Private Law Enforcement, Fine Sharing, and Tax Collection: Theory and Historical Evidence

Metin M. Coşgel
University of Connecticut

Haggay Etkes
Bank of Israel

Thomas J. Miceli
University of Connecticut

Working Paper 2010-03

January 2010
Abstract
This paper contributes to the literature on private law enforcement by proposing a novel solution to the problem of underenforcement by monopolistic enforcers. Monopolistic enforcers underinvest in fine collection because, by maximizing net expected revenue, they ignore the social benefits of deterrence. We show that this problem can be partially resolved by combining the tasks of law enforcement with tax collection because a joint enforcer-collector will have an interest in reducing the crime rate in order to maximize his income from taxes. In support of the theory, we discuss two historical examples of this practice: decentralized law enforcement under European feudalism, and centralized law enforcement in the Ottoman Empire.

Journal of Economic Literature Classification: H11, K42, N40

Keywords: Criminal fines, deterrence, private law enforcement, tax collection
Private Law Enforcement, Fine Sharing, and Tax Collection:
Theory and Historical Evidence

1. Introduction

Economists have devoted a substantial amount of effort to studying optimal law enforcement policies ever since the appearance of Becker’s (1968) original analysis of crime and punishment. (See, for example, the survey by Polinsky and Shavell (2007).) For the most part, this literature has assumed that the enforcement authority internalizes social welfare and hence automatically enacts the optimal policy. One strand of literature, however, has explicitly addressed the incentives of law enforcers by assuming that the apprehension of offenders is delegated to “private” agents who are paid a reward for catching criminals according to a kind of bounty system. That is, enforcers receive as compensation a share of the fines collected from apprehended offenders. As discussed in the next section, an important conclusion of this literature is that such a compensation scheme generally fails to induce enforcers to invest in the socially optimal level of enforcement effort. In particular, if enforcement is delegated to a single monopolistic firm, there will tend to be underenforcement, whereas if it is delegated to a group of competing firms, there will tend to be overenforcement.

Several responses to this inefficiency have been discussed in the literature, some of which are surveyed in the next section. The purpose of this paper is to propose a new solution in the context of monopolistic enforcement. As will be shown, the inefficiency in this case stems from the enforcement firm’s failure to internalize the deterrence benefits of law enforcement. This is true because the firm’s expected income from collected fines will actually decline if its enforcement efforts deter too many potential offenders. Thus, the firm will tend to underinvest in enforcement from a social perspective. We show that this problem can be at least partially
solved by assigning the tasks of law enforcement and tax collection to the same agent. The intuition for this proposal is as follows. If we suppose that the tax base is decreasing in the crime rate (due, for example, to theft), then a private tax collector who is compensated by a share of collected taxes will have an interest in reducing the crime rate in order to increase his income. As a result, by coupling the tasks of tax collection and law enforcement, the government can create an incentive for the enforcer to internalize the benefits of deterrence. The outcome is second-best, however, because total tax revenues represent only a fraction of aggregate wealth.

After deriving the optimal compensation scheme of the joint tax collector-law enforcer in Section 3, the paper goes on to argue that this solution to the problem of underenforcement is of more than mere theoretical interest. In particular, Section 4 describes two historical examples of private law enforcement where something like this practice was actually employed: namely, feudal Europe and the Ottoman Empire.

2. The Literature on Private Law Enforcement

As noted, Becker’s (1968) original model of optimal law enforcement assumed a welfare maximizing government that automatically chooses the socially optimal enforcement scheme. In other words, no allowance is made for the incentives of enforcers to actually carry out the prescribed scheme. In reality, police officers may shirk in their efforts to apprehend offenders, or worse, accept bribes from offenders in return for letting them go free (Polinsky and Shavell, 2001). In response to these problems, Becker and Stigler (1974) proposed that the compensation of enforcers be made dependent on their performance, for example by paying them a reward, or
bounty, for those offenders that they apprehend, thereby effectively “privatizing” the enforcement of law.¹

Several subsequent authors have expanded upon this idea.² Landes and Posner (1975), for example, argue that a system of private enforcement with free entry of enforcers (a perfectly competitive system) would result in overenforcement. The intuition is as follows. In Becker’s model, potential offenders respond to the expected fine, \( pf \), when contemplating a criminal act, where \( p \) is the probability of apprehension and \( f \) is the actual fine. Thus, optimal deterrence involves an appropriate choice of the product of these two variables. Since it is costly to raise \( p \) but not costly to raise \( f \), the optimal (cost-minimizing) scheme achieves the desired value of \( pf \) by raising \( f \) as high as possible (up to the offender’s wealth) and correspondingly lowering \( p \). In a public enforcement regime, this outcome is readily achievable because the enforcement authority controls both \( p \) and \( f \). With private enforcement, however, competitive firms determine \( p \) once the value of \( f \) is announced, and the result will generally be overenforcement since any increase in \( f \) will be met with a corresponding increase in \( p \) as enforcers compete for the higher fine revenue.

Polinsky (1980) and Garoupa and Klerman (2002) showed that the incentives of private enforcers can be made efficient if the government sets a reward, or bounty, that is different from the fine paid by the offender. In other words, the government either taxes or subsidizes enforcers in order to induce them to engage in efficient enforcement. (Also see Besanko and Spulber (1989).) Alternatively, Friedman (1984) showed that competitive private enforcement can be

¹ The literature on “private” law enforcement is therefore really about the incentives of enforcers, rather than about whether they are publicly or privately employed.
² See the survey of this literature by Rajabiun (2009).
made efficient if the government sets the *expected* fine, $E=pf$, rather than the actual fine, thus fixing the optimal crime rate.

An alternative to competitive enforcers is for the government to auction the right to collect fines to a single firm that would act as a monopolist, but again, this regime is not generally capable of achieving an efficient level of enforcement. The problem here is that the monopolist only cares about the expected fine revenue and thus does not internalize the social benefits from deterrence when making its enforcement choice. As a result, it underinvests in enforcement (Polinsky, 1980). Garoupa and Klerman (2002, p. 131) suggest that one way to solve this problem is to combine fine sharing with a reward that is inversely related to the number of offenses. However, they dismiss such a scheme as impractical because it would require the government to know the crime rate. In fact, we show in the next section that such a scheme can be implemented by system that jointly delegates the tasks of tax collection and law enforcement to a single agent.

3. The Model

We consider a simple model in which a community that is subject to thievery by bandits delegates the task of apprehending and recovering the stolen property to a single firm. The objective of the community is to maximize its net wealth, including untaken plus recovered wealth, net of enforcement costs. (The benefit of bandits is thus not counted in social welfare.) The enforcement firm is assumed to maximize its profit, which consists of the reward scheme (to be specified) minus enforcement costs.

---

3 In this sense, the objective function thus differs from that in the standard Becker-Polinsky-Shavell model (Polinsky and Shavell, 2007). It also differs in that it assumes that the harm from crime (theft in this case) is not fixed but can be mitigated by the efforts of the enforcer to recover some stolen wealth.
Bandits are rational and choose whether or not to commit a theft by comparing the expected benefits and costs. To be specific, suppose that an individual bandit can steal $g$ dollars (or property worth that amount), but must incur a cost of $k$ dollars in the effort while facing a probability $p$ of being caught. If caught, we assume that the bandit must turn over the fruits of his crime, reflecting the maximum fine that he is able to pay. (For simplicity, we assume that no other punishment is possible.) Thus, the offender’s net gain from committing the crime is $g(1-p)$. An individual bandit will therefore commit a crime if and only if $g(1-p) \geq k$. Finally, assume that in the population of potential bandits, $k$ is distributed uniformly on $[0,K]$.\(^4\) Thus, given $p$, the total number of crimes (the crime rate) is given by \(R(p) = \Pr(k \leq g(1-p)) = g(1-p)\), while the total amount stolen is

\[
T(p) = gR(p) = g^2(1-p).
\]

Note that \(T' = -g^2 < 0\), reflecting the deterrent effect of greater enforcement (a higher recovery rate).

Let $W$ be the gross wealth of the community, which we assume is fixed. Total net wealth, including recovered wealth, is therefore given by

\[
\]

Let the cost of enforcement be $c(p) = cp$, where $c$ is the constant marginal cost. If it could implement the first best solution (as under “public” enforcement), the community would choose $p$ to maximize $N(p) - cp$. The resulting first-order condition is\(^5\)

\[
-T'(1-p) + T = c.
\]

\(^4\) The distribution of $k$ can reflect either differing opportunity costs, different inclinations to commit crime, or variation in the availability of criminal opportunities across offenders. The assumption of a uniform distribution is made purely for simplicity and to allow an explicit derivation of the optimal recovery rate. The same results hold, however, for more general distributions.

\(^5\) The second order condition for a maximum is satisfied.
The marginal benefit of enforcement, the left-hand side, consists of two terms, both of which are positive. The first term, \(-T'(1-p)\), reflects the deterrence effect of additional enforcement—the fact that greater enforcement lowers the crime rate; while the second term, \(T\), reflects the recovery effect. At the optimum, the sum of these effects is set equal to the marginal cost, \(c\).

Substituting for \(T\) and \(T'\) and solving for \(p\) we obtain the optimal recovery rate

\[
p^* = 1 - \frac{c}{2g^2},
\]

where we assume that \(c < 2g^2\) so that \(p^* > 0\). This represents the first-best rate of enforcement.

3.1. Fine Sharing Only

Now consider the outcome under a delegated enforcement regime where the community auctions the right to apprehend bandits to a single enforcement firm. Assume initially that the firm’s sole revenue is that it is allowed to retain a share \(s\) of the recovered wealth (a variable payment), for which right it pays the community a fixed fee \(F\). The enforcement firm’s expected profit is thus

\[
\pi = -F + spT(p) - cp.
\]

The community’s problem is to choose the share parameter, \(s\), and the fee, \(F\), to maximize its share of expected net wealth, given by

\[
W - T(p) + (1-s)pT(p) + F,
\]

subject to the following constraints:

(i) \(p = \text{argmax } \pi\)

(ii) \(\pi \geq \pi_0\)

(iii) \(0 \leq s \leq 1\).
Constraint (7.i) is the incentive compatibility constraint, (7.ii) is the participation constraint, and (7.iii) limits the firm’s variable payment to the amount of the recovered wealth.\(^6\)

The firm takes \(F\) and \(s\) as given and chooses \(p\) to maximize (5). The resulting first order condition is

\[ s(T + pT') = c. \] (8)

Substituting for \(T\) and \(T'\) and solving for \(p\) as a function of \(s\) yields

\[ \hat{p}(s) = \frac{1}{2} - \frac{c}{s2g^2}. \] (9)

It follows from (9) that \(\hat{p}\) is increasing in \(s\), but clearly \(\hat{p} < p^*\) for \(s \in [0,1]\). That is, the firm underinvests in recovery effort. As noted above, the reason for this inefficiency is that the enforcement firm does not account for the deterrence effect of its efforts, and in fact, actually benefits from a higher crime rate because that increases the available loot to recover. This is the problem of underenforcement by a monopolistic private enforcer first noted by Polinsky (1980).

The best that the community can do in this case as to set \(s=1\) (i.e., allow the enforcer to keep all of the recovered wealth), and then extract any rents by setting \(F\) so that \(\pi = \pi_0\).\(^7\)

### 3.2. Fine Sharing Combined with Tax Collection

As a remedy to this underenforcement problem, we propose a scheme whereby a single firm is delegated the task of both recovering stolen wealth and collecting taxes on untaken wealth, again in return for a fee, \(F\). The firm’s expected profit under this scheme is

\[ \pi = -F + spT(p) + at(W-T(p)) - cp, \] (10)

---

\(^6\) Note that the sum of (5) and (6) equals overall net wealth, \(N(p) - cp\).

\(^7\) It is important to recognize the dependence of this conclusion on the constraint that \(s \leq 1\). To see this, note that the first-best outcome can be achieved by setting \(s = c/(c-g^2)\), which is derived by equating (4) and (9) and solving for \(s\). This expression, however, implies either that \(s > 1\) if \(c > g^2\) (i.e., the enforcer’s collection of fines must be subsidized), or that \(s < 0\) if \(c < g^2\) (i.e., the enforcer must turn over all collected fines plus pay a tax on top of it) (Besanko and Spulber, 1989). Since we rule out both of these solutions, the first-best outcome is not attainable by pure fine sharing.
where $t$ is the tax rate, which we assume is determined by considerations other than law enforcement, and $\alpha$ is the share of taxes retained by the enforcer-collector. The community’s problem is now to choose $s$, $\alpha$, and $F$ to maximize expected net wealth subject to the constraints in (7) along with $\alpha \in [0,1]$.

Proceeding as above, we first derive the enforcer’s privately optimal enforcement rate. The first-order condition for the firm’s problem is

$$s(T + pT') - \alpha t'T = c,$$

which can be solved for $p$ to yield

$$\tilde{p}(s, \alpha, t) = \frac{s + \alpha t}{2s} - \frac{c}{2sg^2}. \tag{12}$$

Note that if $\alpha=0$ and $s=1$, the outcome is identical to that in the previous case. Here, however, setting $\alpha>0$ increases the enforcement rate for any $t>0$. Intuitively, allowing the enforcer to collect (and retain a share of) taxes on untaken wealth induces it to at least partially internalize the deterrence benefits of its efforts. Since $t<1$, the first-best outcome is unattainable, even if $\alpha=1$ (as is optimal), but the level of enforcement is nevertheless closer to first-best than was possible under the pure fine sharing arrangement. In particular, the enforcement rate under the optimal scheme (where $s=\alpha=1$) is

$$\tilde{p}(t) = \frac{1 + t}{2} - \frac{c}{2g^2}, \tag{13}$$

which approaches $p^*$ as $t$ approaches unity.

### 4. Historical Examples of Combined Law Enforcement and Tax Collection

---

8 See, for example, Cosgel and Miceli (2009).

9 We have obviously assumed throughout that the firm is risk-neutral. If it is risk-averse and the community risk-neutral, then setting $s=\alpha=1$ would not be optimal because it transfers all of the risk to the firm. See, for example, Cosgel and Miceli (2009) and Stiglitz (1974).
The preceding scheme resembles a solution to the underenforcement problem suggested by Garoupa and Klerman (2002, p. 131) wherein the enforcer is compensated both by a share of the collected fine and a reward that is inversely related to the number of offenses in the form of taxes on untaken wealth. Although Garoupa and Klerman dismissed such a scheme as too informationally demanding, the preceding shows that there are no special requirements other than that the community needs to raise tax revenues. In the following sections we describe two historical examples of such schemes where law enforcement and tax collection were effectively combined: the decentralized system of European feudalism, and the centralized government of the Ottoman Empire.

4.1. Law Enforcement in Feudal Courts

Law enforcement in the feudal or seignorial courts of Europe was a private system where the lord had jurisdictional authority over his peasants and held a court of his own.\(^\text{10}\) The manor courts of England, for example, resolved disputes and punished misdeeds, dealing not only with questions of property or issues arising from the relationship between the lord and his peasants, but also with criminal matters such as theft, assault, and bloodshed. These courts were typically local and customary in that they had jurisdiction over a geographically limited community and applied law that originated from the specific customs of the manor. Cases arising on the lord’s manor came before him (or his agents), and it was his responsibility to enforce the law by issuing rulings and collecting fines. In exchange for jurisdictional (and other) rights, feudal lords owed the ruler direct allegiance and provided men and arms to fight in his wars.

The lords were the direct beneficiaries of the fiscal returns from feudal law enforcement. Because of the high cost of imprisonment, offenders were generally ordered to pay a fine.

---

\(^{10}\) Ault (1923), Drew (1995), Goebel (1976), Pollock and Maitland (1895).
determined at the discretion of the lord and payable to him. English manorial rolls show that a variety of offenses were punished by fines. For example, in a session of one of Ramsey’s manor courts held in Hemingford in November 17, 1278, we read the case of someone who “carried off wheat that belongs to another from the latter’s land; let satisfaction be made; fine, 6d.,” and in another case “two ale-tasters say that ten women whom they name have broken the assize of beer; the fines are 6d. and 12d.” (Ault, 1923: 147) Winchester Pipe Rolls of the thirteenth century similarly show that offenders were fined 6d. *pro pastura* for harming the lord’s pastures and a wide range of amounts for a variety of other offences, often summarized by the scribes as *pro assisa fracta* (May, 1973). Although the lords’ revenue from jurisdictional fines could vary among manors, these courts were “a source of revenue sufficiently valuable to form a separate paragraph in the great annual rendering of account by the reeve’s *comptus*.” (Bennett, 1937: 219)

In addition to the income that the lords received from fines, they also collected taxes levied as feudal services and payments. The service obligation was primarily for work on the lord’s field. The manor land was divided into two parts, the lord’s demesne and the peasants’ own holdings, and as part of his feudal obligation each peasant was required to work the lord’s land for a certain number of days. Feudal services also included the obligation to maintain roads, fences, and buildings. As feudal payment from the peasants, the lord received a portion of the crop on their own fields. He could also draw income from levies on the peasants’ use of the woods and pasture and from charges they paid for grinding their corn in his mills and baking their bread in his ovens.

Since all fines \((s=1)\) and taxes \((\alpha=1)\) thus went to the lords, feudal courts replicated the optimal (second-best) enforcement incentives described above. True, the lord had enormous
monopoly power in enforcement and taxation, and the efficiency of law enforcement depended entirely on his decisions. But as the theoretical analysis showed, this system induced him to internalize most of the social benefits from law enforcement. In particular, because the tasks of tax collection and law enforcement were jointly delegated to a single agent, that agent benefited from both recovery and deterrence, and he thus had an incentive to choose the (second-best) efficient level of enforcement.

4.2. Law Enforcement in the Ottoman Empire

The Ottoman system of government differed systematically from European feudalism, but what the two systems had in common was the delegation of tax collection and law enforcement to a single agency. In contrast to the decentralized system of European feudalism, where the lords had hereditary rights to land and personal bondage ties to peasants, the Ottoman system of provincial government depended on appointees of the central government with revocable rights. The Ottomans standardized the law in each region, so that the legal basis and procedures of law enforcement depended more on a centralized code than the local custom of the village (though provincial laws typically incorporated common pre-Ottoman customs and practices of the region into the code).11 Despite these systematic differences in the structure of provincial law and government, the Ottoman and feudal systems commonly relied on a single agency for both law enforcement and tax collection. The local appointees of the Ottoman central government enforced the law in the countryside and received income from associated fines as well as from direct tax revenue. In return, they owed military and other service obligations to the sultan, similar to the way the lords met their obligations to the ruler in feudalism.

11 See Heyd (1973) for the Ottoman criminal codes.
In a multi-tiered system of government, the Ottoman Empire consisted of several administrative levels that divided the empire into provinces, the provinces into districts, and the districts into fiefs. In the provinces, government agents typically received income directly from taxable sources assigned by the central government. For example, if a certain cavalryman (sipahi) was appointed as the fiefholder of a village, all peasants in that village paid their taxes directly to him as his remuneration. In return for the right to collect taxes, he provided local protection to the peasants, men and arms to the central government in times of war, and possibly other services such as the maintenance of roads and bridges.

Government appointees in the countryside received income not just from a variety of taxes levied on persons, land, and productive activities, but also from fines collected in the enforcement of law. As in the feudal case, the penalty for most wrongdoings was typically a monetary fine. Expected fine revenues were generally recorded in the tax registers under the general heading of “windfall” revenue (bād-i havā) or in reference to more specific fines (cerīme) expected from widely observed misdemeanors such as crop damage caused by stray cattle. Further details on the types and amounts of fines can be found in the surviving records of actual revenues from these fines. A document pertaining to the fines collected by Hüseyin Subaşı, an official in charge of collecting fines in eight villages around the town of Ramle in 1586, shows that there were thirty-three instances of crimes and misdemeanors in the four months covered by the register, including fighting (with fines ranging between 20 and 540 pāra), stealing rice (200), and drinking wine (160) (Singer, 1990: 138).

As Heyd (1973: 277) shows, the Ottoman legal system prescribed fines for a variety of offences, including theft, fornication, selling unstamped cloth, and even such things as “disturbing falcons in a region where they were bred for the Sultan’s hunt.” For certain crimes,
the fine was determined according to a scale that depended on the financial ability of the offender (Heyd, 1973: 283). In some cases, fines were the only penalty imposed, often as a substitute for capital or severe corporal punishment (prescribed by the shari’a) such as stoning or amputation. In other cases, a monetary fine supplemented the shari’a penalty of chastisement (ta’zīr).

A fundamental principle of the allocation of revenue in the Ottoman Empire was the delegation of taxes and fines to the same recipient under the rule that the fines “belong to the land [on which the offence was committed] (cürm ü cināyet toprağa tābi’dir).” This rule applied directly to the “free” (serbest) lands, on which the agent assigned to collect taxes received all of the fines without interference or demands from the provincial governor. On the lands that were not categorized as “free,” the fiefholder shared the fines with the governor, typically in an equal proportion.12

An Ottoman decree of 1540 supports our argument about the optimality of using the same agent for the collection of taxes and fines. Previous to the decree, it was customary in some regions to farm out the revenue from fines. This was possibly because the revenue was small, and specialized tax-farmers had a lower cost of collection. But farming out of fines often gave rise to a problem because severing the link between taxes and fines added another layer of agency to collection and changed the incentive structure of the collectors. There were, for example, reports of collecting “illegal as well as excessive fines” (Heyd, 1973: 296).

12 See Heyd (1973: 289) and Peirce (2003: 324) for examples of how fine revenues were allocated.
Recognizing the problem, the central government banned the practice in 1540 and ordered that fines be collected only by the office of those assigned to collect taxes.¹³

5. Conclusion

This paper has contributed to the literature on private law enforcement in two ways. First, it proposed a novel solution to the problem of underenforcement in the context of monopolistic collection of criminal fines. It has been shown in the literature that a monopolistic enforcement firm will underinvest in collection of fines because, in pursuit of its goal of maximizing the expected fine revenue, it ignores the social benefits of deterrence. We showed that this problem can be at least partially resolved by combining the tasks of law enforcement with tax collection because the joint enforcer-collector will thereby be given an interest in reducing the crime rate in order to maximize his income from tax collection.

The second contribution of the paper was to point out that various instances of combined tax collection and law enforcement have been observed throughout history. We discussed two examples in the paper: the decentralized system of law enforcement that characterized European feudalism, and centralized law enforcement in the Ottoman Empire.

¹³ A copy and transcription of the ban, known as “para voyvodalari yasaknamesi,” can be found in Akgündüz, ed. (1993: Volume 6, pp. 313-28).
References


