EA & EG

WORKING PAPERS

OPTIMAL CIVIL PENALITIES

John Nash

WORKING PAPER NO. 138

February 1986

BUREAU OF ECONOMICS
FEDERAL TRADE COMMISSION
WASHINGTON, DC 20580

OPTIMAL CIVIL PENALTIES

by John Nash

Federal Trade Commission

The views expressed in this paper are those of the author and do not represent the views or policies of the Federal Trade Commission or of any individual Commissioner.

I. Introduction

The purpose of this paper is to examine in some detail the optimal strategy for setting civil penalties for a regulatory agency that is charged with regulating an industry of risk-neutral firms that impose monetarily measurable social injury when they violate the regulations. In my model, there is no uncertainty in the minds of the firms or the regulator as to what constitutes a violation. The agency's budget, which is outside of its control, determines the probability that it will be able to detect any particular violation, so this probability is exogenous from the point of view of the agency.[1] Thus, the agency's decision variable is the size of the penalty. Its desideratum is to achieve the economically efficient violation rate. That is, it wants to set a penalty structure such that the firm will not violate unless its expected gain from the violation exceeds the expected social injury. The second section of this paper describes the standard approach to this problem, explains why this approach is unsatisfactory from the practical and the theoretical points of view, and describes an alternative approach that is more satisfactory. The third section describes an

---

[1] The problem of establishing an optimal probability of detection is left to the legislative body that sets the budget. For discussion of how this optimal probability might be set, see Becker, Landes and Posner, Polinsky and Shavell, and Keenan and Rubin.

empirical implementation of this alternative approach. Section IV discusses some extensions.

II. The Optimal Penalty

The standard approach in the literature to the optimal penalty problem is to set it equal to (G/y), where $\underline{G}$ is the social injury from violation and $\underline{y}$ is the probability of being fined. This result comes from a one-period model in which the violator gains an amount $\underline{G^*}$ from the violation. Letting $\underline{P}$ represent the penalty imposed if he is caught, his net gain from violation is (G^*-P) if caught[1] and (G^*) otherwise. Thus, his expected gain is $(1-y)(G^*) + y(G^*-P)$. One can verify directly that if the penalty is set equal to (G/y), the expected gain will be positive (and the violation will occur) if and only if the gain exceeds the social injury $(G^* > G)$.

For an agency that actually has to make concrete decisions on the level of penalties, this model is deficient on the levels of both theory and practice. The theoretical problem is that this model looks upon the setting of the penalty as a one-period problem; a potential violator either violates or does not, and there is some set probability of catching him during the period. This is clearly an unrealistic model for the type of on-going multi-period infractions regulated by agencies such as

[1]In this paper, I assume that firms "caught" or "detected" are fined with probability 1.

the Federal Trade Commission. Rather than being one-time violations, as are many crimes, these regulatory violations, once they begin, typically continue until the violator is caught. It is this kind of violation with which this paper is primarily concerned. As I will show later, the one-period model is equivalent to a multi-period model, where the probability of detection remains the same each period. Because of this equivalence, the solution that comes out of these models--the penalty should equal the social injury divided by the probability of detection--is correct if and only if the periodic probability of detection remains constant over time. For some classes of violations, theory and evidence indicate that this may be a good approximation of reality, but for others it is not. For example, many of the Federal Trade Commission's consumer protection cases come from consumer complaints. The likelihood that a firm will be investigated in any given period is positively related to the <u>cumulative</u> number of complaints in its file as of that period. When investigations are based on the number of accumulated complaints, the chance of detection is virtually nil when a violation begins, since complaints do not begin to arrive immediately at the FTC, but this chance increases each year the violation continues and complaints accumulate. If the probability of detection does not remain constant over time, then the penalty calculation based on dividing injury by "the" probability of detection makes no sense, because there is no unique number for this probability.

One could in principle adapt the one-period model to a multi-period context by varying the penalty depending on the violation length in such a way that the standard equality is met in each case, i.e., $P(T) = \frac{G}{y(T)}$, where $y(T)$ is the probability of being caught at time $\underline{T}$ after violation begins. This would mean that, given any length of violation, it would be profitable to continue the violation one more period if and only if $\underline{G}^* > \underline{G}$. Suppose, for example, that the potential injury from violation is $100 per year, and that the probabilities of detection in the first three years of violation are $\frac{1}{5}$, $\frac{1}{4}$, and 1. Then, the appropriate penalty structure for this firm using this approach would be $500, $400, and $100, depending on whether the violation is caught during the first, second, or third year of violation. With such a structure, a firm that has never violated, but is contemplating a violation of one year would find positive expected profits from doing so only if $\underline{G}^* > \underline{G}$. Likewise, for a firm that had violated one year or two years. It is intuitively plausible, and easily shown, that such a penalty structure also makes the expected net gain for a strategy of 'permanent' violation positive if and only if $\underline{G}^* > \underline{G}$. However, while such a variable penalty structure is optimal for a certain class of violation[1], it presents several problems in theory and at least one in practice.

---

[1] For violations that involve a sequential "violate--do not violate" decision each period, the variable penalty structure is superior to a fixed penalty. For violations that involve a once-
(footnote continued)

One problem, of course, is that by recognizing that the periodic probability of detection increases over time, this approach prescribes a penalty that is inversely related to the actual length of the violation. The firm that is caught with its hand in the cookie jar on its first attempt is thus fined an amount approaching or equal to its wealth, while the firm that is finally caught after sneaking cookies for many years is only fined the value of one cookie. The punishment, being tailored to deter prospective violations, does not seem to fit the actual crime. For this reason, it seems extremely dubious that this kind of model would ever be acceptable as a framework for public policy.

In addition, a regime of penalties that depend inversely on the length of violation would be certain to raise the cost of enforcement. Once a violation was caught, the regulatory agency would need to investigate in order to establish the length of the violation to set a penalty. This in itself would be costly. Furthermore, the firm, once it decides to violate, would want to make it appear that its violation had been on-going for a long time, so that when it is caught, its fine will be lower. For many types of violation, it could do this at low cost by minor changes in its old files as soon as it begins to violate. This in turn would raise even higher the cost of the investigation to

---

(footnote continues)

and-for-all decision (i.e., once the violation begins, it continues until the violator is caught, or there is an unexpected change in the gain or the penalty), the fixed or variable structures are equally efficient.

establish the true length of the violation. And if the case finally goes to trial, a variable penalty would introduce another issue into the trial, consuming legal resources as the attorneys for the violator argue (in a rather bizarre turn of events) that their client has actually been violating much longer than the regulators claim and should thus be assessed a lower fine.

Finally, a penalty structure that makes the penalty depend on the violation length also encourages the expenditure of resources to evade detection, particularly in the early periods of the violation. The model I use in the paper is not equipped to handle this issue formally, but an informal argument would be as follows: consider a violation for which there is some fixed gain (G^*) and some probability of detection that increases the longer the violation continues ($y(T)$, where $\underline{T}$ is the period of time since the violation commenced). I will compare two regimes, one in which the penalty depends on the length of the violation when caught ($P(T) = G/y(T)$), and one in which the penalty does not depend on the length of violation, but is the size required to deter a "permanent" violation. (I define the necessary penalty size more explicitly in the following section but here it is only important that it is independent of the violation length.) Call this fixed penalty $\underline{P}^*$. Assuming that $y(T)$ approaches 0 as $\underline{T}$ approaches 0 and unity as $\underline{T}$ approaches infinity, $P(T)$ is very large for $\underline{T}$ close to 0 and approaches $\underline{G}$ for $\underline{T}$ very large. Now, the marginal gain from evading detection is the opportunity value of leaving the penalty money in the bank for one more period (rP^*

or $rP(T)$ for a fixed or variable penalty, respectively, where *r* is the discount rate), plus the direct gain from another period of violation (G^* in either penalty regime), minus the difference between the penalty next period and the penalty this period (0 for a fixed penalty, negative for a penalty that decreases with the length of violation). So the marginal value of evading detection is $(rP^* + G^*)$ for a fixed penalty and $(rP(T) + G^*-dP(T))$, with $dP(T) < 0$, for a variable penalty. For small values of *T*, the latter is clearly greater than the former, both because $P(T)$ is very large for *T* small and because of the negative $dP(T)$ term. Thus, for at least some small values of *T*, the incentive to evade detection is greater under a regime of variable penalties, and more resources will be devoted to this goal at least toward the beginning of the violation than would be devoted under a regime of fixed penalties. While the incentive to evade under a variable penalty regime will eventually (for *T* large) fall below the incentive with a fixed penalty, the net present value of resources devoted to evasion over the lifetime of the violation under the former regime will certainly be larger if the rate of discount is high.

On a practical level also, the prescription to set the penalty equal to the injury divided by the probability of detection is deficient. To implement this, one would need to know the probability of detection, the direct estimation of which has data requirements that are apparently impossible to meet. To directly estimate this magnitude would require knowledge not only of how many violators are caught each year, but also of how many were not caught, a fact which is by definition impossible to know. Other investigators have used proxies for detection probability such as the regulatory agency's budget (Block, Nold, and Sidak), but such measures are obviously inexact.

Because the standard approach to penalty calculation is based on a quantity that is not directly estimatable, and because its implications in a multi-period framework are not suitable for public policy, it seems desirable to develop an alternative approach, one that is based on an explicitly multi-period model and that immediately suggests a way to directly estimate the necessary parameters. I develop such an approach below.

In the context of a firm that violates continuously until it is caught,[1] at which time a penalty is imposed, the expected net present value of violation net of penalty is:

$$(1) \quad E\left[\int_0^T G^* e^{-rt}dt - P\, e^{-rT}\right],$$

---

[1] This formulation assumes that the decision to violate is a once-and-for-all decision, rather than a sequential period-by-period decision. One can make reasonable arguments about whether it is more realistic to model these decisions as once-and-for-all, or sequential. Different types of violations seem to fit more easily into one category or the other. As I noted earlier, this paper is mainly concerned with the former type of violation, since it is this kind of violation with which many regulatory agencies (as opposed to law enforcement agencies) are concerned. However, given the public policy problems discussed above that would come from a variable penalty structure, the fixed penalty that comes from the formulation in (1) may be the best feasible, even in the case of the latter type of violation.

where $\underline{E}$ is the expectations operator, $\underline{T}$ is the number of periods the violation continues before the firm is caught and fined, $\underline{G}^*$ is the gain per period from violation, $\underline{r}$ is the appropriate discount rate, and $\underline{P}$ is the penalty. Performing the integration, (1) is equal to:

(2) $G^*[(1 - E\, e^{-rT})/r] - P\, E(e^{-rT})$

The optimal penalty P^* is:

(3) $P^* = G\,[(1 - Ee^{-rT})/rE(e^{-rT})]$,

where G is the social injury from violation. The optimality can be verified by substituting $\underline{P}^*$ for $\underline{P}$ in (2), indicating that the expected gain net of penalty will be positive (and the violation will occur) if and only if $\underline{G}^* > \underline{G}$. The variable $\underline{T}$ is the "waiting time" until detection, and is regarded here as a random variable from the viewpoint of the firm and the regulatory agency. The term Ee^{-rT} is, of course, the moment generating function for the random variable $\underline{T}$, and its value can be easily looked up in any standard statistical text (e.g., Mood, Graybill and Boes, Appendix B) for any given distribution of $\underline{T}$. Since $\underline{T}$, the time until detection, is historically observable for different classes of cases, we should in principle be able to estimate its distribution.

First, consider a special case. Suppose that the probability of catching a violator each (very short) period is constant (independent of the violation length) and equal to $\underline{\lambda}$. Then, the length of time from the beginning of violation until the violation is caught has an exponential distribution with parmeter λ (c.f. Mood, Graybill, and Boes, p. 121). That is:

(4) $f(T;\lambda) = \lambda\ e^{-\lambda T}$,

(The frequency of such a distribution declines exponentially with T.) The moment generating function for such an exponential distribution is:

(5) $E\ e^{-rT} = \frac{\lambda}{\lambda+r}$.

Substituting (5) into (3) gives:

(6) $P^* = G[1-(\frac{\lambda}{\lambda+r})]/r(\frac{\lambda}{\lambda+r}) = G/\lambda.$

This is the same answer that comes from the familiar 1-period problem: the length of the violation is irrelevant; the penalty is always the one-period gain divided by the (constant) probability of detection. The reason this is true is that, although the problem is multi-period in nature, it is entirely repetitive. That is, the potential violator is faced with the same "game" each period, so a penalty that deters violation one period will do so in every period.

For some classes of violations, we would expect that the probability of detection really does remain more or less constant. Cases against firms that are under surveillance (for whatever reason) would not generally be triggered by complaints, but rather by the (probably random) checking or surveillance process. It would seem, therefore, that the one-period probability of being caught is the same for first-year violators as it is for tenth-year, although the cumulative probability, of course, is much higher over 10 years. For these cases, then, we would expect that the distribution of $\underline{T}$ would be exponential, and to compute the optimal penalty, we would use the maximum likelihood estimate of $1/\lambda$, or the mean of the sample distribution of $\underline{T}$.

For other kinds of investigations, many of which are initiated in response to cumulative complaints, the probability of detection may increase with the length of violation. Then, the distribution of $\underline{T}$ would not be exponential, and the frequency might not even be a monotonically declining function of $\underline{T}$. However, using equation (3), we can still easily develop an optimal penalty formula by examining the sample distribution of $\underline{T}$ to determine what sort of distribution seems to fit, estimating the parameters of the distribution from the sample distribution, then substituting these estimated parameters into the moment generating functions in (3).

For example, we show in the next section that for one very large class of Federal Trade Commission cases (Section 5 redress cases), the frequency of $\underline{T}$ appears to be unimodal. One candiate for modelling this distribution is the gamma distribution, with parameters $\underline{\gamma}$ and $\underline{\lambda}$, i.e.:

$$(7) \quad f(T;\ \lambda,\gamma) = \lambda^{\gamma} T^{\gamma-1}\ e^{-\lambda T}/\Gamma(\gamma) \qquad ;\ 0<T<\infty \quad ,$$

where $\Gamma(\gamma)$ is the gamma function, defined as $\Gamma(\gamma) \equiv \int_0^\infty x^{\gamma-1} e^{-x} dx$ for $\gamma>0$. The moment generating function for this distribution is equal to $(\lambda/\lambda+r)^\gamma$, so the optimal penalty could be computed by substituting this into (3) to derive:

$$(8) \quad P^* = G[(1-(\frac{\lambda}{\lambda+r})^\gamma)/r(\frac{\lambda}{\lambda+r})^\gamma] .$$

A maximum likelihood estimate of P* would be the solution to (8) with the maximum likelihood estimates of λ and γ substituted for λ and γ in the term in [ ].

III. Evidence and Implications

This section discusses some evidence on the distribution of T for several classes of Federal Trade Commission cases, and the implications for estimating the optimal penalty. To claim that the historical distribution of T can be used in this estimation requires certain assumptions about firm beliefs and behavior. I assume that all relevant firms believe that if they *do* violate, the distribution of their own times until detection will match the distribution of T in my sample. This essentially means that the underlying population of firms must be homogeneous in those characteristics that affect the chances of getting caught. (This is why it is important to examine the hypothesis that this distribution may be different for different classes of violations.) In particular, I assume that firms that choose to violate and eventually get caught are representative of the other firms in the population with respect to these characteristics bearing on the chances of getting caught, though not necessarily with respect to other characteristics. I do not make any specific assumption about why one firm chooses to violate and another does not, but I do assume that it has nothing to do with characteristics affecting detection.

The Federal Trade Commission has compiled a list of cases since 1974 in which the FTC has secured a civil penalty or redress, with the salient features of each case, including the dates during which the violations occurred. For different classes of cases, I tabulated the lengths of violation before detection, $\underline{T}$, and the frequency distribution in year-long intervals for these different cases is reported in tables 1-4. In each table, I also report the cumulative sample distribution and the frequency and cumulative distributions of a theoretical distribution, with parameters estimated by maximum likelihood methods, that seems to fit the sample. The results are interesting not only because of what they imply about the probability of detection and appropriate penalty multipliers for different classes of cases, but also because they seem to support the hypothesis that the probability of detection depends on the length of violation for Section 5 cases, but not for some other classes of cases.

Table 1 displays the information described above for order violation cases, that is, cases brought against companies for non-compliance with provisions of consents or litigated court orders. The sample distribution of $\underline{T}$ looks similar to an exponential distribution, being more or less downward sloping. Neither of two goodness-of-fit tests would allow rejection of the hypothesis that the sample of $\underline{T}$ was drawn from an exponential distribution with parameter equal to the inverse of the sample average $\underline{T}$.[1] Since an exponential distribution for $\underline{T}$ would be produced if and only if the periodic probability of detection is independent of the length of violation, this evidence can also be viewed as support for the proposition that this independence holds. The corollary is that the first-best solution to the optimal penalty problem is to make penalties equal to the periodic social injury multiplied by the average length of violation as an estimate of $(1/\lambda)$.[2] The figures in table 1 suggest that the appropriate estimate of $(1/\lambda)$ is 3.10 years for this class of case.[3]

Table 2 contains similar information for cases based on Commission determinations, that is, cases brought against companies for continued violation after receipt of notification that their practices were unfair or deceptive, as determined in

---

1 The two tests were the Kolmogorov-Smirnov test and the chi-square goodness-of-fit test. The former is a test of the simple hypothesis that a sample is drawn from a specific distribution with a pre-specified parameter(s). The latter is a test of the composite hypothesis that the sample is drawn from a class of distributions with the parameter(s) estimated from the sample.

2 The average length of violation in the sample is the maximum likelihood estimate of the inverse of the parameter λ in equation (6).

3 However, there are reasons to believe that this estimate is biased downward by around 14%. A better estimate would be around 3.60 years. See the Appendix for a discussion of this bias.

previous Commission orders.[1] The sample is considerably smaller, and the average violation length shorter than for order violation cases,[2] but once again the sample distribution resembles a drawing from an exponential distribution. Neither the Kolmogorov-Smirnov test nor the chi-square goodness-of-fit test would reject the null hypothesis that the sample was drawn from an exponential distribution with parameter equal to the inverse of the sample average.

Table 3 displays a summary of the data we have on cases that are based on Section 5 of the FTC Act, which prohibits unfair or deceptive practices. These cases are frequently the result of investigations based on consumer complaints. In general, holding firm size constant, the more complaints the FTC has received as of some given time, the more likely is an investigation to be launched at that time. Because there is a lag between the beginning of a firm's injurious practices, and the time when consumers start complaining, and because the number of complaints in a company's file accumulates over time, it would appear that the periodic probability of detection is a positive function of the length of violation for this class of cases. As we noted before, when this probability is a function of the violation

---

[1] The FTC is able to seek civil penalties against firms that act with the knowledge that their practices are unfair or deceptive. The usual practice has been to send firms a synopsis of a previous case in which the Commission determined that their practices were unfair or deceptive. If the firm is detected in violation after receiving the synopsis, a case is brought seeking civil penalties.

[2] The average in the sample is 1.46 years. There are reasons to think this may be a slightly downwardly biased estimate of the true expected waiting period. A better estimate would be around 1.54 years. See the Appendix for a more detailed discussion.

length, the distribution of the random variable $\underline{T}$, the length of violation before detection, is not exponentially distributed, and may be unimodal. Indeed, that is the form of the sample distribution in Table 3. In fact, in contrast with the other cases, both the Kolmogorov-Smirnov and chi-square goodness-of-fit tests would reject at high confidence (greater than 99 percent) the hypothesis that this sample was drawn from an exponential distribution with parameter equal to the inverse of the average value of T, (its maximum likelihood estimate under this hypothesis). Table 3 also displays the frequency and cumulative distributions of a distribution that would seem to be a better candidate for describing the distribution of $\underline{T}$. It is a gamma distribution, with the two parameters estimated by maximum likelihood methods from the sample. Neither of the two goodness-of-fit tests would reject (with 90 percent confidence) the hypothesis that the sample was drawn from this distribution. Substituting the maxi-likelihood parameter estimates ($\gamma = 1.940$, $\lambda = 0.517$) into equation (8), and using an interest rate of 0.05, gives a multiplier figure of 3.92 for calculating the penalty. This figure is fairly insensitive to the interest rate used: $\underline{r} = 0.10$ gives a multiplier of 4.09, while $\underline{r} = 0.025$ gives a multiplier of 3.84.[1]

[1] Applying L'Hopital's rule to the multiplier in (8) shows that the limit of $[(1-(\frac{\lambda}{\lambda+r})^{\gamma})/r(\frac{\lambda}{\lambda+r})^{\gamma}]$ as $\underline{r}$ approaches 0 is equal to γ/λ. For $\underline{\lambda} = 0.517$ and $\underline{\gamma} = 1.940$, this limit is 3.75, which is the sample average of $\underline{T}$. As we would expect, if the future is undiscounted, the penalty should just equal the periodic gain times the number of periods a violator could on average expect to remain undetected.

Finally, Table 4 displays the evidence from a smaller sample of another class of cases brought by the FTC - cases based upon practices that either violated the Truth in Lending Act or that were judged by the Commission to be unfair or deceptive because of concerns closely related to the Truth in Lending Act. This sample is interesting because it is the most homogeneous of any of the classes of cases reported here in terms of the specific violation upon which each case is based. For this group of cases, we would not reject the hypothesis that the sample came from an exponential distribution. The appropriate multiplier for these cases would be 1.86 years.

The implications of this evidence are clear. There do seem to be differences among classes of cases with respect not only to the "average" probability of catching violators, but also with respect to whether this probability remains constant over the period of the violation. The evidence indicates that for some classes of cases, for example, those that depend on complaints for the initiation of the investigation, the probability of detection should not be treated as constant.

IV. Extensions

In this section, I discuss several extensions of this analysis, including the implications of generalizing the model to allow for the possibility that firms may have finite and random lives (with the end of the firm's life meaning that the firm no longer gains from violating, nor can it be caught). I first discuss the implications for the theoretical calculation of the optimal penalty, and then the implications for the empirical estimation of the necessary parameters. At first blush, one might conjecture that this modification in the model would mean that the optimal penalty should be larger, since there is some chance that the firm will not be caught in its life-time and so will avoid paying any penalty. One might also conjecture that the method used in the empirical section to estimate the optimal penalty would give downwardly biased results. For the case of constant probability of detection, I show that the first conjecture would be wrong and the second correct, though the bias in the estimate may not be serious.

In examining the theoretical implications, it is convenient to consider first a world populated with firms with finite and fixed lives, fixed in the sense that the life-time of each firm is the same and is known to both the firm and the regulator. Call the life-time z. The desideratum, once again, is to devise an optimal penalty P^* such that the expected present value of the gain from violation will exceed the expected present value of the penalty if and only if the periodic gain to the firm (G^*) exceeds the periodic social injury (G).

As in Section II, let r be the discount rate (per period), λ be the (constant) probability of detection per period, and T be the number of periods before the violation is detected. (This means that T has the distribution shown in equation (4) except that it is truncated at z.) In this formulation of the problem, for the firm to be caught requires that $\underline{T} \leq \underline{z}$. So, for a firm prospectively considering a strategy of violation, the present value of the penalty the firm will pay is:

$$(9) \quad PVP = P\,e^{-rT} \text{ if } T \leq z$$
$$= 0 \quad \text{if } T > z.$$

If $\underline{T}$ is exponentially distributed, the expected present value of the penalty is therefore:

$$(10) \quad E(PVP) = \int_0^z (P\,e^{-rT})\,\lambda\,e^{-\lambda T}dT + \int_z^\infty (0)\lambda e^{-\lambda T}dT$$
$$= P\left(\frac{\lambda}{r+\lambda}\right)\,(1-e^{-(r+\lambda)z})$$

The firm gains an amount $\underline{G}^*$ each period until it is caught, or it expires, whichever comes first. After it expires, its "gain" is 0 for each period thereafter. The present value of the firm's total gain is:

$$(11) \quad PVG = {}_0\!\int^T G^*\,e^{-rt}dt = \frac{G^*}{r}\,(1-e^{-rT}) \qquad \text{if } T \leq z$$
$$= {}_0\!\int^z G^*\,e^{-rt}dt = \frac{G^*}{r}\,(1-e^{-rz}) \qquad \text{if } T > z.$$

The expected present value of the gain from violation is therefore:

$$(12) \quad E(PVG) = {}_0\!\int^z \left(\frac{G^*}{r}\right)(1-e^{-rT})\,\lambda e^{-\lambda T}\,dT + \left(\frac{G^*}{r}\right)(1-e^{-rz})\,{}_z\!\int^\infty \lambda e^{-\lambda T}dT$$
$$= \left(\frac{G^*}{r}\right)\Big[1-e^{-\lambda z} - \left(\frac{\lambda}{r+\lambda}\right)\,(1-e^{-(r+\lambda)z})$$
$$+ (1-e^{-rz})(e^{-\lambda z})\Big]$$

$$= \left(\frac{G^*}{r}\right)\left[\left(1 - \frac{\lambda}{r+\lambda}\right)\,\left(1 - e^{-(r+\lambda)z}\right)\right]$$

$$= \left(\frac{G^*}{r}\right)\left[\left(\frac{r}{r+\lambda}\right)\ (1 - e^{-(r+\lambda)z})\right]$$

$$= \left(\frac{G^*}{r+\lambda}\right)\ (1 - e^{-(r+\lambda)z})$$

The optimal penalty is:

(13) $P^* = \left(\frac{G}{\lambda}\right)$.

The optimality can be verified by noting that when $\underline{P}^*$ is substituted for $\underline{P}$ in (10), a comparison of (10) and the last equality in (12) shows that the expected present value of the gain exceeds the expected present value of the penalty if and only if $G^* > G$. Comparing (13) with (6), it can be seen that the optimal penalty is the same, whether the firm's life is finite or infinite. This is easily generalizable to the case when the firm's life-time is random. Suppose that $\underline{z}$, instead of being fixed, is random with a frequency distribution $g(z)$. Then the expectations in (10) and (12) would need to be taken over all possible values of $\underline{z}$, as well as over all possible values of $\underline{T}$. That is, the expected present value of the penalty would be:

$$(10') \quad E(PVP) = {}_{o}\!\int^{\infty} P\left(\frac{\lambda}{r+\lambda}\right)\ (1 - e^{-(r+\lambda)z})\ g(z)dz$$

$$= P\ \left(\frac{\lambda}{r+\lambda}\right)\ {}_{o}\!\int^{\infty} (1 - e^{-(r+\lambda)z})\ g(z)dz$$

The expected present value of the gain from violation would be:

$$(12') \quad E(PVG) = {}_{o}\!\int^{\infty} \left(\frac{G^*}{r+\lambda}\right)\ (1 - e^{-(r+\lambda)z})\ g(z)dz$$

$$= \left(\frac{G^*}{r+\lambda}\right)\ {}_{o}\!\int^{\infty} (1 - e^{-(r+\lambda)z})\ g(z)dz.$$

One can easily verify that the optimal penalty in (13) assures that E(PVG) > E(PVP) if and only if $\underline{G}^* > \underline{G}$, even in this case when the lifetime if the firm ($\underline{z}$) is random. The reason for the seemingly counter-intuitive result that the penalty for a firm with a finite lifetime is the same as if the firm were infinitely lived is as follows: Truncation of the firm's life, whether at a random or a fixed time, diminishes the probability that the firm will be fined, but also diminishes the expected value of the violation proportionately, since the firm gets no gains from violation after it expires. The fine necessary to keep it from violating unless $\underline{G}^* > \underline{G}$, therefore, remains unchanged.

Now, consider the statistical question of how the estimator of the optimal penalty used in Section III would be affected if the population of firms from which the estimate is made is finitely lived, with random lifetimes. The optimal penalty in Section III (for cases that are presumed to have a constant probability of detection) was estimated by multiplying the gain per period by the sample average time until detection. This latter value is the maximum likelihood estimate of $(\frac{1}{\lambda})$, the multiplier in (6). The question being considered here is whether the sample average is a good estimator of $(\frac{1}{\lambda})$ if firms are finitely lived. It would seem that this would be a downwardly biased estimate in this case, because large values of $\underline{T}$ are less likely to be observed, having been removed from the sample by attrition (expiration). This is indeed the case. If the probability of a firm's expiration is the same each period, so that the distribution of a firm's lifetime is exponential, then the size of the bias can be exactly derived. Suppose that the probability that a firm will expire in any given period is β, so that its lifetime is distributed as:

(15) $g(z) = \beta\, e^{-\beta z}$.

What we seek is the unconditional expected value of the sample average time until detection, which is what we suggested using as an estimator of $(\frac{1}{\lambda})$ in the optimal penalty calculation (equation (13) above). This expected value can be calculated by first calculating the expected value of T conditional upon z, and then integrating over all possible value of z. First, consider the conditional distribution of T given z. For a given value of z, no value of T greater than z will be observed, since the firm would have been removed from the sample. The probability of observing any value of T less than or equal to z, however, is the unconditional frequency of T divided by the cumulative distribution of T evaluated at z. That is, the conditional distribution of T, given z, can be written:

$$(16) \quad f(T|z) = \frac{\lambda e^{-\lambda T}}{(1 - e^{-\lambda z})} \qquad \text{for } T \leq z$$
$$= 0 \qquad \text{for } T > z.$$

Thus, the conditional expectation of T, given z, is:

$$(17) \quad E(T|z) = \int_0^z \frac{T\lambda e^{-\lambda T}}{1 - e^{-z\lambda}}\, dT = \frac{\frac{1}{\lambda}(1 - e^{-\lambda z}) - ze^{-\lambda z}}{(1 - e^{-\lambda z})}$$

$$= \frac{1}{\lambda} - \frac{z\, e^{-\lambda z}}{(1 - e^{-\lambda z})}$$

The unconditional expectation is:

$$(18) \quad E(T) = \int_0^\infty E(T|z)\, \beta e^{-\beta z}\, dz$$

$$= \int_0^\infty \left(\frac{1}{\lambda}\right) \beta\, e^{-\beta z} dz - \int_0^\infty \frac{z\, e^{-\lambda z}\, \beta e^{-\beta z}}{(1 - e^{-\lambda z})}\, dz$$

$$= \frac{1}{\lambda} - \beta \int_0^\infty \left[\frac{z\, e^{-(\lambda+\beta)z}}{(1-e^{-\lambda z})}\right] dz$$

Since $(\frac{1}{\lambda})$ is what we are trying to estimate, the last term of the right-hand side of the last equality in (18) shows the size of the bias. Obviously, the bias depends on the true values of $\underline{\beta}$ and $\underline{\lambda}$. A realistic estimate of $\underline{\beta}$ seems to be about 0.004 for a sample population of all commercial and manufacturing firms in the economy.[1] For a true value of $(\frac{1}{\lambda})$ of 3.1 (as estimated for order violation cases), the bias would be about 2 percent of the true value. For classes of firms with an average attrition rate, the bias does not seem to be very serious. If, however, we were sampling from a population of firms whose probability of "expiring" each year were substantially higher (perhaps because they

---

[1] This is the average fraction of industrial and commercial failures in the period 1970-1980, as a proportion of all industrial and commercial firms in business in each year. This seems a reasonable proxy for the probability that a firm will go out of business in a given year. Source: Statistical Abstract, p. 535.

act in bad faith, frequently declaring bankruptcy and re-organizing to avoid debtors or suits by injured consumers or regulatory agencies), this bias could be more of a problem. For values of $\underline{\beta}$ = 0.05 and $(\frac{1}{\lambda})$= 3.1, for example, the bias in the estimate of $(\frac{1}{\lambda})$ is about 21 percent of the true value.

The evidence discussed in this paper raises two interesting questions that were not answered here, but could be addressed in the framework set out. One question is whether the FTC has indeed imposed optimal fines historically. To answer this would require a difficult, but not impossible, investigation, since it would require a case-by-case calculation of the social injury from violations. The second question relates to Altrogge and Shughart's observation that FTC penalties have been regressive in the sense that, other factors equal, small firms tend to have larger penalties (in proportion to firm size) imposed than do large firms. They recognize that one possible explanation for this phenomenon is that the probability of detection is smaller for small firms. Whether this is true could be tested by dividing the samples I used in this paper into subsamples by size of firm and performing the same analysis I performed on the whole sample, or just by testing for a correlation (perhaps by regression techniques) between firm size and length of violation before detection.

V. Conclusion

The purpose of this paper was to develop a model for calculating penalties that would be theoretically and practically more appropriate for multi-period violations than the one-period model that is currently the only model in the literature. The model is designed to accommodate classes of violations for which detection probabilities change over time, and there is some evidence to suggest that this class is important. While the model was developed specifically in the context of a violation that, once begun, continues until the violator is caught[1], it may have more general applicability to virtually all multi-period violations, even those that involve sequential decisions, given that the alternative penalty structure for such cases (a penalty inversely related to the length of the violation) is infeasible. An important implication of the model is that the size of the penalty that is socially optimal depends on characteristics of the class of case being considered, and in particular on whether the periodic probability of detection is independent of the violation length. Finally, extensions of the basic model suggest that it is theoretically appropriate for multi-period violations with constant probability of detection, even when violators have finite lives, and that the bias produced by the empirical methods suggested here is not likely to be serious in this context.

---

[1] Or until some unanticipated changes occur in the gain or the penalty.

TABLE 1.--Length of Violation for Order Violation Cases

Length (years)	Number of Cases	Actual Frequency	Theoretical Frequency*	Actual Cumulative	Theoretical Cumulative*
0-1	12	0.273	0.276	0.273	0.276
2	8	0.182	0.200	0.455	0.476
3	6	0.137	0.144	0.592	0.620
4	8	0.182	0.105	0.774	0.725
5	5	0.114	0.076	0.888	0.801
6	1	0.023	0.055	0.911	0.856
7	1	0.023	0.040	0.934	0.896
8	0	0	0.029	0.934	0.925
9	1	0.023	0.020	0.957	0.945
>9	2	0.045	0.055	1.002	1.000
Total	44				

Average: 3.10 years.

* From an exponential distribution with parameter (λ = 0.323) estimated by maximum likelihood.

** Q = 5.92 $\delta(8)>0.50$ $\delta(9)\cong 0.75$

K-S = 0.09 $\delta>>0.20$

is the X^2 goodness-of-fit statistic, $Q = \sum_{i=1}^{y} (\frac{\hat{Z}_i - Z_i}{\hat{Z}_i})^2$, where y is the number of partitions into which the sample is divided, $\hat{Z}_i$ is the predicted number of observations falling in partition i, and Z_i is the actual number of observations in partition i. If r is the number of parameters estimated, Q has distribution bounded between a $X^2(y-1-r)$ and a $X^2(y-1)$ on the null hypothesis regarding the distribution from which the sample was drawn. For a discussion of this statistic and this technique, see, e.g. Mood, Graybill and Boes, p. 442 ff. The K-S statistic is $\max_x (|\hat{F}(x)-F(x)|)$, where $\hat{F}(x)$ is the theoretical cumulative distribution and F(x) is the actual cumulative distribution. The x's are chosen so as to partition the theoretical distribution by decile. For more discussion of this statistic and this technique, see, e.g., Fischer. The number δ is the probability that the sample statistic would exceed its reported value, given that the null hypothesis is true. It is thus the probability of error if we rejected the null hypothesis based on this evidence. For the X^2 statistic, two δ's are given, one for $X^2(y-1-r)$ and one for $X^2(y-1)$.

TABLE 2.--Length of Violation for Commission Determinations (Mainly Synopses)

Length (years)	Number of Cases	Actual Frequency	Theoretical Frequency*	Actual Cumulative	Theoretical Cumulative*
0-1	11	0.611	0.496	0.611	0.496
2	2	0.111	0.250	0.722	0.746
3	3	0.167	0.126	0.889	0.872
4	0	0	0.063	0.889	0.935
5	0	0	0.032	0.889	0.967
>5	2	0.111	0.033	1.000	1.000
Total	18				

Average: 1.46 years.

* From an exponential distribution with parameter ($\lambda = 0.685$) estimated by maximum likelihood.

$Q = 7.146$ $\delta(4)>0.10$ $\delta(5)\cong 0.25$

$K\text{-}S = 0.122$ $\delta >> 0.20$

See Table 1 for explanation of these statistics.

TABLE 3.--Length of Violation for Section 5 Redress Cases

Length (years)	Number of Cases	Actual Frequency	Theoretical Frequency*	Actual Cumulative	Theoretical Cumulative*
0-1	9	0.134	0.105	0.134	0.105
2	7	0.105	0.189	0.239	0.294
3	16	0.239	0.184	0.478	0.478
4	12	0.179	0.151	0.657	0.629
5	10	0.149	0.107	0.806	0.736
6	2	0.030	0.083	0.836	0.819
7	4	0.060	0.058	0.896	0.877
8	3	0.045	0.039	0.941	0.916
9	3	0.045	0.026	0.986	0.942
>9	1	0.014	0.057	1.000	1.000
Total	67				

* From a gamma distribution with parameters (γ = 1.940, λ = 0.517) estimated by maximum likelihood.

Q = 10.836 $\delta(7)>0.10$ $\delta(9)>0.25$

K-S = 0.120 $\delta>0.20$

For the null hypothesis that the sample came from an exponential distribution, the test statistics are:

Q(8) = 24.48 $\delta<0.005$

K-S = 0.236 $\delta<0.01$

See Table 1 for an explanation of these statistics.

TABLE 4.—Length of Violation for Truth-in-Lending Cases

Length (years)	Number of Cases	Actual Frequency	Theoretical Frequency*	Actual Cumulative	Theoretical Cumulative*
0-1	8	0.320	0.416	0.320	0.416
2	9	0.360	0.243	0.680	0.659
3	2	0.080	0.142	0.760	0.801
4	2	0.080	0.083	0.840	0.884
5	3	0.120	0.048	0.960	0.932
>5	1	0.040	0.069	1.000	1.000
Total	25				

Average: 1.86 years.

* From an exponential distribution with parameter (λ = 0.538) estimated by maximum likelihood.

$Q = 4.772$ $\quad \delta(4)>0.25$ $\quad \delta(5)\cong 0.50$

$K\text{-}S = 0.10$ $\quad \delta>>0.20$

See Table 1 for an explanation of these statistics.

REFERENCES

Altrogge, P. and W.F. Shughart II, "The Regressive Nature of Civil Penalties," International Review of Law and Economics, 1984, 4:55-66.

Becker, G.S., "Crime and Punishment: an Economic Approach," Journal of Political Economy, March/April 1968, 76:169-217.

Block, M.K., F.C. Nold, and J.G. Sidak, "The Deterrent Effect of Antitrust Enforcement," Journal of Political Economy, June 1981, 89:429-445.

Fischer, F.E., Fundamental Statistical Concepts, N.Y.: Canfield Press, 1973, p. 300 ff.

Keenan, D.C. and P.H. Rubin, "Criminal Violations and Civil Violations," Journal of Legal Studies, June 1982, 11: 365-377.

Landes, W.M. and R.A. Posner, "The Private Enforcement of Law," Journal of Legal Studies, January 1975, 4:1-46.

Mood, A.M., F.G. Graybill, and D.C. Boes, Introduction to the Theory of Statistics, 3rd ed., N.Y.: McGraw-Hill, 1974.

Polinsky, A.M. and S. Shavell, "The Optimal Tradeoff between Probability and Magnitude of Fines," American Economic Review, December 1979, 69(5):880-91.

APPENDIX: A Note on the Data Source and Potential Bias

I. The Data and Its Compilation

The data used in the estimations done in the paper came from a data set compiled by the Division of Enforcement of the Federal Trade Commission. The data set included, for each case between 1974 and 1984 involving a civil penalty or consumer redress, the dates of the violation, the amount of the penalty or redress, the relevant court dates and court numbers, the statutory authority under which the case was brought, and for many cases, a short synopsis. (For a few cases, one or more pieces of this data was missing or obviously spurious.) In estimating the dates of the violation upon which the case was based, the researcher reviewed the complete files of each case, not just the pleadings or other court records. To the extent that the investigating attorneys recorded the true violation dates for internal FTC use, these dates are the ones that are recorded in the data set. Thus, the dates should be free of the kind of systematic bias that would occur if, for example, the dates recorded were only those dates which fell within the statue of limitations, or only those dates of violations about which the attorneys had sufficient proof to withstand a court-room test. Nevertheless, the data may not be as accurate as we would like, and the results should be interpreted with a certain degree of caution. For some cases, dates of violation were only available from the pleadings; in others, the investigating attorneys may have been unable or felt it unnecessary to find out when a violation really began. These factors may have downwardly biased some of the recorded lengths of violation. In other cases, the attorneys may have recorded for internal FTC use their suspicions about when violations began, even though there was no hard evidence. This could bias the records in the opposite direction.

II. Selection Bias

In sampling from a distribution of waiting times, the estimate of the true mean waiting time will be biased downward to the extent that longer waiting times are unobservable. For example, suppose that all of the cases in the sample of order violation cases were based upon orders issued in 1980, with the data set compiled in 1984. Waiting times longer than four years would be unobservable, and the average waiting time from the data set would be a downwardly biased estimate of the true parameter. If the sample is drawn from an exponential distribution, the size of the bias is given by:[1]

$$\text{(A1)} \quad \frac{\Theta-\hat{\Theta}}{\Theta} = \left[1 + \left(\frac{\bar{\Theta}}{\Theta}\right)\right] e^{-(\bar{\Theta}/\Theta)} \qquad ; \text{ where}$$

Θ = the true waiting time in the underlying distribution;

$\hat{\Theta}$ = the expected sample waiting time; and

1 I owe this point to Russ Porter.

$\bar{\Theta}$ = the maximum observable in the sample (4 years in the above example).

While the potential exists for this kind of bias to be a problem for data of this kind, it is unlikely to seriously affect the conclusion of the current paper for the classes of cases considered herein. The largest class of cases, both in terms of the FTC's case-load and the sample used here, comprises the "Section 5" cases. These cases are based on Section 5 of the FTC act, which has been in effect since 1914.[2] The bias introduced by such a long limit on the maximum waiting time is negligible.

The second largest class of cases are those brought against a firm for violating an outstanding order which proscribes or prescribes certain conduct on its part. Here, there is more potential for bias, since the maximum observable length for a violation is the period between the issuance of the order and 1984. However, a close examination of the data not only indicates that the bias is not likely to be large, but also allows us to approximate its size and adjust the parameter estimate accordingly. First, most (about 58% of the 38 cases where proper order dates were recorded) of the cases dealt with orders issued in 1973 or earlier, meaning that the maximum observable waiting time was at least 10 years. (Only 1 of the 38

---

2 It should be emphasized that although the data set contains only cases brought between 1974 and 1984, the earlier year does not represent a limit on the beginning date of the violation. The violation could have begun at any time, and the true date is the one which would be in the data set, even if earlier than 1974.

had an order date after 1976.) Second, when the pre-1973 sample is considered alone, the estimate of the mean waiting time is not dramatically different from that of the whole sample. For the pre-1973 sample, it is 3.33 years; for the whole sample of 38, it is 2.75.[3] It is probably reasonable to believe that 3.33 is a better estimate of the true mean than is 2.75 (or 3.1, for that matter), but it may still have some bias. How much? Since the average order in the pre-1973 sample was issued about 15.1 years before 1984, we could substitute 15.1 for $\bar{\theta}$ and 3.33 for $\hat{\theta}$ in (A1) and solve for θ. The result is $\theta = 3.6$. Doing this exercise using the numbers from the full sample ($\hat{\theta} = 2.75$ and $\bar{\theta} = 12.7$) gives $\theta = 3.0$. Taken together, all of these estimates seem to indicate that the average waiting time from the larger sample, reported in the text as 3.1 years, is probably a bit low as an estimator of the true average. An unbiased estimate would probably be closer to 3.6 years.

The synopsis cases reported in Table 2 also may appear to be a problem. These cases are based on a firm's continuation of practices after the Commission sends the firm a synopsis of previously litigated cases to inform it that it must desist from the practices or be considered knowingly in violation of Section 5 of the FTC Act. In such cases, the practices would become violations upon receipt of the synopsis. Virtually all of these cases in my data set were based on synopses between 1975 and

---

3 This number (2.75) differs from the mean reported in Table 1 (3.10) because six of the cases used in Table 1 could not be used here because of missing or spurious order dates.

1979, meaning that the maximum observable waiting time is relatively short. Because the cases are so close together in time, it is not feasible to perform the same kind of anaysis on this class of cases as I did for the order violation cases (i.e., comparing a sub-sample to the full sample) to judge the severity of this bias. However, there is one good reason to believe that the bias is not serious. Generally, when synopses are sent out, they are sent to firms that staff believes to be engaging in a certain practice. These firms are then checked soon after being sent the synopsis to determine if they are in violation. (Sometimes synopses are sent to a large number of firms in a particular industry because staff believes that most of the industry is engaging in the practices, even if they do not suspect each individual firm. Even in such cases, staff usually follows up the synopsis by an investigation of each firm to which it has been sent.) Thus, it would appear that the reason for the short violation lengths is the nature of the case, rather than any selection bias. And, of course, if the true mean waiting time is close to the sample mean (1.46 years), even the relatively short maximum waiting time (5-9 years), should not introduce substantial bias. Taking the average time between synopsis and 1984 (7.3 years) as an estimate of $\bar{\Theta}$, and setting $\hat{\Theta}$ = 1.46 years, solving (A1) gives Θ = 1.54.

Some of the "Truth-in-Lending" cases reported in Table 4 are based on synopses, and the same thing that is true for the other synopsis cases is true for these, that is, detection would generally take place very quickly. Others are based on violation of the Truth in Lending Act, which was enacted in 1969. For these cases, the maximum waiting time in the sample would be on the order of 15 years. For both types of cases, therefore, the effect of the bias should be minimal.