Everlasting Fraud*

Vivian W. Fang

Nan Li

Wenyu Wang

Gaoqing Zhang

September 2, 2021

Abstract

This paper models the interdependent mechanisms of corporate fraud and regulation. Our analyses yield two key insights. First, fraud is a never-ending game of cat and mouse because the strength of detection optimally matches the severity of fraud in equilibrium. Second, anti-fraud regulations can tamp down fraud pro tem by sharply decreasing the most fraudulent firms' net benefits from continuing fraud. However, concentration of regulatory resources on these firms allows other firms to be more aggressive. As such, regulations do not eradicate fraud but synchronize firms' otherwise idiosyncratic fraud decisions and lead to fraud waves. Empirical examinations of these insights provide supporting evidence. These results carry strong policy implications, offering a realistic understanding of fraud as a permanent risk in the financial markets and the limited efficacy of anti-fraud regulations.

*Fang, Li, and Zhang are at the Carlson School of Management, University of Minnesota, Minneapolis, MN 55455; and Wang is at the Kelley School of Business, Indiana University, Bloomington, IN 47405. We are grateful to helpful comments from Xu Jiang, Lorien Stice-Lawrence, Anastasia Zakolyukina, and conference participants at the Columbia Burton Accounting Conference. Please send correspondence to fangw@umn.edu.

1 Introduction

From the original Ponzi scheme of 1920 to the collapse of Enron in 2001, Lehman Brothers in 2008, and Wirecard in 2020, the history of the financial markets is marred by a continuous stream of accounting scandals. Billions of dollars were lost as a result of these financial disasters, which shook investors' confidence, destroyed companies, and ruined peoples' lives. In response, reforms in the regulatory framework of financial reporting often followed, with the aim of cracking down on fraud. For example, former president George W. Bush characterized the Sarbanes-Oxley Act of 2002 as "the most far-reaching reforms of American business practice" that include "tough new provisions to deter and punish corporate and accounting fraud and corruption..." The Dodd-Frank Act of 2010 further expanded the efforts to fight fraud. The Act, via its Whistleblower Program, empowered the Securities and Exchange Commission (SEC) to reward whistleblowers in unprecedented ways.

But, what if fraud is a persistent feature of the financial markets? If financial reporting failure is a permanent risk, then to what extent can anti-fraud regulations achieve their stated goals of cracking down on fraud? This study investigates these two questions by probing the interdependent mechanisms of corporate accounting fraud and anti-fraud regulation.

We begin by building a multi-period model featuring a representative firm and a regulator. At the end of each period, the firm manager issues a potentially biased earnings report to the market after privately observing the firm's economic earnings (or fundamental cash flows). Based on the report, the market forms a rational expectation of the firm's current and future economic earnings and estimates firm value. The regulator utilizes a detection technology to inspect the firm's report. With a certain probability, the technology uncovers the fraudulent amount of the report and reveals it to the market.

The manager and the regulator each solve a maximization problem. The manager chooses the fraud amount in each period to maximize firm value, by weighing his marginal benefit(hereafter MB) and marginal cost (hereafter MC) of committing fraud. The MC is linked to detection likelihood. The MB depends on how much the market values the reported earnings and increases with the amount of fraud built to date. An increase in cumulative fraud adds information uncertainty about the firm both in the current period and in future periods, which in turn boosts the value of the new earnings report and potential return from inflating the report. The regulator decides on the amount of resources spent on a detection technology. In doing so, she seeks to maximize the informativeness of the firm's report, by weighing her MB and MC of detecting fraud. The MC is also linked detection likelihood, as a higher likelihood calls for a greater amount of regulatory resources. The MB also increases with the amount of fraud built to date, as it depends on the extent to which detection helps investors unravel fraud, which then allows them to base the conjectured firm value on true economic earnings rather than inflated earnings; catching the firm with a higher level of cumulative fraud clears more information uncertainty both in the current period and in future periods.

Analyses of this single-firm model begin to tell why fraud may never cease to exist. Although the manager and the regulator each solve a maximization problem independently, the two problems are intertwined with their MBs and MCs essentially co-moving. In equilibrium, the regulator chooses the optimal level of detection likelihood (by spending the corresponding amount of resources on detection), anticipating the optimal level of fraud committed by the manager, and vice versa. If the regulator anticipates a low level of fraud built up in the firm, then she would spend little on detection. The manager thus continues to commit fraud as the MB likely outweighs the MC. As fraud gradually builds up, a higher information uncertainty further incentivizes the manager to commit fraud. At the same time, the regulator would increase spending on detection. The two effects go hand-in-hand, simultaneously increasing the manager's MB and MC. When fraud reaches a critical level, the regulator would concentrate resources on the firm and the MC of continuing fraud eventually dominates the MB. Upon detection, fraud is cleared in the firm, and the cycle repeats. This rationale explains the time-series persistence of fraud within firms.

Analyses of an expanded, three-firm model make a separate case for everlasting fraud. H-, M-, and L-firm represent the firm with a high, medium, and low level of cumulative fraud, respectively. As in the single-firm model, the strength of detection matches the severity of fraud in equilibrium. Hence, with three firms in play, the regulator rationally allocates most resources towards H-firm. Ironically, M- and L-firms may factor in the regulator's decision and become more aggressive because their actions would be better masked until H-firm is caught (upon which M-firm becomes next target in line). This rationale explains the cross-sectional persistence of fraud across firms.

The question then arises is whether anti-fraud regulations can still achieve their stated goals of cracking down on fraud. Analyses of the multi-firm model help evaluate the efficacy of such regulations. Indeed, anti-fraud regulations are able to tamp down fraud by effectively lowering H-firm's net benefits from continuing fraud. Before detection, concentration of regulatory resources on H-firm greatly increases its MC of committing fraud. Upon detection, the MB becomes minuscule because a sharply declined uncertainty renders the firm's earnings report less useful and fraudulent reporting less valuable. Yet, the rational allocation of regulatory resources towards the more fraudulent firms may imply less scrutiny of less fraudulent firms, allowing the latter's fraudulent behavior to go undetected and their level of fraud to catch up—a side effect discussed earlier. As such, despite the "cracking-down" on H-firm, anti-fraud regulations do not eradicate fraud. Rather, they synchronize firms' fraud decisions, which may otherwise be idiosyncratic, and induce corporate fraud waves over time.

We take these insights to data. In our multi-firm model, firms are set apart by their level of cumulative fraud. Cumulative fraud directly impacts firms' information uncertainty and we use implied volatility of standardized options to capture fraud-induced information uncertainty. This proxy fits well with the theoretical construct that we intend to capture because it reflects the variance of the market's estimate about a firm's value conditional on all available information.[1] Relying on this proxy, we conduct three analyses.

First, we link implied volatility to the MB of committing fraud. This analysis is a joint test of the model prediction that a rising level of cumulative fraud motivates fraud by exacerbating information uncertainty and our use of implied volatility as a proxy for fraud-induced information uncertainty. We find that analysts' revision of earnings estimates for the next quarter is more responsive to unexpected earnings of the current quarter when implied volatility is higher. This finding is consistent with information uncertainty boosting the value of accounting reports and the potential return from reporting fraudulently, and provides validity

[1] The interpretation of implied volatility as a proxy for conditional variance dates back to the seminal work of Black and Scholes (1973).

for using implied volatility to capture fraud-induced information uncertainty.

Second, we examine a core model prediction that the strength of detection matches the severity of fraud. We show that a firm is more likely to be revealed to have committed fraud in the past (i.e., an earnings restatement is announced or accounting irregularities detected in the current quarter), if the level of implied volatility prior to the quarter is higher. This finding is consistent with the regulator rationally allocating more resources towards more fraudulent firms, thus increasing the likelihood of catching fraud at these firms.

The third analysis intends to show the convergence of fraud level across firms over time. Specifically, we sort firm-quarters in the sample into quintiles based on the firm's level of implied volatility prior to a quarter, and show that firms in a higher-ranked quintile (i.e., those having a higher level of implied volatility prior to a quarter) have a smaller increase in implied volatility during the quarter. This finding supports the model prediction that firms with a higher level of cumulative fraud are more cautious about continuing fraud (because they anticipate closer scrutiny from the regulator) while firms with a lower level of cumulative fraud are more aggressive at committing fraud (because they can hide under the radar). One concern is that this finding merely reflects the mean-reverting nature of fraud. To mitigate the concern, we show that the negative relation between prior level of implied volatility (as measured by quintile rank) and the increase in implied volatility is stronger if a wave of corporate fraud recently surfaced in the firm's industry. If firms do converge in their level of fraud over time, particularly after a regulation manages to crack down on fraud for a group of firms at the same time, corporate fraud waves likely arise.

To our best knowledge, this is the first study to examine the joint mechanisms of corporate fraud and regulation in a dynamic setting. Prior theories of earnings manipulation often assume an exogenous cost related to regulation in a static setting (e.g., Fischer and Verrecchia (2000); Dye and Sridhar (2004)).[2] Closely related to our study, Povel et al. (2007) examine the joint mechanisms of corporate fraud and investor monitoring. Their model, focusing on a single firm in a static setting, does not consider the dynamic features of fraud among multiple firms. Beyer et al. (2019) study earnings manipulation in a dynamic setting but do not

[2]For a comprehensive review of theories on earnings manipulation, please see two recent surveys by Ewert and Wagenhofer (2012) and Stocken (2013).

examine regulators' endogenous detection. By modeling both the firm's fraud decision and the regulator's enforcement decision, our study takes a holistic view in analyzing the formation and evolvement of corporate fraud and evaluating the efficacy of anti-fraud regulations. Our analyses yield two important takeaways. First, fraud is a cat-and-mouse game that is unlikely to end because the manager's fraud decision and the regulator's detection decision are intertwined based on co-moving MBs and MCs. Given that every regulator faces some extent of budget constraints and uncovering corporate fraud inevitably consumes regulatory resources, the amount of resources allocated to a firm should match its level of fraud. Even though maximizing detection intensity at all times is likely the most effective at cracking down on fraud in the economy, it is neither feasible nor socially optimal.

Second, our results offer a more complete picture of fraud, regulation, and their interaction. In particular, our results speak to two prior observations that corporate frauds tend to come in waves and that not only frauds lead regulations but also regulations lead frauds (Hail et al. (2018)). In our model, these patterns arise not because regulations are ineffective. Rather, regulations effectively tamp down fraud in the short term but in the long term, synchronize firms' fraud decisions and allow a wave of frauds to resurface. Hence, fraud remains a permanent risk in the financial markets and the effectiveness of regulations is limited.

Our study also fits in the broad literature of crime in economics. In particular, several studies have offered answers to the question of why maximal penalties are not necessarily desirable in preventing crime. For example, Mookherjee and Png (1992) point out that the enforcement authority should optimally vary its monitoring effort according to a signal of the action selected by the potential offender. Bond and Hagerty (2010) prove that marginal penalties are more attractive in the Pareto inferior crime wave equilibrium. Our results also speak to this point but work through a unique mechanism. As a white-collar crime, fraud is a calculated decision that is fundamentally different from violent crimes. For fraud, we are able to endogenize the economic benefits and costs that enter the manager's calculus. In contrast, the benefits of committing a violent crime are often exogenous by nature (e.g., it is hard to quantify a murderer's marginal utility). Our analyses yield an important insight about accounting fraud—its MB and MC go hand-in-hand—which makes it distinct from other types of crimes. For this reason, a policy that lets punishment fit the crime should work uniquely well in addressing fraud, because once an anti-fraud regulation is sufficiently tough and cracks down on the most fraudulent firms, these firms' MBs of committing fraud also drop sharply upon detection (and so the regulator can safely and should optimally decrease the level of enforcement).

2 Single-firm Model

2.1 Model Setup

We consider a baseline setting in which a representative firm generates economic earnings s_t in each period $t \in \{1, 2, ..., \infty\}$. We assume that s_t follows an AR(1) process such that

$$s_t = \rho s_{t-1} + \varepsilon_t, \tag{1}$$

where the correlation coefficient $\rho \in (0, 1)$ and the random variable $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$. In each period, the firm manager privately learns the realization of the firm's economic earnings s_t and issues a report r_t. Investors use the report to update their expected firm value V_t. We assume that V_t is set by a competitive market and equals the firm's total discounted future earnings in expectation:

$$V_t = \sum_{k=t}^{\infty} \delta^{k-t} E^I \left[s_k | \mathcal{F}_t\right] = \frac{E^I \left[s_t | \mathcal{F}_t\right]}{1 - \delta\rho}, \tag{2}$$

where $E^I \left[\cdot | \mathcal{F}_t\right]$ denotes the investors' expectation, $\mathcal{F}_t \equiv \{r_t, r_{t-1}, ...r_1\}$ denotes the set of the firm's reports up to time t, and $\delta \in (0, 1)$ denotes the discounting factor. The manager may have incentives to manipulate the report r_t to boost V_t, because a greater firm value typically means higher equity compensation and better career prospects for himself.

We model the manager's earnings manipulation decision as follows. In each period t, after observing the true economic earnings s_i, the manager chooses manipulation $m_t \geq 0$ that adds m_t errors $\{\xi_l\}_{l=1}^{m_t}$ to s_i. The choice of manipulation m_t is observable only to the manager.

Each error generates either 0 or 1 with $\Pr(\xi_l = 0) = q \in (0, 1]$. The report is then given by:[3]

$$r_t = s_t + \sum_{l=1}^{m_t} \xi_l. \tag{3}$$

Using the central limit theorem, we can approximate the distribution of $\sum_{l=1}^{m_t} \xi_l$ as

$$\sum_{l=1}^{m_t} \xi_l \sim N\left(m_t\left(1-q\right), m_t q\left(1-q\right)\right). \tag{4}$$

With the manager's manipulation choice $m_t \geq 0$, the report becomes:

$$r_t = s_t + m_t(1-q) + \sqrt{m_t q\left(1-q\right)}\eta_t. \tag{5}$$

$\eta_t \sim N(0, 1)$ is a standard normal random variable that is independent of all other variables in the model. Equation (5) suggests that manipulation has a dual effect on the report: m_t increases the mean of the report but decreases its precision.

We assume that, in each period t, the firm's fraudulent activity is uncovered with an aggregate probability of $d_t = d_0 + d_{rt}$, $d_t \in (0, 1)$. $d_0 \in (0, 1)$ denotes the probability with which fraud is detected in the absence of regulatory involvement, which highlights the fact that other stakeholders (such as external auditors, whistleblowers, and short-sellers) may also play a monitoring role. $d_{rt} \in (0, 1)$ denotes the probability with which fraud is detected with direct regulatory efforts.[4] Specifically, the regulator influences d_{rt} by utilizing a detection technology to inspect the manager's report r_t; the technology consumes regulatory resources of $\frac{\kappa}{2}\left(d_{rt}\right)^2$. If the regulator successfully detects fraud, she would require the manager to restate the report to equal the true earnings, i.e., $r_t = s_t$, and imposes a penalty C_t on the

[3] A firm's earnings aggregate different line items in the financial statements; that is, net income equals sales revenue minus cost of sales and other expenses. By the way that we model the manipulation decision, a manager may choose to add one unit of positive bias to each of the line items (e.g., either over-report a revenue item or under-report an expense item) to inflate the earnings report. However, the manager's manipulation attempts may be blocked by the firm's internal control system, and q denotes the probability with which each of the manager's fraudulent attempts fails.

[4] We acknowledge that there may be some interaction between d_0 and d_{rt}, although the direction is theoretically ambiguous as a higher d_0 may render regulatory efforts less necessary (hence a lower d_{rt}) or it may prompt the regulator to step in (hence a higher d_{rt}). While potentially interesting, this interaction is outside the scope of our model.

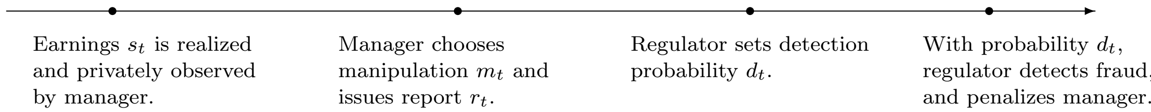

Figure 1: Timeline of the period-t game

manager that is proportional to the fraudulent amount:

$$C_t = c(r_t - s_t), \tag{6}$$

where the coefficient $c > 0$. We assume that the regulator sets the aggregate detection probability d_t by choosing d_{rt} to maximize the informativeness of the set of reports $\mathcal{F}_t \equiv \{r_t, r_{t-1}, ...\}$ about the firm value V_t.[5] Note that since the earnings follow an AR(1) process, the period-t earnings s_t is a sufficient statistic to estimate all of the firm's future earnings and the firm value. Therefore, maximizing the informativeness of $\mathcal{F}_t$ is equivalent to maximizing the informativeness about s_t, or minimizing the conditional variance about s_t:

$$\Phi_t \equiv var\left(s_t|\mathcal{F}_t\right). \tag{7}$$

Note that it is most cost effective for the regulator to focus on detecting fraud in the current period's report r_t and uncovering the true earnings s_t, because s_t is a sufficient statistic for estimating all of the firm's future earnings (as shown in equation (2)). Conditional on the revelation of s_t, detecting fraud in the firm's past reports, $\{r_{t-1}, r_{t-2}, ...\}$, incurs additional costs but does not generate any incremental information benefits.

Figure 1 summarizes the timing of events in each period t.

[5] In solving the model, we substitute d_{rt} with $d_t - d_0$ instead of substituting d_t with $d_0 + d_{rt}$ to simplify the algebra. Since the two are mathematically equivalent, this simplification does not affect any inference.

2.2 Analysis

In this section, we analyze the manager's optimal manipulation choice m_t^* and the regulator's equilibrium detection choice d_t^*.[6] For ease of readability, we present only the equations that illustrate the key intuitions from the model, leaving the detailed derivations to Appendix I.

2.2.1 The manager's problem

We assume that the manager derives utility from his compensation (or career prospects) that is proportional to the firm value. To ease notation, we scale up the manager's utility so that it simply equals the firm value perceived by investors. In each period t, the manager maximizes the total present value of his future expected payoffs by choosing manipulation m_t:

$$U_t = \max_{m_t} E^M \left[\sum_{k=0}^{\infty} \delta^k u_{t+k} \middle| s_t, \mathcal{F}_t \right]. \tag{8}$$

where $E^M [\cdot | s_t, \mathcal{F}_t]$ denotes the manager's expected utility in period t based on his information set, which includes s_t, his privately observed true earnings of the firm for the period, and $\mathcal{F}_t$, the firm's publicly released earnings reports in the past. The manager's period-t payoff is:

$$u_t = d_t^* \left(\frac{s_t}{1-\delta\rho} - C_t \right) + (1 - d_t^*) \frac{E^I [s_t | \mathcal{F}_t]}{1-\delta\rho}, \tag{9}$$

where d_t^* is the regulator's period-t detection probability anticipated by the manager.

The two terms of equation (9) represent the manager's utility under two different scenarios, respectively. In the first scenario, the manager's fraudulent behavior is detected with probability d_t^*. As a result, the firm's true earnings s_t is revealed to investors, who would then update the firm value to $\frac{s_t}{1-\delta\rho}$ based on s_t. The manager suffers a penalty of C_t proportional to m_t as shown in equation (6). In the second scenario, the manager's fraudulent behavior goes undetected with probability $1 - d_t^*$. Thus, the firm's true earnings s_t remains unknown to investors, who would then have to set the firm value to $\frac{E^I[s_t|\mathcal{F}_t]}{1-\delta\rho}$ based on the firm's public

[6] Technically speaking, although the regulator sets the detection probability d_t after the manager chooses m_t, the two essentially play a simultaneous-move game because the regulator does not observe m_t and thus cannot make d_t a function of m_t.

reports $\mathcal{F}_t$. The manager incurs no penalty.

Equation (9) also makes it clear that the manager's manipulation choice m_t only affects firm value when it is undetected. To solve the optimal m_t^*, we first analyze how the investors' conjectured firm value varies with m_t in each period

$$E^I\left[s_t|\mathcal{F}_t\right] = (1-w_t)\rho E^I\left[s_{t-1}|\mathcal{F}_{t-1}\right] + w_t\left[r_t - m_t^*(1-q)\right]. \tag{10}$$

As shown, the investors' conjectured firm value is a weighted average of their prior of s_t (the first term) and the incremental information that they gain from seeing the report r_t (the second term). The prior builds on the AR(1) process of s_t and equals $\rho E^I\left[s_{t-1}|\mathcal{F}_{t-1}\right]$. To extract information from the new report, investors rationally subtract the expected manipulation $m_t^*(1-q)$, leading to a refined signal $r_t - m_t^*(1-q)$. The weight

$$w_t = \frac{\rho^2\Phi_{t-1} + \sigma_\varepsilon^2}{\rho^2\Phi_{t-1} + \sigma_\varepsilon^2 + m_t^* q\,(1-q)} \tag{11}$$

captures the value relevance of the earnings report, with Φ_{t-1} being the inverse precision of the prior, as defined in equation (7). When Φ_{t-1} is larger, the prior is less precise and so the investors have to place a greater weight on the current report to infer firm fundamentals.[7]

Equation (10) suggests that undetected manipulation m_t has a contemporaneous effect as well as an intertemporal effect on the investors' conjectured firm value. To see the contemporaneous effect, note that m_t inflates the current earnings report r_t, which in turn boosts firm value in the current period $E^I\left[s_t|\mathcal{F}_t\right]$; this effect works through the second term of the equation. To see the intertemporal effect, note that $E^I\left[s_t|\mathcal{F}_t\right]$ serves as the prior for the investors to conjecture future earnings s_{t+1}, so as m_t inflates $E^I\left[s_t|\mathcal{F}_t\right]$, it also boosts firm value in the next period $E^I\left[s_{t+1}|\mathcal{F}_t\right]$; this effect works through the first term of the equation. In fact, such bias propagates to all future s_{t+k} for $k > 0$ through the recursive form of equation (10).[8]

[7]It is noteworthy that m_t^* in Equation (10) is the investors' conjectured manipulation by the manager, and the manager factors in the investor's conjecture in his maximization problem. In equilibrium, this conjecture equals the manager's optimal manipulation choice m_t^*.

[8]This can be easily seen by shifting equation (10) forward by k period from t to $t+k$.

We summarize the contemporaneous effect and the intertemporal effect of m_t below as

$$\frac{\partial E\left[s_{t+k}|\mathcal{F}_t\right]}{\partial m_t}=\begin{cases} w_t\left(1-q\right) & \text{if } k=0 \\ \rho^k\left[\prod_{\ell=1}^{k}(1-w_{t+\ell})\right]w_t\left(1-q\right) & \text{if } k>0. \end{cases} \tag{12}$$

Now we solve the manager's optimal choice of manipulation, m_t^*. Taking derivative of U_t in equation (8) with respect to m_t and then substituting in equation (12) derived above, we obtain the first-order condition (F.O.C.) as

$$\begin{aligned}\underbrace{c\left(1-q\right)d_t^*}_{\text{MC of } m_t} = & \underbrace{\left(1-d_t^*\right)\frac{w_t(1-q)}{1-\delta\rho}}_{\text{MB of } m_t \text{ from the contemporaneous effect}} \\ & +\underbrace{\sum_{k=1}^{\infty}\delta^k\left[\prod_{\ell=0}^{k}\left(1-d_{t+\ell}^*\right)\right]\frac{\rho^k\left[\prod_{\ell=1}^{k}(1-w_{t+\ell})\right]w_t(1-q)}{1-\delta\rho}}_{\text{MB of } m_t \text{ from the intertemporal effect}}\end{aligned} \tag{13}$$

The MC of manipulation, expressed on the left hand side (LHS) of the F.O.C., increases with the regulator's optimal choice of detection probability d_t^*, which is correctly conjectured by the manager. The MB of manipulation, expressed on the right hand side (RHS) of the F.O.C., arises from both the contemporaneous effect and the intertemporal effect discussed above. The difference is that the MB from the contemporaneous effect is only affected by the likelihood of no detection in the current period $(1-d_t^*)$, while the MB from the intertemporal effect is affected by the likelihood of no detection up to a future period of interest $\left[\prod_{\ell=0}^{k}\left(1-d_{t+\ell}^*\right)\right]$.

Substituting equation (11) for $w_{t+\ell}$ in equation (13) and solving for $m_{t+\ell}^*$, we find that the manager's current manipulation choice m_t^* depends on his future manipulation choices $\left\{m_{t+1}^*, m_{t+2}^*, ...\right\}$. By induction, we can write m_t^* in a recursive form, as shown in Lemma 1 below. Appendix I provides more details on the derivation.

Lemma 1 *In each period t, given the regulator's equilibrium detection choice d_t^* conjectured* *by the manager, the manager chooses the optimal manipulation*

$$m_t^* = \frac{\rho^2 \Phi_{t-1} + \sigma_\varepsilon^2}{q(1-q)} \left[\frac{1-d_t^*}{c d_t^*} \left(\frac{1}{1-\delta\rho} + \delta\rho c q (1-q) \frac{d_{t+1}^* m_{t+1}^*}{\rho^2 \Phi_t + \sigma_\varepsilon^2} \right) - 1 \right], \tag{14}$$

where Φ_t is the conditional variance of s_t, as defined in equation (7).

Given the manager's optimal manipulation choice, Φ_t evolves endogenously in the model, and standard Bayesian updating yields its law of motion, as shown in Lemma 2 below:

Lemma 2 *In each period t, if the regulator detects fraud, the conditional variance about the firm's earnings s_t drops to zero, i.e., $\Phi_t \equiv 0$. If the regulator fails to detect fraud, Φ_t is a function of the last-period Φ_{t-1} and the manager's period-t manipulation in equilibrium m_t^**

$$\Phi_t(d_t^*, \Phi_{t-1}) = \frac{m_t^* q(1-q)\left(\rho^2 \Phi_{t-1} + \sigma_\varepsilon^2\right)}{\rho^2 \Phi_{t-1} + \sigma_\varepsilon^2 + m_t^* q(1-q)}. \tag{15}$$

The law of motion (15) is intuitive. It states that the uncertainty about the firm's earnings Φ_t is increasing in both the prior uncertainty Φ_{t-1} and the manager's equilibrium manipulation m_t^* in the current period. Iterating (15) over time suggests that Φ_t essentially depends on the manager's undetected manipulation accumulated in the past, i.e., $\{m_t^*, m_{t-1}^*, ...\}$. We thus hereafter refer to the state variable Φ_t as either the information uncertainty about the firm fundamental in period t or the cumulative level of fraud up to period t interchangeably.

2.2.2 The regulator's problem

We then analyze the regulator's choice of detection probability d_t, given the manager's equilibrium manipulation choice m_t^* in equation (14). Specifically, the regulator seeks to maximize her total utility in future periods

$$W_t = \max_{d_t} E^I \left[\sum_{k=0}^{\infty} \delta^k v_{t+k} \middle| \mathcal{F}_t \right]. \tag{16}$$

$E^I[\cdot|\mathcal{F}_t]$ indicates that the regulator has the same information set as the investors. v_t is the regulator's period-t utility

$$v_t = -(1-d_t)\,\Phi_t - \frac{\kappa}{2}(d_t - d_0)^2. \tag{17}$$

Equation (17) sums up the regulator's utility in period t under two scenarios. If detection succeeds with probability d_t, then the true earnings s_t is revealed and the conditional variance Φ_t drops to zero. Alternatively, if detection fails with probability $1-d_t$, then the conditional variance remains at $\Phi_t > 0$. Under either scenario, the regulator incurs a cost for detection of $\frac{\kappa}{2}(d_t - d_0)^2$.

As in the manager's maximization problem, the regulator's choice of detection likelihood in period t also carries two effects. First, a higher d_t increases the regulator's period-t utility by boosting the chance of detection success (upon which Φ_t is decreased to zero); this is the contemporaneous effect of detection. Second, clearing Φ_t also reduces the expected level of $\Phi_{t+\ell}$ for all $\ell > 0$ because Φ_t affects all future $\Phi_{t+\ell}$ through the law of motion specified in equation (7); this is the intertemporal effect of detection.

Now we solve the regulator's choice of optimal detection likelihood, d_t^*. We obtain the F.O.C. as

$$\begin{aligned}\underbrace{\kappa(d_t - d_0)}_{\text{MC of } d_t} = & \underbrace{\Phi_t - 0}_{\text{MB of } d_t \text{ from the contemporaneous effect}} \\ & + \underbrace{\delta\left[W_{t+1}(0) - W_{t+1}(\Phi_t)\right].}_{\text{MB of } d_t \text{ from the intertemporal effect}}\end{aligned} \tag{18}$$

$W_{t+1}(\Phi)$ denotes the regulator's objective function evaluated at an initial level of uncertainty Φ. As in the F.O.C. for the manager's problem, we express the MC of detection on the LHS and the MB of detection on the RHS. The MC is proportional to the amount of detection intensity contributed by the regulator, $d_{rt} = d_t - d_0$. The MB is increasing in the cumulative level of fraud Φ_t as it comes from clearing uncertainty about s_t in the current period (the contemporaneous effect) and decreasing prior uncertainty for all future periods (the intertemporal effect). Specifically, $W_{t+1}(0) - W_{t+1}(\Phi_t) > 0$ represents the capitalized value of resetting the initial uncertainty from Φ_t to zero for all future periods upon successful detection. Finally, we solve d_t^* from the F.O.C, which yields the following lemma.

Lemma 3 *In each period t, the regulator chooses the optimal detection probability*

$$d_t^*\left(\Phi_{t-1}\right) = d_0 + \frac{\Phi_t + \delta\left[W\left(0\right) - W\left(\Phi_t\right)\right]}{\kappa}, \tag{19}$$

where Φ_t is expressed recursively in equation (7).

2.2.3 The equilibrium

Because our model features an infinite horizon and both m_t^* and d_t^* can be written recursively as functions of Φ_{t-1}, we can treat Φ_{t-1} as the state variable for period t and characterize the equilibrium as a dynamic programming problem with the Bellman equations below. For ease of notation, we omit the time subscript and denote variables of the next period with a prime.

Proposition 1 *For a given level of accumulated past fraud Φ, the regulator's equilibrium detection choice $d^*\left(\Phi\right)$ and the manager's equilibrium manipulation choice $m^*\left(\Phi\right)$ are given by the following set of equations, with the two agents rationally anticipating each other's optimal policy function:*

$$d^*\left(\Phi\right) = d_0 + \frac{\Phi' + \delta\left[W\left(0\right) - W\left(\Phi'\right)\right]}{\kappa}, \tag{20}$$

$$m^*\left(\Phi\right) = \frac{\rho^2\Phi + \sigma_\varepsilon^2}{q\left(1-q\right)}\left[\frac{1-d^*}{cd^*}\left(\frac{1}{1-\delta\rho} + \delta\rho c q\left(1-q\right)\frac{d^*\left(\Phi'\right)m^*\left(\Phi'\right)}{\rho^2\Phi' + \sigma_\varepsilon^2}\right) - 1\right], \tag{21}$$

where

$$\Phi'\left(\Phi\right) = \frac{m^*\left(\Phi\right)q\left(1-q\right)\left(\rho^2\Phi + \sigma_\varepsilon^2\right)}{\rho^2\Phi + \sigma_\varepsilon^2 + m^*\left(\Phi\right)q\left(1-q\right)}, \tag{22}$$

$$W\left(\Phi\right) = -\left(1-d^*\right)\Phi' - \frac{\kappa}{2}\left(d^* - d_0\right)^2 + \delta\left[d^* W\left(0\right) + \left(1-d^*\right)W\left(\Phi'\right)\right]. \tag{23}$$

The dynamic programming problem in Proposition 1 does not have a closed-form solution so we solve the full model numerically to analyze the key properties of these policy functions. Appendix I provides details on the numerical method. In the analyses below, we first glean some insights into the solution by approximating it locally using Taylor expansion. We then present the results generated from our numerical solution.

First consider a first-order approximation to the condition on d^* in equation (20). Approximating the condition with a first-order Taylor expansion on the value function W around $\Phi' = 0$ gives that:

$$d^* = d_0 + \frac{1}{\kappa}\left[1 - \delta\left(\left.\frac{dW(\Phi)}{d\Phi}\right|_{\Phi=0}\right)\right]\Phi'. \tag{24}$$

The equation suggests that the regulator's choice of optimal detection strength d^* is increasing in Φ'. That is, the regulator matches the strength of fraud detection with the severity of fraud in equilibrium. This result is intuitive because the manager's manipulation in the past adds noises to the firm's reports and decreases informativeness. Since the regulator's objective is to clear fraud and restore informativeness of the firm's reports, her gains are higher from detecting reports with more extensive fraud. In other words, the regulator's MB of detection is increasing in the cumulative level of fraud and so is her choice of optimal detection strength.

Next, we use the approximated d^* in equation (24) to draw some inferences about the properties of the equilibrium manipulation $m^*(\Phi)$. Most interestingly, we find that m^* can be non-monotonic in Φ. To see this, recall that from Lemma 1, fixing the regulator's detection choice, m^* is increasing in Φ. Intuitively, all else equal, the manager has greater incentives to commit fraud as the market faces a higher uncertainty about the firm and relies on the manager's report to a larger extent. However, our discussion of d^* above suggests that as Φ increases, the regulator would invest more heavily in the detection technology, which deters manipulation and reduces m^*. The two countervailing effects go hand-in-hand, which may lead to a non-monotonic relation between m^* and Φ.

The analysis above builds on a linear approximation of the model solution. Next, we solve the model numerically to verify our findings. We set the six model parameters as follows: the subjective discount rate, δ, equals 0.9, a value commonly used in the literature; the success rate of manipulation, q, equals 0.5, an innocuous assumption in the model; the persistence of the AR(1) process that governs the dynamics of the economic earnings s_t, ρ, equals 0.88; the conditional standard deviation of the AR(1) process, σ_ε, equals 0.15; the detection cost

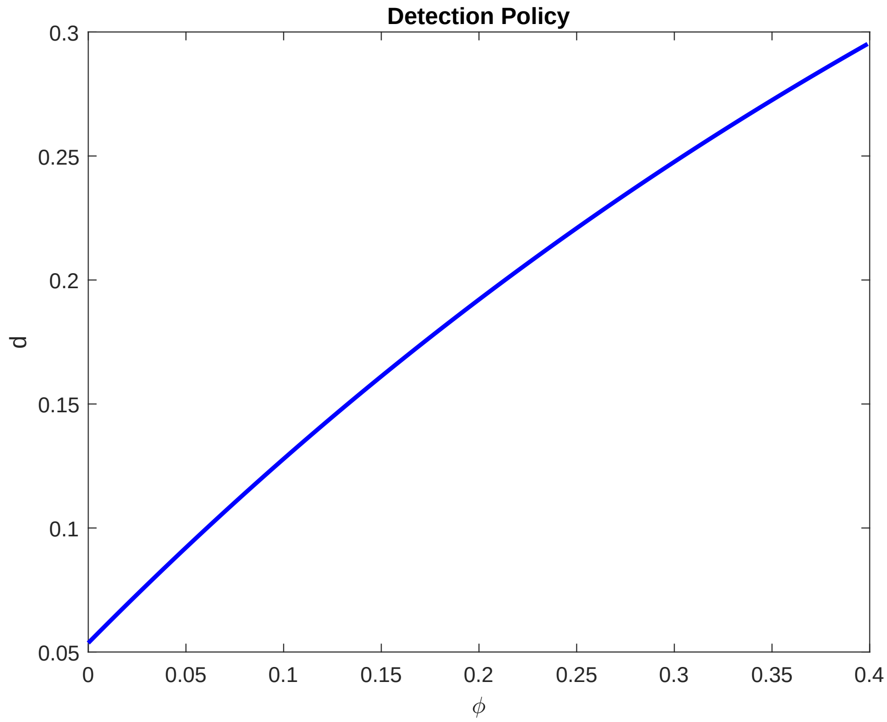

Figure 2: Equilibrium detection probability $d^*(\Phi)$.

parameter, κ, equals 2.5; and the manager's cost parameter, c, equals 3, which suggests that the fine imposed on the manager is three times of his manipulation amount upon detection.

Figure 2 depicts the regulator's optimal detection intensity d_t^* as a function of the firm's state variable Φ_{t-1}. Consistent with our analysis using linear approximation, the numerical solution suggests that a higher level of cumulative fraud increases the regulator's choice of detection intensity, which in turn increases the manager's MC of manipulation.

Figure 3 depicts the manager's optimal manipulation m_t^* also as a function of the state variable Φ_{t-1}. Based on the set of parameter values that we use in the numerical solution, we find that the equilibrium manipulation is hump-shaped in the firm's cumulative fraud. The intuition is clear: when Φ is very low (close to 0), the market is highly informed about the firm's economic earnings and puts little weight on the firm's new report. This implies a low MB of manipulation and fewer incentives for the manager to inflate the report. When Φ is very high, the regulator increases detection efforts, which sharply increase the MC of manipulation.

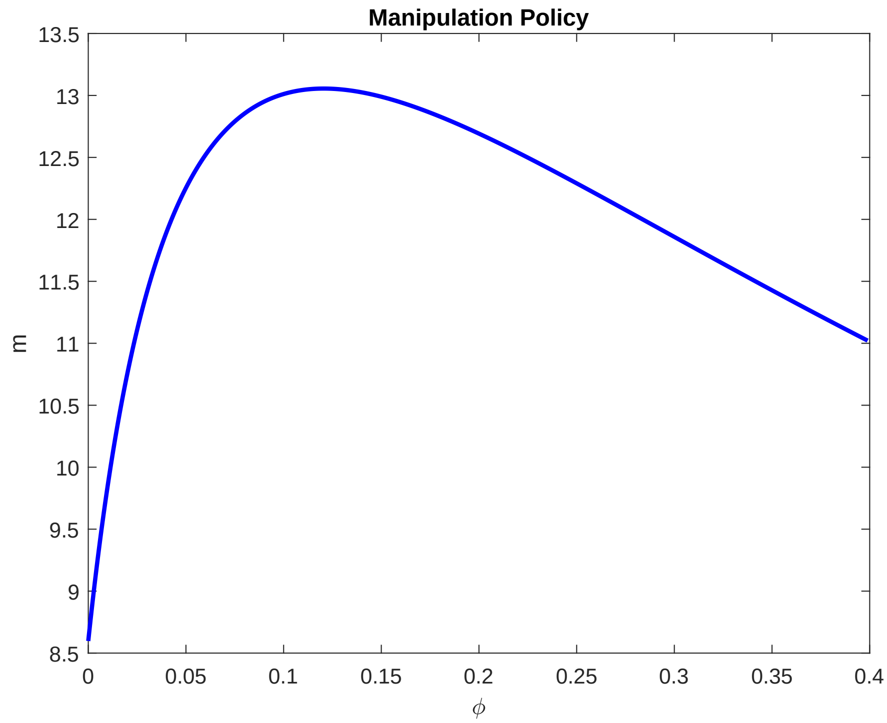

Figure 3: Equilibrium manipulation $m^*(\Phi)$.

Trading off the MB and MC of manipulation, the manager's optimal manipulation may appear in the intermediate range of Φ, leading to a hump-shaped relation between m^* and Φ.

3 Multi-firm Model

3.1 Model Setup

In this section, we expand the single-firm model to study the dynamic features of fraud among multiple firms. The model setup is similar as before with two exceptions. First, the economy contains N firms, and their economic earnings are independent of each other. This assumption allows us to abstract away from the effects of information spillovers, which are not a central focus of this study. Most of our numerical analyses focus on a special case with three firms, i.e., $N = 3$. Second, the regulator has to allocate limited resources among N firms towards fraud detection. Specifically, in each period, the regulator conducts an independent inspection of each firm's report and we denote the probability that the inspection uncovers fraud in firm i's report by $d_{it} = d_0 + d_{irt} \in [d_0, 1]$, where $i \in \{1, 2, ..., N\}$ and d_{irt} represents the regulator's choice of detection technology to influence the probability of detecting fraud at firm i. We assume that the total detection cost for each period is:

$$\frac{\kappa}{2}\left(\sum_{i=1}^{N} d_{irt}\right)^2. \tag{25}$$

The structure of this cost function is consistent with the regulator having a limited budget for fraud detection, in the sense that if she allocates more resources towards inspecting one firm's report, her MC of detecting fraud at other firms goes up.

3.2 Analysis

In the multi-firm model, the manipulation decision of each manager and the detection decision of the regulator can be similarly characterized as in the single-firm model. Both m_{it}^* and d_{it}^* can be written recursively as functions of the cumulative levels of past fraud at all firms, $\{\Phi_{1t-1}, \Phi_{2t-1}, \Phi_{3t-1}\}$. Hence we can treat $\{\Phi_{1t-1}, \Phi_{2t-1}, \Phi_{3t-1}\}$ as the set of state variables for period t and characterize the equilibrium as a dynamic programming problem with the Bellman equations below. For ease of notation, we omit the time subscript and denote variables of the next period with a prime.[9]

Proposition 2 *Consider a three-firm model. Given the levels of accumulated past fraud at the three firms $\{\Phi_1, \Phi_2, \Phi_3\}$, the manager in firm 1 chooses manipulation*

$$m_1^*\left(\Phi_1, \Phi_2, \Phi_3\right) = \frac{\rho^2\Phi_1 + \sigma_\varepsilon^2}{q\left(1-q\right)}\left(\frac{1-d_1^*}{cd_1^*}\left(\frac{1}{1-\delta\rho} + \frac{c\delta\rho q\left(1-q\right)}{\rho^2\Phi_1' + \sigma_\varepsilon^2} \times E\left[m_1^{*\prime} d_1^{*\prime}\right]\right) - 1\right), \tag{26}$$

[9] As in the single-firm model, in solving the multi-firm model, we continue to substitute d_{irt} with $d_{it} - d_0$ and solve for the optimal d_{it} to simplify the algebra.

and the regulator chooses to detect fraud at firm 1 with probability

$$\begin{aligned} d_1^*(\Phi_1,\Phi_2,\Phi_3) = \frac{1}{\kappa}\{&\Phi_1' + \delta(1-d_2^*)(1-d_3^*)\left[W\left(0,\Phi_2',\Phi_3'\right) - W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\ &+ \delta(1-d_2^*)d_3^*\left[W\left(0,\Phi_2',0\right) - W\left(\Phi_1',\Phi_2',0\right)\right] \\ &+ \delta d_2^*(1-d_3^*)\left[W\left(0,0,\Phi_3'\right) - W\left(\Phi_1',0,\Phi_3'\right)\right] \\ &+ \delta d_2^* d_3^*\left[W(0,0,0) - W\left(\Phi_1',0,0\right)\right]\} - (d_2^* + d_3^* - 3d_0), \end{aligned} \tag{27}$$

where

$$\Phi_i'(\Phi_1,\Phi_2,\Phi_3) = \frac{m_i^* q(1-q)\left(\rho^2\Phi_i + \sigma_\varepsilon^2\right)}{\rho^2\Phi_i + \sigma_\varepsilon^2 + m_i^* q(1-q)}, \tag{28}$$

$$\begin{aligned} E\left[m_1^{*\prime} d_1^{*\prime}\right] = {} & (1-d_2^*)(1-d_3^*)\, m_1^*\left(\Phi_1',\Phi_2',\Phi_3'\right) d_1^*\left(\Phi_1',\Phi_2',\Phi_3'\right) \\ & + d_2^*(1-d_3^*)\, m_1^*\left(\Phi_1',0,\Phi_3'\right) d_1^*\left(\Phi_1',0,\Phi_3'\right) \\ & + (1-d_2^*)\, d_3^* m_1^*\left(\Phi_1',\Phi_2',0\right) d_1^*\left(\Phi_1',\Phi_2',0\right) \\ & + d_2^* d_3^* m_1^*\left(\Phi_1',0,0\right) d_1^*\left(\Phi_1',0,0\right), \end{aligned} \tag{29}$$

$$\begin{aligned} W(\Phi_1,\Phi_2,\Phi_3) = {} & -(1-d_1^*)(1-d_2^*)(1-d_3^*)\left[\Phi_1' + \Phi_2' + \Phi_3' - \delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\ & -(1-d_1^*)\, d_2^*(1-d_3^*)\left[\Phi_1' + \Phi_3' - \delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\ & -(1-d_1^*)(1-d_2^*)\, d_3^*\left[\Phi_1' + \Phi_2' - \delta W\left(\Phi_1',\Phi_2',0\right)\right] \\ & -(1-d_1^*)\, d_2^* d_3^*\left[\Phi_1' - \delta W\left(\Phi_1',0,0\right)\right] \\ & - d_1^*(1-d_2^*)(1-d_3^*)\left[\Phi_2' + \Phi_3' - \delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\ & - d_1^* d_2^*(1-d_3^*)\left[\Phi_3' - \delta W\left(0,0,\Phi_3'\right)\right] \\ & - d_1^*(1-d_2^*)\, d_3^*\left[\Phi_2' - \delta W\left(0,\Phi_2',0\right)\right] \\ & + \delta d_1^* d_2^* d_3^* W(0,0,0) - \frac{\kappa}{2}\left(d_1^* + d_2^* + d_3^* - 3d_0\right)^2. \end{aligned} \tag{30}$$

The manipulation choices $\{m_2^*, m_3^*\}$ and the detection choices $\{d_2^*, d_3^*\}$ at firms 2 and 3 can be analogously derived and given in the appendix.

Proposition 2 suggests that the dynamics of the manipulation and the detection decisions in the multi-firm model is largely in line with that in the single-firm model. There are, however, two new insights. First, the managers' manipulation decisions are endogenously linked because the regulator's choices of detection intensity are interdependent across firms. As such, a manager's manipulation choice becomes a function of the cumulative levels of fraud at all firms. Second, while the manager in the single-firm model is able to precisely conjecture the future equilibrium manipulation and detection choices $\left\{m^*_{t+1}, d^*_{t+1}\right\}$ (as shown in equation (14)), managers in the multi-firm model face uncertainty and must form expectations about the two equilibrium choices. This is because, due to the interdependence of detection and manipulation choices across firms, the manager at firm i rationally anticipates that the pair of the future manipulation and detection choices $\left\{m^*_{it+1}, d^*_{it+1}\right\}$ are also functions of the future cumulative levels of fraud at the other firms, $\{\Phi_{it}\}$. However, at the time of choosing m_{it} in period t, the value of Φ_{it} is random as it depends on whether the regulator detects fraud in the other firms later in period t.

We solve the remaining parts of the three-firm model numerically using the same parameter values set for the one-firm model. Based on the numeric solution, we first analyze the regulator's detection decisions. In the three-firm model, the detection intensity imposed by the regulator on a given firm depends on not only the firm's own information uncertainty from cumulative fraud but also how it compares to information uncertainty about the other two firms in the economy. To facilitate our analysis below, we present the model solution for a special case when $\Phi_2 = \Phi_3$. That is, we exemplify our model predictions by analyzing the detection intensity on different firms assuming that firm 2 and 3 have the same level of information uncertainty from cumulative fraud. It is easy to verify that, by model symmetry, the detection intensity on firm 2 and 3 is identical in this case, that is, $d_2 = d_3$.

Figure 4 illustrates the model solution for d_1 and d_2 (d_3) in heatmaps. Specifically, the x-axis represents the information uncertainty for firm 2 and 3, which is assumed to be identical in this example (i.e., $\Phi_2 = \Phi_3$). The y-axis represents the information uncertainty for firm 1 (i.e., Φ_1). The depth of color indicates the detection intensity, with light color representing a higher intensity of detection. The scale bar on the side maps the depth of color to the numerical value of detection intensity. The left (right) panel shows the detection intensity for firm 1 (firm 2 and 3) as a function of the three state variables, Φ_1, Φ_2, and Φ_3.

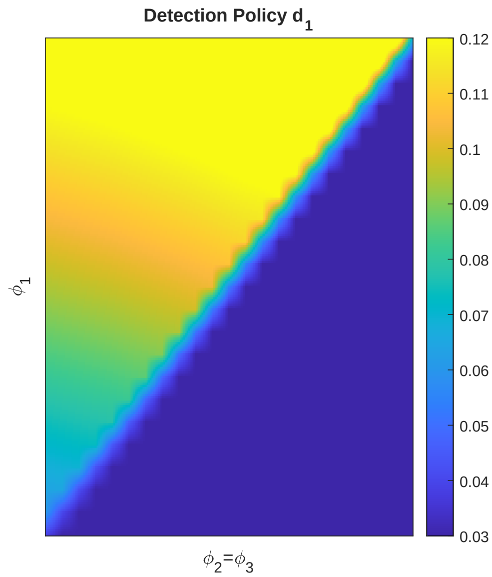

Figure 4: Equilibrium detection probability d_i^* in the three-firm setting.

Three interesting observations emerge. First, the regulator focuses on firm 1 when its cumulative fraud is high and its information uncertainty stands out among the three firms. Specifically, in the northwest corner where $\Phi 1 >> \Phi_2 = \Phi_3$, the regulator invests almost all resources in detecting fraud at firm 1, leaving firm 2 and 3 under the radar. Vice versa, in the southeast corner where firm 2 and 3 both accumulate much fraud and leave firm 1 behind (i.e., $\Phi_2 = \Phi_3 >> \Phi_1$), we observe more regulatory resources directed towards firm 2 and 3 and little regulatory attention is given to firm 1.

Second, the two scenarios are not entirely symmetric as the detection intensity imposed on firm 1 (about 0.12) in the first scenario is much larger than the detection intensity imposed on firm 2 and 3 (about 0.08, respectively) in the second scenario. This is because the regulator's cost of detection is convex in the aggregate detection intensity, as shown in equation (25), and thus the MC of detecting one firm also depends on whether other firms in the economy require close scrutiny. The model implies that detection is the most costly if fraud tends to cluster across firms (i.e., fraud wave), a feature that we will study later in the paper.

Lastly, we observe that when the three firms' information uncertainty converges along the 45-degree line (i.e., $\Phi_1 = \Phi_2 = \Phi_3$), the regulator has to split the detection resource equally among them, which implies $d_1 = d_2 = d_3$.

It is noteworthy that, even though we illustrate the model-implied detection policy above using a special case with $\Phi_2 = \Phi_3$, the intuition is the same in more general cases when the three firms have different levels of information uncertainty from cumulative fraud.[10]

Given the regulator's detection policy discussed above, equation (26) suggests that managers' manipulation decisions are also interdependent in our model. Intuitively, if one firm stands out in its cumulative fraud, it should expect close scrutiny from the regulator and so the MC of further committing fraud likely outweighs the MB, leading the manager to be more conservative. Ironically, as the firm with the highest information uncertainty attracts the most attention by the regulator, other firms are subject to less scrutiny and can afford to become more aggressive in committing fraud. To the extent that manipulation in each period accumulates and adds to the firms' information uncertainty over time, our model predicts an unintended consequence of regulation: it synchronizes managers' manipulation decisions and may eventually lead to fraud waves even in the absence of systematic shocks in the economy.

We next use the model to study the dynamics of the fraud-detection game between the regulator and three firms with different levels of fraud-induced information uncertainty in the initial period. Without loss of generality, we assume that $\Phi_H > \Phi_M > \Phi_L$ at $t = 1$ and denote the three firms H-, M- and L-firm, respectively. We then simulate the magnitude of manipulation committed by each manager m_i, the regulator's detection policy on each firm d_i, and the realization of detection outcomes at the end of each period. As we simulate the model forward, it generates the time series of Φ_{it}, d^*_{it}, and m^*_{it}. Figure 5 plots the three variables over the simulation path.

Starting with L-firm (depicted by the red-dash line), because the regulator anticipates a low level of cumulative fraud in the firm (i.e., a low Φ_L in Panel A), she spends little on detection (i.e., a low d_L in Panel B). The firm manager thus continues to commit fraud (i.e., increasing m_L in Panel C) because the MC is low and fraud starts building up (i.e., increasing Φ_L in Panel A). The first ten periods of the red dash line in Figure 5 illustrate this stage.

[10]For illustration purposes, we solve the model in closed form in a special case with the discounting factor $\delta = 0$, and the equilibrium solution is indeed consistent with the numerical results shown in Figure 4. The detailed analysis is in Appendix II.

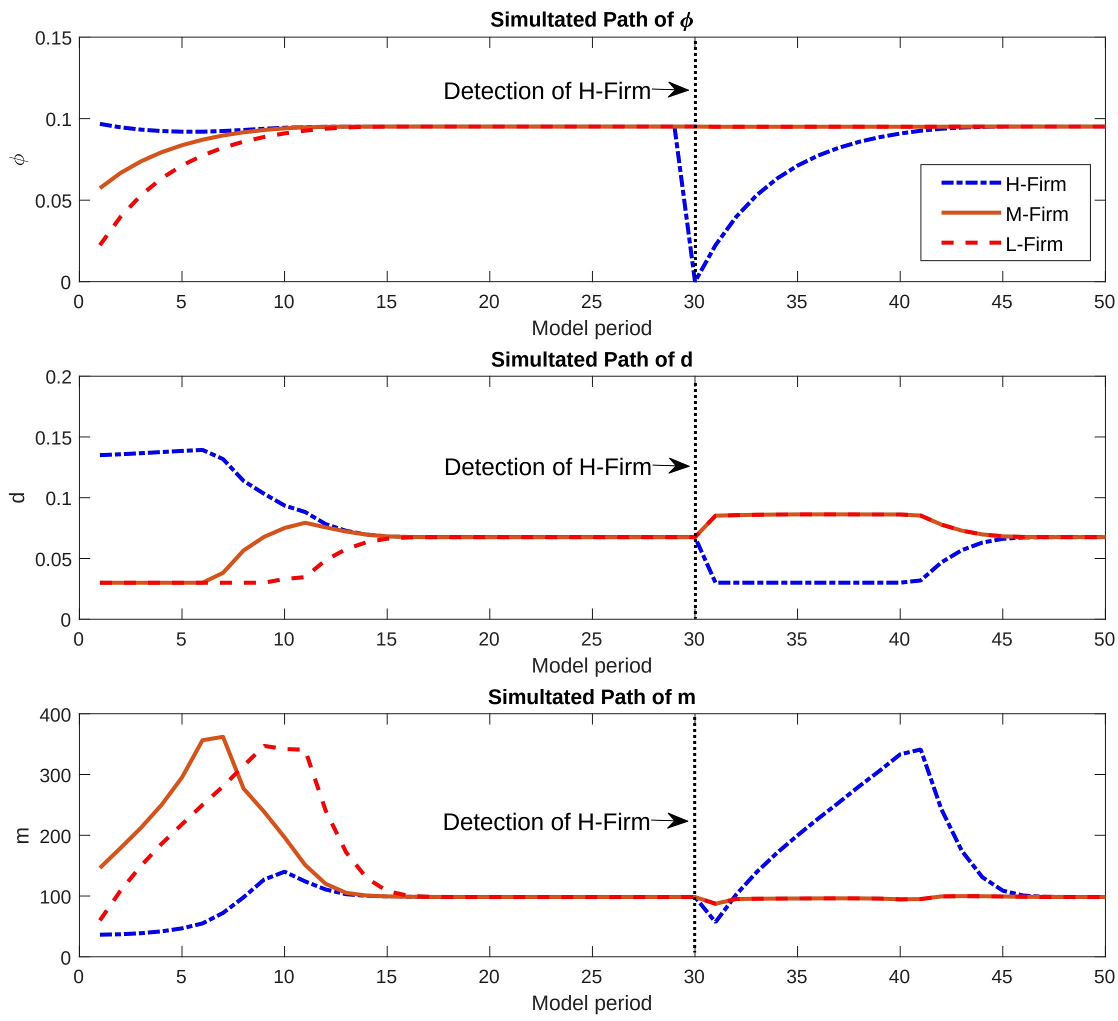

Figure 5: Simulated paths of cumulative fraud Φ, manipulation m^*, and detection intensity d^*.

M-firm (depicted by the brown-solid line) starts with an intermediate level of cumulative fraud. On the one hand, the manager of M-firm has greater incentives to commit fraud than the manager of L-firm, because a higher Φ increases the MB of committing fraud. On the other hand, the regulator invests more heavily in fraud detection of M-firm than L-firm, which suggests a higher MC of committing fraud. The two effects go hand-in-hand. The first five periods of the brown-solid line in Figure 5 show the stage when MB dominates MC, and thus m_M increases over time as Φ_M grows. After the sixth period, we observe that the detection intensity on M-firm quickly rises (see the brown-solid line in Panel B) and MC outweighs MB, leading to a sharp decline in manipulation by M-firm (see the brown-solid line in Panel C). The dynamics in m_M therefore demonstrates the counteracting forces of MB and MC.

Last, H-firm (depicted by the blue-dot line) starts with the highest level of cumulative fraud. Accordingly, it is under the closest scrutiny by the regulator. The regulator concentrates on detecting H-firm in the first five periods until the cumulative fraud of M-firm (and L-firm) catches up and gets close to that of H-firm after the 6th (11th) period, after which the detection intensity of H-firm and M-firm (and L-firm) starts converging. The blue-dot line depicts the trajectory of H-firm's Φ_H, m_H, and d_H in three panels, respectively.

To examine the impact of actual detection, we assume in the simulation trial that H-firm is caught by the regulator at period 30. Upon detection, H-firm's cumulative fraud is cleared and Φ_H drops to zero immediately, as shown in Panel A. As an optimal response, the regulator shifts attention from H-firm to the original M- and L-firms, as shown in Panel B. Interestingly, as the detection intensity on H-firm drops substantially, H-firm faces a low MC of committing fraud and can now afford to become more aggressive in manipulating its report. This explains the sharp increase in m_H and Φ_H in Panel C and A right after period 31. If M- and L-firms remain undetected, cumulative fraud in the three firms will be synchronized again after another few periods. This analysis sheds further light on an unintended consequence of regulation: it may synchronize firms' manipulation decisions and lead to fraud waves even in the absence of aggregate shocks. The intuition is simple: anticipating the optimal allocation of regulatory resources in the economy, firms with a low level of cumulative fraud endogenously choose a high level of manipulation, allowing them to catch up to more fraudulent firms.

4 Data and Sample

This section describes the sample, variables used in our empirical analyses, and data sources used to construct these variables. Detailed variable definitions are provided in Appendix III.

4.1 Sample Selection

We obtain the initial sample of 18,340 accounting restatements from Audit Analytics. These restatements, announced by 10,404 unique firms between 1995Q1 and 2019Q3, cover 105,088 firm-quarters between 1983Q1 and 2019Q2 based on misstating periods. Because the coverage of the Audit Analytics restatement database is relatively narrow before 1999, we focus on the time period starting from 1999Q1. We merge the restating quarters into the universe of Compustat-CRSP. We then obtain implied volatility data from Option Metrics and analyst forecast data from IBES. The final sample, spanning from 1999Q1 to 2017Q4, represents an intersection of the databases that we use. The number of firm-quarter observations used in our main analyses ranges between 134,566 and 151,048.

4.2 Measurement of Information Uncertainty and Detection

As discussed in Section 2, our model analyses center on the interdependence of Φ, the fraud-induced information uncertainty in each period, and d, the fraud detection likelihood in each period. To measure information uncertainty, we extract the implied volatility from options. Since option prices reflect the market's expectations about changes in the firm's value given all available information, implied volatility captures the conditional variance of this information set, which increases with the information uncertainty brought by cumulative fraud. While options typically expire on the third Friday of the contract month, firms make their earnings announcements at various times. Thus, the time between each firm's earnings announcement and its option expiration date differs. To minimize measurement error that may arise because of this non-constant maturity, we use the implied volatility from 90-day standardized option prices provided by Option Metrics. Specifically, we first take the mean of the 90-day call- and put-implied volatility to capture the market's uncertainty about the firm's economic earnings.

We then construct quarterly implied volatility by taking the mean of daily implied values. We denote the variable IV.

To operationalize the second parameter—detection likelihood—we code $DETECT$ as an indicator variable that equals one if an earnings restatement likely to be fraudulent is announced in a quarter, and zero otherwise. One notable finding in the accounting literature is that not all restatements are related to fraud; some are unintentional misapplications of accounting rules and have little effect on stock prices (Hennes et al. (2008); Fang et al. (2017)). We define fraud-related restatements as those that meet at least one of the three following conditions: (1) if the restatement is marked as being fraudulent by Audit Analytics; (2) if the restatement has received a class-action lawsuit as recorded in Audit Analytics; or (3) if the cumulative restated amount (scaled by the total assets as of the last restating period) is in the top decile of the sample. Restatement announcements are usually made through SEC filings or press releases. This proxy builds on the idea that the unconditional probability of fraud getting caught ex post is higher given a higher detection likelihood ex ante.

4.3 Analyst Forecast and Control Variables

We use analyst consensus earnings forecast as a proxy for earnings expectation. To measure how earnings expectation changes in response to reported earnings, we first define $REVISION$ as the difference of one-quarter-ahead earnings forecast issued before and after the earnings announcement. We define earning surprise, SUE, as the difference between reported earnings and the pre-announcement consensus forecast. The $REVISION$-to-SUE sensitivity thus captures how market updates its expectation in response to reported earnings and we expand on the discussion of this sensitivity measure in Section 5.

For controls, we follow prior literature and include four controls previously shown to affect a firm's level of earnings manipulation (e.g., Kothari et al. (2005); Zang (2012)), namely, the natural logarithm of total assets ($SIZE$), market-to-book (MB), return on assets (ROA), and leverage (LEV). Among the four controls, $SIZE$ and MB also help control for firm growth. This is important because prior studies show that growth affects firms' incentives to manipulate earnings (e.g., Povel et al. (2007); Wang et al. (2010); Strobl (2013); Wang and

Winton (2014)). We further include $REVGWTH$, the percentage change of sales from the same quarter of the last year, as an additional control for growth. Firm financials are from the Compustat quarterly files.

4.4 Descriptive Statistics

Table 1 reports the descriptive statistics of the variables used in our analyses. IV has a mean of 0.461, a median of 0.406, and a standard deviation of 0.225. The mean of *Detect* is 0.008, which suggests that, on average, a firm in our sample has a 0.8% likelihood to have at least one fraud-related restatement announced in each quarter.

5 Empirical Analyses

5.1 Information Uncertainty and Analyst Revision

Our model predicts that the MB of committing fraud is positively associated with the firm's cumulative fraud to date, because a higher level of cumulative fraud increases information uncertainty about the firm both in the current period and in future periods, which in turn boosts the value of the new earnings report and potential return from inflating the report.

To test this prediction, we examine the relation between analysts' revision of earnings estimates for the next quarter following earnings announcement of the current quarter and implied volatility immediately prior to the current quarter by estimating the following regression:

$$REVISION_{i,q} = \alpha + \beta_1 SUE_{i,q} \times IV_{i,q} + \beta_3 SUE_{i,q} + \beta_4 IV_{i,q} + \beta_c CONTROLS_{i,q-1}, \quad (31)$$

where subscript i indexes firms and q indexes fiscal quarters. US companies are required to report earnings no later than 45 days after the end of a fiscal quarter and analysts can continue to revise their estimates until the day of earnings announcement. The dependent variable, $REVISION$, thus measures the change in the analyst consensus earnings per share (EPS) forecast for firm i's quarter q, between earnings announcement for quarter $q-1$ (made in quarter q) and that for quarter q (made in quarter $q+1$). Among the regressors, SUE represents standardized unexpected earnings of firm i-quarter $q-1$ announced in quarter q. Unexpected earnings are defined as the difference between the firm's reported EPS and its analyst consensus EPS forecast two days prior to earnings announcement, scaled by stock price two days prior to earnings announcement. As discussed in Section 4.2, IV intends to capture the degree of information uncertainty about firm i brought by cumulative fraud, taken ten trading days before earnings announcement for quarter $q-1$ in quarter q. The interaction term between SUE and IV captures the extent to which implied volatility affects the sensitivity of analyst forecast revision to unexpected earnings. We include year-quarter fixed effects, and cluster standard errors by firm and quarter.

Table 2 column (1) reports the regression results of estimating equation (31). The coefficient of interest β_1 on $SUE \times IV$ is positive and significant at the 1% level, which indicates that analysts are more responsive to the firm's unexpected earnings of the current quarter in their revision of earnings estimates for the next quarter, when the implied volatility of the firm prior to the announcement is higher. This is consistent with our model prediction that the market is more likely to value the reported earnings, particularly the portion that differs from the market's expectations, when information uncertainty is greater because of a higher level of cumulative fraud.

In Table 2 column (2), we reestimate equation (31) including several controls. NEG is an indicator variable denoting whether the reported earnings of firm i-quarter $q-1$ are negative. The interaction term between NEG and IV captures the asymmetric reaction to positive versus negative earnings that analysts may exhibit in their forecast revision. Other controls, measured for firm i-quarter q, include the natural logarithm of total assets ($SIZE$), market-to-book (MB), return on assets (ROA), leverage (LEV), and seasonally adjusted sales growth ($REVGWTH$). In Table 2 column (3), we further include firm fixed effects. The coefficient of interest, β_1, remains positive and significant at the 1% level, in both columns. Again, this result suggests that the MB of committing fraud is larger when the information uncertainty about the firm is higher because unexpected earnings elicit more responsive analyst forecast revision.

Among the controls, the coefficient on $SUE \times NEG$, is negative and significant at the 1% level, which indicates that analyst forecast revision is less responsive to the firm's unexpected earnings when reported earnings are negative. SUE in itself is positively related to $REVISION$, as expected, while IV and NEG are negatively related to $REVISION$. Finally, analyst forecast revision tends to be more positive for firms with stronger growth, but less positive for firms with higher leverage.

5.2 Information Uncertainty and Detection Likelihood

A core prediction from our model is that the strength of detection optimally matches the severity of fraud. To test this prediction, we examine the relation between the likelihood of having fraud revealed in a given quarter and implied volatility immediately prior to the quarter by estimating the following regression:

$$DETECT_{i,q+1} = \alpha + \beta_1 IV_{i,q} + \beta_c CONTROLS_{i,q-1}. \tag{32}$$

The dependent variable, $DETECT$, is an indicator variable that denotes whether a fraud-related accounting restatement is announced for firm i in a given quarter $q+1$. IV is the average daily implied volatility of quarter q. We continue to include year-quarter fixed effects, and cluster standard errors by firm and quarter.

Table 3 column (1) reports the regression results of estimating equation (32). The coefficient of interest, β_1, is positive and significant at the 1% level, supporting the model prediction that fraud detection likelihood is larger when a firm has accumulated a higher level of fraud. In columns (2) and (3), we reestimate equation (32) including five basic firm characteristics. The coefficient of interest, β_1, remains positive and significant at the 1% level in column (2) excluding firm fixed effects and in column (3) including firm fixed effects, respectively. This result suggests that the manager's MC of committing fraud is larger when the information uncertainty about the firm (partly brought by cumulative fraud) is greater because detection likelihood is higher. It is also consistent with the regulator's MB of detecting fraud being larger when a firm's fraud-induced information uncertainty is greater, which would lead the regulator to rationally allocate more resources towards the firm.

A potential concern of the detection indicator is that not all firms in the sample are covered by Audit Analytics so measurement error is likely greater for firms with no recorded restatements in the database. To address this concern, in Table 3 column (4), we focus on a subsample of firms with at least one restatement announcement tracked by Audit Analytics. For each firm, we include the entire time series of quarterly observations during the sample period. Results using this subsample remain similar.

5.3 Convergence of Fraud

One interesting implication from the analyses of our multi-firm model is that anti-fraud regulations are unlikely to eradicate fraud but may synchronize firms' fraud decisions. This is because, while the optimal allocation of regulatory resources towards the more fraudulent firms has a disciplinary effect on these firms, it implies less scrutiny of less fraudulent firms, allowing their fraudulent behavior to go undetected and level of fraud to catch up. As such, firms converge towards each other in their level of fraud.

To study the possible convergence of fraud across firms over time, we sort firm-quarters in the sample into quintiles based on firms' level of implied volatility of prior quarter, and then estimate the following regression:

$$\Delta IV_{i,q\,to\,q+1} = \alpha + \beta_1 IVQ1_{i,q} + \beta_2 IVQ2_{i,q} + \beta_3 IVQ4_{i,q} + \beta_4 IVQ5_{i,q} + \beta_c CONTROLS_{i,q-1}, \tag{33}$$

ΔIV measures the change in the firm's average daily implied volatility from quarter q to $q+1$. $IVQn$ is an indicator variable that denotes whether a firm-quarter falls into the nth-ranked quintile (n=1 to 5), with a higher-ranked quintile representing the subsample with a higher level of average daily implied volatility in quarter q. We omit $IVQ3$ from the regression to avoid multicollinearity so the middle quintile serves as the benchmark group. We include basic controls and year-quarter fixed effects, and cluster standard errors by firm and quarter.

Table 4 columns (1) and (2) report the regression results of estimating equation (33), without and with firm fixed effects. Compared with those in the middle quintile ($IVQ3$=1), firms in a lower-ranked quintile of implied volatility prior to a quarter tend to have a larger increase in implied volatility during the quarter, as evidenced by a positive coefficient estimate on $IVQ2$ and an even larger one on $IVQ1$. Also benchmarked against the middle quintile, firms in a higher-ranked quintile of implied volatility prior to a quarter tend to have a smaller increase in implied volatility during the quarter, as evidenced by a negative coefficient estimate on $IVQ4$ and an even more negative one on $IVQ5$. This finding sheds light on the convergence of corporate fraud across firms over time.

One concern is that this finding merely reflects the mean-reverting nature of IV. To address the concern, we augment equation (33) by further including the interaction terms between $IVQn$ (n=1, 2, 4, and 5) and $WAVE$, an indicator denoting whether a firm-quarter overlaps with a fraud wave in the firm's industry. To define $WAVE$, we first compute $FRAUD\%$, the percentage of firms with restatement announcement in an industry-quarter. We code $WAVE$ as one if the actual $FRAUD\%_{j,q}$ for industry j–quarter q exceeds the 90th percentile of its sample distribution and zero otherwise. The industry classification is based on the Global Industry Classification Standard (GICS) 4-digit industry groups.

Table 4 column (3) reports the regression results of estimating the augmented equation, including firm fixed effects. As in columns (1)-(2), firms in a higher-ranked quintile (i.e., those having a higher level of implied volatility prior to a quarter) have a smaller increase in implied volatility during the quarter, as evidenced by the positive coefficient estimates on $IVQ1$ and $IVQ2$ and the negative coefficient estimates on $IVQ4$ and $IVQ5$. This pattern is more pronounced when a firm-quarter overlaps with a fraud wave in the firm's industry, as evidenced by the positive coefficient estimates on the interaction term between $WAVE$ and $IVQ1$ and that between $WAVE$ and $IVQ2$ and the negative coefficient estimates on the interaction term between WAVE and $IVQ4$ and that between $WAVE$ and $IVQ5$. This finding suggests that the negative relation between prior level of implied volatility (as measured by quintile rank) and the increase in implied volatility in a quarter is not merely reflective of the mean-reverting nature of corporate fraud, or it should not be affected by the existence of an industry-level fraud wave. Rather, this finding is more consistent with the convergence in firms' level of fraud over time.

6 Conclusion

Throughout history, developed and emerging financial markets alike have been booming, crashing, and recovering their way through a wide range of corporate frauds. With the fallout of every major financial scandal comes the public outcry for regulations and reforms to crack down on fraud. This paper aims to lay out a theoretical foundation to better understand the formation and evolvement of accounting fraud, which would then allow for an assessment of anti-fraud regulations.

We first build a dynamic model featuring a representative firm and a regulator. Analyses of this single-firm model show that fraud is unlikely to go extinct, as long as uncovering fraud consumes regulatory resources and such resources are finite. With the regulator rationally directing resources towards the most fraudulent firms, an increasing level of fraud accumulated in the firm attracts scrutiny, but at the same time generates information uncertainty, which gives further incentives to commit fraud. These two effects go hand-in-hand, counteracting each other. As such, the amount of fraud committed in the firm may exhibit repeated cycles of rise, peak, fall, and collapse (upon detection). We present two pieces of evidence in support of these model predictions. First, using implied volatility to capture fraud-induced information uncertainty, we find that analyst forecast revision is more responsive to unexpected earnings when implied volatility is higher. This result explains why a high level of cumulative fraud may further elevate the MB of committing fraud. Second, we find that a firm is more likely to be caught for having committed fraud in the past when implied volatility is higher. This result supports the model prediction that the strength of detection matches the severity of fraud, and explains why a high level of cumulative fraud may also increase the MC of committing fraud.

We then expand the model to consider a regulator and three firms with a high, medium, and low level of cumulative fraud, respectively. Analyses of this multi-firm model offer additional insights. Anti-fraud regulations can be highly effective at lowering the most fraudulent firms' incentives to continue fraud, by not only raising their MC of committing fraud but also sharply decreasing their MB of committing fraud upon detection. However, the rational allocation of regulatory resources towards such firms may imply less scrutiny of less fraudulent firms, allowing the latter's fraudulent behavior to go undetected and their level of fraud to catch up. As such, despite the pro tem "cracking-down," anti-fraud regulations do not eradicate fraud. Rather, they synchronize firms' idiosyncratic fraud decisions and induce corporate fraud waves over time. As supportive evidence of this insight, we show that firms with a higher level of implied volatility prior to a period have a smaller increase in implied volatility during the period. Further, we show that this association is unlikely to be explained by the mean-reverting nature of fraud.

Although consistent with the model predictions, our results are no definitive evidence because the theoretical constructs are abstract and measurement of these constructs is admittedly imperfect. Thus, our inferences are subject to caveats. More research on the joint mechanisms of fraud and regulation is warranted, particularly if better empirical proxies for fraud and detection likelihood become available.

References

Beyer, Anne, Ilan Guttman, and Iván Marinovic, 2019, Earnings management and earnings quality: Theory and evidence, *The Accounting Review* 94, 77–101.

Black, Fischer, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Bond, Philip, and Kathleen Hagerty, 2010, Preventing crime waves, *American Economic Journal: Microeconomics* 2, 138–59.

Dye, Ronald A, and Sri S Sridhar, 2004, Reliability-relevance trade-offs and the efficiency of aggregation, *Journal of Accounting Research* 42, 51–88.

Ewert, Ralf, and Alfred Wagenhofer, 2012, Earnings management, conservatism, and earnings quality, *Foundations and Trends in Accounting* 6, 65–186.

Fang, Vivian W, Allen H Huang, and Wenyu Wang, 2017, Imperfect accounting and reporting bias, *Journal of Accounting Research* 55, 919–962.

Fischer, Paul E, and Robert E Verrecchia, 2000, Reporting bias, *The Accounting Review* 75, 229–245.

Hail, Luzi, Ahmed Tahoun, and Clare Wang, 2018, Corporate scandals and regulation, *Journal of Accounting Research* 56, 617–671.

Hennes, Karen M, Andrew J Leone, and Brian P Miller, 2008, The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover, *The Accounting Review* 83, 1487–1519.

Kothari, Sagar P, Andrew J Leone, and Charles E Wasley, 2005, Performance matched discretionary accrual measures, *Journal of Accounting and Economics* 39, 163–197.

Mookherjee, Dilip, and Ivan PL Png, 1992, Monitoring vis-a-vis investigation in enforcement of law, *The American Economic Review* 556–565.

Povel, Paul, Rajdeep Singh, and Andrew Winton, 2007, Booms, busts, and fraud, *The Review of Financial Studies* 20, 1219–1254.

Stocken, Phillip C, 2013, Strategic accounting disclosure, *Foundations and Trends in Accounting* 7, 197–291.

Strobl, Günter, 2013, Earnings manipulation and the cost of capital, *Journal of Accounting Research* 51, 449–473.

Wang, Tracy Yue, and Andrew Winton, 2014, Product market interactions and corporate fraud, *Working Paper* .

Wang, Tracy Yue, Andrew Winton, and Xiaoyun Yu, 2010, Corporate fraud and business conditions: Evidence from IPOs, *The Journal of Finance* 65, 2255–2292.

Zang, Amy Y, 2012, Evidence on the trade-off between real activities manipulation and accrual-based earnings management, *The Accounting Review* 87, 675–703.

Table 1: Summary Statistics

This table reports summary statistics of the variables used in the analysis. IV is the quarterly average of the daily implied volatility. $DETECT$ is an indicator that denotes whether a firm discloses an accounting restatement that meets at least one of the three conditions: (1) if the restatement is marked as being fraudulent by Audit Analytics; (2) if the restatement has received a class-action lawsuit as tracked by Audit Analytics; or (3) if the cumulative restated amount (scaled by the total assets as of the last restating period) is in the top decile of the sample. $REVISION$ is the change in the analyst consensus EPS forecast for the current quarter surrounding the earnings announcement of the previous quarter. SUE is the earnings surprise of the previous quarter. NEG is an indicator that denotes negative earnings of the previous quarter. $SIZE$ is the natural logarithm of total assets. MB is the market-to-book ratio. LEV is the leverage ratio. ROA is the return on assets. $REVGWTH$ is the sales growth from the same quarter last year. Detailed variable definitions are in Appendix III.

Variables	N	Mean	Std. Dev.	25 Pctl	50 Pctl	75 Pctl
IV	151,048	0.461	0.225	0.296	0.406	0.569
Restatement Variables						
DETECT	151,048	0.008	0.090	0.000	0.000	0.000
Earnings and Forecast Variables						
REVISION	142,054	-0.222%	0.889%	-0.201%	-0.029%	0.040%
SUE	142,054	0.017%	1.072%	-0.033%	0.042%	0.200%
NEG	142,054	0.186	0.389	0.000	0.000	0.000
Firm Characteristics						
SIZE	151,048	7.407	1.826	6.053	7.301	8.616
MB	151,048	1.878	1.743	0.862	1.322	2.225
LEV	151,048	0.223	0.208	0.030	0.189	0.346
ROA	149,108	0.014	0.044	0.005	0.018	0.034
REVGWTH	144,479	0.149	0.426	-0.023	0.075	0.210

Table 2: Analyst Earnings Forecast Revision and Implied Volatility
This table reports the ordinary least squares (OLS) regression results estimating the relation between analyst earnings forecast revision and implied volatility. $REVISION$ is the change in the analyst consensus EPS forecast for the current quarter surrounding the earnings announcement of the previous quarter. IV is the implied volatility ten trading days before earnings announcement. SUE is the earnings surprise of the previous quarter. NEG is an indicator that denotes negative earnings of the previous quarter. Other controls are described in Table 1. Detailed variable definitions are in Appendix III. Columns (1) and (2) include year-quarter fixed effects, and column (3) further includes firm fixed effects. Standard errors are clustered by year-quarter and firm. T-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

Variables	(1) REVISION	(2) REVISION	(3) REVISION
SUE	0.210***	0.163***	0.173***
	(11.34)	(7.76)	(9.36)
IV	-0.008***	-0.008***	-0.005***
	(-18.88)	(-13.60)	(-8.29)
IV×SUE	0.088***	0.106***	0.086***
	(4.18)	(4.78)	(4.25)
NEG		-0.002***	-0.001***
		(-9.51)	(-5.82)
NEG×SUE		-0.001***	-0.001***
		(-7.97)	(-10.33)
SIZE		0.000	-0.001***
		(1.65)	(-6.44)
MB		0.001***	0.000***
		(15.12)	(10.34)
LEV		-0.001***	-0.001
		(-3.16)	(-1.48)
ROA		-0.004*	0.002
		(-1.87)	(0.65)
REVGWTH		0.001***	0.001***
		(9.57)	(7.81)
Observations	142,054	134,873	134,566
Adjusted R-squared	0.175	0.196	0.333
Firm FE	No	No	Yes
Year-Quarter FE	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes

Table 3: Fraud Detection and Implied Volatility

This table reports the OLS regression results estimating the relation between fraud detection likelihood and implied volatility. $DETECT$ is an indicator that denotes whether a firm discloses an accounting restatement that meets at least one of the three conditions: (1) if the restatement is marked as being fraudulent by Audit Analytics; (2) if the restatement has received a class-action lawsuit as tracked by Audit Analytics; or (3) if the cumulative restated amount (scaled by the total assets as of the last restating period) is in the top decile of the sample. IV is the quarterly average of the daily implied volatility, measured in the quarter before $DETECT$. Controls are described in Table 1. Detailed variable definitions are in Appendix III. Columns (1) and (2) include year-quarter fixed effects, and columns (3) and (4) further include firm fixed effects. Column (1)-(3) include the full sample, and column (4) only includes firms with at least one detected restatement from Audit Analytics. Standard errors are clustered by year-quarter and firm. T-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

	(1)	(2)	(3)	(4)
Sample	Full	Full	Full	Detected Firms
Variables	DETECT	DETECT	DETECT	DETECT
IV	0.010***	0.014***	0.007***	0.012***
	(6.16)	(6.41)	(2.74)	(2.96)
SIZE		0.001***	0.004***	0.006***
		(3.22)	(4.01)	(4.11)
MB		-0.000	-0.000	-0.000
		(-0.94)	(-0.14)	(-0.13)
LEV		0.003*	0.003	0.004
		(1.93)	(1.21)	(1.02)
ROA		-0.005	-0.042***	-0.068***
		(-0.66)	(-3.84)	(-3.82)
REVGWTH		0.002*	0.001	0.002
		(1.81)	(1.40)	(1.45)
Observations	151,048	143,252	143,034	86,738
Adjusted R-squared	0.003	0.004	0.022	0.024
Firm FE	No	No	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes	Yes

Table 4: Convergence of Implied Volatility

This table the OLS regression results estimating the relation between the change in implied volatility and the previous level of implied volatility. ΔIV is the change in implied volatility from quarter q to quarter $q+1$. $IVQn$ is an indicator variable that denotes whether a firm-quarter falls into the nth-ranked quintile of IV (n=1 to 5) in quarter q, with quintile five having the highest level of implied volatility. $WAVE$ is an indicator variable that denotes a fraud wave in the firm's industry overlapping quarter q. Controls are described in Table 1. Detailed variable definitions are in Appendix III. Column (1) includes year-quarter fixed effects, and columns (2) and (3) further include firm fixed effects. Standard errors are clustered by year-quarter and firm. T-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

	(1)	(2)	(3)
Variables		$\Delta \mathrm{IV}_{q\ to\ q+1}$	
IVQ1	0.006*	0.019***	0.018***
	(1.90)	(6.93)	(5.90)
IVQ2	0.003*	0.009***	0.008***
	(1.89)	(5.86)	(5.08)
IVQ4	-0.005**	-0.011***	-0.010***
	(-2.20)	(-5.07)	(-4.51)
IVQ5	-0.035***	-0.055***	-0.054***
	(-7.86)	(-11.75)	(-11.25)
WAVE×IVQ1			0.007***
			(2.80)
WAVE×IVQ2			0.003**
			(2.09)
WAVE×IVQ4			-0.007***
			(-2.85)
WAVE×IVQ5			-0.011**
			(-2.35)
WAVE			-0.002
			(-1.00)
SIZE	-0.001	0.005***	0.005**
	(-1.21)	(2.65)	(2.64)
MB	0.001**	0.004***	0.004***
	(2.25)	(3.69)	(3.69)
LEV	0.006***	0.006*	0.006*
	(2.88)	(1.77)	(1.90)
ROA	-0.102***	-0.066***	-0.067***
	(-7.04)	(-3.73)	(-3.76)
REVGWTH	0.003***	0.001	0.001
	(2.72)	(1.07)	(1.08)
Observations	149,665	149,420	149,420
Adjusted R-squared	0.376	0.394	0.395
Firm Fixed Effects	No	Yes	Yes
Year-Quarter Fixed Effects	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes

Appendix I: proofs

Proof. of Lemma 1: Note that (13) can be simplified into

$$\begin{aligned}
&\frac{(1-\delta\rho)\,cd_t^*}{1-d_t^*}\left(1+\frac{m_t^* q\,(1-q)}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}\right)\\
&=1+\delta\rho\left(1-d_{t+1}^*\right)\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q\,(1-q)}\\
&+\delta^2\rho^2\left(1-d_{t+1}^*\right)\left(1-d_{t+2}^*\right)\frac{m_{t+2}^* q\,(1-q)}{\rho^2\Phi_{t+1}+\sigma_\varepsilon^2+m_{t+2}^* q\,(1-q)}\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q\,(1-q)}\\
&+...
\end{aligned}\tag{34}$$

By induction, in period $t+1$, conditional on that the regulator fails to detect fraud in period t, the manager chooses m_{t+1}^* that satisfies:

$$\begin{aligned}
&\frac{(1-\delta\rho)\,cd_{t+1}^*}{1-d_{t+1}^*}\left(1+\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2}\right)\\
&=1+\delta\rho\left(1-d_{t+2}^*\right)\frac{m_{t+2}^* q\,(1-q)}{\rho^2\Phi_{t+1}+\sigma_\varepsilon^2+m_{t+2}^* q\,(1-q)}\\
&+\delta^2\rho^2\left(1-d_{t+2}^*\right)\left(1-d_{t+3}^*\right)\frac{m_{t+3}^* q\,(1-q)}{\rho^2\Phi_{t+2}+\sigma_\varepsilon^2+m_{t+3}^* q\,(1-q)}\frac{m_{t+2}^* q\,(1-q)}{\rho^2\Phi_{t+1}+\sigma_\varepsilon^2+m_{t+2}^* q\,(1-q)}\\
&+...
\end{aligned}\tag{35}$$

Multiplying both sides of equation (35) by $\delta\rho\left(1-d_{t+1}^*\right)\frac{m_{t+1}^* q(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q(1-q)}$ yields:

$$\begin{aligned}
&\delta\rho\,(1-\delta\rho)\,cd_{t+1}^*\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2}\\
&=\delta\rho\left(1-d_{t+1}^*\right)\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q\,(1-q)}\\
&+\delta^2\rho^2\left(1-d_{t+2}^*\right)\left(1-d_{t+1}^*\right)\frac{m_{t+2}^* q\,(1-q)}{\rho^2\Phi_{t+1}+\sigma_\varepsilon^2+m_{t+2}^* q\,(1-q)}\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q\,(1-q)}\\
&+\delta^3\rho^3\left(1-d_{t+1}^*\right)\left(1-d_{t+2}^*\right)\left(1-d_{t+3}^*\right)\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2+m_{t+1}^* q\,(1-q)}\\
&\times\frac{m_{t+3}^* q\,(1-q)}{\rho^2\Phi_{t+2}+\sigma_\varepsilon^2+m_{t+3}^* q\,(1-q)}\frac{m_{t+2}^* q\,(1-q)}{\rho^2\Phi_{t+1}+\sigma_\varepsilon^2+m_{t+2}^* q\,(1-q)}\\
&+...
\end{aligned}\tag{36}$$

Substituting (36) into (34) yields:

$$\frac{(1-\delta\rho)\, cd_t^*}{1-d_t^*}\left(1+\frac{m_t^* q\,(1-q)}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}\right)=1+\delta\rho\,(1-\delta\rho)\, cd_{t+1}^*\frac{m_{t+1}^* q\,(1-q)}{\rho^2\Phi_t+\sigma_\varepsilon^2}. \tag{37}$$

Solving (37) for m_t yields (14) in the lemma. ■

Proof. of Lemma 2: We only consider the case in which the detection fails:

$$\begin{aligned}
\Phi_t &= var\,(s_t|\mathcal{F}_{t-1}) - var\left(E^I\left[s_t|\mathcal{F}_t\right]|\mathcal{F}_{t-1}\right) \qquad (38)\\
&= var\,(\rho s_{t-1}+\varepsilon_t|\mathcal{F}_{t-1}) - var\left(E^I\left[s_t|r_t, r_{t-1}, ...\right]|\mathcal{F}_{t-1}\right)\\
&= \rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 - var\left(\frac{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2}{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2+m_t^* q\,(1-q)} r_t|\mathcal{F}_{t-1}\right)\\
&= \rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 - \left[\frac{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2}{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2+m_t^* q\,(1-q)}\right]^2 var\,(r_t|\mathcal{F}_{t-1})\\
&= \frac{m_t^* q\,(1-q)\left[\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2\right]}{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2+m_t^* q\,(1-q)}\\
&= \frac{m_t^* q\,(1-q)\left(\rho^2\Phi_{t-1}+\sigma_\varepsilon^2\right)}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^* q\,(1-q)}.
\end{aligned}$$

The first equality uses the law of total variance. The third equality uses

$$\begin{aligned}
E^I\left[s_t|\mathcal{F}_t\right] &= E^I\left[s_t|\mathcal{F}_{t-1}\right] + \frac{cov\,(r_t, s_t|\mathcal{F}_{t-1})}{var\,(r_t|\mathcal{F}_{t-1})}\left\{r_t - E\left[r_t|\mathcal{F}_{t-1}\right]\right\} \qquad (39)\\
&= E^I\left[s_t|\mathcal{F}_{t-1}\right] + \frac{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2}{\rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1})+\sigma_\varepsilon^2+m_t^* q\,(1-q)}\left\{r_t - E\left[r_t|\mathcal{F}_{t-1}\right]\right\},
\end{aligned}$$

where

$$\begin{aligned}
var\,(r_t|\mathcal{F}_{t-1}) &= var\,(s_t|\mathcal{F}_{t-1}) + m_t^* q\,(1-q) \qquad (40)\\
&= \rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 + m_t^* q\,(1-q),\\
cov\,(r_t, s_t|\mathcal{F}_{t-1}) &= var\,(s_t|\mathcal{F}_{t-1}) \qquad (41)\\
&= \rho^2 var\,(s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2.
\end{aligned}$$

The last step uses the definition of $\Phi_{t-1} \equiv var\,(s_{t-1}|\mathcal{F}_{t-1})$. ■

Proof. of lemma 3: Using the law of motion (15), we rewrite the regulator's payoff (17) recursively:

$$\begin{aligned} W_t\left(\Phi_{t-1}\right) &= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*,\Phi_{t-1}\right) - \frac{\kappa}{2}\left(d_t-d_0\right)^2 + \delta E^I\left[\sum_{k=t+1}^{\infty}\delta^{k-(t+1)}\left(-\left(1-d_k^*\right)\Phi_k - \frac{\kappa}{2}d_k^{*2}\right)\right] \\ &= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*,\Phi_{t-1}\right) - \frac{\kappa}{2}\left(d_t-d_0\right)^2 + \delta\left[d_t W_{t+1}(0) + \left(1-d_t\right)W_{t+1}\left(\Phi_t\left(d_t^*,\Phi_{t-1}\right)\right)\right], \end{aligned} \tag{42}$$

where $\Phi_t\left(d_t^*,\Phi_{t-1}\right)$ is given in (15). Note that the future cumulative level of fraud Φ_t depends on the equilibrium detection probability d_t^* and not on the actual detection probability d_t. This is because, the manager does not observe the regulator's detection choice at the time of choosing manipulation and his manipulation choice only depends on the equilibrium d_t^*. Taking the first-order condition of W_t with respect to d_t yields (18) in the main text. ■

Proof. of Proposition 1: See the main text. ■

Proof. of Proposition 2: We only derive the manipulation decision by manager 1 as the manipulation decisions by the other managers can be derived analogously.

Taking the first-order condition of m_{1t} gives that:

$$\begin{aligned} & c\left(1-q\right)d_{1t}^* \\ &= \frac{1-d_{1t}^*}{1-\delta\rho}\frac{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2+m_{1t}^*q\left(1-q\right)}\left(1-q\right) \\ &+ \frac{\left(1-d_{1t}^*\right)\delta\rho}{1-\delta\rho}\frac{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2+m_{1t}^*q\left(1-q\right)}\left(1-q\right) \\ &\times E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^*\right)\frac{m_{1t+1}^*q\left(1-q\right)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^*q\left(1-q\right)}\right] \\ &+ \frac{\left(1-d_{1t}^*\right)\delta^2\rho^2}{1-\delta\rho}\frac{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2+m_{1t}^*q\left(1-q\right)}\left(1-q\right) \\ &\times E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+1}^*\right)\left(1-d_{1t+2}^*\right)\frac{m_{1t+2}^*q\left(1-q\right)}{\rho^2\Phi_{1t+1}+\sigma_\varepsilon^2+m_{1t+2}^*q\left(1-q\right)}\frac{m_{1t+1}^*q\left(1-q\right)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^*q\left(1-q\right)}\right] \\ &+ \ldots \end{aligned} \tag{43}$$

Note that we need to take expectations over $\{\Phi_{2t},\Phi_{3t}\}$ because $\left\{d_{1t+1}^*, m_{1t+1}^*\right\}$ depend on

$\{\Phi_{2t}, \Phi_{3t}\}$. Φ_{2t} and Φ_{3t} are random because they can be either 0 or positive, depending on whether the regulator detects fraud at the two firms.

Equation (43) can be simplified into

$$
\begin{aligned}
&\frac{(1-\delta\rho)\, c d_{1t}^{*}}{1-d_{1t}^{*}}\left(1+\frac{m_{1t}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}}\right)\\
&= 1+\delta\rho E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*} q\,(1-q)}\right]\\
&+\delta^{2}\rho^{2} E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+1}^{*}\right)\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+2}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*} q\,(1-q)}\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*} q\,(1-q)}\right]\\
&+\ldots
\end{aligned}
\tag{44}
$$

There are four possible cases of $\{\Phi_{2t}, \Phi_{3t}\}$ in period $t+1$. For each realization of $\{\Phi_{2t}, \Phi_{3t}\}$, by induction, the first-order condition of m_{1t+1} is given by:

$$
\frac{(1-\delta\rho)\, c d_{1t+1}^{*}}{1-d_{1t+1}^{*}}\left(1+\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}}\right)=1+\delta\rho E_{\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+2}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*} q\,(1-q)}\right]+\ldots
\tag{45}
$$

Multiplying both sides by $\left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*} q(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*} q(1-q)}$ gives that:

$$
\begin{aligned}
&(1-\delta\rho)\, c d_{1t+1}^{*}\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}}\\
&= \left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*} q\,(1-q)}\\
&+\delta\rho E_{\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+1}^{*}\right)\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+1}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*} q\,(1-q)}\frac{m_{1t+2}^{*} q\,(1-q)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*} q\,(1-q)}\right]\\
&+\ldots
\end{aligned}
\tag{46}
$$

Taking the expectation over $\{\Phi_{2t}, \Phi_{3t}\}$ gives that:

$$
\begin{aligned}
&\frac{c(1-\delta\rho)\,q(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2} E_{\Phi_{2t},\Phi_{3t}}\left[d_{1t+1}^* m_{1t+1}^*\right] \\
&= E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^*\right)\frac{m_{1t+1}^* q(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^* q(1-q)}\right] \\
&+\delta\rho E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+1}^*\right)\left(1-d_{1t+2}^*\right)\frac{m_{1t+1}^* q(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^* q(1-q)}\frac{m_{1t+2}^* q(1-q)}{\rho^2\Phi_{1t+1}+\sigma_\varepsilon^2+m_{1t+2}^* q(1-q)}\right] \\
&+ \ldots
\end{aligned}
\tag{47}
$$

Substituting (47) into (44) yields:

$$
\frac{(1-\delta\rho)\,cd_{1t}^*}{1-d_{1t}^*}\left(1+\frac{m_{1t}^* q(1-q)}{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}\right) = 1+\frac{c\delta\rho(1-\delta\rho)\,q(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2}E_{\Phi_{2t},\Phi_{3t}}\left[d_{1t+1}^* m_{1t+1}^*\right]. \tag{48}
$$

Solving for m_{1t}^* gives that

$$
m_{1t}^* = \frac{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}{q(1-q)}\left[\frac{1-d_{1t}^*}{cd_{1t}^*}\left(\frac{1}{1-\delta\rho}+\delta\rho cq(1-q)\frac{E_{\Phi_{2t},\Phi_{3t}}\left[m_{1t+1}^* d_{1t+1}^*\right]}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2}\right)\right]-1, \tag{49}
$$

where

$$
\begin{aligned}
E_{\Phi_{2t},\Phi_{3t}}\left[d_{1t+1}^* m_{1t+1}^*\right] &= (1-d_{2t}^*)(1-d_{3t}^*)\,m_{1t+1}^*(\Phi_{1t},\Phi_{2t},\Phi_{3t})\,d_{1t+1}^*(\Phi_{1t},\Phi_{2t},\Phi_{3t}) \\
&+ d_{2t}^*(1-d_{3t}^*)\,m_{1t+1}^*(\Phi_{1t},0,\Phi_{3t})\,d_{1t+1}^*(\Phi_{1t},0,\Phi_{3t}) \\
&+ (1-d_{2t}^*)\,d_{3t}^* m_{1t+1}^*(\Phi_{1t},\Phi_{2t},0)\,d_{1t+1}^*(\Phi_{1t},\Phi_{2t},0) \\
&+ d_{2t}^* d_{3t}^* m_{1t+1}^*(\Phi_{1t},0,0)\,d_{1t+1}^*(\Phi_{1t},0,0).
\end{aligned}
\tag{50}
$$

Analogously, dropping the time subscript, the manipulation choice m_i^* by the manager at firm i can be derived as:

$$
m_i^* = \frac{\rho^2\Phi_i+\sigma_\varepsilon^2}{q(1-q)}\left(\frac{1-d_i^*}{cd_i^*}\left(\frac{1}{1-\delta\rho}+\frac{c\delta\rho q(1-q)}{\rho^2\Phi_i'+\sigma_\varepsilon^2}\times E\left[m_i' d_i'\right]\right)-1\right), \tag{51}
$$

where

$$
\begin{aligned}
E\left[m_1' d_1'\right] &= \left(1-d_2^*\right)\left(1-d_3^*\right) m_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right) d_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right) \\
&+ d_2^*\left(1-d_3^*\right) m_1^*\left(\Phi_1', 0, \Phi_3'\right) d_1^*\left(\Phi_1', 0, \Phi_3'\right) \\
&+ \left(1-d_2^*\right) d_3^* m_1^*\left(\Phi_1', \Phi_2', 0\right) d_1^*\left(\Phi_1', \Phi_2', 0\right) \\
&+ d_2^* d_3^* m_1^*\left(\Phi_1', 0, 0\right) d_1^*\left(\Phi_1', 0, 0\right), \qquad (52)
\end{aligned}
$$

$$
\begin{aligned}
E\left[m_2' d_2'\right] &= \left(1-d_1^*\right)\left(1-d_3^*\right) m_2^*\left(\Phi_1', \Phi_2', \Phi_3'\right) d_2^*\left(\Phi_1', \Phi_2', \Phi_3'\right) \\
&+ d_1^*\left(1-d_3^*\right) m_2^*\left(0, \Phi_2', \Phi_3'\right) d_2^*\left(0, \Phi_2', \Phi_3'\right) \\
&+ \left(1-d_1^*\right) d_3^* m_2^*\left(\Phi_1', \Phi_2', 0\right) d_2^*\left(\Phi_1', \Phi_2', 0\right) \\
&+ d_1^* d_3^* m_2^*\left(0, \Phi_2', 0\right) d_2^*\left(0, \Phi_2', 0\right), \qquad (53)
\end{aligned}
$$

$$
\begin{aligned}
E\left[m_3' d_3'\right] &= \left(1-d_1^*\right)\left(1-d_2^*\right) m_3^*\left(\Phi_1', \Phi_2', \Phi_3'\right) d_3^*\left(\Phi_1', \Phi_2', \Phi_3'\right) \\
&+ d_1^*\left(1-d_2^*\right) m_3^*\left(0, \Phi_2', \Phi_3'\right) d_3^*\left(0, \Phi_2', \Phi_3'\right) \\
&+ \left(1-d_1^*\right) d_2^* m_3^*\left(\Phi_1', 0, \Phi_3'\right) d_3^*\left(\Phi_1', 0, \Phi_3'\right) \\
&+ d_1^* d_2^* m_3^*\left(0, 0, \Phi_3'\right) d_3^*\left(0, 0, \Phi_3'\right). \qquad (54)
\end{aligned}
$$

Dropping the time subscript, the regulator's objective function can be rewritten recursively as:

$$
\begin{aligned}
W\left(\Phi_1,\Phi_2,\Phi_3\right)=\max_{d_1,d_2,d_3} & -\left(1-d_1\right)\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_1'+\Phi_2'+\Phi_3'-\delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& -\left(1-d_1\right)d_2\left(1-d_3\right)\left[\Phi_1'+\Phi_3'-\delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\
& -\left(1-d_1\right)\left(1-d_2\right)d_3\left[\Phi_1'+\Phi_2'-\delta W\left(\Phi_1',\Phi_2',0\right)\right] \\
& -\left(1-d_1\right)d_2d_3\left[\Phi_1'-\delta W\left(\Phi_1',0,0\right)\right] \\
& -d_1\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_2'+\Phi_3'-\delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\
& -d_1d_2\left(1-d_3\right)\left[\Phi_3'-\delta W\left(0,0,\Phi_3'\right)\right] \\
& -d_1\left(1-d_2\right)d_3\left[\Phi_2'-\delta W\left(0,\Phi_2',0\right)\right] \\
& +\delta d_1d_2d_3W\left(0,0,0\right)-\frac{\kappa}{2}\left(d_1+d_2+d_3-3d_0\right)^2, \qquad (55)
\end{aligned}
$$

where

$$
\Phi_i'\equiv\frac{m_i^*q\left(1-q\right)\left(\rho^2\Phi_i+\sigma_\varepsilon^2\right)}{\rho^2\Phi_i+\sigma_\varepsilon^2+m_i^*q\left(1-q\right)}. \qquad (56)
$$

Taking the F.O.C. yields:

$$
\begin{aligned}
d_1^* = & \frac{1}{\kappa}\{\Phi_1'+\delta\left(1-d_2^*\right)\left(1-d_3^*\right)\left[W\left(0,\Phi_2',\Phi_3'\right)-W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& +\delta\left(1-d_2^*\right)d_3^*\left[W\left(0,\Phi_2',0\right)-W\left(\Phi_1',\Phi_2',0\right)\right] \\
& +\delta d_2^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right)-W\left(\Phi_1',0,\Phi_3'\right)\right] \\
& +\delta d_2^*d_3^*\left[W\left(0,0,0\right)-W\left(\Phi_1',0,0\right)\right]\}-\left(d_2^*+d_3^*-3d_0\right), \qquad (57)
\end{aligned}
$$

$$
\begin{aligned}
d_2^* = & \frac{1}{\kappa}\{\Phi_2'+\delta\left(1-d_1^*\right)\left(1-d_3^*\right)\left[W\left(\Phi_1',0,\Phi_3'\right)-W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& +\delta\left(1-d_1^*\right)d_3^*\left[W\left(\Phi_1',0,0\right)-W\left(\Phi_1',\Phi_2',0\right)\right] \\
& +\delta d_1^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right)-W\left(0,\Phi_2',\Phi_3'\right)\right] \\
& +\delta d_1^*d_3^*\left[W\left(0,0,0\right)-W\left(0,\Phi_2',0\right)\right]\}-\left(d_1^*+d_3^*-3d_0\right), \qquad (58)
\end{aligned}
$$

$$
\begin{aligned}
d_3^* = \frac{1}{\kappa}\{&\Phi_3' + \delta\left(1 - d_1^*\right)\left(1 - d_2^*\right)\left[W\left(\Phi_1', \Phi_2', 0\right) - W\left(\Phi_1', \Phi_2', \Phi_3'\right)\right] \\
&+ \delta\left(1 - d_1^*\right) d_2^*\left[W\left(\Phi_1', 0, 0\right) - W\left(\Phi_1', 0, \Phi_3'\right)\right] \\
&+ \delta d_1^*\left(1 - d_2^*\right)\left[W\left(0, \Phi_2', 0\right) - W\left(0, \Phi_2', \Phi_3'\right)\right] \\
&+ \delta d_1^* d_2^*\left[W\left(0,0,0\right) - W\left(0, 0, \Phi_3'\right)\right]\} - \left(d_1^* + d_2^* - 3d_0\right). \qquad (59)
\end{aligned}
$$

■

Appendix II: special case of $\delta = 0$

In this appendix, we consider a special case of our model with three firms and $\delta = 0$. In this special case, we are able to obtain a closed-form solution of our model that is consistent with the numerical results shown in the main text. When $\delta = 0$, dropping the time subscript, the regulator's objective function becomes:

$$W\left(\{\Phi_i\}_{i\in\{1,2,3\}}\right) = -\sum_{i=1}^{3}(1-d_i)\,\Phi_i' - \frac{\kappa}{2}(\sum_{i=1}^{3}(d_i-d_0))^2. \tag{60}$$

In addition, using equation (26) at $\delta = 0$, we can simplify the law of motion for Φ_i (as in (15)) into:

$$\Phi_i' \equiv \left(\rho^2\Phi_i + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_i^*}{1-d_i^*}\right). \tag{61}$$

Taking the first-order condition gives that

$$\frac{\partial W}{\partial d_i} = \Phi_i' - \kappa(\sum_{i=1}^{3}(d_i-d_0)). \tag{62}$$

Without loss of generality, we assume that $\Phi_1 \geq \Phi_2 \geq \Phi_3$. This further implies that $\rho^2\Phi_1 + \sigma_\varepsilon^2 \geq \rho^2\Phi_2 + \sigma_\varepsilon^2 \geq \rho^2\Phi_3 + \sigma_\varepsilon^2$.

Consider three cases. First, suppose that

$$\left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right), \tag{63}$$

that is, Φ_1 is much larger than Φ_2. We will restate condition (63) in terms of exogenous parameters after solving the equilibrium. We now conjecture the equilibrium is that $d_2^* = d_3^* = d_0$ and $d_1^* > d_0$, where d_1^* solves:

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) = \kappa\left(d_1^* - d_0\right). \tag{64}$$

To verify that this is indeed an equilibrium, note first that the solution to (64) is unique because the left-hand side is decreasing in d_1^* whereas the right-hand side is increasing in d_1^*.

In addition, by the implicit function theorem, since the left-hand side is increasing in Φ_1, d_1^* is increasing in Φ_1. Next, using the first-order condition (64), we can rewrite the condition (63) as:

$$\kappa\left(d_1^* - d_0\right) = \Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) > \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right). \tag{65}$$

Since d_1^* is increasing in Φ_1, the condition (63) holds if and only if Φ_1 is sufficiently large and/or Φ_2 is sufficiently small. In other words, we can rewrite the condition (63) as

$$\Phi_1 > H\left(\Phi_2\right), \tag{66}$$

where $H\left(\cdot\right)$ is some given increasing function. Finally, we verify that $d_2^* = d_3^* = d_0$. This is because, at $d_2 = d_3 = d_0$, the first-order condition for d_2 is always negative, i.e.,

$$\begin{aligned}\frac{\partial W}{\partial d_2} &= \Phi_2' - \kappa\left(d_1^* - d_0\right) \\ &= \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) - \kappa\left(d_1^* - d_0\right) \\ &< \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) - \kappa\left(d_1^* - d_0\right) \\ &= 0.\end{aligned} \tag{67}$$

The third step uses (63). The last step uses (64).

Second, suppose that $\Phi_1 \le H\left(\Phi_2\right)$ and

$$\left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right), \tag{68}$$

that is, Φ_1 and Φ_2 are of similar sizes but both are much larger than Φ_3. We will restate condition (68) in terms of exogenous parameters after solving the equilibrium. We now conjecture the equilibrium is that $d_3^* = d_0$, $d_1^* > d_0$ and $d_2^* > d_0$, where the pair of $\{d_1^*, d_2^*\}$

solves:

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) = \kappa\left(D^* - 2d_0\right), \tag{69}$$

$$\Phi_2' = \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_2^*}{1-d_2^*}\right) = \kappa\left(D^* - 2d_0\right), \tag{70}$$

where $D^* = d_1^* + d_2^*$. To verify that this is indeed an equilibrium, note that, since the left-hand side of the two first-order conditions of $\{d_1^*, d_2^*\}$ are increasing in Φ_1 and Φ_2, respectively, applying the implicit function theorem gives that D^* is strictly increasing in Φ_1 and Φ_2. Using the first-order condition of d_1, we can rewrite the condition (68) as:

$$\kappa\left(D^* - 2d_0\right) = \Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) > \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right). \tag{71}$$

Since D^* is increasing in Φ_1 and Φ_2, the condition (68) holds if and only if either Φ_1 or Φ_2 is sufficiently large and/or Φ_3 is sufficiently small. In other words, we can rewrite (68) as

$$L\left(\Phi_1, \Phi_2\right) > \Phi_3, \tag{72}$$

where $L\left(\cdot,\cdot\right)$ is some given increasing function in both Φ_1 and Φ_2. Finally, we verify that $d_3^* = d_0$. This is because, at $d_3 = d_0$, the first-order condition for d_3 is always negative, i.e.,

$$\begin{aligned}
\frac{\partial W}{\partial d_3} &= \Phi_3' - \kappa\left(d_1^* + d_2^* - 2d_0\right) \\
&= \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) - \kappa\left(d_1^* + d_2^* - 2d_0\right) \\
&< \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) - \kappa\left(d_1^* + d_2^* - 2d_0\right) \\
&= 0.
\end{aligned} \tag{73}$$

Lastly, suppose that $\Phi_1 \leq H\left(\Phi_2\right)$ and $L\left(\Phi_1, \Phi_2\right) \leq \Phi_3$. That is, Φ_1, Φ_2 and Φ_3 are of similar sizes. In this case, the equilibrium can only be interior such that the equilibrium is a triplet of $\{d_1^*, d_2^*, d_3^*\} > 0$, which solve:

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_1^*}{1 - d_1^*}\right) = \kappa \left(d_1^* + d_2^* + d_3^* - 3 d_0\right), \tag{74}$$

$$\Phi_2' = \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_2^*}{1 - d_2^*}\right) = \kappa \left(d_1^* + d_2^* + d_3^* - 3 d_0\right), \tag{75}$$

$$\Phi_3' = \left(\rho^2 \Phi_3 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_3^*}{1 - d_3^*}\right) = \kappa \left(d_1^* + d_2^* + d_3^* - 3 d_0\right). \tag{76}$$

Appendix III: variable definitions

IV_q: in equation (29), IV_q is the daily implied volatility of the 90-day standardized option measured 10 trading days before the earnings announcement of $q-1$ (made in q). In equation (30)-(32), IV_q is the quarterly average of the daily implied volatility of the 90-day standardized option in quarter q. IV_q^2 is the squared term of IV_q. .

$REVISION_q$: the EPS consensus forecast for quarter q after earnings announcement (EA) of quarter $q-1$ (made in q) minus the corresponding EPS forecast before EA, scaled by the stock price two days before EA. Pre-EA consensus forecast is the latest forecast for quarter q issued at least two days before EA of quarter $q-1$ (announced in q), averaged cross analysts. Post-EA consensus forecast is the first forecast for quarter q issued within the first 30 days after EA of quarter $q-1$ (announced in q), averaged cross analysts.

SUE_q: reported EPS of quarter $q-1$ (announced in q) minus the pre-EA EPS consensus forecast, scaled by the stock price two days before EA. Pre-EA consensus forecast is the latest forecast for quarter $q-1$ issued at least two days before EA of quarter $q-1$, averaged cross analysts.

NEG_q: an indicator variable that equals one if the reported EPS of quarter $q-1$ (announced in q) is negative and zero otherwise.

$SIZE_{q-1}$: the natural logarithm of total assets at the end of $q-1$.

MB_{q-1}: market value of equity plus book value of debt, divided by book value of assets, at the end of $q-1$.

LEV_{q-1}: book value of total debt divided by book value of total assets, at the end of $q-1$.

ROA_{q-1}: operating income of quarter $q-1$ divided by book value total assets at the end of $q-2$.

$REVGWTH_{q-1}$: sales revenue of quarter $q-1$ divided by sales revenues of quarter $q-5$

(i.e., one-year lag) minus one, in percentage points.

$DETECT_{q+1}$: an indicator variable that equals one if a firm has a restatement disclosure that meets at least one of the following three conditions in quarter $q+1$ and zero otherwise: (1) if the restatement is marked as being fraudulent by Audit Analytics; (2) if the restatement has received a class-action lawsuit as tracked by Audit Analytics; or (3) if the cumulative restated amount (scaled by the total assets as of the last restating quarter) is in the top decile of the sample.

$IVQn_q$: an indicator variable that equals one if a firm-quarter falls into the nth-ranked quintile of IV (n=1 to 5) and zero otherwise, with a higher-ranked quintile representing the subsample with a higher level of average daily implied volatility in quarter q.

$WAVE_q$: an indicator variable that equals one if an industry-quarter's fraud detection rate exceeds the 90^{th} percentile of the empirical distribution based on the industry's fraud detection rates over all quarters in the sample. The fraud detection rate of an industry i in a given quarter q is the number of firms with restatement announcement in industry-quarter j, q divided by the number of firms in industry-quarter j, q. The industry classification is based on the Global Industry Classification Standard (GICS) 4-digit industry groups.