european corporate governance institute

Everlasting Fraud

Law Working Paper N° 617/2021

August 2024

Vivian W. Fang
Indiana University and ECGI

Nan Li
University of Minnesota

Wenyu Wang
Indiana University

Gaoqing Zhang
University of Minnesota

ECGI Working Paper Series in Law

Everlasting Fraud

Working Paper N° 617/2021
August 2024

Vivian W. Fang
Nan Li
Wenyu Wang
Gaoqing Zhang

We are grateful to helpful comments from Jeremy Bertomeu, Philip Bond, Sylvain Chassang, Rebecca Hann, Xu Jiang, Jon Karpoff, Lorien Stice-Lawrence, Ivan Marinovic, Valeri Nikolaev, Hao Xue, Liyan Yang, Anastasia Zakolyukina, Frank Zhou, seminar and conference participants at the Accounting and Economics Society webinar, China International Conference in Finance, Chinese University of Hong Kong, Columbia Burton Accounting Conference, Duke University, FARS meeting, HKUST Accounting Research Symposium, MFA meeting, Rice University, SEC Conference on Financial Market Regulation, University of Maryland, University of Miami, University of Science and Technology of China, and University of Texas, Austin.

Abstract

This paper proposes a new theory to explain the persistence of corporate fraud despite stringent anti-fraud regulations. Our model reveals a “cat-and-mouse” equilibrium within firms, where detection strength optimally matches fraud severity and a “whacka-mole” equilibrium across firms, where regulatory resources are optimally concentrated on the most fraudulent firms. As such, regulations cannot eradicate fraud but synchronize firms’ idiosyncratic fraud levels, contributing to waves. Structural estimation demonstrates the model’s alignment with data. These results hold significant policy implications by highlighting fraud as an enduring risk in the financial markets and the limited efficacy of anti-fraud regulations.

Keywords: Regulation, Corporate Fraud, White-Collar Crime

JEL Classifications: G32, G38, M40, M48

Vivian W. Fang*
Professor
Indiana University
HH 6100 1309 E. 10th St.
Bloomington, IN 47405, USA
e-mail: vfang@iu.edu

Nan Li
Assistant Professor of Accounting
University of Minnesota
321 19th Avenue South
Minneapolis MN 55455, United States
phone: + 1 612-626-8625
e-mail: nanli@umn.edu

Wenyu Wang
Professor of Finance
Indiana University
1309 E. 10th St.
Bloomington, IN 47405, United States
phone: +1 812 855 0189
e-mail: wenywang@indiana.edu

Gaoqing Zhang
Honeywell Professor of Accounting
Carlson School of Management, University of Minnesota
321 19th Avenue South
Minneapolis MN 55455, United States
phone: +1 612 626 3112
e-mail: zhangg@umn.edu

*Corresponding Author

Everlasting Fraud*

Vivian W. Fang
Indiana University, ECGI

Nan Li
University of Minnesota

Wenyu Wang
Indiana University

Gaoqing Zhang
Carnegie Mellon University,
University of Minnesota

July 11, 2024

Abstract

This paper proposes a new theory to explain the persistence of corporate fraud despite stringent anti-fraud regulations. Our model reveals a "cat-and-mouse" equilibrium within firms, where detection strength optimally matches fraud severity and a "whack-a-mole" equilibrium across firms, where regulatory resources are optimally concentrated on the most fraudulent firms. As such, regulations cannot eradicate fraud but synchronize firms' idiosyncratic fraud levels, contributing to waves. Structural estimation demonstrates the model's alignment with data. These results hold significant policy implications by highlighting fraud as an enduring risk in the financial markets and the limited efficacy of anti-fraud regulations.

JEL Classification: G32, G38, M40, M48
Keywords: Regulation, Corporate Fraud, White-Collar Crime

*We are grateful to helpful comments from Jeremy Bertomeu, Philip Bond, Sylvain Chassang, Rebecca Hann, Xu Jiang, Jon Karpoff, Lorien Stice-Lawrence, Ivan Marinovic, Valeri Nikolaev, Hao Xue, Liyan Yang, Anastasia Zakolyukina, Frank Zhou, seminar and conference participants at the Accounting and Economics Society webinar, China International Conference in Finance, Chinese University of Hong Kong, Columbia Burton Accounting Conference, Duke University, FARS meeting, HKUST Accounting Research Symposium, MFA meeting, Rice University, SEC Conference on Financial Market Regulation, University of Maryland, University of Miami, University of Science and Technology of China, and University of Texas, Austin. Please send correspondence to vfang@iu.edu.

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 corporate scandals. Billions of dollars were lost as a result of these financial disasters, which destroyed companies, shook investors' confidence, and ruined people's 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 bolstered efforts to address fraud. The Act, via its Whistleblower Program, empowered the Securities and Exchange Commission (SEC) to reward whistleblowers in unprecedented ways.

Yet, have anti-fraud regulations, even the toughest ones, achieved their stated goals of cracking down on corporate fraud? Empirical studies of fraud history paint a dim picture (Hail, Tahoun, and Wang (2018); Toms (2019)). Notably, Hail et al. (2018) study global corporate fraud over the past two centuries and make two interesting observations. First, while regulation appears effective in tamping down corporate fraud temporarily, fraud always manages to resurface, often in waves. Second, fraud and regulation exhibit interactive dynamics, not only with the former leading the latter but also with the latter leading the former. These observations, particularly the second one, are difficult to explain with either a simple behavioral view that "humans are born greedy" or a rational view that focuses on firms' growth option being the driver of fraud. This paper thus proposes a new theory. Our main contribution is to construct a unified, dynamic model to illustrate the interplay between corporate fraud and regulation. The model shows how regulation effectively curbs fraud as intended, yet fraud continues to persist in financial markets despite 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 financial report to the market after privately observing the firm's economic earnings (or fundamental cash flows).

Based on the report, the market forms rational expectations of the firm's current and future cash flows and estimates firm value. The regulator utilizes a detection technology to inspect the firm's report. With a certain probability, the technology uncovers fraud in 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 boost firm valuation, by weighing his marginal benefit (hereafter MB) and marginal cost (hereafter MC) of committing fraud. The MC is linked to detection likelihood. The MB depends on the extent to which the market values new information and increases with the amount of fraud built to date. As the level of cumulative fraud rises, it introduces greater uncertainty about the firm, prompting investors to place less emphasis on past reports and more on future ones. Consequently, the manager's incentive to inflate forthcoming reports also increases. The regulator decides on the amount of resources spent on a detection technology. In doing so, she seeks to maximize the informativeness of the firm's reports, by weighing her MB and MC of detecting fraud. The MC is also linked to detection likelihood, as a higher likelihood calls for a greater amount of regulatory resources. The MB depends on the degree to which detection reduces uncertainty. This reduction in uncertainty enables investors to base their expectations of firm value on true economic cash flows rather than inflated figures. Notably, catching firm with a higher level of cumulative fraud alleviates more information uncertainty in both current and future periods.

Analyses of this single-firm model begin to tell why regulation can be effective at curbing fraud but fraud never ceases to exist. This is because, although the manager and the regulator each solve a maximization problem independently, their calculus are intertwined in that their MBs and MCs both critically depend on—in fact, co-move with—"the cumulative fraud-induced information uncertainty," or "Φ_t," the key state variable featured in our model.

Starting with the regulator, her goal is to minimize information uncertainty for investors. In pursuit of this goal, she faces a trade-off: a higher Φ_t motivates her because it enhances detection's value in restoring information precision, but increasing detection strength requires more regulatory resources. This trade-off is supported by Section I's motivating evidence, where an empirical proxy for Φ_t—the implied volatility of standard options—plummets upon restatement announcement, indicating that cumulative fraud raises information uncertainty while detection lowers it. Turning to the manager, his aim is to boost firm valuation for personal gain. He similarly faces a trade-off: while a higher Φ_t incentivizes him because investors eager for new information will value even the biased report, it also serves as a constraint because the regulator may increase detection likelihood in response. The cost side of this trade-off is evident, and the benefit side is also supported by prior literature (e.g., Fischer and Verrecchia (2000); Fang et al. (2017)).

In equilibrium, the regulator chooses the optimal level of detection likelihood (by allocating corresponding resources), anticipating the level of information uncertainty due to the manager's cumulative fraud, and vice versa. If the regulator anticipates a low level of information uncertainty, then she would allocate fewer resources to detection, allowing the manager to persist in committing fraud as the manipulation benefits likely outweigh the costs. As fraud accumulates, it results in greater information uncertainty, incentivizing further fraudulent behavior by the manager, yet also prompting the regulator to increase detection spending in response. The two effects go hand-in-hand, simultaneously increasing the manager's benefits and costs of manipulation. When fraud reaches a critical level, the regulator concentrates resources on the firm, tilting the balance so that the costs of continuing fraud eventually dominate the benefits. Upon detection, fraud is cleared in the firm, and the cycle repeats. These analyses point to a non-monotonic relation between Φ_t and the optimal amount of fraud chosen by the manager in a period. This relation gains support from data: in Section I, we document an inverse U-shaped association between a firm's fraud amount in a quarter and the level of implied volatility of the quarter. These analyses also explain the time-series persistence of fraud because the interdependent nature of fraud and regulation essentially results in a cat-and-mouse equilibrium within-firm because the strength of detection optimally matches the severity of fraud.

Analyses of an expanded, three-firm model make a separate case for everlasting fraud despite effective regulation. Let H-, M-, and L-firm represent the firm with a high, medium, and low level of cumulative fraud-induced information uncertainty, respectively. As in the single-firm model, detection strength matches fraud severity in equilibrium, so the regulator rationally allocates most resources towards H-firm. Ironically, with the regulator concentrating detection efforts on H-firm, a whack-a-mole equilibrium emerges in which M- and L-firms may factor in the regulator's decision and become more aggressive because their fraudulent activities are better veiled until H-firm is caught (upon which M-firm becomes next target in line). These analyses explain the cross-sectional persistence of fraud across-firm.

As such, analyses of the multi-firm model offer a nuanced perspective on the efficacy of anti-fraud regulations. On the one hand, these regulations succeed in cracking down on fraud in the most severe cases, exemplified by H-firm, by diminishing its incentives for fraudulent behavior. Before H-firm is detected, concentration of regulatory resources on the firm greatly increases its costs of committing fraud. Once detected, the benefits become minuscule because a sharply declined uncertainty renders the firm's future reports less useful and fraudulent reporting less valuable. On the other hand, the rational allocation of regulatory resources towards the most fraudulent firm (H-firm) diverts attention away from less fraudulent firms (M- and L- firms), allowing their misconduct to remain undetected and potentially catch up to levels seen in more heavily regulated counterparts. This phenomenon, illustrated by our three-firm model, suggests that despite attempts to combat fraud, anti-fraud regulations cannot eradicate it. Rather, they synchronize firms' amount of cumulative fraud, which may otherwise vary, and induce corporate fraud waves.

It is important to clarify that fraud persists in our model not because the regulator necessarily faces high detection costs, but because the regulator's decision to detect fraud is endogenously derived. As described above for both the single- and multi-firm models, the regulator will optimally choose not to detect fraud when the expected level of cumulative fraud is low, due to the low MB, even if the MC is also low. Additionally, persistent fraud need not imply managerial myopia, as managers in our model rationally maximize the sum of expected discounted future earnings net of discounted manipulation costs.

To assess our model's real-world relevance, we conduct a structural estimation using data from 1999 to 2017. Results show that the model-implied moments align well with the empirical counterparts in terms of the frequency of detection (with an average rate of 0.7% per quarter for firms in the lowest tercile of uncertainty and a higher rate of 0.9% per quarter for firms in higher terciles), cumulative fraud levels upon detection (amounting to 11% of total assets), and the duration of uncertainty rebound post-detection (averaging 5.9 quarters). Additionally, parameter estimates suggest that managers face substantial regulatory penalties, nearly three times the size of detected fraud, and also incur significant personal costs for misconduct (e.g., psychic costs and reputation cost), even when fraud remains undetected. In contrast, regulators face relatively low detection costs but focus detection efforts on firms with the highest information uncertainty, reinforcing the model's intuition that persistent fraud is not purely driven by high regulatory costs. Intriguingly, despite the estimation revealing high manipulation costs for managers and low detection costs for regulators—factors that would typically discourage fraud—the model indicates persistent fraud, as we also observe in the data.[1] Lastly, our estimates corroborate observed fraud waves in the data, as reduced regulatory oversight following detection prompts firms to resume fraudulent activities, leading to a convergence of fraud levels across firms within approximately six quarters.

This paper aligns with economic theories highlighting the limitations of using maximal penalties to deter crime. To list a few, Mookherjee and Png (1992) show that enforcement authorities should optimally vary their monitoring efforts based on signals from potential offenders. Bond and Hagerty (2010) show that marginal penalties are more effective in a Pareto inferior crime wave equilibrium. Our results contribute to this discussion through a distinct mechanism. Unlike violent crimes, where benefits are often exogenous, fraud

[1] The finding that managers face high manipulation costs is consistent with prior structural work including Zakolyukina (2018), Bird et al. (2019), Bertomeu et al. (2021), Terry (2023), and Cheynel et al. (2024).

is a calculated decision with endogenous economic benefits and costs. Our model reveals a key insight about corporate fraud: its MB and MC are intrinsically linked. Therefore, policies that align punishment with the crime work uniquely well in addressing fraud. Once regulations are sufficiently stringent and target the most fraudulent firms, these firms' MBs of committing fraud drop sharply upon detection, allowing regulators to optimally decrease enforcement levels. Thus, fraud persists not due to high regulatory costs or ineffective regulations but because enforcement is optimally focused on the most severe fraud cases. Adding to this point, our structural estimation analysis furthers realistic model parameter estimates and sheds light on two results crucial to the model implications: detection strength does not appear maximized despite regulators facing relatively low costs, and fraud persists despite low detection costs for regulators and high total manipulation costs for managers.

Our model furthers evidence for the "whack-a-mole" phenomenon, a term coined in prior studies to describe the situation where a regulation aimed at resolving issues in one market inadvertently leads to emerging challenges in others (Blinder (2008); Blinder (2015); Cai, He, Jiang, and Xiong (2021)). In our model, this scenario arises because maximizing detection intensity at all times is not optimal, even if it is not costly, leading to inevitable strategic reporting behavior in certain sectors of the economy. Moreover, fraud waves in our model are interactive outcomes between managers and monitors; this dynamic mirrors analogous patterns observed in other markets, such as takeover waves (Pound (1992); Harford (2005)) and politics (Bond (2008)).[2]

Additionally, our model both complements and diverges from existing fraud models in notable aspects. First, to our best knowledge, our model is the first to study the joint mechanisms of fraud and regulation in a multi-firm, dynamic setting. Prior theories predominantly focus on modeling the fraud decisions of individual firms, assuming exogenous detection costs. Among these, most study single-period settings (e.g., Fischer and Verrecchia

[2]In a broader context, Varas et al. (2020), Marinovic and Szydlowski (2022), and Marinovic and Szydlowski (2023) build dynamic models to explore the interplay between a monitor and an agent susceptible to misbehavior (moral hazard), with a focus on reputation concerns. Unlike our study, these papers do not investigate managers' financial reporting decisions and are confined to single-firm settings.

(2000), Dye and Sridhar (2004), and Guttman et al. (2006)), while others explore two-period settings with minimal dynamic features (e.g., Sankar and Subramanyam (2001), Kirschenheiter and Melumad (2002)). While a few models extend to incorporate multiple firms, they still operate under the assumption of exogenous detection costs (e.g., Heinle and Verrecchia (2016), Aghamolla et al. (2021)). Two models stand out by endogenizing detection costs: Povel et al. (2007) examine a single-period setting and Beyer et al. (2019) delve into dynamic games. In contrast to these works, our paper combines dynamic fraud decisions with dynamic monitoring. This holistic perspective allows us to analyze the evolution of corporate fraud, while offering a nuanced evaluation of the effectiveness of anti-fraud regulations.[3]

Lastly, several theories highlight growth as an important driver for fraud (e.g., Povel et al. (2007), Strobl (2013)), which explain why corporate frauds vary with business cycles. However, these theories cannot explain why fraud waves occur during both economic boom and bust and why regulations lead or even give rise to fraud waves as observed by Hail et al. (2018). Our model takes the informative perspective and offers a new explanation for the observed interactive dynamics between fraud and regulation, while setting aside firms' idiosyncratic characteristics including their growth options. The information effect of firms' disclosure decisions is crucial, given that information asymmetry has the potential to distort investment and hurt the real economy (see Stein (1989) for a seminal theory and Goldstein and Yang (2017) for a survey of more recent work).

[3]Ye (2023) surveys the theoretical literature on managerial reporting discretion and its interaction with auditors or other internal control measures. Our model deviates from this literature in two ways. First, auditors in this literature are modeled as separate economic agents maximizing their own utilities without internalizing social surplus loss (e.g., Antle (1982), Antle (1984)), while the regulator in our model is dedicated to maximizing social surplus, defined as the aggregate informativeness of corporate reports. Second, even in this literature, multi-firm, dynamic models with endogenize monitoring are rare. Notable attempts include Balachandran and Ramanan (1988), who study optimal allocations of internal control effort in a multi-period model, and Gao and Zhang (2019), who explore endogenous internal control choices in a multi-firm setting.

1 Motivating Evidence

Fraud has persisted throughout recorded history, with the earliest documented case dating back to Ancient Greece in 300 B.C. (Toms (2019)). Although the history of financial markets is shorter in comparison, it has never lacked episodes of corporate fraud. Using a comprehensive worldwide sample from 1800 to 2015, Hail, Tahoun, and Wang (2018) conduct a descriptive study of the intertemporal relation of corporate fraud and regulation. The study reveals a pattern of interactive dynamics, as seen in Figure 1 reproduced using data from Hail, Tahoun, and Wang (2018, pp. 645). As illustrated, fraud often precedes regulatory interventions over extended periods, which is unsurprising given regulators' reactive nature. Interestingly, regulation is also positively related to the incidence of future scandals. This pattern is less anticipated so the authors formulate three conjectures. First, regulators may fail to be fully effective, because they are uninformed, self-interested, captured, or ideological. Second, new anti-fraud measures may be insufficient or contain loopholes that allow managers to adapt and circumvent regulation. Third, regulations can have unintended consequences, although these outcomes remain unclear. The patterns revealed by Hail et al. (2018), which are hard to reconcile with existing theories, motivate our work. Specifically, we aim to build a rational model to explain the interactive dynamics between fraud and regulation and the unintended consequences of regulation that contribute to such dynamics.

In building the model, we take an information perspective, driven by regulators' general consensus that fraud in firms' disclosure and reporting can sway market valuation and deceive investors.[4] To assess the alignment of our proposed theory with real-world practices, we first present motivating evidence on the role of information in the decision-making processes of regulators and firm managers. Detailed discussions of sample selection and variable definitions used to generate this evidence can be found in Appendix B.

[4]For example, the Financial Accounting Standards Board (FASB) states that the objective of financial reporting is to "provide financial information about the reporting entity that is useful to existing and potential investors, lenders, and other creditors in making decisions about providing resources to the entity." (SFAC No. 8, 2010, p.1).

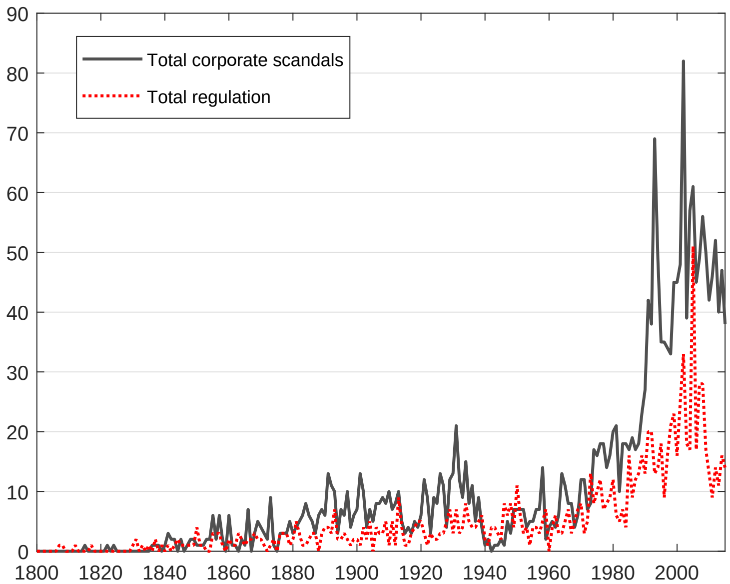

Figure 1: Time-series plot of the yearly number of business news mentions of "scandal" and "regulator" for 24 countries from 1800 to 2015, reproduced from Figure 2 of Hail et al. (2018)

Starting with the regulator, if her goal is to crack down on fraud and restore the accuracy of corporate reports, then it is reasonable to expect that fraud increases information uncertainty while detection decreases it. To examine this, we collect data on restatements announced between 1999 and 2017 from Audit Analytics and implied volatility data from Option Metrics. Since option prices reflect the market's expectation about changes in the firm's value given available information, implied volatility reflects the conditional variance of this information set. Therefore, we expect implied volatility to increase with Φ_t, the key state variable in our model that represents the level of cumulative fraud-induced information uncertainty. Empirically, we calculate the monthly (quarterly) implied volatility from 90-day standardized option prices by averaging the daily values over a given month (quarter) and label the resulting variable monthly (quarterly) IV, respectively. A plot of IV, shown in Figure 2, reveals that implied volatility sharply drops upon revelation of fraud, consistent with the expectation that fraud increases information uncertainty while detection decreases it. Admittedly, IV may reflect factors unrelated to cumulative fraud; we discuss and address this challenge more explicitly in the structural estimation of our model.

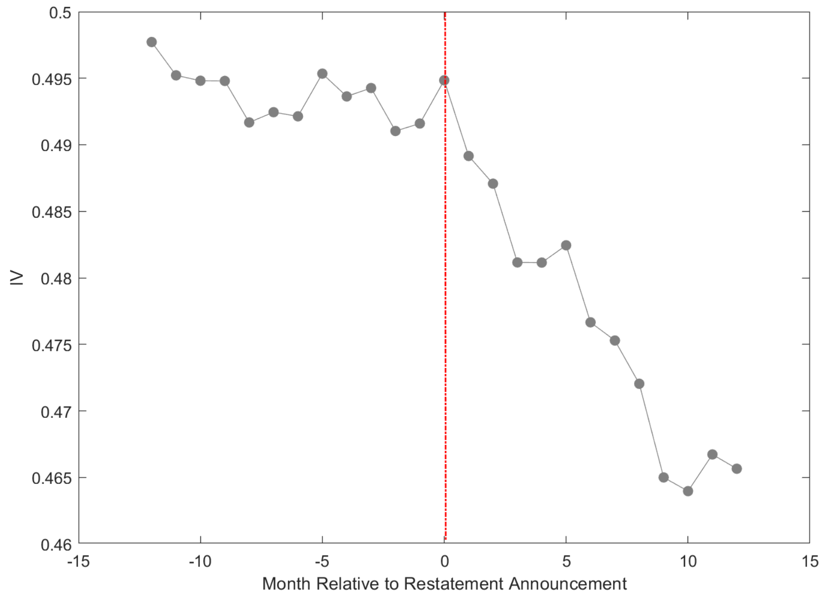

Figure 2: Plot of monthly IV before and after fraud-related restatement announcements

Turning to the manager, the earlier discussion suggests the manager faces a trade-off when it comes to committing fraud: while a higher level of fraud-induced information uncertainty dissuades the manager from engaging in fraudulent activities due to heightened detection likelihood by the regulator, it also provides a stronger incentive because investors are keen for new information. In line with this reasoning, we observe an inverse U-shaped relationship between a firm's quarterly fraud amount (labeled $FRAUD$) and the level of implied volatility within the same quarter (IV), as depicted in Figure 3.

The motivating evidence presented in this section is suggestive of the role that information plays in the calculus of both regulators and firm managers. We are unaware of a pre-existing model that can explain such evidence coherently, which motivates the development of our model. Many of the insights gleaned here gain support from a more rigorous structural estimation of our model in Section 4.

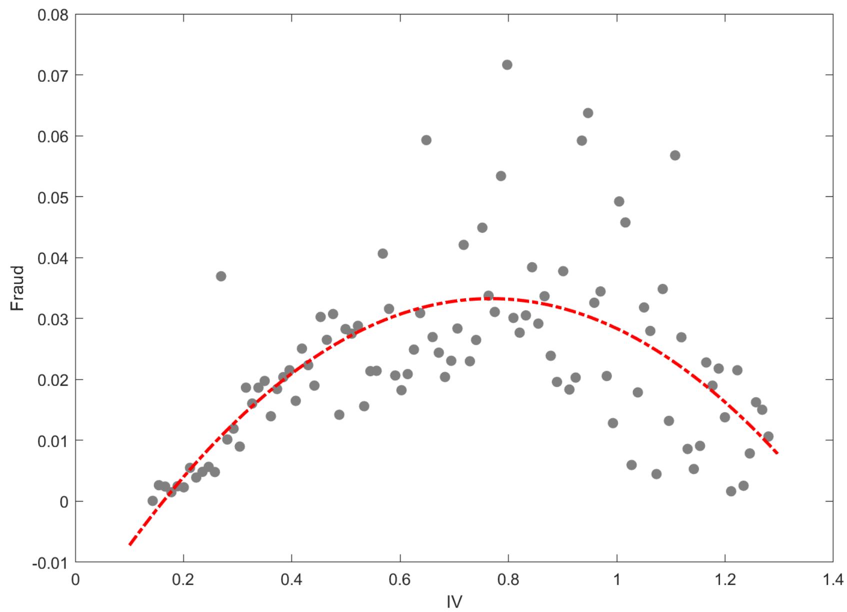

Figure 3: Plot of $FRAUD$ against IV with prediction curve

2 Single-firm Model

2.1 Model Setup

2.1.1 Timing of events

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 $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. In each period, the firm manager privately learns the realization of the firm's economic earnings s_t and issues a report r_t. Investors use the report to update V_t, which is assumed to be set by a competitive market and equals the firm's total discounted future earnings in expectation:

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

where $E^I \left[\cdot | \mathcal{F}_t \right]$ denotes the investors' expectation, $\mathcal{F}_t \equiv \{ r_t, r_{t-1}, ...r_1 \}$ denotes the set of the firm's reports up to time t, and $\delta \in (0, 1)$ denotes the discounting factor. To be clear, since V_t reflects the market's valuation of the firm's economic earnings (or *expected* firm value), it may differ from the firm's fundamental value. The manager has incentives to manipulate the report r_t to boost V_t, because a higher firm valuation typically means higher equity compensation and better career prospects for himself.

We model the manager's earnings manipulation decision following Gao and Zhang (2019). In each period t, after observing the true economic earnings s_t, the manager chooses manipulation $m_t \geq 0$ that generates the following report r_t:

$$r_t = s_t + m_t(1 - q) + \sqrt{m_t q \left(1 - q \right)} \eta_t. \tag{3}$$

The choice of manipulation m_t is observable only to the manager. $\eta_t \sim N\left(0, 1\right)$ is a standard normal random variable that is independent of all other variables in the model. Equation (3) suggests that manipulation has a dual effect on the report: m_t increases the mean of the report but decreases its precision.

Gao and Zhang (2019) show that the generating process of r_t in Equation (3) can be micro-founded as follows. In each period t, upon observing the true economic earnings s_t, the manager chooses manipulation m_t that adds m_t errors $\{\xi_l\}_{l=1}^{m_t}$ to s_t. Each error generates either 0 or 1 with $\Pr(\xi_l = 0) = q \in (0, 1]$. The report is then given by:[5]

$$r_t = s_t + \sum_{l=1}^{m_t} \xi_l. \tag{4}$$

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

$$\sum_{l=1}^{m_t} \xi_l \sim N\left(m_t\left(1-q\right), m_t q\left(1-q\right)\right). \tag{5}$$

Denoting $\eta_t \equiv \left(\sum_{l=1}^{m_t} \xi_l - m_t\left(1-q\right)\right) / \sqrt{m_t q\left(1-q\right)}$, the report becomes Equation (3).

We assume that, in each period t, the firm's fraudulent activity is uncovered by the regulator with a probability of $d_t \in (0, 1)$. Specifically, the regulator influences d_t by utilizing a detection technology to inspect the manager's report r_t; the technology consumes regulatory resources of $\frac{\kappa}{2}\left(d_t\right)^2$. If the regulator successfully detects fraud, she would require the manager to restate the report to equal the true earnings, i.e., $r_t = s_t$, and impose a penalty C_t on the manager that is proportional to the fraudulent amount:

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

where the coefficient $c > 0$, which is assumed to be not too excessive so that the equilibrium manipulation $m_t^* > 0$.

Managers may also incur some personal cost in manipulating the report (e.g., reputation concern, psychic cost, etc.). Accordingly, we assume that the manager incurs a personal manipulation cost proportional to the manipulation amount, i.e., $\mu(r_t - s_t)$, where $\mu > 0$ and is not too large so that the equilibrium manipulation $m_t^* > 0$. A key difference between the penalty C_t and the personal cost is that the penalty is imposed only when a manager is caught

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

by the regulator and thus its efficacy depends on the probability of fraud detection, which we show below is endogenous in our model, while the personal cost is incurred regardless of detection and therefore is taken as an exogenous cost.

We incorporate the personal cost into our model for two reasons. First, previous studies have shown it is an important component of the manipulation cost, because it discourages managers from committing fraud. Second, leaving it out of the model may bias the model towards finding an equilibrium with everlasting fraud. We therefore incorporate the personal cost into both the model and the estimation. We demonstrate that even with a significant estimated personal cost of manipulation, fraud still appears everlasting in equilibrium.

To model the regulator's detection decision, we assume that the regulator sets the detection probability d_t to maximize the informativeness of the set of reports $\mathcal{F}_t \equiv \{r_t, r_{t-1}, ...\}$ about the firm value V_t. This assumption reflects the regulator's objective in ensuring the integrity of financial reporting as discussed earlier. Since earnings follow an AR(1) process, the period-t earnings s_t is a sufficient statistic to estimate all of the firm's future earnings and the firm value. Thus, maximizing the informativeness of $\mathcal{F}_t$ is equivalent to maximizing the informativeness about s_t, or minimizing the conditional variance about s_t:[6]

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

Figure 4 summarizes the timing of events in each period t.

2.1.2 Discussion of assumptions

In this section, we discuss and motivate several assumptions in our model. First, for tractability, we have abstracted away from the accounting property that manipulation in the

[6]Our model focuses on Φ_t, the level of cumulative fraud-induced information uncertainty rather than the level of cumulative fraud per se. This simplifying approach not only allows for tractability but also reflects an insight from prior theories that the market is able to unravel the expected mean of the bias in the report but not necessarily the fraud-induced information uncertainty (see Stein (1989); Fischer and Verrecchia (2000); Dye and Sridhar (2004)). Nevertheless, since fraud and fraud-induced information uncertainty are positively correlated in our model, our inferences are likely qualitatively similar with either concept.

Figure 4: Timeline of the period-t game

report can reverse over time, i.e., the reversal of discretionary accruals (e.g., Beyer, Guttman, and Marinovic (2019)). However, interestingly, because the regulator's equilibrium detection effort increases with the cumulative fraud, a firm who has engaged in more manipulation in the past will have stronger incentive to reduce manipulation in the future, i.e., the manipulated accruals will reverse endogenously in equilibrium due to the threat of increasing regulatory detection.

Second, we assume that the regulator focuses on detecting and penalizing the manager's current manipulation instead of the past or cumulative manipulation. Two points are in order. First, note that it yields the largest information gain for the regulator to focus on detecting fraud in the current period's report r_t and uncovering the true earnings s_t, because s_t is a sufficient statistic for estimating all of the firm's future earnings (as shown in equation (2)). Conditional on the revelation of s_t, detecting fraud in the firm's past reports, $\{r_{t-1}, r_{t-2}, ...\}$, incurs additional costs but does not generate any incremental information benefits. Second, if one assumes alternatively that the regulator can uncover and impose a penalty on past fraud or cumulative fraud, the manager would have weaker incentives to commit fraud. However, in terms of economic forces, this modification will yield similar effects to imposing a greater penalty on the current-period fraud, i.e., a higher c.

2.2 Analysis

In this section, we analyze the manager's optimal manipulation choice m_t^* and the regulator's equilibrium detection choice d_t^*.[7] For brevity, we present only the equations that

[7] 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

illustrate the key intuitions from the model, leaving the detailed derivations to Appendix A.

2.2.1 The manager's problem

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

$$U_t = \max_{m_t} E^M \left[\sum_{k=0}^{\infty} \delta^k u_{t+k} \middle| s_t, \mathcal{F}_{t-1} \right], \tag{8}$$

where $E^M [\cdot|s_t, \mathcal{F}_{t-1}]$ denotes the manager's expected utility in period t based on his information set, which includes s_t, his privately observed true earnings of the firm for the period, and $\mathcal{F}_{t-1}$, the firm's publicly released earnings reports in the past up until period $t-1$. The manager's period-t payoff is:

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

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

The first two terms of equation (9) represent the manager's utility under two different scenarios, respectively. In the first scenario that occurs with a probability of d_t^*, the manager's fraudulent behavior is detected. As a result, the firm's true earnings s_t is revealed to investors, who would then update its valuation of the firm to $\frac{s_t}{1-\delta\rho}$ based on s_t. The manager suffers a penalty of C_t proportional to m_t as shown in equation (6). In the second scenario that occurs with the complement probability of $1-d_t^*$, the manager's fraudulent behavior

thus cannot make d_t a function of m_t. In addition, the regulator's detection choice in equilibrium will be independent of the reported earnings r_t, even though the regulator observes r_t at the time of choosing the detection effort. The reason is that, the regulator sets the detection probability to maximize the reporting informativeness (or equivalently, minimize the conditional variance about firms' economic earnings). Under the assumption of normal distributions, the conditional variance is independent of the realization of the reported earnings. Accordingly, the regulator's detection choice is independent of the reported earnings.

goes undetected. Thus, the firm's true earnings s_t remains unknown to investors, who would then have to set its valuation of the firm to $\frac{E^I[s_t|\mathcal{F}_t]}{1-\delta\rho}$ based on the firm's public reports $\mathcal{F}_t$. The manager incurs no penalty.

Equation (9) also makes it clear that, except for the personal manipulation cost, $\mu(r_t - s_t)$, the manager's manipulation choice m_t only affects firm valuation when it is undetected. To solve the optimal m_t^*, we first analyze how firm valuation varies with m_t in each period:

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

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

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

captures the value relevance of the earnings report, with Φ_{t-1} being the inverse precision of the prior, as defined in equation (7). When Φ_{t-1} is larger, the prior is less precise and so the investors have to place a greater weight on the current report to infer firm fundamentals.[8]

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

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

valuation in the next period $E^I[s_{t+1}|\mathcal{F}_{t+1}]$; this effect works through the first term of the equation. In fact, such bias propagates to all future s_{t+k} for $k > 0$ through the recursive form of equation (10).[9]

Assuming that the firm remains undetected, we can summarize the contemporaneous effect ($k = 0$) and the intertemporal effect ($k > 0$) of m_t below as

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

Now we solve the manager's optimal choice of manipulation, m_t^*. Taking derivative of U_t in equation (8) with respect to m_t and then substituting in equation (12) derived above, we obtain the first-order condition (F.O.C.) as[10]

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

The MC of manipulation, expressed on the left hand side (LHS) of the F.O.C., increases with the regulator's optimal choice of detection probability d_t^*, which is correctly conjectured by the manager. The MB of manipulation, expressed on the right hand side (RHS) of the

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

[10]The F.O.C. implies that, given that the market rationally conjectures the equilibrium manipulation level m_t^*, the MB of manipulation equals the MC of manipulation such that the manager (weakly) prefers the equilibrium manipulation choice m_t^*. Hence m_t^* constitutes a rational-expectation equilibrium in which the manager's equilibrium choice is consistent with the market conjecture. Note that this equilibrium is also the unique rational-expectation equilibrium in our model. We briefly outline the argument here and the detailed analysis is available upon requests. Assume by contradiction that there exists another equilibrium $m_t^{**} \neq m_t^*$. Given the market conjecture m_t^{**}, the MB of manipulation does not equal the MC of manipulation. Accordingly, the manager can only choose a corner solution. Either $m_t^{**} = 0$ if the MB is smaller than the MC, or $m_t^{**} = \infty$ if the MB is larger than the MC. When the manipulation cost c is not too large as assumed earlier, the MB is greater than the MC under the conjecture of $m_t^{**} = 0$. Hence $m_t^{**} = 0$ is not an equilibrium. On the other hand, given the conjecture $m_t^{**} = \infty$, the market believes that the report contains no information and places no weight on the report r_t. Therefore, the MB of manipulation is zero and smaller than the MC, which implies that $m_t^{**} = \infty$ is not an equilibrium.

F.O.C., arises from both the contemporaneous effect and the intertemporal effect discussed above, taking into account the expected likelihood of being detected in the current or a future period. Specifically, the MB from the contemporaneous effect is only affected by the likelihood of not being detected in the current period $(1 - d_t^*)$, while the MB from the intertemporal effect is affected by the likelihood of not being detected up to a future period of interest $\left[\prod_{\ell=0}^{k}\left(1 - d_{t+\ell}^*\right)\right]$.

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

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

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

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

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

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

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

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

2.2.2 The regulator's problem

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

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

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

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

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

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

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

F.O.C. as

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

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

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

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

where Φ_t is expressed recursively in equation (7).

2.2.3 The equilibrium

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

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

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

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

where

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

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

The dynamic programming problem in Proposition 1 does not permit a fully analytical solution, so we take two steps to analyze the model's implications. First, we demonstrate the model's key insights by analytically solving a special case of the model when the discounting factor δ is set to zero. This simplified case represents a situation in which the manager and the regulator are impatient and care only about the contemporaneous effects of their choices. We then confirm that the key insights in the simplified case with $\delta = 0$ generalize to the full model with $\delta > 0$. In doing so, we resort to the numerical method detailed in Appendix A to solve the full model.

When $\delta = 0$, all intertemporal effects vanish and only the contemporaneous effect remains. The F.O.C. of the manipulation decision m_t^* in equation (13) is reduced into:

$$\underbrace{c(1-q)\,d^* + (1-q)\mu}_{\text{MC of } m} = \underbrace{(1-d^*)(1-q)\frac{\rho^2\Phi + \sigma_\varepsilon^2}{\rho^2\Phi + \sigma_\varepsilon^2 + m^*q(1-q)}}_{\text{MB of } m \text{ from the contemporaneous effect}}. \tag{24}$$

The MC of manipulation is the same as in the full model whereas the MB of manipulation includes only the contemporaneous effect of manipulation in boosting the firm valuation in the current period $E^I[s_t|\mathcal{F}_t]$. Therefore, the MB of manipulation in our special case is qualitatively similar to that in the full model but quantitatively smaller as manipulation in the full model also generates an additional intertemporal benefit in boosting the future firm value. Hence, the equilibrium properties of manipulation m^* in the special case extend qualitatively to the full model, as suggested by our later numerical simulations. Solving equation (24) gives the equilibrium manipulation m^* as a function of the fraud-induced uncertainty Φ and the equilibrium detection strength d^*:

$$m^* = \frac{\rho^2\Phi + \sigma_\varepsilon^2}{q(1-q)}\left(\frac{1-d^*}{cd^*+\mu} - 1\right). \tag{25}$$

Equation (25) illustrates some equilibrium properties of manipulation m^*. First, holding d^* constant, m^* is increasing in the level of fraud-induced uncertainty 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. Second, m^* is decreasing in the detection strength d^* as a higher d^* increases the MC of manipulation and deters manipulation in equilibrium.

Substituting equation (15) and equation (25) into the F.O.C. of the detection choice in equation (18) gives the optimal detection choice d^*:

$$\underbrace{\kappa d^*}_{\text{MC of } d} = \underbrace{\left(\rho^2\Phi + \sigma_\varepsilon^2\right)\left(1 - \frac{cd^*+\mu}{1-d^*}\right)}_{\text{MB of } d \text{ from the contemporaneous effect}}. \tag{26}$$

Compared to the first-order condition (equation (18)) in the full model, the MC of detection in the special case is the same whereas the MB of detection includes only the contemporaneous effect of detection in clearing the uncertainty about s_t in the current period. Hence, the MB of detection in the special case is qualitatively similar to that in the full model but quantitatively smaller as detection in the full model also generates an additional intertemporal benefit in reducing prior uncertainty for all future periods. In particular, applying the implicit function theorem to equation (26) shows that the regulator's choice of optimal detection strength d^* is increasing in the fraud-induced uncertainty Φ.[11] That is, the regulator matches the strength of fraud detection with the severity of fraud in equilibrium. This result is intuitive because the manager's manipulation in the past adds noises to the firm's reports and decreases informativeness. Since the regulator's objective is to clear fraud and restore informativeness of the firm's reports, her gains are higher from detecting reports with more extensive fraud. In other words, the regulator's MB of detection is increasing in the cumulative level of fraud and so is her choice of optimal detection strength.

Next, we use the expression of d^* in equation (26) to draw inferences about the properties of the equilibrium manipulation $m^*(\Phi)$ in equation (25). Interestingly, we find that m^* can be non-monotonic in Φ. To see this, recall that from equation (25), fixing the regulator's detection choice, m^* is increasing in Φ. However, equation (25) also suggests that as Φ increases, the regulator would invest more heavily in the detection technology, which deters manipulation and reduces m^*. The two countervailing effects go hand-in-hand, leading to a non-monotonic relation between m^* and Φ (which is consistent with the evidence presented in Figure 3). The following proposition summarizes the equilibrium manipulation choice and detection strength $\{m^*(\Phi), d^*(\Phi)\}$ in our special case of $\delta = 0$.

Proposition 2 *Consider a special case of our model with the discount factor $\delta = 0$. 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, with the two agents rationally anticipating each other's optimal policy function:*

$$\kappa d^*(\Phi) = \left(\rho^2\Phi + \sigma_\varepsilon^2\right)\left(1 - \frac{cd^*(\Phi) + \mu}{1 - d^*(\Phi)}\right), \tag{27}$$

$$m^*(\Phi) = \frac{\rho^2\Phi + \sigma_\varepsilon^2}{q(1-q)}\left(\frac{1 - d^*(\Phi)}{cd^*(\Phi) + \mu} - 1\right). \tag{28}$$

[11] The detailed analysis is included in equation (A.10) of Appendix A.5.

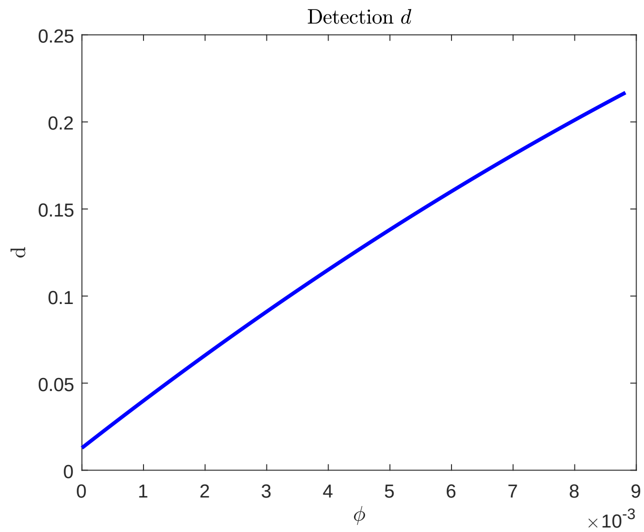

Figure 5: Equilibrium detection probability $d^*(\Phi)$. This figure plots the regulator's optimal choice of detection intensity, d, as a function of the information uncertainty induced by cumulative fraud, Φ. The parameters used in generating this figure are estimated parameters reported in Table 1.

The equilibrium detection $d^*(\Phi)$ is increasing in Φ. The equilibrium manipulation $m^*(\Phi)$ can be non-monotonic in Φ.

The analytical results in Proposition 2 build on a special case of the model solution with $\delta = 0$. Next, we solve the full model numerically to verify our findings. Since we estimate the model in Section 4, we demonstrate the numerical results using the estimated model parameters reported in Table 1: the subjective discount rate, δ, equals 0.974 per quarter, equivalent to 0.9 per annum; the success rate of manipulation, q, equals 0.5; the persistence of the AR(1) process that governs the dynamics of the economic earnings s_t, ρ, equals 0.7; the conditional standard deviation of the AR(1) process, σ_ε, equals 0.015; the detection cost parameter, κ, equals 0.03; the parameter of penalty on fraud, c, equals 2.7; and the personal cost of manipulation, μ, equals 0.37.

Figure 5 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 in the special case with $\delta = 0$, the numerical solution suggests that a higher level of cumulative fraud increases the regulator's choice of detection intensity, which in turn increases the manager's MC of manipulation.

Figure 6 depicts the manager's optimal manipulation m_t^* also as a function of the state variable Φ_{t-1}. Unlike the detection probability d that always increases with Φ_{t-1}, m_t^* is not necessarily monotonic in Φ_{t-1}, and many parameters can drive the relation between these two variables. One interesting example is that this relation depends on the *composition* of the marginal cost in manipulation.

The red line in Figure 6 shows how m_t^* varies with Φ_{t-1} when the penalty on fraud, captured by c, is more dominant compared with the personal cost, captured by μ. In this case, the marginal cost of manipulation is mainly driven by the threat of detection and the resulting penalty: when Φ_{t-1} is low (and thus the detection probability is low), the manager is incentivized to manipulate more aggressively. As Φ_{t-1} becomes high (and thus the detection probability rises significantly), he faces a much higher marginal cost of manipulation and cuts back manipulation. This gives rise to a hump-shaped relation between m_t^* and Φ_{t-1}.

The blue line, on the other hand, depicts the relation between m_t^* and Φ_{t-1} when c is reduced while μ increases. In this case, the marginal cost of manipulation is largely driven by the exogenous personal cost and is thus less sensitive to Φ_{t-1}. Meanwhile, the marginal benefit of manipulation rises with Φ_{t-1}, leading to a monotonically increasing relation between m_t^* and Φ_{t-1}.

Having characterized the equilibrium policies of manipulation and detection decisions with a set of fixed parameters, we next analyze how the equilibrium decisions vary with these parameters through the lens of comparative statics. Figure 7 illustrates how varying the weight on the intertemporal effects (i.e., the discounting factor δ) shifts the equilibrium outcomes. Intuitively, when δ increases, the manager anticipates a higher MB of manipulation, as inflating the current period report r_t boosts firm valuation not only in the current period but also in future periods, as equation (13) shows. Therefore, the manager chooses higher manipulation in equilibrium when he values his future payoffs more. Similarly, the regulator's MB of detection also has both a contemporaneous and an intertemporal components, as equation (18) suggests. Therefore, the MB increases when the regulator places

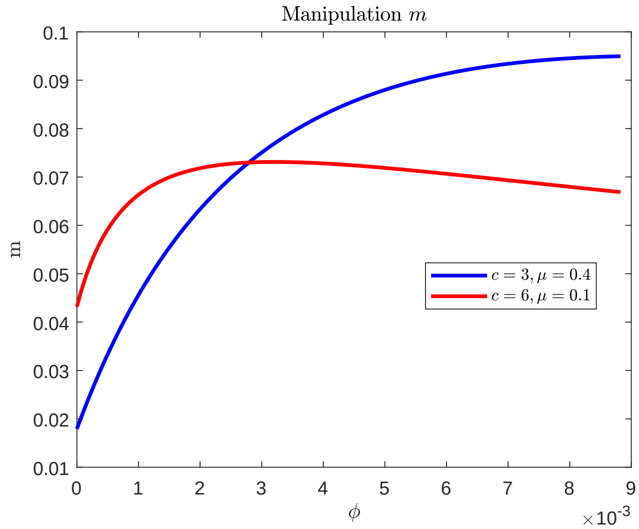

Figure 6: Equilibrium manipulation $m^*(\Phi)$. This figure plots the manager's optimal choice of manipulation, m, as a function of the information uncertainty induced by cumulative fraud, Φ. The red line and blue line use different values of penalty parameter c and manipulation cost μ, and other parameters are set to their estimated values reported in Table 1.

a larger weight on the intertemporal effects, which induces the regulator to choose greater detection effort. However, the last panel of Figure 7 suggests that, despite the regulator exerting greater detection effort, the social surplus W, as defined in equation (17), decreases when the discounting factor δ increases. Intuitively, with a higher δ, the present value of the discounted future information loss is more damaging, thus impairing the social surplus.

Figures 8 and 9 describe the effects of shifting the penalty on fraud and the personal cost. The intuitions for both figures are relatively straightforward. When either the penalty on fraud or the personal cost increases, it reduces the manager's manipulation choice and improves the informativeness of financial reports. The resulting decrease in the fraud-induced uncertainty, in turn, improves the social surplus and allows the regulator to save some of her costly detection effort.

Lastly, Figure 10 characterizes the effect of detection cost. As the cost of detection κ increases, it forces the regulator to shrink her detection effort. This, in turn, lowers the MC of manipulation and the manager's manipulation incentive heightens. Consequently, the social surplus decreases.

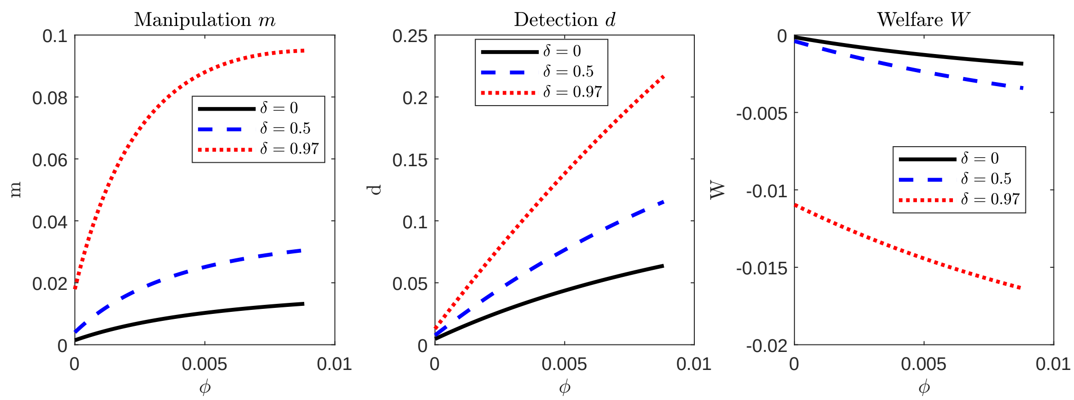

Figure 7: Comparative statics w.r.t. δ. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the discount factor, δ, is set to 0 (black solid line), 0.5 (blue dashed line), and 0.97 (red dotted line). Other parameters used in generating this figure are set to their estimated values reported in Table 1.

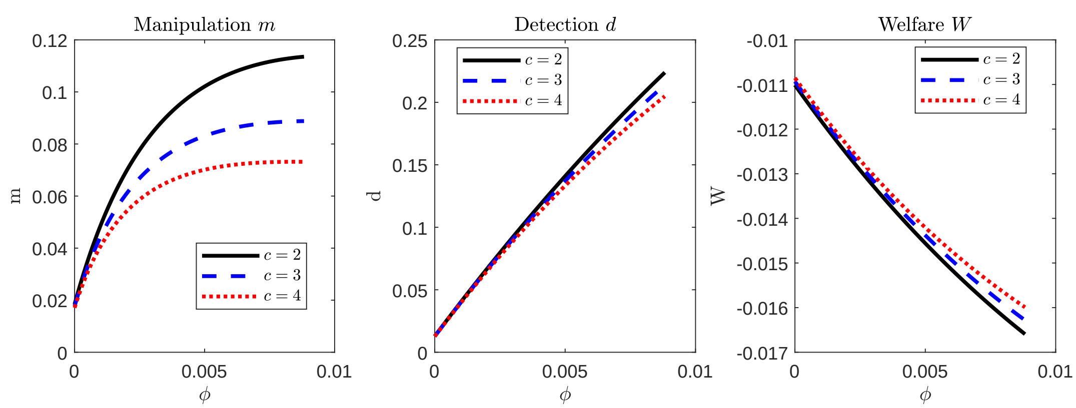

Figure 8: Comparative statics w.r.t. c. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the penalty on fraud, c, is set to 2 (black solid line), 3 (blue dashed line), and 4 (red dotted line). Other parameters used in generating this figure are set to their estimated values reported in Table 1.

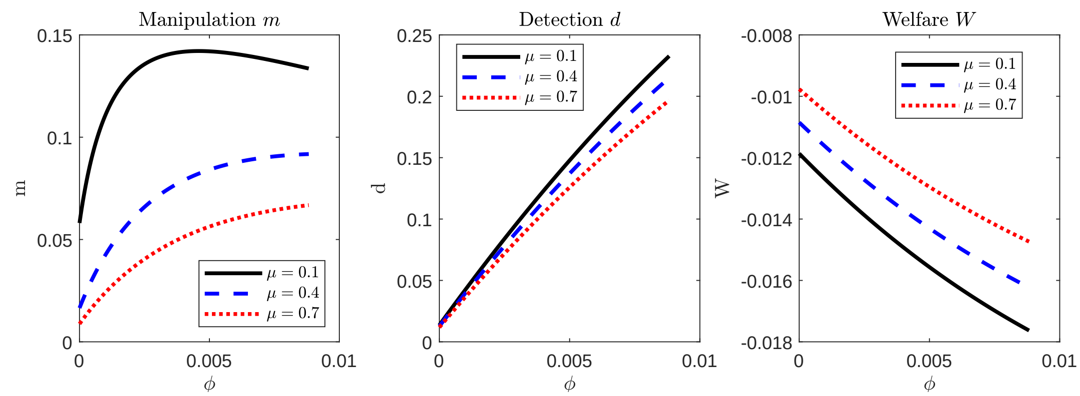

Figure 9: Comparative statics w.r.t. μ. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the manipulation cost, μ, is set to 0.1 (black solid line), 0.4 (blue dashed line), and 0.7 (red dotted line). Other parameters used in generating this figure are set to their estimated values reported in Table 1.

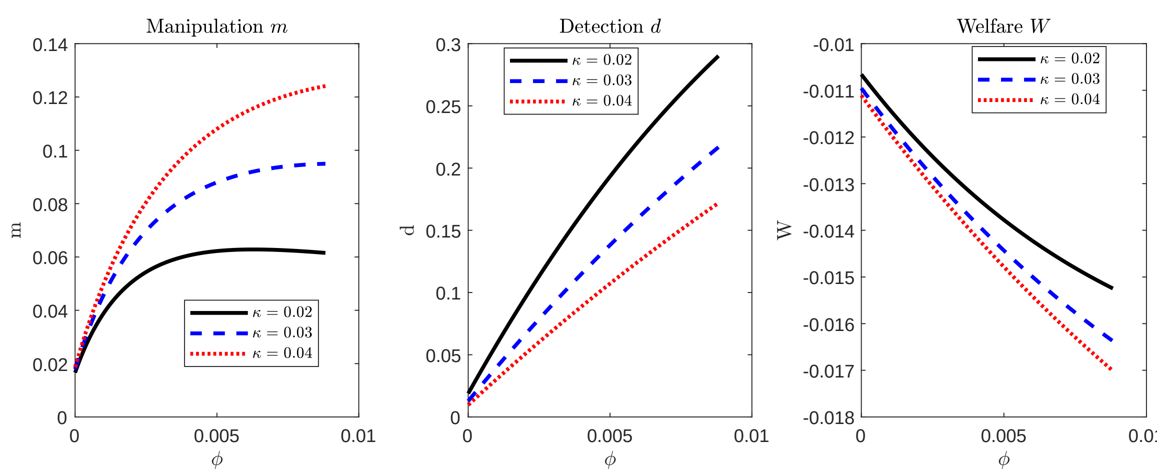

Figure 10: Comparative statics w.r.t. κ. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the cost of detection, κ, is set to 0.02 (black solid line), 0.03 (blue dashed line), and 0.04 (red dotted line). Other parameters used in generating this figure are set to their estimated values reported in Table 1.

3 Multi-firm Model

3.1 Model Setup

In this section, we expand the single-firm model to study the dynamic features of fraud among multiple firms; the model setup is similar with two exceptions. First, the economy contains N firms, and their economic earnings are independent of each other. This assumption rules out the mechanical correlation of fraud among firms due to correlated economic fundamentals. It also allows us to abstract away from the effects of information spillovers, which are not a central focus of this study. Second, the regulator has to allocate limited resources among N firms towards fraud detection. Specifically, in each period, the regulator conducts an independent inspection of each firm's report and we denote the probability that the inspection uncovers fraud in firm i's report by $d_{it} \in [0, 1]$, where $i \in \{1, 2, ..., N\}$ and d_{irt} represents the regulator's choice of detection technology to influence the probability of detecting fraud at firm i. We assume that the total detection cost for each period is:

$$\frac{\kappa}{2}\left(\sum_{i=1}^{N} d_{it}\right)^2. \tag{29}$$

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

3.2 Analysis

Even though the multi-firm setting has no restrictions on the maximum number of firms in the model, most of our analyses in this section focus on the case with three firms (i.e., $N = 3$), because a 3-firm setting is sufficient to deliver the key implications of a general multi-firm model.

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

Proposition 3 *Consider a three-firm model. Given the levels of accumulated past fraud at the three firms* $\{\Phi_1, \Phi_2, \Phi_3\}$, *the manager in firm* 1 *chooses manipulation*

$$m^*_1 (\Phi_1, \Phi_2, \Phi_3) = \frac{\rho^2\Phi_1 + \sigma^2_\varepsilon}{q(1-q)} \left(\frac{1-d^*_1}{cd^*_1+\mu} \left(\frac{1}{1-\delta\rho} + \frac{\delta\rho q(1-q)}{\rho^2\Phi'_1 + \sigma^2_\varepsilon} \times E\left[m^{*\prime}_1 \left(cd^{*\prime}_1 + \mu\right)\right] \right) - 1 \right), \tag{30}$$

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

$$\begin{aligned} d^*_1(\Phi_1, \Phi_2, \Phi_3) = \frac{1}{\kappa} \{ & \Phi'_1 + \delta(1-d^*_2)(1-d^*_3)\left[W(0, \Phi'_2, \Phi'_3) - W(\Phi'_1, \Phi'_2, \Phi'_3)\right] \\ & + \delta(1-d^*_2) d^*_3 \left[W(0, \Phi'_2, 0) - W(\Phi'_1, \Phi'_2, 0)\right] \\ & + \delta d^*_2 (1-d^*_3)\left[W(0,0,\Phi'_3) - W(\Phi'_1, 0, \Phi'_3)\right] \\ & + \delta d^*_2 d^*_3 \left[W(0,0,0) - W(\Phi'_1, 0, 0)\right] \} - (d^*_2 + d^*_3), \end{aligned} \tag{31}$$

where

$$\Phi'_i(\Phi_1, \Phi_2, \Phi_3) = \frac{m^*_i q(1-q)(\rho^2\Phi_i + \sigma^2_\varepsilon)}{\rho^2\Phi_i + \sigma^2_\varepsilon + m^*_i q(1-q)}, \tag{32}$$

$$\begin{aligned} E\left[m^{*\prime}_1 \left(cd^{*\prime}_1 + \mu\right)\right] = & (1-d^*_2)(1-d^*_3) m^*_1(\Phi'_1, \Phi'_2, \Phi'_3)\left(cd^*_1(\Phi'_1, \Phi'_2, \Phi'_3) + \mu\right) \\ & + d^*_2(1-d^*_3) m^*_1(\Phi'_1, 0, \Phi'_3)\left(cd^*_1(\Phi'_1, 0, \Phi'_3) + \mu\right) \\ & + (1-d^*_2) d^*_3 m^*_1(\Phi'_1, \Phi'_2, 0)\left(cd^*_1(\Phi'_1, \Phi'_2, 0) + \mu\right) \\ & + d^*_2 d^*_3 m^*_1(\Phi'_1, 0, 0)\left(cd^*_1(\Phi'_1, 0, 0) + \mu\right), \end{aligned} \tag{33}$$

$$
\begin{aligned}
W\left(\Phi_1,\Phi_2,\Phi_3\right) = &-\left(1-d_1^*\right)\left(1-d_2^*\right)\left(1-d_3^*\right)\left[\Phi_1'+\Phi_2'+\Phi_3'-\delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
&-\left(1-d_1^*\right)d_2^*\left(1-d_3^*\right)\left[\Phi_1'+\Phi_3'-\delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\
&-\left(1-d_1^*\right)\left(1-d_2^*\right)d_3^*\left[\Phi_1'+\Phi_2'-\delta W\left(\Phi_1',\Phi_2',0\right)\right] \\
&-\left(1-d_1^*\right)d_2^*d_3^*\left[\Phi_1'-\delta W\left(\Phi_1',0,0\right)\right] \\
&-d_1^*\left(1-d_2^*\right)\left(1-d_3^*\right)\left[\Phi_2'+\Phi_3'-\delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\
&-d_1^*d_2^*\left(1-d_3^*\right)\left[\Phi_3'-\delta W\left(0,0,\Phi_3'\right)\right] \\
&-d_1^*\left(1-d_2^*\right)d_3^*\left[\Phi_2'-\delta W\left(0,\Phi_2',0\right)\right] \\
&+\delta d_1^*d_2^*d_3^*W\left(0,0,0\right)-\frac{\kappa}{2}\left(d_1^*+d_2^*+d_3^*\right)^2. \qquad (34)
\end{aligned}
$$

The manipulation choices $\{m_2^*, m_3^*\}$ *and the detection choices* $\{d_2^*, d_3^*\}$ *at firms* 2 *and* 3 *can be analogously derived and given in Appendix A.6.*

Proposition 3 suggests that the dynamics of the manipulation and the detection decisions in the multi-firm model is largely in line with that in the single-firm model. There are, however, two new insights. First, the managers' manipulation decisions are endogenously linked because the regulator's choices of detection intensity are interdependent across firms. As such, a manager's manipulation choice becomes a function of the cumulative levels of fraud at all firms. Second, while the manager in the single-firm model is able to precisely conjecture the future equilibrium manipulation and detection choices $\left\{m_{t+1}^*, d_{t+1}^*\right\}$ (as shown in equation (14)), managers in the multi-firm model face uncertainty and must form expectations about the two equilibrium choices. This is because, due to the interdependence of detection and manipulation choices across firms, the manager at firm i rationally anticipates that the pair of the future manipulation and detection choices $\left\{m_{it+1}^*, d_{it+1}^*\right\}$ are also functions of the future cumulative levels of fraud at the other firms, $\{\Phi_{it}\}$. However, at the time of choosing m_{it} in period t, the value of Φ_{it} is random as it depends on whether the regulator detects fraud in the other firms later in period t.

As in the single-firm model, the dynamic programming problem in Proposition 3 does not have a closed-form solution so we solve the full model numerically to analyze the key properties of these policy functions. To glean some analytical results, we, again, first solve a special case of our model in which the discounting factor $\delta = 0$. The following proposition summarizes the regulator's equilibrium allocation of detection strength among the three firms.

Proposition 4 *Consider a special case of our model with the discount factor $\delta = 0$ and three firms. Without loss of generality, assume that the levels of accumulated past fraud at the three firms $\Phi_1 \geq \Phi_2 \geq \Phi_3$. The equilibrium detection strength at the three firms $\{d_1^*, d_2^*, d_3^*\}$ are given as follows:*

1. *If Φ_1 is much larger than Φ_2 (i.e., $\Phi_1 > H(\Phi_2)$), $d_2^* = d_3^* = 0$ and d_1^* solves*

$$\left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^* + \mu}{1 - d_1^*}\right) = \kappa d_1^*, \tag{35}$$

   and is increasing in Φ_1;

2. *If Φ_1 and Φ_2 are of similar sizes but both much larger than Φ_3 (i.e., $\Phi_1 \leq H(\Phi_2)$ and $L(\Phi_1, \Phi_2) > \Phi_3$), $d_3^* = 0$ and the pair of $\{d_1^*, d_2^*\}$ solves*

$$\left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^* + \mu}{1 - d_1^*}\right) = \kappa\left(d_1^* + d_2^*\right), \tag{36}$$

$$\left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_2^* + \mu}{1 - d_2^*}\right) = \kappa\left(d_1^* + d_2^*\right). \tag{37}$$

   d_1^* is increasing in Φ_1 and decreasing in Φ_2 whereas d_2^* is increasing in Φ_2 and decreasing in Φ_1;

3. *If Φ_1, Φ_2 and Φ_3 are of similar sizes (i.e., $\Phi_1 \leq H(\Phi_2)$ and $L(\Phi_1, \Phi_2) \leq \Phi_3$), the*

triplet of $\{d_1^*, d_2^*, d_3^*\}$ *solves:*

$$\left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right), \tag{38}$$

$$\left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_2^* + \mu}{1 - d_2^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right), \tag{39}$$

$$\left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_3^* + \mu}{1 - d_3^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right). \tag{40}$$

d_1^* *is increasing in* Φ_1 *and decreasing in* $\{\Phi_2, \Phi_3\}$, d_2^* *is increasing in* Φ_2 *and decreasing in* $\{\Phi_1, \Phi_3\}$ *and* d_3^* *is increasing in* Φ_3 *and decreasing in* $\{\Phi_1, \Phi_2\}$.

$H(\Phi_2)$ *and* $L(\Phi_1, \Phi_2)$ *are strictly increasing functions defined in the proof.*

Proposition 4 generates two insights. First, as in the single-firm model, the regulator matches the strength of fraud detection at a given firm with the severity of fraud at that firm. Second, since the regulator needs to allocate the regulatory resources across the three firms, the detection intensity imposed by the regulator on a given firm depends on not only the firm's own information uncertainty from cumulative fraud but also how it compares to information uncertainty about the other two firms in the economy. Stated differently, when there are multiple firms, the regulator determines the strength of fraud detection at each firm based on the *relative* severity of fraud. When the fraud-induced uncertainty at one firm is much higher than that at the other firms, the regulator will devote most of the regulatory resource to detect fraud at that firm and almost none to the other firms. Furthermore, even when the regulator directs regulatory resources to all three firms, she will direct fewer regulatory resources to a firm when she anticipates the fraud at the other firms to rise. This, in turn, boosts the incentive of the underlying firm to manipulate. In this light, an increase of fraud at a firm generates spillovers to the other firms through the interdependence of detection and manipulation choices across firms. We will later explore how these spillovers may drive fraud waves across firms.

Next, we solve the full model with the three firms ($\delta > 0$) numerically using the same

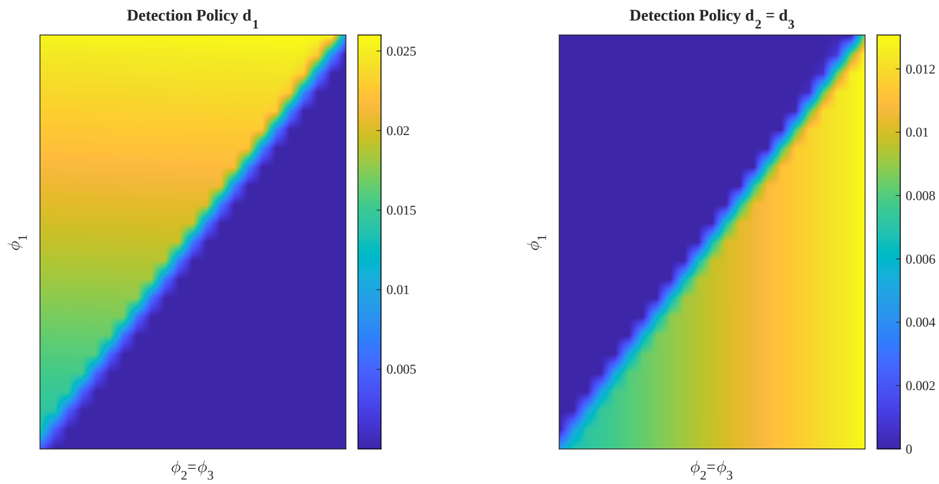

Figure 11: Equilibrium detection probability d_i^* in the three-firm setting. This figure plots the manager's optimal choice of manipulation m for firm 1 (left panel) and firm 2 and 3 (right panel) as a function of Φ_1, Φ_2, and Φ_3 in a heatmap. We assume that $\Phi_2 = \Phi_3$ in this figure. The parameters used in generating this figure are set to their estimated values reported in Table 1.

parameter values set for the one-firm model: $\delta = 0.974$, $q = 0.5$, $\rho = 0.70$, $\sigma_\varepsilon = 0.015$, $k = 0.03$, $c = 2.7$, and $\mu = 0.37$. Based on the numeric solution, we first analyze the regulator's detection decisions. To facilitate our analysis below, we present the model solution for a special case when $\Phi_2 = \Phi_3$. That is, we exemplify our model predictions by analyzing the detection intensity on different firms assuming that firm 2 and 3 have the same level of information uncertainty from cumulative fraud. It is easy to verify that, by model symmetry, the detection intensity on firm 2 and 3 is identical in this case, that is, $d_2 = d_3$.

Figure 11 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 focuses on firm 1 when its cumulative fraud is high and its information uncertainty stands out among the three firms. Specifically, in the northwest corner where $\Phi_1 >> \Phi_2 = \Phi_3$, the regulator invests almost all resources in detecting fraud at firm 1, leaving firm 2 and 3 under the radar. Vice versa, in the southeast corner where firm 2 and 3 both accumulate much fraud and leave firm 1 behind (i.e., $\Phi_2 = \Phi_3 >> \Phi_1$), we observe more regulatory resources directed towards firm 2 and 3 and little regulatory attention is given to firm 1. Note that this is in line with the analytical results derived in the special case of $\delta = 0$ (i.e., Parts 1 and 2 of Proposition 4).

Second, the two scenarios are not entirely symmetric as the detection intensity imposed on firm 1 (about 0.025) in the first scenario is much larger than the detection intensity imposed on firm 2 and 3 (about 0.012, respectively) in the second scenario. This is because the regulator's cost of detection is convex in the aggregate detection intensity, as shown in equation (29), and thus the MC of detecting one firm also depends on whether other firms in the economy require close scrutiny. The model implies that detection is the most costly if fraud tends to cluster across firms (i.e., fraud wave), a feature that we will study later in the paper.

Lastly, we observe that when the three firms' information uncertainty converges along the 45-degree line (i.e., $\Phi_1 = \Phi_2 = \Phi_3$), the regulator has to split the detection resource equally among them, which implies $d_1 = d_2 = d_3$. Note that this result corresponds to Part 3 of Proposition 4.

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

Given the regulator's detection policy discussed above, equation (30) suggests that managers' manipulation decisions are also interdependent in our model. Intuitively, if one firm stands out in its cumulative fraud, it should expect close scrutiny from the regulator and so the MC of further committing fraud likely dominates the MB, leading the manager to be more conservative. Ironically, as the firm with the highest information uncertainty attracts the most attention by the regulator, other firms are subject to less scrutiny and can afford to become more aggressive in committing fraud. To the extent that manipulation in each period accumulates and adds to the firms' information uncertainty over time, our model predicts an unintended consequence of anti-fraud regulation: it leads to synchronization of firms' cumulative fraud by allowing less fraudulent firms to catch up, which may eventually lead to fraud waves even in the absence of systematic shocks in the economy.

We next use the model to study the dynamics of the fraud-detection game between the regulator and three firms with different levels of fraud-induced information uncertainty in the initial period. Without loss of generality, we assume that $\Phi_H > \Phi_M > \Phi_L$ at $t = 1$ and denote the three firms H-, M- and L-firm, respectively. We then simulate the magnitude of manipulation committed by each manager m_i, the regulator's detection policy on each firm d_i, and the realization of detection outcomes at the end of each period. As we simulate the model forward, it generates the time series of Φ_{it}, d^*_{it}, and m^*_{it}. Figure 12 plots the three variables over the simulation path.

For the L-firm (depicted by the red-dash line), since the regulator anticipates a low level of cumulative fraud in the firm (i.e., a low Φ_L in Panel A), she spends little on detection (i.e., a low d_L in Panel B). The firm manager thus continues to commit fraud (i.e., increasing m_L in Panel C) because the MC is low and fraud starts building up (i.e., increasing Φ_L in Panel A). The first ten periods of the red dash line in Figure 12 illustrate this stage.

M-firm (depicted by the black-solid line) starts with an intermediate level of cumulative fraud. On the one hand, the manager of M-firm has greater incentives to commit fraud than the manager of L-firm, because a higher Φ increases the MB of committing fraud. On the other hand, the regulator tends to invest more in fraud detection of M-firm than L-firm, which suggests a higher MC of committing fraud. The two effects go hand-in-hand. The first three periods of the black-solid line in Figure 12 show the stage when the benefit of manipulation dominates its cost, and thus m_M increases over time as Φ_M grows. After the third period, we observe that the detection intensity on M-firm quickly rises (see the black-

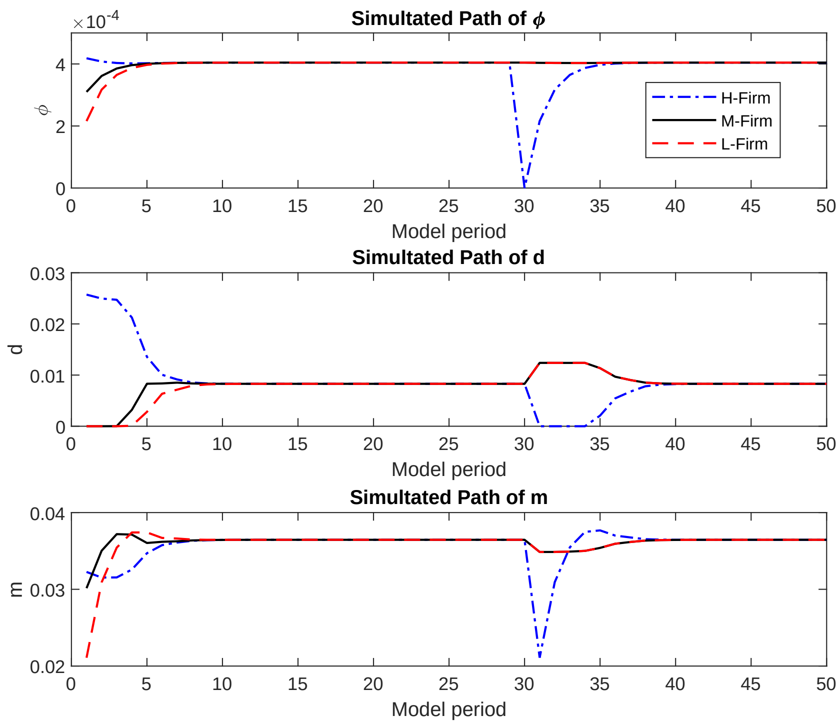

Figure 12: Simulated paths of the three-firm model. This figure plots the simulated path of information uncertainty induced by cumulative fraud Φ, the regulator's optimal choice of detection intensity d, and the manager's optimal choice of manipulation m for firms with high- (H), medium- (M), and low- (L) level of initial Φ. In the simulation, H-firm is detected at period 30. The parameters used in generating this figure are set to their estimated values reported in Table 1.

solid line in Panel B) and thus the cost of manipulation outweighs the benefit, leading to a decline in manipulation by M-firm (see the black-solid line in Panel C). The dynamics in m_M therefore demonstrates the counteracting forces of MB and MC.

Last, H-firm (depicted by the blue-dot line) starts with the highest level of cumulative fraud. Accordingly, it is under the closest scrutiny by the regulator. The regulator concentrates on detecting H-firm in the first five periods until the cumulative fraud of M-firm (and L-firm) catches up and gets close to that of H-firm after the 6th (9th) period, after which the detection intensity of H-firm and M-firm (and L-firm) starts converging. The blue-dot line depicts the trajectory of H-firm's Φ_H, m_H, and d_H in three panels, respectively.

To examine the impact of actual detection, we assume in the simulation trial that H-firm is caught by the regulator at period 30. Upon detection, H-firm's cumulative fraud is cleared and Φ_H drops to zero immediately, as shown in Panel A. As an optimal response, the regulator shifts attention from H-firm to the original M- and L-firms, as shown in Panel B. Interestingly, as the detection intensity on H-firm drops substantially, H-firm faces a low MC of committing fraud and can now afford to become more aggressive in manipulating its report. This explains the sharp increase in m_H and Φ_H in Panel C and A right after period 31. If M- and L-firms remain undetected, cumulative fraud in the three firms will be synchronized again after another few periods. This analysis sheds further light on an unintended consequence of regulation: it may synchronize firms' cumulative fraud 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 Estimation

To illuminate whether our model insights are reflected in the data, we employ the Simulated Method of Moments (SMM) to estimate the key model parameters. As is customary with SMM, we first identify empirical patterns (moments) that are likely informative of model parameters and then search for a set of parameters that minimize the distance between the model-implied moments and their empirical counterparts.

4.1 Sample and Variables

The sample used for estimation represents an intersection of restatement data from Audit Analytics, implied volatility data from Option Metrics, and quarterly financial data from Compustat-CRSP spanning from 1999 to 2017, as previously outlined in Section 1. As discussed, a significant challenge we face is that cumulative fraud-induced uncertainty is unobservable in the data. Implied volatility, while the best empirical proxy available, may reflect factors unrelated to fraud. In other words, implied volatility is not the precise empirical counterpart of cumulative fraud-induced uncertainty in the model.

To address this challenge, we do not aim to match the model-implied uncertainty directly to implied volatility in the estimation. Instead, we focus on matching other moments, such as detection probability, cumulative fraud upon detection, and the duration of uncertainty recovery post detection. Doing so requires a more modest identifying assumption – the rank of fraud-induced uncertainty is preserved in the rank of implied volatility. Building on this assumption, we construct several moments and measures necessary for estimating the model. First, we sort firms in the data into tercile groups based on their IV in each quarter, and we create an indicator that takes value of L, M, or H for the low-, medium-, and high-group, respectively. Since the rank of fraud-induced uncertainty is preserved in the rank of IV (i.e., the identifying assumption), we can map these groups of firms in the data to the firms with different levels of uncertainty in the model. Second, we assess detection probability for each group using $DETECT$, an indicator denoting the announcement of a fraud-related restatement. Third, for each restatement announced by a firm, we measure the firm's cumulative fraud by taking the restatement's aggregate dollar impact on income statements provided by Audit Analytics and scaling it by total assets at the beginning of the earliest affected quarter (as disclosed in the announcement). Fourth, we define the duration of uncertainty recovery as the time it takes for IV to rebuild from its lowest level within the two-year period following detection to its highest level within the subsequent five years.[12] To the extent that the rank of fraud-induced uncertainty in the model is preserved in the rank of IV in the data, the duration of IV recovery in the data is mapped to the duration of uncertainty recovery in the model. Lastly, we estimate quarter-to-quarter earnings persistence by utilizing industry- and seasonally-adjusted earnings. To perform the seasonal adjustment, we apply the X11-Autoregressive Integrated Moving Average (X11-ARIMA) method.

4.2 Simulation

After solving the model, we proceed to simulate a large number of three-firm panels. Within each panel, we randomly assign initial values of ϕ to the L-, M-, or H-firm, drawn from the low-, medium-, or high-tercile of the ϕ distribution, respectively. These individual panels are then pooled to form our simulation sample. Using the sample, we begin the simulation process by computing the model-implied manipulation amount, m, for each firm from equation (30) and update the value of ϕ based on the law of motion specified in equation (32). We then randomly draw the realization of detection probability following equation (31). If a firm is detected, its ϕ is reset to zero. This simulation process is repeated for 500 periods (quarters), with the first 50 periods disregarded as burn-in. Following this, we utilize the

[12] Section 1 suggests that, on average, IV experiences a decline of approximately one year before reaching its lowest level post-detection. To accommodate potential variations in this duration, we utilize a two-year window to locate the trough. With this choice, the average duration to reaching the trough is 1.04 years, with a standard deviation of 0.66 years. Following the trough, we apply an extended window of five years to capture the peak, allowing sufficient time for fraud to rebuild within a firm post-detection and for uncertainty to rebound. The average duration between the trough and peak is 17.8 months, or 5.93 quarters.

simulated panel data to compute the model-implied moments, sorting firms into terciles based on their realized ϕ in each period.

In the data, we analogously categorize firms into terciles based on their observed IV each quarter. We then construct empirical moments using a similar approach as outlined for the model moments.

4.3 Identification

For identification, we first set the parameters of ρ and σ_ε to match the persistence and the standard deviation of residual in the AR(1) process for earnings observed in the data. We also preset the annualized discount rate to be 0.9, implying a quarterly discount rate $\delta = 0.974$, and the likelihood of success manipulation, q, to be 0.5.

We have three model parameters left to estimate: one related to the regulator's detection cost (κ), and the other two linked to the manager's manipulation cost, including the regulatory penalty imposed upon fraud detection (c) and the manager's personal cost (μ). We focus the estimation of these parameters to best align three model moments with their observable empirical counterparts, including (1) the frequency of fraud detection, (2) the magnitude of cumulative fraud upon detection, and (3) the duration of uncertainty recovery.

In our model, detection cost and manipulation cost exert distinct effects on these moments. A high detection cost reduces detection intensity, thereby decreasing the frequency of detection. Managers, anticipating lower detection risk, manipulate more aggressively, thus increasing the magnitude of cumulative fraud upon detection. However, high manipulation costs, stemming from either the regulator's imposed penalty or the managers' personal cost, discourage manipulation, resulting in a lower level of cumulative fraud upon detection. The regulator, anticipating a lower level of manipulation, reduces detection intensity, which in turn reduces the frequency of detection. Thus, analyzing the combination of detection frequency and cumulative fraud upon detection helps distinguish detection cost (κ) from manipulation cost (μ and c).

To separately identify the two parameters linked to the manager's manipulation cost, we note their primary difference: the manager's personal cost (μ) continuously deters fraud, while the deterring effect of regulatory penalty (c) grows with detection risk, as the penalty is only imposed upon detection. Because detection risk is minimized immediately after fraud detection, the rate at which ϕ accumulates post-detection provides insight into the relative weight of these two costs: rapid accumulation followed by gradual slowing indicates a more dominant role of c, whereas smooth accumulation suggests μ is likely more influential. As shown in Section 1, the data reveals a significant drop in uncertainty (proxied using IV) after detection followed by a gradual buildup post-detection. Building on this evidence, we measure the time it takes for uncertainty to rebound from the trough to the post-detection peak, labeled duration of uncertainty recovery. A shorter duration suggests a more significant role played by c.

In addition, we examine the nine moments represented in the 3-by-3 transition matrix for ϕ. These moments, although untargeted by our estimation, represent the probability of ϕ remaining in the same tercile or moving into other terciles. Like the targeted moments, these moments are also influenced by all model parameters and reflect the model's primary mechanisms of manipulation, detection, and uncertainty dynamics. Therefore, assessing how well the model aligns with these moments sheds light on the model's fit.

4.4 Estimation Results

In this section, we present the estimation results for the six targeted moments, the point estimates of model parameters, and the overall model fit. Panel A of Table 1 indicates a close alignment between the model and the data for the targeted moments. Starting with the frequency of fraud detection, both the model and empirical estimates show that firms in the bottom tercile of uncertainty exhibit the lowest detection rate, approximately 0.7% per quarter, which corresponds to an annualized fraud detection rate of 2.8%. In contrast, firms in the mid- and top-tercile experience higher detection rates of around 0.9% per quarter or

3.6% annually, suggesting increased regulatory scrutiny for these firms. These detection rates are estimated with high precision, as evidenced by the small standard errors associated with them. In addition, the differences between the model-implied detection rates and empirical counterparts are statistically insignificant.

Turning to the magnitude of cumulative fraud upon detection, we observe that firms' fraudulent activities persist over various horizons in the data. The average length of the restatement period, covering all affected quarters disclosed in a restatement announcement, is eight quarters. As outlined in Section 4.2, we compute the cumulative fraud amount for an announced restatement in the data as a percentage of total assets immediately prior to the restatement period. Similarly, we measure cumulative fraud upon detection in the model by aggregating the total manipulation amount ($m(1-q)$) over the eight-quarter period preceding detection and scaling it by total assets measured at the beginning of the period. The model-implied magnitude of cumulative fraud upon detection closely resembles its empirical counterpart, representing approximately 11% of total book assets; the difference between the two is statistically insignificant. This moment is again estimated with high precision.

Third, our model closely mirrors the data in estimating the duration of uncertainty recovery. Specifically, the empirical estimate suggests that it takes an average of 5.93 quarters for a firm's IV to rebound from its lowest level to its highest level post-detection, a duration that aligns closely with the model's implied number.

Lastly, we examine whether the model also fits the data for moments that are untargeted by our estimation. For the model, we construct a 3-by-3 transition matrix of firm uncertainty (ϕ), where diagonal elements denote the likelihood of firms remaining in the same tercile and off-diagonal elements denote the probability of moving into a different tercile. Empirically, we sort firms into terciles based on their uncertainty in a given quarter and then compute the likelihoods for a given firm to either remain in the same tercile or move into a different tercile in the subsequent quarter. We observe strong persistence in firm uncertainty, with both the data and the model indicating a probability of staying in the same uncertainty tercile above 80%. Moreover, firms in the top and bottom terciles appear to exhibit greater persistence in uncertainty compared to those in the middle-tercile.

Panel B of Table 1 presents the parameter estimates, which minimize the distance between the model-implied moments and their empirical counterparts reported in Panel A. We preset four parameters that are not essential to our study and estimate the remaining three parameters. The coefficient of detection cost, κ, is estimated to be 0.03. To evaluate its magnitude, we note that the detection cost, $\frac{1}{2}\kappa d^2$, is specified as the regulator's disutility relative to the aggregate information uncertainty that it attempts to minimize. Our estimate of $\kappa = 0.03$ suggests that, from the regulator's perspective, the cost of increasing the detection rate by one percentage-point is equivalent to seeing the aggregate information uncertainty to increase by 1%.[13]

We estimate the coefficient of penalty on detected fraud, c, to be 2.7. This implies that for each unit of fraud detected, the imposed regulatory penalty is nearly three times as large. Additionally, we estimate the manager's personal manipulation cost, μ, to be 0.37, which corresponds to approximately 14% of the regulatory penalty cost.

Based on the estimated parameters, our model indicates that eradicating fraud is unlikely due to a dynamic interplay between regulatory actions and firm behavior. After detecting fraud, the regulator optimally reduces the scrutiny on the offending firm. However, this reduced oversight enables the firm to resume fraudulent activities. In equilibrium, firms facing heightened regulatory scrutiny decrease their manipulation, while those hiding under the radar increase their fraudulent activities. This leads to a convergence in fraud levels and information uncertainty across firms. Quantitatively, our estimation suggests that this convergence occurs rather swiftly, typically within six quarters. Therefore, rather than eradicating fraud, anti-fraud regulations serve to synchronize fraudulent activities among firms, contributing to the observed waves of fraud in the data.

[13] Increasing the detection rate by one percentage-point from its sample mean incurs an incremental cost of $\frac{1}{2}\kappa[(d+0.01)^2 - d^2] = 4 \cdot 10^{-6}$, which is about 1% of the steady-state uncertainty $\phi = 4 \cdot 10^{-4}$.

5 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 build a theoretical foundation to better understand the formation and evolvement of 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. This is because, the interdependent nature of fraud and regulation essentially presents a cat-and-mouse equilibrium in which the strength of detection optimally matches the severity of fraud. As such, 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. Hence, the amount of fraud committed in the firm may exhibit repeated cycles of rise, peak, fall, and collapse (upon detection).

We then expand the model to consider a regulator and three firms with a high, medium, and low level of cumulative fraud, respectively. Analyses of this multi-firm model offer additional insights. Anti-fraud regulations can be highly effective at lowering the most fraudulent firms' incentives to continue fraud, by not only raising their MC of committing fraud but also sharply decreasing their MB of committing fraud upon detection. However, the rational allocation of regulatory resources towards such firms may imply less scrutiny of less fraudulent firms (a whack-a-mole equilibrium), 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.

A structural estimation validates our model's alignment with empirical data regarding detection frequency, cumulative fraud levels, and the duration of uncertainty recovery post-detection. Additionally, both the model and data emphasize the persistence of firm uncertainty, the significant penalties managers face for conducting fraud, as well as the constraints regulators face in optimizing detection intensity. Notably, our estimates also corroborate observed patterns of fraud waves, indicating a gradual convergence of fraud levels across firms over time.

These results carry strong policy implications. In our model, regulations lead frauds not because regulations are ineffective. Rather, regulations effectively tamp down fraud in the short term but in the long term, they synchronize firms' level of fraudulent activities and allow a wave of frauds to resurface. Hence, fraud remains a permanent risk in the financial markets and the efficacy of regulation is limited.

References

Aghamolla, Cyrus, Ilan Guttman, and Evgeny Petrov, 2021, Sequential reporting bias, *The Accounting Review* 1–33.

Antle, Rick, 1982, The auditor as an economic agent, *Journal of Accounting Research* 503–527.

Antle, Rick, 1984, Auditor independence, *Journal of Accounting Research* 1–20.

Balachandran, Bala V, and Ramachandran Ramanan, 1988, Optimal internal control strategy under dynamic conditions, *Journal of Accounting, Auditing & Finance* 3, 1–13.

Bertomeu, Jeremy, Edwige Cheynel, Edward Xuejun Li, and Ying Liang, 2021, How pervasive is earnings management? evidence from a structural model, *Management Science* 67, 5145–5162.

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

Bird, Andrew, Stephen A Karolyi, and Thomas G Ruchti, 2019, Understanding the "numbers game", *Journal of Accounting and Economics* 68, 101242.

Blinder, Alan S, 2008, From the new deal, a way out of a mess, *New York Times* February 24.

Blinder, Alan S, 2015, Financial entropy and the optimality of over-regulation, *The New International Financial System: Analyzing the Cumulative Impact of Regulatory Reform* 3–35.

Bond, Philip, 2008, Persistent court corruption, *The Economic Journal* 118, 1333–1353.

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

Cai, Jinghan, Jibao He, Wenxi Jiang, and Wei Xiong, 2021, The whack-a-mole game: Tobin taxes and trading frenzy, *Review of Financial Studies* 34, 5723–5755.

Cheynel, Edwige, Davide Cianciaruso, and Frank S Zhou, 2024, Fraud power laws, *Journal of Accounting Research* 62, 833–876.

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

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.

Gao, Pingyang, and Gaoqing Zhang, 2019, Accounting manipulation, peer pressure, and internal control, *The Accounting Review* 94, 127–151.

Goldstein, Itay, and Liyan Yang, 2017, Information disclosure in financial markets, *Annual Review of Financial Economics* 9, 101–125.

Guttman, Ilan, Ohad Kadan, and Eugene Kandel, 2006, A rational expectations theory of kinks in financial reporting, *The Accounting Review* 81, 811–848.

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

Harford, Jarrad, 2005, What drives merger waves?, *Journal of Financial Economics* 77, 529–560.

Heinle, Mirko S, and Robert E Verrecchia, 2016, Bias and the commitment to disclosure, *Management Science* 62, 2859–2870.

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.

Kirschenheiter, Michael, and Nahum D Melumad, 2002, Can "big bath" and earnings smoothing co-exist as equilibrium financial reporting strategies?, *Journal of Accounting Research* 40, 761–796.

Marinovic, Iván, and Martin Szydlowski, 2022, Monitoring with career concerns, *The RAND Journal of Economics* 53, 404–428.

Marinovic, Iván, and Martin Szydlowski, 2023, Monitor reputation and transparency, *American Economic Journal: Microeconomics* 15, 1–67.

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

Pound, John, 1992, On the motives for choosing a corporate governance structure: A study of corporate reaction to the pennsylvania takeover law, *Journal of Law, Economics & Organization* 8, 656–672.

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

Sankar, Mandira Roy, and K.R. Subramanyam, 2001, Reporting discretion and private information communication through earnings, *Journal of Accounting Research* 39, 365–386.

Stein, Jeremy C, 1989, Efficient capital markets, inefficient firms: A model of myopic corporate behavior, *Quarterly Journal of Economics* 104, 655–669.

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

Terry, Stephen J, 2023, The macro impact of short-termism, *Econometrica* 91, 1881–1912.

Toms, Steven, 2019, Financial scandals: a historical overview, *Accounting and Business Research* 49, 477–499.

Varas, Felipe, Iván Marinovic, and Andrzej Skrzypacz, 2020, Random inspections and periodic reviews: Optimal dynamic monitoring, *The Review of Economic Studies* 87, 2893–2937.

Ye, Minlei, 2023, The theory of auditing economics: Evidence and suggestions for future research, *Foundations and Trends® in Accounting* 18, 138–267.

Zakolyukina, Anastasia A, 2018, How common are intentional GAAP violations? estimates from a dynamic model, *Journal of Accounting Research* 56, 5–44.

Table 1: Estimation Results

This table shows the overall model fit (Panel A) and the parameter estimates (Panel B). In each quarter, we sort firms into tercile groups based on IV, and L (M and H) indicates the group of firms with low (medium and high) IV. The targeted moments include the quarterly likelihood of fraud detection, the cumulative fraud committed upon detection, and the duration of uncertainty recovery post detection. We measure quarterly detection probability for each IV tercile group based on $DETECT$, an indicator for the announcement of fraud-related restatement. For each detected fraud, we measure cumulative fraud as the aggregate impact on income statement from Audit Analytics divided by the total assets at the beginning of restatement period. We define the duration of uncertainty recovery as the time it takes for IV to recover from its lowest level in the two-year window post-detection to the highest level in the five-year window after reaching the lowest point. The untargeted moments include the transition probability of a firm falling in a certain IV tercile in quarter q to stay in the same tercile or move to a different tercile in quarter $q+1$.

Panel A. Model Fit

Targeted Moments				
		Model	**Data**	**S.E.**
	L	0.69	0.64	0.04
Detection probability (%)	M	0.84	0.90	0.04
	H	0.91	0.89	0.04
Cumulative fraud upon detection		0.108	0.109	0.012
Duration of uncertainty recovery (qtr)		6.00	5.93	4.57

Untargeted Moments

	Transition Matrix					
		Data			**Model**	
	L	M	H	L	M	H
L	0.908	0.089	0.003	0.945	0.042	0.013
M	0.102	0.801	0.098	0.028	0.847	0.125
H	0.002	0.112	0.886	0.027	0.111	0.862

Panel B: Parameter Estimate

Pre-set parameters		
δ	discount rate	0.974
q	probability of successful manipulation	0.5
ρ	persistence of earnings	0.7
σ_ϵ	conditional volatility of earnings	0.015
Estimated parameters		
κ	detection cost	0.03
c	penality on detected fraud	2.7
μ	variable cost of manipulation	0.37

For Online Publication

Internet Appendix of "Everlasting Fraud"

Appendix A. Proofs

A.1 Lemma 1

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

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

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

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

Multiplying both sides of equation (A.2) by $\delta\rho\left(1-d^*_{t+1}\right)\frac{m^*_{t+1}q(1-q)}{\rho^2\Phi_t+\sigma^2_\varepsilon+m^*_{t+1}q(1-q)}$ yields:

$$
\begin{aligned}
&\delta\rho\left(1-\delta\rho\right)\left(cd^*_{t+1}+\mu\right)\frac{m^*_{t+1}q\left(1-q\right)}{\rho^2\Phi_t+\sigma^2_\varepsilon} \\
&=\delta\rho\left(1-d^*_{t+1}\right)\frac{m^*_{t+1}q\left(1-q\right)}{\rho^2\Phi_t+\sigma^2_\varepsilon+m^*_{t+1}q\left(1-q\right)} \\
&+\delta^2\rho^2\left(1-d^*_{t+2}\right)\left(1-d^*_{t+1}\right)\frac{m^*_{t+2}q\left(1-q\right)}{\rho^2\Phi_{t+1}+\sigma^2_\varepsilon+m^*_{t+2}q\left(1-q\right)}\frac{m^*_{t+1}q\left(1-q\right)}{\rho^2\Phi_t+\sigma^2_\varepsilon+m^*_{t+1}q\left(1-q\right)} \\
&+\delta^3\rho^3\left(1-d^*_{t+1}\right)\left(1-d^*_{t+2}\right)\left(1-d^*_{t+3}\right)\frac{m^*_{t+1}q\left(1-q\right)}{\rho^2\Phi_t+\sigma^2_\varepsilon+m^*_{t+1}q\left(1-q\right)} \\
&\times\frac{m^*_{t+3}q\left(1-q\right)}{\rho^2\Phi_{t+2}+\sigma^2_\varepsilon+m^*_{t+3}q\left(1-q\right)}\frac{m^*_{t+2}q\left(1-q\right)}{\rho^2\Phi_{t+1}+\sigma^2_\varepsilon+m^*_{t+2}q\left(1-q\right)} \\
&+\ldots
\end{aligned}
\tag{A.3}
$$

Substituting (A.3) into (A.1) yields:

$$
\frac{\left(1-\delta\rho\right)\left(cd^*_t+\mu\right)}{1-d^*_t}\left(1+\frac{m^*_tq\left(1-q\right)}{\rho^2\Phi_{t-1}+\sigma^2_\varepsilon}\right)=1+\delta\rho\left(1-\delta\rho\right)\left(cd^*_{t+1}+\mu\right)\frac{m^*_{t+1}q\left(1-q\right)}{\rho^2\Phi_t+\sigma^2_\varepsilon}. \tag{A.4}
$$

Solving (A.4) for m_t yields (14) in the lemma. ■

A.2 Lemma 2

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

$$\begin{aligned}
\Phi_t &= var\left(s_t|\mathcal{F}_{t-1}\right) - var\left(E^I\left[s_t|\mathcal{F}_t\right]|\mathcal{F}_{t-1}\right) \\
&= var\left(\rho s_{t-1} + \varepsilon_t|\mathcal{F}_{t-1}\right) - var\left(E^I\left[s_t|r_t, r_{t-1}, ...\right]|\mathcal{F}_{t-1}\right) \\
&= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 - var\left(\frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right)} r_t|\mathcal{F}_{t-1}\right) \\
&= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 - \left[\frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right)}\right]^2 var\left(r_t|\mathcal{F}_{t-1}\right) \\
&= \frac{m_t^* q\left(1-q\right)\left[\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2\right]}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right)} \\
&= \frac{m_t^* q\left(1-q\right)\left(\rho^2 \Phi_{t-1} + \sigma_\varepsilon^2\right)}{\rho^2 \Phi_{t-1} + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right)}.
\end{aligned} \tag{A.5}$$

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

$$\begin{aligned}
E^I\left[s_t|\mathcal{F}_t\right] &= E^I\left[s_t|\mathcal{F}_{t-1}\right] + \frac{cov\left(r_t, s_t|\mathcal{F}_{t-1}\right)}{var\left(r_t|\mathcal{F}_{t-1}\right)}\left\{r_t - E\left[r_t|\mathcal{F}_{t-1}\right]\right\} \\
&= E^I\left[s_t|\mathcal{F}_{t-1}\right] + \frac{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2}{\rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right)}\left\{r_t - E\left[r_t|\mathcal{F}_{t-1}\right]\right\},
\end{aligned} \tag{A.6}$$

where

$$\begin{aligned}
var\left(r_t|\mathcal{F}_{t-1}\right) &= var\left(s_t|\mathcal{F}_{t-1}\right) + m_t^* q\left(1-q\right) \\
&= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2 + m_t^* q\left(1-q\right),
\end{aligned} \tag{A.7}$$

$$\begin{aligned}
cov\left(r_t, s_t|\mathcal{F}_{t-1}\right) &= var\left(s_t|\mathcal{F}_{t-1}\right) \\
&= \rho^2 var\left(s_{t-1}|\mathcal{F}_{t-1}\right) + \sigma_\varepsilon^2.
\end{aligned} \tag{A.8}$$

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

A.3 Lemma 3

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

$$
\begin{aligned}
W_t\left(\Phi_{t-1}\right) &= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*,\Phi_{t-1}\right)-\frac{\kappa}{2}d_t^2 \\
&+\delta E^I\left[\sum_{k=t+1}^{\infty}\delta^{k-(t+1)}\left(-\left(1-d_k^*\right)\Phi_k-\frac{\kappa}{2}d_k^{*2}\right)\right] \\
&= \max_{d_t} -\left(1-d_t\right)\Phi_t\left(d_t^*,\Phi_{t-1}\right)-\frac{\kappa}{2}d_t^2 \\
&+\delta\left[d_t W_{t+1}\left(0\right)+\left(1-d_t\right)W_{t+1}\left(\Phi_t\left(d_t^*,\Phi_{t-1}\right)\right)\right],
\end{aligned}
\tag{A.9}
$$

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

A.4 Proposition 1

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

A.5 Proposition 2

Proof. of Proposition 2: To complete the proof, we prove that d^* increases in Φ. Applying the implicit function theorem to equation (26) yields that:

$$
\frac{\partial d^*}{\partial \Phi}=\frac{\rho^2\left(1-\frac{cd^*(\Phi)+\mu}{1-d^*(\Phi)}\right)}{\frac{(c+\mu)(\rho^2\Phi+\sigma_\varepsilon^2)}{(1-d^*)^2}+\kappa}>0. \tag{A.10}
$$

The result that m^* can be non-monotonic in Φ is verified by numerical examples. ■

A.6 Proposition 3

Proof. of Proposition 3: We only derive the manipulation decision by manager 1 as the manipulation decisions by the other managers can be derived analogously.

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

$$\begin{aligned}
&c\,(1-q)\,d_{1t}^{*}+(1-q)\mu \\
&=\frac{1-d_{1t}^{*}}{1-\delta\rho}\frac{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}}{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}+m_{1t}^{*}q\,(1-q)}\,(1-q) \\
&+\frac{\left(1-d_{1t}^{*}\right)\delta\rho}{1-\delta\rho}\frac{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}}{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}+m_{1t}^{*}q\,(1-q)}\,(1-q) \\
&\times E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*}q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\,(1-q)}\right] \\
&+\frac{\left(1-d_{1t}^{*}\right)\delta^{2}\rho^{2}}{1-\delta\rho}\frac{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}}{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}+m_{1t}^{*}q\,(1-q)}\,(1-q) \\
&\times E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}} \\
&\times\left[\left(1-d_{1t+1}^{*}\right)\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+2}^{*}q\,(1-q)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*}q\,(1-q)}\frac{m_{1t+1}^{*}q\,(1-q)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\,(1-q)}\right] \\
&+...
\end{aligned} \tag{A.11}$$

Note that we need to take expectations over $\{\Phi_{2t},\Phi_{3t}\}$ because $\left\{d_{1t+1}^{*},m_{1t+1}^{*}\right\}$ depend on $\{\Phi_{2t},\Phi_{3t}\}$. Φ_{2t} and Φ_{3t} are random because they can be either 0 or positive, depending on whether the regulator detects fraud at the two firms.

Equation (A.11) can be simplified into

$$
\begin{aligned}
&\frac{(1-\delta\rho)\left(cd_{1t}^{*}+\mu\right)}{1-d_{1t}^{*}}\left(1+\frac{m_{1t}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t-1}+\sigma_{\varepsilon}^{2}}\right)\\
&=1+\delta\rho E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\left(1-q\right)}\right]\\
&+\delta^{2}\rho^{2}E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}}\\
&\times\left[\left(1-d_{1t+1}^{*}\right)\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+2}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*}q\left(1-q\right)}\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\left(1-q\right)}\right]\\
&+\ldots
\end{aligned}
\tag{A.12}
$$

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

$$
\begin{aligned}
&\frac{(1-\delta\rho)\left(cd_{1t+1}^{*}+\mu\right)}{1-d_{1t+1}^{*}}\left(1+\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}}\right)\\
&=1+\delta\rho E_{\Phi_{2t+1},\Phi_{3t+1}}\left[\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+2}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*}q\left(1-q\right)}\right]+\ldots
\end{aligned}
\tag{A.13}
$$

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

$$
\begin{aligned}
&(1-\delta\rho)\left(cd_{1t+1}^{*}+\mu\right)\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}}\\
&=\left(1-d_{1t+1}^{*}\right)\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\left(1-q\right)}\\
&+\delta\rho E_{\Phi_{2t+1},\Phi_{3t+1}}\\
&\times\left[\left(1-d_{1t+1}^{*}\right)\left(1-d_{1t+2}^{*}\right)\frac{m_{1t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t}+\sigma_{\varepsilon}^{2}+m_{1t+1}^{*}q\left(1-q\right)}\frac{m_{1t+2}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{1t+1}+\sigma_{\varepsilon}^{2}+m_{1t+2}^{*}q\left(1-q\right)}\right]\\
&+\ldots
\end{aligned}
\tag{A.14}
$$

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

$$\begin{aligned}
&\frac{(1-\delta\rho)\, q\,(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2} E_{\Phi_{2t},\Phi_{3t}}\left[\left(cd_{1t+1}^*+\mu\right) m_{1t+1}^*\right] \\
&= E_{\Phi_{2t},\Phi_{3t}}\left[\left(1-d_{1t+1}^*\right)\frac{m_{1t+1}^* q\,(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^* q\,(1-q)}\right] \\
&+\delta\rho E_{\Phi_{2t},\Phi_{3t},\Phi_{2t+1},\Phi_{3t+1}} \\
&\times\left[\left(1-d_{1t+1}^*\right)\left(1-d_{1t+2}^*\right)\frac{m_{1t+1}^* q\,(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2+m_{1t+1}^* q\,(1-q)}\frac{m_{1t+2}^* q\,(1-q)}{\rho^2\Phi_{1t+1}+\sigma_\varepsilon^2+m_{1t+2}^* q\,(1-q)}\right] \\
&+\ldots \qquad (A.15)
\end{aligned}$$

Substituting (A.15) into (A.12) yields:

$$\frac{(1-\delta\rho)\left(cd_{1t}^*+\mu\right)}{1-d_{1t}^*}\left(1+\frac{m_{1t}^* q\,(1-q)}{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}\right)=1+\frac{\delta\rho\,(1-\delta\rho)\, q\,(1-q)}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2}E_{\Phi_{2t},\Phi_{3t}}\left[\left(cd_{1t+1}^*+\mu\right) m_{1t+1}^*\right]. \qquad (A.16)$$

Solving for m_{1t}^* gives that

$$m_{1t}^*=\frac{\rho^2\Phi_{1t-1}+\sigma_\varepsilon^2}{q\,(1-q)}\left[\frac{1-d_{1t}^*}{cd_{1t}^*+\mu}\left(\frac{1}{1-\delta\rho}+\delta\rho q\,(1-q)\,\frac{E_{\Phi_{2t},\Phi_{3t}}\left[m_{1t+1}^*\left(cd_{1t+1}^*+\mu\right)\right]}{\rho^2\Phi_{1t}+\sigma_\varepsilon^2}\right)-1\right], \qquad (A.17)$$

where

$$\begin{aligned}
E_{\Phi_{2t},\Phi_{3t}}\left[\left(cd_{1t+1}^*+\mu\right) m_{1t+1}^*\right] &= \left(1-d_{2t}^*\right)\left(1-d_{3t}^*\right) m_{1t+1}^*\left(\Phi_{1t},\Phi_{2t},\Phi_{3t}\right)\left(cd_{1t+1}^*\left(\Phi_{1t},\Phi_{2t},\Phi_{3t}\right)+\mu\right) \\
&+d_{2t}^*\left(1-d_{3t}^*\right) m_{1t+1}^*\left(\Phi_{1t},0,\Phi_{3t}\right)\left(cd_{1t+1}^*\left(\Phi_{1t},0,\Phi_{3t}\right)+\mu\right) \\
&+\left(1-d_{2t}^*\right) d_{3t}^* m_{1t+1}^*\left(\Phi_{1t},\Phi_{2t},0\right)\left(cd_{1t+1}^*\left(\Phi_{1t},\Phi_{2t},0\right)+\mu\right) \\
&+d_{2t}^* d_{3t}^* m_{1t+1}^*\left(\Phi_{1t},0,0\right)\left(cd_{1t+1}^*\left(\Phi_{1t},0,0\right)+\mu\right). \qquad (A.18)
\end{aligned}$$

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

$$m_i^* = \frac{\rho^2 \Phi_i + \sigma_\varepsilon^2}{q(1-q)} \left(\frac{1-d_i^*}{cd_i^* + \mu} \left(\frac{1}{1-\delta\rho} + \frac{\delta\rho q(1-q)}{\rho^2 \Phi_i' + \sigma_\varepsilon^2} \times E\left[m_i'\left(cd_i' + \mu\right)\right] \right) - 1 \right), \tag{A.19}$$

where

$$\begin{aligned} E\left[m_1'\left(cd_1' + \mu\right)\right] &= \left(1-d_2^*\right)\left(1-d_3^*\right) m_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right)\left(cd_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right) + \mu\right) \\ &+ d_2^*\left(1-d_3^*\right) m_1^*\left(\Phi_1', 0, \Phi_3'\right)\left(cd_1^*\left(\Phi_1', 0, \Phi_3'\right) + \mu\right) \\ &+ \left(1-d_2^*\right) d_3^* m_1^*\left(\Phi_1', \Phi_2', 0\right)\left(cd_1^*\left(\Phi_1', \Phi_2', 0\right) + \mu\right) \\ &+ d_2^* d_3^* m_1^*\left(\Phi_1', 0, 0\right)\left(cd_1^*\left(\Phi_1', 0, 0\right) + \mu\right), \end{aligned} \tag{A.20}$$

$$\begin{aligned} E\left[m_2'\left(cd_2' + \mu\right)\right] &= \left(1-d_1^*\right)\left(1-d_3^*\right) m_2^*\left(\Phi_1', \Phi_2', \Phi_3'\right)\left(cd_2^*\left(\Phi_1', \Phi_2', \Phi_3'\right) + \mu\right) \\ &+ d_1^*\left(1-d_3^*\right) m_2^*\left(0, \Phi_2', \Phi_3'\right)\left(cd_2^*\left(0, \Phi_2', \Phi_3'\right) + \mu\right) \\ &+ \left(1-d_1^*\right) d_3^* m_2^*\left(\Phi_1', \Phi_2', 0\right)\left(cd_2^*\left(\Phi_1', \Phi_2', 0\right) + \mu\right) \\ &+ d_1^* d_3^* m_2^*\left(0, \Phi_2', 0\right)\left(cd_2^*\left(0, \Phi_2', 0\right) + \mu\right), \end{aligned} \tag{A.21}$$

$$\begin{aligned} E\left[m_3'\left(cd_3' + \mu\right)\right] &= \left(1-d_1^*\right)\left(1-d_2^*\right) m_3^*\left(\Phi_1', \Phi_2', \Phi_3'\right)\left(cd_3^*\left(\Phi_1', \Phi_2', \Phi_3'\right) + \mu\right) \\ &+ d_1^*\left(1-d_2^*\right) m_3^*\left(0, \Phi_2', \Phi_3'\right)\left(cd_3^*\left(0, \Phi_2', \Phi_3'\right) + \mu\right) \\ &+ \left(1-d_1^*\right) d_2^* m_3^*\left(\Phi_1', 0, \Phi_3'\right)\left(cd_3^*\left(\Phi_1', 0, \Phi_3'\right) + \mu\right) \\ &+ d_1^* d_2^* m_3^*\left(0, 0, \Phi_3'\right)\left(cd_3^*\left(0, 0, \Phi_3'\right) + \mu\right). \end{aligned} \tag{A.22}$$

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

$$
\begin{aligned}
W\left(\Phi_1,\Phi_2,\Phi_3\right)=\max_{d_1,d_2,d_3}\; & -\left(1-d_1\right)\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_1'+\Phi_2'+\Phi_3'-\delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& -\left(1-d_1\right)d_2\left(1-d_3\right)\left[\Phi_1'+\Phi_3'-\delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\
& -\left(1-d_1\right)\left(1-d_2\right)d_3\left[\Phi_1'+\Phi_2'-\delta W\left(\Phi_1',\Phi_2',0\right)\right] \\
& -\left(1-d_1\right)d_2d_3\left[\Phi_1'-\delta W\left(\Phi_1',0,0\right)\right] \\
& -d_1\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_2'+\Phi_3'-\delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\
& -d_1d_2\left(1-d_3\right)\left[\Phi_3'-\delta W\left(0,0,\Phi_3'\right)\right] \\
& -d_1\left(1-d_2\right)d_3\left[\Phi_2'-\delta W\left(0,\Phi_2',0\right)\right] \\
& +\delta d_1d_2d_3W\left(0,0,0\right)-\frac{\kappa}{2}\left(d_1+d_2+d_3\right)^2, \qquad \text{(A.23)}
\end{aligned}
$$

where

$$
\Phi_i'\equiv\frac{m_i^*q\left(1-q\right)\left(\rho^2\Phi_i+\sigma_\varepsilon^2\right)}{\rho^2\Phi_i+\sigma_\varepsilon^2+m_i^*q\left(1-q\right)}. \qquad \text{(A.24)}
$$

Taking the first-order condition yields:

$$
\begin{aligned}
d_1^*=\frac{1}{\kappa}\{ & \Phi_1'+\delta\left(1-d_2^*\right)\left(1-d_3^*\right)\left[W\left(0,\Phi_2',\Phi_3'\right)-W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& +\delta\left(1-d_2^*\right)d_3^*\left[W\left(0,\Phi_2',0\right)-W\left(\Phi_1',\Phi_2',0\right)\right] \\
& +\delta d_2^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right)-W\left(\Phi_1',0,\Phi_3'\right)\right] \\
& +\delta d_2^*d_3^*\left[W\left(0,0,0\right)-W\left(\Phi_1',0,0\right)\right]\}-\left(d_2^*+d_3^*\right), \qquad \text{(A.25)}
\end{aligned}
$$

$$
\begin{aligned}
d_2^*=\frac{1}{\kappa}\{ & \Phi_2'+\delta\left(1-d_1^*\right)\left(1-d_3^*\right)\left[W\left(\Phi_1',0,\Phi_3'\right)-W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
& +\delta\left(1-d_1^*\right)d_3^*\left[W\left(\Phi_1',0,0\right)-W\left(\Phi_1',\Phi_2',0\right)\right] \\
& +\delta d_1^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right)-W\left(0,\Phi_2',\Phi_3'\right)\right] \\
& +\delta d_1^*d_3^*\left[W\left(0,0,0\right)-W\left(0,\Phi_2',0\right)\right]\}-\left(d_1^*+d_3^*\right), \qquad \text{(A.26)}
\end{aligned}
$$

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

■

A.7 Proposition 4

Proof. of Proposition 4: When $\delta = 0$, dropping the time subscript, the regulator's objective function becomes:

$$
W\left(\{\Phi_i\}_{i\in\{1,2,3\}}\right) = -\sum_{i=1}^{3}\left(1-d_i\right)\Phi_i' - \frac{\kappa}{2}\Big(\sum_{i=1}^{3} d_i\Big)^2. \qquad \text{(A.28)}
$$

In addition, using equation (30) at $\delta = 0$, we can simplify the law of motion for Φ_i (as in (15)) into:

$$
\Phi_i' \equiv \left(\rho^2\Phi_i + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_i^* + \mu}{1-d_i^*}\right). \qquad \text{(A.29)}
$$

Taking the first-order condition gives that

$$
\frac{\partial W}{\partial d_i} = \Phi_i' - \kappa\Big(\sum_{i=1}^{3} d_i\Big). \qquad \text{(A.30)}
$$

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

Consider three cases. First, suppose that

$$
\left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1-\mu\right) < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1-d_1^*}\right), \qquad \text{(A.31)}
$$

that is, Φ_1 is much larger than Φ_2. We will restate condition (A.31) 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:

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) = \kappa d_1^*. \tag{A.32}$$

To verify that this is indeed an equilibrium, note first that the solution to (A.32) 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 (A.32), we can rewrite the condition (A.31) as:

$$\kappa d_1^* = \Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) > \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)(1 - \mu). \tag{A.33}$$

Since d_1^* is increasing in Φ_1, the condition (A.31) 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 (A.31) as

$$\Phi_1 > H(\Phi_2), \tag{A.34}$$

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

$$\begin{aligned}
\frac{\partial W}{\partial d_2} &= \Phi_2' - \kappa d_1^* \\
&= \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)(1 - \mu) - \kappa d_1^* \\
&< \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) - \kappa d_1^* \\
&= 0.
\end{aligned} \tag{A.35}$$

The third step uses (A.31). The last step uses (A.32).

Second, suppose that $\Phi_1 \leq H(\Phi_2)$ and

$$\left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)(1-\mu) < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right), \tag{A.36}$$

that is, Φ_1 and Φ_2 are of similar sizes but both are much larger than Φ_3. We will restate condition (A.36) 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:

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) = \kappa D^*, \tag{A.37}$$

$$\Phi_2' = \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_2^* + \mu}{1 - d_2^*}\right) = \kappa D^*, \tag{A.38}$$

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. In addition, applying the implicit function theorem gives that d_1^* is increasing in Φ_1 and decreasing in Φ_2 whereas d_2^* is increasing in Φ_2 and decreasing in Φ_1. Using the first-order condition of d_1, we can rewrite the condition (A.36) as:

$$\kappa D^* = \Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^* + \mu}{1 - d_1^*}\right) > \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)(1-\mu). \tag{A.39}$$

Since D^* is increasing in Φ_1 and Φ_2, the condition (A.36) 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 (A.36) as

$$L(\Phi_1, \Phi_2) > \Phi_3, \tag{A.40}$$

where $L(\cdot,\cdot)$ is some given increasing function in both Φ_1 and Φ_2. 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.,

$$\begin{aligned} \frac{\partial W}{\partial d_3} &= \Phi_3' - \kappa\left(d_1^* + d_2^*\right) \qquad \text{(A.41)} \\ &= \left(\rho^2 \Phi_3 + \sigma_\varepsilon^2\right)(1-\mu) - \kappa\left(d_1^* + d_2^*\right) \\ &< \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^* + \mu}{1 - d_1^*}\right) - \kappa\left(d_1^* + d_2^*\right) \\ &= 0. \end{aligned}$$

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

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^* + \mu}{1 - d_1^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right), \qquad \text{(A.42)}$$

$$\Phi_2' = \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_2^* + \mu}{1 - d_2^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right), \qquad \text{(A.43)}$$

$$\Phi_3' = \left(\rho^2 \Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_3^* + \mu}{1 - d_3^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^*\right). \qquad \text{(A.44)}$$

Applying the implicit function theorem gives that d_1^* is increasing in Φ_1 and decreasing in $\{\Phi_2, \Phi_3\}$, d_2^* is increasing in Φ_2 and decreasing in $\{\Phi_1, \Phi_3\}$ and d_3^* is increasing in Φ_3 and decreasing in $\{\Phi_1, \Phi_2\}$. ■

Appendix B. Data and Sample

We obtain accounting restatement data from Audit Analytics, implied volatility data from Option Metrics, and firm financial data from Compustat quarterly. Due to the limited coverage of the Audit Analytics database before 1999, our focus is on the period from 1999 to 2017. The baseline sample, after merging data, includes 151,048 firm-quarter observations. The number of observations in each analysis varies depending on the specification and variables used.

Our model centers on the interdependence of Φ, the fraud-induced information uncertainty; d, the fraud detection likelihood; and m, the fraud amount. To measure information uncertainty, we extract the implied volatility from options. 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 possible measurement error due to non-constant maturity, we use the implied volatility from 90-day standardized option prices provided by Option Metrics. Specifically, we first take the mean of the 90-day call- and put-implied volatility to capture the market's uncertainty about the firm's economic earnings. We then construct monthly (quarterly) implied volatility by taking the mean of daily implied values over a given month (quarter). We denote the resulting variable monthly (quarterly) IV, respectively.

To measure detection likelihood, we code $DETECT$ as an indicator variable that equals one if a fraud-related earnings restatement is announced in a quarter, and zero otherwise. Prior literature finds that not all restatements are related to fraud and some are unintentional misapplications of accounting rules (Hennes et al. (2008); Fang et al. (2017)). We define fraud-related restatements as those meeting at least one of the three following conditions: (1) if the restatement is marked as 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) is in the top decile of the restatement sample.

To measure fraud amount, we calculate $FRAUD$ as the firm's magnitude of fraud-related restatement in net income in the misstating quarter (proxied using the average restatement amount of all involved quarters), scaled by the standard deviation of quarterly operating income (after restatement, if any) measured over the most recent eight quarters. $FRAUD$ is coded as zero for all other firm-quarters. Again, we limit the calculation of $FRAUD$ to firms that are associated with fraud-related restatements as defined above. If the dollar amount of a fraud-related restatement is missing in Audit Analytics, we remove the firm-quarters that are associated with the restatement from the analyses involving $FRAUD$. Table B.I reports the descriptive statistics of the variables.

Table B.I: Descriptive Statistics

This table reports the firm-quarter-level statistics of the variables used in the analysis. IV is the quarterly average of the daily implied volatility. $DETECT$ is an indicator that denotes whether a firm discloses an accounting restatement that meets at least one of the three conditions: (1) if the restatement is marked as being fraudulent by Audit Analytics; (2) if the restatement has received a class-action lawsuit as tracked by Audit Analytics; or (3) if the cumulative restated amount (scaled by the total assets) is in the top decile of the restatement sample. $FRAUD$ is the magnitude of fraud-related restatement in net income in the misstating quarter (proxied using the average restatement amount of all involved quarters) scaled by the standard deviation of quarterly operating income of the most recent eight quarters.

Variables	N	Mean	SD	P25	Median	P75
IV	151,048	0.461	0.225	0.296	0.406	0.569
DETECT	151,048	0.008	0.090	0.000	0.000	0.000
FRAUD	147,234	0.016	0.152	0.000	0.000	0.000

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

european corporate governance institute

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (https://ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html