Y$ for each guessed value of $T$ in a matter of seconds. One can thereby arrive at the optimum value for the policy variable more quickly and accurately using this repetitive guessing approach than using a more sophisticated calculus/algebra approach.
As an alternative, one might suggest making before-and-after comparisons in a single state or set of states in order to control for the kinds of characteristics which do not generally change so much over short periods of time. If the conviction rate before was 64 percent with twelve-person juries, the conviction rate afterwards with six-person juries might be substantially lower rather than higher, although most criminal attorneys would predict a higher conviction rate with six-person juries. The conviction rate might, however, fall by virtue of the fact that if defense attorneys predict that six-person juries are more likely to convict, then they will be more likely to plea bargain their clients and to bring only their especially prodefendant cases before the six-person juries. Thus the nature of the new cases, not the change in jury size, would cause at least a temporary drop in the conviction rate, and there would be no way to hold constant the type of cases heard by the new six-person juries.

As another alternative, one might suggest working with experimental juries, all of whom would hear exactly the same case. Half the juries would be six-person juries, and half would be twelve-person juries. This experimental analysis, however, involves a sample of only one case, no matter how many juries are used. Whatever differences or similarities are found may be peculiar to that one case, such as being proprosecution, prodefense, highly divisive, or simply unrealistic, and the results may thus not be generalizable. What is needed is about one hundred different trial cases on audio or video tape selected in such a way that 64 percent of them have resulted in unanimous conviction before twelve-person juries and 36 percent in acquittals or hung juries, as tends to occur in real jury trials. It would, however, be too expensive a research design to obtain and play so many trials before both a large set of twelve-person juries and a large set of six-person juries, especially if the experiment lacks representative realism.

As an alternative to the cross-sectional, the before-and-after, and the simulation approaches, we could try a deductive approach to determine the impact of jury size on the probability of conviction. Figure 6 shows in syllogistic form how such a deduction might be made. The basic premise is that twelve-person juries tend to convict 64 percent of the time and individual jurors on twelve-person juries tend to vote to convict 67.7 percent of the time. If jury decision-making involved an independent probability model like coin-flipping, then individual jurors would vote to convict 96.4 percent of the time in order for twelve-person juries to convict 64 percent of the time. If, on the other hand, jury decision-making involved an averaging model analogous to bowling, where the twelve pins tend to stand or fall depending on what happens to the average pin, then individual voters would vote to convict 64.0 percent of the time in order for the twelve-person juries to convict 64 percent of the time. Since individual jurors actually vote to convict 67.7 percent of the time, that means jury decision-making is much more like the
bowling model than the coin-flipping model, or to be more exact, it is about 1.00 to 0.13, or eight to one, more like the bowling model. That information and some simple calculations analogous to calculating a weighted average between the two models enable us to deduce (as Figure 6 shows) that if we switch from a twelve-person jury to a six-person jury and everything else remains constant, the conviction rate should rise from 64 percent to 66 percent.

**FIGURE 6 THE IMPACT OF JURY SIZE ON THE PROBABILITY OF CONVICTION**

I. Basic Symbols

\[
PAC = \text{probability of an average defendant before an average jury being convicted (empirically equals .64 for a twelve-person jury shown to two decimal places).}
\]

\[
pac = \text{probability of an average defendant receiving from an average juror a vote for conviction (empirically equals .677 for a juror shown to three decimal places).}
\]

II. Implications of the Coin-Flipping Analogy (Independent Probability Model)

\[
PAC = (pac)^{NJ}.
\]

\[
.64 = (pac)^{12}, \text{ which deductively means the coin-flipping pac is .964.}
\]

III. Implications of the Bowling Analogy (Averaging Model)

\[
PAC = pac.
\]

\[
.64 = pac, \text{ which deductively means the bowling pac is .640.}
\]

IV. Weighting and Combining the Two Analogies

Actual pac = \[
\frac{\text{[weight (coin-flipping pac) + (bowling pac)]}}{\text{(weight + 1)}}
\]

\[
.677 = \frac{\text{[weight (.964) + (.640)]}}{\text{(weight + 1)}}, \text{ which deductively means that the relative weight of the coin-flipping analogy to the bowling analogy is .13.}
\]

V. Applying the Above to a Six-Person Jury

\[
PAC = \frac{\text{[weight (coin-flipping PAC) + (bowling PAC)]}}{\text{(weight + 1)}}
\]

\[
PAC = \frac{.13 (.964)^6 + (.64)}{1.13}, \text{ which deductively means PAC with a six-person jury is .66.}
\]

VI. Applying the Above to a Decision Rule Allowing Two of Twelve Dissenters for a Conviction

\[
PAC = \frac{\text{[weight (coin-flipping PAC) + (bowling PAC)]}}{\text{(weight + 1)}}
\]

\[
PAC = \frac{.13 (.99) + (.64)}{1.13}, \text{ which deductively means that PAC with a 10/12 rule is .68.}
\]
The reason the conviction rate goes up so little when jury size is reduced from twelve to six is that jury decision-making is more like the bowling or averaging model than it is like the independent probability or coin-flipping model. The reason the conviction rate goes up at all is probably that nonconvicting hung juries decrease with a six-person jury, since holdouts are less likely to be reinforced by other holdouts than with a twelve-person jury, and the number of reinforcing supporters is more important in maintaining a holdout than the number of opponents within the six to twelve range.

B. Finding an Optimum Jury Size

Knowing the relation between jury size and conviction rates can be used as input into the empirical premises of an optimizing model designed to arrive at an optimum jury size that minimizes the weighted sum of type one errors of convicting the innocent plus type two errors of not convicting the guilty. Figure 7 summarizes what is involved in that kind of analysis. The curve of “weighted errors of innocent convicted” was arrived at by using the analysis of Figure 6, but operating on the tentative assumption that a twelve-person jury would be likely to convict an innocent person only about forty percent of the time (rather than sixty-four percent), and that about five out of one-hundred defendants tried by juries may be actually innocent. The curve of “errors of guilty not convicted” was arrived at with the tentative assumption that a twelve-person jury would be likely to convict a guilty person about seventy percent of the time (rather than sixty-four percent), and that about ninety-five out of one-hundred defendants tried by juries are probably actually guilty.

The “weighted sum of errors” curve was calculated by simply summing the other two curves after multiplying the points on the first curve by ten to indicate that convicting the innocent is traditionally considered ten times as bad as not convicting the guilty. With those tentative assumptions and that analysis, the weighted sum of errors bottoms out at jury sizes between six and eight. Justice Blackmun in announcing the Court’s decision in Ballew v. Georgia referred favorably to that analysis in deciding that it was meaningful to allow juries to be smaller than twelve, but no smaller than six. The model, however, is not particularly sensitive to that trade-off weight, since the weight can be reasonably varied and the weighted errors curve will still bottom out at about seven. The model, however, is particularly sensitive to the assumption that prosecutors operate at a .95 level of confidence in deciding whether a

---

defendant is guilty such that he or she can be brought to trial. If prosecutors are operating substantially below a .95 level, then one would need larger juries to protect those additional innocent defendants from being convicted. If prosecutors are operating substantially above a .95 level, then one could have smaller juries to increase the probability of convicting those additional guilty defendants.

Thus, the optimum jury size mainly depends on what assumption one is willing to make regarding how many jury-tried defendants out of one-hundred are truly guilty. People who favor twelve-person juries probably perceive the number of innocent jury-tried defendants as being greater than the number perceived by people who favor six-person juries. Thus, the optimum jury size ultimately depends on normative values, rather than on observed relations or relations deduced from empirical premises. Nevertheless, the combination of the predictive model of Figure 6 and the optimizing model of Figure 7 enables us to see better what is involved in choosing among alternative jury sizes. The same analysis can be applied to choosing among alternative fractions required to convict, such as unanimity, a 10/12 rule, or a 9/12 rule.19

---

19 Ballew, 435 U.S. at 235.
19 For further detail on relating conviction rates to jury size and finding an optimum jury size, see Nagel & Neef, Deductive Modeling to Determine an Optimum Jury Size and Fraction Required to Convict, 1975 WASH. U.L.Q. 933-78 (1975).
V. ARRIVING AT AN OPTIMUM LEVEL BY USING THE PREVIOUS AVERAGE POLICY

A. The Relations Between Societal Costs and Sentence Length

Figure 8 shows in theory how one should be able to arrive at an optimum sentence length through benefit-cost analysis. The figure shows the assumed relation between sentence length and holding cost, releasing cost, and total cost. As sentence length increases, the holding cost goes up at a roughly constant rate if we adjust for inflation. In other words, it costs about twice as much to hold a person in jail for ten years as it does for five years. As sentence length increases, the releasing cost supposedly goes down, since the longer a convict is held in prison the less damage he or she is likely to do upon being released. This releasing cost goes down because of an increase in maturity, deterrence, and rehabilitation with each passing year, although those effects tend to plateau out with time. The releasing cost also goes down as a result of delaying the post-release criminal behavior since that decreases the total amount of subsequent crime the defendant can commit by decreasing the remainder of his or her criminal career.

FIGURE 8 RELATING SOCIETAL COSTS TO SENTENCE LENGTH

The total cost curve simply represents the sum of the holding cost and the releasing cost at each sentence length. The total cost curve goes down at first and then goes up, as a result of (1) the positively-sloped linear relation between length and holding cost, and (2) the negatively sloped nonlinear relation between length and releasing cost. Where the total cost curve bottoms out is the optimum sentence length. That point is semantically the same as the point where one maximizes holding benefits minus holding costs, or releasing benefits minus releasing costs. Holding benefits are simply the releasing costs saved by holding a defendant. In theory, the kind of curve-drawing shown in Figure 8 makes sense. When one analyzes actual data, however, the result may not be so sensible.

Table 1 shows the results of analyzing approximately 1,000 federal
criminal cases in which we know for each case:

1. The crime for which the defendants was convicted, including the eight crimes analyzed.
2. The number of months which the defendant actually served in prison.
3. The number of months the defendant previously served in prison as a measure of the severity of his prior record.
4. The number of months the defendant subsequently served in prison for a felonious conviction after being released, as part of the data gathering of a twenty-year followup study.
5. The number of months the defendant delayed committing the felony for which he was subsequently convicted.

### TABLE 1. "OPTIMUM" SENTENCES BY CRIME AND PRIOR RECORD

<table>
<thead>
<tr>
<th>Crime</th>
<th>N</th>
<th>a</th>
<th>b₁</th>
<th>b₂</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Theft</td>
<td>313</td>
<td>9.49</td>
<td>-.20</td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>Narcotics Offense</td>
<td>125</td>
<td>213.75</td>
<td>-1.05</td>
<td>-.35</td>
<td>.26</td>
</tr>
<tr>
<td>Burglary/Larceny</td>
<td>120</td>
<td>2.17</td>
<td>-.13</td>
<td>.18</td>
<td>.08</td>
</tr>
<tr>
<td>Robbery/Kidnap</td>
<td>28</td>
<td>.17</td>
<td>.01</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>Checks/Counterfeiting</td>
<td>63</td>
<td>.26</td>
<td>.35</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Tax/Embezzlement</td>
<td>25</td>
<td>2.80</td>
<td>.57</td>
<td>-.38</td>
<td>.33</td>
</tr>
<tr>
<td>Non-Robbery Assault</td>
<td>37</td>
<td>19.82</td>
<td>-.55</td>
<td>.24</td>
<td>.21</td>
</tr>
<tr>
<td>Moonshine</td>
<td>104</td>
<td>6.85</td>
<td>-.24</td>
<td>.06</td>
<td>.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crime</th>
<th>L* (R = 10)</th>
<th>L* (R = 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car Theft</td>
<td>1.91(R)</td>
<td>2</td>
</tr>
<tr>
<td>Narcotics</td>
<td>53.85(R)</td>
<td>51</td>
</tr>
<tr>
<td>Burglary</td>
<td>.21(R)</td>
<td>0</td>
</tr>
<tr>
<td>Robbery</td>
<td>-.24(R)</td>
<td>0</td>
</tr>
<tr>
<td>Checks</td>
<td>-.02(R)</td>
<td>0</td>
</tr>
<tr>
<td>Tax</td>
<td>-.17(R)</td>
<td>0</td>
</tr>
<tr>
<td>Assault</td>
<td>24.67(R)</td>
<td>50</td>
</tr>
<tr>
<td>Moonshine</td>
<td>1.68(R)</td>
<td>2</td>
</tr>
</tbody>
</table>

Notes to the table:

b₁ = Relation between releasing costs (S²/D) and length of sentence (L).

b₂ = Relation between releasing costs and prior record (R) in the equation S²/D = a(L)b₁(R)b₂.

N = Number of cases.

r = Multiple correlation coefficient between S²/D and both L and R together.

L* = Optimum sentence in light of those relations and the specified prior record.

L* = -1/(b₁a(R)b₂) raised to the power 1/(b₁ - 1).

The results do not make sense for at least three general reasons, namely the relation between the releasing costs and the length of sentence (b₁) is not sufficiently negative, the relation between releasing costs and prior record (b₂) is not sufficiently positive, and no adjustments in the data analysis sufficiently improve b₁ and b₂ to make this approach meaningful. In more verbal, less quantitative terms, the relation between sentence length and releasing...
cost should be negative as shown in the figure. This relation, however, is positive for robbery/kidnap, checks/counterfeiting, and taxes/embezzlement. It is almost positive in burglary/larceny and moonshine. The explanation may be that “bad” defendants get longer sentences and commit the worse subsequent crimes even when we hold prior record constant because prior record is not sufficient for separating out the relatively “bad” defendants. Even if we also hold constant job record and age, however, we still do not obtain consistently negative relations between sentence length and subsequent crime severity. In other words, the longer the sentence the worse the subsequent record either because we are not sufficiently controlling for “badness,” or because long prison sentences increase rather than decrease subsequent criminal behavior.

One would likewise expect the relation between prior record and subsequent crime severity to be a positive relation in the sense that defendants with the worse prior records should have the worse subsequent records. This relation, however, is negative for narcotics offenses and taxes/embezzlement, and is almost negative for robbery/kidnap, and moonshine. When the relation is negative, the optimizing analysis in effect says to give shorter sentences to the defendants with the worse prior records. The explanation for the negative relations between prior and subsequent record may be that defendants with the worse prior records are older, and for certain crimes older defendants are not so likely to recommit serious crimes. That may be true of narcotics offenses. Another explanation may be that for certain crimes, longer prison sentences are more deterrent, and the defendants with the worse prior records get those longer, more deterrent sentences. That may be true of taxes/embezzlement, where middle-class jobs are more jeopardized. Holding constant age and job record as well as prior record, however, does not consistently generate positive $b_2$ relations. However, the worse the prior record, the better the subsequent record either because we are not sufficiently controlling for other variables, or because prison time cannot practically be used to predict subsequent severity ($S$) and delay ($D$), regardless whether the prison time is prior record ($R$) or the mean sentence length ($L$). These largely meaningless relations stay meaningless even when variations are tried on the basic data analysis, such as varying the weight of $S$ relative to $D$, using other kinds of non-linear relations to relate $S$ and $D$ to $L$ and $R$, and controlling for additional variables.

B. The Averaging Approach

A possibly more meaningful approach to arriving at determinate sentences for each of the same eight crime categories (or an alternative set of crime categories) is simply to calculate the average actual time served by convicted defendants. That reflects the collective wisdom of (1) the legislators who have written the traditional sentencing laws, (2) the judges who have imposed sentences within those statutory limits, and (3) the parole board members and the prison administrators, who can reduce judicially imposed
sentences through early release. One could argue that the collective wisdom of all those decision-makers may arrive at optimum decisions in the sense of maximizing societal benefits minus costs when one averages across all the cases involving a given crime. In any given case, any one of those decision-makers can be quite wrong, but by averaging across both the decision-makers and the cases, the errors may balance out so that the averages make sense from a benefit-cost perspective.

To use that approach in determinate sentencing, a sentencing judge (or a probation officer preparing a pre-sentence report) would calculate a predicted sentence for the crime for which the defendant was convicted, taking into consideration the defendant's prior record. The calculation would be made by using formulas like those shown in the first \( L^* \) column at the bottom of Table 2. The decision-maker would then add as much as about twenty-five percent to that base figure for aggravating circumstances, or deduct a like amount for mitigating circumstances. The adjusted figure might then be doubled to take into consideration that, under many determinate sentencing laws, the defendant gets one day off for every day of good behavior. Thus, by doubling the sentence, the defendant who behaves well serves the average amount of time, and other defendants serve more than the average.

### TABLE 2. AVERAGE SENTENCES BY CRIME AND BY PRIOR RECORD

<table>
<thead>
<tr>
<th>Crime Category</th>
<th>No Prior Record</th>
<th>Some Prior Record</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M_0 )</td>
<td>( N_0 )</td>
</tr>
<tr>
<td>Car Theft</td>
<td>15 (80)</td>
<td>21 (261)</td>
</tr>
<tr>
<td>Narcotics</td>
<td>20 (43)</td>
<td>27 (89)</td>
</tr>
<tr>
<td>Burglary</td>
<td>16 (41)</td>
<td>21 (99)</td>
</tr>
<tr>
<td>Robbery</td>
<td>37 (17)</td>
<td>71 (15)</td>
</tr>
<tr>
<td>Checks</td>
<td>16 (13)</td>
<td>22 (55)</td>
</tr>
<tr>
<td>Tax</td>
<td>10 (19)</td>
<td>30 (6)</td>
</tr>
<tr>
<td>Assault</td>
<td>38 (28)</td>
<td>46 (12)</td>
</tr>
<tr>
<td>Moonshine</td>
<td>10 (33)</td>
<td>13 (73)</td>
</tr>
<tr>
<td></td>
<td>( L^* )</td>
<td>( L^* )</td>
</tr>
<tr>
<td></td>
<td>(( R = 10 ))</td>
<td>(( R = 100 ))</td>
</tr>
<tr>
<td>Car Theft</td>
<td>14(11)</td>
<td>28</td>
</tr>
<tr>
<td>Narcotics</td>
<td>18(10)</td>
<td>23</td>
</tr>
<tr>
<td>Burglary</td>
<td>13(20)</td>
<td>16</td>
</tr>
<tr>
<td>Robbery</td>
<td>22(26)</td>
<td>41</td>
</tr>
<tr>
<td>Checks</td>
<td>15(97)</td>
<td>17</td>
</tr>
<tr>
<td>Tax</td>
<td>9(24)</td>
<td>16</td>
</tr>
<tr>
<td>Assault</td>
<td>32(43)</td>
<td>34</td>
</tr>
<tr>
<td>Moonshine</td>
<td>8(13)</td>
<td>11</td>
</tr>
</tbody>
</table>

Notes to the table:
1. Prior record is expressed in terms of months of previous imprisonment.
2. The numbers in parentheses indicate the number of cases on which those averages are based.
The main justification for an averaging approach to criminal sentencing is the one already given, namely that it represents the collective wisdom of a lot of decision-makers across a lot of cases. Those decisions are based on factual perceptions and normative values. One implication of the averaging approach is that those decision-makers have reasonably accurate factual perceptions, at least relative to alternative predictive methods, such as the kind of statistical analysis on which Table 1 is based. A second implication of the averaging approach is that those decision-makers apply values that roughly reflect what society wants and that such a democratic reflection is desirable. This does not mean that criminal justice decision-makers or other governmental decision-makers are generally perceptive or are usually likely to apply representative values. It only means that in the criminal sentencing context, they seem to be operating in line with those implications, at least collectively and relatively speaking.

In other decisional contexts, one may find a very different situation. In pretrial release, for example, judges may consistently misperceive the likelihood of defendants appearing in court, such that a statistical prediction method may be substantially more accurate there than relying on judicial perceptions. More importantly, in the pretrial release context, an individual judge may be seeking to maximize personal benefits minus costs, more so than in the sentencing context. Personal values manifest themselves in pretrial release when judges hold defendants in jail prior to trial because the judge would be more embarrassed by a releasing error than by a holding error. Society may, however, suffer more by wrongly holding people who would appear in court, than wrongly releasing people who would fail to appear. That is less of a problem in reaching a decision on length of sentence.20

The averaging approach can also be justified on the grounds that it implicitly considers the effect of sentences on deterring the general public from crime-committing, as contrasted to deterring only the specific defendant being sentenced. The benefit-cost analysis could include a statistical analysis designed to relate crime rates to average sentence-length across states or cities, but doing so produces unreliable results due to the influence of confounding variables. The averaging approach is also more politically feasible because it represents less deviation from the status quo, and is much simpler to develop, explain, and apply.21

VI. SOME CONCLUSIONS

Optimum level analysis can be applied to a great variety of legal problems where doing too much or too little is undesirable. The same general

20 See supra Section II., A.
21 For further detail on arriving at an optimum sentence length, see Nagel & Levy, The Average May Be the Optimum in Determinate Sentencing, 42 U. Pitt. L. Rev. (1981).
method\textsuperscript{22} can be used to obtain an understanding of how much due process should be provided in legal proceedings. By due process in this context, we mean safeguards for the innocent, such as rights to have an attorney, present witnesses, cross-examine one's accusers, receive reasons for the decisions

\begin{enumerate}
\item Determine the policy \((X)\) on which we are trying to arrive at an optimum level. The policy may be a composite of sub-policies.
\item Determine the goals \((Y_1, Y_2, \ldots)\) that are relevant to arriving at an optimum level on the above policy. Some of the goals may be composite goals.
\item Determine the relation between the policy and each goal. Each of those relations should ideally be capable of being expressed as an equation of the form \(Y = f(X)\), where the functional relation can be linear or various types of non-linear relations. Those equations can then be plotted as curves. One can also work backwards from plotting data to fitting a curve to the data, and then fitting an equation to the curve.
\item Sum the right side of those equations to obtain an overall goal to be minimized or to be maximized. The alternative is to sum the right side of the equations that represent benefits, and then sum those that represent costs. The overall goal is then the difference between those two sums. Where the overall goal involves subgoals that are measured on two or more different dimensions, one can (1) convert all the subgoals into money, (2) convert the subgoals into a common unit other than money, such as probabilities, months, errors, etc., and use numerical weights to indicate the relative importance of the subgoals, or (3) keep the subgoals in their original units and do a paired comparison of whether policy 1 is preferred over policy 2, with the winner playing against policy 3, and so on with the uneliminated policy being the overall winner after going through \(N-1\) paired comparisons where there are \(N\) policy alternatives. The problems of dealing with goals on multiple dimensions are discussed in Nagel, \textit{Nonmonetary Variables in Benefit-Cost Evaluation, 7 Evaluation Review 37} (1983) and “Combining and Relating Goals” in S. Nagel & M. Neff, \textit{Policy Analysis: In Social Science Research} (1979).
\item With an overall goal equation from step 4, one can now find the optimum value for the policy variable by just experimenting with different values. After each new policy value is tried, determine the numerical value of the overall goal to make sure one is moving in the right direction in guessing values for the policy-variable. After a short time with a hand calculator, one can converge on the value of the policy-variable that will minimize or maximize the goal-variable within as many decimal places as one desires.
\item If the above steps do not produce meaningful results, one might be able to arrive at a more meaningful optimum policy level by averaging the decisions that have been reached by many decision-makers in many relevant situations in the past. That approach, however, assumes that the decision-makers have goals similar to those in step 2 above. Another approach that differs substantially from steps 1 through 5 is the approach of part/whole percentaging. It involves scoring each alternative level on each of the goals using whatever scoring dimensions seem comfortable. The scores for all the alternatives on a given goal are then totaled. Each score or part is then divided by that total or whole. The best alternative is the one that has the highest sum of its unweighted or weighted part/whole percentages. See “Multiple Goals and Policies” in S. Nagel, \textit{Public Policy: Goals, Means and Methods} (1983).
\end{enumerate}
reached, be able to take an appeal, and other such rights. If too much due process is provided, it may become too difficult to establish wrongdoing, where wrongdoing has occurred. If, however, too little due process is provided, it may be too easy to establish wrongdoing where wrongdoing has not occurred. The same general methods can also be used to obtain a better understanding of how much enforcement should be applied in legal regulation. By enforcement in this context, we mean (1) how much money should be spent to secure compliance with certain legal regulations, (2) how severe the negative sanctions, fines, or jail sentences should be, or (3) how high the standards of compliance should be set. If too much compliance is demanded, the enforcement costs may exceed the benefits. If too little compliance is demanded, great societal damage might be done that could have been prevented, as for example in the field of environmental regulation. To be more specific, demanding zero water pollution for bio-degradable wastes, as the 1972 legislation recommends, may be extremely expensive, with little harm done from small amounts of such pollution. On the other hand, allowing water pollution to the extent it was allowed prior to 1972 may jeopardize public health, commercial fishing, water recreation, industrial water uses, and other water uses, when only a small effort could have made a big difference.

From the above examples, one can see that optimum levels analysis, if meaningfully applied, can enable the legal system to operate more effectively and efficiently. If society is operating below or to the left of the optimum policy level, it may be incurring excessive costs or unnecessarily missing benefits. Likewise, if we are operating above or to the right of an optimum policy level, we may be incurring other excessive costs or unnecessarily missing other benefits. Only at the optimum level is one getting the most for one's policy input. Only at that point is one minimizing the total costs when all the effects are expressed in terms of costs, or maximizing benefits minus costs when some of the effects are expressed in terms of benefits. What may be needed are more legal policy-makers and legal policy-appliers who think in terms of policy degrees and optimum policy levels, and who have some awareness of the general ideas involved in arriving at an optimum policy level. Such people benefit the legal system by making it more productive in terms of maximizing societal benefits minus societal costs.