The Optimal Use of Fines and Imprisonment when Wealth is Unobservable

by

A. Mitchell Polinsky, Stanford

THURSDAY, December 8, 2005
3:40-5:15
Room 236 Hutchins Hall

Additional hard copies of the paper are available in Room 972LR
or available electronically at http://www.law.umich.edu/centersandprograms/olin/workshops.htm
THE OPTIMAL USE OF FINES AND IMPRISONMENT 
WHEN WEALTH IS UNOBSERVABLE

A. Mitchell Polinsky*  

*Stanford Law School, Stanford, CA 94305, United States  
National Bureau of Economic Research, Cambridge, MA, United States

Abstract

This article studies the optimal use of fines and imprisonment when an offender’s level of wealth cannot be observed by the enforcement authority. I employ a model in which there are two types of offenders — a low-wealth type and a high-wealth type. The consequence of the unobservability of wealth depends on whether the enforcement authority would employ fines alone, or would also impose imprisonment sentences, if wealth were observable. In the former case, the inability to observe wealth lowers social welfare. But in the latter case, the unobservability of wealth does not lower social welfare. In both cases, offering offenders a choice of sanctions can induce high-wealth offenders to pay higher fines even though their wealth is unobservable. Specifically, a relatively high imprisonment sentence must accompany the payment of a low fine, so that high-wealth offenders will prefer to pay a higher fine and bear a lower (possibly no) imprisonment sentence.

JEL classification: D31; D62; H23; K14; K42

Keywords: Fines; Imprisonment; Public enforcement; Penalties; Unobservable wealth

* E-mail address: polinsky@stanford.edu
1. Introduction

Individuals have many ways of hiding assets from government enforcers, including by hoarding cash, transferring assets to relatives or friends, or moving money to offshore bank accounts. Notwithstanding these opportunities, prior analyses of optimal sanctions nearly always have assumed that an offender’s level of wealth is costlessly observable by the enforcement authority. Knowing wealth levels, the enforcement authority then chooses the sanctions to impose, fines and/or imprisonment sentences. The contribution of the present article is to consider optimal sanctions when an offender’s level of wealth cannot be observed by the enforcement authority.1

My analysis is based on a model in which there are two types of risk-neutral potential offenders — a low-wealth type and a high-wealth type. I first derive optimal sanctions when the offender’s level of wealth is observable, to provide a point of comparison to the outcome when wealth is unobservable. Two cases are considered, depending on whether the enforcement authority would employ fines alone, or would also impose imprisonment sentences.

If fines alone would be used if wealth were observable, higher fines would be imposed on higher-wealth offenders. The inability to observe wealth obviously precludes duplicating this outcome. When wealth is unobservable, it may be optimal to continue to rely exclusively on fines, though then the fine must be set at the wealth level of low-wealth offenders. Alternatively, it may be optimal to offer a choice between a high fine and a lower fine combined with an imprisonment sentence, with the burden of the latter combination exceeding the burden of the high fine. Given this choice, high-wealth offenders will elect to pay the higher fine even though

1 I discuss in Section 5 the most relevant prior literature on this topic — Levitt (1997), Chu and Jiang (1993), and Polinsky (2004) — and how this article relates to it.
their wealth is unobservable. Low-wealth offenders will, by necessity, choose the low fine and imprisonment sentence.

In other words, when wealth is not observable, it may be desirable to impose a costly sanction — imprisonment sentences — on low-wealth offenders in order to better deter high-wealth offenders through a cheap sanction — fines. If wealth were observable, it would not be necessary to incur this cost.

If both fines and imprisonment sentences would be employed if wealth were observable, the fines would be maximal for both low-wealth and high-wealth offenders, but the burden of the combined sanctions would be higher for low-wealth offenders. The explanation of the latter result, demonstrated below, is essentially the following. To equally deter low-wealth and high-wealth offenders requires a higher imprisonment sentence on low-wealth offenders. Consequently, there is a greater benefit from increasing the deterrence of low-wealth offenders, since there is a greater savings in imprisonment cost for each low-wealth offender deterred compared to each high-wealth offender deterred.

In this case, perhaps surprisingly, the inability to observe wealth is not detrimental. This is because, by offering the same combination of sanctions employed when wealth is observable, the same outcome can be achieved when wealth is unobservable. In particular, high-wealth offenders will prefer to pay a fine equal to their wealth and bear a relatively low imprisonment sentence rather than to pay the lower fine and bear the higher imprisonment sentence since, as noted above, the burden of the latter combination is higher. Low-wealth offenders will have to choose the combination of the lower fine and higher imprisonment sentence.

In sum, information about wealth levels is useful only if the enforcement authority would want to impose a higher burden of sanctions on high-wealth individuals than on low-wealth
individuals, for if this is the case, high-wealth individuals would pretend to be low-wealth individuals if wealth could not be observed. Conversely, if the enforcement authority would want to impose lower sanctions on high-wealth individuals, they will voluntarily bear such sanctions. Both cases can occur — the first if fines alone would be used if wealth were observable, and the second if both fines and imprisonment sentences would be employed if wealth were observable.

Section 2 presents the general framework used in this article. Section 3 examines the case in which fines alone would be used if wealth were observable, and Section 4 the case in which both fines and imprisonment sentences would be employed. Section 5 discusses the prior literature in relation to the analysis here. Section 6 considers various generalizations.

2. General Framework

I first describe the model employed when wealth is observable. Its modification to accommodate unobservable wealth is discussed at the end of this section.

Observable wealth. In the model, individuals contemplate whether to commit an offense that causes harm. Each individual is identified by the benefit he would obtain from committing the offense and by his level of wealth. For simplicity, I assume that there are two levels of wealth and that an individual’s benefit is independent of his wealth level. An individual who commits the offense is detected with a probability that is determined by the enforcement expenditures of the state. If detected, he may be sanctioned with a fine and/or an imprisonment sentence.

---

2 I discuss in Section 6 why my general results when wealth is unobservable should continue to hold if there is a continuum of wealth levels.

3 This probability is assumed to be independent of the individual’s level of wealth. This makes sense in many, if not most, enforcement contexts.
sentence. I also assume for simplicity that individuals are risk neutral with respect to fines and
imprisonment sentences, and that they bear the same disutility from a given sentence.\footnote{An
individual who is risk neutral with respect to an imprisonment sentence only cares about the expected
value of the sentence length. I explain in Section 6 why my general results would not be affected if
individuals are not risk neutral with respect to fines or imprisonment sentences, or if they differ in
terms of their disutility from time in prison.} Fines are
treated as socially costless sanctions, while the cost of imprisonment to the state is assumed to be
proportional to the length of the sentence.

The following notation will be used.

\( h = \) harm caused if the offense is committed; \( h > 0; \)

\( b = \) benefit from committing the offense; \( b \geq 0; \)

\( r(b) = \) probability density of \( b; r \) is positive for all \( b \geq 0; \)

\( w_L, w_H = \) level of wealth of low-wealth and high-wealth individuals; \( 0 \leq w_L < w_H; \)

\( \theta = \) fraction of individuals with a low level of wealth;

\( p = \) probability of detection;

\( e(p) = \) enforcement expenditures required to achieve \( p; e(0) = 0; e'(p) > 0; e''(p) > 0; \)

\( f_L, f_H = \) fine imposed on low-wealth and high-wealth individuals; \( f_L \leq w_L; f_H \leq w_H; \)

\( s_L, s_H = \) imprisonment sentence imposed on low-wealth and high-wealth individuals; and

\( c = \) cost to the state per unit of imprisonment sentence; \( c > 0. \)

Imprisonment sentences are measured in units of time such that one unit of an imprisonment
sentence corresponds to one dollar’s worth of disutility to an individual. In other words, an
imprisonment sentence of length \( s \) imposes a cost on an individual equal to \( s. \)
A low-wealth individual will commit the offense if and only if his benefit equals or exceeds \( p(f_L + s_L) \), and similarly for a high-wealth individual.\(^5\)

Social welfare is the sum of the benefits obtained by individuals who commit the offense, less the harm done, less the private and public cost of imprisonment, and less the cost of detection:

\[
\theta \int \frac{[b - h - p(1 + cs_L)]r(b)db}{p(f_L + s_L)} + (1 - \theta) \int \frac{[b - h - p(1 + cs_H)]r(b)db - e(p)}{p(f_H + s_H)}
\]

The first term reflects that every low-wealth individual whose benefit equals or exceeds \( p(f_L + s_L) \) commits the offense, resulting in a benefit \( b \) and a harm \( h \). With probability \( p \) such individuals are detected and made to bear an imprisonment sentence \( s_L \). The social cost of imprisonment is the sum of the private cost \( s_L \) and public cost \( cs_L \). The second term is interpreted similarly for high-wealth individuals. The state also incurs enforcement expenditures \( e(p) \).

The state’s problem is to maximize (1) through the choice of the fines \( f_L \) and \( f_H \), the imprisonment sentences \( s_L \) and \( s_H \), and the probability of detection \( p \), subject to the constraints that \( f_L \) and \( f_H \) cannot exceed \( w_L \) and \( w_H \), respectively.\(^6\) I do not impose an upper bound on imprisonment sentences; it will be evident below that there is no reason to expect the optimal imprisonment sentences to be maximal, so any corner solution resulting from limiting imprisonment sentences would not be particularly interesting.

---

\(^5\) I assume without loss of generality that he will commit the offense if he is indifferent.

\(^6\) Because my objective is to derive the socially optimal enforcement policy, I ignore the issue of whether the state’s policy is credible.
I assume that the state’s problem is strictly concave in the policy instruments; asterisks are used to denote the unique solution. I also assume that some enforcement is optimal, that is, \( p^* > 0 \); otherwise, the problem is uninteresting.

Unobservable wealth. When wealth is unobservable, the state is assumed to know the distribution of wealth levels among the population, but not the wealth level of a particular individual. The state might then impose a specific fine and imprisonment sentence on all offenders, designated \((f, s)\), where \( f \leq w_L \).\(^7\) Social welfare in this case is

\[
\int_{p(f+s)}^{\infty} \left[ b - h - p(1 + c)s \right] r(b) db - e(p). \tag{2}
\]

Alternatively, the state might offer offenders a choice among combinations of sanctions; the combinations will be designated \((f_1, s_1), (f_2, s_2), \text{ and so forth, ordered from the lowest to the highest fine. Social welfare in this case is discussed below.}

3. Fines Alone Would be Used if Wealth Were Observable

In this section, I first characterize the optimal enforcement system when fines alone would be used if wealth were observable (say because imprisonment has high social costs), and then describe the optimal enforcement system when wealth cannot be observed.

**Proposition 1.** (wealth observable).\(^8\) Assume that wealth is observable and that the optimal enforcement system would rely exclusively on fines to sanction offenders. In this case,

(a) the optimal fine on low-wealth offenders equals their wealth level: \( f_L^* = w_L \); and

---

\(^7\) I assume, without formally incorporating this assumption into the model, that when wealth is unobservable the state can impose a fine on everyone of up to and including \( w_L \) (for example, by threatening to impose an additional imprisonment sentence greater than \( w_L \) if this fine is not paid).

\(^8\) The results in this proposition were demonstrated by Polinsky and Shavell (1984, pp. 96-97). Because the proof is short, I include it here for completeness.
(b) the optimal fine on high-wealth offenders is higher: $f_H^* > w_L$.

Proof. (i) I show in this step that $f_L^* = w_L$. Assume, to the contrary, that $f_L^* < w_L$, and first suppose that $f_H^* < w_H$. Then there would exist a $p < p^*$, an $f_L > f_L^*$, and an $f_H > f_H^*$ such that $pf_L = p^*f_L^*$ and $pf_H = p^*f_H^*$. Since the behavior of low-wealth and high-wealth individuals will not have changed, but enforcement costs will have declined, $f_L^* < w_L$ could not have been optimal. Now suppose that $f_L^* < w_L$ and $f_H^* = w_H$. Observe first that it must be that $p^*f_H^* \leq h$; otherwise, the resulting overdeterrence could be eliminated at no cost by reducing $f_H$. Since $w_L < w_H$, it must be that $p^* w_L < h$. Thus, the underdeterrence associated with $f_L^* < w_L$ could be reduced at no cost by raising $f_L$, contradicting the presumed optimality of $f_L^*$. Hence, it must be that $f_L^* = w_L$.

(ii) I demonstrate in this step that $f_H^* > w_L$. Assume otherwise, that $f_H^* \leq w_L$. Suppose first that $f_H^* < w_L$. Note that $p^*f_H = p^*w_L \leq h$, since otherwise the resulting overdeterrence of low-wealth individuals could be eliminated at no cost by reducing $f_L$. This implies that $p^*f_H < h$. But then the underdeterrence of high-wealth individuals could be eliminated at no cost by raising $f_H$, contradicting the presumed optimality of $f_H^* < w_L$. Now suppose $f_H^* = w_L$. Then the first-order condition for determining $p^*$ from (1) can be written as $(h - pw_L)[dR(pw_L)/dp] = e'(p)$, which implies that $pw_L < h$. But then the underdeterrence of the high-wealth individuals could be reduced at no cost by raising $f_H$, contradicting the presumed optimality of $f_H^* = w_L$. Therefore, $f_H^* > w_L$.

To understand this result most easily, suppose all individuals were identical and had low wealth. By well-known logic — what I will refer to as “the Becker argument” — the optimal fine would be set as high as possible, equal to $w_L$, so that the probability of detection could be set as low as possible for any given level of deterrence. Moreover, the optimal probability would result in some underdeterrence, for if $p$ were such that $pw_L = h$, a slight reduction in $p$ would save
enforcement costs but not have any adverse first-order effect on social welfare due to the increased number of offenses because the marginal offenders would be individuals whose gains equal the harm.

Now suppose there are some high-wealth individuals. If the fine applicable to them also were \( w_L \), they too would be underdeterred. Because they can bear a higher fine, it is optimal to raise the fine imposed on them to reduce the extent of underdeterrence (possibly eliminating it).

I next consider how the optimal solution differs when wealth is not observable.

**Proposition 2.** (wealth unobservable). Assume that if wealth were observable, the optimal enforcement system would rely exclusively on fines to sanction offenders. In this case, if wealth is unobservable,

(a) social welfare is lower; and

(b) the optimal enforcement system might rely exclusively on fines equal to the wealth level of low-wealth individuals: \((w_L, 0)\); or

(c) the optimal enforcement system might offer offenders a choice between a fine equal to the wealth level of low-wealth offenders combined with an imprisonment sentence, and a fine exceeding the wealth level of low-wealth offenders without an imprisonment sentence: \((w_L, s^\dagger)\), where \( s^\dagger > 0 \), and \((f^\ddagger, 0)\), where \( f^\ddagger > w_L \).

**Proof.** (i) Part (a) is obvious since it is no longer possible to duplicate the outcome when wealth is observable.

(ii) I will prove part (b) in several steps. First observe that if the enforcement authority imposes a specific set of sanctions on offenders when wealth is unobservable (as opposed to offering them a choice among combinations of sanctions), \( f^* = w_L \). If \( s = 0 \), this result follows from the Becker argument. If \( s > 0 \), this result follows from observing that, otherwise, \( f \) could be
raised and $s$ lowered so as to keep their sum constant, thereby lowering imprisonment costs without affecting deterrence.

(iii) Given $f^* = w_L$, a sufficient condition for an imprisonment sentence to be undesirable when wealth is unobservable is that the derivative of (2) with respect to $s$ is negative at $s = 0$ — that is, that

\[ [h - p\hat{p}w_L]r(p\hat{p}w_L)p\hat{p} - p\hat{p}(1 + c)][1 - R(p\hat{p}w_L)] < 0, \tag{3} \]

where $p\hat{p}$ is the optimal probability of detection when wealth is unobservable.

(iv) The hypothesis of this proposition implies that at the optimal solution when wealth is observable, the derivative of (1) with respect to $s_L$ is non-positive at $s = 0$. Assume that it is negative — that is, that

\[ [h - p^*w_L]r(p^*w_L)p^* - p^*(1 + c)][1 - R(p^*w_L)] < 0. \tag{4} \]

(v) Given (4), (3) will hold if $p\hat{p}$ is sufficiently close to $p^*$. I now show that if $\theta$ (the fraction of low-wealth individuals) is sufficiently close to 1, then $p^*$ can be made arbitrarily close to $p\hat{p}$. Let $p^*(\theta)$ be the optimal probability if wealth is observable, given $\theta$; let $SW\hat{p}(\theta)$ be the corresponding maximum level of social welfare attainable; and let $SW^*\hat{p}$ be the maximum level of social welfare attainable if wealth is unobservable. It is obvious that the limit of $SW\hat{p}(\theta)$ as $\theta \rightarrow 1$ is $SW\hat{p}(1)$ and that $SW\hat{p}(1) = SW^\hat{p}$. Thus, for $\theta$ sufficiently close to 1, the difference between $SW\hat{p}(\theta)$ and $SW^\hat{p}$ can be made arbitrarily small. This could not be true if the limit of $p^*(\theta)$ as $\theta \rightarrow 1$ is anything other than $p\hat{p}$. Thus, if $\theta$ is sufficiently close to 1, $p^*(\theta)$ can be made arbitrarily close to $p\hat{p}$, which completes the proof of part (b).

(vi) I will prove part (c) in three steps. First note that if wealth is observable and all individuals have high wealth ($\theta = 0$), the optimal fine is $w_H$ (by the arguments used in step (ii)
above). Consistent with the hypothesis of this proposition, assume that at the optimal solution when wealth is observable, the derivative of (1) with respect to $s_H$ is negative at $s_H = 0$:

$$[h - p^*(0)w_H]r(p^*(0)w_H)p^*(0) - p^*(0)(1 + c)[1 - R(p^*(0)w_H)] < 0. \quad (5)$$

(vii) Now suppose that wealth is unobservable and that offenders are offered a choice between $(w_L, w_H - w_L + \varepsilon)$, where $\varepsilon > 0$, and $(w_H, s_2)$. If $s_2 < \varepsilon$, the burden of $(w_H, s_2)$ is less than the burden of $(w_L, w_H - w_L + \varepsilon)$, so high-wealth offenders will choose $(w_H, s_2)$ and low-wealth offenders will, by necessity, choose $(w_L, w_H - w_L + \varepsilon)$. Social welfare then is

$$\theta \int_{p(w_H + \varepsilon)}^{\infty} [b - h - p(1 + c)(w_H - w_L + \varepsilon)]r(b)db \quad (6)$$

$$+ (1 - \theta) \int_{p(w_H + s_2)}^{\infty} [b - h - p(1 + c)s_2]r(b)db - e(p).$$

A sufficient condition for $s_2^* = 0$ is that the derivative of (6) with respect to $s_2$ is negative at $s_2 = 0$ — equivalently, that

$$[h - p\bar{p}(\theta)w_H]r(p\bar{p}(\theta)w_H)p\bar{p}(\theta)(1 + c)[1 - R(p\bar{p}(\theta)w_H)] < 0, \quad (7)$$

where $p\bar{p}(\theta)$ is the optimal probability of detection.

(viii) Given (5), (7) will hold if $p\bar{p}(\theta)$ is sufficiently close to $p^*(0)$. I now show that if $\theta$ is sufficiently close to 0, then $p\bar{p}(\theta)$ can be made arbitrarily close to $p^*(0)$. If $p = p^*(0)$ and $s_2 = 0$, it is clear from (6) that the limit of $SW_L(\theta)$ as $\theta \to 0$ is $SW(0)$. Thus, it must be that the limit of $SW_L(\theta)$ as $\theta \to 0$ is $SW(0)$, which in turn implies that the limit of $p\bar{p}(\theta)$ as $\theta \to 0$ is $p^*(0)$. Consequently, if $\theta$ is sufficiently close to 0, $p\bar{p}(\theta)$ can be made arbitrarily close to $p^*(0)$, completing the proof of part (c).

Because the optimal system of fines if wealth were observable would impose higher fines on higher-wealth individuals, it is obvious that the inability to observe wealth levels lowers
social welfare. It also is clear that, due to the cost of imprisonment, the optimal enforcement system when wealth is unobservable might involve the use of fines alone, with a fine equal to the wealth level of low-wealth individuals. This is especially likely if there are relatively few high-wealth individuals, or if the difference between wealth levels is small, because then the outcome when wealth is unobservable is not much different from that when wealth is observable, where it is assumed that the use of fines alone is optimal.

The notable aspect of the present proposition is that it may be worthwhile when wealth is unobservable to use an imprisonment sentence even though imprisonment is so costly that it would not be used if wealth were observable. When wealth is unobservable, there is an extra benefit from adding an imprisonment sentence to a fine equal to the wealth level of low-wealth individuals — high-wealth offenders can thereby be induced to pay a higher fine to avoid bearing the imprisonment sentence. In other words, when wealth is not observable, it may be desirable to impose a costly sanction — imprisonment sentences — on low-wealth offenders in order to better deter high-wealth offenders through a cheap sanction — fines. If wealth were observable, it would not be necessary to incur this cost.9

4. Both Fines and Imprisonment Sentences Would be Used if Wealth Were Observable

I first characterize the optimal enforcement system when fines and imprisonment sentences would be used together if wealth were observable, and then show that the optimal enforcement system when wealth cannot be observed leads to the same outcome.

9 It also seems clear that if \( w_L \) is low enough and the public cost of imprisonment \( c \) is high enough, the optimal solution when wealth is not observable may be to not enforce at all. Another logical possibility is that even if fines alone would be employed if wealth were observable, imprisonment sentences might be imposed on both low-wealth and high-wealth offenders if wealth is unobservable. I have been unable to construct an example in which this would be optimal or to prove that it cannot be optimal.
Proposition 3. (wealth observable). Assume that wealth is observable and that the optimal enforcement system would rely on both fines and imprisonment sentences to sanction offenders. In this case,

(a) the optimal fines for low-wealth and high-wealth offenders equal their respective wealth levels: \( f_L^* = w_L \) and \( f_H^* = w_H \);

(b) the optimal imprisonment sentence on low-wealth offenders exceeds that on high-wealth offenders: \( s_L^* > s_H^* \); and

(c) low-wealth offenders bear a higher burden of sanctions than high-wealth offenders: \( w_L + s_L^* > w_H + s_H^* \).

Proof. (i) To prove part (a) for low-wealth offenders, suppose \( s_L > 0 \) and \( f_L < w_L \). Then it would be possible to raise \( f_L \) and lower \( s_L \) so as to keep \( f_L + s_L \) constant without affecting behavior, but raising social welfare by reducing the cost of imprisonment. Thus, \( f_L^* = w_L \). By the same argument, \( f_H^* = w_H \).

(ii) I next prove part (c). The first-order condition with respect to \( s_H \) derived from (1) is, with \( f_H = w_H \) and after dividing through by \(- (1 - \theta)p\),

\[
[h - b_H + p(1 + c)s_H]r(b_H) - (1 + c)[1 - R(b_H)] = 0, \tag{8}
\]

where \( b_H = p(w_H + s_H) \). The derivative of social welfare with respect to \( s_L \) is

\[
\theta p \{(h - b_L + p(1 + c)s_L)r(b_L) - (1 + c)[1 - R(b_L)]\}, \tag{9}
\]

where \( b_L = p(w_L + s_L) \). Consider this derivative at the value of \( s_L \) such that \( w_L + s_L = w_H + s_H^* \), that is, at \( s_L = s_H^* + (w_H - w_L) \). At this value of \( s_L \), (9) can be written as

\[
\theta p \{(h - b_L + p(1 + c)s_H^*)r(b_L) - (1 + c)[1 - R(b_L)]\}
+ \theta p^2 [(1 + c)(w_H - w_L)]r(b_L), \tag{10}
\]
where \( b_L = p(w_L + s_L) = p(w_H + s_H) \). Since \( b_L \) at this \( s_L \) equals \( b_H \) at \( s_H \), (8) implies that the first term in (10) is zero. The second term in (10) clearly is positive, so (10) is positive at \( s_L = s_H + (w_H - w_L) \), implying that \( s_L^* > s_H^* + (w_H - w_L) \) or, equivalently, \( w_L + s_L^* > w_H + s_H^* \). This establishes part (c).

(iii) Part (b) follows immediately from part (c).

That the fines should be maximal before imprisonment is used follows from well-known logic. Specifically, since fines are socially cheaper sanctions than are imprisonment sentences, fines should be used to their fullest extent before the state resorts to using imprisonment.

The key insight from Proposition 3 is that the optimal imprisonment sentence imposed on low-wealth offenders is sufficiently higher than that imposed on high-wealth offenders that low-wealth offenders bear a greater total burden of sanctions. This result can be explained as follows (in a way that parallels the proof). Consider an imprisonment sentence \( s_L \) borne by low-wealth individuals that, combined with a fine equal to their wealth level \( w_L \), creates the same degree of deterrence as for high-wealth individuals, given their payment of a fine equal to their wealth level \( w_H \) and bearing of the imprisonment sentence \( s_H \). Obviously, the imprisonment sentence borne by the low-wealth individuals must be higher. Now raise the imprisonment sentence imposed on each group. The effect on social welfare from the additional deterrence is the same for each group since the level of deterrence initially is the same (and the distribution of benefits is assumed to be the same for both groups). But because the imprisonment sentence borne by low-wealth individuals is greater, the social benefit of raising their imprisonment sentence is greater — society saves more on imprisonment costs for each low-wealth individual deterred from committing the offense. Consequently, it is optimal to impose a higher imprisonment
sentence on low-wealth individuals than that which would equalize the deterrence of both groups.

**Proposition 4.** (wealth unobservable). Assume that if wealth were observable, the optimal enforcement system would rely on both fines and imprisonment sentences to sanction offenders. In this case, if wealth is unobservable,

(a) social welfare is not affected;

(b) the optimal enforcement system offers a choice between two combinations of sanctions, the combinations that would be imposed on low-wealth and high-wealth offenders if wealth were observable: \((w_L, s_{L}^{*})\) and \((w_H, s_{H}^{*})\);

(c) low-wealth offenders choose, by necessity, the sanctions \((w_L, s_{L}^{*})\); and

(d) high-wealth offenders choose the sanctions \((w_H, s_{H}^{*})\).

**Proof.** If offenders are given the choice between \((w_L, s_{L}^{*})\) and \((w_H, s_{H}^{*})\), part (c) is obvious, and part (d) follows from the result of part (c) of Proposition 3. Then, if the same probability of detection is chosen when wealth is unobservable as would be chosen if wealth were observable, part (a) holds, which implies that this probability and the sanction combinations \((w_L, s_{L}^{*})\) and \((w_H, s_{H}^{*})\) must be optimal when wealth is unobservable, proving part (b).

This result shows, perhaps surprisingly, that the inability to observe wealth is not detrimental when both fines and imprisonment sentences would be used to sanction offenders if wealth were observable. By offering the same combination of sanctions employed when wealth is observable, the same outcome can be achieved when wealth is unobservable. In particular, high-wealth offenders will prefer to pay a fine equal to their wealth and bear a relatively low imprisonment sentence rather than to pay the lower fine and bear the higher imprisonment sentence.
sentence since, as noted above, the burden of the latter combination is higher. Low-wealth offenders have to choose the combination of the lower fine and higher imprisonment sentence.

5. Prior Literature

To my knowledge, the only prior article to have systematically studied optimal sanctions under the assumption that an offender’s level of wealth is unobservable is Levitt (1997).10 His model differs in significant respects, however, from mine. In Levitt’s framework, “rich” and “poor” individuals are distinguished not in terms of their wealth levels (both types are assumed to be capable of paying any fine the state imposes), but instead in terms of their disutility from time in jail (rich individuals suffer greater disutility). He also assumes that the benefit from committing the offense is the same for every individual of a given type, but differs between the types, and that the benefit is less than the harm from the offense, so that ideal deterrence involves complete deterrence. In his formal analysis, Levitt considers the choice between an imprisonment sentence and a fine, not the optimal combination of the two, as I do here (though he does discuss this issue informally).

Levitt’s main point is that, in contrast to the situation when individuals’ wealth levels can be costlessly observed, the availability of fines as an alternative to imprisonment might not result in higher social welfare. This is because, in his analysis, the introduction of a fine as an alternative to a jail term cannot increase deterrence — if the fine were more burdensome, the

---

10 Several articles have studied optimal sanctions when variations in wealth are assumed to be costlessly observable. See Friedman (1981), Polinsky and Shavell (1984, pp. 95-98), Polinsky and Shavell (1991), and Garoupa (1998, pp. 484-87). (Garoupa examines the case in which the enforcement authority costlessly observes an underestimate of individuals’ wealth levels.)
offender would choose the jail term. Consequently, Levitt suggests that the case for using imprisonment sentences as sanctions is stronger than is generally appreciated.

In an article slightly predating Levitt’s, Chu and Jiang (1993) examine optimal sanctions when there are three types of individuals in terms of wealth levels and a continuum of offenses that individuals can commit, each corresponding to a different level of harm. Because Chu and Jiang assume that the fine is proportional to the level of harm, they in effect assume that the enforcement authority cannot observe offenders’ wealth levels, though this is not the focus of their attention. Their main point is that, because of marginal deterrence considerations and differences among individuals in their responses to imprisonment (wealthier individuals have a higher opportunity cost of time in prison), it may be desirable to use imprisonment sentences combined with less-than-maximal fines. Like Levitt, they emphasize the desirability of imprisonment sentences relative to fines. My analysis differs from theirs in that I do not impose any restrictions on the choice of fines (aside from wealth constraints), whereas, as noted, they assume that the fine is proportional to the level of harm. Our conclusions differ as well. I find, contrary to Chu and Jiang’s result, that imprisonment sentences should not be used unless the party subject to the sentence also makes a monetary payment equal to his wealth level.11

In a companion article — Polinsky (2004) — I derive optimal fines when an offender’s level of wealth can be determined after a costly audit. In contrast to the present analysis, I do not consider imprisonment sanctions. The focus of my other article is on characterizing the optimal

11 Also of some relevance to the present article is the analysis by Lott (1987) of the question of whether individuals should be allowed to spend freely on their defense in criminal cases. In his model, offenders bear the same imprisonment sentence regardless of their wealth level, and higher-wealth individuals are assumed to suffer greater disutility from the sentence. Lott’s argument is that allowing higher-wealth individuals to spend more on their defense reduces the overdeterrence that otherwise would occur. For further discussion of this point, see Garoupa and Gravelle (2003).
audit rate and on deriving the optimal fine for misrepresenting one’s wealth level, issues ignored here.

6. Concluding Remarks

I made several simplifying assumptions in order to keep the analysis tractable. The most notable ones are that individuals are risk neutral with respect to fines and imprisonment; that there are only two levels of wealth; and that individuals bear the same disutility from time in prison regardless of their wealth level. I discuss these briefly here and suggest why modifying them is not likely to change my principal points regarding optimal sanctions when wealth is unobservable.

Consider first the results when fines alone would be employed if wealth were observable (Section 3). If individuals are risk averse with respect to fines, the result that optimal fines increase with wealth (Proposition 1) still could hold; indeed, because consideration of risk aversion tends to lower optimal fines, this result may be reinforced if, as is commonly assumed, risk aversion declines with wealth. Moreover, if wealth is continuous over a range of values, it still could be optimal when wealth is unobservable to rely on a fine equal to the lowest level of wealth (Proposition 2, part (b)). Similarly, the argument used to demonstrate that it may be desirable to offer combinations of sanctions that induce higher-wealth individuals to pay higher fines and lower-wealth individuals to bear imprisonment sentences (Proposition 2, part (c)) holds regardless of individuals’ risk preferences with respect to imprisonment and regardless of
whether individuals bear the same disutility from time in prison independent of their wealth level.\footnote{This claim follows from the fact that the argument used to prove part (c) of Proposition 2 is a limiting argument in which the fraction of offenders bearing an imprisonment sentence goes to zero. See step (viii) of the proof.}

Next consider the results when both fines and imprisonment sentences would be imposed on offenders if wealth were observable (Section 4). The key result in this case is that, although the optimal fine rises with wealth, the burden of the fine and the imprisonment sentence combined declines with wealth (Proposition 3); hence, even if wealth is unobservable, high-wealth offenders can be induced to pay higher fines because they would bear more burdensome sanctions if they did not (Proposition 4).

Suppose that high-wealth individuals have a greater distaste for jail time than low-wealth individuals, as the other authors discussed in Section 5 assumed. The following argument shows that optimal sanctions when wealth is unobservable still will induce high-wealth offenders to pay higher fines (although the level of social welfare when wealth is observable may no longer be attainable). Given this assumption, suppose the sanctions were such that both groups chose to pay an amount of money equal to the wealth level of low-wealth individuals $w_L$ and to bear an imprisonment sentence $s_1 > 0$. This outcome could not be optimal, for it would be possible to choose some fine $f_2 > w_L$ and some imprisonment sentence $s_2 < s_1$ such that high-wealth offenders prefer $(f_2, s_2)$ to $(w_L, s_1)$ by a negligibly small margin. Hence, they could be induced to choose $(f_2, s_2)$, resulting in an increase in social welfare due to a reduction in the imprisonment sentence borne by high-wealth offenders. In other words, optimally chosen sanctions still will induce high-wealth offenders to pay a fine that exceeds the wealth level of low-wealth individuals.
If individuals are not risk neutral with respect to fines or imprisonment, the same logic implies that optimal sanctions when wealth is unobservable will lead high-wealth offenders to pay higher fines. Suppose, given a different set of assumptions about the risk preferences of low-wealth and high-wealth individuals (but assuming that all high-wealth individuals have the same preferences and all low-wealth individuals have the same preferences), that the sanctions were such that both groups chose to pay $w_L$ and to bear an imprisonment sentence $s_i > 0$. Again, this outcome could not be optimal, for it would be possible to offer another sanction combination with a higher fine, a lower imprisonment sentence, and with essentially the same total burden, that would be chosen by high-wealth offenders and result in lower imprisonment costs.

Finally, suppose wealth levels vary continuously among individuals. Based on reasoning employed in this article (specifically, step (ii) of the proof of Proposition 3), the optimal sanctions if wealth is observable should have the property that the burden of the sanctions declines with the fine paid. Thus, if the enforcement authority offers the same menu of choices to offenders when wealth is unobservable, each offender will choose to pay the highest fine he is capable of paying, and the outcome will be the same as that when wealth is observable.

**Acknowledgements**

This research was supported by the John M. Olin Program in Law and Economics at Stanford Law School. I received helpful comments from Nuno Garoupa, Jason Johnston, Louis Kaplow, Daniel Rubinfeld, Steven Shavell, Kathryn Spier, Robert Weisberg, and an anonymous referee.
References


Friedman, D., 1981. Reflections on optimal punishment, or: should the rich pay higher fines? Research in Law and Economics 3, 185-205.


Garoupa, N., Gravelle, H., 2003. Efficient deterrence does not require that the wealthy should be able to buy justice. Journal of Institutional and Theoretical Economics 159, 545-552.


OPTIMAL FINES AND AUDITING
WHEN WEALTH IS COSTLY TO OBSERVE

A. Mitchell Polinsky*

Abstract: This article studies optimal fines when an offender’s wealth is private information that can be obtained by the enforcement authority only after a costly audit. I derive the optimal fine for the underlying offense, the optimal fine for misrepresenting one’s wealth level, and the optimal audit probability. I demonstrate that the optimal fine for misrepresenting wealth equals the fine for the offense divided by the audit probability, and therefore generally exceeds the fine for the offense. The optimal audit probability is positive, increases as the cost of an audit declines, and equals unity if the cost is sufficiently low. If the optimal audit probability is less than unity, there are some individuals who are capable of paying the fine for the offense who misrepresent their wealth levels. I also show that the optimal fine for the offense results in underdeterrence due to the cost of auditing wealth levels.

* Stanford University and National Bureau of Economic Research. This research was supported by the John M. Olin Program in Law and Economics at Stanford Law School. I received helpful comments from Nuno Garoupa, Jason Johnston, Daniel Rubinfeld, Steven Shavell, Kathryn Spier, and Robert Weisberg.
I. Introduction

Despite the obvious ability of individuals to hide assets (for example, by hoarding cash or distributing property to relatives or friends), prior analyses of optimal fines nearly universally assume that an offender’s level of wealth is costlessly observable by the enforcement authority.1 The contribution of this article is to consider optimal fines when an offender’s wealth is private information that can be obtained by the enforcement authority only after a costly audit. Obviously, the conventional analysis is a special case of the present analysis, when the audit cost is zero.

I employ a model in which there is a continuum of individuals with respect to their level of wealth and their potential benefit from committing an offense. If an individual commits the offense and is caught, he is sanctioned with a fine. If he claims that he cannot pay the fine, he may be audited. If the audit determines that he misrepresented his wealth level, he can be fined for having lied about his wealth. The enforcement authority’s problem is to choose the probability of detecting the offense, the fine for the offense, the probability of an audit, and the fine for misrepresentation of wealth so as to maximize social welfare.

Among other things, I demonstrate that the optimal fine for misrepresenting one’s wealth level equals the fine for the offense divided by the audit probability, and therefore generally exceeds the fine for the offense. The optimal audit probability is positive, increases as the cost of an audit declines, and equals unity if the cost is sufficiently low. If the optimal audit probability is less than unity, there are some individuals who are capable of paying the fine for the offense who misrepresent their wealth levels. I also

---

1 See note 2 below for a discussion of the few exceptions.
show that the optimal fine for the offense results in underdeterrence due to the cost of auditing wealth levels.

Section II presents the model. Section III derives the optimal fines and audit probability. Section IV contains some concluding observations about the use of imprisonment to sanction misrepresentation of wealth and the offense.2

II. Model

In the model, risk-neutral individuals contemplate whether to commit an offense that causes harm. Each individual is identified by the benefit he would obtain from committing the offense and by his level of wealth. An individual who commits the offense is detected with some probability (which is costly for the state to maintain) and fined. If he claims that he cannot pay the fine because he has insufficient wealth, he may be audited at some cost to the state to determine his wealth level (audits are assumed to be accurate). If an audit reveals that the individual misrepresented his wealth level, he may be sanctioned further, by a higher fine. The fine imposed on an individual, whether for the offense or for misrepresenting his wealth level, cannot exceed his wealth. Fines are assumed to be socially costless to impose.

The following notation will be used.

\[ h = \text{harm caused if the offense is committed}; \ h > 0; \]

---

2 To my knowledge, this article is the first to analyze optimal fines when the offender’s level of wealth can be observed by the enforcement authority only after a costly audit. Several studies of related interest should be mentioned, however. Chu and Jiang (1993) and Levitt (1997) consider the choice between fines and imprisonment when wealth cannot be discovered by the enforcement authority at any cost (in Chu and Jiang’s case, this assumption is implicit). Garoupa (1998) investigates optimal fines when the enforcement authority is assumed to costlessly observe an underestimate of offenders’ wealth levels. See also the discussion of Polinsky (2004) in section IV below. The large theoretical literature on the auditing of wealth levels in the context of controlling tax evasion — see Mookherjee (1997, pp. 207-31) and Andreoni, Erard, and Feinstein (1998, pp. 823-35) — is only tangentially relevant to the analysis here because its focus is on controlling offenses that affect the distribution of
\[ b = \text{benefit from committing the offense}; \quad b \geq 0; \]

\[ r(b) = \text{probability density of } b; \quad r(b) \text{ is positive for all } b \geq 0; \]

\[ w = \text{wealth of an individual}; \quad w \geq 0; \]

\[ g(w) = \text{probability density of } w; \quad g(w) \text{ is positive for all } w \geq 0; \]

\[ p = \text{probability of detection}; \]

\[ e(p) = \text{enforcement expenditures of the state}; \quad e(0) = 0; \quad e'(p) > 0; \]

\[ f_O = \text{fine for committing the offense}; \quad f_O \geq 0; \]

\[ q = \text{probability of an audit}; \quad q \geq 0; \]

\[ k = \text{cost to the state of an audit}; \quad k > 0; \quad \text{and} \]

\[ f_M = \text{fine for misrepresenting one’s wealth level}; \quad f_M \geq 0. \]

The distributions of benefits and of wealth are assumed to be independent and known by the state. Without loss of generality, I assume that if an offender is found to have misrepresented his wealth level, the fine \( f_M \) is imposed instead of the fine \( f_O \) that otherwise would be applicable.

**Behavior of individuals.** First consider the decision of an individual who has been caught committing the offense whether to misrepresent his wealth level \( w \). If he does not do so, he will pay the fine for the offense \( f_O \) if \( w \geq f_O \), and \( w \) otherwise. In other words, he will pay \( \min[w, f_O] \).

If he pays \( w \) because his wealth is insufficient to pay \( f_O \) and he is subsequently audited, he will be found to have not misrepresented his wealth level.

Suppose, however, that the individual does misrepresent his wealth level, claiming he is only capable of paying \( w '< f_O \). If he is not audited, he will pay \( w' \). If he is audited, he will pay

---

3 This is the nominal fine. An individual might not have sufficient wealth to pay this fine. The same statement applies to the fine for misrepresenting one’s wealth level.
if he has sufficient wealth to pay this fine; otherwise he will pay his wealth \( w \). Thus, if he misrepresents his wealth level, his expected payment is \((1 - q)w' + q\min[w, f_M]\). If an individual misrepresents his wealth level, he will do so to the greatest extent possible, claiming that he has no wealth.\(^4\) Hence, his expected payment would be \(q\min[w, f_M]\).

Given the preceding discussion, an individual will misrepresent his wealth level if and only if\(^5\)

\[
q\min[w, f_M] < \min[w, f_O]. \tag{1}
\]

Let

\[ \hat{w} = \text{critical level of wealth below which individuals who are detected committing the offense will misrepresent their wealth levels and at and above which individuals will not misrepresent their wealth levels.} \]

It can be demonstrated from (1) that

\[
\hat{w} = \begin{cases} 
  f_O/q & \text{if } qf_M \geq f_O \\
  \infty & \text{if } qf_M < f_O. \tag{2}
\end{cases}
\]

In other words, if the expected fine for misrepresenting one’s wealth level equals or exceeds the fine for the offense, individuals with wealth less than \( f_O/q \) will misrepresent their wealth levels.

\(^4\) This occurs because the fine for misrepresentation is assumed not to depend on the extent of misrepresentation and the probability of being audited is assumed not to depend on the reported wealth level. The analysis would be much more complicated if these assumptions were not made, because it would become necessary to choose the optimal schedules for the fine for misrepresentation and the audit probability, as well as to derive the behavior of individuals with different levels of wealth in response to these schedules.

\(^5\) It is convenient to assume that if an individual is indifferent between misrepresenting his wealth level and not, he does not misrepresent his wealth level. None of the results depend on this assumption.

\(^6\) To prove this requires evaluating (1) in three cases — whether \( qf_M \) exceeds, equals, or is less than \( f_O \). Because a complete treatment of these cases is tedious, I consider only one here. Suppose that \( qf_M > f_O \). In this case, \( f_O < f_O/q < f_M \). If \( w < f_O \), (1) becomes \( qw < w \), which clearly holds. If \( f_O < w < f_O/q \), (1) becomes \( qw < f_O \), which holds because \( w < f_O/q \). If \( f_O/q \leq w < f_M \), (1) becomes \( qw < f_O \), which does not hold because \( f_O/q \leq w \). If \( w \geq f_M \), (1) becomes \( qf_M < f_O \), which does not hold because \( f_O/q < f_M \). Hence, if \( qf_M > f_O \), individuals with wealth levels less than \( f_O/q \) will misrepresent their wealth levels, while individuals with wealth levels equal to or greater than \( f_O/q \) will not.
(claiming to be zero-wealth individuals), while individuals with wealth equal to or greater than $f_0/q$ will pay the fine for the offense $f_0$. Note that if $q < 1$, there are some individuals who are capable of paying $f_0$ who choose to misrepresent their wealth levels (individuals with wealth between $f_0$ and $f_0/q$).\(^7\) If, however, the expected fine for misrepresenting one’s wealth level is less than the fine for the offense, all individuals will misrepresent their wealth levels.

Next consider the decision of an individual whether to commit the offense. Let

$$b(w) = \text{critical value of benefit below which an individual whose wealth is } w \text{ will not commit the offense and at and above which he will.}$$\(^8\)

Thus, the higher $b(w)$ is, the greater the level of deterrence.

An individual will commit the offense if his benefit equals or exceeds the expected fine he faces, which depends on whether he will misrepresent his wealth level if he is caught. Since he will misrepresent his wealth if and only if (1) holds, it follows that the critical value of benefit is

$$b(w) = p \min \{q \min \{w, f_M\}, \min \{w, f_0\}\}. \quad (3)$$

**Social welfare.** Social welfare is the sum of the benefits obtained by individuals who commit the offense, less the harm done, less the cost of detection, and less the cost of auditing. Thus, social welfare is

$$\int_{\hat{w}}^{\infty} \left\{ \int_{\min(b(w), 0)}^{\infty} [(b - h - pqk)r(b)db]g(w)dw \right\} + \int_{\hat{w}}^{\infty} \left\{ \int_{\min(b(w), 0)}^{\infty} [(b - h)r(b)db]g(w)dw - e(p) \right\}. \quad (4)$$

\(^7\) Individuals whose wealth is less than $f_0$ obviously will not pay anything because the worst that can happen to them is that they are audited and lose their wealth.

\(^8\) I assume without loss of generality that an individual commits the offense if he is indifferent.
where \( b(w) \) is given by (3) and \( \hat{w} \) by (2). The first term in (4) is the contribution to social welfare associated with individuals with relatively low levels of wealth (less than \( \hat{w} \)) who misrepresent their wealth levels and do not pay the fine for the offense \( f_0 \). Every individual in this group whose benefit equals or exceeds \( b(w) \) commits the offense, resulting in a benefit \( b \), a harm \( h \), and an expected auditing cost of \( pqk \). The second term is the contribution to social welfare associated with relatively high-wealth individuals who pay the fine for the offense \( f_0 \). The state also incurs enforcement expenditures \( e \).

**The state’s problem.** The state’s problem is to maximize social welfare through the choice of the probability of detection \( p \), the fine for the offense \( f_0 \), the audit probability \( q \), and the fine for misrepresenting one’s wealth level \( f_M \).\(^9\) The fines actually paid by individuals cannot exceed their levels of wealth. Asterisks will be used to denote the solution to the state’s problem, which I assume is unique.

Because the choice of the optimal probability of detection \( p^* \) does not bear in any interesting way on the analysis of optimal fines and auditing, I will not analyze the choice of \( p^* \) below. However, I assume that some enforcement is optimal, that is, \( p^* > 0 \); otherwise, the problem is uninteresting.

**III. Analysis**

In this section I describe the optimal enforcement system through four propositions. Following the proof of each proposition is a brief informal discussion of the result.\(^{10}\)

---

\(^9\) Because my objective is to derive the socially optimal enforcement policy, I do not consider whether the state’s policy is credible.

\(^{10}\) Although some of the proofs convey useful intuition, they may be skipped with little loss of continuity.
Propositions 1 and 2 describe the optimal fine for misrepresenting wealth, and its effect on individuals’ decisions to engage in misrepresentation, when the audit probability is, respectively, less than or equal to unity. Proposition 3 characterizes the optimal fine for the offense. Proposition 4 establishes that the optimal audit probability is positive and may be less than or equal to unity.

**Proposition 1.** If the optimal audit probability is less than unity, $q^* < 1$, then:

(a) the optimal fine for misrepresenting one’s wealth level $f_M^*$ exceeds the optimal fine for the offense; in particular, it equals the fine for the offense divided by the audit probability: $f_M^* = f_O^*/q^* > f_O^*$;\(^{11}\)

(b) individuals with wealth below the fine for misrepresentation $f_M^*$ misrepresent their wealth levels, while individuals with wealth at and above $f_M^*$ pay the fine for the offense $f_O^*$; and

(c) there are some individuals who are capable of paying the fine for the offense $f_O^*$ who misrepresent their wealth levels (individuals with wealth between $f_O^*$ and $f_M^*$).

**Proof:** (i) I first show that the optimal enforcement system is characterized by $q f_M^* > f_O^*$. Assume otherwise, that $q f_M^* < f_O^*$. If this condition were to hold, then individuals would misrepresent their wealth regardless of their level of wealth (see (2)). Consequently, everyone will be audited with probability $q$ and their expected fine, conditional on having been detected committing the offense, will be $q w$ for $w < f_M^*$ and $q f_M^*$ for $w \geq f_M^*$. It is possible, however, to raise social welfare by lowering $f_O^*$ to $q f_M^*$. If $f_O = q f_M^*$, then $f_O < f_M^*$. If $w < f_O$, (1) becomes $q w < w$, which holds. If $f_O < w < f_M^*$, (1) becomes $q w < f_O^*$, which holds because $f_O^* = q f_M^*$ and $w < f_M^*$. If $w \geq f_M^*$, (1) becomes $q f_M^* < f_O^*$, which does not hold because $f_O^* = q f_M^*$. Thus, if $f_O$ is lowered to $q f_M^*$,

\(^{11}\) As will be seen, $f_M^*$ is not unique. Any $f_M^*$ equal to or greater than $f_O^*/q^*$ is optimal. An analogous observation applies to Proposition 2 below.
individuals for whom \( w < f_M \) will continue to misrepresent their wealth levels, be audited with probability \( q \), and face an expected fine conditional on detection of \( qw \). But individuals for whom \( w \geq f_M \) will now choose to pay the fine for the offense \( f_O \) and not misrepresent their wealth levels. Since \( f_O = qf_M \), they are deterred to the same extent as before. But since they pay the fine \( f_O \), they are not audited. Thus, the behavior of individuals is not affected, but auditing costs are reduced, thereby raising social welfare. In the optimal enforcement system, therefore, it must be that \( qf_M \geq f_O \).

(ii) I next show that in the optimal enforcement system one can assume without loss of generality that \( qf_M = f_O \). Since step (i) rules out \( qf_M < f_O \), suppose \( qf_M > f_O \). Then, by (2), individuals for whom \( w < f_O/q \) would misrepresent their wealth levels and individuals for whom \( w \geq f_O/q \) would not misrepresent their wealth. Consequently, an individual in the former group will be audited with probability \( q \) and his expected fine, conditional on detection, will be \( qw \) (he is unable to pay \( f_M \) since \( w < f_O/q < f_M \)). Individuals in the latter group will pay the fine for the offense \( f_O \) (they are able to because \( w \geq f_O/q > f_O \)) and not be audited. If \( f_M \) is lowered until \( qf_M = f_O \), it is easy to see that individuals’ decisions about misrepresenting their wealth levels, and their expected fines, are not affected. Hence, \( f_M \) such that \( qf_M = f_O \) is optimal (as is any higher \( f_M \)), establishing part (a) of the proposition.

(iii) Given \( qf_M = f_O \), (2) implies that \( \hat{w} = f_O/q = f_M \). In other words, individuals with wealth below \( f_M \) misrepresent their wealth levels, while individuals with wealth at or above \( f_M \) do not. Since \( f_M = f_O/q > f_O \), the latter individuals pay the fine \( f_O \). This establishes part (b).

(iv) Part (c) follows immediately from step (iii) since individuals with wealth between \( f_O \) and \( f_M = f_O/q > f_O \) misrepresent their wealth levels.
Comment: It is easy to see why the optimal fine for misrepresenting one’s wealth level $f_M$ must be at least equal to the fine for the offense divided by the audit probability, $f_O/q^*$. If this were not the case, then the expected fine for misrepresenting one’s wealth level, $q^*f_M$, would be less than the fine for the offense, $f_O$, and no one would pay the fine for the offense. To induce individuals to pay the fine for the offense, the fine for misrepresentation must be at least equal to $f_O/q^*$. Any higher fine for misrepresentation has the same effect. Even if $f_M = f_O/q^*$, not everyone will pay the fine for the offense. In particular, individuals whose wealth is below $f_M$ are unable to pay the full fine for misrepresentation, and thus cannot be induced to pay the fine for the offense; since $f_M > f_O$, this includes some individuals who are capable of paying $f_O$. But everyone else can be threatened with a fine for misrepresentation of $f_M$ and therefore will pay the fine for the offense.

**Proposition 2.** If the optimal audit probability equals unity, $q^* = 1$, then:

(a) the optimal fine for misrepresenting one’s wealth level equals the optimal fine for the offense: $f_M = f_O$;

(b) individuals do not misrepresent their wealth levels; and

(c) everyone who is able to pay the fine for the offense $f_O$ does so.

**Proof:** The proof of part (a) parallels steps (i) and (ii) in the proof of Proposition 1. Part (b) follows from (1) given $f_M = f_O$. Part (c) follows immediately from part (b).

Comment: Clearly, if an audit is certain if one does not pay the fine for the offense, then a fine for misrepresentation equal to the fine for the offense will induce those who are able to pay the fine for the offense to do so. Individuals who are unable to pay the fine for the offense do not have an incentive to misrepresent their wealth levels since they will have to pay their wealth whether they are subject to the fine for the offense or the fine for misrepresentation.
**Proposition 3.** The optimal fine for the offense $f_0^*$

(a) is less than the harm divided by the probability of detection: $f_0^* < h/p$;
(b) results in underdeterrence due to the cost of auditing; and
(c) declines as the cost of auditing $k$ increases.

Proof: (i) To show that $f_0^* < h/p$, first rewrite social welfare (4) using the result from part (a) of Proposition 1 that $qf_M = f_0$. It follows from (2) that $\hat{w} = f_0/q$. For $w \leq f_0/q = f_M$, it follows from (3) that $b(w) = p\min\{qw, \min[w, f_0]\} = pqw$. For $w > f_0/q = f_M$, $b(w) = p\min\{qf_M, f_0\} = pf_0$. Thus, social welfare can be rewritten as:

$$\int_{0}^{f_0/q} \int_{pqw}^{\infty} \{b - h - pqk\}r(b)db\}g(w)dw$$

$$+ \int_{f_0/q}^{\infty} \int_{pf_0}^{\infty} \{b - h\}r(b)db\}g(w)dw - e(p).$$

From (5), the first-order condition with respect to the fine for the offense $f_0$ is:

$$-pqk[1 - R(pf_0)][dG(f_0/q)/df_0] - [pf_0 - h][1 - G(f_0/q)][dR(pf_0)/df_0] = 0,$$

where $G(.)$ is the cumulative distribution of wealth and $R(.)$ is the cumulative distribution of benefits. Since the first term in (6) is negative, the second term must be positive, which requires $pf_0 - h$ to be negative. In other words, it must be that $f_0 < h/p$, establishing part (a).

(ii) If auditing were costless, $k = 0$, (6) would require that $pf_0 - h = 0$, or $pf_0 = h$; since the expected fine for the offense equals the harm, this would result in first-best deterrence (for everyone who pays $f_0$). But if auditing is costly, $k > 0$, step (i) showed that $pf_0 < h$, resulting in underdeterrence. This establishes part (b).

(iii) By the implicit function theorem and the assumption that the second-order condition for a maximum is satisfied with respect to the fine for the offense $f_0$, the sign of $df_0/dk$ is the
same as the sign of the derivative of the left-hand side of (6) with respect to $k$. The latter
derivative is $-pq[1 - R(pfO)][dG(fO/q)/dfO] < 0$. Thus, $dfO/dk < 0$, proving part (c).

Comment: To achieve first-best deterrence, the expected fine for the offense $pfO$ would
have to equal the harm $h$ or, equivalently, the fine for the offense $fO$ would have to equal $h/p$.
The reason it is optimal to employ a lower fine, and tolerate some underdeterrence, is that doing
so reduces auditing costs. Specifically, if $fO$ is lowered, more individuals will be willing to pay
$fO$ rather than misrepresent their wealth levels (Proposition 1 established that individuals with
wealth levels above $fM = fO/q$ will pay the fine for the offense). These individuals no longer will
be audited, thereby saving auditing costs. There is no first-order reduction in social welfare due
to the marginal individuals who now commit the offense as a result of the lower fine for the
offense, because they were individuals whose benefits just equaled the harm. Thus, it is optimal
to set the fine for the offense $fO$ such that there is some underdeterrence, in order to reduce
auditing costs. Obviously, the higher the cost of an audit, the greater the motive to lower $fO$ for
this reason.

Note that, although a fine usually is treated as a socially costless sanction because it is a
mere transfer of wealth, it is no longer socially costless if auditing is required to induce
individuals to pay the fine. In effect, there is a marginal social cost incurred from imposing a
higher fine due to the need for a higher audit rate to induce the payment of the fine. As a
consequence, the fine for the offense should not be as high as it would be in the absence of
auditing costs.\textsuperscript{12}

\textsuperscript{12} The results in Polinsky and Shavell (1991), where it was assumed that offenders’ wealth levels could be
observed without cost, are a special case of the present analysis — when the cost of an audit, $k$, is zero. There it was
shown that the optimal fine for the offense is $h/p$, implying that individuals with wealth levels equal to or greater
than $h/p$ are deterred to the first-best extent. Here, due to the cost of auditing, the optimal fine is less than $h/p$,
implying that all individuals are underdeterred.
Proposition 4. The optimal audit probability $q^*$

(a) is positive: $q^* > 0$;

(b) rises as the cost of auditing $k$ declines; and

(c) equals unity for $k$ sufficiently low.

Proof: (i) To demonstrate that the optimal audit probability $q^*$ is positive, suppose it were zero. Then everyone would misrepresent their wealth levels. The critical value of benefit $b(w)$ would be zero for all wealth levels, regardless of the probability of detection $p$ (see (3)). But then it would be optimal for the state to spend nothing on enforcement, contradicting the assumption that $p^* > 0$. Thus, if the offense is worth deterring to any extent — that is, if $p^* > 0$ — it must be that $q^*$ is positive, establishing part (a).¹³

(ii) To see that $q^*$ rises as $k$ declines, consider the first-order condition with respect to $q$; using (5), and after some manipulation, it can be written as:

$$
k \left[ - \left[ \int_{pqw}^{f_0/q} \frac{dG}{dq} \right] \frac{f_0/q}{pqw} + \left[ \int_{pqw}^{f_0/q} \frac{dR}{dw} - \frac{pr(b)db}{dw} \right] g(w)dw \right] = 0.

(7)

By the implicit function theorem and the assumption that the second-order condition for a maximum is satisfied with respect to $q$, the sign of $dq/dk$ is the same as the sign of the derivative of (7) with respect to $k$. The latter derivative is the expression in large brackets in (7). To see that this expression must be negative, observe that $pqw - h < 0$ for $w < f_0/q$ because, by part (a)

---

¹³ This result does not require that the lowest level of wealth is zero, only that it is sufficiently low. Even if the state could impose a fine for the offense equal to the lowest level of wealth without auditing individuals, if this level is sufficiently low, the value of the resulting deterrence will be less than the cost of detection.
of Proposition 3, $f_0/q < h/pq$. Thus, the last term in (7) is positive, which implies that the expression in large brackets must be negative. This proves part (b).

(iii) To see that $q^* = 1$ for $k$ sufficiently low, I will first show that if $k = 0$, $q^* = 1$. If $k = 0$, the derivative of social welfare with respect to $q$ evaluated at $q = 1$ is the second term on the left-hand side of (7) evaluated at $q = 1$, which is

$$f_0 - \int_0^1 \{pw - h\}[dR(pw)/dq]g(w)dw.$$ 

(8)

If $k = 0$, step (ii) of the proof of Proposition 3 showed that $f_0 = h/p$. Since $pw - h < 0$ for all $w < h/p$, (8) is positive. This implies that $q^* = 1$ if $k = 0$. Clearly, if $k$ is sufficiently small, the derivative of social welfare with respect to $q$ evaluated at $q = 1$ will continue to be positive, in which case $q^*$ will continue to equal unity. This establishes part (c). □

Comment: If it is optimal to expend resources on trying to detect offenders, it must be optimal to audit them with some probability if they are caught, since otherwise the expenditures on detection would be wasted. (Of course, it might be optimal not to deter the harmful activity at all if the cost of detection and/or the cost of auditing is sufficiently great.) For any given probability of detection, a higher audit probability is beneficial because it induces more offenders to pay the fine for the offense $f_0$ rather than misrepresent their wealth levels (see parts (a) and (b) of Proposition 1), and thereby deters them to a better degree. Thus, if the cost of auditing were to decline, the optimal audit probability would rise, and if the cost were low enough, the optimal audit probability would be unity.
IV. Concluding Remarks

Although this article has considered monetary sanctions only, I want to observe in closing that the appropriate use of an imprisonment sanction for misrepresenting one’s wealth could increase social welfare. Specifically, suppose an imprisonment sentence, in addition to the fine $f_M$, is imposed on individuals who are found to have been capable of paying the fine for the offense $f_O$ but who nonetheless misrepresented their wealth level. (Individuals who are not able to pay the fine $f_O$ are sanctioned only with a fine equal to their wealth if they are found to have misrepresented their wealth level, as above.) If the imprisonment sentence is high enough, such individuals now can be induced to pay $f_O$ instead of misrepresenting their wealth level; and since they pay $f_O$, they will not be audited and will not bear the sentence. The behavior of all other individuals is unaffected under this sentencing policy. Hence, social welfare rises because the extent of underdeterrence, as well as the cost of auditing, declines for the affected individuals.

The preceding paragraph establishes that some use of the threat of imprisonment to deter misrepresentation of wealth is socially desirable, but it does not derive the optimal use of imprisonment sanctions for this purpose. It may be optimal also to impose imprisonment sentences for misrepresentation on individuals who are not capable of paying the fine for the offense $f_O$. Although some of these individuals will be audited and bear the imprisonment sentence, the additional deterrence due to use of an imprisonment sanction can allow the auditing probability $q$ or the probability of detection $p$ to be lowered, thereby saving auditing or enforcement costs. Thus, the optimal use of imprisonment to sanction misrepresentation of wealth might involve sentences that depend on whether an individual is found to have been capable of paying the fine for the offense $f_O$. 
Another possibility is to use imprisonment to sanction the offense. In a companion paper — Polinsky (2004) — I consider the optimal use of fines and imprisonment when wealth is assumed to be impossible to observe. I demonstrate in the model employed there that when imprisonment sentences are used, the optimal choice of sanctions induces high-wealth offenders to pay a fine that exceeds the wealth level of low-wealth offenders; by paying the fine, they avoid a less desirable combination of sanctions they would bear if they paid only as much as low-wealth offenders and faced a longer imprisonment sentence.\textsuperscript{14} Thus, an appropriately designed sanctioning policy that includes imprisonment sentences for the offense can serve as an alternative to auditing — both can reduce the incentive of offenders to misrepresent their wealth levels.

\textsuperscript{14} This result is similar in spirit to a point developed by Levitt (1997), though he formally considers fines and imprisonment sentences as alternatives, rather than the optimal combination of the sanctions.
References


