Everlasting Fraud

Vivian W. Fang*[1, 2]

Nan Li[1]

Wenyu Wang[3]

Gaoqing Zhang[1]

[1]Carlson School of Management, University of Minnesota

[2]ECGI

[3]Kelley School of Business, Indiana University

March 10, 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 Carlson School of Management, University of Minnesota, Minneapolis, MN 55455; and Wang is at Kelley School of Business, Indiana University, Bloomington, IN 47405. We are grateful to helpful comments from Xu Jiang and Lorien Stice-Lawrence. 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. 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 the unexpected earnings (i.e., reported earnings minus earnings expected by the market), by weighing his benefits and costs of committing fraud. The marginal cost is linked to detection likelihood. The marginal benefit depends on how much the market values the reported earnings and increases with the amount of fraud built to date. A high level of cumulative fraud increases information uncertainty about the firm, which in turn boosts the value of accounting reports and potential return from reporting fraudulently. The regulator decides on the amount of resources spent on a detection technology; the more regulatory resources spent, the higher the detection likelihood. The regulator seeks to maximize the informativeness of the firm's report minus detection cost.

Analyses of this single-firm model begin to tell why fraud may never cease to exist. Although the manager and the regulator independently solve their own maximization problem, their calculus are intertwined such that the regulator correctly conjectures the level of fraud in the firm's report and then decides how much to spend on the detection technology. If the regulator anticipates a low level of fraud built up in the firm, then she would spend few resources on detection. The manager thus continues to commit fraud as the marginal benefit likely outweighs the marginal cost. As fraud gradually builds up, a higher information uncertainty gives the manager greater incentives to commit fraud. At the same time, the regulator would increase spending on detection. The two effects go hand-in-hand, simultaneously increasing the marginal benefit and cost. When fraud reaches a critical level, the regulator would concentrate resources on the firm and the marginal cost of continuing fraud eventually dominates the marginal benefit. 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. Thus, 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 marginal cost of committing fraud. Upon detection, the marginal benefit 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 four analyses.

First, we link implied volatility to the marginal benefit 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 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

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

finding is consistent with the regulator rationally allocating more resources towards more fraudulent firms, thus increasing the likelihood of catching fraud at these firms.

Results of the first two analyses point to a non-monotonic relation between the amount of fraud committed in a quarter and implied volatility, because a high level of cumulative fraud increases both the marginal benefit of continuing fraud (by exacerbating information uncertainty) and the marginal cost (by attracting regulatory scrutiny). The third analysis estimates this relation in a quadratic regression. We observe a hump-shaped association between a firm's restatement amount in a quarter (set to zero if there is no restatement) and the level of implied volatility of the quarter, which is consistent with the two countervailing effects illustrated by the model.

Our final 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

[2]For a comprehensive review of theories on earnings manipulation, please see two recent surveys by Ewert

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 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 between firms and regulators that is unlikely to end. In essence, given that every regulator faces finite resource 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,

and Wagenhofer (2012) and Stocken (2013).

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 marginal benefit and marginal cost 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' marginal benefits 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. The report is used to update the firm's valuation V_t. We assume that V_t is set by a competitive market and equals the firm's total discounted future earnings in expectation:

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

where the discounting factor $\delta \in (0, 1)$ and $\mathcal{F}_t \equiv \{r_t, r_{t-1}, ...r_1\}$ stands for the set of the firm's reports up to time t. We assume that the manager's compensation in period t depends on the extent to which the firm outperforms the market expectation, i.e., $V_t - E\left[V_t | \mathcal{F}_{t-1}\right]$. This assumption is reasonable since most firm managers' compensation is either directly written on earnings metrics or indirectly tied to earnings through stock price (and positive earnings surprise typically leads to favorable market reaction).

To boost firm valuation and his compensation, the manager has incentives to inflate the report r_t. 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 error $\{\xi_l\}_{l=1}^{m_t}$ into 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]

$$r_i = s_i + \sum_{l=1}^{m_t} \xi_l. \tag{3}$$

Using the central limit theorem, we can approximate the distribution of $\sum_{l=1}^{m_t} \xi_l$ as $\sum_{l=1}^{m_t} \xi_l \sim N(m_t(1-q), m_t q(1-q))$. With the manager's manipulation choice $m_t \geq 0$, the report becomes:

$$r_t = s_t + m_t(1-q) + \sqrt{m_t q(1-q)}\eta_t. \tag{4}$$

η_t is a standard normal random variable that is independent of all other variables in the model. Equation (4) suggests that manipulation has two effects on the report. m_t increases the mean of the report but reduces its precision.

The regulator utilizes a detection technology to inspect the manager's report r_t. In each period t, the regulator chooses, at a cost of $\frac{\kappa}{2}d_t^2$, the probability $d_t \in [0, 1]$ that the inspection uncovers the fraudulent amount of the report and reveals the firm's true earnings. If the regulator successfully detects fraud, she requires 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 manager that is proportional to the fraudulent amount:

$$C_t = c(r_t - s_t), \tag{5}$$

where the coefficient $c > 0$. We assume that the regulator chooses the detection probability d_t to maximize the informativeness of the set of reports $\mathcal{F}_t$ about the firm's valuation V_t. Note that since the earnings follow an AR(1) process, the period-t earnings s_t is a sufficient statistic

[3]To illustrate our modeling of the manager's manipulation decision, consider the following example. The firm's earnings represent the sum of different line items in the financial statements (e.g., sales revenue and cost of goods sold). The manager can choose to add one unit of positive bias to each of the line items (e.g., either over-report revenue items or under-report expense items) to inflate the earnings report. However, the manager's manipulation attempt may be blocked by the firm's internal control system, where the probability that the internal control system successfully prevents each of the manager's fraudulent attempts is q.

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, i.e., $\Phi_t \equiv var\left(s_t|\mathcal{F}_t\right)$.[4]

Finally, we specify the payoffs of the players in our model. The manager's period-t payoff includes two parts. Conditional on detection of the manager's fraudulent behavior, the firm's true economic earnings s_t is revealed and the firm value becomes:

$$V_t = \frac{s_t}{1-\delta\rho}. \tag{6}$$

The manager incurs a penalty C_t and receives a compensation that equals the firm value minus the prior expected level. Therefore, the manager's payoff is:

$$\frac{E\left[s_t|\mathcal{F}_t\right]}{1-\delta\rho} - \frac{E\left[s_t|\mathcal{F}_{t-1}\right]}{1-\delta\rho} - C_t = \frac{s_t - \rho s_{t-1}}{1-\delta\rho} - C_t. \tag{7}$$

Alternatively, if the detection fails, the manager incurs no penalty and obtains a compensation that equals:

$$\frac{E\left[s_t|\mathcal{F}_t\right]}{1-\delta\rho} - \frac{E\left[s_t|\mathcal{F}_{t-1}\right]}{1-\delta\rho}. \tag{8}$$

Thus, the manager's expected period-t payoff is given by:

$$d_t\left(\frac{s_t - \rho s_{t-1}}{1-\delta\rho} - C_t\right) + (1-d_t)\left(\frac{E\left[s_t|\mathcal{F}_t\right]}{1-\delta\rho} - \frac{E\left[s_t|\mathcal{F}_{t-1}\right]}{1-\delta\rho}\right). \tag{9}$$

The manager seeks to maximize the present value of the sum of his future expected payoffs:

$$U_t = E\left[\sum_{k=t}^{\infty}\delta^{k-t}\left[d_k\left(\frac{s_k - \rho s_{k-1}}{1-\delta\rho} - C_k\right) + (1-d_k)\frac{E\left[s_k|\mathcal{F}_k\right] - E\left[s_k|\mathcal{F}_{k-1}\right]}{1-\delta\rho}\right]\right]. \tag{10}$$

The regulator's period-t payoff also includes two parts. Conditional on detection, the true earnings s_t is revealed and the conditional variance Φ_t drops to zero. Alternatively, if the detection fails, the conditional variance remains at $\Phi_t > 0$. Therefore, considering the cost

[4] As such, 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. 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.

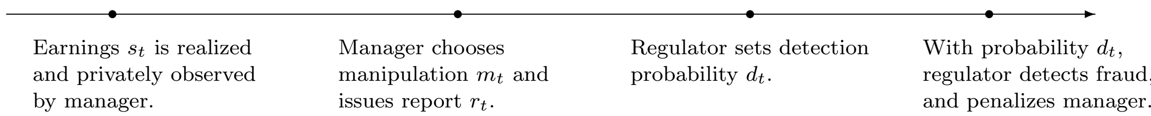

Figure 1: Timeline of the period-t game

of detection, the regulator's expected period-t payoff is given by

$$-\left(1-d_t\right)\Phi_t-\frac{\kappa}{2}d_t^2. \tag{11}$$

The regulator seeks to maximize the present value of the sum of her future expected payoffs:

$$W_t=\sum_{k=t}^{T}\delta^{k-t}\left(-\left(1-d_k\right)\Phi_k-\frac{\kappa}{2}d_k^2\right). \tag{12}$$

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

2.2 Analysis

We first derive, in each period t, the manager's manipulation decision m_t, anticipating the regulator's equilibrium detection choice d_t^*.[5] In doing so, we first compute the valuation of the firm conditional on the manager's report r_t. If the regulator detects fraud, the report reveals the true earnings s_t and the firm value equals

$$V_t=\frac{s_t}{1-\delta\rho}, \tag{13}$$

as discussed previously. Alternatively, the report is a noisy measure of s_t because of the manipulation. Note that since earnings s_t follow an AR(1) process, the market can update its expectation about s_t in two steps. First, the market uses the set of past reports up to period $t-1$, $\mathcal{F}_{t-1}$, to update the expectation about the last period's earnings s_{t-1}. Denote such expectation as $E\left[s_{t-1}|\mathcal{F}_{t-1}\right]$ and the (inverse) precision of past reports about the last-

[5] 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.

period's earnings s_{t-1} as $\Phi_{t-1} \equiv var\left(s_{t-1}|\mathcal{F}_{t-1}\right)$. Second, treating $E\left[s_{t-1}|\mathcal{F}_{t-1}\right]$ as a new prior for s_{t-1}, the market then uses report r_t to update the expectation about s_t. This two-step procedure gives the firm value as:

$$V_t = \frac{E\left[s_t|\mathcal{F}_t\right]}{1-\delta\rho} = \frac{1}{1-\delta\rho}\left[\rho E\left[s_{t-1}|\mathcal{F}_{t-1}\right] + \frac{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^* q\left(1-q\right)}\left(r_t - m_t^*(1-q)\right)\right]. \tag{14}$$

Note that since the market does not observe the manager's actual choice of manipulation m_t, it forms a conjecture about the manager's manipulation choice in equilibrium m_t^*. Rational expectations require that the conjecture is consistent with the manager's equilibrium choice.

Equation (14) characterizes how the market forms its conjecture about firm value based on the current report, r_t, and m_t^* plays two roles in this process. First, anticipating that the manager has chosen manipulation m_t^* and inflated the mean of the report by $m_t^*\left(1-q\right)$, the market rationally subtracts the report by the average bias $m_t^*\left(1-q\right)$; this is reflected by the term $r_t - m_t^*(1-q)$ in the equation. Second, the value relevance of the report is affected by two countervailing forces. On the one hand, manipulation adds noise to the current report and reduces precision, and thus the market places a smaller weight on the report if m_t^* is larger; the term $m_t^* q(1-q)$ in the denominator measures the variance of the noise added. On the other hand, information uncertainty regarding the firm's fundamental value (i.e., Φ_{t-1}), induced by cumulative fraud in the past, increases the value of the current report. In other words, if past reports are noisier (or Φ_{t-1} is larger), the market has to rely more on the current report to value the firm, thus placing a greater weight on the report.

Substituting (14) into the manager's payoff (10) gives that:

$$U_t = E\left[\sum_{k=t}^{\infty}\delta^{k-t}\left[d_k^*\left(\frac{s_k-\rho s_{k-1}}{1-\delta\rho} - C_k\right) + \frac{1-d_k^*}{1-\delta\rho}\frac{\rho^2\Phi_{k-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{k-1}+\sigma_\varepsilon^2+m_k^* q\left(1-q\right)}\left(r_k - m_k^*(1-q)\right)\right]\right]. \tag{15}$$

Note that the manager's actual choice of manipulation in period t, m_t, only affects his payoff, U_t, in the same period. Based on this observation, taking the first-order condition of U_t with respect to m_t gives that:[6]

$$\frac{1-d_t^*}{1-\delta\rho}\frac{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^* q(1-q)}=cd_t^*. \tag{16}$$

The right-hand side of equation (16) represents the marginal cost of manipulation, which is related to the penalty that the regulator imposes upon detection of fraud. The marginal cost, therefore, is increasing in the equilibrium detection probability d_t^*. The left-hand side of equation (16) represents the marginal benefit of manipulation. When the regulator fails to detect fraud in the report (with probability $1-d_t^*$), the manipulation by the manager inflates the report, the firm valuation, and the manager's own compensation. The larger the weight placed on the report, the stronger the manager's incentives to inflate the report. Since the lack of informativeness in the firm's past reports increases the weight placed on the current report, it increases the marginal benefit of manipulation (i.e., all else equal, m_t^* is increasing in the conditional variance Φ_{t-1}). Solving (16) yields the manager's manipulation decision m_t^*, which we state formally in the following lemma.

Lemma 1 *In each period t, given the regulator's equilibrium detection choice d_t^* and the (inverse) informativeness of the firm's past reports Φ_{t-1}, the manager chooses manipulation*

$$m_t^*\left(d_t^*,\Phi_{t-1}\right)=\frac{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}{q(1-q)}\max\left\{\frac{1-d_t^*}{c(1-\delta\rho)\,d_t^*}-1,0\right\}. \tag{17}$$

m_t^* is decreasing in d_t^* and increasing in Φ_{t-1}.

Next, we derive the regulator's choice of detection probability d_t, given the manager's equilibrium manipulation choice m_t^* in (17). We first rewrite the regulator's payoff in a recursive form using the (inverse) informativeness of the firm's past reports Φ_{t-1} as the state variable. Standard Bayesian updating yields the law of motion for the state variable Φ_{t-1}, which we state in the following lemma.

[6] In deriving equation (16), we have also used that the effect of manipulation on the report $\frac{\partial E[r_t]}{\partial m_t}=1-q$ and the effect of manipulation on the penalty $\frac{\partial E[C_t]}{\partial m_t}=c(1-q)$.

Lemma 2 *In each period t, if the regulator detects fraud, the conditional variance about the firm's earnings s_t, $\Phi_t \equiv 0$. 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^*:*

$$\Phi_t\left(d_t^*, \Phi_{t-1}\right) = \frac{m_t^* q\left(1-q\right)\left(\rho^2\Phi_{t-1}+\sigma_\varepsilon^2\right)}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^* q\left(1-q\right)}, \tag{18}$$

where $m_t^*\left(d_t^*, \Phi_{t-1}\right)$ is given in (17).

The law of motion (18) is intuitive. It states that the uncertainty about the firm's earnings is increasing in both the manager's equilibrium manipulation m_t^* and the prior uncertainty about the earnings Φ_{t-1}. Iterating (18) over time suggests that the uncertainty about the earnings Φ_t is essentially determined by accumulating the manager's undetected manipulations in the past, i.e., $\left\{m_t^*, m_{t-1}^*, ...\right\}$. As such, we interpret the state variable Φ_t as the cumulative level of fraud (up to period t).

We now rewrite the regulator's payoff (12) recursively:

$$\begin{aligned} W_t\left(\Phi_{t-1}\right) &= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*, \Phi_{t-1}\right) - \frac{\kappa}{2}d_t^2 + \delta E\left[\sum_{k=t+1}^{T}\delta^{k-(t+1)}\left(-\left(1-d_k^*\right)\Phi_k - \frac{\kappa}{2}d_k^{*2}\right)\right] \\ &= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*, \Phi_{t-1}\right) - \frac{\kappa}{2}d_t^2 + \delta\left[d_t W_{t+1}\left(0\right) + \left(1-d_t\right)W_{t+1}\left(\Phi_t\left(d_t^*, \Phi_{t-1}\right)\right)\right], \end{aligned} \tag{19}$$

where $\Phi_t\left(d_t^*, \Phi_{t-1}\right)$ is given in (18). 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 gives that

$$\Phi_t + \delta\left[W_{t+1}\left(0\right) - W_{t+1}\left(\Phi_t\right)\right] = \kappa d_t. \tag{20}$$

The right-hand side of equation (20) represents the marginal cost of increasing the detection likelihood whereas the left-hand side represents the marginal benefit. Equation (20) suggests that increasing the detection likelihood yields two benefits. First, detection uncovers the firm's true earnings for the current period, and reduces the uncertainty about firm earnings to zero as opposed to a higher level Φ_t had fraud not been detected. Second, going forward, the regulator's value function is reset to be a fraud-free level $W_{t+1}(0)$ as opposed to a lower level with cumulative fraud $W_{t+1}(\Phi_t)$. This reflects the future benefit of fraud detection. We summarize the equilibrium detection probability in the following lemma.

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

$$d_t^* = \frac{\Phi_t + \delta\left[W(0) - W(\Phi_t)\right]}{\kappa}. \tag{21}$$

Combining the first-order condition (20) and the manager's equilibrium manipulation (17) characterizes the equilibrium of the single-firm model. Note that both m_t^* and d_t^* can be written recursively as functions of Φ_{t-1}. Thus, treating Φ_{t-1} as a state variable for period t and suppressing the time subscript, we state the equilibrium in the following proposition.

Proposition 1 *For a given level of accumulated past fraud Φ, the regulator's equilibrium detection choice $d^*(\Phi)$ and the manager's equilibrium manipulation choice $m^*(\Phi)$ are given by the following set of equations:*

$$d^*(\Phi) = \frac{\Phi' + \delta\left[W(0) - W(\Phi')\right]}{\kappa}, \tag{22}$$

$$m^*(\Phi) = \frac{\rho^2\Phi + \sigma_\varepsilon^2}{q(1-q)} \max\left\{\frac{1-d^*}{c(1-\delta\rho)d^*} - 1, 0\right\}, \tag{23}$$

where

$$\Phi'(\Phi) = \left(\rho^2\Phi + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c(1-\delta\rho)d^*}{1-d^*}, 0\right\}, \tag{24}$$

$$W(\Phi) = -(1-d^*)\Phi' - \frac{\kappa}{2}(d^*)^2 + \delta\left[d^* W(0) + (1-d^*) W\left(\Phi'\right)\right]. \tag{25}$$

The dynamic programming problem in Proposition 1 does not have a closed-form solution, and thus we will solve the full model using the numerical method. However, to glean some insights into the properties of the equilibrium detection probability and manipulation, we first consider a first-order approximation to the first-order condition (20) on d^*. Approximating (20) with a first-order Taylor expansion on the value function W around $\Phi = 0$ gives that:

$$\left[1 - \delta W'(0)\right] \left(\rho^2 \Phi + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1 - \delta\rho)\, d^*}{1 - d^*}, 0\right\} = \kappa d^*. \tag{26}$$

Note that the marginal benefit of detection is increasing in the cumulative level of fraud. An application of the implicit function theorem thus implies that, the regulator should match the strength of fraud detection with the severity of fraud in equilibrium (i.e., d^* is increasing in Φ). This is intuitive as the regulator seeks to maximize the informativeness of the firm's reports and all reporting noises in our model are endogenously generated from the manager's fraudulent reporting actions. When the manager has engaged excessively in fraudulent reporting, the gains from detecting fraud are substantial.

Next, we use the approximated d^* in (26) to draw some inference 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 also suggests that as Φ increases, the regulator would invest more heavily in the detection technology, which deters the manager's manipulation incentives 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 to the true model solution, and we then solve the model numerically to verify our findings. We set the 6 model parameters as follows in our numerical solution: we set the subjective discount rate, δ, to be 0.9, a value commonly used in the literature. We set the success rate of manipulation, q, to be 0.5, an innocuous assumption in the model. Since the regulator's objective function relates the detection cost parameter κ to the conditional uncertainty Φ that in turn is determined by the parameter σ_ε, we can normalize κ to be 1 and then set σ_ε to be 0.09. Last, we set the manager's cost of being caught on involving in manipulation, c, to be 2. This parameter value

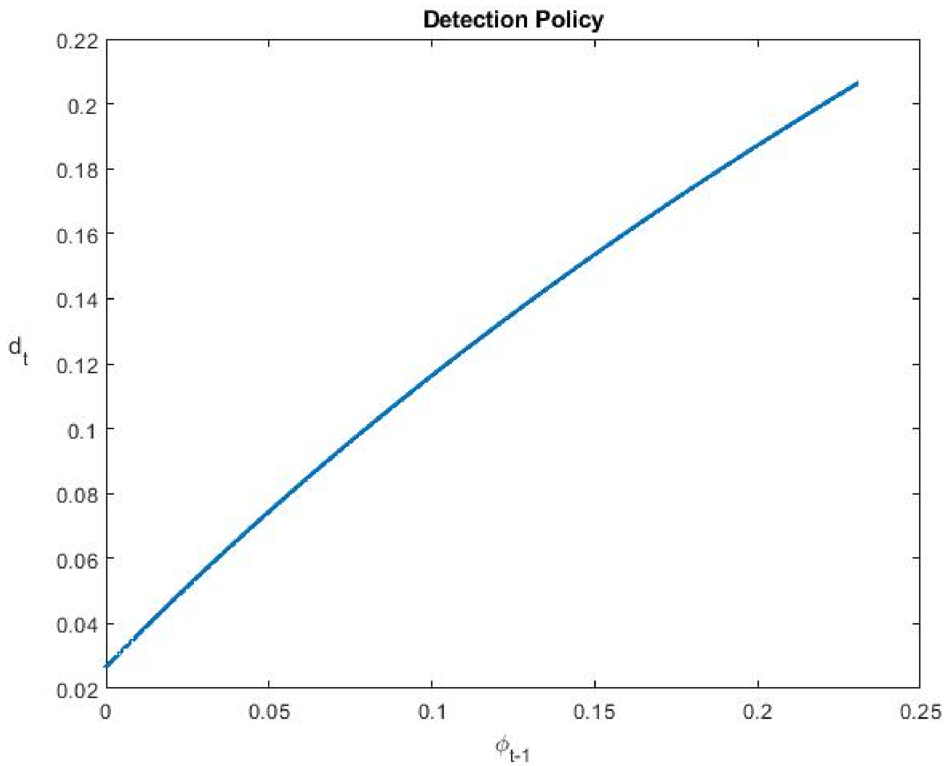

Figure 2: Equilibrium detection probability $d^*(\Phi)$.

suggests that the fine on manipulation is twice as large as the magnitude of manipulation quantity.

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 the detection intensity increases with the firm's cumulative fraud level. The model therefore predicts that the marginal cost of manipulation increases with the firm's cumulative fraud.

Figure 3 illustrates the manager's optimal manipulation decision, m_t^*, as a function of the firm's state variable Φ_{t-1}. For the set of parameter values we use in the numerical solution, we find that the equilibrium manipulation is hump-shaped in the firm's cumulative fraud. To see the intuition, note that when Φ is very low (e.g., close to 0), the market is highly informed about the firm's financial conditions, and thus puts little weight on the firm's report. This implies a low marginal benefit of manipulation and few incentives for the manager to

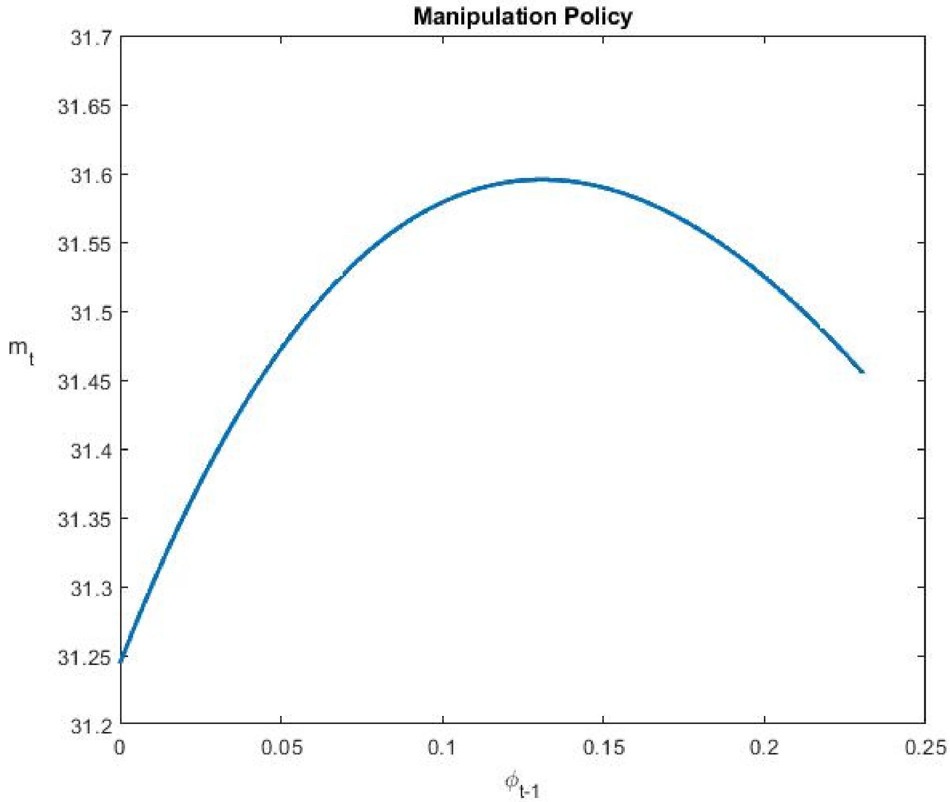

Figure 3: Equilibrium manipulation $m^*(\Phi)$.

inflate the report. When Φ is very high, the regulator would spend heavily on detection, which sharply increases the marginal cost of manipulation. Trading off the marginal benefit and marginal cost of manipulating earnings, the maximal 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

We now expand our single-firm model to study the dynamic features of fraud among multiple firms. The model setup is similar to the single-firm setting, 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. For most of our numerical analysis, we will focus on a special case of three firms, i.e., $N = 3$. Second, the regulator needs to allocate resources on detecting fraud among N firms. 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_i, where $i \in \{1, 2, ..., N\}$. We assume that the total detection cost for each period is:

$$\frac{\kappa}{2}(\sum_{i=1}^{N} d_i)^2. \tag{27}$$

Note that this structure of the cost function implies a limited budget for the regulator, in the sense that if the regulator allocates more resources towards inspecting one firm's report, her marginal cost of detecting fraud at other firms goes up.

3.2 Analysis

In the setting with multiple firms, the manipulation decision of each manager is similar to the one in the single-firm setting. Thus, following similar steps, we obtain the same first-order condition on the manipulation decision m_i^*:

$$\frac{1 - d_i^*}{1 - \delta\rho} \frac{\rho^2\Phi_i + \sigma_\varepsilon^2}{\rho^2\Phi_i + \sigma_\varepsilon^2 + m_i^* q\,(1 - q)} = c d_i^*. \tag{28}$$

Note that, fixing the regulator's detection decision d_i^*, the managers' manipulation decisions are independent of each other. However, because the regulator's choices of detection probability are interdependent across the firms, there exists an endogenous link among the managers' manipulation decisions.

We solve the three-firm model numerically and use the same parameter values as we set for the one-firm model. 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 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. It is easy to verify by model symmetry that, in this case, the detection intensity on firm 2 and 3 is identical, 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.

Three interesting observations emerge. First, the regulator concentrates on detecting firm 1 when its cumulative fraud is high and its information uncertainty stands out among the three firms. Specifically, at the northwestern corner where $\Phi 1 >> \Phi_2 = \Phi_3$, the regulator invests almost all detection resource on firm 1, which leaves firm 2 and 3 under the radar. Second, on the contrary, at the southeastern 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 that the regulator imposes intensive detection on both firm 2 and 3 and pays little attention to firm 1.

Second, even though the two scenarios above seem symmetric in pattern, the detection intensity imposed on firm 1 (about 0.2) in the first scenario is much larger than the detection intensity imposed on firm 2 and 3 respectively (about 0.12) in the second scenario. This is because the regulator's cost of detection is convex in the aggregate detection intensity, as shown in Equation (27), and thus the marginal cost of detecting one firm also depends on whether other firms in the economy require close scrutiny. The model implies that it is the most costly to the regulator if fraud clusters 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 worth noting that, even though we demonstrate the model-implied detection policy above using a special case in which $\Phi_2 = \Phi_3$, the intuition remains the same in more general

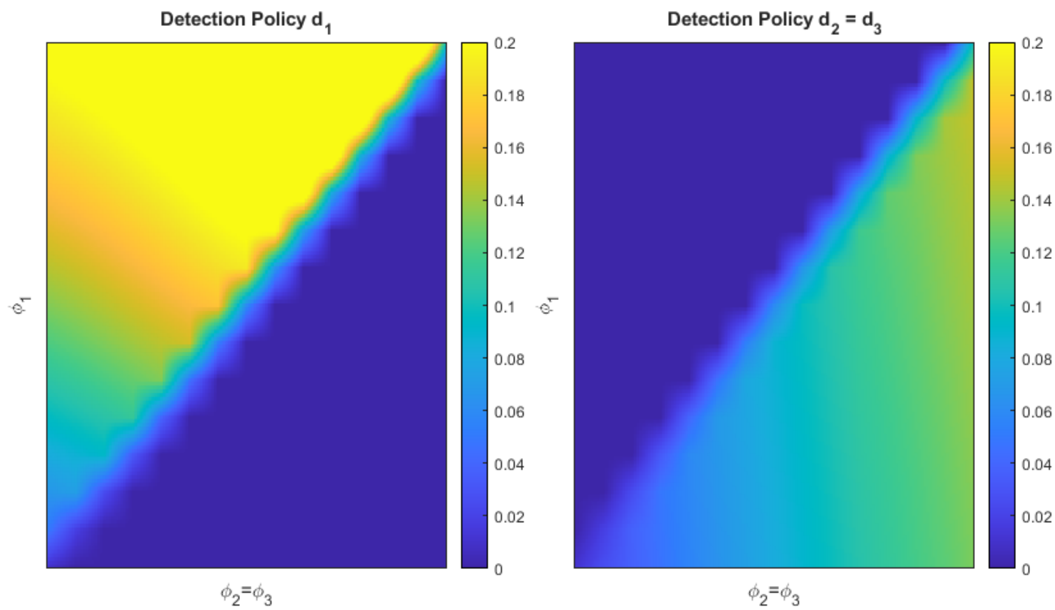

Figure 4: Equilibrium detection probability d_i^* in the three-firm setting.

cases when the three firms have different levels of information uncertainty.[7]

Given the regulator's detection policy discussed above, Equation (28) suggests that managers' manipulation decisions are also interdependent in our model. Intuitively, if one firm stands out in its cumulative fraud, it expects to face intensive detection from the regulator and thus the cost of committing fraud outweighs the benefit, making the firm's manager more conservative in manipulating the report. Meanwhile, as the firm with the highest information uncertainty attracts the most attention by the regulator, other firms are subject to less scrutiny and thus become aggressive in manipulation. To the extent that manipulation accumulates in each period and adds to the firms' information uncertainty over time, our model predicts an unintended consequence of regulation — it results in an endogenous synchronization of fraud across firms and may eventually lead to fraud waves even without the presence of systematic shocks in the economy.

We next use the model to study the dynamics of the fraud-detection game between the

[7]For illustrational purposes, we solve the model in closed form in a special case of 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.

regulator and three firms with different levels of cumulative fraud (or 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 each manager commits 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 $\Phi_{i,t}$, $d_{i,t}$, and $m_{i,t}$. Figure 5 plots the three variables over the simulation path.

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

M-firm 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 marginal benefit 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 marginal cost of committing fraud. The two effects go hand-in-hand. The first five periods of the yellow solid line in Figure 5 shows the stage when marginal benefit dominates marginal cost, 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 (i.e., the yellow solid line in Panel C) and marginal cost outweighs marginal benefit, leading to a sharp decline in manipulation by M-firm (i.e., the yellow solid line in Panel B). The dynamics in m_M therefore demonstrate the counteracting forces of marginal benefit and marginal cost.

Last, note that H-firm 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 further examine the impact of realized detection on the dynamics of firm fraud, we assume in the simulation trial that H-firm is caught by the regulator at period 15. Upon detection, H-firm's cumulative fraud is cleared and thus Φ_H drops to zero immediately, as shown in Panel A. As an optimal response, the regulator reduces attention to H-firm and refocuses detection on M- and L-firms, as suggested in Panel C. Interestingly, as the detection intensity on H-firm drops substantially, H-firm faces a low marginal cost of committing fraud and thus becomes aggressive in manipulating its report. We observe a sharp increase in m_H and Φ_H in Panel B and A right after period 15. If M- and L-firms remain undetected, cumulative fraud in the three firms will be synchronized again after another few periods.

Our model and simulation therefore reveal an unintended consequence: regulation may synchronize corporate fraud across firms 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. Depending on the analysis, we merge either the misstating quarters or 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 1996 to 2017, represents an intersection of the databases that we use. The number of firm-quarter observations used in our analyses ranges between 124,426 and 165,501.

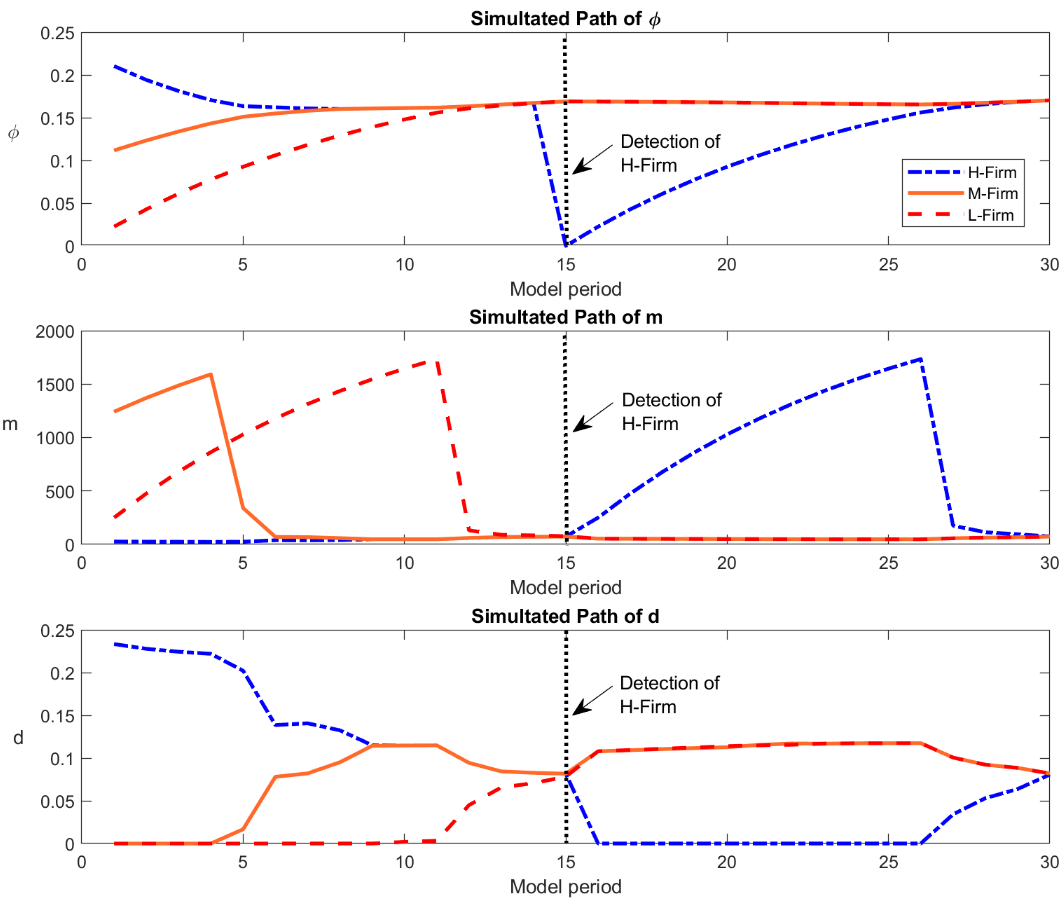

Figure 5: Simulated path of accumulated fraud Φ^*, manipulation m^*, and detection probability d^*.

4.2 Measurement of Cumulative Fraud, Detection, and Fraud Amount

As discussed in Section 2, our model analyses center on the interdependence among the three key parameters: Φ, the cumulative fraud up to each period; d, the fraud detection likelihood in each period; and m, the amount of fraud committed in each period. The first parameter—the firm's cumulative fraud to date—affects the manager's fraud decision because it impacts the information uncertainty about the firm. To measure fraud-induced 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 might arise because of this non-constant maturity, we use the implied volatility from 30-day standardized option prices provided by Option Metrics. Specifically, we take the mean of 30-day call-implied volatility and 30-day 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 is announced or accounting irregularities discovered in a quarter, and zero otherwise. 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.

To operationalize the third parameter—fraud amount—we calculate $FRAUD$ as the firm's magnitude of restatement in net income scaled by the absolute value of operating income (after restatement, if any) in the misstating quarter. $FRAUD$ is coded as zero for all other firm-quarters. If the dollar amount of a restatement is missing in Audit Analytics, we remove the firm-quarters that are associated with the restatement from the analyses involving

$FRAUD$. In robustness checks, we further refine the measure to capture the magnitude of restatement only if the restatement is likely to be fraudulent as classified by Audit Analytics.

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. As discussed in Section 5, the $REVISION$-to-SUE sensitivity captures how market updates its expectation in response to reported earnings.

For controls, we follow the accounting 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 0.495, a median of 0.439, and a standard deviation of 0.242. The mean of $Detect$ is 0.017, meaning on average a firm in our sample has a 1.7% likelihood to have at least one restatement announced in a quarter. 13,944 firm quarters, or 8.8% of the sample, have positive $FRAUD$. Conditional on $FRAUD$ being positive, the average restated amount in a quarter is 9.4% of the quarterly operating income, suggesting the restatements in the sample have nontrivial impacts on the reported earnings on average.

5 Empirical Analyses

5.1 Cumulative Fraud and Marginal Benefit of Fraud

Our model predicts that the marginal benefit of committing fraud is positively associated with the firm's cumulative fraud to date, because a high level of cumulative fraud increases the information uncertainty about the firm, which in turn boosts the value of accounting reports and the potential return from reporting fraudulently.

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:

$$REVISION_{i,q} = \alpha + \beta_1 SUE_{i,q} \times IV_{i,q} + \beta_3 SUE_{i,q} + \beta_4 IV_{i,q} + \beta_c CONTROLS_{i,q-1}, \quad (29)$$

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 (29). 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 the level of cumulative fraud is higher because of a larger degree of information uncertainty.

In Table 2 column (2), we reestimate equation (29) 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 marginal benefit 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 Cumulative Fraud 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:

$$DETECT_{i,q+1} = \alpha + \beta_1 IV_{i,q} + \beta_c CONTROLS_{i,q-1}. \tag{30}$$

The dependent variable, $DETECT$, is an indicator variable that denotes whether an 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 (30). 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 (30) including five basic firm characteristics. The coefficient of interest, β_1, remains positive; it is significant at the 1% level in column (2) excluding firm fixed effects and 5% level in column (3) including firm fixed effects, respectively. This result suggests that the marginal cost of committing fraud is larger when the information uncertainty about the firm (partly brought by cumulative fraud) is higher because the regulator would rationally allocate more resources towards such firms.

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 pricing (Hennes et al. (2008); Fang et al. (2017)). In column (4), we replace $DETECT$ with $DETECT\_ALT$, which is an indicator variable that denotes whether an fraud-related restatement is announced for firm i in a given quarter q. 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. Results using this refined indicator to denote fraud detection remain similar.

Among the controls, one variable that is worth highlighting is market-to-book MB, which carries a significantly negative coefficient in columns (2)-(4). This result suggests that fraud detection likelihood is lower for growth firms.

5.3 Cumulative Fraud and Additional Fraud

Results of Section 5.1-5.2 point to a non-monotonic relation between the amount of fraud committed in a period and the amount of fraud accumulated to date, because a high level of cumulative fraud increases both the marginal benefit and marginal cost of further committing fraud. Intuitively, when the level of fraud is initially low in a firm, the marginal benefit likely outweighs the marginal cost because the likelihood of the firm being detected is small, thus allowing fraud to accumulate. As fraud gradually builds up and reaches a critical level, the marginal cost of committing fraud eventually outweighs the marginal benefit because the regulator would now concentrate efforts on catching fraud at the firm. Figure 3 graphically illustrates these patterns.

The above reasoning suggests a possible hump-shaped relation between a firm's misstatement amount in a quarter and the level of implied volatility prior to the quarter. We examine the relation by estimating the following quadratic regression:

$$FRAUD_{i,q} = \alpha + \beta_1 IV_{i,q} + \beta_2 IV_{i,q}^2 + \beta_c CONTROLS_{i,q-1}. \tag{31}$$

The dependent variable, $FRAUD$, measures firm i's earnings misstatement in its fiscal quarter q. We compute $FRAUD$ as the firm's magnitude of restatement in net income in the misstating quarter, scaled by the absolute value of the restated operating income of the quarter. $FRAUD$ is coded as zero for all other firm-quarters. IV is the average daily implied volatility of quarter q, and IV^2 is the squared term of IV. As before, we include year-quarter fixed effects, and cluster standard errors by firm and quarter.

Table 4 column (1) reports the regression results of estimating equation (31). As shown, IV exhibits a positive coefficient and its squared term exhibits a negative coefficient, both significant at the 1% level. This result demonstrates a hump-shaped relation between the amount of fraud committed in a quarter and the level of fraud accumulated up to the quarter.

In our model, such a hump-shaped relation arises because the level of cumulative fraud has countervailing effects on the manager's incentives to further commit fraud: the marginal benefit dominates when the level of cumulative fraud is low, and the marginal cost takes over when the level of cumulative fraud is sufficiently high. Table 4 column (2) repeats the analysis, including five basic firm characteristics as controls. Table 4 column (3) further includes firm fixed effects. The inference that we draw on IV and its squared term remains qualitatively similar. In column (4), we replace $FRAUD$ with $FRAUD\_ALT$, which is defined similarly as $FRAUD$ but based on fraud-related restatements. Results using this alternative measure of fraud amount are also qualitatively similar.

5.4 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:

$$\Delta IV_{i,q\ to\ q+1} = \alpha + \beta_1 IVQ1_{i,q} + \beta_2 IVQ2_{i,q} + \beta_3 IVQ4_{i,q} + \beta_4 IVQ5_{i,q} + \beta_c CONTROLS_{i,q-1}, \tag{32}$$

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 5 columns (1) and (2) report the regression results of estimating equation (32), 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 (32) 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 misstating firms 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.

Table 5 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 three 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 marginal benefit 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 marginal cost of committing fraud. Third, consistent with the existence of two countervailing effects, we document a hump-shaped relation between the amount of fraud committed in a period and implied volatility in the period.

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 marginal cost of committing fraud but also sharply decreasing their marginal benefit 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 these insights, 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 in a quarter. $FRAUD$ is the restated amount in the quarter if there is a restatement or zero otherwise. $DETECT\_ALT$ and $FRAUD\_ALT$ are alternative constructs of $DETECT$ and $FRAUD$ based on fraud-related restatements. $REVISION$ 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	165,444	0.495	0.242	0.318	0.439	0.613
Restatement Variables						
DETECT	165,444	0.017	0.129	0.000	0.000	0.000
DETECT_ALT	165,444	0.007	0.082	0.000	0.000	0.000
FRAUD	158,302	0.008	0.045	0.000	0.000	0.000
FRAUD (non-zero)	13,933	0.094	0.120	0.012	0.034	0.126
FRAUD_ALT	158,302	0.004	0.026	0.000	0.000	0.000
Earnings and Forecast Variables						
REVISION	156,458	-0.217%	0.853%	-0.194%	-0.027%	0.037%
SUE	156,458	0.014%	1.008%	-0.030%	0.038%	0.187%
NEG	156,458	0.180	0.384	0.000	0.000	0.000
Firm Characteristics						
SIZE	165,444	7.388	1.819	6.048	7.278	8.590
MB	165,270	1.839	1.648	0.865	1.315	2.184
LEV	165,444	0.227	0.209	0.034	0.195	0.351
ROA	165,289	0.014	0.043	0.005	0.019	0.034
REVGWTH	160,504	0.152	0.428	-0.022	0.076	0.214

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.

	(1)	(2)	(3)
Variables		REVISION_q	
SUE_q	0.200***	0.144***	0.166***
	(9.13)	(5.87)	(7.44)
IV_q	-0.008***	-0.007***	-0.005***
	(-21.25)	(-14.65)	(-9.54)
$\text{IV}_q \times \text{SUE}_q$	0.120***	0.148***	0.111***
	(4.71)	(5.45)	(4.36)
NEG_q		-0.002***	-0.001***
		(-9.95)	(-6.16)
$\text{NEG}_q \times \text{SUE}_q$		-0.001***	-0.001***
		(-8.03)	(-9.94)
SIZE_{q-1}		0.000**	-0.001***
		(2.13)	(-6.04)
MB_{q-1}		0.001***	0.000***
		(15.50)	(10.74)
LEV_{q-1}		-0.001***	-0.001**
		(-3.87)	(-2.13)
ROA_{q-1}		-0.000	0.006**
		(-0.01)	(2.43)
REVGWTH_{q-1}		0.001***	0.001***
		(9.64)	(7.54)
Observations	148,352	148,352	148,075
Adjusted R-squared	0.190	0.196	0.336
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 in a quarter. $DETECT\_ALT$ is an alternative indicator of restatement announcement based on fraud-related restatements. IV is the 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. 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) DETECT_{q+1}	(2) DETECT_{q+1}	(3) DETECT_{q+1}	(4) DETECT_ALT_{q+1}
IV_q	0.009***	0.012***	0.007**	0.008***
	(5.22)	(5.30)	(2.21)	(3.99)
SIZE_{q-1}		-0.000	0.001	0.002***
		(-0.97)	(0.65)	(2.82)
MB_{q-1}		-0.002***	-0.002***	-0.001***
		(-7.12)	(-4.50)	(-2.69)
LEV_{q-1}		0.008***	0.004	0.003
		(3.86)	(1.05)	(1.24)
ROA_{q-1}		0.005	-0.023	-0.011
		(0.49)	(-1.27)	(-0.94)
REVGWTH_{q-1}		0.002	0.001	0.000
		(1.48)	(1.24)	(0.28)
Observations	165,444	160,259	160,035	160,035
Adjusted R-squared	0.009	0.010	0.020	0.017
Firm FE	No	No	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes	Yes

Table 4: Fraud Amount and Implied Volatility

This table reports the quadratic regression results estimating the relation between fraud amount and implied volatility. $FRAUD$ is the magnitude of accounting restatement in a quarter, scaled by the absolute value of operating income. $FRAUD\_ALT$ is an alternative construct of $FRAUD$ based on fraud-related restatements. IV is the implied volatility. IV^2 is the squared term of IV. 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. 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) FRAUD_q	(2) FRAUD_q	(3) FRAUD_q	(4) FRAUD_ALT_q
IV_q	0.050***	0.049***	0.027***	0.012***
	(8.76)	(7.21)	(4.75)	(3.55)
IV_q^2	-0.026***	-0.025***	-0.013***	-0.005**
	(-7.17)	(-6.08)	(-3.51)	(-2.50)
SIZE_{q-1}		-0.000	0.003***	0.002***
		(-0.23)	(3.25)	(3.05)
MB_{q-1}		-0.000	0.001**	0.001***
		(-0.67)	(2.26)	(3.37)
LEV_{q-1}		-0.000	0.001	0.001
		(-0.11)	(0.33)	(0.29)
ROA_{q-1}		-0.002	-0.044***	-0.019***
		(-0.34)	(-5.12)	(-3.50)
REVGWTH_{q-1}		-0.000	-0.000	0.000
		(-0.25)	(-0.32)	(1.36)
Observations	158,302	153,359	153,114	153,114
Adjusted R-squared	0.024	0.025	0.302	0.331
Firm FE	No	No	Yes	Yes
Year-Quarter FE	Yes	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes	Yes

Table 5: 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. Columns (1) and (2) include year-quarter fixed effects, and columns (3) and (4) 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		$\Delta IV_{q\,to\,q+1}$	
$IVQ1_q$	0.012***	0.029***	0.027***
	(3.84)	(9.41)	(6.86)
$IVQ2_q$	0.005***	0.013***	0.012***
	(3.38)	(8.02)	(6.28)
$IVQ4_q$	-0.007***	-0.017***	-0.016***
	(-2.93)	(-6.48)	(-5.41)
$IVQ5_q$	-0.042***	-0.068***	-0.066***
	(-8.98)	(-13.71)	(-11.77)
$WAVE_q$*$IVQ1_q$			0.006**
			(2.51)
$WAVE_q$*$IVQ2_q$			0.004**
			(2.14)
$WAVE_q$*$IVQ4_q$			-0.008***
			(-2.96)
$WAVE_q$*$IVQ5_q$			-0.021***
			(-3.76)
$WAVE_q$			-0.001
			(-0.45)
$SIZE_{q-1}$	-0.003***	-0.002	-0.000
	(-3.87)	(-1.03)	(-0.17)
MB_{q-1}	-0.000	0.000	0.000
	(-0.69)	(0.32)	(0.52)
LEV_{q-1}	0.006***	0.003	0.004
	(2.79)	(0.72)	(0.90)
ROA_{q-1}	-0.118***	-0.078***	-0.076***
	(-7.32)	(-4.80)	(-4.41)
$REVGWTH_{q-1}$	0.005***	0.003***	0.002**
	(5.30)	(3.39)	(2.23)
Observations	153,450	153,200	124,426
Adjusted R-squared	0.366	0.384	0.418
Firm Fixed Effects	No	Yes	Yes
Year-Qtr Fixed Effects	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes

Appendix I: proofs

Proof. of Lemma 1: For our convenience, we reorganize the first-order condition (16) below:

$$\frac{1-d_t^*}{c\left(1-\delta\rho\right)d_t^*}\frac{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}=1. \tag{33}$$

Note that if $\frac{1-d_t^*}{c(1-\delta\rho)d_t^*}<1$, the left-hand side is always smaller than the right-hand side. Therefore, the manager will never manipulate the report, i.e., $m_t^*=0$. If $\frac{1-d_t^*}{c(1-\delta\rho)d_t^*}\geq 1$, solving the first-order condition gives that

$$m_t^*\left(d_t^*,\Phi_{t-1}\right)=\frac{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2}{q\left(1-q\right)}\left(\frac{1-d_t^*}{c\left(1-\delta\rho\right)d_t^*}-1\right). \tag{34}$$

Combining the two cases proves the lemma. ■

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

$$\begin{aligned}\Phi_t &= var\left(s_t|\mathcal{F}_{t-1}\right)-var\left(E\left[s_t|\mathcal{F}_t\right]|\mathcal{F}_{t-1}\right) \qquad (35)\\ &= var\left(\rho s_{t-1}+\varepsilon_t|\mathcal{F}_{t-1}\right)-var\left(E\left[s_t|r_t,r_{t-1},...\right]|\mathcal{F}_{t-1}\right)\\ &= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2-var\left(\frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}r_t|\mathcal{F}_{t-1}\right)\\ &= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2-\left[\frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}\right]^2 var\left(r_t|\mathcal{F}_{t-1}\right)\\ &= \frac{m_t^*q\left(1-q\right)\left[\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2\right]}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}\\ &= \frac{m_t^*q\left(1-q\right)\left(\rho^2\Phi_{t-1}+\sigma_\varepsilon^2\right)}{\rho^2\Phi_{t-1}+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}.\end{aligned}$$

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

$$\begin{aligned}E\left[s_t|\mathcal{F}_t\right] &= E\left[s_t|\mathcal{F}_{t-1}\right]+\frac{cov\left(r_t,s_t|\mathcal{F}_{t-1}\right)}{var\left(r_t|\mathcal{F}_{t-1}\right)}\left\{r_t-E\left[r_t\right]\right\} \qquad (36)\\ &= E\left[s_t|\mathcal{F}_{t-1}\right]+\frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_\varepsilon^2+m_t^*q\left(1-q\right)}\left\{r_t-E\left[r_t\right]\right\},\end{aligned}$$

where

$$\begin{aligned} var\left(r_t|\mathcal{F}_{t-1}\right) &= var\left(s_t|\mathcal{F}_{t-1}\right) + m_t^* q\left(1-q\right) \\ &= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right), \end{aligned} \tag{37}$$

$$\begin{aligned} cov\left(r_t, s_t|\mathcal{F}_{t-1}\right) &= var\left(s_t|\mathcal{F}_{t-1}\right) \\ &= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2. \end{aligned} \tag{38}$$

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

Proof. of Lemma 3: See the main text. ■

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

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 $\delta = 0$, the regulator's objective function becomes:

$$W\left(\{\Phi_i\}_{i=1}^N\right) = -\sum_{i=1}^{3}(1-d_i)\,\Phi_i' - \frac{\kappa}{2}(\sum_{i=1}^{3} d_i)^2. \tag{39}$$

where

$$\Phi_i' \equiv \left(\rho^2\Phi_i + \sigma_\varepsilon^2\right)\max\left\{1 - \frac{c\,(1-\delta\rho)\,d_i^*}{1-d_i^*}, 0\right\}. \tag{40}$$

Taking the first-order condition gives that

$$\frac{\partial W}{\partial d_i} = \Phi_i' - \kappa(\sum_{i=1}^{3} d_i). \tag{41}$$

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

$$\rho^2\Phi_2 + \sigma_\varepsilon^2 < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\max\left\{1 - \frac{c\,(1-\delta\rho)\,d_1^*}{1-d_1^*}, 0\right\}, \tag{42}$$

that is, Φ_1 is much larger than Φ_2. We will restate condition (42) in terms of exogenous parameters after solving the equilibrium. We now conjecture the equilibrium is that $d_2^* = d_3^* = 0$ and $d_1^* > 0$, where d_1^* solves:

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\max\left\{1 - \frac{c\,(1-\delta\rho)\,d_1^*}{1-d_1^*}, 0\right\} = \kappa d_1^*. \tag{43}$$

To verify that this is indeed an equilibrium, note first that the solution to (43) 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 (43), we can rewrite the condition

(42) as:

$$\kappa d_1^* = \Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\left(1-\delta\rho\right) d_1^*}{1-d_1^*}, 0\right\} > \rho^2 \Phi_2 + \sigma_\varepsilon^2. \tag{44}$$

Since d_1^* is increasing in Φ_1, the condition (42) 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 (42) as

$$\Phi_1 > U\left(\Phi_2\right), \tag{45}$$

where $U\left(.\right)$ is some given increasing function. Finally, we verify that $d_2^* = d_3^* = 0$. This is because, at $d_2 = 0$, the first-order condition for d_2 is always negative, i.e.,

$$\begin{aligned} \frac{\partial W}{\partial d_2} &= \Phi_2' - \kappa d_1^* \qquad (46) \\ &= \rho^2 \Phi_2 + \sigma_\varepsilon^2 - \kappa d_1^* \\ &< \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\left(1-\delta\rho\right) d_1^*}{1-d_1^*}, 0\right\} - \kappa d_1^* \\ &= 0. \end{aligned}$$

The third step uses (42). The last step uses (43).

Second, suppose that $\Phi_1 \leq U\left(\Phi_2\right)$ and

$$\rho^2 \Phi_3 + \sigma_\varepsilon^2 < \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\left(1-\delta\rho\right) d_1^*}{1-d_1^*}, 0\right\}, \tag{47}$$

that is, Φ_1 and Φ_2 are of similar sizes but both are much larger than Φ_3. We will restate condition (47) in terms of exogenous parameters after solving the equilibrium. We now conjecture the equilibrium is that $d_3^* = 0$, $d_1^* > 0$ and $d_2^* > 0$, where the pair of $\{d_1^*, d_2^*\}$ solves:

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\left(1-\delta\rho\right) d_1^*}{1-d_1^*}, 0\right\} = \kappa D^*, \tag{48}$$

$$\Phi_2' = \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\left(1-\delta\rho\right) d_2^*}{1-d_2^*}, 0\right\} = \kappa D^*, \tag{49}$$

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 (47) as:

$$\kappa D^* = \Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1-\delta\rho)\, d_1^*}{1-d_1^*}, 0\right\} > \rho^2 \Phi_3 + \sigma_\varepsilon^2. \tag{50}$$

Since D^* is increasing in Φ_1 and Φ_2, the condition (47) 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 (47) as

$$L\left(\Phi_1, \Phi_2\right) > \Phi_3, \tag{51}$$

where $L\,(.)$ is some given increasing function. Finally, we verify that $d_3^* = 0$. This is because, at $d_3 = 0$, the first-order condition for d_3 is always negative, i.e.,

$$\begin{aligned} \frac{\partial W}{\partial d_3} &= \Phi_3' - \kappa\left(d_1^* + d_2^*\right) \qquad (52) \\ &= \rho^2 \Phi_3 + \sigma_\varepsilon^2 - \kappa\left(d_1^* + d_2^*\right) \\ &< \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1-\delta\rho)\, d_1^*}{1-d_1^*}, 0\right\} - \kappa\left(d_1^* + d_2^*\right) \\ &= 0. \end{aligned}$$

Lastly, suppose that $\Phi_1 \leq U\left(\Phi_2\right)$ 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:

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1-\delta\rho)\, d_1^*}{1-d_1^*}, 0\right\} = \kappa\left(d_1^* + d_2^* + d_3^*\right), \tag{53}$$

$$\Phi_2' = \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1-\delta\rho)\, d_2^*}{1-d_2^*}, 0\right\} = \kappa\left(d_1^* + d_2^* + d_3^*\right), \tag{54}$$

$$\Phi_3' = \left(\rho^2 \Phi_3 + \sigma_\varepsilon^2\right) \max\left\{1 - \frac{c\,(1-\delta\rho)\, d_3^*}{1-d_3^*}, 0\right\} = \kappa\left(d_1^* + d_2^* + d_3^*\right). \tag{55}$$

Appendix III: variable definitions

IV_q: in equation (29), IV_q is the daily implied volatility of the 30-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 30-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 in quarter $q+1$ and zero otherwise.

$DETECT\_ALT_{q+1}$: An alternative construct of $DETECT_{q+1}$, based on restatements that meet 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.

$FRAUD_q$: the dollar amount of restatement's impact on net income in quarter q, scaled by the absolute value of the restated operating income of quarter q.

$FRAUD\_ALT_q$: An alternative construct of $FRAUD_q$, based on restatements that meet 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.

$IVQn_q$: 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 restatement announcements in industry-quarter i, q divided by the number of firms in industry-quarter i, q.