Everlasting Fraud

Law Working Paper N° 617/2021

June 2022

Vivian W. Fang
University of Minnesota 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
June 2022

Vivian W. Fang
Nan Li
Wenyu Wang
Gaoqing Zhang

We are grateful to helpful comments from Jeremy Bertomeu, Rebecca Hann, Xu Jiang, Jon Karpoff, Lorien Stice-Lawrence, Valeri Nikolaev, Anastasia Zakolyukina, Frank Zhou, an anonymous reviewer of the 2022 FARS meeting, and conference/seminar participants at the Accounting and Economics Society webinar, 9th SEC Annual Conference on Financial Market Regulation, Chinese University of Hong Kong, 2021 Columbia Burton Accounting Conference, 2022 FARS meeting, 2021 HKUST Accounting Research Symposium, 2022 MFA meeting, University of Maryland, University of Miami, and University of Science and Technology of China.

Abstract

This paper proposes a new theory to explain why corporate fraud waves always resurface despite tough anti-fraud regulations. Our model offers two insights. First, the interdependent nature of fraud and regulation presents a cat-and-mouse equilibrium within-firm because detection strength optimally matches fraud severity. Second, it yields a whack-a-mole equilibrium across-firm because regulatory resources are optimally concentrated on the most fraudulent firms. Therefore, regulations cannot eradicate fraud but synchronize firms' idiosyncratic fraud decisions, contributing to waves. These results carry strong policy implications by highlighting fraud as a permanent risk in the financial markets and the limited efficacy of anti-fraud regulations.

Keywords: Regulation, Financial Reporting, Accounting Fraud, Crime

JEL Classifications: G32, G34, G38, M40, M41, M48

Vivian W. Fang*
Associate Professor of Accounting
University of Minnesota, Carlson School of Management
321 19th Avenue South
Minneapolis, MN 55455, United States
phone: +1 612 301 9288
e-mail: fangw@umn.edu

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

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

Gaoqing Zhang
Associate 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
University of Minnesota, ECGI

Nan Li
University of Minnesota

Wenyu Wang
Indiana University

Gaoqing Zhang
University of Minnesota

June 6, 2022

Abstract

This paper proposes a new theory to explain why corporate fraud waves always resurface despite tough anti-fraud regulations. Our model offers two insights. First, the interdependent nature of fraud and regulation presents a cat-and-mouse equilibrium within-firm because detection strength optimally matches fraud severity. Second, it yields a whack-a-mole equilibrium across-firm because regulatory resources are optimally concentrated on the most fraudulent firms. Therefore, regulations cannot eradicate fraud but synchronize firms' idiosyncratic fraud decisions, contributing to waves. These results carry strong policy implications by highlighting fraud as a permanent 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

*Fang, Li, and Zhang are at the Carlson School of Management, University of Minnesota, Minneapolis, MN 55455; and Wang is at the Kelley School of Business, Indiana University, Bloomington, IN 47405. We are grateful to helpful comments from Jeremy Bertomeu, Rebecca Hann, Xu Jiang, Jon Karpoff, Lorien Stice-Lawrence, Valeri Nikolaev, Anastasia Zakolyukina, Frank Zhou, an anonymous reviewer of the 2022 FARS meeting, and conference/seminar participants at the Accounting and Economics Society webinar, 9th SEC Annual Conference on Financial Market Regulation, Chinese University of Hong Kong, 2021 Columbia Burton Accounting Conference, 2022 FARS meeting, 2021 HKUST Accounting Research Symposium, 2022 MFA meeting, University of Maryland, University of Miami, and University of Science and Technology of China. Please send correspondence to fangw@umn.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 expanded the efforts to fight fraud. The Act, via its Whistleblower Program, empowered the Securities and Exchange Commission (SEC) to reward whistleblowers in unprecedented ways.

Yet, did anti-fraud regulations, even the toughest ones, achieve 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)). In particular, 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 coming 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 build a unified, rational model of corporate fraud and regulation to illustrate the interdependent nature of the two and reconcile the actuality of regulation working to curb fraud as intended with the inevitability of fraud remaining a persistent feature of the financial markets despite the enactment of anti-fraud regulations.

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. An increase in the level of cumulative fraud raises information uncertainty about the firm, leading investors to place less weight on past reports and more weight on future reports, which in turn boosts the manager's incentive to inflate the forthcoming report. The regulator decides on the amount of resources spent on a detection technology. In doing so, she seeks to maximize the informativeness of the firm's 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 extent to which detection brings down uncertainty, which then allows investors to base the expected firm value on true economic cash flows rather than inflated figures; catching the firm with a higher level of cumulative fraud clears more information uncertainty in both the 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.[1]

Starting with the regulator, her goal is to minimize information uncertainty for investors. With this task at hand, she faces a trade-off: a higher Φ_t motivates her because fraud detection is more valuable in restoring information precision (MB) but increasing detection

[1]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)). Neverthless, since fraud and fraud-induced information uncertainty are positively correlated in our model, our inferences are likely qualitatively similar with either concept.

strength inevitably consumes more regulatory resources (MC). Indeed, as motivating evidence for this trade-off, we show that an empirical proxy for Φ_t—the implied volatility of standard options—drops sharply upon revelation of fraud, suggesting that cumulative fraud raises information uncertainty and detection lowers it. Turning to the manager, his goal is to boost firm valuation for private benefits. He similarly faces a trade-off: a higher Φ_t not only incentivizes him because investors are eager for new information and will value even the biased report (MB) but also disciplines him because the regulator also chooses a higher detection likelihood (MC). While the cost side of this trade-off is clear, the benefit side requires some justification. As evidence consistent with information uncertainty boosting the value of earnings reports and the potential return from reporting fraudulently, we show that financial analysts' revision of future earnings estimates is more responsive to unexpected earnings of the current reports when implied volatility (the empirical proxy for Φ_t) is higher. We detail motivating and supporting evidence in Section I.

In equilibrium, the regulator chooses the optimal level of detection likelihood (by spending a corresponding amount of resources on detection), anticipating the cumulative fraud committed by the manager, and vice versa. If the regulator anticipates a low level of information uncertainty, then she would spend little on detection. As such, the manager continues to commit fraud because his MB likely outweighs MC. As fraud gradually builds up, it induces more information uncertainty that further incentivizes the manager to report fraudulently as discussed earlier. Meanwhile, the regulator would increase spending on detection as more fraud attracts closer scrutiny. The two effects go hand-in-hand, simultaneously increasing the manager's MB and MC. When fraud reaches a critical level, the regulator would concentrate resources on the firm and the MC of continuing to commit fraud eventually outweighs the MB. 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 effective regulation but everlasting fraud. 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 regulatory resources concentrated 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 behavior is better masked 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.

The question then arises is whether regulations can still achieve their stated goals of cracking down on fraud. Analyses of the multi-firm model show that the efficacy of anti-fraud regulations is quite nuanced. On the one hand, such regulations are able to effectively tamp down fraud by lowering H-firm's net benefits from continuing fraud. Before detection, concentration of regulatory resources on H-firm greatly increases its MC of committing fraud. Upon detection, the MB becomes minuscule because a sharply declined uncertainty renders the firm's future reports less useful and fraudulent reporting less valuable. On the other hand, the rational allocation of regulatory resources towards the more fraudulent firms may imply less scrutiny of less fraudulent firms, allowing the latter's fraudulent behavior to go undetected and their level of fraud to catch up—a side effect revealed by our three-firm model. As a result, despite the "cracking-down" on H-firm, anti-fraud regulations do not eradicate fraud. Rather, they synchronize firms' fraud decisions, which may otherwise be idiosyncratic, and induce corporate fraud waves—the convergence of firms' fraud amount—at certain times. To lend support to this unique insight from our model, we show that firms with a higher (lower) level of implied volatility in the prior period experience a greater (smaller) increase in implied volatility in the current period, suggesting that firms with more (less) fraud-induced information uncertainty are more cautious (aggressive) about committing fraud.

This paper fits in the broad literature of crime and regulation in finance and economics in three ways. First, it offers consistent evidence for the "whack-a-mole" game, a term adopted by prior studies to describe the situation in which a regulation intended to address problems in one market may inadvertently lead to new problems cropping up in other markets (Blinder (2008); Blinder (2015); Cai, He, Jiang, and Xiong (2021)). In our model, this situation arises because uncovering corporate fraud inevitably consumes regulatory resources. As a result, maximizing detection intensity at all times is neither fiscally feasible nor socially optimal, so strategic reporting behavior in certain parts of the economy is inescapable. Additionally, fraud waves, as predicted by our model, are interactive outcomes between the defensive calculated behaviors of incumbent managers and the offensive tactical innovations of entrepreneurial monitors; similar outcomes are observed in other markets such as takeover waves (Pound (1992); Harford (2005)).

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

Lastly, our model complements prior theories of fraud but also differs from them in two notable aspects. First, to our best knowledge, our model is the first to examine the joint mechanisms of corporate fraud and regulation in a multi-firm, dynamic setting. In contrast, prior theories either assume an exogenous detection cost for a single-firm game (e.g., see Fischer and Verrecchia (2000) and Dye and Sridhar (2004) for two static models and Beyer, Guttman, and Marinovic (2019) for a dynamic model) or endogenize detection cost in a single-firm, static setting (e.g., Povel, Singh, and Winton (2007)). By modeling both the manager's and regulator's decisions, our study takes a holistic view in analyzing the formation and evolvement of corporate fraud and evaluating the efficacy of anti-fraud regulations. Second, several prior theories highlight growth as an important driver for fraud (e.g., Povel et al. (2007); Strobl (2013)), which are useful in explaining why corporate fraud may come in waves depending on business cycle but not why regulations lead or even give rise to fraud waves. 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 as a reduction in price informativeness is capable of distorting investment and hurting the real economy (see Goldstein and Yang (2017) for a survey).

I. Motivating Evidence

Fraud has existed for as long as recorded history, with the earliest documented case dating back to Ancient Greece, 300 B.C. (Toms (2019)). The history of financial markets is shorter in comparison but never lacked episodes of corporate fraud. Using a comprehensive sample worldwide from 1800 to 2015, Hail, Tahoun, and Wang (2018) conduct a descriptive study of the intertemporal relation of corporate fraud and regulation, and reveal a pattern of interactive dynamics between the two. Figure 1 reproduces Figure 2 of Hail, Tahoun, and Wang (2018, pp. 645). As the figure shows, fraud is an antecedent to regulation over long stretches of time, which is fully expected as regulators are typically reactive rather than proactive. Interestingly, regulation is 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,

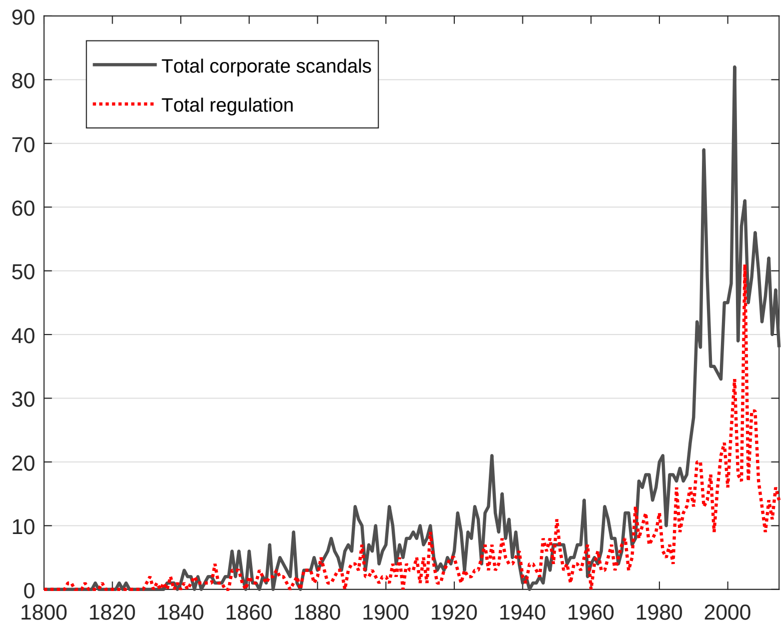

Figure 1: Time-series plot of the yearly number of business news mentions of "scandal" and "regulator" for 24 countries from 1800 to 2015, reproduced from Figure 2 of Hail et al. (2018)

new anti-fraud measures may not be exhaustive or contain loopholes that allow managers to adapt and circumvent regulation. Third, regulations can have unintended consequences, although it is not clear what these consequences may be. 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, which is motivated by the view held by regulators worldwide that fraud in firms' disclosure and reporting choices can influence market valuation and mislead investors.[2] As a screening test of whether our proposed theory indeed reflects practice, we conduct two sets of analyses to separately shed light on the role that information may play in the regulator and the firm manager's calculus. Detailed sample

[2]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).

selection, variable definitions, and empirical results are discussed in Appendix II.

Starting with the regulator, if her goal is to crack down on fraud and restore the accuracy of corporate reports (which is a stated objective of many anti-fraud regulations), then it is reasonable to expect that fraud builds up information uncertainty while subsequent detection decreases it. To check this prior, we obtain a sample of accounting restatements announced between 1999 and 2019 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 all available information, implied volatility captures the conditional variance of this information set. We thus expect implied volatility to increase with Φ_t, the key state variable in our model that represents the level of information uncertainty brought by cumulative fraud. 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. We document two new stylized facts. First, as Figure 2 shows, implied volatility drops sharply upon revelation of fraud, consistent with our prior that detection clears fraud and lowers information uncertainty.

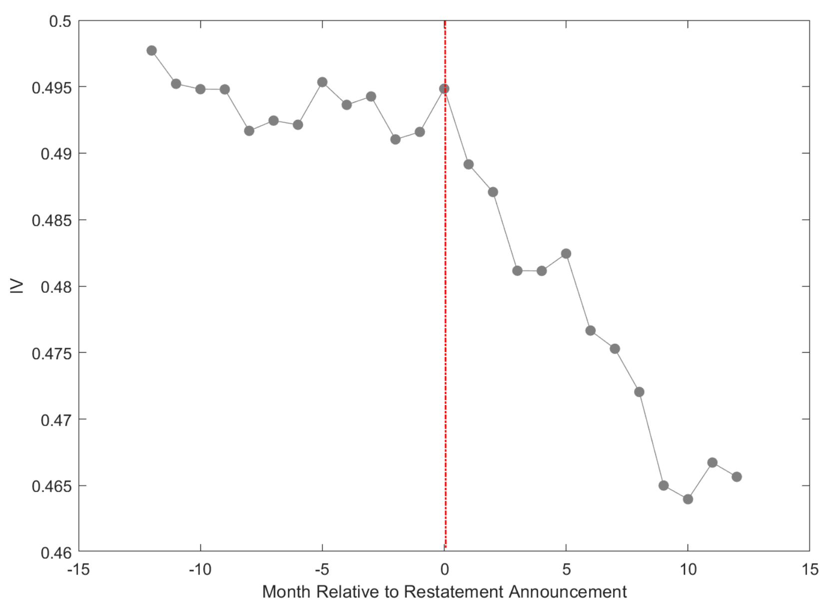

Figure 2: Plot of monthly IV before and after fraud-related restatement announcements

Second, Table I shows that the likelihood of having fraud revealed in a given quarter is positively related to the level of implied volatility during the prior quarter, suggesting that regulatory resources are concentrated on firms with higher levels of fraud-induced information uncertainty. Combined, these two findings point to the role that information plays in the regulator's calculus: as detection helps restore information precision, she benefits more by directing efforts towards firms with higher levels of fraud-induced information uncertainty.

Turning to the manager, we first note that since detection likelihood increases with implied volatility (the empirical proxy for fraud-induced information uncertainty, Φ_t), the manager's cost of committing fraud is also expected to increase with implied volatility. To shed light on how information uncertainty affects the manager's benefit from committing fraud, we link implied volatility to analyst forecast revision, a commonly used proxy for changes in the market's expectation about the firm's value. In Table II, we show that analysts' revision of earnings estimates for the next quarter following earnings announcement of the current quarter is positively related to implied volatility during the prior quarter. This result suggests that the market is likely to put a greater weight on the new information contained in the firm's financial report when information uncertainty is higher. This discussion thus suggests the manager faces a trade-off: while a higher level of fraud-induced information uncertainty deters the manager from committing fraud because the regulator chooses a higher detection likelihood, it also incentivizes him because investors are eager for new information. Consistent with this intuition, we find an inverse U-shaped association between a firm's fraud amount in a quarter (labeled $FRAUD$) and the level of implied volatility of the quarter, both in the univariate plot (Figure 3) and regression analyses (Table III).

To summarize, although the empirical evidence documented in this section are far from conclusive, they are suggestive of the role that information plays in the regulator and the firm manager's calculus. We are unaware of a pre-existing model that can explain these findings, which motivates the development of our model in the next section.

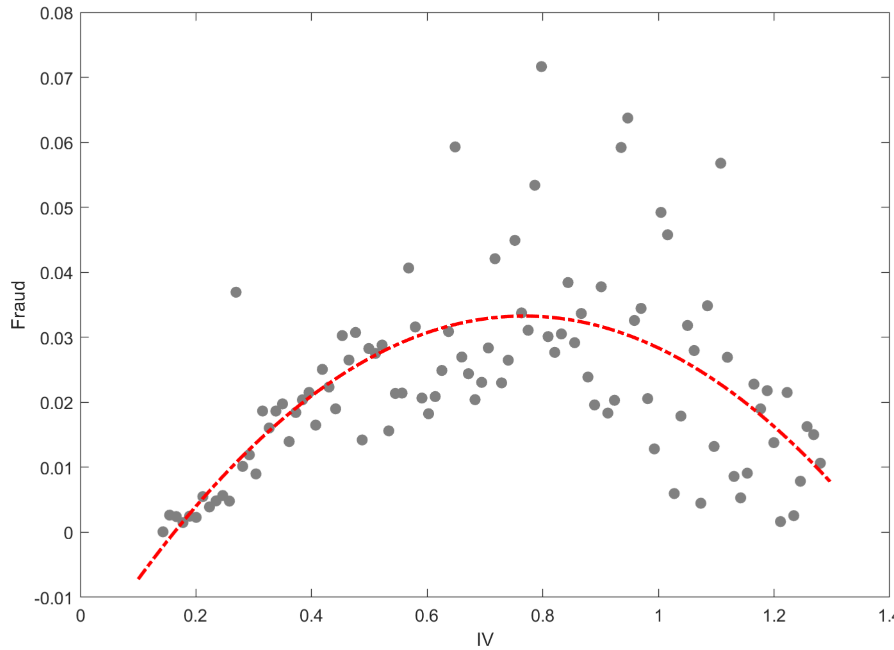

Figure 3: Plot of $FRAUD$ against IV with prediction curve

II. Single-firm Model

A. Model Setup

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

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

where the correlation coefficient $\rho \in (0, 1)$ and the random variable $\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$. In each period, the firm manager privately learns the realization of the firm's economic earnings s_t and issues a report r_t. Investors use the report to update 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 $E^I[\cdot|\mathcal{F}_t]$ 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 as follows. In each period t, after observing the true economic earnings s_t, the manager chooses manipulation $m_t \geq 0$ that adds m_t errors $\{\xi_l\}_{l=1}^{m_t}$ to s_t. The choice of manipulation m_t is observable only to the manager. Each error generates either 0 or 1 with $\Pr(\xi_l = 0) = q \in (0,1]$. The report is then given by:[3]

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

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

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

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

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

$\eta_t \sim N(0,1)$ is a standard normal random variable that is independent of all other variables in the model. Equation (5) suggests that manipulation has a dual effect on the report: m_t increases the mean of the report but decreases its precision. Note that we do not impose the restriction that manipulation in the report must reverse at some fixed point in time, i.e., the mechanical reversal of discretionary accruals. However, because the regulator's equilibrium detection effort increases with the cumulative fraud, a firm who has engaged in more

[3] 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.

manipulation in the past will have stronger incentive to reduce manipulation in the future.

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

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

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

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

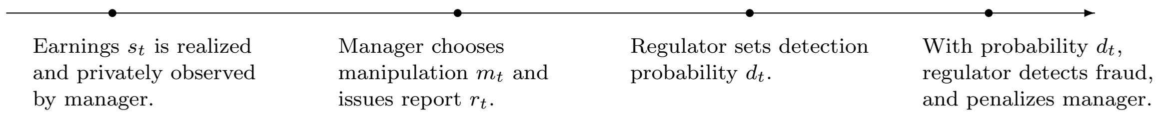

Figure 4: Timeline of the period-t game

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.[5]

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

B. Analysis

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

B.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,

[5]If the regulator can uncover and impose a penalty on past fraud, the manager will have weaker incentives to commit fraud. However, in terms of the qualitative predictions of the model, this modification will yield similar effects to imposing a greater penalty on the current-period fraud, i.e., a higher c.

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

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:

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

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

The two terms of equation (9) represent the manager's utility under two different scenarios, respectively. In the first scenario 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 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 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:

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

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 $\rho E^I [s_{t-1}|\mathcal{F}_{t-1}]$. To extract information from the new report, investors rationally subtract the expected manipulation $m_t^*(1-q)$, leading to a refined signal $r_t - m_t^*(1-q)$. The weight

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

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

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

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

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

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

obtain the first-order condition (F.O.C.) as[9]

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

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 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 I.A 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*

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

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

[9]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 one and the only rational-expectation equilibrium. Note that by adding even a small convexity in the cost of manipulation, e.g., making the manipulation penalty $C_t = c(r_t - s_t) + \frac{\epsilon}{2}(r_t - s_t)^2$, where $\epsilon > 0$ can be arbitrarily small, we can show that, at the market conjecture of m_t^*, the manager strictly prefers to choose m_t^*. Detailed analysis is available upon requests.

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^**

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

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., $\left\{m_t^*, m_{t-1}^*, ...\right\}$. 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.

B.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

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

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

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

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 $\frac{\kappa}{2}(d_t - d_0)^2$.

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

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

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

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

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

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

where Φ_t is expressed recursively in equation (7).

B.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:*

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

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

where

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

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

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 I to solve the full model.

When $\delta = 0$, all intertemporal effects exit and only the contemporaneous effect remains.

The F.O.C. of the manipulation decision m_t^* in equation (13) is reduced into:

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

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 $E^I\left[s_t|\mathcal{F}_t\right]$. 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^*:

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

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^*:

$$\underbrace{\kappa\,(d^*-d_0)}_{\text{MC of } d} = \underbrace{\left(\rho^2\Phi+\sigma_\varepsilon^2\right)\left(1-\frac{cd^*}{1-d^*}\right)}_{\text{MB of } d \text{ from the contemporaneous effect}}. \tag{26}$$

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 equation (A.10) to equation (26) shows that the regulator's choice of optimal detection strength d^* is increasing in the fraud-induced uncertainty Φ. That is, the regulator matches the strength of fraud detection with the severity of fraud in equilibrium. This result is intuitive because the manager's manipulation in the past adds noises to the firm's reports and decreases informativeness. Since the regulator's objective is to clear fraud and restore informativeness of the firm's reports, her gains are higher from detecting reports with more extensive fraud. In other words, the regulator's MB of detection is increasing in the cumulative level of fraud and so is her choice of optimal detection strength.

Next, we use the 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 and Table III). 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:*

$$\kappa\left(d^*(\Phi) - d_0\right) = \left(\rho^2\Phi + \sigma_\varepsilon^2\right)\left(1 - \frac{cd^*(\Phi)}{1 - d^*(\Phi)}\right), \tag{27}$$

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

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. We set the seven model parameters as follows: the subjective discount rate, δ, equals 0.9, a value commonly used in the literature; the success rate of manipulation, q, equals 0.5, an innocuous assumption in the model; the persistence of the AR(1) process that governs the dynamics of the economic earnings s_t, ρ, equals 0.88; the conditional standard deviation of the AR(1) process, σ_ε, equals 0.15; the detection cost parameter, κ, equals 2.5; and the manager's cost parameter, c, equals 3, which suggests that the fine imposed on the manager is three times of his manipulation amount upon detection. Last, we set the exogenous detection rate by the market participants other than the regulator, d_0, to be 3%.

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}. Based on the set of parameter values that we use in the numerical solution, we find that the equilibrium manipulation is non-monotonic in the firm's cumulative fraud. The intuition is clear: when Φ is very low (close to 0), the market is highly informed about the firm's economic earnings and puts little weight on the firm's new report. This implies a low MB of manipulation and weaker incentives for the manager to inflate the report. When Φ is very high, the regulator increases detection efforts, which sharply increase the MC of manipulation. Trading off the MB and MC of manipulation, the manager's maximum manipulation may appear in the intermediate range of Φ, leading to an inverse U-shaped relation between m^* and Φ.

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

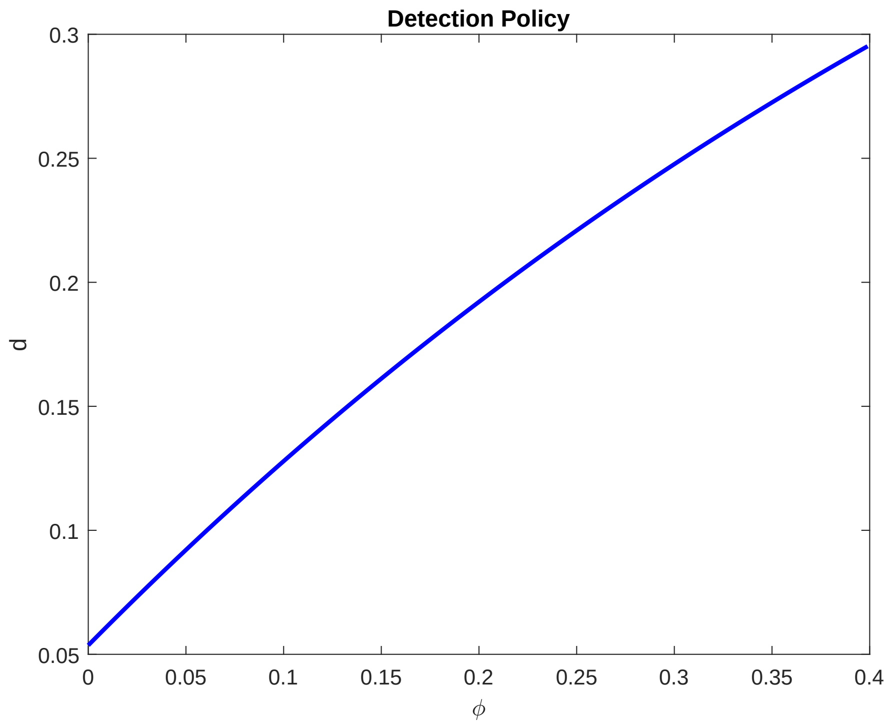

Figure 5: Equilibrium detection probability $d^*(\Phi)$. This figure plots the regulator's optimal choice of detection intensity, d, as a function of the information uncertainty induced by cumulative fraud, Φ. The parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

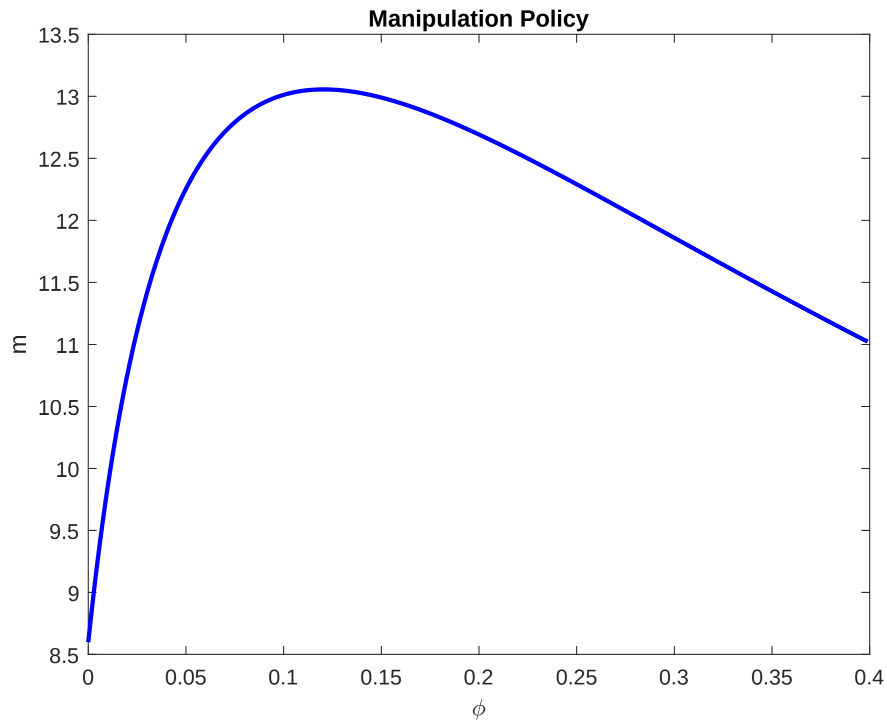

Figure 6: Equilibrium manipulation $m^*(\Phi)$. This figure plots the manager's optimal choice of manipulation, m, as a function of the information uncertainty induced by cumulative fraud, Φ. The parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

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 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 of 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.

Another interesting observation from Figure 7 is that when the discounting factor increases, the equilibrium manipulation is more likely to be inverse U-shaped in the firm's cumulative fraud. To see the intuition, recall that the equilibrium manipulation is non-monotonic in the cumulative fraud Φ only when the MB and the MC of manipulation move in tandem with the level of Φ. From the second panel of Figure 7, the regulator is most likely to respond to a jump in Φ and increase her detection effort when the discounting factor is large and any fraud-induced uncertainty is highly detrimental. This sharper response by the regulator makes the MC of manipulation more likely to countervail the positive effect of Φ on the MB of manipulation, leading to the inverse U-shaped relationship.

Figure 8 sheds light on how increasing the persistence in the firm's cash flows across different periods (i.e., the correlation ρ) alters the equilibrium outcomes. It suggests that when the cash flows become more persistent over time, it will reinforce both the manager's manipulation and the regulator's detection incentives, but reduce the social surplus. The economic forces driving the effect of ρ are largely in line with that of δ. Intuitively, when cash flows are more correlated across periods, the manager's inflation of the current-period report boosts firm valuation in the future by a greater extent, thus contributing to a stronger intertemporal effect of manipulation and resulting in more manipulation in equilibrium. Similarly, when

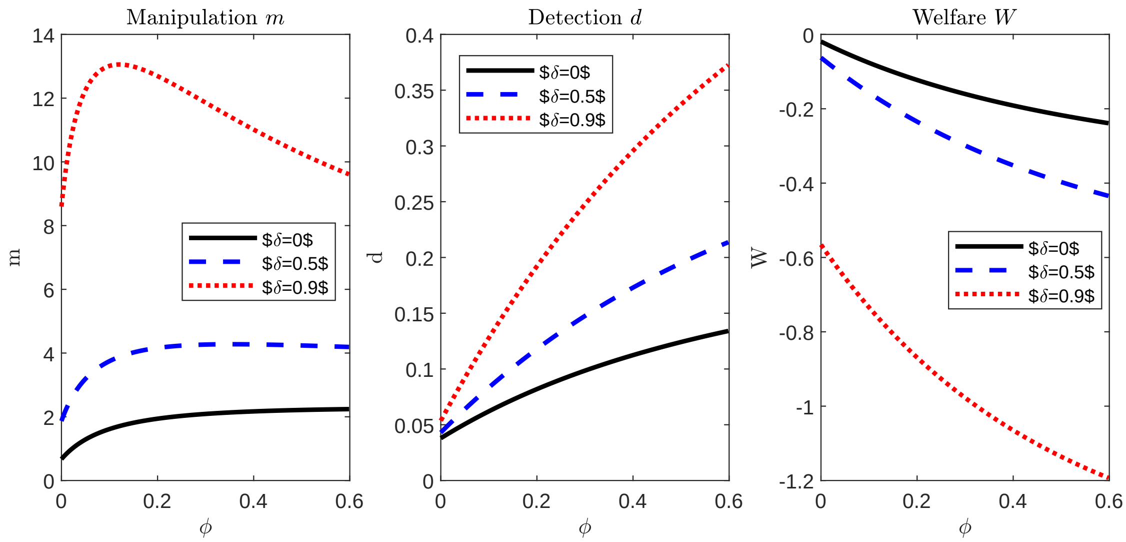

Figure 7: Comparative statics w.r.t. δ. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the discount factor, δ, is set to 0 (black solid line), 0.5 (blue dashed line), and 0.9 (red dotted line). Other parameters used in generating this figure are: $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

cash flows are more correlated, the regulator has stronger incentives to detect manipulation in the current period as diminishing the current-period uncertainty translates into a greater reduction in the future uncertainty due to the higher persistence in the cash flows. In the mean time, the social surplus decreases in ρ since a higher cash flow persistence also implies that any uncertainty unresolved in this period will persist in the future.

Similar to the effect of δ, Figure 8 also suggests that the equilibrium manipulation is more likely to be inverse U-shaped in the firm's cumulative fraud when the cash flows are more persistent. Intuitively, the second panel of Figure 8 suggests that the detection effort by the regulator increases more sharply in Φ when the persistence parameter ρ is high. An important thing to notice is that, when $\rho = 0$, firm performance becomes i.i.d and thus investors would not use past performance to infer the current and future performance. As a result, the effect of fraud committed in the past does not build up uncertainty in the firm's market value, and thus the regulator's detection effort would be independent of Φ. When the persistence parameter is large, the regulator also cares about the cumulative fraud as the

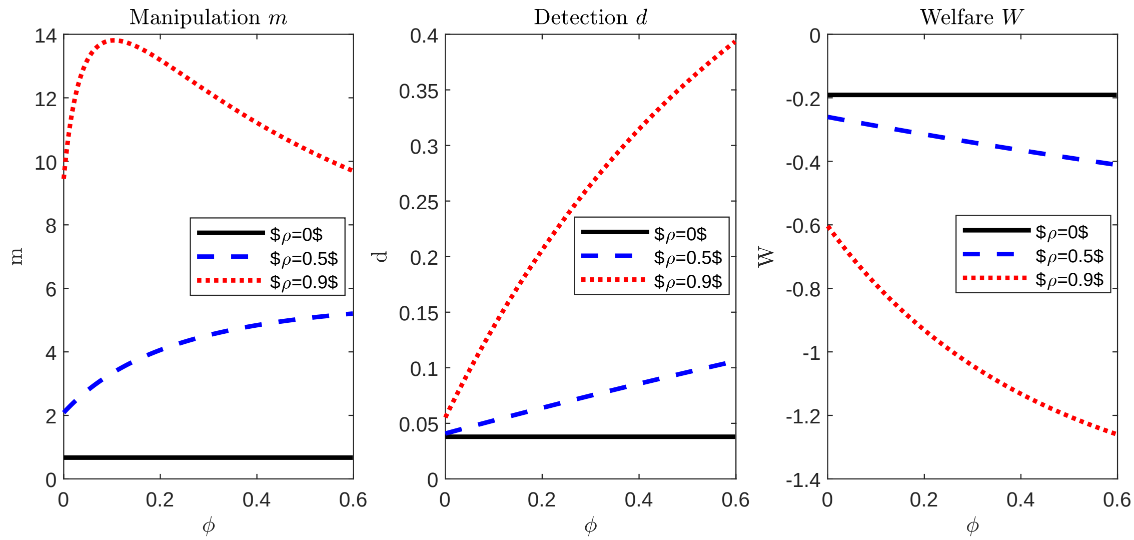

Figure 8: Comparative statics w.r.t. ρ. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the discount factor, ρ, is set to 0 (black solid line), 0.5 (blue dashed line), and 0.9 (red dotted line). Other parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

cash flows are correlated across periods. As a result, the equilibrium detection effort becomes more responsive to Φ, which partly offsets the boosting effect of Φ on the MB of manipulation and makes the equilibrium manipulation non-monotonic in Φ.

Figures 9 and 10 describe the effects of shifting the costs of manipulation and detection. The intuitions for both figures are relatively straightforward. When the MC of manipulation c increases, it disciplines the manager's manipulation incentive 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. Similarly, when the MC of detection k 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.

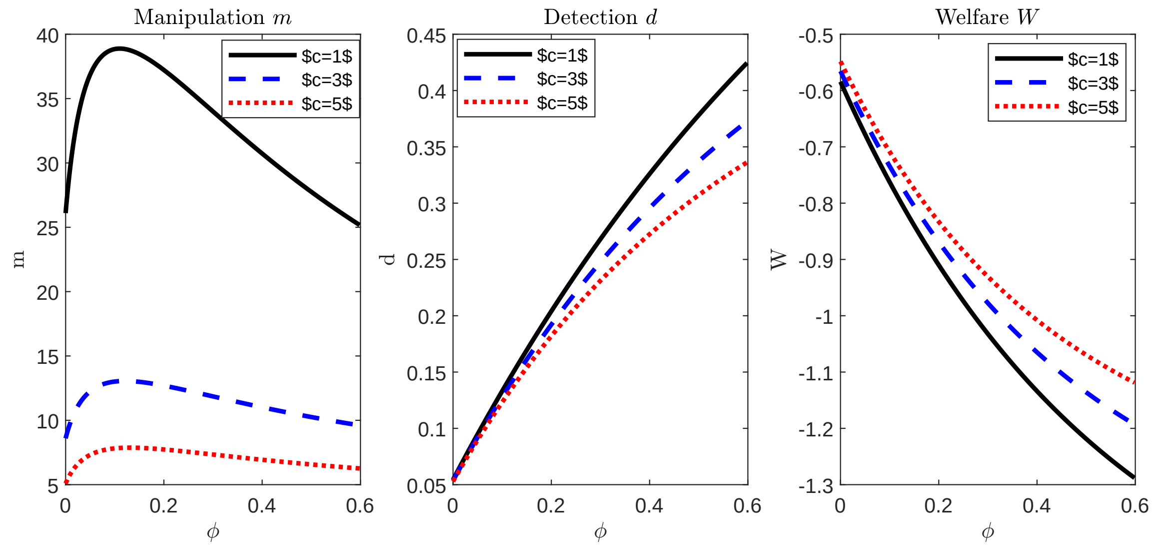

Figure 9: Comparative statics w.r.t. c. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the penalty on fraud, c, is set to 1 (black solid line), 3 (blue dashed line), and 5 (red dotted line). Other parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $d_0 = 0.03$.

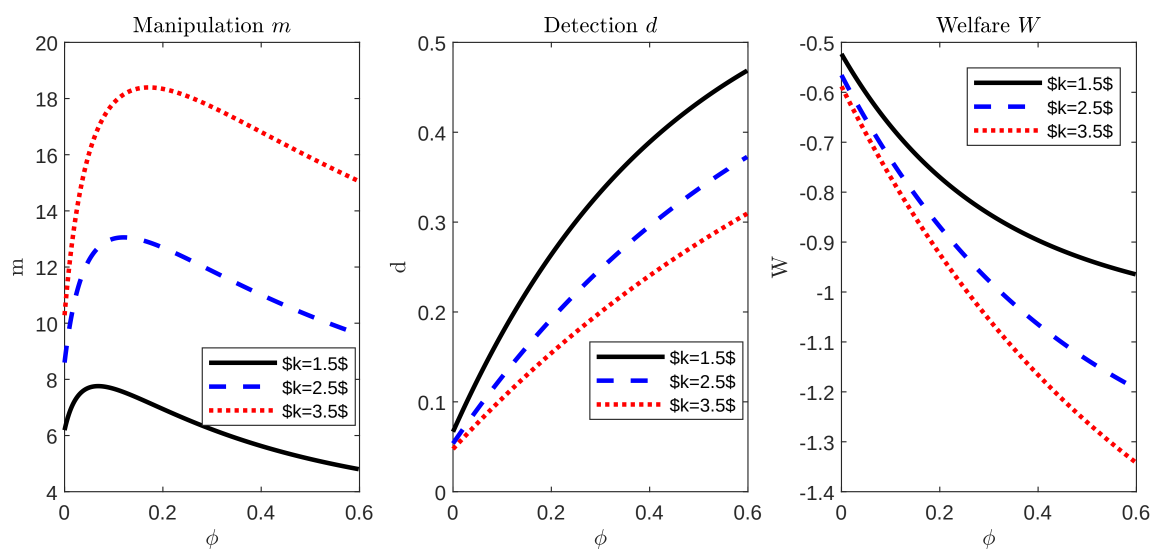

Figure 10: Comparative statics w.r.t. k. This figure plots the manager's optimal choice of manipulation m, the regulator's optimal choice of detection intensity d, and the equilibrium welfare W, when the cost of detection, d, is set to 1.5 (black solid line), 2.5 (blue dashed line), and 3.5 (red dotted line). Other parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $c = 3$, $d_0 = 0.03$.

III. Multi-firm Model

A. 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 $d_{it} = d_0 + d_{irt} \in [d_0, 1]$, where $i \in \{1, 2, ..., N\}$ and d_{irt} represents the regulator's choice of detection technology to influence the probability of detecting fraud at firm i. We assume that the total detection cost for each period is:

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

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.

B. 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*

$$m_1^*\left(\Phi_1, \Phi_2, \Phi_3\right) = \frac{\rho^2\Phi_1 + \sigma_\varepsilon^2}{q\left(1-q\right)}\left(\frac{1-d_1^*}{cd_1^*}\left(\frac{1}{1-\delta\rho} + \frac{c\delta\rho q\left(1-q\right)}{\rho^2\Phi_1' + \sigma_\varepsilon^2} \times E\left[m_1^{*\prime} d_1^{*\prime}\right]\right) - 1\right), \quad (30)$$

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

$$\begin{aligned} d_1^*\left(\Phi_1, \Phi_2, \Phi_3\right) = \frac{1}{\kappa}\{&\Phi_1' + \delta\left(1-d_2^*\right)\left(1-d_3^*\right)\left[W\left(0, \Phi_2', \Phi_3'\right) - W\left(\Phi_1', \Phi_2', \Phi_3'\right)\right] \\ &+ \delta\left(1-d_2^*\right)d_3^*\left[W\left(0, \Phi_2', 0\right) - W\left(\Phi_1', \Phi_2', 0\right)\right] \\ &+ \delta d_2^*\left(1-d_3^*\right)\left[W\left(0, 0, \Phi_3'\right) - W\left(\Phi_1', 0, \Phi_3'\right)\right] \\ &+ \delta d_2^* d_3^*\left[W\left(0,0,0\right) - W\left(\Phi_1', 0, 0\right)\right]\} - \left(d_2^* + d_3^* - 3d_0\right), \end{aligned} \quad (31)$$

where

$$\Phi_i'\left(\Phi_1, \Phi_2, \Phi_3\right) = \frac{m_i^* q\left(1-q\right)\left(\rho^2\Phi_i + \sigma_\varepsilon^2\right)}{\rho^2\Phi_i + \sigma_\varepsilon^2 + m_i^* q\left(1-q\right)}, \quad (32)$$

$$\begin{aligned} E\left[m_1^{*\prime} d_1^{*\prime}\right] = &\left(1-d_2^*\right)\left(1-d_3^*\right)m_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right)d_1^*\left(\Phi_1', \Phi_2', \Phi_3'\right) \\ &+ d_2^*\left(1-d_3^*\right)m_1^*\left(\Phi_1', 0, \Phi_3'\right)d_1^*\left(\Phi_1', 0, \Phi_3'\right) \\ &+ \left(1-d_2^*\right)d_3^* m_1^*\left(\Phi_1', \Phi_2', 0\right)d_1^*\left(\Phi_1', \Phi_2', 0\right) \\ &+ d_2^* d_3^* m_1^*\left(\Phi_1', 0, 0\right)d_1^*\left(\Phi_1', 0, 0\right), \end{aligned} \quad (33)$$

$$
\begin{aligned}
W\left(\Phi_1,\Phi_2,\Phi_3\right) = &-\left(1-d_1^*\right)\left(1-d_2^*\right)\left(1-d_3^*\right)\left[\Phi_1'+\Phi_2'+\Phi_3'-\delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
&-\left(1-d_1^*\right)d_2^*\left(1-d_3^*\right)\left[\Phi_1'+\Phi_3'-\delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\
&-\left(1-d_1^*\right)\left(1-d_2^*\right)d_3^*\left[\Phi_1'+\Phi_2'-\delta W\left(\Phi_1',\Phi_2',0\right)\right] \\
&-\left(1-d_1^*\right)d_2^*d_3^*\left[\Phi_1'-\delta W\left(\Phi_1',0,0\right)\right] \\
&-d_1^*\left(1-d_2^*\right)\left(1-d_3^*\right)\left[\Phi_2'+\Phi_3'-\delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\
&-d_1^*d_2^*\left(1-d_3^*\right)\left[\Phi_3'-\delta W\left(0,0,\Phi_3'\right)\right] \\
&-d_1^*\left(1-d_2^*\right)d_3^*\left[\Phi_2'-\delta W\left(0,\Phi_2',0\right)\right] \\
&+\delta d_1^*d_2^*d_3^*W\left(0,0,0\right)-\frac{\kappa}{2}\left(d_1^*+d_2^*+d_3^*-3d_0\right)^2. \qquad (34)
\end{aligned}
$$

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 I.F.*

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., $\Phi_1 > H(\Phi_2)$), $d_2^* = d_3^* = d_0$ and d_1^* solves*

$$\left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^*}{1 - d_1^*}\right) = \kappa\left(d_1^* - d_0\right), \tag{35}$$

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 $L(\Phi_1, \Phi_2) > \Phi_3$), $d_3^* = d_0$ and the pair of $\{d_1^*, d_2^*\}$ solves*

$$\left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^*}{1 - d_1^*}\right) = \kappa\left(d_1^* + d_2^* - 2d_0\right), \tag{36}$$

$$\left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_2^*}{1 - d_2^*}\right) = \kappa\left(d_1^* + d_2^* - 2d_0\right). \tag{37}$$

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:*

$$\left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_1^*}{1 - d_1^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right), \tag{38}$$

$$\left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_2^*}{1 - d_2^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right), \tag{39}$$

$$\left(\rho^2 \Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{c d_3^*}{1 - d_3^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right). \tag{40}$$

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 parameter values set for the one-firm model: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, and $d_0 = 0.03$. 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,

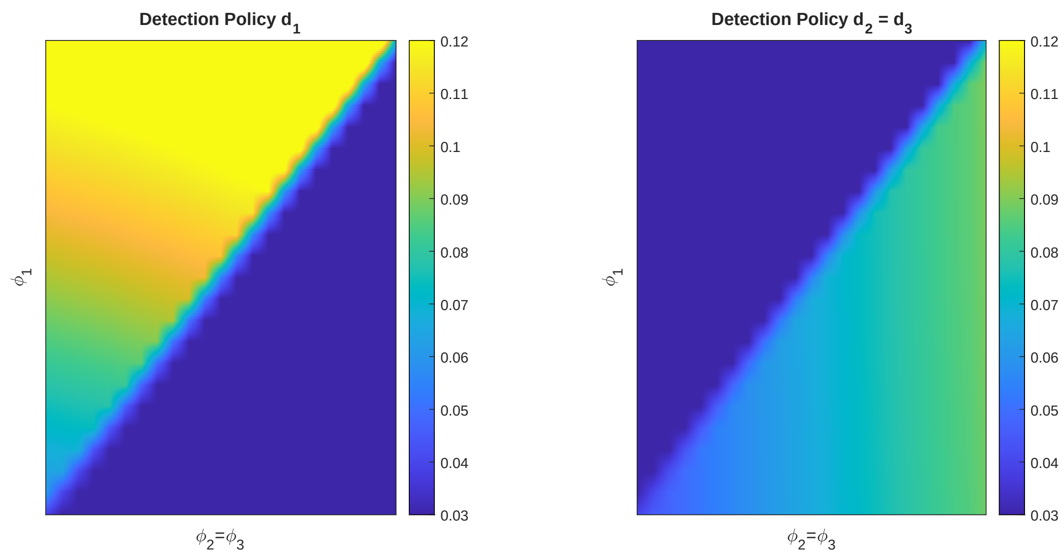

Figure 11: Equilibrium detection probability d_i^* in the three-firm setting. This figure plots the manager's optimal choice of manipulation m for firm 1 (left panel) and firm 2 and 3 (right panel) as a function of Φ_1, Φ_2, and Φ_3 in a heatmap. We assume that $\Phi_2 = \Phi_3$ in this figure. The parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

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 0.12) in the first scenario is much larger than the detection intensity imposed on firm 2 and 3 (about 0.08, respectively) in the second scenario. This is because the regulator's cost of detection is convex in the aggregate detection intensity, as shown in equation (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 outweighs the MB, leading the manager to be more conservative. Ironically, as the firm with the highest information uncertainty attracts the most attention by the regulator, other firms are subject to less scrutiny and can afford to become more aggressive in committing fraud. To the extent that manipulation in each period accumulates and adds to the firms' information uncertainty over time, our model predicts an unintended consequence of regulation: it synchronizes managers' manipulation decisions and may eventually lead to fraud waves even in the absence of systematic shocks in the economy.

We next use the model to study the dynamics of the fraud-detection game between the regulator and three firms with different levels of fraud-induced information uncertainty in the initial period. Without loss of generality, we assume that $\Phi_H > \Phi_M > \Phi_L$ at $t = 1$ and denote the three firms H-, M- and L-firm, respectively. We then simulate the magnitude of manipulation committed by each manager m_i, the regulator's detection policy on each firm

d_i, and the realization of detection outcomes at the end of each period. As we simulate the model forward, it generates the time series of Φ_{it}, d^*_{it}, and m^*_{it}. Figure 12 plots the three variables over the simulation path.

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

M-firm (depicted by the brown-solid line) starts with an intermediate level of cumulative fraud. On the one hand, the manager of M-firm has greater incentives to commit fraud than the manager of L-firm, because a higher Φ increases the MB of committing fraud. On the other hand, the regulator invests more heavily in fraud detection of M-firm than L-firm, which suggests a higher MC of committing fraud. The two effects go hand-in-hand. The first five periods of the brown-solid line in Figure 12 show the stage when MB dominates MC, and thus m_M increases over time as Φ_M grows. After the sixth period, we observe that the detection intensity on M-firm quickly rises (see the brown-solid line in Panel B) and MC outweighs MB, leading to a sharp decline in manipulation by M-firm (see the brown-solid line in Panel C). The dynamics in m_M therefore demonstrates the counteracting forces of MB and MC.

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

To examine the impact of actual detection, we assume in the simulation trial that H-firm is caught by the regulator at period 30. Upon detection, H-firm's cumulative fraud is cleared and Φ_H drops to zero immediately, as shown in Panel A. As an optimal response, the regulator shifts attention from H-firm to the original M- and L-firms, as shown in Panel B.

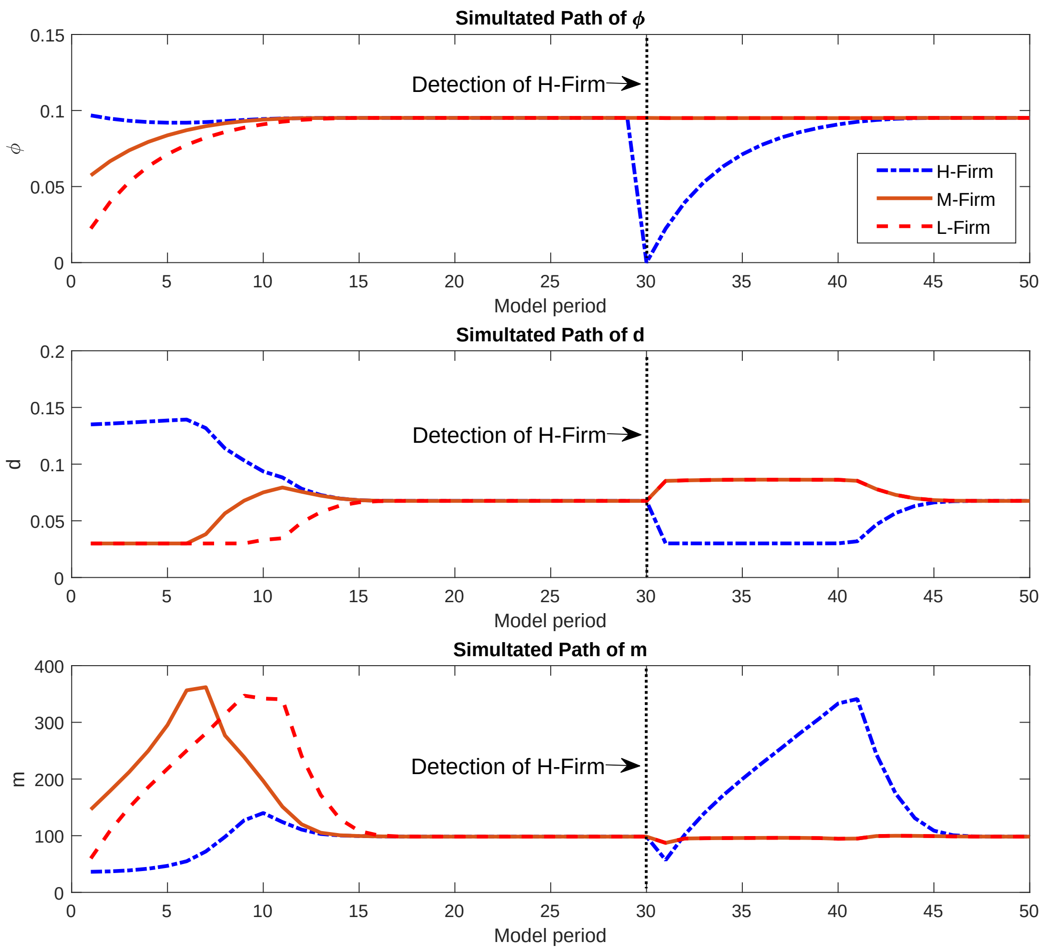

Figure 12: Simulated paths of the three-firm model. This figure plots the simulated path of information uncertainty induced by cumulative fraud Φ, the regulator's optimal choice of detection intensity d, and the manager's optimal choice of manipulation m for firms with high- (H), medium- (M), and low- (L) level of initial Φ. In the simulation, H-firm is detected at period 30. The parameters used in generating this figure are: $\delta = 0.9$, $q = 0.5$, $\rho = 0.88$, $\sigma_\varepsilon = 0.15$, $k = 2.5$, $c = 3$, $d_0 = 0.03$.

Interestingly, as the detection intensity on H-firm drops substantially, H-firm faces a low MC of committing fraud and can now afford to become more aggressive in manipulating its report. This explains the sharp increase in m_H and Φ_H in Panel C and A right after period 31. If M- and L-firms remain undetected, cumulative fraud in the three firms will be synchronized again after another few periods. This analysis sheds further light on an unintended consequence of regulation: it may synchronize firms' manipulation decisions and lead to fraud waves even in the absence of aggregate shocks. The intuition is simple: anticipating the optimal allocation of regulatory resources in the economy, firms with a low level of cumulative fraud endogenously choose a high level of manipulation, allowing them to catch up to more fraudulent firms.

Indeed, the predicted convergence of fraud level across firms over time gains support from the data. When we sort firm-quarters in the sample into quintiles based on the firm's level of implied volatility prior to a quarter (our empirical proxy for Φ_t)), we show that firms in a higher-ranked quintile (i.e., those having a higher level of implied volatility prior to a quarter) have a smaller increase in implied volatility during the quarter. This finding supports the model prediction that firms with a higher level of cumulative fraud are more cautious about continuing fraud (because they anticipate closer scrutiny from the regulator) while firms with a lower level of cumulative fraud are more aggressive at committing fraud (because they can hide under the radar). The results are presented in Table IV and detailed in Appendix II.B.

IV. Conclusion

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

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' fraud decisions 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

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

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, 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.

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.

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

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.

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.

Lind, Jo Thori, and Halvor Mehlum, 2010, With or without u? the appropriate test for a u-shaped relationship, *Oxford Bulletin of Economics and Statistics* 72, 109–118.

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.

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.

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

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

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

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

Table I: Implied Volatility and Fraud Detection
This table reports the ordinary least squares (OLS) regression results estimating the relation between implied volatility and fraud detection likelihood. IV is the quarterly average of the daily implied volatility, measured in the quarter before $DETECT$. $DETECT$ is an indicator that denotes whether a firm discloses a fraud-related 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 is in the top decile of the sample. Detailed variable definitions are in Appendix II.C. Columns (1) and (2) include year-quarter fixed effects, and columns (3) and (4) further include firm fixed effects. Column (1)-(3) include the full sample, and column (4) only includes firms with at least one detected restatement from Audit Analytics. Standard errors are clustered by year-quarter and firm. T-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

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

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

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

Table III: Implied Volatility and Fraud Magnitude
This table reports the OLS regression results estimating the relation between implied volatility and the magnitude of fraud. IV is the quarterly average of the daily implied volatility. $FRAUD$ is the magnitude of fraud-related restatement scaled by the standard deviation of quarterly operating income. A restatement is defined as fraud-related if it meets 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 is in the top decile of the sample. Detailed variable definitions are in Appendix II.C. Columns (1) and (2) include year-quarter fixed effects, and columns (3) and (4) further include firm fixed effects. Column (1)-(3) include the full sample, and column (4) only includes firms with at least one detected restatement from Audit Analytics. Standard errors are clustered by year-quarter and firm. T-statistics are reported in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively, using two-tailed tests.

	(1)	(2)	(3)	(4)
Sample	Full	Full	Full	Detected Firms
Variables	FRAUD	FRAUD	FRAUD	FRAUD
IV	0.105***	0.114***	0.040*	0.076**
	(5.48)	(4.59)	(1.76)	(2.06)
IV^2	-0.066***	-0.071***	-0.027*	-0.053**
	(-5.03)	(-4.45)	(-1.95)	(-2.30)
SIZE		0.002*	0.009***	0.014***
		(1.80)	(2.73)	(2.71)
MB		0.003***	0.005***	0.008***
		(2.76)	(3.39)	(3.53)
LEV		0.002	-0.010	-0.019
		(0.36)	(-0.87)	(-1.05)
ROA		0.019	-0.031	-0.060
		(0.88)	(-1.10)	(-1.26)
REVGWTH		0.004**	0.003**	0.005**
		(2.01)	(2.49)	(2.28)
INCOMESTD		-0.000**	-0.000**	-0.000***
		(-2.46)	(-2.29)	(-2.77)
Observations	147,234	139,681	139,432	84,059
Adjusted R-squared	0.008	0.010	0.345	0.348
Firm Fixed Effects	No	No	Yes	Yes
Year-Qtr Fixed Effects	Yes	Yes	Yes	Yes
Two-way Clustering	Yes	Yes	Yes	Yes
Lind-Mehlum U-shape Test				
Extreme Point	0.792	0.800	0.730	0.718
T-statistics	4.04	3.77	1.70	1.98
P-value	0.000	0.000	0.047	0.025

Table IV: Convergence of Implied Volatility

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

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

Internet Appendix of "Everlasting Fraud"

Appendix I: Proofs

A. Lemma 1

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

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

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:

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

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

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

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

$$\frac{\left(1-\delta\rho\right)cd_{t}^{*}}{1-d_{t}^{*}}\left(1+\frac{m_{t}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{t-1}+\sigma_{\varepsilon}^{2}}\right)=1+\delta\rho\left(1-\delta\rho\right)cd_{t+1}^{*}\frac{m_{t+1}^{*}q\left(1-q\right)}{\rho^{2}\Phi_{t}+\sigma_{\varepsilon}^{2}}. \qquad \text{(A.4)}$$

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

B. Lemma 2

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

$$\begin{aligned}
\Phi_{t}&=var\left(s_{t}|\mathcal{F}_{t-1}\right)-var\left(E^{I}\left[s_{t}|\mathcal{F}_{t}\right]|\mathcal{F}_{t-1}\right) \qquad \text{(A.5)}\\
&=var\left(\rho s_{t-1}+\varepsilon_{t}|\mathcal{F}_{t-1}\right)-var\left(E^{I}\left[s_{t}|r_{t},r_{t-1},...\right]|\mathcal{F}_{t-1}\right)\\
&=\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}-var\left(\frac{\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}}{\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}+m_{t}^{*}q\left(1-q\right)}r_{t}|\mathcal{F}_{t-1}\right)\\
&=\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}-\left[\frac{\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}}{\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}+m_{t}^{*}q\left(1-q\right)}\right]^{2}var\left(r_{t}|\mathcal{F}_{t-1}\right)\\
&=\frac{m_{t}^{*}q\left(1-q\right)\left[\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}\right]}{\rho^{2}var\left(s_{t-1}|\mathcal{F}_{t-1}\right)+\sigma_{\varepsilon}^{2}+m_{t}^{*}q\left(1-q\right)}\\
&=\frac{m_{t}^{*}q\left(1-q\right)\left(\rho^{2}\Phi_{t-1}+\sigma_{\varepsilon}^{2}\right)}{\rho^{2}\Phi_{t-1}+\sigma_{\varepsilon}^{2}+m_{t}^{*}q\left(1-q\right)}.
\end{aligned}$$

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

$$E^I [s_t|\mathcal{F}_t] = E^I [s_t|\mathcal{F}_{t-1}] + \frac{cov (r_t, s_t|\mathcal{F}_{t-1})}{var (r_t|\mathcal{F}_{t-1})} \{r_t - E [r_t|\mathcal{F}_{t-1}]\} \tag{A.6}$$

$$= E^I [s_t|\mathcal{F}_{t-1}] + \frac{\rho^2 var (s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2}{\rho^2 var (s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 + m_t^* q (1-q)} \{r_t - E [r_t|\mathcal{F}_{t-1}]\} ,$$

where

$$var (r_t|\mathcal{F}_{t-1}) = var (s_t|\mathcal{F}_{t-1}) + m_t^* q (1-q) \tag{A.7}$$

$$= \rho^2 var (s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 + m_t^* q (1-q) ,$$

$$cov (r_t, s_t|\mathcal{F}_{t-1}) = var (s_t|\mathcal{F}_{t-1}) \tag{A.8}$$

$$= \rho^2 var (s_{t-1}|\mathcal{F}_{t-1}) + \sigma_\varepsilon^2 .$$

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

C. Lemma 3

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

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

$$= \max_{d_t} - (1-d_t) \Phi_t (d_t^*, \Phi_{t-1}) - \frac{\kappa}{2} (d_t - d_0)^2 + \delta [d_t W_{t+1} (0) + (1-d_t) W_{t+1} (\Phi_t (d_t^*, \Phi_{t-1}))] , \tag{A.9}$$

where $\Phi_t (d_t^*, \Phi_{t-1})$ 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. ■

D. Proposition 1

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

E. 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:

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

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

F. 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:

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

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

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

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:

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

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:

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

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

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

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

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

Solving for m_{1t}^* gives that

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

where

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

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

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

where

$$\begin{aligned}E\left[m_1' d_1'\right] &= (1-d_2^*)\,(1-d_3^*)\, m_1^*\left(\Phi_1',\Phi_2',\Phi_3'\right) d_1^*\left(\Phi_1',\Phi_2',\Phi_3'\right)\\ &+ d_2^*\,(1-d_3^*)\, m_1^*\left(\Phi_1',0,\Phi_3'\right) d_1^*\left(\Phi_1',0,\Phi_3'\right)\\ &+ (1-d_2^*)\, d_3^* m_1^*\left(\Phi_1',\Phi_2',0\right) d_1^*\left(\Phi_1',\Phi_2',0\right)\\ &+ d_2^* d_3^* m_1^*\left(\Phi_1',0,0\right) d_1^*\left(\Phi_1',0,0\right), \end{aligned} \quad \text{(A.20)}$$

$$
\begin{aligned}
E\left[m_2'd_2'\right] = &\left(1-d_1^*\right)\left(1-d_3^*\right)m_2^*\left(\Phi_1',\Phi_2',\Phi_3'\right)d_2^*\left(\Phi_1',\Phi_2',\Phi_3'\right) \\
&+ d_1^*\left(1-d_3^*\right)m_2^*\left(0,\Phi_2',\Phi_3'\right)d_2^*\left(0,\Phi_2',\Phi_3'\right) \\
&+ \left(1-d_1^*\right)d_3^* m_2^*\left(\Phi_1',\Phi_2',0\right)d_2^*\left(\Phi_1',\Phi_2',0\right) \\
&+ d_1^* d_3^* m_2^*\left(0,\Phi_2',0\right)d_2^*\left(0,\Phi_2',0\right), \qquad \text{(A.21)}
\end{aligned}
$$

$$
\begin{aligned}
E\left[m_3'd_3'\right] = &\left(1-d_1^*\right)\left(1-d_2^*\right)m_3^*\left(\Phi_1',\Phi_2',\Phi_3'\right)d_3^*\left(\Phi_1',\Phi_2',\Phi_3'\right) \\
&+ d_1^*\left(1-d_2^*\right)m_3^*\left(0,\Phi_2',\Phi_3'\right)d_3^*\left(0,\Phi_2',\Phi_3'\right) \\
&+ \left(1-d_1^*\right)d_2^* m_3^*\left(\Phi_1',0,\Phi_3'\right)d_3^*\left(\Phi_1',0,\Phi_3'\right) \\
&+ d_1^* d_2^* m_3^*\left(0,0,\Phi_3'\right)d_3^*\left(0,0,\Phi_3'\right). \qquad \text{(A.22)}
\end{aligned}
$$

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

$$
\begin{aligned}
W\left(\Phi_1,\Phi_2,\Phi_3\right) = \max_{d_1,d_2,d_3} &-\left(1-d_1\right)\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_1'+\Phi_2'+\Phi_3'-\delta W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
&-\left(1-d_1\right)d_2\left(1-d_3\right)\left[\Phi_1'+\Phi_3'-\delta W\left(\Phi_1',0,\Phi_3'\right)\right] \\
&-\left(1-d_1\right)\left(1-d_2\right)d_3\left[\Phi_1'+\Phi_2'-\delta W\left(\Phi_1',\Phi_2',0\right)\right] \\
&-\left(1-d_1\right)d_2 d_3\left[\Phi_1'-\delta W\left(\Phi_1',0,0\right)\right] \\
&-d_1\left(1-d_2\right)\left(1-d_3\right)\left[\Phi_2'+\Phi_3'-\delta W\left(0,\Phi_2',\Phi_3'\right)\right] \\
&-d_1 d_2\left(1-d_3\right)\left[\Phi_3'-\delta W\left(0,0,\Phi_3'\right)\right] \\
&-d_1\left(1-d_2\right)d_3\left[\Phi_2'-\delta W\left(0,\Phi_2',0\right)\right] \\
&+\delta d_1 d_2 d_3 W\left(0,0,0\right)-\frac{\kappa}{2}\left(d_1+d_2+d_3-3d_0\right)^2, \qquad \text{(A.23)}
\end{aligned}
$$

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 F.O.C. yields:

$$
\begin{aligned}
d_1^* = \frac{1}{\kappa}\{\Phi_1' &+ \delta\left(1-d_2^*\right)\left(1-d_3^*\right)\left[W\left(0,\Phi_2',\Phi_3'\right) - W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
&+ \delta\left(1-d_2^*\right)d_3^*\left[W\left(0,\Phi_2',0\right) - W\left(\Phi_1',\Phi_2',0\right)\right] \\
&+ \delta d_2^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right) - W\left(\Phi_1',0,\Phi_3'\right)\right] \\
&+ \delta d_2^* d_3^*\left[W\left(0,0,0\right) - W\left(\Phi_1',0,0\right)\right]\} - \left(d_2^* + d_3^* - 3d_0\right),
\end{aligned} \tag{A.25}
$$

$$
\begin{aligned}
d_2^* = \frac{1}{\kappa}\{\Phi_2' &+ \delta\left(1-d_1^*\right)\left(1-d_3^*\right)\left[W\left(\Phi_1',0,\Phi_3'\right) - W\left(\Phi_1',\Phi_2',\Phi_3'\right)\right] \\
&+ \delta\left(1-d_1^*\right)d_3^*\left[W\left(\Phi_1',0,0\right) - W\left(\Phi_1',\Phi_2',0\right)\right] \\
&+ \delta d_1^*\left(1-d_3^*\right)\left[W\left(0,0,\Phi_3'\right) - W\left(0,\Phi_2',\Phi_3'\right)\right] \\
&+ \delta d_1^* d_3^*\left[W\left(0,0,0\right) - W\left(0,\Phi_2',0\right)\right]\} - \left(d_1^* + d_3^* - 3d_0\right),
\end{aligned} \tag{A.26}
$$

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

■

G. Proposition 4

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

$$
W\left(\{\Phi_i\}_{i\in\{1,2,3\}}\right) = -\sum_{i=1}^{3}\left(1-d_i\right)\Phi_i' - \frac{\kappa}{2}\left(\sum_{i=1}^{3}\left(d_i - d_0\right)\right)^2. \tag{A.28}
$$

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

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

Taking the first-order condition gives that

$$\frac{\partial W}{\partial d_i} = \Phi_i' - \kappa (\sum_{i=1}^{3} (d_i - d_0)). \tag{A.30}$$

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

$$\left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_0}{1 - d_0}\right) < \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_1^*}{1 - d_1^*}\right), \tag{A.31}$$

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 $d_2^* = d_3^* = d_0$ and $d_1^* > d_0$, where d_1^* solves:

$$\Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_1^*}{1 - d_1^*}\right) = \kappa \left(d_1^* - d_0\right). \tag{A.32}$$

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:

$$\kappa \left(d_1^* - d_0\right) = \Phi_1' = \left(\rho^2 \Phi_1 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_1^*}{1 - d_1^*}\right) > \left(\rho^2 \Phi_2 + \sigma_\varepsilon^2\right) \left(1 - \frac{c d_0}{1 - d_0}\right). \tag{A.33}$$

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 $d_2^* = d_3^* = d_0$. This is because, at $d_2 = d_3 = d_0$, the first-order condition for d_2 is always negative, i.e.,

$$\begin{aligned}
\frac{\partial W}{\partial d_2} &= \Phi_2' - \kappa\left(d_1^* - d_0\right) \qquad\qquad \text{(A.35)}\\
&= \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) - \kappa\left(d_1^* - d_0\right)\\
&< \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) - \kappa\left(d_1^* - d_0\right)\\
&= 0.
\end{aligned}$$

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

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

$$\left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) < \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right), \tag{A.36}$$

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 $d_3^* = d_0$, $d_1^* > d_0$ and $d_2^* > d_0$, where the pair of $\{d_1^*, d_2^*\}$ solves:

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) = \kappa\left(D^* - 2d_0\right), \tag{A.37}$$

$$\Phi_2' = \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_2^*}{1-d_2^*}\right) = \kappa\left(D^* - 2d_0\right), \tag{A.38}$$

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:

$$\kappa\left(D^* - 2d_0\right) = \Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) > \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right). \tag{A.39}$$

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

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

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

$$\begin{aligned}\frac{\partial W}{\partial d_3} &= \Phi_3' - \kappa\left(d_1^* + d_2^* - 2d_0\right) \qquad\qquad (A.41)\\ &= \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_0}{1-d_0}\right) - \kappa\left(d_1^* + d_2^* - 2d_0\right)\\ &< \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) - \kappa\left(d_1^* + d_2^* - 2d_0\right)\\ &= 0.\end{aligned}$$

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

$$\Phi_1' = \left(\rho^2\Phi_1 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_1^*}{1-d_1^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right), \tag{A.42}$$

$$\Phi_2' = \left(\rho^2\Phi_2 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_2^*}{1-d_2^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right), \tag{A.43}$$

$$\Phi_3' = \left(\rho^2\Phi_3 + \sigma_\varepsilon^2\right)\left(1 - \frac{cd_3^*}{1-d_3^*}\right) = \kappa\left(d_1^* + d_2^* + d_3^* - 3d_0\right). \tag{A.44}$$

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 II: Empirical Analyses

A. Data and Sample

A.1. Sample Selection

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

A.2. Variable Measurement

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 as of the last restating period) is in the top decile of the 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 $FRAUD$.

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

For controls, we follow prior literature and include four controls previously shown to affect a firm's level of earnings manipulation (e.g., Zang (2012)), namely, the natural logarithm of total assets ($SIZE$), market-to-book (MB), return on assets (ROA), and leverage (LEV). Among the four controls, $SIZE$ and MB also help control for firm growth. This is important because prior studies show that growth affects firms' incentives to manipulate earnings (e.g., Povel et al. (2007); Wang et al. (2010); Strobl (2013); Wang and Winton (2014)). We further include $REVGWTH$, the percentage change of sales from the same quarter of the last year, as an additional control for growth. Firm financials are from the Compustat quarterly files.

B. Results

B.1. Information Uncertainty and the Regulator's Decision

We conduct two analyses to shed light on the role that information plays in the regulator's calculus. First, we plot monthly implied volatility surrounding restatement announcements. As Figure 2 shows, implied volatility drops sharply upon revelation of fraud, suggesting that detection clears fraud and lowers information uncertainty. Second, we link detection likelihood in a given quarter to the level of implied volatility during the prior quarter by estimating the following regression:

$$DETECT_{i,q+1} = \alpha + \beta_1 IV_{i,q} + \beta_c CONTROLS_{i,q-1}. \tag{A.45}$$

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

Table I column (1) reports the regression results of estimating equation (A.45) excluding controls. The coefficient of interest, β_1, is positive and significant at the 1% level. In columns (2) and (3), we reestimate equation (A.45) including controls. The coefficient of interest, β_1, remains positive and significant at the 1% level in column (2) excluding firm fixed effects and in column (3) including firm fixed effects, respectively. This result suggests that fraud detection likelihood is larger when the information uncertainty about a firm is greater possibly because of a higher level of cumulative fraud. A potential concern is that not all firms in the sample are consistently covered by Audit Analytics so measurement error is likely greater for firms with no recorded restatements in the database. To address this concern, in Table I column (4), we focus on a subsample of firms with at least one restatement announcement tracked by Audit Analytics. For each firm, we include the entire time series of quarterly observations during the sample period. Results using this subsample remain similar.

Results from these two analyses point to the role that information plays in the regulator's calculus: as detection helps restore information precision, she benefits more by directing detection efforts towards firms with higher levels of fraud-induced information uncertainty.

B.2. Information Uncertainty and the Manager's Decision

We then conduct two analyses to shed light on the role that information plays in the manager's calculus. We first note that results in Table I indicate that the manager's cost of committing fraud is expected to increase with information uncertainty as detection likelihood is higher. To shed light on how information uncertainty affects the manager's benefit from committing fraud, we link implied volatility to analyst forecast revision, a proxy for changes in the market's expectation about the firm's value, by estimating the following regression:

$$REVISION_{i,q} = \alpha + \beta_1 SUE_{i,q} \times IV_{i,q} + \beta_3 SUE_{i,q} + \beta_4 IV_{i,q} + \beta_c CONTROLS_{i,q-1}, \quad \text{(A.46)}$$

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

Table II column (1) reports the regression results of estimating equation (A.46) excluding controls. The coefficient of interest β_1 on $SUE \times IV$ is positive and significant at the 1%

level. Column (2) repeats the analysis including controls and column (3) further includes firm fixed effects. The coefficient of interest, β_1, remains positive and significant at the 1% level, in both columns. This result suggests that the MB of committing fraud is larger when the information uncertainty about the firm is higher because unexpected earnings elicit more responsive analyst forecast revision.

Together, results in Table I and Table II point to a non-monotonic relation between the amount of fraud committed in a period and the amount of fraud accumulated to date, because a high level of cumulative fraud increases both the MB and MC of further committing fraud. Indeed, when we plot the empirical proxy for m_t^* ($FRAUD$) against the empirical proxy for Φ_{t-1} (IV), with $FRAUD$ being the amount of fraud in the reported net income for a given quarter q and IV being the level of implied volatility measured prior to the earnings announcement of quarter q, we observe an inverse U-shaped relation between the two. The plot is shown in Figure 3.[10]

Based on the univariate plot, we then examine the relation by estimating the following multivariate quadratic regression:

$$FRAUD_{i,q} = \alpha + \beta_1 IV_{i,q} + \beta_2 IV_{i,q}^2 + \beta_c CONTROLS_{i,q-1}. \tag{A.47}$$

Again, the dependent variable, $FRAUD$, measures firm i's fraud amount in its fiscal quarter q. IV is the average daily implied volatility of quarter q prior to the earnings release, and IV^2 is its squared term. As for controls, we continue to include the five basic firm characteristics. Since $FRAUD$ is scaled by the standard deviation of operating income, we also include this scaling factor (labeled $INCOMESTD$) as an additional control to alleviate the concern that any observed relation between IV and $FRAUD$ is entirely driven by a denominator effect. As before, we include year-quarter fixed effects, and cluster standard errors by firm and quarter.

Table III column (1) reports the regression results of estimating equation (A.47). As shown, IV exhibits a positive coefficient and its squared term exhibits a negative coefficient,

[10] For ease of presentation, we sort the number of observations into 100 bins based on the level of IV. Each marker then represents the average level of $FRAUD$ for the observations in a bin. We fit a quadratic curve to the plotted data.

both significant at the 1% level. This result demonstrates an inverse U-shaped relation between the amount of fraud committed in a quarter and the level of information uncertainty induced by past fraud. Column (2) repeats the analysis including controls and column (3) further includes firm fixed effects. The inference that we draw on IV and its squared term remains qualitatively similar, although the statistical significance weakens with the inclusion of firm fixed effects. Column (4) again focuses on a subsample of firms with at least one restatement announcement tracked by Audit Analytics. Results using this subsample become stronger. In the last two rows of the table, we formally test the relation using the U-shape test developed in Lind and Mehlum (2010). The p-values of the Lind-Mehlum test statistics reject the null of no inverse U-shaped in all columns at the 5% level or lower.

B.3. Convergence of Fraud

One important prediction from our multi-firm model is that anti-fraud regulations are unlikely to eradicate fraud but may synchronize firms' fraud decisions. To test this prediction, we sort firm-quarters in the sample into quintiles based on firms' level of implied volatility of prior quarter, and then estimate the following regression:

$$\Delta IV_{i,q\,to\,q+1} = \alpha + \beta_1 IVQ1_{i,q} + \beta_2 IVQ2_{i,q} + \beta_3 IVQ4_{i,q} + \beta_4 IVQ5_{i,q} + \beta_c CONTROLS_{i,q-1}, \tag{A.48}$$

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

Table IV columns (1) and (2) report the regression results of estimating equation (A.48), without and with firm fixed effects. Compared with those in the middle quintile ($IVQ3$=1), firms in a lower-ranked quintile of implied volatility prior to a quarter tend to have a larger increase in implied volatility during the quarter, as evidenced by a positive coefficient estimate on $IVQ2$ and an even larger one on $IVQ1$. Also benchmarked against the middle quintile, firms in a higher-ranked quintile of implied volatility prior to a quarter tend to have a smaller increase in implied volatility during the quarter, as evidenced by a negative coefficient estimate on $IVQ4$ and an even more negative one on $IVQ5$. This finding sheds light on the convergence of corporate fraud across firms over time.

One concern is that this finding merely reflects the mean-reverting nature of IV. To address the concern, we augment equation (A.48) by further including the interaction terms between $IVQn$ (n = 1, 2, 4, and 5) and $WAVE$, an indicator denoting whether a firm-quarter overlaps with a fraud wave in the firm's industry. To define $WAVE$, we first compute $FRAUD\%$, the percentage of firms with restatement announcement in an industry-quarter. We code $WAVE$ as one if the actual $FRAUD\%_{j,q}$ for industry j–quarter q exceeds the 90th percentile of its sample distribution and zero otherwise. The industry classification is based on the Global Industry Classification Standard (GICS) 4-digit industry groups.

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

C. Variable Definitions

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

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

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

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

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

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

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

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

$REVGWTH_{q-1}$: sales revenue of quarter $q-1$ divided by sales revenues of quarter $q-5$ (i.e., one-year lag) minus one, in percentage points.

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

$FRAUD_q$: the absolute magnitude of fraud-related restatement in the misstating quarter (proxied using the cumulative net income impact of the restatement divided by the number of restating quarters), scaled by the standard deviation of quarterly operating income (after restatement, if any) measured over the most recent eight quarters. If a quarter does not have any fraud related restatement, $FRAUD$ is coded as zero. Fraud-related restatements are defined above. If the dollar amount of a fraud-related restatement is missing in Audit Analytics, the observations associated with that restatement are removed from the analyses involving $FRAUD$.

$INCOMESTD_q$: the standard deviation of quarterly operating income (after restatement, if any) measured over the most recent eight quarters (from $q-7$ to q).

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

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

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

ECGI Working Paper Series in Law

https://ecgi.global/content/working-papers

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