Coordinated Private and Public Enforcement of Law: Deterrence under Qui Tam

David Kwok

February 8, 2010

Abstract

Inducing private enforcement of law through civil litigation and bounty percentages can help scale enforcement efforts towards more harmful offenses. For a social planner focused upon deterrence, an optimal coordinated strategy would result in public enforcement targeting civil penalties towards comparatively less harmful offenses. I test this deterrence model against data from the federal False Claims Act and find that executive efforts do not appear to result in a coordinated deterrence strategy.

1 Introduction

Economists have compared theories regarding the superiority or equivalence of private versus public enforcement (see, e.g., Polinsky, 1980; Shavell, 1993; Landes & Posner, 1975; Becker & Stigler, 1974). In reality, although modern criminal law enforcement generally focuses upon the efforts of public agents, private actors play at least limited roles in the detection and resolution of legal violations. As Polinsky and Shavell (2000) note, coordinating these public and private efforts merits examination. This paper looks at the potential coordination of private enforcement via civil financial penalties with traditional public enforcement. Private enforcers may not always have a big picture view of law enforcement, but their collective efforts may allow public enforcers to better utilize limited resources. Using government fraud as an example, I will look at the design of a coordinated enforcement regime.

The coordination of public and private enforcement efforts makes particular sense in the area of financial crimes within a complex economy. Following the concern regarding relative enforcement costs from Polinsky (1980), detection alone of such crimes may be the most difficult and costly step of the enforcement process. We do not have good victimization data to help determine background rates of offenses. Without such data, a social planner concerned about deterrence has a very difficult problem in determining the sufficiency of enforcement efforts. Prompt detection by public authorities may require intrusive and costly monitoring techniques that could generate hostility and political backlash. In contrast, existing employees within such businesses and organizations have regular, real-time visibility into potential legal violations. Personal detection by such employees is nearly costless, although for them to reveal this personal knowledge and pursue enforcement action is a very different story.

Prior models have emphasized a uniform harm per violation, resulting in concerns of marginal enforcement costs per harm. In reality, we may be more concerned about crimes that result in greater harm to society. For financial crimes we can see a range of external harms; deterring large scale fraud may be more important than stopping a single bad check. Rewarding private enforcement via a percentage bounty may improve the distribution of enforcement resources. If civil penalties correspond to the level of societal harm, a percentage reward can help ensure that greater resources are against legal violations that cause greater harm.

One historical class of cases showcasing the coordination of private and public enforcement efforts is qui tam litigation. In qui tam cases, private enforcers (known as "relators") litigate on behalf of the government (historically, the king). Successful relators receive a portion of the penalty paid by the defendant. Although qui tam actions have a long history in the common law tradition, the major remaining U.S. federal statute with a qui tam provision is the False Claims Act (FCA). The FCA proscribes fraud against the federal government, most of which today is Medicare and defense contractor fraud. My data regarding FCA actions will help illustrate the results of one coordinated enforcement regime.

Part II of this paper will review some of the history and procedure behind

FCA qui tam actions. Part III describes a model of government fraud with private enforcement on a percentage bounty. Part IV considers the government response. Part V applies this model to the FCA and considers the result of the government response. Part VI considers some extensions, and Part VII concludes with some potential policy recommendations.

2   FCA Background

Qui tam provisions allow a private citizen, also known as a "relator," to bring a civil action in the name of the government. Qui tam is shorthand for the Latin phrase "qui tam pro domino rege quam pro se ipso in hac parte sequitur," which translates to "who as well for the king as for himself sues in this matter." Such provisions date back thirteenth century England, by which private citizens gained access to royal courts. While England's American colonies adopted statutes with qui tam provisions, most have not survived.

The main remaining modern statute with a qui tam provision is the False Claims Act. The False Claims Act (FCA), 31 U.S.C. §§3729 - 3733 was enacted in 1863 by a Congress concerned that suppliers of goods to the Union Army during the Civil War were defrauding the Army. Since then, the FCA has been amended several times, most recently in 2009. The statute generally proscribes fraud against the federal government, and, unusual in today's statutory environment, it includes a qui tam provision allowing for private parties to litigate on behalf of the federal government. Many states have followed the federal lead in establishing state-level FCA qui tam provisions.

Watergate-related investigations led to the most significant amendments of the FCA in 1986. Compared to the stagnant situation pre-1986, today approximately 400 such cases are filed at the federal level each year. The amendments granted that prior government knowledge of the allegations does not automatically prevent a relator from filing a qui tam action. More importantly, the 1986 amendment provided that even if the government joins the lawsuit and has "primary responsibility for prosecuting the action," the relator "shall have the right to continue as a party to the action." Second, it increased a relator's recovery for a successful suit to a maximum of 30% if the government does not intervene, and to a maximum of 25% if it does, and increased the overall damages and penalties that can be imposed on a defendant from double to treble damages. Finally, the 1986 amendment protected a relator from retaliatory actions by employers, making it safer for an individual to bring qui tam actions by adding whistleblower protection language to the statute.

§§3729(a)(1) and (2) set forth FCA liability for any person who knowingly submits a false claim to the Government or causes another to submit a false claim to the Government or knowingly makes a false record or statement to get a false claim paid by the Government. Section 3729(a)(7) is known as the reverse false claims section; it provides liability where one makes a false record or statement - not to get money from the Government - but to avoid having to pay money to the Government.

Those found liable must pay a civil penalty of between $5,000 and $10,000 for each false claim (those amounts are adjusted from time to time; the current amounts are $5,500 to $11,000) and treble the amount of the Government's damages. Where a person who has violated the FCA reports the violation to the Government under certain conditions, the FCA provides that the person shall be liable for not less than double damages.

2.1 Qui Tam Procedure

The qui tam provisions begin at §3730(b) of the FCA; §3730(b)(1) states that a person may file a qui tam action. Section 3730(b)(2) provides that a qui tam complaint must be filed with the court under seal. Most relators are represented by counsel on a contingency fee basis; they are not responsible for attorneys' fees if the case is unsuccessful. The complaint and a written disclosure of all the relevant information known to the relator must be served on the U.S. Attorney for the judicial district where the qui tam was filed and on the Attorney General of the United States.

The qui tam complaint is initially sealed for 60 days. The Government is required to investigate the allegations in the complaint; if the Government cannot complete its investigation in 60 days, it can seek extensions of the seal period while it continues its investigation. Such extensions are generally granted, as the typical time under seal is around 13 months. When the Government has made a decision, also known as the time of "election," the Government must then notify the court that it is proceeding with the action (generally referred to as "intervening" in the action) or declining to take over the action, in which case the relator can proceed with the action. The Government can also dismiss the case outright, although it does not choose to do so often. If the Government has not made a decision, but the court does not grant an extension, the Government will file a "notice of no decision," which courts generally treat as non-intervention. Accompanying such a notice of no decision is typically a proposed order unsealing materials for litigation, although it is still within the court's discretion to actually unseal the case.

When the DOJ receives the initial complaint, it generally forwards copies to the potentially defrauded agency, typically the inspector general's office. The results of the investigation can be shared with both criminal and civil divisions. After the investigation is complete, the case can be "shopped" to different attorneys, either within the local US Attorney's office or at the DOJ. By regulation, the local US Attorney's office has the authority to handle cases under $5 million and does not have to defer to the DOJ. Otherwise, the local office must defer to the DOJ. Within the DOJ Civil Division, it is the individual attorney's decision whether to accept the case.

If the Government intervenes in the qui tam action it has the primary responsibility for prosecuting the action. It can dismiss the action, even over the objection of the relator so long as the court gives the relator an opportunity for a hearing and it can settle the action even if the relator objects so long as the relator is given a hearing and the court determines that the settlement is fair.

The relator's attorneys can work alongside government attorneys in pursuing the case, and the relator may also provide expertise and support. If a relator seeks to settle or dismiss a qui tam action, it must obtain the consent of the Government. When the case is proceeding, the Government and the defendant can ask the court to limit the relator's participation in the litigation. Even if the government decides against intervention during the sealed complaint period, it can do so at a later time upon a showing of "good cause."

If the government decides against intervention (and thus implicitly deciding against dismissal), the relator can then litigate the case with counsel. The Government will often notify the relator and counsel of its non-intervention decision prior to formally notifying the court, thus allowing the relator and counsel time to make a decision regarding the case.

Current DOJ policy is to request for courts to unseal cases even if relators withdraw except in cases of unusual hardship-for example, if the relator has reason to fear severe retaliation. I do not have information at this time on U.S. Attorney compliance with this policy, and courts still retain discretion as to the unsealing of cases.

2.2 Award to the relator

If the Government intervenes in the qui tam action, the relator is entitled to receive between 15 and 25 percent of the amount recovered by the Government through the qui tam action. If the Government declines to intervene in the action, the relator's share is increased to 25 to 30 percent. If the case turns primarily on publicly disclosed information, the relator's share may be reduced to an amount no more than ten percent. If the relator planned and initiated the fraud, the court may reduce the award without limitation. The relator's share is paid to the relator by the Government out of the payment received by the Government from the defendant. If a qui tam action is successful, regardless of intervention, the relator also is entitled to be paid his legal fees and other expenses of the action by the defendant. The FCA also provides that if the Government chooses to obtain a recovery from the defendant in certain types of proceedings other than the relator's FCA suit, this is known as an alternate remedy and the relator is entitled to same share of the recovery as if the recovery was obtained through the relator's FCA suit. Even if the government does not immediately intervene in the qui tam action, the attorney general may be able to veto settlements otherwise approved by the court, relator, and defendant.

2.3 Statutory bars to qui tam actions

Under several circumstances, the FCA prohibits pursuit of a qui tam action:

1. The relator was convicted of criminal conduct arising from her role in the FCA violation.

2. Another qui tam case concerning the same conduct already has been filed (the "first to file bar").

3. The Government already is a party to a civil or administrative proceeding concerning the same conduct.

4. The qui tam action is based upon information that has been disclosed to the public through any of several means included criminal, civil, or administrative hearings, Government hearings, audits, or reports, or through the news media (this is known as the "public disclosure bar.") There is an exception to the public disclosure bar where the relator was the original source of the information.

2.4   Parallel Criminal Actions

The civil enforcement provisions of the FCA run in parallel with the criminal provisions in 18 U.S.C. §287. Although the civil provisions explicitly proscribe tax fraud, §287 has been used for a variety of claims including tax fraud. Criminal violations of the FCA can incur sentences of up to five years and fines up to \$250,000 per charge. Qui tam proceedings, like other civil FCA proceedings, generally defer to criminal proceedings for a number of reasons. First, a defendant has a different set of rights under civil discovery rules. Second, while a civil case is not entitled to knowledge of an ongoing grand jury process, the civil division can obtain a court order to learn about the grand jury proceedings after the criminal case is complete. Finally, a criminal conviction can be the basis for summary judgment in a civil case.

3   The Model

Notation
c = contract amount

$f_c(.)$ = probability density of contracts
$F_c(.)$ = cumulative distribution of $f_c(.)$
γ = amount of contract fraud, $[0, c]$
$p(\gamma)$ = probability of detection & punishment given reward of the fraud, $p' > 0, p'' < 0$
α = damage multiplier for punishment

Using the FCA as an example, we will follow the Becker model of crime. The potential fraudsters here are contractors, providing some service or good, and the federal government compensates them financially. There may be various motivations leading a contractor to commit fraud. Nonetheless, as an ostensibly profit-maximizing entity considering an economic crime, the Becker model has some relevance to the contractor's decision process. These contractors are distinguished only on the size of the government contract they hold. Thus, the cumulative distribution of contracts, F_c, is determined exogenously. We can interpret the distribution to be some technical distribution of contracts that leads to the optimum provision of services. Although c is a contract amount,

it does not necessarily correspond to a single government contract. Rather, it should be envisioned as the opportunity for fraud given the contractual relationship with the government. For example, if this "contractor" is a doctor filing for Medicare reimbursement, he may be filing numerous claims on behalf of a number of different patients. Since the opportunity for fraud extends across the various claims, his contract amount c would be the sum of the limits of Medicare reimbursement for those patients.

Each contractor, having been allocated a particular contract amount c, then chooses the amount of fraud to commit, γ. The maximum amount of fraud he can commit equals c, which represents a situation in which he claims to perform all services or provide all goods but actually does nothing besides collect federal money. If he opts to properly fulfill the contract according to its terms, he selects $\gamma = 0$. We assume a competitive situation such that performance of the contract, either in part or in whole, results in no profit to the contractor. Thus, the contractor's only opportunity for profit lies in committing fraud γ. Although some might view fraud as simply a transfer rather than a net loss for society, here we will assume that the net loss to society is at least as large as the amount of fraud γ.

This risk neutral contractor, facing his decision on the amount of fraud to commit, is concerned about the probability of detection and punishment. If his fraud is discovered and punished, he will lose $\alpha\gamma$, the amount of fraud times a damage multiplier. For the FCA, treble damages means $\alpha = 3$. $\alpha\gamma$ is the only loss he suffers; there is no reputational loss nor loss associated with the overall value of the contract. We also assume that the contractor is not judgment proof, thus the $\alpha\gamma$ penalty is meaningful.

His probability of detection and punishment, p, is a function of γ. The probability is an abstraction of the entire detection and judicial process. Here we emphasize the role of the relator in stepping forward with her knowledge of the fraud. We assume that the relator's decision is based upon the amount of the fraud γ and that her chance of success in relating does not depend on the amount of the fraud. The relator's action is always honest, accurately revealing fraud. Thus, $\lim_{\gamma \to 0} p(\gamma) = 0$. In effect, this is a fixed-cost model of action by the relator; if she reveals her knowledge of the fraud, she will face some future professional and social stigma as a whistleblower than does not depend upon the amount of fraud. She receives benefits corresponding to the amount of the fraud: both in terms of the bounty percentage paid after successful litigation and the psychic benefits of helping stop large levels of fraud. Thus, as the level of fraud increases, she is more likely to relate. Also note that the contractor has no ability to hide or fight the relator's actions; his only method of decreasing the detection & penalty is to decrease the amount of fraud.

Steps:

1. Nature chooses $C = c, F_c(.)$

2. Contractor chooses $\gamma \in [0, c]$; illicit gain of γ

3. Relator relates successfully with probability $p(\gamma)$, which causes Contractor to lose $\alpha\gamma$.

Contractor: Maximize expected profit

The contractor, knowing p, has a decision rule to maximize expected net profit in choosing γ.

$$\max{}_\gamma E[netprofit] = p(\gamma)[-\alpha\gamma] + \gamma$$

if $E[netprofit] < 0$, then $\gamma = 0$.

To make things interesting, I assume p has a form sufficient to induce a non-trivial result. It may be possible that p times α is sufficiently high to deter all fraud. Thus, selecting $\gamma \neq 0$ consistently results in a negative expected profit for the contractor. Similarly, p might have a form such that it generates no deterrence; contractors would always select $\gamma = c$ if p is consistently and sufficiently low. Besides $p' > 0, p'' < 0$ and $\lim_{\gamma \to 0} p(\gamma) = 0$, I assume that $\lim_{\gamma \to 0} p'(\gamma) = \infty$, $\lim_{\gamma \to \infty} p(\gamma) = 1$, $\lim_{\gamma \to \infty} p'(\gamma) = 0$, and $\alpha > 1$. Furthermore, I need a constraint on $\frac{p''}{p'}$ to ensure that the second order condition is satisfied.

The first order condition to maximize the contractor's expected profit is

$$1 - p(\gamma)\alpha - \gamma p'(\gamma)\alpha = 0$$

which resolves to

$$\gamma^* = \frac{1 - p(\gamma^*)\alpha}{p'(\gamma^*)\alpha}$$

Ensuring that this is a maximum through the second order condition:

$$-p'(\gamma)\alpha - p'(\gamma)\alpha - \gamma p''(\gamma)\alpha < 0$$

reduces to

$$-\frac{2}{\gamma} < \frac{p''(\gamma)}{p'(\gamma)}$$

Under this γ^* solution, we see that all contractors are motivated to commit fraud at the exact same level. Contractors who have large c opportunity for fraud will still commit γ^*. Contractors for whom $c < \gamma^*$ will simply commit $c\gamma = c$ fraud. By itself, private enforcement through the qui tam relator process tends to deter high levels of fraud while leaving low levels of fraud untouched.

4   The Role of Government

Before addressing what government might do, I consider the objective function for a hypothetical unified government regulator. Here I looking at fraud as a proxy for the net harm to society caused by the fraud. This simplifies the calculations, as we can treat the government as a selfish actor-minimizing fraud is already in its own best interest.

4.1 Objective Functions

1. Minimize expected total fraud

$$\min E[F_c(\gamma^*) + \gamma^*(1 - F_c(\gamma^*))]$$

This objective function corresponds to a strong focus on deterrence. Once a contractor commits fraud, society sustains the loss, regardless of ex-post detection or remediation. The fact that a contractor is successfully held liable for fraud after the fact does not make a difference to the government.

2. Minimize total expected unrecovered fraud

$$\max E[-F_c(\gamma^*) - \gamma^*(1 - F_c(\gamma^*)) + \gamma^* p(\gamma^*)(1 - F_c(\gamma^*)) + \int_0^{\gamma^*} \gamma p(\gamma) f_c(\gamma) d\gamma]$$

Although this objective function also focuses on deterrence, it distinguishes between fraud that is detected and undetected. If fraud is successfully detected, this objective function suggests that the post-detection remediation is sufficient to offset the harm incurred by the commission of fraud. This calculation excludes the welfare of the contractor held liable for the fraud, as he will bear a significant loss through this process.

3. Maximize expected fraud recovery - expected total fraud

$$\max E[-F_c(\gamma^*) - \gamma^*(1 - F_c(\gamma^*)) + (\alpha - \beta)\gamma^* p(\gamma^*)(1 - F_c(\gamma^*)) + \int_0^{\gamma^*} (\alpha - \beta)\gamma p(\gamma) f_c(\gamma) d\gamma]$$

Similar to option 2, this objective function suggests that we can offset the losses due to fraud by recovering money from defendants. It is distinct, however, in that it considers the cost of paying the bounty, β, to the successful relator. I will refer to option 3 as a mixed deterrence and compensation objective.

4. Maximize expected fraud recovery

$$\max E[(\alpha - \beta)\gamma^* p(\gamma^*)(1 - F_c(\gamma^*)) + \int_0^{\gamma^*} (\alpha - \beta)\gamma p(\gamma) f_c(\gamma) d\gamma]$$

We might imagine a bureaucratic structure that might be narrowly focused strictly on fraud recovery. It does not consider the deterred fraud as part of the objective function.

4.2 What can government change?

1. Damage multiplier α

   The traditional optimal deterrence argument is to maximize the penalty, since the penalty loss is born primarily by the offender and does not require costly investment (in comparison to increasing the chance of detection). Increasing the damage multiplier would depress the expected value of fraud to the contractors. The ideal for deterrence would be to calibrate the penalty to the full wealth of the defendant, but considering the defendant's wealth in determining damages is somewhat problematic in U.S. jurisprudence. The FCA is a treble damages statute, so formal adjustments would require legislative change, but government officials might push for more expansive interpretation of damage estimates that would effectively increase the perceived multiplier. Nonetheless, besides the limitations under Supreme Court jurisprudence for high damage multipliers in civil litigation, we still have the concern for judgment-proof contractors who may not be able to pay if the multiplier becomes too high.

2. Probability of detection p

   The probability of detection and success is a function of gamma, the amount of the fraud. The fact that the probability of detection can be calibrated to the level of harm allows us to bypass wealth constraints in deterrence, assuming that we can raise detection to sufficient levels satisfying the reciprocal standard. The relator receives a bounty payment equal to a portion of gamma, the amount of the fraud. To increase p, the government might adjust the relator's share of the recovery, up to the entire amount of the current treble damages. A higher share of the recovery might cause marginal relators to step forward with information. A government operating under options 1 or 2 would grant relators the full amount of the recovery, since the government's share of the recovery is not part of the objective function. Of course, other considerations might limit this full grant: we might be concerned about backroom deals between relator & contractor if the government does not receive any funds.

   Under option 3, the marginal responsiveness of the relators to increased bounty share becomes important. To the extent the relator is motivated by the actual amount of the fraud rather than her particular share of the recovery, increasing her bounty share might not make sense to such a government. One example of such a relator would be a well-to-do idealist whose concern is primarily stopping large fraudsters.

   Outside of manipulating the relator's reward share, the government might be able to shift the probability of success through procedural or resource benefits, thus enticing greater relator participation. A government that shares information resources, provides assurances regarding confidentiality and security, or helps in the settlement/negotiating process may induce higher probabilities of people becoming relators.

3. Contract distribution F_c

   Under this regime, if the value a social planner receives from contract performance is $c - \gamma$, it is in society's interest to avoid any contracts in which $c \leq \gamma^*$, as those provide 0 value, and to maximize contract sizes. Another way of interpreting this result is the desire for consolidation of liability for fraud-we prefer giving our business to a smaller number of large corporations, and we want generous extension of liability for fraud. For example, we would want longer statutes of limitations that would help magnify the amount of fraud liability. The tradeoff would be deviation from the technical optimum of Fc, as society would be receiving less value from these contracts. This may be due to reduced competitiveness of the market, or the reduced diversity of contractors that can precisely satisfy a broad range of desirable attributes such as ownership diversity, capability, and geography.

   Beyond the technical constraints of F_c, increasing liability for fraud is also constrained by defendant liquidity. If the statute of limitations is relatively large, a company may become effectively judgment proof. It will be accumulating liability for large damages it cannot pay.

   The objective function of the government likely does not matter under this situation. Even if p for certain values is sufficiently high such that the expected dollars from catching fraud outweigh the fraud in that bracket, the contractors would similarly realize this and not commit fraud at those levels.

4.3 Coordinating non-private enforcement

Perhaps the most relevant consideration for the government in this paper is its deployment of its own enforcement resources. The government can conduct its own investigations, and it can also apply criminal sanctions. Criminal sanctions generally are reserved for the worst cases, although deterrence theory would suggest that extremely infrequent but harsh sanctions could be useful even for weak offenses. Following general criminal law principles, though, if we accept that there is correlation between wrongfulness and the absolute level of fraud, we would expect to see criminal sanctions applied against the highest levels of fraud. Similar to the private enforcement qui tam system, this would provide deterrence against high fraud but provide little against low level fraud. If, however, we believe that criminal sanctions are a possibility for low-level fraud, it is possible that particularly prison-risk-averse contractors might be deterred even from low-level fraud. On the other hand, a real threat of criminal sanctions to low-level offenders might trigger more fraud-if contractors are going to commit any fraud, they are going to go big. For purposes of this argument, it might be sufficient to believe that criminal prison sanctions will not be particularly relevant to contractors considering small levels of fraud. It seems politically difficult to justify imprisoning individuals who file small but bogus claims with the government.
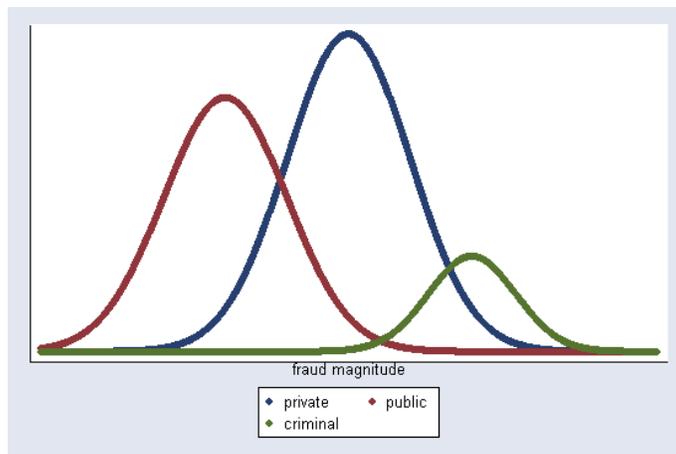
Figure 1: Coordinated Deterrence Strategy

This leaves us with government-initiated civil sanctions. To the extent that civil sanctions overlap with criminal sanctions (i.e., if we apply criminal sanctions, we'll also apply civil sanctions), we expect to see large civil sanctions in parallel with the criminal convictions. For a government focused on deterrence, however, we would expect to see civil sanctions targeted at a level lower than the γ^* of the qui tam system, since those are the cases that are not deterred at all.

Thus, looking at Figure 1, the rightmost portion of cases draws primarily criminal sanctions. These are the most severe cases, and we may see civil sanctions overlapping with these criminal sanctions. The middle portion of cases are primarily handled by the private actors. The leftmost portion of cases will be public civil enforcement. As offense severity goes towards zero, even public civil enforcement will fall as the marginal costs of detecting and litigating these cases exceeds even the deterrence value.

5   Evidence from FCA

My data come from the DOJ's Civil Division, Commercial Litigation Branch's Fraud Section. The data set contains all of the unsealed civil cases since 1986, the time at which the Fraud Section began keeping statistics, up until July 2009. I do not have information regarding criminal cases at this time.

The Fraud Section maintains two distinct databases in tracking qui tam claims. The first database follows the progress of every qui tam claim, starting at the time it is first filed under seal. The second database tracks impositions-final judgments against defendants. Since not all cases are run through the Civil Division, the first database is only accurate to the extent PACER and the local AUSAs keep the Civil Division informed. The second database, containing only

successful results, is considered to be more accurate since it focuses primarily on the identity of the defendant and their imposition, but it contains minimal other information regarding the cases. Although the official website's data stems from the same databases I am using, there is currently substantial mismatch in case volume from the first database.

I will be looking at the coordinated public civil sanction system and also provide some small evidence on the adjustment of p.

The DOJ is not a unitary government actor in this process, but as an umbrella organization, it is the point through which all FCA cases must go at some point.

5.1 Combined deterrence strategy?

Does the government appear to be targeting resources towards lower amounts of fraud in comparison to the purely private cases? Ideally we would want to have information regarding the size of the alleged fraud prior to the government's decision on intervention. Once the government becomes involved, it is very likely it affects the size of case resolution. The government settles the vast majority of cases, bringing only a handful to trial each year. The negotiating power of the government, with the ability to threaten to cut off existing or future contracts, may be rather strong in comparison to the relator operating by herself. Since we are only observing the final imposition in each case, the cases with government involvement may settle for a comparatively larger amount. On the other hand, the government is likely to be a repeat player with the defendants, which may put some downward pressure on the settlement amount. The private relator, who will likely never be able to obtain sensitive information or employment again from that defendant, has a stronger incentive to go for broke in bringing down the defendant.

Using the imposition data as a proxy for the scale of the fraud, the following shows the distribution of successful cases-cases in which there had been final judgment against the defendant. This differs from an actual recovery against the defendant, as they may have been judgment proof. Figure 2 shows the density of intervened cases, cases in which the government intervened on behalf of the relator sorted by the natural log of the imposition. Figure 3 shows the density of non-intervened cases, representing purely private efforts.

Figure 4 shows a direct comparison of the two types of cases. For intervened cases, n=729 Mean: 14.18336 Std Dev: 2.02045 Min: 8.517193 Max: 20.1577. For non-intervened cases, n=192 Mean: 12.05372 Std. Dev.: 2.200645 min: 5.74665 max: 18.77655.

This comparison is of intervened cases against non-intervened qui tam cases. The DOJ is applying more effort on behalf of intervened cases, although this is not to say that the DOJ applies zero effort on non-intervened cases. I await non-qui tam FCA civil actions for a separate reference distribution.

The data suggest that the cases in which the DOJ is involved have higher impositions than those without DOJ participation. We do not see a double-humped distribution that might suggest that the DOJ is pursuing both higher
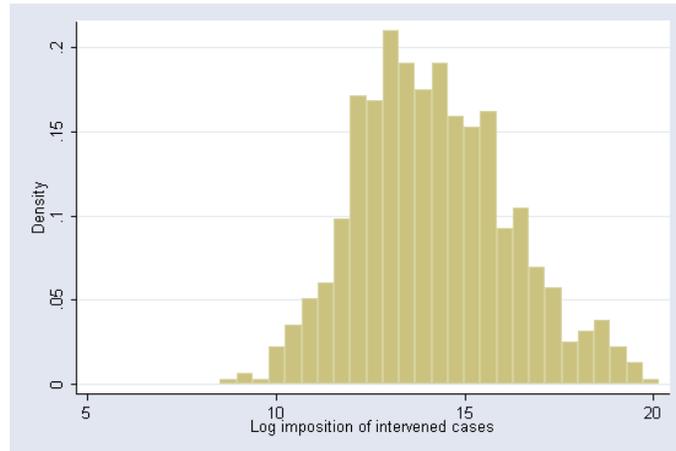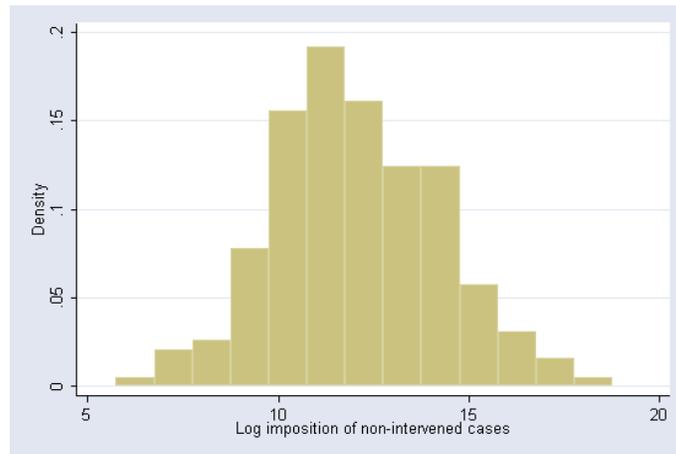
Figure 2: Density of intervened cases

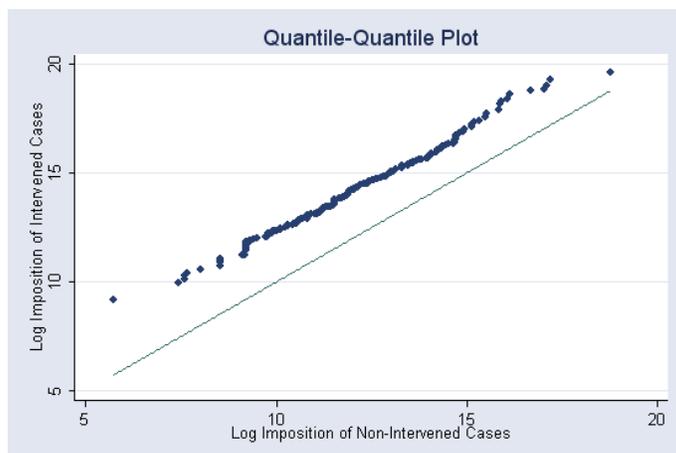Figure 3: Density of non-intervened cases

Figure 4: quartile-quartile plot of intervened vs non-intervened cases

and lower value cases than those run solely by relators.

5.2 Attempts to raise p?

The DOJ has some flexibility with regards to bounty percentage for the relator. If the Government intervenes in the qui tam action, the relator is entitled to receive between 15 and 25 percent of the amount recovered by the Government through the qui tam action. If the Government declines to intervene in the action, the relator's share is increased to 25 to 30 percent.

From the data, the median reward for intervened cases is 18% with a mean of 17% and a standard deviation of .06%. Only about 5% of cases receive clost to the 25% limit. For the non-intervened cases, the median reward is 28% with a mean of 25% and a standard deviation of .08%.

The DOJ does not appear to be aggressively pushing bounty percentages higher to induce greater information disclosure by relators.

5.3 What are we seeing?

In contrast to the combined deterrence strategy, the government appears to be investing in higher value cases of fraud in comparison to the purely private cases. There are various reasons we might not see the predicted strategies for deterrence. It may be that the qui tam process is not providing any deterrence value, or that γ^* is just really high. If the DOJ is simply overwhelmed with the amount of fraud, it might just focus narrowly on deterring or just catching some of the large fraudsters. [test overload w/ density function discontinuity? ln \$5million cutoff = 15.4]

Fraud deterrence may not be the primary objective of the government. It may be that qui tam litigation is not driven by the fraud levels, but rather by

other offenses. The government might only take on cases in which the defendant is committing some other wrong for which the government is really concerned, and the FCA claim is tacked on for additional leverage/penalty. A similar result would occur if the government were simply emphasizing the imposition amounts-objective function 3. Large recoveries provide visible evidence that a government agency is doing useful work.

Alternatively, we may be observing an ill-behaved p function. Unlike the basic model, contractors may exert effort to conceal and fight against allegations of fraud. This effort might scale with the amount of fraud committed, which may deter relators and their attorneys from pursuing these cases aggressively. It is difficult to presume what the relators would have done had the government not intervened in these high value cases, but it is certainly possible that the government believed these high value cases were marginal cases that would not have proceeded had it not intervened. Perhaps the government was also concerned about the protracted litigation and time-value of money for these cases. If it expected the defendants to fight particularly hard against large fraud allegations, the government might intervene to bring its superior threat values in negotiating a prompt settlement. [compare time to settle/resolve?]

6 Extensions

6.1 Deceptive relators & the role of attorneys

One extension to this basic model of relator action is to bring in the possibility of relator error or deception. In this extension, a relator can choose to relate with deceptive information, either exaggerating the extent of fraud or claiming fraud when none exists. This creates a burden on various parties, first of all for the DOJ and the administrative agency reviewing the relator's claim, as they will spend time and effort investigating questionable accusations.

The possibility of deceptive relators may help justify the judicial restraint on pro se qui tam actions. Although not specified in the statute, various jurisdictions forbid pro se relators, thus requiring relators to obtain legal representation. Unlike whistleblowers, attorneys tend to be repeat players in the judicial system, particularly with regards to the qui tam process. Whistleblowers may only gain access to sensitive information once in their careers, particularly if future employers learn of the whistleblower's qui tam past. In contrast, attorneys at a qui tam firm build a reputation with both the judicial system and the DOJ. Requiring whistleblowers to earn an attorney's trust can shift some of the investigative burden from the government to the private sector. It is not unheard of for the DOJ to request that a particular qui tam law firm participate in an action before becoming involved itself.

6.2 Investment model

Perhaps instead of a simple decision of whether or not the potential relator files a qui tam action, the system is designed to induce investment into the qui tam process. Relators and their attorneys first signal the quality of their case through investment. The government then decides whether to invest or not through intervention. The way the government ensures that relators present quality information is by selecting only high value cases in which it is confident to win. If it were to invest in lower quality cases with weaker chances of success, the value of earning the government's intervention would be reduced to the relator. As a result, relators would invest less and offer weaker cases to the government. This particular relator-government equilibrium may induce particular types of fraud behavior among contractors.

7 Recommendations

7.1 Within existing law

7.1.1 Maximizing bounty percentages

If the DOJ is not achieving sufficient deterrence of large fraud through private efforts, it should consider increasing the bounty percentages within the discretionary window allowed under the current statute. Potential relators and their attorneys may be sensitive to marginal increases of compensation, but the DOJ's track record on bounty percentages tends towards the lower end of the spectrum. Understandably, the DOJ wants to maintain an incentive for relators to continue cooperating with the government after the intervention decision. This desire must still be balanced against the importance of signaling to potential new relators of the importance of their participation in addressing fraud.

7.1.2 Pre-intervention announcement settlement negotiation

If the government is currently overloaded and believes there are significant numbers of legitimate cases which merit attention, it might consider revising its procedure prior to announcing its intervention decision. Currently the DOJ informally announces its intervention decision to the relator before making its formal announcement. The defendant learns of the decision at the formal announcement. If the DOJ declines to intervene, this is a negative signal to the relator regarding the merits of the case and a positive signal to the defendant regarding its future prospects. This action may be discouraging private enforcement of cases the DOJ wishes to see proceed.

Instead, for cases in which the DOJ believes the law and facts are relatively clear, it could bring the defendant and relator into the discussion before making its intervention announcement. If all the parties could pursue a settlement amount before intervention, a defendant might be more likely to settle given the uncertainty about the government's upcoming intervention decision. The

DOJ would have to credibly commit, however, that it would not settle for a similar or lower amount in the future should it decide to intervene; otherwise, it would likely be in the defendant's best interest to wait for the intervention decision before acting.

7.2   Requiring legislative changes

7.2.1   Increasing the bounty percentage

If we are not deterring large scale fraudsters because private relators are not coming forward with important information regarding large amounts of fraud, Congress should consider increasing the bounty percentage framework. As the 1986 FCA amendments showed, increasing the potential bounty payout along with legal protection for relators triggered hundreds of new cases each year. If the large cases go undetected, saving money on an uncollectible bounty does not make sense.

7.2.2   Scaling low-end bounty percentage

Perhaps more importantly for deterrence purposes, however, is the lower end of fraud cases. If the political or organizational structure of the government prevents pursuit of the lower level fraudsters, perhaps we should implement an incentive structure to encourage the private sector to do so. One method would be a minimum recovery amount for successful qui tam actions, an amount sufficient to cover at least some estimate of the private harms to the relator. For example, we might establish the minimum recovery as $200,000. If, after successful litigation, the damages are less than the minimum recovery, the Department of Justice will be responsible for paying the differential to the relator. If the damages are greater, the relator will receive the 15% (or higher, as determined by the statute) of the remaining recovery. Alternatively, we can implement a "transition share" that would provide a little more money to the government. If the damages are greater than the minimum recovery amount, the relator will receive the minimum recovery amount plus a small percentage share of the remaining recovery (which we will call the transition share). This formula stays in effect until the damages are great enough that the relator can receive a straightforward 15% (or higher) share of the full damage amount. As an example, let us assume the minimum recovery level is $200,000, the transition share percentage is 2% [maybe call this a "clawback" percentage], and, for simplicity's sake, a normal relator share of 25%. In every case, the relator would receive attorneys' fees upon successful litigation. If the damages after trial are $100,000, the relator would receive $100,000 from the defendant along with $100,000 from the DOJ. If damages after trial are $250,000, the relator would receive $200,000 plus 2% of $50,000, which totals $201,000 from the defendant, while the DOJ would recover $49,000 from the defendant. If damages are $1 million, the relator would receive $250,000 (25% of the total), and the remainder would go to the DOJ. The "switchover" point in this case would be approximately $870,000 in

damages, under which the transition share percentage would be in play. Over $870,000, the straightforward 25% would be the relator's share. The minimum recovery amount serves two purposes: first, it severs the link between relator recovery and the fraud she observes. As the magnitude of damages from fraud decreases, we do not expect the magnitude of the private harm to the relator to decrease at the same rate. This gives the relator some increased certainty of compensation. Second, the minimum recovery amount gives the Attorney General incentive to evaluate smaller value claims more thoroughly-in particular, to consider seriously the option of dismissal. Since the government will be on the hook for some portion of the damages, the attorney general has an increased motivation to not allow qui tam litigation to proceed unless the social value of the action exceeds the cost of payout for the DOJ. We can still imagine cases in which a relator identifies a type of difficult to detect fraud by a particular defendant for a limited amount of money, but the revelation of this previously unknown type of fraud in an entire industry provides much larger social value to the public and the government-even if the damages from that one defendant are limited. The transition share percentage provides some marginal incentive to the relator to not settle for the minimum recovery amount. This applies some linkage between the relator's interest and the government's interest in recovery. Thus, even though she is guaranteed to receive the minimum recovery amount, since she continues to receive a percentage of the damages above the minimum recovery, she has some incentive to ensure that the government also collects.

8   Conclusion

Criminal deterrence can bring society to a happy equilibrium in which we enjoy low crime combined with low amounts of costly litigation and remediation. As a social planner selecting some combination of private and public enforcement tools, we can see that rewarding private enforcement through percentage bounties can help generate deterrence against more harmful crimes. To balance against an outbreak of less harmful crimes, we might expect public enforcement efforts to focus on such smaller cases for which the private agents might have insufficient incentive. Such a combined deterrence strategy might be particularly valuable in fighting financial crimes in a complex economy due to the detection capabilities of the private sector.

As we have seen under the False Claims Act, private relators responded to the bounty percentages by bringing in thousands of new cases since 1986. The corresponding government agencies, however, in concert with the DOJ, appear to have invested their resources in higher dollar fraud cases, contrary to the coordinated deterrence strategy.

If the trend for public enforcement tends towards more harmful crimes, it may be that we should reconsider the percentage bounty system for motivating private enforcement. In a qui tam system, enhancing the low end of bounties may improve the deterrence and enforcement effects of private relators. Such a design may complement limitations within the public enforcement sphere.

References

[1] The Economic Theory of Public Enforcement of Law A. Mitchell Polinsky and Steven Shavell Journal of Economic Literature, Vol. 38, No. 1 (Mar., 2000), pp. 45-76

[2] Becker, Gary S. and George J. Stigler. 1974. "Law Enforcement, Malfeasance, and Compensation of Enforcers," J. Legal Stud., 3, pp. 1-18.

[3] Landes, William M. and Richard A. Posner. 1975. "The Private Enforcement of Law," J. Legal Stud., 4, pp. 1-46.

[4] Polinsky, A. Mitchell. 1980. "Private versus Public Enforcement of Fines," J. Legal Stud., 9, pp. 105-27.

[5] Shavell, Steven. 1993. "The Optimal Structure of Law Enforcement," J. Law Econ., 36, pp. 255-87.