A Theory of "Crying Wolf":
The Economics of Money Laundering Enforcement

Előd Takáts*

April 20, 2006

Abstract

The paper shows how excessive reporting, called "crying wolf", can dilute the information value of reports. Excessive reporting is investigated by undertaking the first formal analysis of money laundering enforcement. Banks monitor transactions and report suspicious activity to government agencies, which use these reports to identify investigation targets. Banks face fines should they fail to report money laundering. However, excessive fines force banks to report transactions which are less suspicious, thereby diluting information. The empirical evidence is shown to be consistent with the model's predictions. The model is used to suggest implementable corrective policy measures, such as decreasing fines and introducing reporting fees. Furthermore, crying wolf is shown to be a general economic problem relevant to corporate finance, especially after the Sarbanes-Oxley Act.

JEL classification: G28, K23, L51, M21
Keywords: Money Laundering, Patriot Act, Disclosure, Auditing, Sarbanes-Oxley Act

1 Introduction

Reports identifying problems are ubiquitous in the economy. Generalist CEOs depend on reports from specialists to identify pressing legal or technical issues. The President receives security reports identifying the most relevant threats. The public and investors turn to auditors to identify transactions which fundamentally affect the value of listed firms, especially so after the Sarbanes-Oxley Act. Law enforcement agencies combating money laundering and terrorism financing rely on banks to identify suspicious activity.

The paper shows formally how excessive reporting in such situations fails to identify what is truly important by diluting the information value of reports. The intuition can be best understood through an analogy with the tale: "The boy who cried wolf". In the tale, the boy rendered his cries useless by resorting to them too often, and failing to identify the wolf's presence. Similarly, excessive reporting, which will be referred to as "crying wolf", fails to identify what is truly relevant. More generally, the crying wolf phenomenon shows that information is not only data, but also able and expert identification of truly important data.

The reporting problem is investigated through the first formal analysis of money laundering enforcement. Choosing money laundering enforcement as the leading example is motivated by the fact that the identification role of reports is particularly strong. Furthermore, money laundering is an economically significant crime. Several hundred billion dollars are washed through the financial sector in the United States, and money laundering facilitates crimes as harmful as drug trafficking and terrorism, as detailed in the next section. Investigating reporting in this context is also motivated by a highly publicized case, when Bob Dole, former Senate majority leader and presidential candidate, was falsely reported for money laundering (Wall Street Journal, 2004a).

The model explores the agency problem between the bank and government law enforcement agencies. The bank monitors transactions and reports suspicious activity to the government, which identifies targets for investigations based on these reports. The bank undertakes costly monitoring and reporting, because the government fines it if money laundering is successfully prosecuted and the bank did not report the transaction. Though Masciandaro (1999) abstracted from this agency problem in the first economic analysis of money laundering, it is crucial as the later survey by Masciandaro and Filotti (2001) shows.

The formal model builds on five main economic building blocks. First, communication is coarse between the bank and the government, as the bank cannot communicate in a short report all the local information it has. This communication problem is similar in spirit to the information hardening problem in Stein (2002), though here the problem is not with verifying the information, but rather with telling it precisely. Second, the bank's incentives to report are coarse; the bank is fined only for false negatives, i.e. for not reporting transactions which are prosecuted later as money laundering. Third, the bank is always uncertain about the transaction's true nature, i.e. every transaction can be potential money laundering. Fourth,

2

the bank faces dual tasks: it has to monitor all transactions in order to report the suspicious ones. Fifth, the bank's information, i.e. its signal on the transaction, is not verifiable ex-post, because the local information at the time of the judgment cannot be reproduced later.

The model shows that harmful excessive reporting, called crying wolf, can arise in this setup. As the bank cannot share its signal with the government, the government must make decisions based on whether or not it observes the report. Intuitively, if the bank identifies all transactions as suspicious, then it fails to identify any one of them - exactly as if it would not have identified a single one. Thus, crying wolf can fully eliminate the information value of reports. Crying wolf can arise because excessively high fines for false negatives force the uncertain bank to err on the safe side and report transactions which are less suspicious. In the extreme case the bank is forced to report all transactions, thereby fully diluting the information value of reports.

The paper shows that the model's findings are consistent with the available empirical evidence in the United States. Fines have increased in the last ten years, especially so after the Patriot Act. In response, banks have reported an increasing number of transactions. However, the number of money laundering prosecutions has fallen - even though the estimates of money laundering volumes have been stable. Furthermore, regulatory agencies have identified 'defensive filing' which exhibits striking similarities with what happens under crying wolf.

The model also provides implementable policy implications on how to stop crying wolf and thereby increase the efficiency of money laundering enforcement. First, the model calls for reduced fines to cease crying wolf, as optimal and not maximal fines are needed. The bank needs some fines in order to monitor and report, but excessively strong ones result in crying wolf. The intuition behind deviating from Becker's (1968) seminal proposal of maximal deterrence is that banks are not criminals, but rather honest informants. Thus, excessively strong incentives do not deter banks from committing a crime, but rather distort their information provision.

Second, reporting fees might be needed to elicit optimal reporting. As in Holmström and Milgrom (1991), single dimensional incentives might not be able to elicit efficient two-dimensional banking effort to monitor and report. Reporting fees provide an implementable second dimension of incentives by punishing the bank for false positives. Furthermore, reporting fees can be thought of as pricing reporting externalities. Each report dilutes the value of all other reports, and reporting fees would make banks internalize these externalities.

Third, the model's most important comparative static result shows that fines should decline in the harm caused by money laundering. The intuition is that the bank's effective incentives have two components: government investigation to identify false negatives and nominal fines. As money laundering becomes more harmful, optimal government investigation increases as the marginal benefit of prosecuting money laundering increases. However, the bank's incentives should be constant so as not to trigger crying wolf. Thus, the model shows that the fine increases of the Patriot Act could have been intuitive, but mistaken measures.

As it was argued earlier, crying wolf is a general economic problem. The main economic

building blocks identified in money laundering enforcement can be found in many other situations. For instance, product information provision is very similar to suspicious activity reporting. Firms must use their expertise to identify the most relevant dangers related to using their product. There, coarse incentives are provided by the legal structure: omissions of warnings can result in damages and following lawsuits. False positives have no such easily identifiable victims. However, there is a real damage from crying wolf, because information is lost, as customers disregard warnings and accept contracts without reading the fine print.

The agency setup of the model is very similar to that of the auditing problems analyzed first in Tirole (1986). The government uses both auditors and banks to obtain information about their clients. More precisely, the government is interested in learning if there are problems such as accounting omissions or signs of money laundering. Naturally, both auditors and banks are reluctant to provide such negative information, which creates an agency problem. These similarities in the agency setup allow building on the Kofman and Lawarrée (1993) auditing model in setting up the action set. A major difference is, however, that the auditing literature focused on the disclosure of verifiable or certifiable information as reviewed in Verrecchia (2001).

The model's focus on coarse communication of unverifiable information is particularly relevant to investigate the increased role of auditors after the Sarbanes-Oxley Act. Auditors are not only supposed to disclose verifiable information, but they also have to identify material transactions, i.e. transactions that fundamentally affect the firm's value. Identifying material transactions is very similar to identifying suspicious activities, because in both cases reporting involves identification and coarse communication of unverifiable and uncertain information. Furthermore, auditors are also sanctioned for false negatives, i.e. for not disclosing transactions which later substantially affect the firm's value. Thus, excessive fines might make auditors report more transactions as material, thereby failing to identify the truly important ones. Lengthening auditing reports and firm disclosures, documented in Gordon (2005), might well signal crying wolf.

The rest of the paper is organized as follows. Section 2 details how money laundering enforcement works. Section 3 sets up the model. Section 4 solves the model and demonstrates crying wolf. Section 5 analyzes comparative statics. Section 6 links the model to available empirical evidence. Section 7 discusses the policy implications and other contexts where the model could be used. Section 8 concludes. The appendix details most proofs and formal extensions such as numerical estimation of a non-linear version of the model.

2 Money Laundering Enforcement

Money laundering is defined as an illicit money transfer. There are two main kinds of illicit money transfers. First, traditional money laundering entails transferring illegally obtained funds to conceal their origins and make them appear legal. For example, drug dealers deposit cash revenues in banks and later transfer them until the funds appear to originate from legit-

imate sources. Second, terrorism financing entails transferring mostly legal funds for illegal purposes. For instance, legal charity donations are transferred to fund terrorist attacks. In sum, both forms of money laundering are characterized by illicit and socially harmful fund transfers. Money laundering causes social harm because it facilitates crime and enables criminals to enjoy criminal revenues.

Money laundering can happen through various intermediaries. Bank transfers, both by wire and check, are the most common channels for illicit money transfers as described in Reuter and Truman (2004). Money transmitting businesses, such as Western Union, are also used for money laundering as detailed in The Wall Street Journal (2004b). These businesses are typically franchised or owned by individuals, who might have stronger incentives to turn a blind eye to money laundering than bank branch-managers. In the greyer area of finance, informal value transfer systems (IVTSs) provide money transmitting services usually without a proper paper trail. The *hawala* or *hindi* systems used by different ethnic communities are described, for instance, in El-Qorchi (2002).

Money laundering is an economically significant crime, though precise estimates are hard to obtain. According to Camdessus (1998), the consensus range of money laundering volume is between 2 and 5 percent of the global GDP. The FBI (2001) estimates the volume of globally laundered funds as falling between \$600 billion and \$1.5 trillion. Laundering fees, i.e. what money launderers charge their criminal clients, are estimated at 5-15% of the laundered amount according to Lal (2003) and Reuter and Truman (2004). Thus, money laundering, including self-laundering, is estimated to be a \$30 to \$225 billion global 'industry'. Moreover, the harm caused by money laundering and its predicate crimes shows that money laundering is even more significant than what volumes and laundering revenues would suggest.

Money laundering enforcement is particularly relevant for the United States.[1] According to some lawmakers' estimates, half of the globally laundered funds are transferred through US Banks (FBI, 2001). Three known money laundering cases highlight the point. First, \$7 billion of Russian funds were washed through the Bank of New York until 1999 (Reuter and Truman, 2004). Second, Stephen Saccoccia alone laundered up to \$550 million of drug money for both the Calí and Medellin cartels until he was prosecuted in 1993 (Reuter and Truman, 2004). Third, terrorists transferred \$0.5 million for the 9/11 attack (9/11 Commission Report, 2004). The examples also show that the volumes involved in laundering proceedings of tax evasion and drug trafficking are enormous compared to terrorism financing.

Money laundering enforcement relies primarily on bank reporting to law enforcement and

---

[1]Responding to the threat of money laundering, the United States has developed one of the strongest anti-money laundering regulation. The Banking Secrecy Act (1970), which in fact curbed banking secrecy to fight money laundering, was followed by a series of laws, each one of them further strengthening money laundering enforcement: The Money Laundering Control Act (1986), the Annunzio-Wylie Money Laundering Act (1992), the Money Laundering Suppression Act (1994), The Money Laundering and Financial Crimes Strategy Act (1998) and finally the USA Patriot Act (2001).

government agencies as reviewed in Reuter and Truman (2004). Banks provide two kinds of reports: rule-based and discretionary reports. For instance, banks file the rule-based currency transaction report (CTR) for any cash transactions exceeding $10,000. Enforcing rule-based reporting is a standard disclosure problem as bank actions are ex-post verifiable. They are, however, insufficient, because money launderers are aware of the rules and can circumvent them. For instance, money launderers usually 'smurf', i.e. break down large cash deposits over $10,000 into smaller deposits below the reporting threshold. Nevertheless, rule-based reporting makes money laundering more cumbersome.

The weaknesses of rule-based reports led to the introduction of a discretionary report, the suspicious activity report (SAR), in 1996. The suspicious activity report is filed for any activity that the bank considers to be 'suspicious'. For instance, if the bank spots several transactions just below $10,000, it can identify them as suspicious because they hint at smurfing. The notion of suspicion is intentionally left vague so as to leave both money launderers and banks uncertain. Thus, money launderers cannot rely on simple rules to avoid being reported. Furthermore, banks are forced to constantly improve their understanding of how money laundering is done. This intentional vagueness can be understood as another form of constructive ambiguity. The SARs are filed to the Financial Crime Enforcement Network (FinCEN), which forwards them to law enforcement agencies for further investigation.

Banks incur costly screening, monitoring and reporting because of the threat of sanctions.[2] Failing to file SARs led to fines of $25 million for Riggs Bank (FinCEN, 2004a), $24 million for Arab Bank (FinCEN, 2005g) and $50 million for AmSouth Bank (Wall Street Journal, 2005 and FinCEN, 2004b). Most importantly, sanctions or fines are levied for false negatives, i.e. for not reporting transactions which are later prosecuted as money laundering or judged to be suspicious ex-post.[3] Banks are not fined for false positives, i.e. for reporting legal transactions as money laundering. There are no 'judgment calls' to second guess the validity of reporting.

3   Model Setup

The model explicitly investigates the agency problem between the government and the bank, where the government needs information from the bank to investigate the transaction which may or may not be money laundering The setup is illustrated in Figure (1). If the bank monitors the transaction, then it receives information, a signal about the transaction. The informed bank, which has received the signal is able to inform the government by filing a

---

[2]Sanctions include nominal fines (civil money penalties), the costs of public law enforcement actions (cease and desist orders or written agreements), the costs of private law enforcement actions (memoranda of understanding), and also implicit reputation costs.

[3]Regulatory language is somewhat different: banks are fined if suspicious activity detection and reporting procedures are not in place. Yet, the test of these policies is whether banks are able to identify and report those transactions which are considered to be suspicious ex-post.
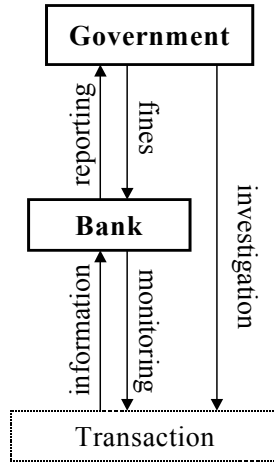
Figure 1: Model scheme

suspicious activity report. The government provides incentives, by means of fines, for the bank to monitor and report suspicious transactions.

3.1   Economy and Players

The economy of the model consists of a single money transfer. The transaction is either money laundering or a legitimate transfer. The prior probability that the transaction is money laundering is $\alpha \in (0, 1/2)$. Money laundering causes harm ($h > 0$) to society.

Two players are modeled explicitly: the government and the bank, who form a principal-agent relationship. Both players are risk neutral. The bank maximizes private profit and the government maximizes social welfare.

The costs of monitoring and reporting, which are defined formally later, decrease the bank's private profit. The profit is naturally decreased by the fine (F). The transaction fee and the cost of transaction are normalized to zero, and thus they do not affect the bank's profit.

The harm caused by money laundering (h) decreases social welfare. However, prosecuting money laundering increases social welfare by ρh, where $\rho > 0$. Parameter ρ represents the recovery rate, the portion of harm prevented by the prosecution of money laundering. Utility ρh can be interpreted as a reduced form representation of utilities from asset seizure, deterring money laundering and predicate crime, and finally from preventing future crimes by intercepting the money flow. Naturally, prosecuting money laundering is more useful the more harmful money laundering is. Fines do not affect social welfare, as they represent simple transfers from the bank to the government. Finally, social welfare is decreased by government investigation costs and bank monitoring and reporting costs.

The agency problem arises because the bank does not internalize the social gains stemming from the prosecution of money laundering, ρh. Thus, to implement socially desirable policies

7

(and to make the bank monitor and report), the government uses fines.

3.2 Timing

There are five periods in the model:

1. Nature selects the true nature of the transaction (legal or money laundering).

2. The bank exerts monitoring effort to learn the signal.

3. The informed bank, which has observed the signal, decides to report. (If the bank has not observed the signal, it can not report.)

4. The government observes the report and sets investigation effort.

5. The government fines the bank, if money laundering was prosecuted and the bank did not report the transaction.

3.3 Signal Structure

The bank might receive an informative signal (σ) about the transaction. The signal is assumed to be binary, and it takes either high (1) or low (0) value, $\sigma \in \{0, 1\}$. In case of money laundering, the signal takes the high value 1 with probability δ, and the low value 0 with probability $1 - \delta$. For the legal transaction, the signal takes the low value 0 with probability δ, and the high value 1 with probability $1 - \delta$. Under symmetry probability δ can be interpreted as the precision of the signal. The higher δ, the more likely that the high signal indicates money laundering, and the low signal indicates a legal transaction. The probabilities are summarized below:

	Money Laundering (α)	Legal Transaction $(1 - \alpha)$
Low Signal (0)	$1 - \delta$	δ
High Signal (1)	δ	$1 - \delta$

Probability δ is restricted to being more than one-half: $\delta \in (1/2, 1)$. This implies that money laundering is more likely to trigger the high signal, and legal transaction the low signal. Thus, observing a high signal implies that the transaction is more likely to be money laundering.

The posterior probabilities of money laundering can be determined through straightforward Bayesian updating. The posterior probabilities, β_0 and β_1 denote, respectively, the likelihood of money laundering given that the signal is low (β_0) or high (β_1).

$$\beta_0 = \Pr(\text{ML}|\sigma = 0) = \frac{\alpha(1 - \delta)}{\alpha + \delta - 2\alpha\delta}$$

$$\beta_1 = \Pr(\text{ML}|\sigma = 1) = \frac{\alpha\delta}{1 - \alpha - \delta + 2\alpha\delta} > \beta_0$$

3.4 Action Sets

The government investigates transactions and fines the bank. The investigation effort (I) exerted by the government determines the probability that the government discovers the truth about the transaction, i.e. whether the transaction is legal or money laundering. Uncovered money laundering is prosecuted. As investigation effort I represents a probability, the government's effort choice is naturally constrained to the unit interval. The investigation effort is costly, and it is assumed to be quadratic in the investigation effort with parameter $k > 0$: kI^2. Furthermore, the government can condition its investigation effort only upon receipt of the bank's report. The investigation effort chosen given no reporting is indexed as I_0, the investigation effort given reporting is indexed as I_1. The government is able to commit to equilibrium, best response investigation actions. Thus, when multiple equilibria are possible the state can implement the equilibrium with the highest social welfare.

The government can also impose fines ($F \geq 0$) on the bank. As discussed earlier, fines are socially costless to impose, as fines represent welfare transfers from the bank to the government. The government imposes fines if the bank does not report the transaction, which is later prosecuted as money laundering. The government is able to commit to levying preset fines. In sum, the action set of the government has three elements: $(I_0, I_1, F) \in \left([0,1]^2 \times [0,\infty) \right)$.

The bank monitors the transaction and reports to the government. Bank monitoring effort (M) is assumed to be binary, and it takes either high ($M = 1$) or low values ($M = 0$), thus $M \in \{0,1\}$. The monitoring effort determines - as in Kofman and Lawarrée (1993) - the probability that the bank receives its signal (σ). Thus, with high effort the bank always observes the signal, whereas with low effort it never observes the signal. The bank's monitoring cost for high effort ($M = 1$) is normalized to $m > 0$ and for low effort ($M = 0$) to zero.

If the bank did not receive the signal, it cannot report.[4] If the bank received the signal, it decides on reporting based on the signal. Action R_0 denotes the bank's filing decision if the observed signal is low (0). If the bank reports after observing the low signal, then $R_0 = 1$, or else $R_0 = 0$. Similarly, after observing the high signal, the bank either reports ($R_1 = 1$) or does not report ($R_1 = 0$). The bank incurs cost $c > 0$ by reporting.[5] Thus, the action set of the bank has three elements: $(M, R_0, R_1) \in \left(\{0,1\}^3 \right)$.

3.5 Final Assumptions

Four assumptions finalize the problem setup. First, the solution focuses on pure strategy subgame perfect Nash equilibria in order to ease interpretation. Second, it is assumed that parameters are such that in the first best equilibrium, when the agency problem is abstracted away, the bank monitors and reports. This is consistent with the basic premise behind the

---

[4] The assumption relies on the empirical fact that banks need to provide detailed information in order to file the three page SAR.

[5] According to FinCEN completing a SAR filing takes between 1/2 and 3 hours for a compliance officer.

money laundering enforcement regime, i.e. banks need to monitor their clients. The necessary and sufficient parameter restrictions for the second assumption are:[6]

$$\frac{[\alpha(1-\delta)\rho h]^2}{4k\,(\alpha+\delta-2\alpha\delta)} + \frac{(\alpha\delta\rho h)^2}{4k\,(1-\alpha-\delta+2\alpha\delta)} - \frac{(\alpha\rho h)^2}{4k} > (\alpha+\delta-2\alpha\delta)\,c + m \qquad (1)$$

The parameter restriction implies that bank monitoring and reporting (right hand side) is relatively cheap compared to the gains derived from prosecuting money laundering (left hand side).

Third, it is assumed that government investigation is sufficiently costly to ensure interior investigation solutions. The formal parameter restriction is:

$$\frac{\alpha\delta\rho h}{2\,(1-\alpha-\delta+2\alpha\delta)} < k \qquad (2)$$

In other words, restriction (2) implies that the government never sets first best investigation to unity. The reason for this is that investigation is so costly, that some uncertainty in prosecution is preferred to spending the resources required to establish the truth with certainty.

Finally, for tie-breaking it is assumed that whenever they are indifferent, both players take actions which are better for the other player. This can be interpreted as a weak understanding of common goals.

4 Solving the Model

The model is solved in four steps. The first subsection rewrites the model in an equivalent form, which eases the exposition. The second subsection derives the benchmark first best equilibrium. The third subsection characterizes second best equilibria, where the bank's incentive problem is analyzed. The fourth one analyzes the second best game under exogenously set fines. Most importantly, this subsection demonstrates the crying wolf problem. Namely, that excessively high fines are not only costly in terms of investigation, but also decrease the expected number of money laundering events that are prosecuted.

4.1 Equivalent Problem

The bank's action set can be rewritten as (M, T), where $T \in \{0, 1\}$ is a reporting threshold, instead of (M, R_0, R_1).[7] The bank reports all signals weakly higher than T, and does not report signals below it. This rewriting of the actions set, however, excludes the $(R_0 = 1, R_1 = 0)$ action pair, while allowing for all other combinations. Thus, it has to be shown that the $(R_0 = 1, R_1 = 0)$ action pair can be excluded without loss of generality both from the first best and second best equilibria.

---

[6] Both equation (1) and (2) follow from the proof of Proposition (1) in the Appendix.

[7] Notice that constraining $T \in \{0, 1\}$ is without the loss of generality. Allowing for $T \in [0, 1]$ would produce exactly the same equilibrium reporting thresholds, i.e. either zero or one.

In the first best equilibrium social welfare is maximized while the bank's incentive considerations are abstracted away. Consequently, in the maximization and government investigation setting only the signal's information content matters. Reporting only the low signal and not the high signal ($R_0 = 1, R_1 = 0$) provides the same information as reporting only the high signal and not the low signal ($R_0 = 0, R_1 = 1$). However, as money laundering is rare and the signals are sufficiently precise, the low signal is more likely. Thus, reporting the low signal and not reporting the high signal is more expensive, and therefore this reverse reporting cannot be part of a first best equilibrium. The reasoning is formalized in Lemma (1).

Lemma 1 *The ($R_0 = 1, R_1 = 0$) action pair does not arise in the first best equilibrium.*

Proof. Follows from the discussion above. To see that the low signal is more likely:

$$\Pr(\sigma = 1) = 1 - \alpha - \delta + 2\alpha\delta < \alpha + \delta - 2\alpha\delta = \Pr(\sigma = 0)$$

It is equivalent with

$$1/2 < (1 - \alpha)\,\delta + \alpha\,(1 - \delta)$$

which directly follows from $0 < \alpha < 1/2 < \delta < 1$. ■

In second best equilibria, the social welfare maximization is constrained by incentive considerations. The bank is motivated to report because, absent reporting, it might get fined. Thus, whenever the expected fine is higher than the reporting cost, the bank reports. However, the expected fine depends on the likelihood of money laundering. Thus, if the bank reports given the low signal (when the probability of laundering is low), then it should also report under the high signal (when the probability of laundering is high). Lemma (2) shows formally that the ($R_0 = 1, R_1 = 0$) pair can be excluded from the second best equilibrium without loss of generality.

Lemma 2 *In second best equilibria, the ($R_0 = 1, R_1 = 0$) action pair does not arise, if in equilibrium $M = 1$. Furthermore, if in equilibrium $M = 0$, then the ($R_0 = 1, R_1 = 0$) action pair can be replaced with any other action pair without loss of generality.*

Lemmas (1) and (2) allow for rewriting the problem as formalized in Corollary (1).

Corollary 1 *The action set of the bank can be rewritten as (M, T), where $T \in \{0, 1\}$ is a reporting threshold. The bank reports, if $\sigma \geq T$, and does not report, if $\sigma < T$.*

Proof. Follows from Lemmas (1) and (2) and noticing that except for ($R_0 = 1, R_1 = 0$) the new action set can replicate all actions of the original action set. ■

Next, using the equivalent action set six probability shortcuts are introduced. These probabilities will be used to present the model in a concise form. The probability notations conditional on reporting threshold T are summarized below:

p_T probability of reporting

q_{0T} probability of money laundering given no reporting

q_{1T} probability of money laundering given reporting

The probability that the informed bank reports is denoted by p_T where $T \in \{0, 1\}$. If $T = 0$, then the informed bank reports all signals, and thus $p_0 = 1$. If $T = 1$, then the bank only reports the high signal, thus $p_1 = 1 - \alpha - \delta + 2\alpha\delta$.

The probability of money laundering conditional on reporting is denoted by q_{1T} and conditional on no reporting by q_{0T}. The probabilities depend on the reporting threshold T. Unit value of reporting threshold $(T = 1)$ implies that the report perfectly conveys the bank's signal. This would mean that the fact of reporting implies that the signal is high, thus $q_{11} = \beta_1$. Similarly, no reporting implies that the signal is low, consequently $q_{01} = \beta_0$. However, zero reporting threshold $(T = 0)$ renders reporting uninformative. Thus, the probability of money laundering is the same with or without the report, and equals the unconditional probability $q_{10} = q_{00} = \alpha$.

4.2   First Best Benchmark

The first best case, when the bank's incentive problem is abstracted away, is investigated first. Using the equivalent problem and the probability shorthands derived before, the first best problem can be set as a social welfare (W) maximization problem:

$$\max_{I_0, I_1, F, M, T} W = M\left[(1 - p_T)(q_{0T}I_0\rho h - kI_0^2) + p_T(q_{1T}I_1\rho h - kI_1^2 - c) - m\right] \qquad (3)$$
$$+(1 - M)\left[\alpha I_0\rho h - kI_0^2\right] - \alpha h$$

Interpreting equation (3) is straightforward. The first term shows welfare, when the bank is informed, which happens with probability M. The informed bank, however, does not report with probability $(1 - p_T)$. In this case the probability that the transaction is money laundering is q_{0T}. Money laundering is prosecuted then with probability I_0, which yields utility ρh. The government's investigation cost is kI_0^2. The informed bank reports with probability p_T. Then the probability that the transaction is money laundering is q_{1T} and money laundering is prosecuted with probability I_1. The government's investigation cost is kI_1^2 and the bank incurs reporting cost c. Finally, the bank has to incur cost m to be informed.

The second term depicts welfare when the bank is uninformed, which happens with probability $(1 - M)$. Then the probability that the transaction is money laundering is the unconditional probability α. Money laundering is prosecuted with probability I_0. The government's

investigation cost is kI_0^2. The third and last term summarizes the social losses (h) stemming from money laundering, which happens with probability α.

Equation (3) immediately shows two properties of the first best problem. First, if there is no money laundering enforcement regime in place, i.e. the bank does not monitor or report and the government does not investigate, then social welfare is $-\alpha h$. Second, in the first best problem redistributive fines (F) do not play any role.

The first best solution can be determined in an intuitive manner by focusing on the limited number of bank actions. The government investigation best responses to these bank actions can be determined straightforwardly. Furthermore, these government best responses allow easy quantification of the welfare resulting from each bank action pair. The bank action pair $(M = 1, T = 0)$ can be immediately excluded as it cannot be the first best equilibrium. Under this action pair the bank provides useless information (reports all signals) at a positive monitoring (m) and reporting (c) cost. Welfare could be trivially improved by not reporting and not monitoring.

This leaves two possible first best equilibria. First, the equilibrium where the bank does not investigate $(M = 0)$ and does not provide any information to the government $(T \in \{0, 1\})$. Note that, if the bank is uninformed, then the reporting threshold choice is irrelevant, because the bank never reports. Second, the bank investigates $(M = 1)$ and reports only with high threshold $(T = 1)$. However, the parameter restriction set in equation (1) implies that the equilibrium with bank investigation and high reporting threshold provides higher social welfare. Thus, it is the unique pure strategy first best equilibrium. Proposition (1) formalizes the argument.

Proposition 1 *In the first best equilibrium, given condition (1) the bank investigates $(M = 1)$ and reports only the suspicious, high signal transaction $(T = 1)$. The government's best responses conditional on no reporting (I_0^*) and reporting (I_1^*) are determined in terms of parameters in the following:*

$$I_0^* = \frac{q_{01}\rho h}{2k} = \frac{\alpha(1-\delta)\rho h}{2k\left(\alpha + \delta - 2\alpha\delta\right)} < I_1^* \tag{4}$$

$$I_1^* = \frac{q_{11}\rho h}{2k} = \frac{\alpha\delta\rho h}{2k\left(1 - \alpha - \delta + 2\alpha\delta\right)} < 1 \tag{5}$$

The government does not levy fines $(F = 0)$ in the first best equilibrium.

Finally, in order to ease further discussion, equilibria which implement the first best welfare and actions are defined formally.

Definition 1 (First Best equilibrium) *The equilibrium which implements the first best social welfare (denoted as W^*) and actions derived in Proposition (1), except for fines, is called First Best equilibrium with fine F.*

4.3 Second Best

In the second best problem, the government aims to maximize social welfare as in the first best problem (3) subject to the bank's profit maximization (7):

$$\Pi(M, T, I_0, I_1) = -M\left[(1 - p_T)\, q_{0T} I_0 F + p_T c + m\right] - (1 - M)\alpha I_0 F \qquad (6)$$

$$IC \qquad \{M, T\} = \arg\max \Pi(M, T, I_0, I_1) \qquad (7)$$

Banking profit (Π) shown in equation (6) has two main terms. First, the bank is informed with probability M. The informed bank does not report with probability $(1 - p_T)$, however, in this case the transaction might still be money laundering with probability q_{0T}. The government uncovers money laundering with probability I_0, and fines the bank with F. The bank reports with probability p_T at filing cost c. Finally, in order to be informed, the bank incurs monitoring cost m.

Second, the bank is uninformed with probability $(1 - M)$, and then it cannot report. The transaction is money laundering with probability α, and the government prosecutes money laundering with probability I_0, thereby fining the bank with F.

Obviously, the government prefers to implement the First Best equilibrium, which requires actions ($M = 1$, $T = 1$, $I_0 = I_0^*$, $I_1 = I_1^*$). The problem with implementing the first best actions is that incentive compatibility requires both weak and strong fines. On the one hand, fines should be high enough that the bank would not deviate to no monitoring ($M = 0$). On the other hand, fines should be sufficiently low that the bank would not deviate and report all signals ($T = 0$). If fines are set sufficiently low and high, then the first best is implemented, because the government can commit to First Best equilibrium investigations. Lemma (3) summarizes the results.

Lemma 3 *The First Best equilibrium can be implemented with fines on the $[F^*, F^{**}]$ interval, only if $F^* \leq F^{**}$, where*

$$F^* \equiv 2k\left(\alpha + \delta - 2\alpha\delta\right) \frac{(1 - \alpha - \delta + 2\alpha\delta)\, c + m}{\alpha^2 \delta(1 - \delta)\rho h}$$

$$F^{**} \equiv \frac{2kc\left(\alpha + \delta - 2\alpha\delta\right)^2}{\left[\alpha(1 - \delta)\right]^2 \rho h}$$

Proof. The first best welfare level is implemented by equilibrium actions ($M = 1, T = 1$, $I_0 = I_0^*, I_1 = I_1^*$). These actions are implementable in the second best, if they are incentive compatible and maximize the bank's profit. Formally:

$$IC_1 \qquad \Pi(M = 0, T \in \{0, 1\}, I_0 = I_0^*, I_1 = I_1^*) \leq \Pi(M = 1, T = 1, I_0 = I_0^*, I_1 = I_1^*)$$

$$IC_2 \qquad \Pi(M = 1, T = 0, I_0 = I_0^*, I_1 = I_1^*) \leq \Pi(M = 1, T = 1, I_0 = I_0^*, I_1 = I_1^*)$$

Starting with IC_1:

$$-\alpha I_0^* F \leq -(1 - p_1) q_{01} I_0^* F - p_1 c - m$$

$$F^* \equiv \underbrace{\frac{p_1 c + m}{(\alpha - (1 - p_1) q_{01}) I_0^*}}_{\alpha \delta} = 2k (\alpha + \delta - 2\alpha\delta) \frac{(1 - \alpha - \delta + 2\alpha\delta) c + m}{\alpha^2 \delta (1 - \delta) \rho h} \leq F$$

Then IC_2:

$$-c - m \leq -(1 - p_1) q_{01} I_0^* F - p_1 c - m$$

$$F \leq \frac{c}{q_{01} I_0^*} = \frac{2kc (\alpha + \delta - 2\alpha\delta)^2}{[\alpha(1 - \delta)]^2 \rho h} \equiv F^{**}$$

Finally, the First Best equilibrium is surely implemented with fines such that: $F^* \leq F \leq F^{**}$, because the government can commit to $(I_0 = I_0^*, I_1 = I_1^*)$ investigation levels. ∎

Fines can be both weak and strong if reporting is sufficiently costly (c) compared to monitoring costs (m). Intuitively, if reporting is overly cheap, then fines strong enough to make the bank monitor (and incur cost m) will also force them to report all the signals (at cost c). Thus, fines can take intermediate values if the monitoring costs are not too large compared to filing costs. The reasoning is formalized in Lemma (4).

Lemma 4 *The First Best equilibrium is implementable, if and only if reporting is sufficiently costly compared to bank monitoring*

$$m \leq \frac{(1 - \alpha)(2\delta - 1)}{(1 - \delta)} c \tag{8}$$

Proof. Follows from Lemma (3) and noting that condition (8) is equivalent with $F^* \leq F^{**}$. ∎

Lemma (4) is a simple result, but it has profound implications. It shows that the reporting cost penalizes for false negatives, and thus discourages the bank from crying wolf. Consequently, it provides one of the model's main policy insights, namely that reporting might be crucial in implementing the First Best equilibrium. Reporting fees could raise the right hand side of condition (8) until the first best is implementable.

In order to ease the following discussion, note that all possible equilibria with zero bank monitoring $(M = 0)$ share a number of properties and, most importantly, yield the same social welfare as summarized in Corollary (2).

Corollary 2 *All equilibria with no bank monitoring $(M = 0)$ yield the same social welfare (W^{**}), and government investigation conditional on no reporting (I_0) is:*

$$I_0 = I^{**} \equiv \frac{\alpha \rho h}{2k}$$

Bank reporting threshold and government investigation conditional on reporting are undetermined, $T \in \{0, 1\}$ and $I_1 \in [0, 1]$). Furthermore the resulting welfare (W^{**}) is lower than the First Best welfare: $W^{**} < W^*$.

Proof. Follows from the proof of Proposition (1). ∎

The equilibria characterized in Corollary (2) are called Second Best equilibria. The reason is that, as it will be proven in Proposition (2), one of these equilibria will prevail if the First Best is not implementable.

Definition 2 (Second Best equilibria) *Equilibria with no bank monitoring* $(M = 0)$ *and implemented by fine* F *are called Second Best equilibria with fine* F.

Next, it is shown that Second Best equilibria indeed prevail if the First Best is not implementable. The intuitive reason is that, if the First Best is not implementable, then bank monitoring $(M = 1)$ implies that the bank reports all signals $(T = 0)$. However, if the SAR is filed irrespective of the signal, then it does not convey any information. Thus, the government rather implements no bank monitoring $(M = 0)$, which is exactly the second best. The second best is always implementable with zero fines $(F = 0)$. Proposition (2) summarizes the results.

Proposition 2 *The First Best equilibrium is implementable (and it is implemented) with* $F = F^*$, *if condition (8) holds. Otherwise, Second Best equilibria are implemented with zero fines* $(F = 0)$.

4.4 Second Best with Exogenous Fines

In this subsection the second best incentive compatible game is analyzed under exogenous fines. The question is asked: "How does fine setting affect the efficiency of the money laundering enforcement regime?" For the sake of tractability, this subsection focuses on the parameter setup where the first best is implementable, i.e. inequality (8) holds. Equilibria for the parameter setup when the first best is not implementable are detailed in the appendix.

The equilibria under exogenous fine setting follow mostly from the characterization of the Second Best equilibria. First, as Lemma (3) shows, fines on the $[F^*, F^{**}]$ interval implement the First Best equilibrium. Second, the Second Best equilibria are implemented with low fines, when the bank prefers paying fines to exerting monitoring and reporting effort. Sufficiently low fines are set formally in Corollary (3).

Corollary 3 *Second best equilibria prevail if fines are sufficiently low:* $F \leq F'$ *where*

$$F' \equiv \min \left\{ 2k \frac{(1 - \alpha - \delta + 2\alpha\delta)\,c + m}{\alpha\delta\rho h}, \frac{2k\,(\alpha + \delta - 2\alpha\delta)\,(c + m)}{\alpha^2\rho h} \right\} \tag{9}$$

Third, if neither the First Best nor the Second Best equilibria can be implemented, then only one more pure strategy bank action pair remains: when the bank monitors $(M = 1)$ and reports all transactions $(T = 0)$. As it turns out, all equilibria in which the bank monitors and reports are characterized by the same social welfare and similar government investigation levels. Corollary (4) summarizes the result.

Corollary 4 *All equilibria in which the bank monitors $(M = 1)$ and reports all transactions $(T = 0)$ yield identical welfare (W^{***}). Moreover, in all such equilibria the government investigates, given reporting, with effort $I_1 = I^{**}$. Furthermore, the resulting welfare (W^{***}) is lower than that of Second Best equilibria: $W^{***} < W^{**} < W^*$*

Proof. The proof follows from the proof of Proposition (1). ∎

The similarities among the equilibria entailing bank action $(M = 1, T = 0)$ allow these equilibria to be called Third Best according to their welfare rank.

Definition 3 (Third Best equilibria) *The equilibria in which the bank monitors $(M = 1)$ and reports all transactions $(T = 0)$, and which are implemented with fines F and with no reporting investigation I_0 are called Third Best equilibria with F and I_0.*

Third best equilibria are implemented by high fines. Intuitively, strong enough fines cause the bank not only to monitor, but also to report all signals. Lemma (5) formalizes the argument.

Lemma 5 *Sufficiently large fines $F \geq F^{**}$, implement the Third Best equilibria, if the First Best equilibrium is implementable. Then government investigations given no reporting are set larger than $I_0'(F)$, $I_0 \geq I_0'(F)$ where*

$$I_0'(F) \equiv \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha\delta F}, \frac{c + m}{\alpha F} \right\}$$

Furthermore, $I_0'(F) < 1$ for $\forall F \geq F^{**}$.

The above results allow for characterizing the Second Best equilibria with exogenous fine setting. Proposition (3) summarizes the results of the subsection and characterizes the equilibria with exogenous fines.

Proposition 3 *Fines implement the following equilibria if the First Best equilibrium is implementable*

$$
\begin{array}{ll}
F \leq F' & \text{Second Best equilibria} \\
F^* \leq F \leq F^{**} & \text{First Best equilibrium} \\
F^{**} < F & \text{Third Best equilibria with } I_0 \geq I_0'(F)
\end{array}
$$

There is no pure strategy equilibrium with fines $F' < F < F^*$.

Proof. The proof follows directly from Lemma (3), Corollary (3) and Lemma (5). ∎

Finally, there is no pure strategy equilibrium with fines in the (F', F^*) region. Fines which are too high for the second best are too low for the first best. The reason is that fines represent incentives only with government investigation conditional on no reporting. In Second Best equilibria investigation conditional on no reporting is higher than in the First Best equilibrium $(I_0^* > I^{**})$. Thus, fines which are too strong with high investigation (I_0^*), turn out to be too weak with low investigation (I^{**}).

4.4.1 Crying Wolf

Proposition (3) clearly shows that fines first increase, but later on decrease welfare. Moreover, very high fines lead to lower social welfare (W^{***}) than no fines at all (W^{**}).

This subsection shows that the effects are graver than a simple reduction in welfare. Even the prosecution rate, i.e. the probability that money laundering is prosecuted, decreases with excessive fines. Prosecution rates are the highest in the First Best equilibrium. However, the rate is the same in the Second and the Third Best equilibria. Intuitively, what matters for the success of government investigation is information supplied by the bank. However, reporting no transaction or reporting all transactions is equally uninformative in identifying the most likely suspects. Lemma (6) formalizes the result.

Lemma 6 *Fines first increase and later decrease the likelihood that money laundering is prosecuted (χ) if the First Best equilibrium is implementable. The expected prosecution rate in the three possible equilibria is set as*

$F \leq F'$	*Second Best equilibria*	$\chi^{**} = \alpha I^{**} > \chi^{**}$
$F^* \leq F \leq F^{**}$	*First Best equilibrium*	$\chi^* = \alpha(1-\delta)I_0^* + \alpha\delta I_1^*$
$F^{**} < F$	*Third Best equilibria with $I_0 \geq I_0'(F)$*	χ^{**}

Lemma (6) highlights the crying wolf problem. In the case of coarse communication, some scarcity of reporting is desirable. Providing more reports can be detrimental because the additional reports dilute the value of existing ones. As shown in the model, if the bank reports all transactions, then the information value of the reports is eliminated. This dilution of information is called crying wolf and is defined formally below.

Definition 4 (Crying Wolf) *Crying wolf arises when excessive reporting dilutes the information value of reports. In the extreme case of crying wolf, reports become completely uninformative.*

Formally, the extreme version of crying wolf arises in the Third Best equilibria when the bank monitors, and also reports all transactions, as reports become completely uninformative.

4.4.2 Information Laffer Curve

The result of Lemma (6) can also be interpreted as the Information Laffer curve. The prosecutions (output of the money laundering enforcement regime) can be drawn as a function of fines. Lemma (6) shows that the number of expected prosecutions first rises with fines, but eventually it falls back to the original low level. This is called the Information Laffer curve. The qualitative results are illustrated in Figure (2). In order to illustrate the workings of the model, the expected reporting figures are also charted as a function of fines.
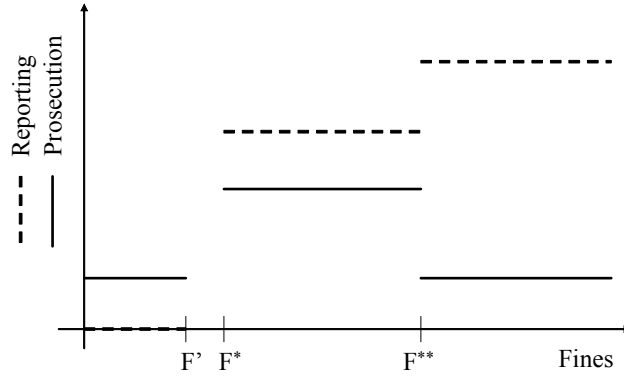
Figure 2: Information Laffer curve

The Information Laffer curve shows an unusual property of the model. In most law enforcement problems increasing incentives increases prosecution. Thus, if fighting a particular type of crime is important, uncertain decision makers might want to err on the safe side and introduce strong incentives. In the money laundering enforcement problem this logic produces counterproductive results. In fact, as Second Best equilibria welfare dominate Third Best equilibria, uncertain decision makers should actually prefer lower fines.

5 Comparative Statics

5.1 First Best Investigation: I_0^* and I_1^*

As expected, investigation efforts (I_0^* and I_1^*) increase both in the harm caused by money laundering (h) and in the recovery rate (ρ). Intuitively, the marginal benefit from the prosecution of money laundering is higher, and so the government prefers to undertake more investigation. Similarly, investigation efforts (again both I_0^* and I_1^*) decrease in the cost of government investigation (k). Clearly, if the marginal cost of investigation is higher, then less investigation is undertaken.

More interestingly, investigation efforts also depend on the prior likelihood of money laundering (α) and the precision of the signal (δ). The more likely money laundering is, the higher both investigation efforts are. The intuition is that higher likelihood of money laundering, ceteris paribus, increases the marginal benefit of investigation as more laundering is prosecuted with the same effort expenditure. Formally:

$$
\begin{aligned}
\frac{\partial I_0^*}{\partial \alpha} &= \frac{\delta(1-\delta)\rho h}{2k\left(\alpha + \delta - 2\alpha\delta\right)^2} > 0 \\
\frac{\partial I_1^*}{\partial \alpha} &= \frac{\rho h}{2k\left(1 - \alpha - \delta + 2\alpha\delta\right)^2} > 0
\end{aligned}
\tag{10}
$$

19

However, the precision of the signal affects the two investigation efforts disparately:

$$\begin{aligned}
\frac{\partial I_0^*}{\partial \delta} &= \frac{-\alpha \left(1 - \alpha\right) \rho h}{2k \left(\alpha + \delta - 2\alpha\delta\right)^2} < 0 \\
\frac{\partial I_1^*}{\partial \delta} &= \frac{\alpha \left(1 - \alpha\right) \rho h}{2k \left(1 - \alpha - \delta + 2\alpha\delta\right)^2} > 0
\end{aligned} \tag{11}$$

The intuition is that the more precise the signal is, the more reliable first best reporting is. Thus, the government investigates more vigorously given reporting, and less so in the absence of it.

5.2   Minimal Optimal Fine: F^*

Two properties of the minimal optimal fine are crucial to understand the comparative statics. First, the minimal optimal fine provides incentives for the bank to stay informed and not deviate from monitoring ($M = 1$), as can be seen from the IC_1 constraint in the proof of Lemma (3). Second, the minimal optimal fine depends on the first best investigation effort conditional on no reporting (I_0^*). The reason is, that fines represent incentives only together with I_0^*, as the bank is fined only if money laundering is prosecuted. Equation (12) depicts fines as a function of first best investigation (I_0^*), bank costs (c, m), prior probability of money laundering (α), and signal precision (δ):

$$F^* \equiv \frac{\left(1 - \alpha - \delta + 2\alpha\delta\right)c + m}{\alpha\delta I_0^*} \tag{12}$$

Evidently from (12), the minimal optimal fine decreases in the first best investigation level (I_0^*), which has three unexpected consequences.

First, the minimal optimal fine decreases in the harm caused by money laundering (h) and in the recovery rate (ρ). Equation (4) shows that higher harm and higher recovery rates increase investigation. Thus, increased investigation provides additional incentives for the bank to stay informed, and lower fines suffice.

Second, the minimal optimal fine also decreases in the prior probability of money laundering (α). Two factors simultaneously decrease the optimal fine. First, increasing α increases the likelihood of uncovering money laundering should the bank choose to be uninformed. This means that smaller fines suffice, which is demonstrated in the following equation:

$$\frac{\partial}{\partial \alpha} \frac{\left(1 - \alpha - \delta + 2\alpha\delta\right)c + m}{\alpha\delta} = -\frac{\left(1 - \delta\right)c + m}{\alpha^2\delta} < 0$$

Second, increasing the prior probability of money laundering increases the investigation effort, which is apparent in equation (10). Increasing investigation, as before, further reduces the fine level. Thus, both effects cause the minimal optimal fine to decrease in the prior likelihood of money laundering.

Third, the minimal optimal fine increases in investigation costs k. Higher investigation costs decrease first best investigation by equation (4). Thus, higher fines are needed to preserve the bank's incentives to stay informed.

The minimal optimal fine also decreases in the precision of the signal (δ). However, the intuition of this result is not immediately obvious. On the one hand, as shown in equation (11), higher precision of the signal implies lower investigation effort. This effect requires increased fines to preserve incentives. On the other hand, higher precision also increases the value of being informed for the bank. The informed bank is more certain about its information, and the likelihood of fines, given the low signal, is lower. Formally, this can be demonstrated as:

$$\frac{\partial}{\partial \delta} \frac{(1 - \alpha - \delta + 2\alpha\delta)\,c + m}{\alpha\delta} = -\frac{(1 - \alpha)c + m}{\alpha\delta^2} < 0$$

In order to evaluate the relative strength of the contradicting forces, F^* is formally differentiated:

$$\frac{\partial F^*}{\partial \delta} = \frac{-1}{(\alpha\delta)^2 I_0^*} \left[\frac{\alpha\,(1 - \alpha)\,(2\delta - 1)}{(1 - \delta)(\alpha + \delta - 2\alpha\delta)}c + \frac{\alpha(2\delta - 1) + \delta^2(1 - 2\alpha)}{(1 - \delta)(\alpha + \delta - 2\alpha\delta)}m \right] < 0$$

The result shows that the second effect is dominant, and the optimal fine decreases in signal precision.

The remaining comparative statics are straightforward. The optimal minimal fine increases in the bank's costs of monitoring (m) and reporting cost (c). This is intuitive, as the greater the effort costs are, the higher incentives must be.

5.3   Maximal Optimal Fine: F^{**}

Two properties of the maximal optimal fine are crucial in understanding the comparative statics. First, the maximal optimal fine shows how large the fine can be before the bank deviates to reporting every transaction ($T = 0$), as can be seen from the IC_2 constraint in the proof of Lemma (3). Second, the maximal optimal fine depends on the first best investigation effort given no reporting (I_0^*), because fines represents incentives only with investigation. Thus, fine F^{**} can be written formally as follows:

$$F^{**} = \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha(1 - \delta)I_0^*} \tag{13}$$

Equation (13) allows the analysis of F^{**} similarly to that of F^*. The results are very similar, and show the same three unexpected consequences.

First, maximal optimal fine F^{**} decreases in the harm caused by money laundering (h) and in the recovery rate (ρ). The reason, in both cases, is that equilibrium investigation I_0^* increases in h and ρ as equation (4) shows. Thus, the bank needs smaller nominal fines so as not to cry wolf, because the probability of prosecuting money laundering is higher.

Second, the maximal optimal fine also decreases in the prior probability of money laundering (α). The reason is twofold. First, higher prior likelihood of money laundering increases

21

investigation I_0^* as equation (10) shows. Second, higher prior money laundering probability implies that even with the low signal the transaction is more likely to be money laundering. Thus, ceteris paribus, it is more likely that the bank will be fined if it does not report. So, the bank needs smaller fines so as not to report all transactions. This effect is demonstrated by:

$$\frac{\partial}{\partial \alpha} \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha(1-\delta)} = \frac{-\delta c}{\alpha^2(1-\delta)} < 0$$

Thus, both effects decrease maximal optimal fine F^{**} in the prior likelihood of money laundering.

Third, maximal optimal fine increases in investigation costs k. Higher investigation costs affect maximal optimal fines through decreasing investigation I_0^*. Thus, higher fines are needed to keep incentives constant.

The maximal optimal fine also increases in the precision of the signal (δ). First, investigation I_0^* decreases in δ by equation (11). Intuitively, the more precise the signal, the less investigation is needed if no report was filed in the First Best equilibrium. Second, higher precision decreases the likelihood that the low signal activity is indicative of money laundering. Thus, the bank has lower incentives to report low signal transactions. This latter effect is demonstrated formally by:

$$\frac{\partial}{\partial \delta} \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha(1-\delta)} = \frac{(1-\alpha)c}{\alpha(1-\delta)^2} > 0$$

Thus, both effects increase maximal optimal fine F^{**} in signal precision.

The remaining results are trivial. The maximal optimal fine increases in the cost of reporting (c), as more costly reporting provides disincentives against crying wolf. The maximal optimal fine is unaffected by changes in the monitoring cost m, because the bank keeps monitoring even if it deviates from the first best to cry wolf.

Finally, the comparative statics are summarized concisely on Table (1) below.

	I_0^*	I_1^*	F^*	F^{**}
h	+	+	−	−
ρ	+	+	−	−
α	+	+	−	−
δ	−	+	−	+
k	−	−	+	+
c	0	0	+	+
m	0	0	+	0

Table 1: Comparative Statics

6   Empirical Evidence

The model provides two predictions which can be investigated empirically:[8]

Prediction 1 (Fines) *Excessive fine increases can trigger crying wolf, that is an increasing number of suspicious activity reports (SARs) results in fewer prosecutions.*

Prediction 2 (Harm) *Even with constant fines, the increasing harm caused by money laundering can increase government investigation efforts enough to trigger crying wolf. Thus, increased investigation efforts result in lower or stagnant number of prosecutions.*

The predictions are evaluated in three steps. First, it is shown that the available evidence is consistent with the model's underlying assumption that money laundering volumes are unchanged and deterrence effects are secondary. Second, data on fines, reports, and prosecutions are analyzed. In particular, it is shown that the available evidence is consistent with the model's predictions. Finally, further supporting testimonial evidence is discussed.

The volume of money laundering in the United States is estimated to be constant during the period of interest. Microeconomic estimates show money laundering volumes around 8% of the US GDP according to Reuter and Truman (2004). Though criminal earnings (without tax evasion) declined from 2.8% of the GDP in 1995 to 2.3% in 2000, increases in estimated tax evasion (from 5.2% to 5.6%) compensated the decline. Besides these microeconomic estimates, macroeconomic estimates also show negligible variance and confirm that the underground economy is roughly 8.8% of the US GDP (Schneider, 2002). Of course, money laundering is a clandestine activity and it cannot be estimated precisely. Nevertheless, the available evidence is consistent with the assumption that the volume of money laundering is essentially unchanged.

Fines show a steep increase during the period under consideration. Though fines are very difficult to measure precisely because data on sanctions are scattered among regulatory agencies, all possible measures increase. All fines and restitutions, including fines on non-depository institutions as well as individuals, increased six-fold between 1996 and 2001 on the basis of US Sentencing Commission information (Reuter and Truman, 2004). GAO (2004) also confirms steep increases in civil money penalties, though without providing exact figures. FinCEN provides detailed statistics of its fines for neglecting SAR filing duties (at http://www.fincen.gov/). FinCEN started to levy fines in 2002, and the volume grew quickly from $0.1 million in 2002 to $24.5 million in 2003 and finally to $35 million in 2004.

Reporting has also increased steeply. Reporting data are published in FinCEN (2005d), and the data are depicted on Figure (3). In order to correct for the widening range of reporting institutions, SARs filed by depository institutions are primarily discussed. As it is evident from Figure (3), depository SARs grew exponentially, and the growth does not seem to saturate. In fact, according to The Wall Street Journal (2005), the growth has further accelerated in the first half of 2005.

---

[8]In order to link national level data to the single bank reporting model, the available data is interpreted as the aggregate outcome of many reporting problems.
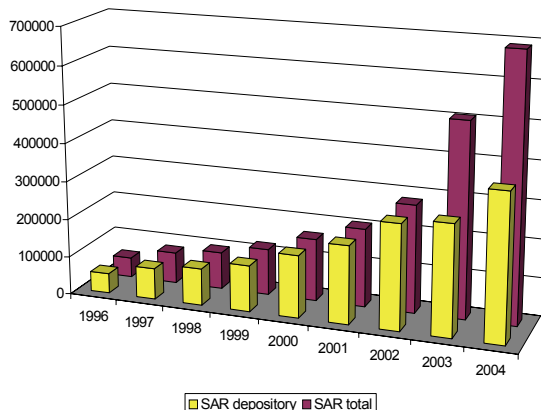
Figure 3: Reporting trends

This strong growth is, however, missing from prosecution figures. Money laundering prosecution peaked in 1999 and declined thereafter.[9] Two measures of money laundering prosecutions are depicted in Figure (4). First: the number of cases *filed* in each year with money laundering as the most serious terminating offense. This measures the efficiency of the law enforcement agencies in bringing cases which can be prosecuted. Second: the number of cases *terminated* in each year with money laundering as the most serious terminating offense. This measures the efficiency of law enforcement in bringing cases in which the defendants can also be convicted. In theory, the two different measures could allow for different comparisons. However, in practice, they show the same qualitative picture as evidenced in Figure (4). According to both measures prosecutions rose until 1999 and declined thereafter.

The available evidence is consistent with Prediction (1): increasing fines first increase prosecutions and reporting. However, further fine increases result in crying wolf and increasing reports result in lower number of prosecutions.

Prediction (2) can be evaluated using data pre- and post-9/11, as the harm of money laundering was clearly updated upwards after 9/11. As the model predicts, government investigation efforts increased, as documented in Reuter and Truman (2004). Furthermore, fines continued their increase. All this resulted in accelerated reporting. Between 1997 and 2000, the number of SAR filings by depository institutions grew by 20,381 per year on average. Between

---

[9]Federal prosecution data for money laundering are available from two sources. First, the Money Laundering Special Report (2003) published data for 1994-2001 on money laundering prosecution. Second, prosecution data are obtained from the Federal Justice Statistics Resource Center website query system until 2003 (at http://fjsrc.urban.org/index.cfm). The two databases show the same qualitative picture. According to Mark Motivans, author of the Money Laundering Special Report (2003), the web query system reports the title and section of the most serious terminating offense, while the report uses the title and section of the most serious filing offense. In the analysis the longer query data are discussed, because they better fit the scope of the analysis.
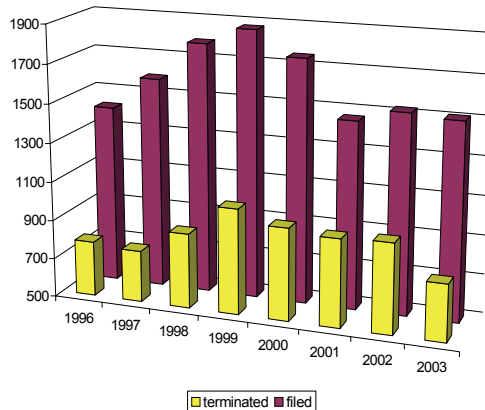
Figure 4: Money laundering prosecution

2001 and 2004, depository SAR filings grew on average by 44,531 per year.[10] This increased reporting seems to have diminished the effects of increased government investigations. The number of prosecutions stayed essentially unchanged which supports the existence of crying wolf in reporting.

Finally, FinCEN testimonies also indicate the existence of crying wolf. FinCEN (2005b) describes 'defensive filing' as strikingly similar to the effects of crying wolf. First, the number of SARs filed skyrockets. Second, banks file SARs regardless of the level of suspicion. Third, law enforcement efforts are compromised by the large number of filings. A FinCEN (2005a, p3) quote illustrates the point:

> We estimate that if current filing trends continue, the total number of Suspicious Activity Reports filed this year will far surpass those filed in the previous years. [...] it fuels our concern that financial institutions are becoming increasingly convinced that the key to avoiding regulatory and criminal scrutiny under the Banking Secrecy Act is to file more reports, regardless of whether the conduct or transaction identified is suspicious. [...] If this trend continues, consumers of the data - law enforcement, regulatory agencies, and intelligence agencies - will suffer.

7 Implementable Policy Implications

The model is able to identify implementable policy solutions to the crying wolf problem. However, as the empirical evidence is scattered, the policy implications are suggestions for future

---

[10] The raw figures might even understate the 9/11 effect as the 1997-2000 data contain the initial SAR growth due to setting up the reporting systems. Note also that total SARs grew much faster: the jump is from 20,486 report-increase/year to 121,125. However, the increase is due, in great part, to changes in the regulation requiring other institutions, like casinos and money service businesses, to report.

policy debates rather than strict prescriptions.

The model identifies that the crying wolf problem might be remedied by reducing fines. If the first best is implementable, then as Proposition (3) shows, decreasing fines will implement it. There are some signs that fines are indeed being reduced. Fines were decreased, for instance, by centralizing the prosecution of banks for Banking Secrecy Act violations by modifying the US Attorney General's manual (2005). According to industry opinions, the changes were motivated by the case of AmSouth Bank which was criminally prosecuted by a Mississippi state attorney. The prosecution disturbed banks because they felt threatened to be fined by too many organizations, as Federal Reserve Banks, Office of the Comptroller of the Currency (OCC), FinCEN and state attorneys. Furthermore, the OCC (2005) also signaled decreasing fines by intentionally leaking an internal memorandum stressing that there is no zero tolerance policy in effect.

Another policy suggestion advocated here calls for a 'safe harbor' or fine-free zone for banks. Banking supervision could guarantee that well-reporting banks will not be fined for Banking Secrecy Act violations for some period. This safe harbor would give peace of mind to banks, and reduce the pressure to cry wolf. Furthermore, such safe harbor is specifically useful to preserve the power of banking supervision in setting fines.

The model also highlights that the first best can always be implemented if the bank pays an additional reporting fee. As Lemma (4) has shown, the first best is always implementable if reporting is sufficiently costly. Furthermore, an extension of the model with continuous signals and monitoring efforts, discussed in the appendix, demonstrates that reporting fees are generally necessary to implement the first best. Moreover, changing the cost of reporting is easily implementable, as regulatory authorities can charge a reporting fee. The economic argument for reporting fees can also be understood through the externalities caused by reporting. Each report dilutes the value of all the other reports. However, banks do not realize these reporting externalities, i.e. the shadow costs of information dilution, and reporting fees are pricing these dilution costs.

Reporting fees might be seen as putting undue burden on the banking industry. This is, however, not necessarily so. Reporting fees are not needed for their redistributive effect (which might be the major concern of the industry), but rather for their incentive effect. For instance, fees might be channeled back to the industry, eliminating the redistributive effect and still maintaining the incentives. This could eliminate placing undue burden on the industry, though redistribution between individual banks might not be eliminated.

The comparative statics also provide lessons for setting policy in response to changes in the economy. Three examples are discussed here. First, the increased harm caused by money laundering should lead to decreased fines. Second, the increased likelihood of money laundering should also lead to decreased fines. In a sense, the graver the problem, the weaker the incentives should be that banks are facing. Third, technological changes are likely to decrease reporting costs which can lead to crying wolf. In order to avoid this, policy makers might want to

introduce or further increase reporting fees.

8 Conclusion

The paper identifies the crying wolf problem in money laundering enforcement. This is important for three main reasons. First, the paper is the first to formally analyze money laundering enforcement. The research helps to understand the specificities of how this harmful financial crime is fought. Furthermore, it also provides a solid ground for further studies analyzing economic issues related to money laundering and its enforcement.

Second, the paper provides implementable policy implications, many of which seem counterintuitive at first sight. The paper calls for decreased fines against banks found to be negligent in carrying out their anti-money laundering duties. Furthermore, the model shows that fines against banks should decrease as the harm caused by money laundering increases. The paper also makes the case for introducing reporting fees, i.e. charging banks for informing law enforcement agencies. The paper also highlights how well-intentioned and seemingly sensible money laundering regulations, such as the Patriot Act, could have backfired.

Third, the paper explores the general economic problem of crying wolf. The model, for instance, can be used to address product development questions. Firms, fearing litigation, might provide excessive information on product hazards thereby diluting the information value of their warnings. Customers, in order to save on costly precautionary actions, might disregard the information provided. For example, customers commonly sign or click that they have read and understood the terms of service without even looking at them.

The model can also be used to further think about the possible adverse consequences of the Sarbanes-Oxley Act. Auditors reporting material transactions face a similar problem as banks disclosing suspicious activities. The act has increased auditing fines on auditors much like the Patriot Act has increased fines on banks. The question whether fines are already excessively high or they are in the optimal range, deserves further investigation. Interestingly, some preliminary findings hint that the crying wolf problem is potentially relevant. For instance, Gordon (2005) reports that the average length of Fortune 500 firms' 10-K reports has grown from 16 pages in 1950 to 126 pages in 2000, even before the Sarbanes-Oxley Act.

Finally, the problem of crying wolf sheds some light on the fundamental question of what qualifies as information. The paper focuses on identification, and shows formally how excessive reporting fails to identify and thus inform. The role of identification highlights that in certain situations filtering, in other words withholding specific data, is necessary to efficiently inform. Such filtering of data is becoming increasingly important, as data are increasingly easier to store and forward. This suggests that crying wolf problems might surface in many other applications in the foreseeable future.

9   Appendix

9.1   Exogenous Fine Equilibria with $F^* > F^{**}$

If the first best is not implementable, then the government prefers to implement one of the Second Best equilibria. Corollary (3) shows that this is possible, if the fine is sufficiently low. (Note that Corollary (3) does not depend on whether the first best is implementable or not.) If the second best is not implementable, because the fine is too high, then Third Best equilibria are implemented. Lemma (7) formalizes the argument.

Lemma 7 *Sufficiently high fines which satisfy both: $F \geq F''$ and $F > F'$, implement the Third Best equilibria if the First Best equilibrium is not implementable. Then government investigation is set $I_0 \geq I_0'(F)$. Furthermore F' is as defined in Lemma (5) and*

$$F'' \equiv \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha\delta}, \frac{c+m}{\alpha} \right\}$$

and $I_0'(F) < 1$ for $\forall F \geq F''$.

Proof. As First Best equilibrium is not implementable, the government prefers to implement the second best, which is implementable by Corollary (3) with $F \leq F'$. Otherwise, the government can only implement Third Best equilibria. By IC_1 and IC_2 formalized in the proof of Lemma (5), Third Best equilibria are implementable if

$$F(I_0) \geq \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha\delta I_0}, \frac{c+m}{\alpha I_0} \right\}$$

Thus, the Third Best is implementable with fines larger than F'' where

$$F'' \equiv \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)\,c}{\alpha\delta}, \frac{c+m}{\alpha} \right\}$$

However, depending on the parameter values, this F'' can be both higher and lower than F', as straightforward examples can show. If $F'' > F'$, then there is no pure strategy equilibrium with $F' < F < F''$, and Third Best equilibria prevail with $F \geq F''$. If $F'' \leq F'$, then the Third Best is implemented with $F > F'$. ∎

Consequently, the resulting equilibria can be characterized.

Proposition 4 *Low fines, $F \leq F'$ implement Second Best equilibria and high fines, $F'' < F$ implement the Third Best equilibria with $I_0 \geq I_0'(F)$, if the First Best is not implementable.*

Proof. Follows from Corollary (3) and Lemma (7). ∎

The solution shows that increasing fines decrease social welfare. In this case, however, prosecution rates do not change, because welfare decrease only through wasteful monitoring and reporting. Corollary (5) formalizes the result.

Corollary 5 *The expected prosecution rate is constant in fines: χ^{**}, if the First Best equilibrium is not implementable.*

Proof. Follows directly from Lemma (6) and Proposition (4) of the appendix. ∎

Thus, even if the first best can not take place, crying wolf arises as defined earlier. Excessively high fines still remain detrimental because they increase the social costs of operating the money laundering enforcement regime without improving the prosecution of money laundering. Nonetheless, in this case crying wolf does not decrease expected prosecution.

9.2 Proofs

Proof of Lemma (2). Suppose that $(R_0 = 1, R_1 = 0)$. Let us assume further, without loss of generality, that I'_0 and F' are the government's best responses.

First, consider that the bank monitors, $M = 1$ in equilibrium. Then, the expected value of fines conditional on non-reporting is $I'_0 F' \beta_0$ with the low, and $I'_0 F' \beta_1$ with the high signal. In the second best, the bank reports only if the expected fines conditional on non-reporting are at least weakly higher than the reporting cost c. Thus, reporting implies that the expected fine conditional on non-reporting is higher than the reporting cost:

$$R_0 = 1 \implies I'_0 F' \beta_0 \geq c$$
$$R_1 = 0 \implies I'_0 F' \beta_1 \leq c$$

which is impossible if both $I'_0, F' > 0$ as $\beta_0 < \beta_1$. Furthermore, if the bank monitors $(M = 1)$, then the government does not set $I'_0 = 0$ or $F' = 0$ in any second best equilibria. The reason is that if either $I'_0 = 0$ or $F' = 0$, then the bank is never fined. This implies, however, that the bank does not monitor (and sets $M = 0$) because monitoring reduces private profits by $m > 0$, but it does not save on fines.

Second, consider equilibrium bank monitoring $M = 0$. In this case the bank is never informed, and it never reports. Thus, equilibrium $(R_0 = 1, R_1 = 0)$ can be replaced with any other reporting actions. The government's best response, social welfare, and observed bank actions do not change. ∎

Proof of Proposition (1). Notice, first, that in the first best fines do not play any role. Thus, as the government prefers to make the bank as well-off as possible and it sets the fine to zero, $F = 0$.

In order to explore the possible equilibria, start with the bank action set. There are four possible bank action pairs:

$$(M = 0, T = 0) \qquad (M = 0, T = 1) \qquad (M = 1, T = 0) \qquad (M = 1, T = 1)$$

First, if $M = 0$, then the bank never reports. Setting T is therefore inconsequential, and it does not affect welfare or best response. Thus, bank action pairs $(M = 0, T \in \{0, 1\})$ can be

treated together. The social welfare function then is written as:

$$W(I_0, I_1, F = 0, M = 0, T \in \{0, 1\}) = \alpha I_0 \rho h - k I_0^2 - \alpha h \tag{14}$$

which yields the following government investigation best responses:

$$I_0 = \frac{\alpha \rho h}{2k} \equiv I^{**}$$
$$I_1 \in [0, 1]$$

Substituting back to the welfare function yields:

$$\begin{aligned} W^{**} &= I^{**} \rho h - k (I^{**})^2 - \alpha h \\ &= \frac{(\alpha \rho h)^2}{4k} - \alpha h \end{aligned}$$

Second, consider the $(M = 1, T = 0)$ action pair. Then the social welfare function can be written as:

$$W = q_{1T} I_1 \rho h - k I_1^2 - c - m - \alpha h$$

which is maximized by setting

$$I_1 = \frac{q_{1T} \rho h}{2k} = \frac{\alpha \rho h}{2k} \equiv I^{**}$$

and is not affected by I_0, as the bank always reports. Thus the social welfare:

$$\begin{aligned} W^{***} &= I^{**} \rho h - k (I^{**})^2 - c - m - \alpha h \\ &= W^{**} - c - m \end{aligned}$$

Thus, the $(M = 1, T = 0)$ action pair cannot be part of the First Best equilibrium.

Under bank action pair $(M = 1, T = 1)$, the social welfare function can be written as:

$$\begin{aligned} W(I_0, I_1, F = 0, M = 1, T = 1) = {}&(\alpha + \delta - 2\alpha\delta)(\beta_0 I_0 \rho h - k I_0^2) \\ &+ (1 - \alpha - \delta + 2\alpha\delta)(\beta_1 I_1 \rho h - k I_1^2 - c) - \alpha h \end{aligned} \tag{15}$$

which yields the following investigation best responses:

$$\begin{aligned} I_0^* &= \frac{\alpha(1 - \delta)\rho h}{2k(\alpha + \delta - 2\alpha\delta)} \\ I_1^* &= \frac{\alpha\delta\rho h}{2k(1 - \alpha - \delta + 2\alpha\delta)} \end{aligned}$$

Substituting back to the welfare function yields:

$$W^* = \frac{[\alpha(1 - \delta)\rho h]^2}{4k(\alpha + \delta - 2\alpha\delta)} + \frac{(\alpha\delta\rho h)^2}{4k(1 - \alpha - \delta + 2\alpha\delta)} - (\alpha + \delta - 2\alpha\delta)c - m - \alpha h$$

Finally, the equilibrium with positive bank investigation provides higher welfare as

$$W^* - W^{**} = \frac{[\alpha(1-\delta)\rho h]^2}{4k\,(\alpha + \delta - 2\alpha\delta)} + \frac{(\alpha\delta\rho h)^2}{4k\,(1 - \alpha - \delta + 2\alpha\delta)} - \frac{(\alpha\rho h)^2}{4k} - (\alpha + \delta - 2\alpha\delta)\,c - m > 0$$

The expression is positive by assumption set in equation (1).

Finally, note that all possible government best response investigations are on the unit interval. First, all (I^{**}, I_0^*, I_1^*) are non-negative. Second, by $\delta > (1 - \delta)$:

$$I^{**} = \frac{\alpha\rho h}{2k} < \frac{\alpha\delta\rho h}{2k\,(1 - \alpha - \delta + 2\alpha\delta)} = I_1^*$$

Third, again by $\delta > (1 - \delta)$:

$$I_0^* = \frac{\alpha(1-\delta)\rho h}{2k\,(\alpha + \delta - 2\alpha\delta)} < \frac{\alpha\rho h}{2k} = I^{**}$$

Finally, by the assumption stated in equation (2): $I_1^* < 1$. Thus, all three investigation probabilities are on the unit interval. Furthermore, they can be ranked as: $I_0^* < I^{**} < I_1^*$. ∎

Proof of Proposition (2). There are three possible best response pairs by the proof of Proposition (1). In terms of actions and social welfare:

First Best	$(M = 1, T = 1)$	$(I_0 = I_0^*, I_1 = I_1^*)$	W^*
Second Best	$(M = 0, T \in \{0,1\})$	$(I_0 = I^{**}, I_1 \in [0,1])$	W^{**}
best response pair 3	$(M = 1, T = 0)$	$(I_0 \in [0,1], I_1 = I^{**})$	W^{***}

The welfare implied by the three best response pairs is derived in Proposition (1):

$$W^{***} < W^{**} < W^* \tag{16}$$

Thus, the government would like to first implement (if it is possible) the First Best equilibrium. By Lemma (4) the first best is implementable if

$$m \leq \frac{(1-\alpha)(2\delta - 1)}{(1-\delta)}c$$

As the government prefers to increase banking profits, as long as social welfare is not affected, it sets fines as low as possible. The lowest possible fine which still implements the First Best is F^*, as demonstrated by Lemma (3). As the government can commit to investigation action, the First Best equilibrium is uniquely implemented with $F = F^*$ fines.

If the First Best is not implementable, the government prefers to implement one of the Second Best equilibria, all of which require the very same incentive compatibility constraints:

IC_1 $\Pi(M = 1, T = 1, I_0 = I^{**}, I_1 \in [0,1]) \leq \Pi(M = 0, T \in \{0,1\}, I_0 = I^{**}, I_1 \in [0,1])$

IC_2 $\Pi(M = 1, T = 0, I_0 = I^{**}, I_1 \in [0,1]) \leq \Pi(M = 0, T \in \{0,1\}, I_0 = I^{**}, I_1 \in [0,1])$

Starting with IC_1:

$$-(1-p_1)\,q_{01}I^{**}F - p_1c - m \;\leq\; -\alpha I^{**}F$$

$$F \;\leq\; \frac{p_1c+m}{(\alpha-(1-p_1)\,q_{01})I^{**}} = 2k\frac{(1-\alpha-\delta+2\alpha\delta)\,c+m}{\alpha\delta\rho h}$$

Then IC_2:

$$-c-m \;\leq\; -\alpha I^{**}F$$

$$F \;\leq\; \frac{c+m}{\alpha I^{**}} = \frac{2k\,(\alpha+\delta-2\alpha\delta)\,(c+m)}{\alpha^2\rho h}$$

Thus, the Second Best equilibria are implementable with fines such that:

$$F \leq \min\left\{2k\frac{(1-\alpha-\delta+2\alpha\delta)\,c+m}{\alpha\delta\rho h}, \frac{2k\,(\alpha+\delta-2\alpha\delta)\,(c+m)}{\alpha^2\rho h}\right\}$$

The government again uses the smallest fine necessary, which in this case is zero, $F = 0$. Government commitment along with the satisfied IC constraints guarantee that one of the Second Best equilibria will prevail in equilibrium. Finally, the government never implements Best Response Pair 3. It is welfare dominated by Second Best equilibria, which are always implementable. ∎

Proof of Corollary (3). Follows from the proof of Proposition (2) and observing that

$$F' < F^*$$

as

$$2k\frac{(1-\alpha-\delta+2\alpha\delta)\,c+m}{\alpha\delta\rho h} \;<\; 2k\,(\alpha+\delta-2\alpha\delta)\,\frac{(1-\alpha-\delta+2\alpha\delta)\,c+m}{\alpha^2\delta(1-\delta)\rho h}$$

$$\alpha(1-\delta) \;<\; (\alpha+\delta-2\alpha\delta) = \alpha(1-\delta)+\delta(1-\alpha)$$

which implies that with $F \leq F'$ the government cannot implement the First Best equilibrium, thus it prefers to implement Second Best equilibria. ∎

Proof of Lemma (5). The proof is lengthy and divided into four parts. First, by Lemma (3), if fines are higher than F^{**}, then the First Best equilibrium is not implementable.

Second, notice that if inequality (8) holds and fines are higher than F^{**}, then the Second Best equilibria cannot prevail either. The necessary fine level for the Second Best equilibria is F' derived in equation (9) which is lower than F^* by the proof of Corollary (3). Furthermore, inequality (8) implies that $F^* \leq F^{**}$, which means, by induction, that $F' < F^* \leq F^{**}$. Thus, with $F \geq F^{**}$, Second Best equilibria are not implementable.

Third, with fines $F \geq F^{**}$, only the Third Best equilibria can be implemented. By proof of Proposition (2), the following incentive compatibility constraints are necessary:

IC_1 $\qquad \Pi(M=1, T=0, I_0 \in [0,1], I_1 = I^{**}) \leq \Pi(M=1, T=1, I_0 \in [0,1], I_1 = I^{**})$

IC_2 $\qquad \Pi(M=0, T \in \{0,1\}, I_0 \in [0,1], I_1 = I^{**}) \leq \Pi(M=1, T=1, I_0 \in [0,1], I_1 = I^{**})$

Starting with IC_1:

$$- (1 - p_1) q_{01} I_0 F - p_1 c - m \leq -c - m$$

$$\frac{(1 - p_1)c}{(\alpha - (1 - p_1) q_{01})I_0} = \frac{(\alpha + \delta - 2\alpha\delta)c}{\alpha\delta I_0} \leq F$$

Then IC_2:

$$\lambda g - \alpha I_0 F \leq \lambda g - c - m$$

$$\frac{c + m}{\alpha I_0} \leq F$$

Thus, Third Best equilibria are implementable with fines such that:[11]

$$F(I_0) \geq \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)c}{\alpha\delta I_0}, \frac{c + m}{\alpha I_0} \right\}$$

$F(I_0)$ can be smaller than F^{**} with some $I_0 \in (0,1]$. First consider:

$$\frac{(\alpha + \delta - 2\alpha\delta)c}{\alpha\delta I_0} \leq \frac{c}{q_{01}I_0^*} = F^{**}$$

$$\frac{1 - \delta}{\delta}I_0^* \leq I_0$$

and notice that both I_0^* and $\frac{1-\delta}{\delta}$ are smaller than one, by Proposition (1) and the $1/2 < \delta < 1$ assumption respectively. Second:

$$\frac{c + m}{\alpha I_0} \leq \frac{c}{q_{01}I_0^*} = F^{**}$$

$$\frac{q_{01}}{\alpha}\left(I_0^* + \frac{m}{c}\right) \leq I_0$$

because by inequality (8)

$$\frac{1 - \delta}{\alpha + \delta - 2\alpha\delta}\left(I_0^* + \frac{m}{c}\right) \leq \frac{1 - \delta}{\alpha + \delta - 2\alpha\delta}\left(I_0^* + \frac{(1 - \alpha)(2\delta - 1)}{(1 - \delta)}\right) < 1$$

and the inequality follows from $I_0^* < 1$. In sum, if inequality (8) holds, the third best is implemented with no reporting investigation such that

$$I_0 \geq I_0'(F) \equiv \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)c}{\alpha\delta F}, \frac{c + m}{\alpha F} \right\} < 1$$

where $I_0'(F) \leq 1$ with $F \geq F^{**}$.

Finally, it is shown that Third Best equilibria cannot be implemented with fines $F < F^*$. To see this, consider the necessary fines for the third best:

$$F(I_0) \geq \max \left\{ \frac{(\alpha + \delta - 2\alpha\delta)c}{\alpha\delta I_0}, \frac{c + m}{\alpha I_0} \right\} \geq \frac{c + m}{\alpha}$$

---

[11] Notice that the state can freely set I_0 because there is no reporting in equilibrium.

Consequently, it suffices to show that

$$\frac{c+m}{\alpha} > \frac{(1-\alpha-\delta+2\alpha\delta)\,c+m}{\alpha\delta I_0^*} = F^*$$

which is equivalent to:

$$c \geq \frac{(1-\delta I_0^*)}{\delta I_0^* - (1-\alpha-\delta+2\alpha\delta)}m \qquad (17)$$

Rearranging is possible, because

$$\delta I_0^* - (1-\alpha-\delta+2\alpha\delta) > 0 \qquad (18)$$

to see that (18) holds:

$$\delta I_0^* > (1-\alpha-\delta+2\alpha\delta)$$

$$\frac{\alpha\delta(1-\delta)\rho h}{(\alpha+\delta-2\alpha\delta)\,(1-\alpha-\delta+2\alpha\delta)} > k$$

This latter statement follows from the assumption stated in equation (2):

$$k\frac{(1-\delta)}{(\alpha+\delta-2\alpha\delta)} > \frac{\alpha\delta(1-\delta)\rho h}{(\alpha+\delta-2\alpha\delta)\,(1-\alpha-\delta+2\alpha\delta)}$$

and observing that

$$\frac{(1-\delta)}{(\alpha+\delta-2\alpha\delta)}k > k$$

which trivially follows from $1-\alpha > \alpha$.

Now, turning back to (17) - note that by equation (8):

$$c \geq \frac{(1-\delta)}{(1-\alpha)(2\delta-1)}m$$

So, it suffices to show that (8) implies (17). To show this, first notice that the right hand side of (17) is decreasing in I_0^* :

$$\frac{\partial}{\partial I_0^*}\frac{(1-\delta I_0^*)}{\delta I_0^* - (1-\alpha-\delta+2\alpha\delta)} = \frac{-\delta^2(\alpha+\delta-2\alpha\delta)}{[\delta I_0^* - (1-\alpha-\delta+2\alpha\delta)]^2} < 0$$

Consequently, it is enough to show that (8) implies (17) for $I_0^* = 1$, because it is implied then for all the other I_0^*. To see this, consider:

$$\frac{(1-\delta)}{\delta - (1-\alpha-\delta+2\alpha\delta)} = \frac{(1-\delta)}{(1-\alpha)(2\delta-1)} < 1$$

Thus, with $I_0^* = 1$, (8) and (17) are equivalent, and with any other I_0^*, (8) implies (17), which finishes the proof as $I_0^* < 1$ by the proof of Proposition (1). ∎

Proof of Lemma (6).   In the Second Best and Third Best equilibria, the expected prosecution rate is trivially:

$$\chi^{**} = \alpha I^{**} = \frac{\alpha^2\rho h}{2k}$$

In the First Best, from equation (15):

$$
\begin{aligned}
\chi^* &= (1 - p_1)\, q_{01} I_0^* + p_1 q_{11} I_1^* \\
&= \alpha(1 - \delta) I_0^* + \alpha \delta I_1^* \\
&= \left(\frac{(1 - \delta)^2}{\alpha + \delta - 2\alpha\delta} + \frac{\delta^2}{1 - \alpha - \delta + 2\alpha\delta} \right) \frac{\alpha^2 \rho h}{2k}
\end{aligned}
$$

To see that $\chi^* > \chi^{**}$, consider the following. First, the statement is equivalent to:

$$
1 < \frac{(1 - \delta)^2}{\alpha + \delta - 2\alpha\delta} + \frac{\delta^2}{1 - \alpha - \delta + 2\alpha\delta} \tag{19}
$$

Notice further that, with $\alpha = 1/2$, the inequality can be written as

$$
0 < 1/2 + 2\delta(1 - \delta) \tag{20}
$$

and it always holds when $\alpha = 1/2$.

Now, differentiate the right hand side of (19) with respect to α:

$$
\frac{\partial}{\partial \alpha} \left(\frac{(1 - \delta)^2}{\alpha + \delta - 2\alpha\delta} + \frac{\delta^2}{1 - \alpha - \delta + 2\alpha\delta} \right) = (1 - 2\delta) \left[\frac{\delta}{p_1} - \frac{1 - \delta}{1 - p_1} \right] \left[\frac{\delta}{p_1} + \frac{1 - \delta}{1 - p_1} \right] < 0 \tag{21}
$$

because

$$
\begin{aligned}
p_1 &= 1 - \alpha - \delta + 2\alpha\delta \\
(1 - 2\delta) &< 0 \\
\left[\frac{\delta}{p_1} + \frac{1 - \delta}{1 - p_1} \right] &> 0
\end{aligned}
$$

The first two statements are obvious. For the third, consider that it is equivalent to

$$
\delta + p_1 > 0
$$

which can be written as

$$
\delta + 1 - \alpha - \delta + 2\alpha\delta > 0
$$

Simplifying yields

$$
2\delta(1 - \alpha) > (1 - \alpha)
$$

which always holds because $\delta > 1/2$.

Thus, by the negativity of derivative (21), inequality (19) holds with $\forall \alpha \in (0, 1/2)$. ∎

References

The 9/11 Commission Report (2004) by the National Commission on Terrorist Attacks, US Government Printing Office, Washington D.C.

Becker G. (1968) Crime and Punishment: An Economic Approach, *Journal of Political Economy*, 76. (March/April) pp 169–217

Camdessus, M (1998) Speech to the Financial Action Task Force, Paris

Economic Perspectives (2001) The Fight against Money Laundering, US Department of State, Vol. 6. No 2, May 2001, Electronic Journal

El-Qorchi, M. (2002) Hawala, Finance and Development, December, Vol. 39, No. 4

FBI (2001) Money Laundering, FBI Law Enforcement Bulletin, by William R. Schroeder, May, Vol 70, No 5, pp. 1-9.

FinCEN (2004a) $25 Million Civil Money Penalty Against Riggs Bank N.A., May 13, http://www.fincen.gov/riggs6.pdf

FinCEN (2004b) In the Matter of AmSouth Bank, October 12 http://www.fincen.gov/amsouthassessmentcivilmoney.pdf

FinCEN (2005a) The SAR Activity Review, Trends, Tips and Issues, #8, April http://www.fincen.gov/sarreviewissue8.pdf

FinCEN (2005b) Statement of William J. Fox, Director, before the United States Senate Committee on Banking, Housing and Urban Affairs, April 26, http://www.fincen.gov/foxtestimony042605.pdf

FinCEN (2005c) SAR Activity Review, Issue 4, May, http://www.fincen.gov/sarreviewmay2005.pdf

FinCEN (2005d) FinCEN's 314(a) Fact Sheet, June 08, http://www.fincen.gov/314afactsheet.pdf

FinCEN (2005e) FinCEN and OCC (Office of the Comptroller of Currency) Joint Release, August 17, http://www.fincen.gov/pressrelease08172005.pdf

GAO (2004) US General Accounting Office: Anti-Money Laundering: Issues Concerning Depository Institution Regulatory Oversight, GAO-04-833T, June 3, http://www.gao.gov/new.items/d04833t.pdf

Gordon, N. J. (2005) The Rise of Independent Directors, 1950-2000: Towards a New Corporate Governance Paradigm, mimeo, Columbia Law School

Holmström, B. and P. Milgrom (1991) Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics, and Organization*, Spec. Issue, 7, pp. 24-52.

Kofman, F. and J. Lawarrée (1993) Collusion in Hierarchical Agency, *Econometrica*, 61, pp. 629-56.

Lal, B. (2003) *Money Laundering: An Insight into the Dark World of Financial Frauds*, Siddharth Publications, Delhi, India

Money Laundering Special Report (2003) Bureau of Justice Statistics, complied by M. Motivans, July 2003, NCJ 199574

Masciandaro, D. (1999) Money Laundering: The Economics of Regulation, *European Journal of Law and Economics*, 7, no. 3: 225-240.

Masciandaro, D., and U. Filotti (2001) Money Laundering Regulation and Bank Compliance Costs: What do your Customers Know? Economics and the Italian Experience, Journal of Money Laundering Control, Vol. 5, No. 2, pp. 133-145

Masciandaro, D. and A. Portolano (2002) Inside the Black (list) Box: Money Laundering, Lax Financial Regulation, Non-cooperative Countries. A Law and Economics Approach. Paulo Baffi Center, Bocconi University and Bank of Italy

OCC (2005) Memorandum, Subject: Bank Secrecy Act, From: Wayne Rushton, Tim Long, and Doug Roeder, Committee on Bank Supervision, April 25

Posner, R. A. (2004) The 9/11 Report: A Dissent, The New York Times, August 29, S7, p1

Reuter, P. and E. M. Truman (2004) *Chasing Dirty Money – The Fight against Money Laundering*, Institute for International Economics, Washington DC

Schneider, F. (2002) The Size and Development of the Shadow Economies of 22 Transition and 21 OECD Countries, IZA Discussion Paper, No. 514

Stein J. (2002) Information Production and Capital Allocation: Decentralized versus Hierarchical Firms, *Journal of Finance*, Vol. LVII, No.5 October, pp 1891-1922

Tanzi, V. (1996) Money Laundering and the International Financial System, IMF Working Paper No. 55, Washington: International Monetary Fund

Tirole, J. (1986) Hierarchies and Bureaucracies: On the Role of Collusion in Organizations, *Journal of Law, Economics and Organization*, 2, pp. 181-214

United Nations Office on Drugs and Crime (2005) Money Laundering
web: http://www.unodc.org/unodc/en/money_laundering.html

US Attorneys' Manual (2005) http://www.usdoj.gov/usao/eousa/foia_reading_room/usam/

Wall Street Journal (2004a) Bob Dole Goes Banking – and Trips the Alarm, Sept. 3, C1

Wall Street Journal (2004b) Expanding in an Age of Terror, Western Union Faces Scrutiny, October 20, A1

Wall Street Journal (2005) US Banks Overreport Data for the Patriot Act, July 07, C1

Verrecchia R. E. (2001) Essays on Disclosure, *Journal of Accounting and Economics*, 32, pp. 97-180.