Risk Management Failures*

Matthieu Bouvard[†] Samuel Lee[‡]

November 2015

Abstract

We present a model in which financial firms are engaged in preemptive competition for trading opportunities and the cost of risk management increases with time pressure in financial markets. Because time pressure is in turn endogenous to risk management choices, strategic complementarities can trigger a race to the bottom: firms' decisions to abandon risk management, while individually rational, are collectively inefficient and cause a misallocation of risks among financial intermediaries. Externalities operate through opportunity costs and agency costs, and provide a rationale for regulation that views risk management both as a coordination problem (among firms) and as a governance problem (inside firms). *JEL Classification: G20; G32; G01.*

*We thank Bruno Biais, Adolfo de Motta, Nadya Malenko, Albert Menkveld, Martin Oehmke, Adriano Rampini, Hersh Shefrin, Basil Williams and Liyan Yang for helpful discussions. We also thank participants to seminars at the Tinbergen Institute, Université Laval, McGill University, Toulouse School of Economics, Santa Clara University, University of Calgary, and to the 24th Mitsui Finance Symposium, the 2015 China International Conference in Finance, the 2015 OxFIT conference, the 2015 Wharton Conference on Liquidity and Financial Crises and the 2015 Washington University Corporate Finance Conference. Matthieu Bouvard acknowledges financial support from FRQSC (Young Researcher Grant #175728).

†McGill University, Desautels Faculty of Management. E-mail: matthieu.bouvard@mcgill.ca.

‡Santa Clara University, Leavey School of Business, Swedish House of Finance, and European Corporate Governance Institute. E-mail: sslee@scu.edu.

> There are three ways to make a living in this business: Be first, be smarter, or cheat. Well I don't cheat, and even though I like to think we have got some pretty smart people in this building, of the two remaining options it sure is a hell of a lot easier to just be first.
>
> (*Margin Call*, J.C. Chandor, 2011)

1 Introduction

The standard risk management problem in finance describes a single firm trying to optimize a portfolio of risk exposures. In that context, risk management deficiencies in financial institutions have been linked to two underlying problems: The "risk shifting" problem that leverage, implicit or explicit government guarantees, or other channels of externalizing losses induce excessive risk taking, and the "black swan" problem that firms miscalibrate their risk exposures because they put too much trust in models that do not adequately account for certain types of risks. In this paper, we show that a third and distinct source of inefficiency emerges when risk management is envisioned not only as an individual optimization problem, but as an interrelated set of decisions made by financial institutions that compete in the same markets: the value of risk management to a financial firm depends on the dynamics of capital markets, which, in turn, is affected by financial firms' choices of risk management. Through this loop arises the possibility that risk management deficiencies constitute the outcome of a *coordination* failure.

The premise of our theory is that risk management introduces frictions in the operations of a financial institution that affect its ability to compete in capital markets. This view emerges from two vantage points. First, risk management is an organization-wide effort to gather, analyze and transmit information. In the words of practitioners, risk management

> reduces to the basics of getting the right information, at the right time, to the right people, such that those people can make the most informed judgments possible... and that critical information flowing into and out of risk monitoring processes can be distilled and compiled in a coherent and timely manner and made available, not only to the risk managers, but to key business leaders across the institution and to top management.[1]

[1] This definition is from a report entitled "Containing Systemic Risk: The Road to Reform" (CRMPG III, 2008, 70) submitted to the US Treasury and the Financial Stability Board by a policy group composed of top executives and chief risk officers of leading global banks.

Indeed, risk management failures commonly uncover communication or information deficiencies inside firms (Stulz, 2008).[2]

Second, the design of risk management processes in financial firms cannot be decoupled from the requirements of a fast-paced competitive environment. In a survey of risk management issues in the financial industry (Ernst & Young, 2013, 14), executives cite the need to strike a "balance between a sales-driven front-office culture and a risk-focused culture" and to "achieve higher-quality and more timely reports" as their two top challenges. More generally, chief risk officers emphasize the challenge of translating the risk appetite statement of a firm into decision-making tools that provide effective guidance to market operatives (8-25).

These observations suggest a view of risk management in financial institutions as an information process, and of deficiencies in this process as neither inevitable nor mistakes but rather the outcome of organizational trade-offs. In other words, the design of risk management processes reflects the costs and benefits of producing, transmitting and using information about risks across the organization. This view raises a simple question which is the point of entry of our analysis: What cost(s) of acquiring risk information could outweigh the benefits of that information even when the financial amounts at stake are substantial?[3] The theory we propose along these lines seems particularly suitable for financial markets. The key assumption is that information "costs" *time*, which delays investment decisions. To firms that face short-lived investment opportunities, delays entail costs that never dwindle in importance. Quite the contrary, these opportunity costs inherently scale up with the size or frequency of investments.

We construct a model of trading under time pressure to trace out the implications of this assumption: Competing firms search for scarce trading opportunities. Before beginning to search, each firm decides whether to activate a risk management system. An active system investigates any located opportunity on its "fit" with the firm's risk profile, which improves trading decisions but takes time. This generates a basic tension between trade execution and risk management.

This tension can affect both firms' decisions to set up risk management protocols and traders' decisions to comply with them. Our analysis is hence divided into two parts. The first part abstracts from agency problems in firms and begins with characterizing the risk management choice

[2] In Stulz's classification of risk management failures, every category but one relates to information or communication: (1) mismeasurement of known risks, (2) failure to take risks into account, (3) failure in communicating risks to top management, (4) failure in monitoring risks, (5) failure in managing risks, and (6) failure to use appropriate metrics.

[3] To put this question into perspective, it is useful to consider financial contexts in which investors explicitly pay for information to reduce uncertainty. Buyers in mergers and acquisitions hire advisors to conduct due diligence, often for several weeks, to detect potential issues before a deal is closed, and in 2012, paid on average .85% of the deal value for this information service (Barba, 2013). For a $100 million acquisition (1.67% of the $6 billion J.P. Morgan Chase lost in the 2012 London Whale scandal), this would amount to $850,000.

of a single firm as a function of time pressure. Time pressure renders all trading strategies less profitable. However, it affects slower ones more severely for two distinct reasons: the information benefit of risk management is contingent on trade execution, and the execution probability of slower strategies is more sensitive to available time. This in turn implies monotonic strategies: a firm discards risk management (only) if time pressure exceeds a unique threshold, in which case prioritizing trade execution over risk management maximizes firm value even though it provokes more "trading debacles."

However, these privately optimal choices can lead to a constrained inefficient equilibrium outcome. The monotonicity of best responses implies that risk management choices are strategic complements: Firms that abandon risk management to accelerate trade execution raise the time pressure on other firms, which then become more prone to do the same. This can support competitive equilibria where firms compel one another to prioritize trade execution against their collective interest – a race to the bottom. The root of this market failure is that every firm, while optimizing its own time use, fails to internalize that its speed contributes to time pressure in the market. Such failures are likelier in "hot" markets where trading is ex ante more attractive and thus the opportunity cost of risk management to each firm larger.

These strategic complementarities translate the firm-level tensions between trade execution and risk management into a market-wide trade-off between trading activity and risk allocation. We use global games to study comparative statics with respect to competition and a speed ratio that relates the speed at which firms locate trading opportunities in the market to the speed at which information is processed inside firms. Improvements in market access or information technology can trigger a feedback loop between market immediacy, or trading volume, and cutbacks in risk management. As a result, equilibrium liquidity and trading activity can be *excessive*, and by virtue of the feedback loop, both cause and consequence of real misallocation.

The second part of our analysis studies how agency problems interact with this time-based tension between risk management and trade execution. For this purpose, we embed a multi-task problem in the model that subjects search and risk management to moral hazard: Traders choose whether to invest effort into search. In addition, they choose to what extent to comply with risk management protocols that delay and possibly interfere with trading (pre-trade controls). Firms design compensation contracts that specify wages contingent on trading profits and periodical reviews of trading activity that imperfectly detect non-compliance (post-trade controls).

The optimal contract can be implemented with bonuses and clawbacks. To incentivize search, firms must pay bonuses tied to trading profits. To incentivize compliance, they must clawback wages if non-compliance is detected. There is, however, an endogenous conflict between these two tasks: While necessary for search, the bonuses create a preference for trading that endogenously inclines traders against complying with risk management. Hence, raising search incentives makes compliance more costly to incentivize at the margin.

The *partial* effect of this multi-task conflict is to create agency costs inside each firm, which, all else equal, further discourages risk management. In addition, delegation produces an *equilibrium* effect. When the probability that trading opportunities vanish while being screened rises, traders are more tempted to circumvent risk management and boost trading profit. The tension between front-office search and risk management compliance thus causes agency rents *to increase with time pressure*. This, in turn, makes it optimal for each firm to incentivize less compliance under higher time pressure. However, when optimizing risk management incentives for their own traders, firms ignore that these individually optimal arrangements, through their effect on time pressure, affect the agency cost of risk management in other firms. That is, delegation gives rise to *contractual externalities*.

This reinforces the monotonicity of firms' best responses and the strategic complementarities of the basic model without agency. With agency, time pressure devalues risk management along two fronts: the "opportunity cost" effect reduces its value to the firm as a whole, while the "agency rent" effect transfers that value increasingly to the firm's traders. Since both effects feed back into time pressure and thus reinforce each other, it is moot to debate in this setting whether risk management failures are to be blamed on the moral hazard of traders or the negligence of firms – the answer is both.[4]

The agency problem is not merely an amplification mechanism. The contractual externalities can create coordination failures among firms *even if the risk management latency goes to zero* (i.e., the opportunity cost effect disappears). Time pressure in the market depends on compliance even without latency because risk management causes firms to trade more *selectively*. Since the cost of incentivizing compliance depends on time pressure, contractual externalities alone can cause incentive structures geared toward trade execution to be a self-enforcing market outcome. We interpret these strategic complementarities that sustain a systematic lack of risk management absent any tangible friction as front-office "cultures."

Further, the agency problem can render the outcome inefficient even if firms could coordinate. The agency rent effect amounts to a redistribution of surplus from the firms' principals to the traders, and in sufficiently fast or competitive markets, can make risk management too costly

[4]This is invariably the main point of contention when rogue traders stand trial. The trader at the heart of the 2008 Société Générale trading scandal, Jerome Kerviel, insisted that his superiors had been intentionally negligent and tacitly supportive. An audit report revealed that the bank had failed to follow through on more than 74 internal alerts about his trading activities dating back to 2006 (Bennhold and Clark, 2008).

for each principal irrespective of the other principals' choices. The different sources and types of inefficiencies, and their interactions, suggest that several policy levers may be needed for effective regulation.

While incentive contracts are instrumental in the implementation of risk management, our analysis cautions against the inference that high-powered contracts – low ratios of base wage to bonus – indicate weak compliance incentives. We show that the equilibrium relationship between *salary-incentive ratios* and risk management quality is ambiguous. Because the severity of the agency conflict is endogenous to overall compliance (via time pressure), traders' compensation can become higher or lower-powered when shocks to the trading environment decrease the equilibrium level of compliance. Therefore one cannot infer from high-powered contracts whether a firm finds it cheap or too expensive to elicit compliance, at least not without additional information.

Risk management failures in our model are rooted not in the capital structures of financial firms but in the markets they operate in. Unlike risk shifting or black swans, they are impervious to capital and liquidity requirements. Some alternative policies have ambiguous effects as they deter valuable as well as excessive trades, but the following two-pronged approach seems reasonable: The first prong views deficient risk management as a governance problem, mandates and supervises standards on risk controls and incentive compensation, and makes *both* firms and individual traders liable for violations to counteract both opportunity costs and agency rents of risk management. The second prong views it as a public goods problem, subsidizes risk management investments, and integrates risk controls into market processes and trading platforms to alleviate the coordination failure.

The rest of the paper is organized as follows. Section 2 discusses evidence consistent with time-related tensions between risk management and trading. Section 3 introduces a baseline model of risk management without agency and derives equilibrium predictions. Section 4 introduces agency and shows that optimal contracting creates a new externality channel. Section 5 discusses the robustness of our results to a number of alternative specifications. Section 6 contains a regulatory discussion. We postpone the literature review to Section 7 and conclude in Section 8. The Appendix contains the proofs of the main results; proofs of auxiliary results and robustness checks are relegated to the Online Appendix.

2 Motivating observations

2.1 Occupational characteristics of financial trading

The first basic premise of our theory is that financial traders (*i*) make decisions with potentially material impact on their firms (*ii*) under significant time pressure (*iii*) in a highly competitive environment. We use the employee survey data of the U.S. Department of Labor's Occupational Information Network (O*Net) to check whether this description is accurate. Most relevant to our analysis are the four "structural job characteristics" (from the category "work context") displayed in Table 1. Each characteristic is measured on a 100-point scale split into five intervals of equal length that corresponds to qualitative assessments. A score in the top interval (80-100) indicates that the respective characteristic is extremely salient in the occupation under consideration.

[Table 1 about here.]

Financial traders are subsumed under the (group of) occupation(s) *Sales Agents, Securities and Commodities*. Table 1 compares the score of this occupation on each of the characteristics against the mean, median, and standard deviation across the 922 occupations that O*Net reports these characteristics for. *Sales Agents, Securities and Commodities* scores far above the mean or median in each of the characteristics.

[Figure 1 about here.]

Furthermore, in the three-dimensional plots of Figure 1, almost no other occupation is located in the vicinity of *Sales Agents, Securities and Commodities* (indicated by the arrow) in the front-upper-right corner. Indeed, only three out of the occupations score in the 80-100 interval on time pressure, competition, and decision impact simultaneously. The average score of *Sales Agents, Securities and Commodities* score across these characteristics is the highest and a statistical outlier.

2.2 Time-related tensions in risk management

The second premise of our theory is that the process of collecting, transmitting, and aggregating information to operationalize risk control takes time and creates a tension between trading and risk management. In a recent survey, the chief risk officers of major financial institutions identify as their top challenges (i) the need for (investment in) information technology and data systems that can deliver "timely transparency" and (ii) the conflict between trading desks ("front office") and risk management (Ernst & Young, 2013, 2-5 & 14). For example,

> Executives interviewed warned that there is a tendency for a sales-driven culture to take a minimum-compliance approach to risk, particularly as revenue pressures grow... Many have added new metrics to better measure risks and concentrations, but the quality and timeliness of reporting is an area that continues to require significant ongoing IT investment. (14)

The gist of (i), expressed at various points in the survey, is that collecting higher-quality information takes more time and that risk management cannot fulfill its role properly unless that time is reduced. A tacit premise in this statement is that timeliness is *paramount* – the converse idea of delaying other operations to match the pace of risk management never features. This priority order suggests that latency is costly, being a possible cause of (ii), and that "sales" have primacy over risk management. Below, we describe a few contexts in which time concerns seem to have relegated risk management to a lesser role.

2.2.1 Value-at-risk analysis and stress tests: comprehensiveness vs. timeliness

Value-at-Risk (VaR) is the main method banks use to evaluate the risk of losses due to changes in market prices (market risk). Two basic choices in VaR design are whether to simulate price scenarios with the *Monte Carlo* or *historical* method, and whether to track losses using *full revaluation* or *sensitivities.*

A McKinsey report (Mehta et al., 2012) highlights the key trade-off inherent in these choices. Most of the surveyed banks use historical simulations although the Monte Carlo method is widely considered to provide a more comprehensive picture of risks in the "tails" of the distribution and allows risk factors and assumed correlations to be modified with some precision.

> But Monte Carlo, which typically requires about 10,000 simulations per risk factor, places a burden of complexity on the bank... the result is often a computational bottleneck that leads to much longer *reaction times* [emphasis added] compared with the easier but less accurate historical simulation. (4)

Similarly, sensitivities reduce trading positions to a few parameters ("Greeks") to approximate the impact of market movements, whereas a full revaluation accounts for the details of positions and specific pricing models. Here, too, banks lean toward the cursory approach:

> Approximations or sensitivities are often deployed; the computing power needed to reprice a typical trading portfolio is so enormous that banks sometimes cannot do it

> in a *timely manner* [emphasis added]... the survey found that average VaR run time ranges between 2 and 15 hours; in stressed environments, it can take much longer. (5)

As the report further mentions, the time frame of risk measurement naturally extends with the thoroughness and comprehensiveness of the tests, making it very difficult to aggregate and verify market risk in real time across an entire institution. This concern is echoed in the survey by Ernst & Young (2013):

> Banks are beginning to question the approaches they are taking to stress testing, with the aim of *speeding up the process* [emphasis added]. 38% of survey respondents indicated it takes a month to complete a group-wide test, 22% said it takes two months, and for 25%, it takes three months or more... slow results are a barrier to using the tests as an effective management tool. (37)

2.2.2 Algorithmic trading: machine speed vs. human speed

Two fundamental reasons that financial markets have become so much faster in recent time are advances in quantitative finance modeling and computer technology. In discussing the automatization of trading and its implications for financial market regulation, Kirilenko and Lo (2013) point to the dangers of the resulting mismatch between "machine speed" and "human speed."

> Automated trading systems provide enormous economies of scale and scope in managing large portfolios, but trading errors can now accumulate losses at the speed of light before they're discovered and corrected by human oversight. Indeed, the enhanced efficiency, precision, and scalability of algorithms may diminish the effectiveness of those risk controls and systems safeguards that rely on experienced human judgment and are applied at human speeds. (60)

They further argue that the primacy of speed may erode *incentives* for risk management.

> In a competitive trading environment, increased speed of order initiation, communication, and execution become a source of profit opportunities for the fastest market participants. Given these profit opportunities, some market participants... may choose to engage in a "race to the bottom," forgoing certain risk controls that may slow down order entry and execution. This vicious cycle can lead to a growing misalignment of incentives as greater profits accrue to the fastest market participants with less-comprehensive safeguards. (61)

Policy notes by the Federal Reserve Bank of Chicago voice similar concerns about the propensity of high-speed trading firms – who are under immense time pressure to capture desired prices – to dispense with safeguards that slow down an order (Clark, 2010, 2012):

> Errors have been the result of the removal of pre-trade risk controls to decrease latency. For example, futures broker MF Global suffered \$141.5 million in losses in February 2008, when a rogue trader initiated transactions during off hours... One breakdown in MF Global's internal risk systems was the removal of trade limits, which had been done to increase trading speeds.[5] (2010, 3)

3 Risk management in equilibrium

3.1 Baseline model

A mass M of risk-neutral firms (traders), indexed by k, competes for trading opportunities. Time is continuous, and a generic trading opportunity takes the form of a mispricing $\pi > 0$ that appears in the market at $t = 0$. For example, demand shocks to agents in segmented markets can create price discrepancies between assets with correlated cash flows, as in Shleifer and Vishny (1997) or Gromb and Vayanos (2002).

A friction interferes with the ability of traders to instantly take advantage of this mispricing. This friction can take two forms. First, discovery can take time: there can be delay after the mispricing appears until the time at which a trader becomes aware of it. Second, execution can take time: there can be a delay between a trader's decision to trade and its physical execution. Since these two frictions are equivalent in our framework, we model only the first one.[6] Specifically, traders discover the trading opportunity at random times that are identically and independently distributed according to an exponential distribution:

$$\tilde{t}_k \sim \text{Exp}(\lambda^{-1}).$$

The traders' discount factors are normalized to 1.

Upon locating an opportunity, a trader can request a trade (of one unit). A trade pays the sum of the mispricing π, which is a common value across traders, and a private value α_k.

[5] Kirilenko and Lo (2013) provide similar anecdotes.

[6] More precisely, a specification of our framework in which all traders locate the trading opportunity immediately and execution introduces latency is isomorphic to one in which locating the opportunity introduces latency and execution is instantaneous.

We interpret α_k as the "fit" between the trade and the risk profile of that particular trader, desk, or firm. For example, the firm may prefer trades that hedge rather than amplify existing exposures.[7] α_k can also reflect the shadow cost of mobilizing collateral to guarantee positions (e.g., Rampini and Viswanathan, 2010), or frictions that justify risk management because they amplify the impact of cash flow shocks on a firm, such as bankruptcy costs or financial constraints (e.g., Froot, Scharfstein and Stein, 1993).

There is uncertainty about the private values. At the time of discovery, k merely knows that

$$\tilde{\alpha}_k = \begin{cases} \alpha_+ & \text{with probability } \rho \\ \alpha_- & \text{with probability } 1-\rho. \end{cases}$$

The private values $\{\alpha_k\}_{k\in[0,M]}$ have a mean of zero and are independent across traders. We also assume $-\alpha_- > \pi > -\alpha_+$, that is, a trade is profitable if and only if $\alpha_k = \alpha_+$. We call "risk management" the process of producing information on α_k. Specifically, before $t=0$, each firm simultaneously decides whether to activate a risk management technology. The technology investigates any requested trade and executes it only if $\alpha_k = \alpha_+$. However, investigation takes a deterministic time ι, and hence delays execution.

Last, the mispricing is sensitive to trading pressure: it disappears once the mass of trades exploiting this opportunity reaches I. For example, I can be interpreted as the net order flow that eliminates the difference between local demands across segmented markets, as in Kondor (2009).[8] The finite size of the trading opportunity creates preemptive competition among the traders, the intensity of which is captured by the ratio

$$i \equiv \frac{I}{M}.$$

The smaller this ratio is, the more intense is the competition. To focus on the case where the finite size of the trading opportunity always generates concern about preemption, we assume that I is strictly smaller than the mass of traders ρM for whom $\alpha_k = \alpha_+$, that is, $i < \rho$. This assumption simplifies the exposition for now but is not essential. We will later endogenize M by introducing an entry stage.[9]

[7] One of the most important operational risk management metrics are so-called concentration limits, the role of which is to ensure that the bank is not exposed too heavily to one particular idiosyncratic risk (Ernst & Young, 2013, 20).

[8] In Section 5, we discuss cases in which the mispricing decreases smoothly as the mass of trades increases, or is directly dependant on the time t since it appeared in the market.

[9] $i < \rho$ ensures that the trading opportunity is exhausted in finite time in any equilibrium. The results are qualitatively the same if $\rho \leq i < 1$. For $i > 1$ time pressure disappears.

We conclude the description of this basic framework with three remarks on modeling choices.

First, risk management blocks trades that create undesirable risks. In practice, some risks can be hedged. Allowing firms to hedge risks and then proceed with trades would not change the gist of our analysis insofar as identifying risks and implementing hedges still takes time. As a pre-trade process, it would delay trade execution, and as a post-trade process, it would leave the firm vulnerable to risks in the interim.

Second, activating risk management is an ex ante decision. Our model thus analyzes firms' incentives to set up risk management controls that systematically oversee trading activity, that is, considers risk management as an organizational choice.[10] In Section 4, we allow traders that operate within a firm's chosen risk management framework to manipulate the controls. In Section 5, we further discuss a specification in which firms make risk management choices "on the fly" as trading opportunities are discovered.

Finally, assuming a continuum of firms and i.i.d. random variables makes the model highly tractable: Every firm knows how many in total locate the opportunity over time and how many of those firms will have positive private values. In fact, the only aggregate uncertainty is strategic: To infer how many trades are *executed* over time, firms must form beliefs about everyone else's risk management choices. We later add exogenous aggregate uncertainty to sharpen the predictions of the model (Section 3.4).

3.2 Privately optimal risk management

Consider a trader who believes that trading opportunities stay alive for a period of length T, which we take as exogenous for the moment. Without risk management, his expected profit is π conditional on locating the opportunity before T, which happens with probability

$$p_h(T) \equiv 1 - e^{-T/\lambda}.$$

We will refer to this strategy as "hasty." This strategy is obviously an irrelevant alternative for $\pi < 0$. It is only when trading without risk management is (on average) profitable that risk management has an opportunity cost.

If the trader activates risk management, his expected profit is $\rho(\pi + \alpha_+)$ conditional on

[10]The Basel Committee uses the term risk management *framework* to describe systematic risk controls in firms.

locating the opportunity and identifying α_k before T, which happens with probability

$$p_d(T) \equiv \begin{cases} 0 & \text{if } T < \iota, \\ 1 - e^{-(T-\iota)/\lambda} & \text{otherwise.} \end{cases}$$

We will refer to this strategy as "deliberate." We also refer to either strategy as "implemented" once the trader has the possibility to execute the trade. Without risk management implementation amounts to locating the opportunity. With risk management it further requires identifying α_k, and does not entail execution if $\alpha_k = \alpha_-$.

The difference between the unconditional expected profits of the two strategies as a function of T is

$$\begin{aligned} \Delta(T) &\equiv p_d(T)\rho(\pi + \alpha_+) - p_h(T)\pi \\ &= p_h(T)(1-\rho)|\pi + \alpha_-| - [p_h(T) - p_d(T)]\rho|\pi + \alpha_+|. \end{aligned} \tag{1}$$

We refer to $\Delta(.)$ as the private net value of risk management. On the bottom line, the first term reflects the benefit of risk management: avoiding bad trades that would occur with probability $p_h(T)(1-\rho)$ under the hasty strategy. This benefit depends on the implementation probability $p_h(.)$, which means that the value of risk management is *contingent* on the option of executing the trade. The second term reflects the cost of risk management: failing to capture good trades with probability $[p_h(T) - p_d(T)]\rho$ that would be executed under the hasty strategy. This opportunity cost depends on the *difference* in the implementations probabilities, $p_h(T) - p_d(T)$, and hence the relative speeds of the two strategies.

Since both implementation probabilities $p_h(.)$ and $p_d(.)$ are increasing functions, time pressure has a negative effect on both a hasty and a deliberate trader. The slope of $\Delta(.)$ is thus a priori ambiguous. For $T < \iota$, time pressure affects only the hasty strategy: the implementation probability is zero under the deliberate strategy, while it strictly increases with T under the hasty one. For $T > \iota$, the deliberate strategy is more sensitive to time pressure than the hasty one for reasons related to both the benefit and the cost of risk management:

(i) *Value of information.* Conditional on implementation, the deliberate strategy pays off more than the hasty one. Hence, even if the implementation probabilities were to decrease equally, raising time pressure would lower the unconditional expected profit on the margin more under the deliberate strategy. The difference in conditional profits, $\rho(\pi + \alpha_+) - \pi =$

$(1-\rho)|\pi+\alpha_-|>0$, is precisely the conditional benefit of risk management – avoiding bad trades – the value of which shrinks as the probability of trade decreases with time pressure, reflecting that the information value of risk management is contingent on the ability to execute trades.

(ii) *Value of time.* The implementation probability of the deliberate strategy is more sensitive to changes in the deadline that that of the hasty strategy: $p'_d(T) > p'_h(T)$. In order to implement her strategy, a hasty trader must find the opportunity before T, whereas a deliberate trader must find it before $T-\iota$. A trader who can search from 0 to T gains less from a marginal increase in search time than one who can search only from 0 to $T-\iota$. Intuitively, additional time matters more to those who have less to begin with. The difference $p_h(.)-p_d(.)$ and so the opportunity cost of risk management hence increase with time pressure.

[Figure 2 about here.]

From the above, it follows that $\Delta(.)$ is U-shaped and reaches a minimum at $T=\iota$ (see Figure 2).[11] Note also that $\Delta(.)$ goes to 0 as $T\to 0$ reflecting that implementation probabilities vanish under both strategies as time pressure intensifies. Conversely, when $T\to\infty$, $\Delta(T)$ converges to $(1-\rho)|\pi+\alpha_-|>0$ as implementation probabilities tend to 1, so that the opportunity cost of risk management vanishes. In sum, these observations imply the existence of a *unique* point $T^*>0$ at which a trader is indifferent between the two strategies: $\Delta(T^*)=0$. When time pressure is high, $T<T^*$, it is optimal for a trader to abandon risk management and prioritize execution, i.e., $\Delta(T)<0$. Conversely, when time pressure is low, $T>T^*$, the benefit of informed decision-making under risk management overcomes the loss in execution speed, i.e., $\Delta(T)>0$.

Lemma 1. *The private value of risk management $\Delta(T)$ is strictly decreasing for $T<\iota$ and strictly increasing for $T>\iota$. Furthermore, there exists a unique threshold T^* such that firms choose to activate risk management if and only if $T>T^*$, and T^* is an increasing function of π.*

This contrasts with explanations of risk management deficiencies based on risk-shifting arguments. The lack of risk management in our model is not driven by differences in the financial claims held by various stakeholders (e.g., managers, shareholders, or creditors), nor does it pit

[11] The shape of $\Delta(.)$ does not depend on investigation time ι being deterministic. $\Delta(.)$ is similarly U-shaped when investigation time is exponentially distributed, as search time is. The non-monotonicity of $\Delta(.)$ will later complicate the global games refinement (Section 3.4).

their interests against each other. On the contrary, a firm discards risk management in order to stay competitive under time pressure and to thereby *maximize firm value* for all stakeholders.

Lemma 1 also establishes that T^* increases in π, that is, risk management is less attractive when trading is ex ante more attractive. This is also apparent in (1): When π increases, the losses from bad trades, $|\pi+\alpha_-|$, become smaller, while the forgone profits from trading at a lower pace, $[p_h(T) - p_d(T)]\rho(\pi + \alpha_+)$, become larger. Thus, the benefit of risk management decreases, while its opportunity cost increases.

By Lemma 1, a firm's best response to time pressure is given by

$$\Pr(h) = q(T) \equiv \begin{cases} 1 & \text{if } T < T^*(\pi) \\ [0,1] & \text{if } T = T^*(\pi) \\ 0 & \text{otherwise.} \end{cases}$$

Given a continuum of ex ante identical traders, q also denotes the fraction of hasty traders.

Notice that $q(T)$ is monotonic, which will be key to generating strategic complementarity. Also, T^* is independent of scale in the sense that multiplying the payoff $\pi + \tilde{\alpha}_k$ does not change the sign of $\Delta(.)$ – that is, the costs (and benefits) of risk management scale up with the size of the trading opportunity. Last, any and all interaction between firms goes through T but every firm, being infinitesimal, takes T as given.

3.3 Collectively inefficient risk management

In equilibrium, the deadline T is endogenously determined. Let q denote the fraction of traders that choose the hasty strategy. The deadline T by which the trading opportunity is exhausted satisfies

$$qp_h(T) + (1-q)\rho p_d(T) = i. \tag{2}$$

Let $T(q)$ denote the solution to (2). As is intuitive, $T(q)$ is decreasing: The deadline becomes tighter when the proportion of hasty traders is higher. Note that there are two reasons for this monotonicity. First, hasty traders have a faster implementation because they avoid the risk management delay. Second, conditional on locating the opportunity, hasty traders execute with probability 1, rather than only when $\alpha = \alpha_+$, that is, with probability ρ. As a result, the trading opportunity is depleted at a faster rate.

We let

$$T_h \equiv T(1) \qquad \text{and} \qquad T_d \equiv T(0)$$

denote, respectively, the shortest possible deadline when all traders are hasty, and the longest possible deadline when all traders are deliberate. T_h and T_d bound the range of deadlines that can arise in the market. They can be used to identify parameter regions where strategic dominance arises. We know from Lemma 1 that the threshold T^* below which the private value of risk management is negative increases with π. For high enough π, there may be an equilibrium in which all traders are hasty (hereafter, hasty equilibrium). Indeed, let $\overline{\pi}$ be defined by $T^*(\overline{\pi}) = T_d$. If $\pi > \overline{\pi}$, then $T_d < T^*(\pi)$: Even if everyone else were to be deliberate, T_d would be lower than the threshold T^* below which the hasty strategy is the best response, making it strictly dominant for any trader to be hasty. An analogous argument applies to low π. If $\pi < \underline{\pi}$, where $T^*(\underline{\pi}) = T_h$ if $T_h > \iota$ or else $\underline{\pi} = 0$, being deliberate is strictly dominant for all traders (hereafter, deliberate equilibrium).

In the intermediate region $(\underline{\pi}, \overline{\pi})$, the equilibrium prediction is ambiguous: $q(T)$ and $T(q)$ intersect at three (fixed) points. Suppose a trader believes everyone else is hasty. The implied deadline T_h is then smaller than $T^*(\pi)$ so that his best response is also to be hasty. Thus, a hasty equilibrium exists. If the trader instead believes that everyone else is deliberate, the implied deadline T_d is larger than $T^*(\pi)$ and supports a deliberate equilibrium. In the third equilibrium, the fraction of hasty traders q is exactly such that the implied deadline renders everyone indifferent between the two strategies, i.e., $T(q) = T^*(\pi)$ or $\Delta(T(q)) = 0$. The different equilibrium configurations are illustrated in Figure 3.

Proposition 1. *There exists a unique interval* $[\underline{\pi}, \overline{\pi}] \neq \emptyset$ *such that the hasty equilibrium exists for all* $\pi \geq \underline{\pi}$*, the deliberate equilibrium exists for all* $\pi \leq \overline{\pi}$*, and one equilibrium in which both strategies are used exists for all* $\pi \in (\underline{\pi}, \overline{\pi})$*.*

[Figure 3 about here.]

The source of equilibrium multiplicity is the positive feedback loop created by $T(q)$ and $q(T)$ being both decreasing, which makes risk management choices strategic complements: When more traders are deliberate, the trading opportunity is depleted at a slower pace as risk management delays execution and blocks executions whenever $\alpha_k = \alpha_-$. This lowers the time pressure on others, making them more inclined to be deliberate as well. Conversely, more hastiness raises time pressure, which begets more hastiness.[12]

[12]Compared with bank runs or currency attacks, the strategic complementarities in our model operate through

The strategic complementarity also makes the interior equilibrium unstable: Any shock that causes $T(q)$ to deviate from T^* breaks the indifference condition and pushes all traders to the same strategy. We will later comment on extensions of the model that may accommodate stable interior equilibria (Section 4).

Corollary 1. *Only corner equilibria are stable.*

In every equilibrium, payoffs are symmetric across traders. Since aggregate payoffs decrease in the mass of executed trades with $\alpha_k = \alpha_-$, we obtain a simple Pareto ranking:

Corollary 2. *Every trader being deliberate strictly Pareto-dominates any non-deliberate equilibrium.*

Any non-deliberate equilibrium is thus a coordination failure – a "race to the bottom." The reason that traders can be trapped in the hasty equilibrium is that their individual objective functions fail to internalize the social value of risk management. Consider trader k's net gain from deviating to the deliberate strategy when everyone else is hasty:

$$\Delta(T_h) = p_d(T_h)(1-\rho)|\pi + \alpha_-| - [p_h(T_h) - p_d(T_h)]\pi. \tag{3}$$

The second term captures the preemption motive: By switching to the deliberate strategy, the trader becomes less likely to capture the common value π. This private loss to k is not a social one, since another trader will capture the common value in lieu of k. The first term corresponds to the private benefit of risk management: Under the deliberate strategy, k can avoid a loss of $|\pi + \alpha_-|$. But the social benefit of risk management in this event further depends on the private value of the (hasty) trader k' who then executes in lieu of k: The social gain is zero if $\alpha_{k'} = \alpha_-$ but otherwise $\alpha_+ - \alpha_-$, which happens with probability ρ. Since $\rho(\alpha_+ - \alpha_-) > |\pi + \alpha_-|$, trader k does not fully internalize the allocative efficiency gain of risk management.

3.4 Market structure and technology

While equilibrium multiplicity highlights strategic complementarity, it generates ambiguity about the influence of structural parameters. Hence, a preliminary to studying the effect of changes in market competition or technology is to sharpen the predictions of Proposition 1.

time pressure and learning incentives as opposed to direct payoff externalities: by choosing to be hasty, each trader increases the opportunity cost of acquiring information for every other trader.

3.4.1 Global games treatment

The indeterminacy in games of strategic complementarities can usually be resolved with global games (Carlsson and van Damme, 1993; Morris and Shin, 1998). Standard techniques apply in the presence of *global* strategic complementarity (see, e.g., Morris and Shin, 2003), which in our model would require that raising the fraction of deliberate traders always increases the private value of risk management. This requirement is violated: While T is monotonically increasing in the fraction of deliberate traders, $\Delta(.)$ is not monotonic in T. However, our model satisfies a weaker form of strategic complementarity: By Lemma 1, $\Delta(.)$ crosses 0 once, and is monotonic when positive. These properties define *one-sided* strategic complementarity, as in Goldstein and Pauzner (2005). Borrowing from their approach, we show that the equilibrium uniqueness of global games extends to our setting.

We introduce aggregate uncertainty in the model by assuming that the common value π, instead of being a fixed parameter, is a random variable uniformly distributed over $(-\alpha_+, -\alpha_-)$.[13] We also assume that traders have dispersed information about the realization of π. Specifically, before making a decision on risk management, each privately observes a noisy signal, $s_k \equiv \pi + \xi_k$, where $\{\xi_k\}_{k\in[0,1]}$ are uniformly and independently distributed on $[-\varepsilon, +\varepsilon]$. As we shall see, this information structure will keep traders from knowing exactly what others know and thereby from perfectly coordinating on one strategy.

The equilibrium derivation is, due to the absence of global strategic complementarity, more involved but very similar to Goldstein and Pauzner (2005) and hence relegated to the Online Appendix. Equilibrium strategies retain the threshold form that is typical of global games: Every trader is hasty if his signal s_k lies below a unique common threshold $s^\star$, and deliberate otherwise. Since traders hold different beliefs about π, hasty and deliberate traders may coexist. Indeed, a higher realization of π lowers the mass of traders with signals below $s^\star$, and hence the share of deliberate traders. Finally, a well-known property of global games is that equilibrium uniqueness is preserved even in information structures arbitrarily close to π being common knowledge. That is, even when the noise in traders' signals vanishes, $\varepsilon \to 0$, the equilibrium remains unique and takes the following form:

Proposition 2. *When ε tends to 0, all traders follow a hasty strategy if $\pi > \pi^*$ and a deliberate*

[13] As is usual in global games, these bounds guarantee the existence of regions in which being hasty is a strictly dominant strategy ($\pi > \overline{\pi}$), and being deliberate is strictly dominant ($\pi < \underline{\pi}$).

strategy if $\pi < \pi^*$. π^* is strictly positive and satisfies

$$\int_0^1 \Delta[T(q), \pi^*]dq = 0. \tag{4}$$

Proposition 2 is the counterpart of Proposition 1 in the richer environment of global games.[14] To understand equation (4), consider the non-limit case where ε is bounded away from 0. The marginal trader who receives the threshold signal s^* believes that π is uniformly distributed between $[s^* - \varepsilon, s^* + \varepsilon]$. Since every trader plays the threshold strategy s^*, and the errors in traders' signals are independently distributed, every π maps one-to-one into a proportion q of traders who play a hasty strategy. Accordingly, the marginal trader can transform his posterior on π into a posterior on q. This posterior distribution holds for ε arbitrarily small, and in turn, generates a distribution of deadlines $T(q)$, as defined in equation (2). (4) captures the indifference condition of the marginal trader under these deadlines when ε tends to 0, and his signal $s_k = s^*$ becomes arbitrarily close to an underlying realization of the common value, π^*.[15]

3.4.2 Market immediacy and allocative efficiency

Using the unique equilibrium of Proposition 2, we can study the effect of structural changes. Since preemption plays a key role in the model, we begin by endogenizing the intensity of competition. To this end, we assume that, before observing their signals, traders make an entry decision that involves a fixed cost $\chi > 0$. We will maintain this assumption until the end of this section.

Suppose that in equilibrium traders are hasty, that is, $\pi < \pi^*$. Then, each trader's implementation probability is i, and the profit conditional on implementation is π. Conversely, if traders are deliberate, $\pi \geq \pi^*$, the probability of implementation is the ratio of the investment opportunity size I to the mass of traders with a high private value, ρM, and the profit conditional on implementation is $\rho(\pi + \alpha_+)$. Overall the expected profit from entering is

$$\Pr[\pi \geq \pi^\star(i)]i\mathbb{E}(\pi|\pi \geq \pi^*(i)) + \Pr[\pi < \pi^*(i)]i\mathbb{E}(\pi + \alpha_+|\pi < \pi^*(i)) = i\{\mathbb{E}(\pi) + \Pr[\pi < \pi^*(i)]\alpha_+\}$$

[14] In equation (4), $\Delta(.,.)$ is defined as in (1), with the addition of a second argument explicitly recognizing the dependence on the common value π.

[15] The result that firms are more likely to compromise on risk management in favorable market conditions (when π is large) is consistent with a survey of chief risk officers (Ernst & Young, 2013, 18). More than half of the respondents cite market conditions as a determinant of the leniency of risk governance. The survey suggests that risk appetite is often reined in during difficult times and expanded when markets pick up.

In equilibrium, the intensity of competition i must satisfy the zero-profit condition

$$i\{\mathbb{E}(\pi) + \Pr[\pi < \pi^*(i)]\alpha_+\} = \chi. \tag{5}$$

Recall that the threshold π^* depends on i through $T(q)$ (see Proposition 2). The effect of a change in the cost of entry on equilibrium strategies depends hence on the direction in which the equilibrium threshold π^* moves when competition intensity i changes. One can show that this effect is monotonic: π^* decreases when competition intensifies, i.e., when i decreases. As explained earlier, the marginal trader, the one receiving the signal $s^* = \pi^*$, forms posterior beliefs about the distribution of deadlines $T(q)$, which spans $[T_h(i), T_d(i)]$ and includes T^*. By (2), when i decreases, $T(q)$ shifts down for all q, that is, the deadline shortens for every realization of q. This in turn shifts more probability mass into the region below T^* where the marginal trader prefers being hasty (i.e., $\Delta[T(q), \pi^*] < 0$), so that the integral in (4) turns negative and π^* must decrease for the indifference condition to remain satisfied.[16]

That $\pi^*(i)$ is increasing implies that the zero-profit condition (5) pins down a unique equilibrium competition intensity i^*, which, in turn, reacts to changes in the cost of entry.

Proposition 3. *A decrease in the cost of entry χ increases the mass of traders, decreases the expected deadline, and makes risk management less likely.*

Lower entry costs mechanically raise market immediacy: The larger the mass of active traders, the more trades occur in any time interval. However, there is also an equilibrium effect: The rise in time pressure lowers the threshold for π above which traders abandon risk management, which further accelerates the market as fewer trades are delayed or blocked. Thus, market immediacy is not only *inversely* related to allocative efficiency but, due to this positive feedback loop, both *cause* and *consequence* of misallocation.

The equilibrium risk management response to increased entry involves another feedback effect: Higher i decrease traders' expected profit not only because each of them is less likely to trade, but also because conditional on trading, the value of the trade is smaller due to the lower quality of risk management. As a result, the mass of entry following a drop in the entry cost χ is moderated by the reduction in risk management benefits, which lowers the aggregate value traders capture in the market.

[16] The fact that $\Delta(\pi^*, .)$ is negative below T^* and positive otherwise drives the monotonicity of $\pi^*(i)$. However, the proof cannot rely solely on this observation because $\Delta(\pi^*, .)$ is non-monotonic below T^* (see the proof of Proposition 3 in the Online Appendix).

3.4.3 Internal and external speed

We next study the effect of "technological" shocks to the speed at which information is processed. Since both search and risk management may be affected by such changes, we consider an external-internal speed ratio, $\frac{\iota}{\lambda}$, that measures how fast search is *relative* to risk management.

Proposition 4. *An increase in the external-internal speed ratio $\frac{\iota}{\lambda}$ makes risk management less likely and reduces the mass of active traders.*

To understand why *only* relative speed matters, consider how an increase in search speed impacts the marginal trader: On one hand, it raises the probability that he discovers the trading opportunity before any given deadline $T(q)$; on the other hand, all other traders also locate the opportunity faster, which shortens the deadlines $\{T(q)\}_{q\in[0,1]}$. These two effects offset each other such that, under hasty strategies, no trader gains or loses any advantage. That is, higher search speed does not advantage search *per se*. It does, however, increase the opportunity cost of risk management: A trading opportunity becomes more likely to vanish between time $\tilde{t}_k$ at which a trader discovers it and time $\tilde{t}_k + \iota$ at which he can trade on it under the deliberate strategy. As the implementation probability of the deliberate strategy decreases relative to that of the hasty one, so must π^* for indifference condition (4) to hold. Key to our results is thus the latency risk management imposes relative to the time it takes traders to discover opportunities.

In addition, by making the hasty strategy more prevalent, a relative increase in search speed lowers a trader's expected (average) profit in the market so that, by zero-profit condition (5), the mass of active traders shrinks. This produces a countervailing effect – since less entry favors the deliberate equilibrium – though of second-order importance. Overall, changes that favor external speed lead to a more concentrated market in which traders exercise less risk management.

In practice firms can invest in speed. Such investments inherit the strategic complementarities identified in our model. In particular, it is easy to conceive of self-fulfilling equilibria in which each firm invests only in external speed and chooses the hasty strategy because its expects all other firms to do so. Moreover, the priority that trading has over risk management in our model spills into the investment incentives in the sense that investment in internal speed is dispensable in the hasty equilibrium whereas (some) investment in external speed will always occur irrespective of the equilibrium.

Finally, note that as the external-internal speed ratio $\frac{\iota}{\lambda}$ goes to zero, the opportunity cost of risk management vanishes, and with it, externalities across firms. In that case, risk management becomes a dominant strategy. We show in the next section that introducing an agency friction within firms creates another externality channel that magnifies the impact of the external-internal speed ratio, and creates the possibility of coordination failures even when the risk management latency disappears.

4 Delegated trading and risk management

We have so far ignored agency conflicts assuming traders act in the best interest of their firms. However, many risk management failures are rooted in misaligned incentives within organizations, as much as deficiencies of information systems.[17] For instance, in a joint document, the U.S. Treasury, the Federal Reserve Board (Fed) and the Federal Deposit Insurance Corporation (FDIC) contend that compensation arrangements in financial organizations may prompt agents to disregard risk controls, and recommend an integrated approach to risk management and pay.[18]

In this section, firms must provide traders with incentives to identify trading opportunities *and* to comply with risk management protocols. The agency cost created by this multitasking problem *within firms* generates strategic complementarities *across firms*: Time pressure exacerbates traders' moral hazard problem, making it more expensive for firms to write contracts that incentivize compliance. Thus, delegation expands the scope for coordination failure through a *contractual externality.*

4.1 Stationary model of delegated trading and risk management

A preliminary to studying delegation is to introduce stationarity in the model. In the original setup, calendar time provides information on the probability that a trading opportunity can be found or executed. While this feature is innocuous when search is exogenous and the risk management protocol is decided ex ante, it generates economically unappealing artefacts when an agent within the firm can continuously affect search intensity and risk management quality.[19] To allow for stationarity, we consider overlapping sequences of trading rather than a single round: Trading opportunities now appear in the market at a continuous rate, normalized to one. (I.e., in an interval of time dt, the market gives birth to a mass dt of new trading opportunities.) As in the original model, trading opportunities have a size I, and if trader k locates one of them

[17] While rogue trading scandals present the most visible aspect of this problem, the broader concept of "operational risk" in trading-related activities includes moral hazards such as unauthorized leverage and price manipulation. It is one of the areas banks rate themselves least effective: Only 45% of the surveyed financial institutions rate their operational risk management as "effective," compared to 85% for asset-liability risk, 85% for liquidity risk, 83% for credit risk, and 72% for market risk (Deloitte, 2013, 24).

[18] "Guidance on Sound Incentive Compensation Policies," Federal Register, Vol. 75, No. 122, Friday, June 25, 2010, Notices.

[19] For instance, incentives to search dwindle as time passes.

at time t, the value of executing the trade is the sum of a common value π and a private value α_k^t that is independently drawn from $\{\alpha_-, \alpha_+\}$.[20] Finally, we let $e_t \in [0, \lambda^{-1}]$ denote the search intensity that a trader applies between t and $t + dt$: If a trading opportunity is still alive in the market at t, the probability that the trader locates it between t and $t + dt$ is $e_t dt$.

We focus on *steady states* of the model, that is, equilibria such that the lifetime of a trading opportunity T is constant over time. In such an equilibrium, this extended model easily compares to the original one. To see this, suppose the search intensity is, as earlier, exogenously set to λ^{-1}. Then, T increases with the fraction of deliberate traders, as in the original model. Further, an opportunity found at t may have been born at any point between $t - T$ and t, so that the probability of being able to trade on it subject to running it through risk management is

$$p \equiv T^{-1} \left[\int_{t-T}^{t-T+\iota} 0 d\tau + \int_{t-T+\iota}^{t} 1 d\tau \right] = 1 - \frac{\iota}{T}. \tag{6}$$

Hence, risk management is optimal if T is above a threshold given by

$$p\rho(\pi + \alpha_+) > \pi \Leftrightarrow T > T^* \equiv \iota \frac{\pi + \alpha_+}{\alpha_+ - (1 - \rho)\pi}. \tag{7}$$

It follows that this extended model delivers the same equilibrium predictions as the original one, and in particular, Proposition 1 readily applies.[21]

We now introduce delegation: each firm is indexed by k and to realize trading profits ($\pi + \alpha_k$ per trade) needs to hire one trader. To search at intensity e_t between t and $t + dt$, the trader must bear a personal cost $\kappa e_t dt$. In addition, traders can interfere with risk management: Each chooses the probability $m_t \in [0, 1]$ with which a trade discovered at t is executed *without* risk management. With probability $1 - m_t$, risk management takes place, which delays execution by ι but blocks trades with a low private value $\alpha_k^t = \alpha_-$. While the model basically captures a delegation of risk management to the trader, it can also describe an organizational design where a distinct entity ("back-office") is in charge of pre-trade controls, but can be circumvented by the trader. In that spirit, we will refer to m_t as the level of "manipulation." In sum, delegation has an initiative dimension (search) and compliance dimension (risk management).

We close this description of the model by specifying the set of admissible contracts between a firm and a trader. First, while firms are infinitely lived, traders have finite tenures of length

[20] We index the private values by the trader identity k and the discovery time t, even though it could also depend directly on the trading opportunity. Since trading opportunities are found sequentially an all α_k^t are independently distributed, indexing by trading opportunity would be redundant.

[21] Note that functional forms differ (e.g., T^*).

D: a trader hired at t leaves the firm at $t + D$, and is replaced at no cost by an identical trader. Given a zero discount rate, this simply rules out that compensation is backloaded indefinitely. This assumption is merely technical. As we will show, in our setup, the tenure's length does not affect equilibrium outcomes, so that D can be arbitrarily long or short, or differ across firms and traders.

Every firm makes a take-it-or-leave-it contract offer to a trader. Neither search nor compliance is contractible, which generates the agency problem. However, a firm can write contracts on the instantaneous *common* value of trades that were initiated by its trader at any point in time. In a steady state equilibrium, the mass of trading opportunities that are "alive" at any t is

$$\int_{t-T}^{t} d\tau = T. \tag{8}$$

Hence, a trader who searches at intensity e_t generates an instantaneous common-value profit

$$\Pi_t = Te_t \left[m_t + (1 - m_t)p\rho\right] \pi \equiv \Pi(e_t, m_t), \tag{9}$$

where p is the probability that a trade subjected to risk management can be executed, as defined in (6). Since π is common knowledge, the profit measure Π_t perfectly reveals trading volume, which is therefore implicitly contractible.[22] Note also that the realization of private values α_k^t is not contractible, an assumption made both for convenience and because defining a contractible measure of the fit between different, time-varying risk exposures may in practice be difficult.[23]

To incentivize compliance, firms can (imperfectly) verify manipulation ex post (i.e., at the end of the trader's tenure). Let m denote the trader's average manipulation throughout his tenure, that is, the fraction of trading opportunities that were executed without risk management relative to the total mass of trading opportunities discovered by the trader:

$$m \equiv \int_{\tau}^{\tau+D} m_t e_t dt \left[\int_{\tau}^{\tau+D} e_t dt\right]^{-1}.$$

The firm can contract on a signal that detects manipulation with probability γm, with $\gamma \in [0, 1]$.[24] We interpret this signal as "post-trade controls" that may find verifiable evidence of misconduct

[22] Assuming that π is random and volume is imperfectly contractible increases the scope for agency problems.

[23] This assumption is not crucial: Even with α_k^t contractible, a trader has incentives to manipulate if trades that would be blocked have some (small) probability of generating α_+.

[24] The fact that the probability is linear in m is not crucial: All of our results hold for increasing $\gamma(.)$ that satisfy $\gamma''(.) \geq -[\gamma'(.)]^2/[1 - \gamma(.)]$, which includes all but not only increasing, weakly convex $\gamma(.)$.

after the fact and allow the firm to claw back wages from the trader.[25]

Last, we assume that the marginal cost of search is not too high:

$$\kappa < \pi i \lambda. \tag{10}$$

This ensures that firms always find it worthwhile to incentivize some search. It also implies that in the absence of delegation, search intensity would always be set to λ^{-1}, as in the original model.

4.2 Optimal contracts for initiative and compliance

We sketch here the derivation of an optimal contract between a firm and a trader, taking time pressure T (or equivalently $p = 1 - \frac{\iota}{T}$) as given. The full analysis is in the Appendix.

In a first step, we show that in a steady-state equilibrium, the dynamic delegation problem can be formulated as a static one. Letting $e \equiv \frac{1}{D}\int_0^D e_t dt$ denote average search effort per unit of time between 0 and D, the average (contractible) profit over the trader's tenure writes

$$\hat{\Pi} \equiv \frac{1}{D}\int_0^D \Pi_t(e_t, m_t)dt = e[m + (1-m)p\rho]\pi = \Pi(e, m). \tag{11}$$

Lemma 2. *There exists an optimal contract such that*

1. *if the trader generates average profit $\hat{\Pi}$, he is paid $\overline{W}(\hat{\Pi})$ if no manipulation is detected, and $\underline{W}(\hat{\Pi})$ if manipulation is detected,*
2. *the trader exerts constant search and manipulation: For every t, $e_t = e$ and $m_t = m$.*

In short, there is no loss for the firm in making payments to the trader contingent on average profit $\hat{\Pi}$ rather than on a specific profit path $\{\Pi_t\}_{t\in[0,D]}$. Indeed, the total value of trading to the firm, gross of wages is

$$\int_0^D e_t T[m_t\pi + (1-m_t)p\rho(\pi + \alpha_+)]dt = eT[m\pi + (1-m)p\rho(\pi + \alpha_+)]D. \tag{12}$$

So, keeping payments to the trader fixed, only average search and manipulation matter to the firm. In addition, imposing a specific profit path does not constrain the trader's choice of *average*

[25] The Committee of European Banking Supervisors (2010) stresses *deferred pay* and *post-trade audits* as instruments that promote good decisions without delaying the execution of trades. In practice, valuation control groups examine ex post the assumptions used by traders for complex trading strategies or in illiquid markets. Valuation control is often imperfect. In its settlement with the SEC over the London Whale trading scandal, J.P. Morgan Chase acknowledged that its valuation control unit in London was unable to properly oversee trading activities (Trindle, 2013).

search and manipulation more than specifying average profit only, and therefore has no additional incentive effect.

Thus, the firm's problem is to choose and incentivize a pair (e, m). The agency problem lies in the multitasking nature of the trader's activity. To see this, suppose the firm could impose a level of risk management m on the trader. Any level of effort e can then be obtained by offering a contract that promises $\overline{W}(\hat{\Pi}) = \underline{W}(\hat{\Pi}) = D\kappa e$ if $\hat{\Pi} \geq \Pi(e, m)$ and $\overline{W}(\hat{\Pi}) = \underline{W}(\hat{\Pi}) = 0$ otherwise. Since this contract only compensates the trader for his cost of effort, delegating search does not in itself create agency costs for the firm.[26] However, it constrains the shape of the contract by imposing that payments be sensitive to profit: The trader is paid a bonus if he reaches a certain threshold performance. Conversely, if the firm could impose any level of search e on the trader, his pay could be decoupled from profit $\hat{\Pi}$, which would make him indifferent with respect to the level of compliance, m.

Now consider the case in which the firm must incentivize risk management as well as search. Risk management lowers profit $\hat{\Pi}$ because some of the trading opportunities the trader generates are either blocked or preempted while risk management takes place. Since incentivizing search requires a bonus when $\hat{\Pi}$ reaches a certain threshold, the trader endogenously has incentives to inflate profits through manipulation rather than costly effort. To elicit compliance, the contract must punish him when manipulation is detected ex post: $\overline{W}(\hat{\Pi}) - \underline{W}(\hat{\Pi}) > 0$. $\overline{W}(\hat{\Pi}) - \underline{W}(\hat{\Pi})$ can be interpreted as wages that are clawed back from the trader. Because ex-post verification is imperfect, that amount needs to be high enough, which may conflict with the limited liability of the agent ($\underline{W}(\hat{\Pi}) \geq 0$). In that case, the conflict between incentivizing search and compliance creates an agency cost for the firm. Let

$$m_0 \equiv \min \left\{ \frac{1}{2} \left[\frac{1}{\gamma} - \frac{p\rho}{1 - p\rho} \right], 1 \right\}.$$

Lemma 3. *A cost-minimal contract that incentivizes search effort e and manipulation m features*

- *a bonus: $\overline{W}(\hat{\Pi}) = 0$ if $\hat{\Pi} < \Pi(e, m)$ and $\overline{W}(\hat{\Pi})$ is strictly positive and non-decreasing if $\hat{\Pi} \geq \Pi(e, m)$,*
- *a claw-back: if $m < 1$ and $\hat{\Pi} \geq \Pi(e, m)$, then $\overline{W}(\hat{\Pi}) - \underline{W}(\hat{\Pi}) > 0$; furthermore if $m < m_0$, then $\underline{W}(\hat{\Pi}) = 0$ for any $\hat{\Pi}$.*

[26] The contract is incentive compatible because the trader is indifferent between exerting no search effort and searching at intensity e. Any other level of effort is strictly dominated.

In that contract, the expected payment to the trader per unit of time is

$$\begin{cases} \kappa e & \text{if } m \geq m_0, \\ \kappa e + r(e, m) & \text{if } m < m_0, \end{cases}$$

where $r(e, m) \equiv \dfrac{2(1 - p\rho)(m_0 - m)}{p\rho + (1 - p\rho)m}\kappa e$ *represents the trader's rent.*

Let us highlight two consequences of Lemma 3. First, as hinted earlier, the cost of delegation to the firm is independent of the trader's tenure D. On one hand, when D increases, clawbacks are a stronger deterrent to non-compliance because more total compensation is at stake. On the other hand, it becomes easier for a trader to "hide" (a given mass of) manipulated trades as they are spread out across more trades in total. (Recall that the probability of detection depends on the *average* manipulation rate m.) In our setup, these two effects, neither of which is crucial to our qualitative results, cancel each other out conveniently.

Second, the firm's gross revenue from trading (12) and the trader's expected compensation are both linear in e. This linearity implies that equilibrium search choice is "bang-bang," which together with assumption (10) yields the next corollary.[27]

Corollary 3. *In an optimal contract, the trader exerts maximum search effort,* $e^* = \lambda^{-1}$.

In the remainder of the paper, we will omit search effort in the expression of the trader's rent and write $r(m) \equiv r(\lambda^{-1}, m)$.

4.3 Contractual externalities and risk cultures

The next proposition highlights the central mechanism that underlies the equilibrium effect of delegation: Time pressure in the market exacerbates agency problems within firms.

Proposition 5. *If* $m < m_0$*, the firm's marginal cost of incentivizing compliance,* $-r'(m)$*, strictly increases as time pressure rises (*T *decreases).*

In words, the agency cost of risk management is *sensitive to time pressure.* Intuitively, as the threat of preemption rises, manipulation becomes more attractive to the trader. Formally, when $p = 1 - \frac{\iota}{T}$ decreases, the difference between execution probability 1 (without risk management) and execution probability $p\rho$ (with risk management) increases, and with it the marginal effect

[27]Note that this property is driven by the linearity of the effort cost function κe. It allows separating the analysis of the optimal effort from the analysis of the optimal compliance.

of manipulation on profits. Since the trader's marginal cost of manipulation (the increase in the probability of ex-post detection) is independent of time pressure, the net incentive for the trader to manipulate increases. Note that m_0 is itself increasing with time pressure, so that heightened time pressure does not only cause the rent to increase for values of m where it is already strictly positive, it also widens the range of manipulation levels for which the rent is strictly positive, as illustrated in Figure 4.

[Figure 4 about here.]

Coupled with preemptive competition, the agency problem thus creates *contractual externalities*: each firm ignores the impact his chosen incentive scheme has on the agency rent other firms must leave with their agents to incentivize risk management. Agency reinforces the race to the bottom not only through the "partial equilibrium" effect that risk management becomes costlier within each firm but also through the "general equilibrium" effect that firms make it costlier *for each other*. The former speaks to the shape of the "second-best" contract inside a firm, whereas the latter contributes to the constrained inefficiency of the market outcome.

Note also that while the agency rent $r(m)$ depends on time pressure, as explained above, through the probability p that a trade is preempted during risk management, it does not vary directly with the (stationary) mass T of discoverable trading opportunities in the market (see equation (8)). Intuitively, a trader's rent depends on the rate at which he can substitute manipulation for costly search for a given profit level. When trading opportunities are more abundant, the marginal impact of search and manipulation on profit both increase linearly in T (keeping p constant), but the marginal rate of substitution between the two remains the same.

However, a lower T erodes firms' revenues as trading opportunities are scarcer. As a result, a firm's marginal benefit from compliance,

$$\lambda T[p\rho(\pi + \alpha_+) - \pi], \tag{13}$$

scales down with T (once again, keeping p constant). This effect is irrelevant in the absence of agency costs, because only the sign of (13) matters to the decision, not the scale (see equation (7)). Here, the scale of (13) *relative* to the marginal cost of incentivizing compliance, $-r'(m)$, affects the firm's optimal strategy: Heightened time pressure increases the *average cost* of risk management by making compliance per unit of identified trading opportunity more expensive at the margin. This effect further reinforces the contractual externality.[28]

[28]The key feature here is that, for a given p, agency rents do not scale with the mass of trading opportunities,

One interpretation of the strategic complementarities that operate solely through contracts is that they formalize the notion of risk "cultures" and sales-driven front office "cultures," which have been cited as intangible obstacles to reforming risk management. In fact, these contractual externalities can sustain coordination failures even when the opportunity cost of risk management (preemption) vanishes, that is, *even when the "physical" friction ι is taken to 0.*

Proposition 6. *Even for $\iota = 0$, the equilibrium level of risk management is lower than under full coordination. There is a region $[\underline{\pi}', \overline{\pi}'] \neq \{\emptyset\}$ where this inequality is strict.*

Firms do not internalize that, even in the absence of risk management latency, contract choices still affect time pressure: Trading opportunities are depleted at a faster rate when compliance deteriorates, as trades that sidestep pre-trade control are indiscriminately executed as opposed to only when $\alpha_k^t = \alpha_+$. Time pressure, in turn, affects agency costs, which closes the loop.[29] This special case highlights that through contractual externalities, incentive structures that prioritize execution over risk control can be self-enforcing.

In the general case where $\iota > 0$, delegation amplifies the strategic complementarities present in the original model. As time pressure increases, higher preemption risk raises the opportunity cost of risk management while heightened incentive problems amplify the agency cost of risk management. Both effects induce lower compliance which in turn raises time pressure. Overall, equilibria with a high level of risk management become more difficult to sustain.[30]

Proposition 7. *For any given set of parameters, the highest and lowest risk management quality that can be supported as an equilibrium outcome is weakly, and sometimes strictly, lower in the setting with agency than in the setting without agency.*

Finally, delegation does not only make the coordination problem more severe ex post, it also makes firms less inclined to agree to a high level of risk management ex ante, even if they could commit to it.

Proposition 8. *Principals may not implement $m = 0$ even if they can coordinate ex ante.*

i.e., they are a *fixed cost*. This suggests that other types of fixed costs (e.g., maintaining an information system) could also reinforce strategic complementarity. However, such costs would differ from agency costs and opportunity costs, in the presence of risk management latency, insomuch as they neither increase in time pressure p nor scale up with the value of (possibly forgone) trading opportunities π.

[29] When $\iota = 0$, the probability of preemption during risk management is 0 ($p = 1$), so that the absolute agency cost (the trader's rent) does not depend on T. However, as explained above, the average agency cost per identified trading opportunity increases when time pressure intensifies.

[30] See Lemma 4 in the Appendix for an explicit derivation of best response functions.

As ex-post verification becomes less effective (γ decreases), principals may find the provision of compliance incentives prohibitively expensive even if all the other firms were fully deliberate. Here, the inefficiency does not stem from a coordination problem, but from firms' failure to internalize the entire surplus from risk management, part of which needs to be paid as rents to traders.

Propositions 5-8 suggest that agency problems increase the potential for systematic lack of risk management and the scope for regulation. We discuss the implications of the various sources of externalities and inefficiencies for regulation in Section 6.

4.4 Traders' compensation and risk management quality

The preceding analysis implies that different risk management regimes require different incentive contracts. Since contracts or compensation are (to some extent) observable, they offer a potential instrument for regulating risk management inside firms. For instance, in 2010, the European Parliament introduced a limitation of the ratio between the variable and fixed components of remuneration to 100% with the explicit objective of limiting incentives for risk-taking induced by profit-sensitive remunerations within banks.[31]

In the model, when the trader perceives a rent ($m^* < m_0$), the binding incentive constraint is the one that governs the level of compliance. The incentive constraint that governs search and imposes that pay be sensitive to profit is then slack so that part of the trader's remuneration could be paid as a fixed wage.[32] In the optimal contract with the highest fixed component, we derive the ratio of fixed pay to total pay (i.e., fixed pay plus bonus). In practice, this measure is sometimes referred to as the "pay mix" or the "salary/incentive ratio," and is higher if the contract is lower-powered (less sensitive to profit). It takes a simple analytical form in our model:

$$\text{S/I-ratio} = 2\gamma \max\{m_0 - m^*, 0\} \tag{14}$$

From a casual inspection of (14), one might conclude that high risk management (low m^*) is indeed associated with a high S/I-ratio (low-powered incentives). However, we show next that equilibrium effects make the mapping between incentives and risk management quality difficult to interpret.

Proposition 9. *In the Pareto-superior equilibrium, when competition intensifies (i decreases), risk management deteriorates. Meanwhile, traders' expected compensation and the S/I-ratio first*

[31] Directive 2010/76/EU of the European Parliament and the Council of 24 November 2010.

[32] See the proof of Proposition 9 in the Appendix.

increase and then decrease.

The traders' compensation and the S/I-ratio are driven by agency rents. On one hand, since rents grow with compliance *keeping T fixed*, both measures tend to increase when the level of risk management is higher. On the other hand, higher *equilibrium* compliance is related to lower time pressure which mitigates the agency problem, leading to lower rents. (In (14), T enters the S/I-ratio through m_0.) When i is large and time pressure is low, risk management is high and cheap to incentivize. As a result, rents are low. When competition intensifies, risk management deteriorates but becomes more expensive at the margin. The second effect first dominates causing rents to increase in spite of declining compliance before the first effect takes over causing rents to revert and eventually reach 0.

While compensation regulation can be helpful, Proposition 9 points to some caveats: A high S/I-ratio does not rule out a role for regulation, as it need not be a sign of (constrained) efficient incentives. At the same time, a low S/I-ratio does not necessarily indicate a need for regulation, and in that case, constraining firms to a higher level can impose superfluous costs on principals, and may even destroy equilibria with high levels of risk management.[33] Moreover, even if useful, mandating a higher S/I-ratio may only improve incentives marginally instead of shifting the market to a Pareto-superior equilibrium – "culture" may be crucial. The crux is that compensation per se is neither a sufficient signal of inefficiencies nor always a sufficient remedy, an issue we will discuss further in Section 6.

5 Robustness and extensions

Time-dependent deadline. In our model, a trading opportunity persists until it is fully exploited by traders. In a more general setup, the lifetime of a trading opportunity could also be directly dependent on the time since it appeared in the market. For instance, if the counterparty seeks to hedge a specific change in a risk factor, the change may materialize before a trader offers a hedge; or the counterparty may have a liquidity demand that, if missed, can disappear over time. We introduce time-dependence by assuming that the trading opportunity may disappear with some constant probability in any small interval of time dt before it is exhausted (see Online Appendix).

This enriches our analysis in two ways. First, it makes traders internalize a social benefit of fast execution, that is, being more likely to seize the common value π before it disappears from

[33] We can show that in our model mandating a S/I-ratio may both increase the trader's pay and decrease the equilibrium level of risk management.

the market. Indeed, when the trading opportunity can disappear before it is fully exhausted, it is no longer true that all traders being deliberate is Pareto-optimal. This offers partial justification for a market design that rewards speed by processing trades according to some time priority rule, which is tacitly assumed in our model and true for many financial markets in practice. However, coordination failure remains a problem: A hasty equilibrium can still coexist with a deliberate one, and whenever this is the case, it is Pareto-dominated.

Second, this extension delivers volume implications. In expectation, there is more trading in a hasty equilibrium than in a deliberate one, implying an ambiguous relationship between trading volume and allocative efficiency: A higher trading volume reflects both the social benefit of fast execution as well as the social cost of less risk management. In the latter respect, trading volume can be "excessive" in that it manifests a worse allocation of risks in the market.

Alternative modeling of trading. Our model of trading is stylized. For instance, firms cannot choose how many units to trade, nor can they reverse previous trades. While these assumptions may seem restrictive, they are not crucial for our results. Allowing firms to trade more units tends to reinforce the threat of preemption and hence the fragility of the deliberate equilibrium under time pressure. (This is true even when a firm's willingness to trade more units increases with active risk management.) Furthermore, our results are robust to re-trading as long as trades are *partly* irreversible, for example, due to (a duplication of) transaction costs. In fact, the effect of partial reversibility is ambiguous: By allowing efficient *re*-allocations, it also lowers the private value of ex ante risk management and makes hasty trading more likely to begin with.

The way in which trading pressure affects the magnitude of the trading opportunity could also be modeled differently. In particular, the common value π could continuously decrease as more traders execute the trade, reflecting a price-sensitive demand for liquidity as in Kondor (2009). This has two countervailing effects on the strategic complementarities between traders. On one hand, when more traders are hasty, the expected common value at which a trader can execute the trade is lower, which makes risk management more desirable. On the other hand, when more traders are hasty, the common value shrinks at a faster rate, which heightens the preemption motive and thus makes risk management less desirable. One can show that when π decreases linearly with the mass of (traders who) executed trades, the second effect dominates the first one, thereby reinforcing strategic complementarities.

Risk accumulation. Beyond measuring risk at the level of a specific activity or portfolio, the ultimate objective of risk management in large financial firms is to monitor the cumulative build-up of risk exposures throughout the firm. As mentioned in the motivating remarks (Section 2.2.1), a comprehensive and thorough evaluation of aggregate exposures in large financial institutions (e.g., stress tests) can take several months. While the length of this time frame reflects the sheer complexity of financial products and the difficulty of measuring correlations across divisions and businesses, it could also be the product of organizational tensions. Activities that are performed under time pressure may require decentralized controls. In a survey of risk management practices (Deloitte, 2013, 11), 54 percent of financial institutions reported that trading transaction approval was determined at the business unit level, compared to 28 percent who cited independent risk management.[34] The time it takes a firm to aggregate risks then depends on trading desks' efforts to adapt their own procedures and activities to facilitate the measurement and communication of risks. However, such adaptation often involves a conflict of interest (Mehta et al., 2012, 9):

> While the front-office teams prized high flexibility and finely calibrated pricing models to *facilitate innovation in quickly changing markets* [emphasis added], the finance function and risk group were focused on meeting regulatory, accounting, and internal standards. As business complexity increased, these separate systems agreed less and less often... At many banks today, aggregating and verifying market risk across the bank in real time has become a significant challenge.

Our model can be modified to capture this alternative manifestation of the conflict between trading and risk management. Suppose the time it takes a firm to get a snapshot of its total risk exposure depends on resources spent by decentralized trading desks on complying with protocols to feed information about their trading operations into the firm's risk management system. At the same time, suppose that such resources spent on compliance come at the expense of search.

This creates the following trade-off. Implementing more compliance slows down trading but allows the firm to detect accumulated risks earlier. Conversely, if the firm compromises on risk management, it promotes trading but allows undesirable risks to accumulate (i) for a longer time before being detected and (ii) more quickly as a result of faster trading. Such a trade-off can lead to best responses and strategic complementarities similar to the ones in our main analysis.[35]

[34] By contrast, two-thirds of institutions said counterparty risk limit excess approval and credit policy exception approval were determined by independent risk management, while only about 10 percent said these were determined by their business units. The fact that risk management is more centralized for counterparty risk and credit policy than for trading transactions is consistent with our model if the latter decisions are more time-sensitive.

[35] Stulz (2008) provides an account of a massive loss at the Union Bank of Switzerland that resulted from the failure to fully aggregate information from all trading desks because it had decided not to integrate the computers of a very successful derivatives trading group into its risk management information system. The loss was in part responsible for the bank having to merge with another Swiss bank. The reason stated for why the risk management system was not fully integrated is that it was deemed "more important to let the traders make money than disrupt what they were doing" (63).

On-the-fly risk management choice. In our baseline model the decision whether to activate risk management is made ex ante. Our model thus analyzes firms' incentives to set up risk management protocols that systematically oversee trading activity. In the Online Appendix, we study a setting in which risk management decisions are made "on the fly," at the time trading opportunities are located. Using iterated deletion of strictly dominated strategies, we show that this specification has a unique pure-strategy equilibrium, even in the absence of aggregate uncertainty, unlike the model studied above (see Proposition 1). This alternative version still generates a coordination failure in that there is too little risk management in equilibrium, and thus a misallocation of risk.

Continuous risk management choice. The risk management choice in Section 3 is binary. A richer model could let firms choose, on a continuous scale, the extent to which they trade off speed against the accuracy of risk management signals. Such a model would resemble the agency extension in Section 4 where risk management quality is captured by the compliance rate m: There could be stable equilibria with interior levels of risk management quality, and since best responses are a correspondence, even with heterogenous levels across firms.[36]

6 Risk management regulation

Practitioners commonly note that financial firms must balance "business needs and risk appetite" (Ernst & Young, 2013, 12). While this is true for each firm in our model, our analysis also suggests that if all firms freely strike their own balance, the market can be constrained inefficient. This creates scope for regulation to curb excessive trading and to improve the risk allocation. Below we discuss regulatory approaches from this perspective, some used in practice and others so far only debated.

Capital and liquidity requirements. Leverage or maturity play no explicit role in our model. Rather, any impact they might have is implicit in the private value $\tilde{\alpha}_k$. There are two possible interpretations: First, management acts in the interest of all investors, and $\tilde{\alpha}_k$ reflects potential *deadweight* losses from undesirable risks given the firm's capital structure. In this case, $|\tilde{\alpha}_k|$ may increase with leverage (e.g., due to bankruptcy costs). Second, management is biased toward specific investors (e.g., shareholders), whose vulnerability alone is represented by $\tilde{\alpha}_k$, in which

[36]Of course, equilibria with heterogenous risk management also obtain with exogenous differences in speed.

case $|\tilde{\alpha}_k|$ may decrease with leverage (e.g., due to risk shifting).[37] Constrained inefficiencies in our model are *conditional* on $\tilde{\alpha}_k$, irrespective of *whose* "need" for risk management it reflects. Thus, capital structure regulation in our model would modify the need for risk management, but not correct any inefficiencies conditional on that need.

Pigouvian approaches. In the constrained inefficient outcome of our model, firms trade "too much" in that trades are not selective enough. One countermeasure is hence to levy a tax τ on every trade.[38] This decreases the value of a trading opportunity to $\hat{\pi} = \pi - \tau$, which has two effects: On one hand, it lowers the opportunity cost of risk management. On the other hand, it makes trades with $\hat{\pi} + \alpha_+ < 0$ unprofitable. The tax can thus deter excessive as well as valuable trade.

Given the role of the external-internal speed ratio $\frac{\iota}{\lambda}$, one could also "tax" speed investments.[39] Discouraging external speed (increasing λ) is also a double-edged sword: The decrease in time pressure promotes risk management, but lowers market immediacy and trading volume for valuable trades. By contrast, subsidizing internal speed (decreasing ι) helps risk management, and conditional thereon, also raises speed. Technological investment into information processes inside banks is currently a risk management priority (Ernst & Young, 2013, 68f):

> Systems and data vied for the top spot on the challenges to internal transparency... and, indeed, have been raised as among the top challenges throughout this report. "There is a huge effort underway to redo all the plumbing, data aggregation, accuracy, quality of information," one executive said. "That's the framework in which a lot of our future-state risk systems will be addressed... a huge, multiyear, gazillion-dollar effort."

In our model, such investments in internal speed, in practice often only made to meet regulatory requirements, contribute to a "public good" by decreasing the likelihood of coordination failure (Section 3.4.3).

Governance regulation. To be effective against both of the mutually reinforcing externalities in our model, opportunity costs and agency rents, regulation must address risk management in

[37] Maćkowiak and Wiederholt (2012) show that a management with risk-shifting incentives, being less concerned about uncertainty, may underinvest in information. Correcting management incentives would in this case resolve the lack of information acquisition. The converse is not true: Better information per se would not curb risk-shifting. Quite the contrary, it would enable management to shift risks more effectively.

[38] Tobin (1978) proposed transaction taxes based on the argument that excess currency speculation undermines the allocative role of exchange rates for hedging purposes, i.e., risk allocation.

[39] Budish, Cramton and Shim (2013) provide examples of such arms races between trading firms.

conjunction with incentive compensation. This resonates with the view taken by the Fed in its *Guidance on Sound Incentive Compensation Policies* (see footnote 18):

> [S]trong and effective risk-management and internal control functions are critical to the safety and soundness of banking organizations. However,... poorly designed or managed incentive compensation arrangements can themselves be a source of risk to banking organizations and undermine the controls in place. Unbalanced incentive compensation arrangements can place substantial strain on the risk-management and internal control functions of even well-managed organizations... [and] encourage employees to take affirmative actions to weaken the organization's risk-management or internal control functions. (36401)

The guidelines recommend supervising compensation practices (e.g., deferred pay and clawbacks) and associated risk control and governance processes together (36397).[40] Similarly, a qualitative inspection of risk management is now part of regulatory assessments of bank capital adequacy. For instance, in 2014, the Fed rejected the capital plans of four large financial institutions on grounds of qualitative deficiencies in their risk management processes.[41]

Our model supports deferred pay and clawbacks but provides a more nuanced view on "bonus cultures." In particular, it suggests that an exclusive reliance on bonuses in trader compensation need not imply weak compliance incentives, provided the bonuses are deferred and can be clawed back. In fact, high-powered contracts can be efficient when a firm's risk controls are strong, risk management is not a major concern, or time pressure is low. A blanket requirement to use lower-powered contracts would then impose unnecessary agency costs on those trading activities with unintended side effects (such as discouraging the activity or investment in other risk controls). Instead, our analysis suggests confining regulatory intervention to "problem areas" that display high-powered compensation in conjunction with other characteristics, such as a high degree of time pressure, reliance on soft information, slow pre-trade controls, and weak post-trade controls (see Figure 5).

[Figure 5 about here.]

[40] See also SR letter 12-17, "Consolidated Supervision Framework for Large Financial Institutions," released by the Fed in 2012.

[41] The banks are Citibank, HSBC, RBS Citizens, and Santander. In the case of Citibank, the capital plan was rejected because of concerns about its ability to properly aggregate risk information across its business units, or more precisely, its "ability to project revenue and losses under a stressful scenario for material parts of the firm's global operations, and its ability to develop scenarios for its internal stress testing that adequately reflect and stress its full range of business activities and exposures" (Federal Reserve Board, Comprehensive Capital Analysis and Review 2014: Assessment Framework and Results, 7).

Still, ex ante restrictions may be problematic. For one, whether risk controls are (too) slow or weak can be hard to assess, and is often discernible only after the fact. Moreover, since agency rents may constrain risk management even in the constrained efficient equilibrium, compensation regulation may face the dilemma that lax rules provoke weak risk management while mandating sufficiently low-powered compensation renders the trading activity unprofitable for firms.

It may be more effective to enforce standards or duties through *ex post* liability, in which case the question arises whether institutions or individuals should be held liable for risk management failures. According to our model, the answer is both. On one hand, individual liability fails to address the "opportunity cost effect" that firms may (tacitly) tolerate or even want lax controls and non-compliance. For instance, in the London Whale scandal, J.P. Morgan Chase was fined nearly $1 billion for unsound risk protocols (Dayen, 2013):

> The internal controls – the key fraud prevention device inside the company – were a joke. The Chief Investment Office in London had a Valuation Control Group (VCG) that was supposed to act as a check on mis-marking or other violations. But it had only one employee for a large trading desk. And the employee would get price quotes from the traders themselves, like asking the fox for statistics on the hen house...
>
> This was a license to cheat, and the VCG guidelines could only have come from the risk management officers at the bank. Traders "took full advantage" of the VCG's laissez-faire approach to valuations,... and would lobby successfully for even more leeway. Essentially, there was no risk management at the Chief Investment Office, and senior executives were all too happy to not be apprised of the details.

Afterwards, the bank substantially expanded its valuation control group and pledged to impose stricter discipline on valuations as well as "check traders' valuations more frequently than its previous practice of once a month" (Zuckerman and Fitzpatrick, 2012).[42]

At the same time, institutional liability fails to address the "agency cost effect." Here individual liability helps: it relaxes incentive constraints, making it cheaper for firms to simultaneously provide search and compliance incentives. In cases where compliance is high even without regulation, this is of no consequence. But where agency rents constrain risk management, it helps firms mitigate the multi-task conflict that drives the inefficiencies in both the constrained inefficient and second-best equilibria of our model.

[42] Following the scandal and the regulatory fine, J.P. Morgan Chase invested close to $1 billion in 2013 on strengthening internal controls and assigned more than 5,000 employees to compliance.

Market design. Regulators may also want to reconsider market processes. For one, the coordination failure in our model is driven by preemptive competition, which is a result of time priority in market rules or interactions. Moreover, as markets speed up further, algorithmic trading becomes increasingly attractive, making traditional notions of internal governance obsolete; machines are fast and demand no agency rents. Regulators would have to assess to what extent an algorithmic strategy or computer code sidesteps risk controls, which may prove difficult.

Kirilenko and Lo (2013) caution that the gap between "machine speed" and "human speed" created by algorithmic trading may aggravate the lack of risk management; and suggest that a growing focus on speed and machines may necessitate *systemwide* risk management regulation "translated into computer code and executed by automated systems" with "safeguards at multiple levels of the system." In other words, interventions into market processes may become integral. Among measures currently debated are proposals (i) to discretize trading time to eliminate preemptive competition at very short intervals and (ii) to build safeguards, such as trade limits and pre-trade protocols, into the system at the level of intermediaries or central counterparties, such as dealer-brokers, clearing houses, or exchanges (see, e.g, Kirilenko and Lo, 2013; Budish, Cramton and Shim, 2013; Clark, 2012).[43]

To place the case for market design interventions in context, consider Grossman and Miller (1988)'s framework in which *equilibrium market structure* is the outcome of the tradeoff between the costs (to intermediaries) of maintaining a continuous presence in a market and the benefits (to traders) of being able to transact as immediately as possible. In their setting, there is no over-provision of immediacy.[44] In our setting, immediacy can invoke a race to the bottom with respect to risk management by traders, which is why interventions in market design to reduce immediacy can be Pareto-improving.

Regulatory competition. When competing firms belong to different jurisdictions, regulators inherit the race-to-the-bottom incentives identified in our model insofar as they care about the competitiveness of "their" firms. Concerns about such regulatory competition indeed exist (Clark,

[43]In a survey on the risks of high-frequency trading, proprietary trading firms – when asked what they would change for "the betterment of the markets" – mention inter alia that (i) "requiring trading venues to uniformly apply pre-trade risk checks for all market participants would consistently apply latency to and level the playing field for all trading firms" and that (ii) "every trading venue should have limits on maximum positions, quantity per order, and credit... [and] on number of messages that can be sent to the trading venue within a specified period of time... per product/customer" (Clark and Ranjan, 2012, 13f).

[44]In Pagnotta and Philippon (2013), immediacy is determined by technological investments of the exchanges on which investors trade, and the exchanges' choices, which shape market structure in their model, are driven by differentiation incentives. In aggregate, these investments can be too high relative to the welfare optimum since they not only accelerate trading but also (are meant to) relax competition between the exchanges.

2010, 4):

> Issues related to risk management of these technology-dependent trading systems are numerous and complex and cannot be addressed in isolation within domestic financial markets. For example, placing limits on high-frequency algorithmic trading or restricting unfiltered sponsored access and co-location within one jurisdiction might only drive trading firms to another jurisdiction where controls are less stringent.

Similar concerns are voiced in the aforementioned *Guidance on Sound Incentive Compensation Policies* (36399).

7 Related literature

Smith and Stulz (1985) and Froot, Scharfstein and Stein (1993) were first to formalize why and how firms should hedge exposures to idiosyncratic risk in the presence of external financing frictions.[45] Rampini and Viswanathan (2010) refine this theory qualifying when risk management is optimal if hedging is subject to the same frictions as financing. In their model, risk management incurs opportunity costs in that collateral committed to hedging contracts reduces a firm's capacity to finance current investment. While our paper shares a focus on the costs of risk management, there are two main differences. First, Rampini and Viswanathan consider a firm's problem of whether or how to insure known risks. By contrast, we focus on firms' decisions to set up systems to monitor the take-up of, if unmonitored, unknown risks. Second, in our model, the resource that firms commit to risk management is time and opportunity costs of risk management arise from preemption in financial markets.

Preemption is similar to the first-come-first-served rule in bank run models (Bryant, 1980; Diamond and Dybvig, 1983), from which our model departs in two noteworthy ways.[46] First, risk management design is the outcome of long-run decisions that do not coincide with on-the-spot preemptive actions: such organizational choices precede individual trades. Yet since we model trade as randomly staggered through time as a result of independent search processes, preemption motives pass via "time pressure" to risk management choices, which through this medium inherit

[45] In reduced form, the benefit of risk management is a private value (firm-specific benefit of hedging idiosyncratic risk) of entering a financial contract that is traded at a common value (market price of hedging contract). In earlier work, Stulz (1984) analyzes optimal hedging policies from the perspective of a risk-averse manager (employee).

[46] See also Postlewaite and Vives (1987), Chari and Jagannathan (1988), Jacklin and Bhattacharya (1988), Allen and Gale (1998), Chen (1999), Peck and Shell (2003), and additional references in the survey by Gorton and Winton (2003).

the strategic complementarities known from bank runs. Because of this structural similarity, we can also adapt global games techniques used to refine bank run equilibria (Goldstein and Pauzner, 2005) to risk management equilibria by dispersing firms' expectations of time pressure.

Second, since risk management choices and traders' actions are distinct, we can further introduce agency problems that firms must address to implement their chosen risk management framework. This allows us to study the *interaction* of agency problems across firms subject to "bank run" externalities. Costly monitoring theories are common in financial intermediation, internal capital markets, and corporate governance (e.g., Diamond, 1984; Gertner, Scharfstein and Stein, 1994; Holmstrom and Tirole, 1997; Burkart, Gromb and Panunzi, 1997). We are, however, unaware of existing work where the cost of monitoring is time, or depends on competition or on others' monitoring choices.

Our theory provides a rationale for risk management regulation – as distinct from capital or liquidity regulation – similar to theories that justify corporate governance regulation based on externalities. This literature has focused on pecuniary externalities (Acharya and Volpin, 2010; Dicks, 2012) and learning externalities (Nielsen, 2006; Raff, 2011; Cheng, 2011; Acharya, Pagano and Volpin, 2013) in the context of managerial labor markets.

Links between competition and risk taking have been studied in the banking literature with a focus on the effect of competition on bank franchise values (Keeley, 1990; Hellmann, Murdock and Stiglitz, 2000; Boyd and De Nicolo, 2005; Martinez-Miera and Repullo, 2010) and the returns to screening (Ruckes, 2004; Dell'Ariccia and Marquez, 2004).[47] The key mechanism in our paper is that competition raises the (opportunity and agency) *costs* of screening. Our analysis of agency is reminiscent of Heider and Inderst (2012) who examine a multi-task conflict between screening and loan "prospecting," which – despite lacking an explicit time dimension – bears similarity to preemptive competition.

A literature in industrial organization studies more broadly how competition interacts with agency and has identified a variety of effects operating through information revelation, marginal returns to managerial effort, and total firm income.[48] The overall effect is generally ambiguous, qualifying the "Hicks conjecture" that product market competition curbs managerial slack. These papers typically study one-dimensional moral hazard in oligopolistic models. Our results rely on multi-dimensional moral hazard, and in this respect, are closer to Benabou and Tirole (2014) who study how labor market competition skews contractual incentives across different types of

[47] A different perspective is taken in Parlour and Rajan (2001) where lenders to the same borrower exert negative externalities on each other by raising the borrower's overall default incentives.

[48] See, e.g., Nalebuff and Stiglitz (1983); Scharfstein (1988); Hermalin (1992); Martin (1993); Schmidt (1997); Raith (2003); Golan, Parlour and Rajan (2014).

tasks.

Time-based competition is essential to the sizable literature on innovation and patent races. Most of this literature uses sequential games or real options models in which strategic choices coincide with the acts of preemption.[49] As mentioned earlier, the strategic choice in our model – whether to run risk management – is made *ex ante*. Our model is hence more similar to the one in Askenazy, Thesmar and Thoenig (2006) in which firms that compete on innovation choose ex ante between "mechanistic" and "organistic" organizational designs that differ in production efficiency and "time-to-market."

Recently, time-based competition has become the focus of research on high-frequency (low-latency) trading in financial markets. Apart from showing that the race to reduce latency spurs overinvestment in technology, these papers trace out the impact on market liquidity, asset prices, and trading volume (Budish, Cramton and Shim, 2013; Biais, Foucault and Moinas, 2014; Dugast and Foucault, 2014).[50] Our analysis is not specific to high-frequency trading but shares a similar view, and adds to the list of concerns that competing on speed may impair the *risk allocation* in financial markets by undermining governance processes inside firms.

This connects our paper to the literature on the allocative role of secondary capital markets. Most existing results in this literature revolve around (efficient) prices as a source of information that can destroy risk-sharing opportunities, guide investment decisions, enhance incentive contracts, and frustrate takeovers.[51] In our model, allocation is driven not by information revealed through market prices but by processes inside firms, which are, however, affected by the market's speed.

Our focus on a risk management link between markets and organizations combines the perspectives of two recent papers. Garleanu and Pedersen (2007) study the role of risk management protocols in creating liquidity feedback loops in the market, whereas Landier, Sraer and Thesmar (2009) focus on the "dissent" function of risk management inside an organization and when this function may be compromised. In our model, market interactions and organizational choices are jointly determined by a trade-off between immediacy in the market and "dissent" in organizations.

[49] For work on real options in competitive environments, see, e.g., Leahy (1993), Caballero and Pindyck (1996), Grenadier (2002), and Aguerrevere (2009), who examine how strategic interactions affect firms' payoffs from exercising their options.

[50] A notable exception is Pagnotta and Philippon (2013) who, instead of focusing on preemption among traders, study the incentives of securities exchanges to offer trading platforms of different speed to heterogenous traders.

[51] See, e.g., Bond, Edmans and Goldstein (2012) for a survey of this literature.

8 Conclusion

The implementation of risk management requires monitoring and information processes to collect the relevant information inside firms. These processes take time and can delay investment decisions, which represents an opportunity cost that scales up with the size of the firms' investment opportunities when those opportunities are *short-lived.* Based on this premise, this paper has presented a theory to explain why risk management failures occur and also why there may be scope for risk management regulation not warranted by the capital structure of financial firms but rather the type of markets they compete in.

Financial markets are a natural context for the speed-information trade-off in our model, and there are several avenues we have left unexplored. We have expressed the private value of risk management in reduced form. A micro-foundation of the source of this value in multi-divisional firms could link risk management to questions about the boundaries of the firm. To focus our attention on risk management, we have also abstracted from learning about the common value of traded assets. It would be of interest to study how the strategic complementarities in our framework affect information aggregation. One could also examine the speed-information trade-off from the perspective of other market participants, such as managers that disclose information or learn from prices, or securities exchanges that can affect the speed at which trading unfolds.

Our framework could also be extended to other contexts. First, our formalization of "time pressure" lends itself to the analysis of strategic complementarities akin to those in bank runs or financial panics without the connotation of frenzy. It may be useful in modeling long-term organizational choices in a variety of settings with time-based competition other than financial markets, thereby expanding the applicability of the theoretical apparatus that has been developed for models of panics.

Second, costly monitoring or state verification models are common in principal-agent theory. The notion that the relevant cost of such information processes is time, and that this may determine optimal contracts in environments where time is of the essence, is more generally applicable beyond risk management. In particular, as we have shown, it naturally creates a tension between monitoring (by the principal) and initiative (by the agent), akin to those analyzed in the literature on delegation, but dependent on time pressure.

McGill University, New York University, Stanford Institute for Economic Policy Research, and European Corporate Governance Institute.

Appendix: Proofs of main results

Proof of Lemma 1

See Online Appendix.

Proof of Proposition 2

We derive the equilibrium of the general case in which ε can be bounded away from 0. The proof is in several steps, and we only show here the existence of a unique equilibrium *in threshold strategies*. The proof that any equilibrium is in threshold strategies follows Goldstein and Pauzner (2005) and is in the Online Appendix.

For a given realization of π, the proportion of hasty traders under a threshold strategy $\hat{s}$ is

$$q(\pi,\hat{s}) \equiv \begin{cases} 0 & \text{if } \pi \leq \hat{s}-\varepsilon, \\ \dfrac{\pi+\varepsilon-\hat{s}}{2\varepsilon} & \text{if } \hat{s}-\varepsilon < \pi < \hat{s}+\varepsilon, \\ 1 & \text{if } \pi \geq \hat{s}+\varepsilon \end{cases} \tag{15}$$

For a proportion q of hasty traders, the mass of trade executed by time T is

$$m(q,T) \equiv q p_h(T) + (1-q)\rho p_d(T),$$

Hence the time at which the trading opportunity is exhausted, $\tau(\pi,\hat{s})$, is solution to $m[q(\pi,\hat{s}),\tau] = i$. Finally, the net expected benefit of a deliberate strategy given a signal s_k and a threshold $\hat{s}$ is

$$u(s_k,\hat{s}) \equiv \mathbb{E}_\pi\{\Delta[\pi,\tau(\pi,\hat{s})]|s_k\} = \frac{1}{2\varepsilon}\int_{s_k-\varepsilon}^{s_k+\varepsilon} \Delta[\pi,\tau(\pi,\hat{s})]d\pi, \tag{16}$$

Step 1: Existence of a unique threshold equilibrium.

Claim 1. *$\tau(\pi,\hat{s})$ is decreasing in π and increasing in $\hat{s}$. Furthermore, $\tau(\pi+a,\hat{s}+a) = \tau(\pi,\hat{s})$.*

Proof. $q(\pi,\hat{s})$ is increasing in π and decreasing in $\hat{s}$. Furthermore, $m(q,T)$ is increasing in T, and since $p_h(T) > \rho p_d(T)$, increasing in q. Therefore, $\tau(\pi,\hat{s})$ is decreasing in π and increasing in $\hat{s}$. Finally, from (15), $q(\pi+a,\hat{s}+a) = q(\pi,\hat{s})$, which in turn implies $\tau(\pi+a,\hat{s}+a) = \tau(\pi,\hat{s})$. □

Claim 2. *There exists a unique s^* such that $u(s^*,s^*) = 0$.*

Proof. We first show the existence of s^* using upper- and lower-dominance regions. Suppose that $s < \underline{\pi} - \varepsilon$, then of any $\pi \in [s-\varepsilon, s+\varepsilon]$, $T^*(\pi) < T_h \leq \tau(\pi,s)$, therefore $\Delta[\pi,\tau(\pi,s)] > 0$ and hence $u(s,s) > 0$. Similarly, if $s > \overline{\pi} + \varepsilon$, then $u(s,s) < 0$. The continuity of $u(.,.)$ then implies the existence of s^*, which proves existence.

Furthermore,

$$
\begin{aligned}
u(s,s) &= \frac{1}{2\varepsilon}\int_{s-\varepsilon}^{s+\varepsilon} \Delta[\pi, \tau(\pi, s)]d\pi \\
&= \frac{1}{2\varepsilon}\int_{s-\varepsilon}^{s+\varepsilon} \Delta[\pi, \tau(\pi + a, s + a)]d\pi \\
&= \frac{1}{2\varepsilon}\int_{s+a-\varepsilon}^{s+a+\varepsilon} \Delta[\pi - a, \tau(\pi, s + a)]d\pi \\
&< \frac{1}{2\varepsilon}\int_{s+a-\varepsilon}^{s+a+\varepsilon} \Delta[\pi, \tau(\pi, s + a)]d\pi = u(s + a, s + a),
\end{aligned}
$$

where the second equality follows from Claim 1 and the last inequality follows from $\frac{\partial \Delta}{\partial \pi} < 0$. Hence, $u(s, s)$ is strictly decreasing in s, which proves uniqueness. □

Finally, to complete the proof we show the following result.

Claim 3. *$u(s, s^*) > 0$ for $s < s^*$ and $u(s, s^*) < 0$ for $s > s^*$.*

Proof. From (16), $u(s^*, s^*) = 0$ implies that $\Delta[., \tau(., s^*)]$ changes sign on $[s^* - \varepsilon, s^* + \varepsilon]$. Therefore, by continuity, there exists $\hat{\pi} \in [s^* - \varepsilon, s^* + \varepsilon]$ such that $\Delta[\hat{\pi}, \tau(\hat{\pi}, s^*)] = 0$, and hence, $\tau(\hat{\pi}, s^*) = T^*(\hat{\pi}) > 0$.[52] Suppose $\pi < \hat{\pi}$, then $\tau(\pi, s^*) \geq T^*(\hat{\pi})$, and therefore using the single crossing property (Lemma 1), $\Delta[\hat{\pi}, \tau(\pi, \hat{s})] \geq 0$. Furthermore, since $\frac{\partial \Delta}{\partial \pi} < 0$, $\Delta[\pi, \tau(\pi, \hat{s})] > \Delta[\hat{\pi}, \tau(\pi, \hat{s})] \geq 0$. Similarly, if $\pi > \hat{\pi}$, then $\Delta[\pi, \tau(\pi, \hat{s})] < 0$. This also shows that $\hat{\pi}$ is uniquely defined.

Suppose $s < s^*$. If $s < \hat{\pi} - \varepsilon$, for any $\pi \in [s - \varepsilon, s + \varepsilon]$, $\Delta[\pi, \tau(\pi, s^*)] > 0$, and thus $u(s, s^*) > 0$. If $\hat{\pi} - \varepsilon \leq s < s^*$,

$$
\begin{aligned}
u(s, s^*) - u(s^*, s^*) &= \frac{1}{2\varepsilon}\int_{s-\varepsilon}^{s+\varepsilon} \Delta[\pi, \tau(\pi, s^*)]d\pi - \frac{1}{2\varepsilon}\int_{s^*-\varepsilon}^{s^*+\varepsilon} \Delta[\pi, \tau(\pi, s^*)]d\pi \\
&= \frac{1}{2\varepsilon}\int_{s-\varepsilon}^{s^*-\varepsilon} \Delta[\pi, \tau(\pi, s^*)]d\pi - \frac{1}{2\varepsilon}\int_{s+\varepsilon}^{s^*+\varepsilon} \Delta[\pi, \tau(\pi, s^*)]d\pi.
\end{aligned}
$$

$\pi < s^* - \varepsilon$ implies $\pi < \hat{\pi}$, and therefore $\Delta[\pi, \tau(\pi, s^*)] > 0$. Thus, $\int_{s-\varepsilon}^{s^*-\varepsilon} \Delta[\pi, \tau(\pi, \hat{s})]d\pi > 0$. $\pi \geq s + \varepsilon$ implies $\pi \geq \hat{\pi}$ and therefore $\Delta[\pi, \tau(\pi, s^*)] < 0$. Hence, $\int_{s+\varepsilon}^{s^*+\varepsilon} \Delta[T(\pi, s^*)]d\pi < 0$. Therefore $u(s, s^*) - u(s^*, s^*) = u(s, s^*) > 0$. Symmetrically, if $s > s^*$, $u(s, s^*) < 0$. □

Step 2: Any equilibrium is a threshold equilibrium.

See Online Appendix.

Proof of Proposition 3

We show here that π^* is an increasing function of i, which together with the discussion in the main text, proves Proposition 3.

[52] If $T^*(\hat{\pi}) = 0$, then $\hat{\pi} \leq 0$ and $\Delta[\hat{\pi}, T] > 0$ for any $T > 0$, a contradiction.

Let $T(q,i)$ be defined as in (2), with the addition of the second argument explicitly recognizing its dependence on i. Let

$$U(\pi) \equiv \int_0^1 \Delta[\pi, T(q,i)]dq.$$

Note that π^* solves $U(\pi) = 0$. Note also that $U'(.) < 0$, and from (2), $T(q,i)$ is increasing in i. Consider two cases,

(a) $T_h(i) \geq \iota$

Then, for any $q \in [0,1)$, $T(q,i) > \iota$, and therefore $\frac{\partial \Delta}{\partial T}[\pi^*, T(q,i)] > 0$. This, in turn, implies

$$\frac{\partial U}{\partial i}(\pi^*) = \int_0^1 \frac{\partial \Delta}{\partial T}[\pi^*, T(q,i)] \frac{\partial T}{\partial i}(q,i)dq > 0,$$

and finally, by the implicit function theorem,

$$\frac{\partial \pi^*}{\partial i} = -\frac{\frac{\partial U}{\partial i}(\pi^*)}{U'(\pi^*)} > 0.$$

(b) $T_h(i) < \iota$

Let

$$\hat{q}(i) \equiv \frac{i}{p_h(\iota)}. \tag{17}$$

$$\begin{aligned} U(\pi^*) &= \int_0^{\hat{q}(i)} \{\rho(\pi^* + \alpha_+)p_d[T(q,i)] - \pi^* p_h[T(q,i)]\}dq - \int_{\hat{q}(i)}^1 \pi^* p_h[T(q,i)]dq \\ &= \hat{q}(i)\rho(\pi^* + \alpha_+) - \pi^* - \int_0^{\hat{q}(i)} \{\rho(\pi^* + \alpha_+)[1 - p_d[T(q,i)]] - \pi^*[1 - p_h[T(q,i)]]\}dq \\ &\quad + \int_{\hat{q}(i)}^1 \pi^*[1 - p_h[T(q,i)]]dq \end{aligned} \tag{18}$$

If $q < \hat{q}(i)$, then $1 - p_d[T(q,i)] = e^{\iota/\lambda}[1 - p_h[T(q,i)]]$. Using $qp_h[T(q,i)] + (1-q)\rho p_d[T(q,i)] = i$, we get

$$1 - p_d[T(q,i)] = e^{\iota/\lambda}[1 - p_h[T(q,i)]] = e^{\iota/\lambda}\frac{q + (1-q)\rho - i}{q + (1-q)\rho e^{\iota/\lambda}}. \tag{19}$$

Hence, the first integral in (18) becomes

$$[\rho e^{\iota/\lambda}(\pi^* + \alpha_+) - \pi^*] \int_0^{\hat{q}(i)} \frac{q + (1-q)\rho - i}{q + (1-q)\rho e^{\iota/\lambda}} dq. \tag{20}$$

If $q > \hat{q}(i)$, then $p_d[T(q,i)] = 0$. Using this in $qp_h[T(q,i)] + (1-q)\rho p_d[T(q,i)] = i$, we get $1 - p_h[T(q,i)] = \frac{q-i}{q}$. Hence, the second integral in (18) becomes

$$\pi^* \int_{\hat{q}(i)}^1 \frac{q-i}{q} dq \tag{21}$$

Consider the first line of equation (18). $U(\pi^*)$ depends on i both through the boundaries of the integrals (via $\hat{q}(i)$) and through the integrands (via $T(q,i)$). However, since $p_d[T(\hat{q}(i)), i] = 0$, the effect of a marginal change in i that goes through $\hat{q}(i)$ cancels out. Hence, using (19) and (20) to substitute into the second line of (18), we obtain

$$\frac{\partial U}{\partial i}(\pi^*) = [\rho e^{\iota/\lambda}(\pi^* + \alpha_+) - \pi^*] \int_0^{\hat{q}(i)} \frac{1}{q + (1-q)\rho e^{\iota/\lambda}} dq - \pi^* \int_{\hat{q}(i)}^1 \frac{1}{q} dq. \tag{22}$$

Now, if $q > \hat{q}(i)$,

$$\frac{1}{q} = \frac{p_h[(T(q)]}{i}, \tag{23}$$

In addition, rearranging (19),

$$p_d[T(q)] = \frac{q(1 - e^{\iota/\lambda}) + e^{\iota/\lambda} i}{q + (1-q)\rho e^{\iota/\lambda}} \text{ and } p_h[T(q)] = \frac{(1-q)\rho(e^{\iota/\lambda} - 1) + i}{q + (1-q)\rho e^{\iota/\lambda}},$$

which, since $e^{\iota/\lambda} > 1$, implies

$$\frac{p_d[T(q)]}{e^{\iota/\lambda} i} < \frac{1}{q + (1-q)\rho e^{\iota/\lambda}} < \frac{p_h[T(q)]}{i}. \tag{24}$$

Finally, using (22), (23) and (24),

$$i \frac{\partial U}{\partial i}(\pi^*) > \rho(\pi^* + \alpha_+) \int_0^{\hat{q}(i)} p_d[T(q)] dq - \pi^* \int_0^1 p_h[T(q)] dq$$

The RHS of this last inequality is $U(\pi^*) = 0$, and using again the implicit function theorem concludes the proof. □

Proof of Proposition 4

We show here that π^* is a decreasing function of $\frac{\iota}{\lambda}$, which together with the discussion in the main text, proves Proposition 4.

As in (17), let

$$\hat{q}(\iota/\lambda) \equiv \frac{i}{1 - e^{-\iota/\lambda}}.$$

We have

$$U(\pi^*) = \int_0^{\min\{\hat{q}(\iota/\lambda),1\}} \{\rho(\pi^* + \alpha_+) p_d[T(q)] - \pi^* p_h[T(q)]\} dq - \int_{\min\{\hat{q}(\iota/\lambda),1\}}^1 \pi^* p_h[T(q)] dq$$

Note that $U(\pi^*)$ depends on λ and ι both through the boundaries of the integrals and, implicitly, through the functions $p_d(.)$ and $p_h(.)$, that is, the probabilities of execution under each strategy. However, $p_d[T(\hat{q}(\iota/\lambda))] = 0$, and hence, the effect of a marginal change in λ or in ι on the integral boundaries cancels out. As a result, differentiating $U(\pi^*)$ with respect to λ or ι only requires differentiating the integrands.

Let $x \equiv e^{\iota/\lambda}$. Using equations (18), (20) and (21), one obtains

$$\frac{\partial U(\pi^*)}{\partial x} = -\frac{\partial}{\partial x} \left\{ [\rho x(\pi^* + \alpha_+) - \pi^*] \int_0^{\min\{\hat{q}(\iota/\lambda),1\}} \frac{q + (1-q)\rho - i}{q + (1-q)\rho x} dq \right\}. \tag{25}$$

It is easy to check that the expression between brackets is increasing in x. Therefore

$$\frac{\partial \pi^*}{\partial x} = -\frac{\frac{\partial U}{\partial x}(\pi^*)}{U'(\pi^*)} < 0.$$

This, in turn, implies that π^* is decreasing in ι and increasing in λ. □

Proof of Lemma 2

See Online Appendix.

Proof of Lemma 3

See Online Appendix.

Preliminary to the proofs of Propositions 5 to 8

We start by deriving the firm's choice of compliance $m(T)$ given a deadline T. That is, $m(T)$ is the firm's best response.

Lemma 4. *There exist three ordered thresholds $T^* < T^{**} < T^{***}$ such that firm's best response to time pressure is*

- *If $T < T^*$, $m(T) = 1$, the firm abandons risk management.*
- *If $T = T^*$, $m(T^*) \in [m_0(T^*), 1]$.*
- *If $T^* < T \leq T^{**}$, $m(T) = m_0(T)$.*
- *If $T^{**} < T < T^{***}$, $m(T)$ is a strictly decreasing function such that $m(T) < m_0(T)$.*
- *If $T \geq T^{***}$, $m(T) = 0$.*

Proof. It is suboptimal for the principal to implement positive compliance unless it has strictly positive value for the whole firm. Thus, for $T < T^*$, he optimally sets $m = 1$, and for $T = T^*$, he is indifferent between all $m \in [m_0(T^*), 1]$ where $m_0(T^*)$ is the smallest amount of manipulation that she can implement "for free," i.e., without paying the trader an agency rent. For $T > T^*$, the marginal value of compliance to the firm is positive, but to increase compliance (reduce m) below $m_0(T^*)$, the principal must pay an agency rent. The marginal agency rent is given by

$$\frac{\partial r}{\partial m}(\lambda^{-1}, m) = \begin{cases} -\frac{\frac{1}{\gamma}+\frac{p\rho}{1-p\rho}}{\left(\frac{p\rho}{1-p\rho}+m\right)^2}\lambda^{-1}\kappa & \text{if } m < m^0, \\ 0 & \text{otherwise.} \end{cases}$$

The value of this derivative at $m_0(T)$ is

$$\frac{\partial r}{\partial m}(\lambda, m_0) = -\frac{4}{\frac{1}{\gamma} + \frac{p\rho}{1-p\rho}}\kappa\lambda^{-1}.$$

Thus, even when compliance has positive value for the firm, the principal will not start increasing compliance at $m_0(T)$ unless the marginal value of compliance to the firm exceeds the above marginal increase in agency rent, i.e., unless

$$[p\rho(\pi+\alpha_+)-\pi]T > \frac{4}{\frac{1}{\gamma}+\frac{p\rho}{1-p\rho}}\kappa. \tag{26}$$

If $T > T^*$, the LHS is strictly positive, strictly increasing in T, tends to 0 as T tends to T^* and to $+\infty$ as T tends to $+\infty$. The RHS is also strictly positive but strictly decreasing in T (through p). Taken together, this implies that there exists a unique $T^{**} > T^*$ such that (i) for any $T \in (T^*, T^{**})$, the firm implements $m(T) = m_0(T)$, which is (weakly) decreasing (ii) for any $T > T^{**}$ the firm implements $m(T) < m_0(T)$.

More precisely, for $T > T^{**}$, the optimal compliance rate is given by the larger of 0 and the unique solution to the first-order condition

$$[p\rho(\pi+\alpha_+)-\pi]T = \frac{\frac{1}{\gamma}+\frac{p\rho}{1-p\rho}}{\left(\frac{p\rho}{1-p\rho}+m\right)^2}\kappa, \tag{27}$$

which identifies a maximum because $\Omega''(m) = \frac{\partial^2 r(m,p)}{\partial m^2} < 0$.

To see that the solution for m identified by the first-order condition decreases in T, define the LHS as A, the RHS as $-B$, and then

$$g(m,T) \equiv A + B = 0.$$

It is easy to verify that $\frac{\partial A}{\partial T} > 0$ (taking into account $p = 1 - \frac{\iota}{T}$). Next define $x \equiv \frac{p\rho}{1-p\rho}$ and note that $\frac{\partial B}{\partial x} = \frac{\frac{2}{\gamma}+x-m}{(x+m)^3} > 0$ since $m < 1 < \frac{2}{\gamma}$ and $x > 0$. This implies $\frac{\partial B}{\partial T} > 0$ since $\frac{\partial x}{\partial T} = \frac{\partial x}{\partial p}\frac{\partial p}{\partial T} > 0$. In sum, we thus have $\frac{\partial g}{\partial T} > 0$. Moreover, it is easy to verify that $\frac{\partial g}{\partial m} > 0$. By the implicit function theorem, we hence have

$$\operatorname{sgn}\left(\frac{\partial m}{\partial T}\right) = -\frac{\operatorname{sgn}\left(\frac{\partial g}{\partial T}\right)}{\operatorname{sgn}\left(\frac{\partial g}{\partial m}\right)} < 0.$$

Finally, as $T \to +\infty$, the LHS of (27) tends to $+\infty$. It follows that the solution m to (27) tends to $-\frac{\rho}{1-\rho} < 0$ as $T \to +\infty$, which, in turn, implies the existence of T^{***} such that the principal implements $m(T) = 0$ iff $T > T^{***}$. □

The lifetime of a trading opportunity given an aggregate level of manipulation m is

$$T(m) \equiv \frac{\lambda i + \rho(1-m)\iota}{m + (1-m)\rho} \tag{28}$$

Keeping notation consistent with the previous section, the shortest possible deadline is $T_h = T(1) = \lambda i$ and the longest possible deadline is $T_d = T(0) = \frac{\lambda i}{\rho} + \iota$.

Finally, the level of manipulation that generates a deadline T (the inverse of $T(m)$) is:

$$\hat{m}(T) \equiv \frac{\lambda i - \rho(T-\iota)}{T - \rho(T-\iota)}. \tag{29}$$

Proof of Proposition 6

When $\iota = 0$, the best response function simplifies to

- $m(T) = m_0$ if $T \leq T^{**}$, where m_0 is independent from T,
- $m(T)$ continuously decreases from m_0 to 0 when T goes from T^{**} to T^{***},
- $m(T) = 0$ if $T > T^{***}$.

An equilibrium is an intersection of $m(T)$ with $\hat{m}(T)$ (i.e., the level of manipulation that generates a deadline T defined in (29)). We first show that *there exists a unique equilibrium* (if $\iota = 0$). Note first that

$$\hat{m}'(T) = -\frac{\frac{\rho}{1-\rho} + \hat{m}(T)}{T}.$$

In addition, using (27), if $T \in (T^{**}, T^{***})$,

$$m'(T) = \begin{cases} 0 & \text{if } T < T^{**}, \\ -\frac{\frac{\rho}{1-\rho} + m(T)}{2T} & \text{if } T^{**} < T < T^{***} \\ 0 & \text{if } T > T^{***}, \end{cases}$$

Hence, if $m(T)$ and $\hat{m}(T)$ cross for some T (and hence, $m(T) = \hat{m}(T)$), then $\hat{m}'(T) < m'(T)$ so that $\hat{m}(T)$ crosses $m(T)$ from above. This, together with the continuity of $m(T)$ and $\hat{m}(T)$, implies that there can be at most one intersection, hence one equilibrium. Finally, $\hat{m}(T_h) = 1 \geq m(T_h)$, $\hat{m}(T_d) = 0 \leq m(T_d)$ and continuity imply that there is at least one equilibrium.

From (26) and (27), T^{**} and T^{***} are continuous and increasing functions of π, and so is $m(T)$. Since $\hat{m}(T)$ or $T(m)$ are independent from π, it follows that the equilibrium level of manipulation $m^*(\pi)$ is increasing in π. Furthermore, there exists $\underline{\pi}'$ such that if $\pi > \underline{\pi}'$, $m > 0$, and $\overline{\pi}' > \underline{\pi}''$ such that if $\pi < \overline{\pi}'$, $m < m_0$. Every firm's net revenue is

$$T(m)[m\pi + (1-m)\rho(\pi + \alpha_+)] - r(m).$$

The derivative with respect to m is

$$-T(m)[\rho(\pi + \alpha_+) - \pi] - r'(m) + T'(m)[m\pi + (1-m)\rho(\pi + \alpha_+)]$$

if $m^* \in (0, m_0)$, every trader's best response is set by the first-order condition (27), which implies that the sum of the first two terms is negative for any $m > m^*$ and 0 for $m = m^*$. Furthermore, $T'(m) < 0$ implies that the level of risk management that maximizes firms' net profit is strictly lower than m^*.

Using similar arguments, one shows that if $\pi < \underline{\pi}'$, $m^* = 0$ and Pareto-dominates $m > 0$, and if $\pi > \overline{\pi}'$, $m^* = m_0$ is either Pareto-dominated by $m < m_0$ or Pareto-optimal.

Proof of Proposition 7

Let $T^{***} > T^{**} > T^*$ be defined as in Lemma 4. Let $\mathcal{S}_m$ and $\mathcal{S}_m^{ag}$ denote the set of m that can be supported in equilibrium, respectively, in the setting with and without agency.

Note that

$$\min \mathcal{S}_m \begin{cases} = 0 & \text{if } T_d \geq T^* \\ = 1 & \text{if } T_d < T^* \end{cases}$$

and

$$\min \mathcal{S}_m^{ag} \begin{cases} = 0 & \text{if } T_d \geq T^{***} \\ > 0 & \text{if } T^* \leq T_d < T^{***} \\ = 1 & \text{if } T_d < T^*. \end{cases}$$

The result regarding the minima of the equilibrium sets follows from simple inspection.

Similarly, note that

$$\max \mathcal{S}_m \begin{cases} = 0 & \text{if } T_h > T^* \\ = 1 & \text{if } T_h \leq T^* \end{cases}$$

and

$$\max \mathcal{S}_m^{ag} \begin{cases} \geq 0 & \text{if } T_h > T^* \\ = 1 & \text{if } T_h \leq T^*. \end{cases}$$

The weak inequality result regarding the maxima of the equilibrium sets follows from simple inspection. The inequality is strict, for example, for $T(m)$ such that $T(0) < T^{***}$ and $T(1) > T^*$. □

Proof of Proposition 9

First, if several equilibria coexist, the one with highest compliance (lowest m) Pareto-dominates the other ones. Indeed, suppose m_1^* and $m_2^* > m_1^*$ are equilibria. m_1^* being a best response to $T(m_1^*)$ implies

$$T(m_1^*)[m_1^*\pi + (1 - m_1^*)\rho(\pi + \alpha_+)] - r[m_1^*, T(m_1^*)] > T(m_1^*)[m_2^*\pi + (1 - m_2^*)\rho(\pi + \alpha_+)] - r[m_2^*, T(m_1^*)].$$

$m_2^* > m_1^*$ implies $T(m_2^*) < T(m_1^*)$ and $r[m_2^*, T(m_2^*)] \geq r[m_2^*, T(m_1^*)]$, which, in turn implies

$$T(m_1^*)[m_1^*\pi + (1 - m_1^*)\rho(\pi + \alpha_+)] - r[m_1^*, T(m_1^*)] > T(m_2^*)[m_2^*\pi + (1 - m_2^*)\rho(\pi + \alpha_+)] - r[m_2^*, T(m_2^*)].$$

That is, firms' net profits are higher in the equilibrium with m_1^* than in the one with m_2^*. In the rest of the proof, we let m^* denote the *smallest equilibrium manipulation.*

Second, since $\hat{m}(T_h) = 1 \geq m(T_h)$, $\hat{m}(T_d) = 0 \leq m(T_d)$, and $m(T)$ is downward slopping (see Lemma 4), $\hat{m}'[T(m^*)] < m'[T(m^*)]$, that is, $\hat{m}(T)$ crosses $m(T)$ from above at m^*. Note also that as i decreases (competition intensifies), the deadline $T(m)$ shrinks (see (29)), that is, $\frac{\partial T(m)}{\partial i} > 0$. Now $m[T(m^*)] = m^*$ implies

$$\frac{\partial m^*}{\partial i} = m'[T(m^*)]\frac{\partial T(m)}{\partial i}/[1 - m'[T(m^*)]T'(m^*)] = m'[T(m^*)]\frac{\partial T(m)}{\partial i}/\left[1 - \frac{m'[T(m^*)]}{\hat{m}'[T(m^*)]}\right] < 0.$$

That is, the equilibrium level of manipulation increases as competition intensifies (i decreases).

Third, for i sufficiently large, $T(0) = T_d > T^{***}$ and hence, $m^*(i) = 0$. In that case, the trader's equilibrium rent is

$$r[m^*(i)] = r(0) = \frac{2m_0}{\frac{p\rho}{1-p\rho}} = \frac{\frac{1}{\gamma}}{\frac{p\rho}{1-p\rho}} - 1,$$

which is decreasing in p, and hence, in $T[m^*(i)] = T(0) = T_d$. Since T_d is increasing in i, when i decreases, the trader's rent, $r[m^*(i)]$, increases.

As i decreases further $m^*(i)$ becomes strictly positive but strictly lower than m_0 (i.e., $T[m^*(i)] \in (T^{**}, T^{***})$). Then, $m^*(i)$ is pinned down by the first-order condition (27), which, after rearranging, yields

$$m^*(i) = \sqrt{\frac{\frac{1}{\gamma} + \frac{p\rho}{1-p\rho}}{[p\rho(\pi + \alpha_+) - \pi]T}\kappa} - \frac{p\rho}{1-p\rho}.$$

The trader's equilibrium rent is,

$$r[m^*(i)] = \frac{2[m_0 - m^*(i)]}{\frac{p\rho}{1-p\rho} - m^*(i)} = \frac{\frac{1}{\gamma} - m^*(i)}{\frac{p\rho}{1-p\rho} + m^*(i)} - 1$$

Plugging $m^*(i)$ into $r[m^*(i)]$ yields

$$r[m^*(i)] = \sqrt{\left(\frac{1}{\gamma} + \frac{p\rho}{1-p\rho}\right)\frac{[p\rho(\pi + \alpha_+) - \pi]T}{\kappa}} - 2$$

It follows that $r[m^*(i)]$ is increasing in T, both directly and through p. In equilibrium, when i decreases, T decreases both because $T(m)$ directly decreases in i, keeping m fixed, and because a decrease in i causes $m^*(i)$ to increase, and $T(m)$ is decreasing in m. Hence, when i decreases, the trader's rent, $r[m^*(i)]$ decreases.

The same reasoning applies to the S-I-ratio with simpler computations as its expression is simpler than the one for the rent. The derivation of the S-I-ratio is as follows. Let F denote the fixed component of trader's wage. The incentive compatibility constraints respectively for search and compliance become

$$(1 - \gamma m)\left\{\overline{W}[\Pi(\lambda, m)] + F\right\} \geq F + \kappa D\lambda,$$

$$\gamma\left\{\overline{W}[\Pi(\lambda, m)] + F\right\} \geq \frac{(1 - p\rho)}{m + (1-m)p\rho}\kappa D\lambda.$$

1. If $m \geq m_0$, the incentive compatibility constraint for search is binding when $F = 0$, which is therefore the highest possible fixed wage in an optimal contract.
2. If $m < m_0$, the incentive compatibility constraint for search is not binding when $F = 0$. The highest fixed wage in an optimal contract is such that both incentive compatibility constraints (for search and compliance) are binding, which yields

$$\frac{F}{F + \overline{W}[\Pi(\lambda, m)]} = 2\gamma(m_0 - m^*).$$

References

Acharya, V., and P. Volpin, "Corporate Governance Externalities," *Review of Finance* 14 (2010), 1–33.

Acharya, V., M. Pagano, and P. Volpin, "Seeking Alpha: Excess Risk Taking and Competition for Managerial Talent," 2013, NBER Working Paper No. 18891.

Aguerrevere, F., "Real Options, Product Market Competition, and Asset Returns," *Journal of Finance* 64 (2009), 957–83.

Allen, F., and D. Gale, "Optimal Financial Crises," *Journal of Finance* 53 (1998), 1245–1284.

Askenazy, P., D. Thesmar and M. Thoenig "On the Relation Between Organisational Practices and New Technologies: the Role of (Time Based) Competition," *Economic Journal*, 115 (2006), 128–154.

Barba, R., "Bank M&A Fees Have Skyrocketed in the Past Decade," *American Banker*, February 15, 2013.

Benabou, R., and J. Tirole "Bonus Culture: Competitive Pay, Screening, and Multitasking," *Journal of Political Economy*, forthcoming.

Bennhold, K. and N. Clark, "Report Says French Bank Trader's Managers Were Negligent," *New York Times*, August 2, 2008.

Biais, B., T. Foucault and S. Moinas "Equilibrium Fast Trading," *Journal of Financial Economics*, forthcoming.

Bond, P., A. Edmans and I. Goldstein "The Real Effects of Financial Markets," *Annual Review of Financial Economics*, 4 (2012), 339-360.

Boyd, J., and G. De Nicole "The Theory of Bank Risk Taking and Competition Revisited," *Journal of Finance*, 60 (2005), 1329–1343.

Bryant, J., "A Model of Reserves, Bank Runs and Deposit Insurance," *Journal of Banking and Finance*, 4 (1980), 335–344.

Budish, E., P. Cramton and J. Shim "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," 2013, Working Paper.

Burkart, M., D. Gromb and F. Panunzi "Large Shareholders, Monitoring, and the Value of the Firm," *Quarterly Journal of Economics*, 112 (1997), 693–728. (2005), 1329–1343.

Caballero, R., and R. Pindyck, "Uncertainty, Investment, and Industry Evolution," *International Economic Review* 37 (1996), 641–62.

Carlsson, H., and E. van Damme "Global Games and Equilibrium Selection," *Econometrica*, 61 (1993), 989–1018.

Chari, V., and R. Jagannathan, "Banking Panics, Information, and Rational Expectations Equilibrium," *Journal of Finance*, 43 (1988), 749–761.

Chen, Y., "Banking Panics: The Role of the First-Come, First-Served Rule and Information Externalities," *Journal of Political Economy*, 107 (1988), 946–968.

Cheng, I., "Corporate Governance Spillovers," 2011, Working Paper, `http://ssrn.com/abstract=1299652`.

Clark, C., "Controlling Risk in a Lightning-Speed Trading Environment," *Chicago Fed Letter*, 272 (2010).

Clark, C., "How to Keep Markets Safe in the Era of High-Speed Trading," *Chicago Fed Letter*, 303 (2012).

Clark, C., and R. Ranjan "How Do Proprietary Trading Firms Control the Risks of High Speed Trading?" Working Paper, Federal Reserve Bank of Chicago.

Dayen, D., "The 'London Whale's Case Lands Another Minnow: Criminal charges against JP Morgan traders let their bosses off the hook," *New Republic*, August 15, 2013.

Dell'Ariccia, G. and R. Marquez, "Information and bank credit allocation," *Journal of Financial Economics*, 72 (2004), 185–214.

Deloitte, *Global risk management survey, 8th edition: Setting a higher bar* (2013).

Diamond, D., "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51 (1984), 393–414.

Dicks, D., "Executive Compensation and the Role for Corporate Governance Regulation," *Review of Financial Studies*, 25 (2012), 1971–2004.

Diamond, D., and P. Dybvig, “Bank Runs, Deposit Insurance, and Liquidity,” *Journal of Political Economy*, 91 (1983), 401–419.

Dugast, J., and T. Foucault, “False News, Informational Efficiency, and Price Reversals,” 2014, Working Paper, http://ssrn.com/abstract=2398904.

Ernst & Young Global Banking and Capital Markets Center, “Remaking financial services: risk management five years after the crisis. A survey of major Financial Institutions,” 2013.

Froot, K., D. Scharfstein and J. Stein “Risk Management: Coordinating Corporate Investment and Financing Policies,” *Journal of Finance*, 48 (1993), 1629–1658.

Garleanu, N., and L. Pedersen “Liquidity and Risk Management,” *American Economic Review*, 97 (2007), 193-197.

Gertner, R. H., D. S. Scharfstein, and J. Stein “Internal versus External Capital Markets,” *Quarterly Journal of Economics*, 109 (1994), 1211-1230.

Golan, L., C. A. Parlour, and U. Rajan “Competition, Managerial Slack, and Corporate Governance,” *Review of Corporate Finance Studies*, forthcoming.

Goldstein, I., and A. Pauzner “Demand–Deposit Contracts and the Probability of Bank Runs,” *Journal of Finance*, 60 (2005), 1293–1327.

Gorton, G., and A. Winton “Financial Intermediation,” *Handbook of Economics and Finance*, 1 (2003), 431–552.

Grenadier, S., “Option Exercise Games: An Application to the Equilibrium Investment Strategies of Firms,” *Review of Financial Studies*, 15 (2002), 691–721.

Gromb, D., and D. Vayanos “Equilibrium and welfare in markets with financially constrained arbitrageurs,” *Journal of Financial Economics*, 66 (2002), 361–407.

Grossman, S., and M. Miller “Liquidity and Market Structure,” *Journal of Finance*, 43 (1988), 617–633.

Heider, F., and R. Inderst “Loan Prospecting,” *Review of Financial Studies*, 25 (2012), 2381–2415.

Hellmann T., K. Murdock and J. Stieglitz, "Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough?" *American Economic Review*, 90 (2000), 147–165.

Hermalin, B. E., "The Effects of Competition on Executive Behavior," *Rand Journal of Economics*, 23 (1992), 350–365.

Holmstrom, B. and J. Tirole, "Financial Intermediation, Loanable Funds and the Real Sector," *Quarterly Journal of Economics*, 112 (1997), 663–691.

Jacklin, J. and S. Bhattacharya, "Distinguishing Panics and Information-based Bank Runs: Welfare and Policy Implications," *Journal of Political Economy*, 96 (1988), 568–592.

Keeley, M., "Deposit Insurance, Risk, and Market Power in Banking," *American Economic Review*, 80 (1990), 1183–1200.

Kirilenko, A., and A. Lo "Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents," *Journal of Economic Perspectives*, 27 (2013), 51–72.

Kondor, P., "Risk in Dynamic Arbitrage: The Price Effects of Convergence Trading," *Journal of Finance*, 64 (2009), 631–655.

Landier, A., D. Sraer and D. Thesmar "Financial Risk Management: when does independence fail?" *American Economic Review*, 99 (2013), 454–458.

Maćkowiak, B., and M. Wiederholt "Information Processing and Limited Liability," *American Economic Review*, 102 (2012), 30–34.

Leahy, J., "Investment in Competitive Equilibrium: The Optimality of Myopic Behavior," *Quarterly Journal of Economics*, 108 (1993), 1105–1133.

Martin, S., "Endogenous Firm Efficiency in a Cournot Principal-Agent Model," *Journal of Economic Theory*, 59 (1993), 445–450.

Martinez-Miera, D., and R. Repullo "Does Competition Reduce the Risk of Bank Failure?" *Review of Financial Studies*, 23 (2010), 3638–3664.

Mehta, A., M. Neukirchen, S. Pfetsch and T. Poppensieker "Managing market risk: Today and tomorrow," 2012, McKinsey Working Papers on Risk No 32.

Morris, S., and H. Shin "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review*, 88 (1998), 587–97.

——— "Global Games: Theory and Applications," *Advances in Economics and Econometrics (Proceedings of the Eighth World Congress of the Econometric Society)*, (2003).

Nalebuff, B. J., and J. Stiglitz, "Information, Competition, and Markets," *American Economic Review*, 73 (1983), 278–283.

Pagnotta, E., and T. Philippon "Competing on speed," 2013, NBER Working Paper No 17652.

Parlour, C., and U. Rajan "Competition in Loan Contracts," *American Economic Review*, 91 (2001), 1311–1328.

Peck, J., and K. Shell "Equilibrium Bank Runs," *Journal of Political Economy*, 111 (2003), 103–123.

Plantin, G. "Learning by Holding and Liquidity," *Review of Economic Studies*, 76 (2009), 395–412.

Postlewaite, A. and X. Vives, "Bank runs as an equilibrium phenomenon," *The Journal of Political Economy*, 95 (1987), 485–491.

Raff, K., "Information Externalities in Corporate Governance," 2011, Working Paper, `http://ssrn.com/abstract=1787130`.

Raith, M., "Competition, Risk, and Managerial Incentives," *American Economic Review*, 93 (2003), 1425–1436.

Rampini, A., and S. Viswanathan "Collateral, Risk Management, and the Distribution of Debt Capacity," *Journal of Finance*, 65 (2010), 2293–2322.

Ruckes, M. "Bank Competition and Credit Standards,"*Review of Financial Studies*, 17 (2004), 1073–1102.

Scharfstein, D. "Product-Market Competition and Managerial Slack,"*Rand Journal of Economics*, 19 (1988), 147–155.

Schmidt, K. M. "Managerial Incentives and Product Market Competition,"*Review of Economic Studies*, 64 (1997), 191–213.

Shleifer, A., and R. Vishny "The Limits of Arbitrage," *Journal of Finance*, 52 (1997), 35–55.

Smith, C., and R. Stulz "The Determinants of Firm's Hedging Policies," *Journal of Financial and Quantitative Analysis*, 20 (1985), 391–406.

Stulz, R. "Optimal Hedging Policies," *Journal of Financial and Quantitative Analysis*, 19 (1984), 127–140.

Stulz, R. "Risk management failures: What are they and when do they happen?," *Journal of Applied Corporate Finance*, 20 (2008), 39–48.

Tobin, J. "A Proposal for International Monetary Reform," *Eastern Economic Journal*, 4 (1978), 153–159.

Trindle, J., "A whale of a trade: one man to check the books," *Wall Street Journal*, September 19, 2013.

Zuckerman, G. and D. Fitzpatrick, "J.P. Morgan 'Whale' Was Prodded: Bank's Probe Concludes Trader's Boss Encouraged Boosting Values of Bets That Were Losing," *Wall Street Journal*, August 3, 2012.

Tables and figures

Table 1: O*Net characteristics.

O*Net characteristics	All 922 occupations: Average	Median	Std. dev.	Sales Agents, Securities & Commodities
1. Time pressure	70.6	72.0	12.4	89
2. Level of competition	51.8	52.0	14.5	93
3. Impact of decisions on co-workers or company results	71.3	71.0	12.6	95
4. Frequency of decision making	71.5	72.0	14.3	91

Note: Each characteristic is scored on a scale from 0 to 100. When the workers are surveyed, the scale is divided into five 20-point intervals that correspond to different qualitative statements:

1. Time pressure: *How often does this job require the worker to meet strict deadlines?* The qualitative statements range from *never* (< 20) to *every day* (> 80).
2. Level of competition: *To what extent does this job require the worker to compete or to be aware of competitive pressures?* The qualitative statements range from *not at all competitive* (< 20) to *extremely competitive* (> 80).
3. Impact of decisions on co-workers or company results: *How do the decisions an employee makes impact the results of co-workers, clients or the company?* The qualitative statements range from *no results* (< 20) to *very important results* (> 80).
4. Frequency of decision making: *How frequently is the worker required to make decisions that affect other people, the financial resources, and/or the image and reputation of the organization?* The qualitative statements range from *never* (< 20) to *every day* (> 80).

Source: O*Net Online, http://www.onetonline.org/

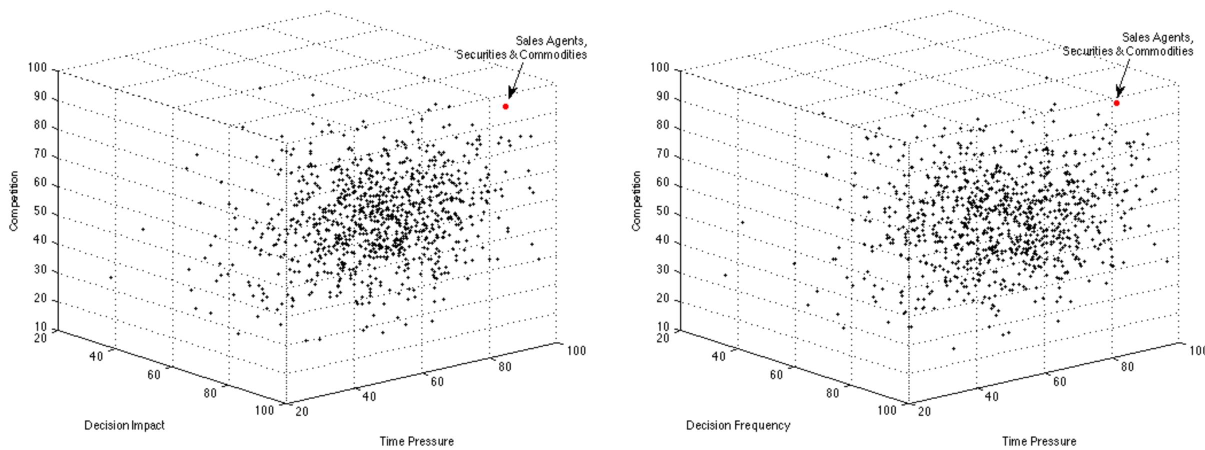

Figure 1: O*Net characteristics—scatter plots.

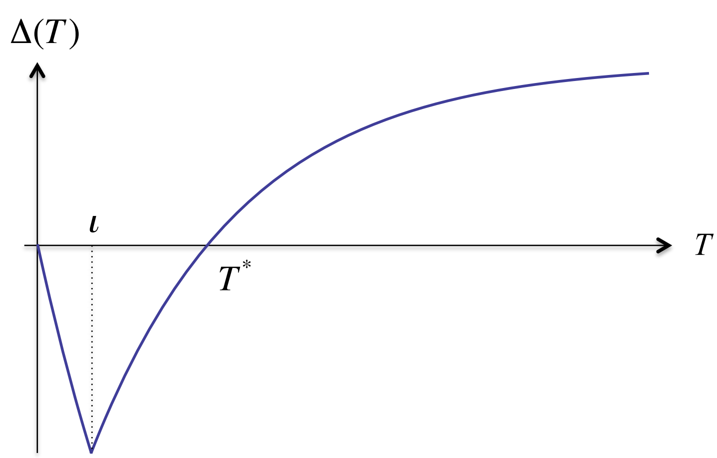

Figure 2: Time pressure and private net value of risk management.

T is an inverse measure of time pressure: The lower T, the faster do trading opportunities disappear. ι is the risk management delay. $\Delta(T)$ measures the firm's expected profit from trading with risk management relative to trading without risk management.

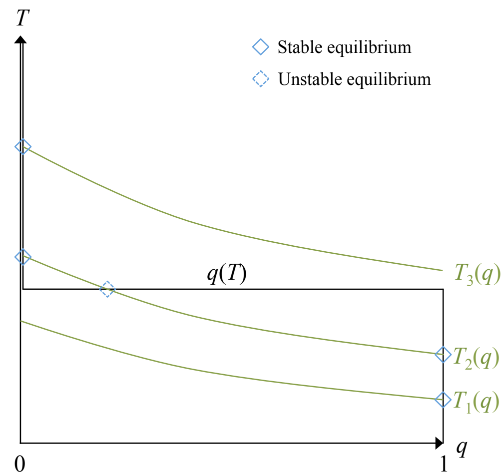

Figure 3: Strategic complementarities in risk management.

The step function $q(T)$ depicts the firms' optimal risk management choice as a function of time pressure, while $T_{\cdot}(q)$ depicts how time pressure depends on the firms' risk management choices for various levels of the common value π (i.e., the ex ante value of a trade). Lower $T_{\cdot}(q)$ correspond to higher values of π. Intersections between $q(T)$ and $T_{\cdot}(q)$ constitute equilibria.

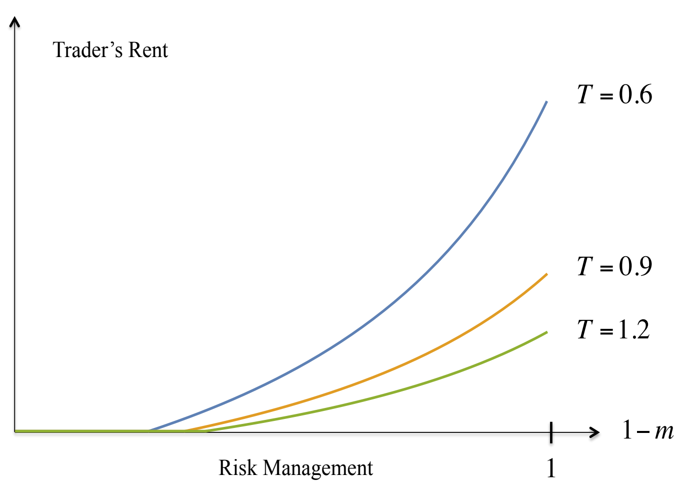

Figure 4: Traders' rents

Trader's rent per identified trading opportunity for different levels of time pressure.

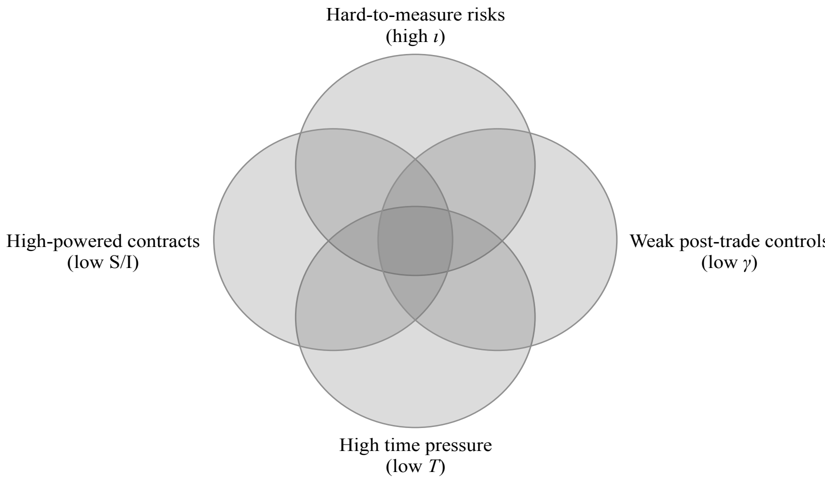

Figure 5: Problem areas.

Darker areas are where risk management is more likely (constrained) inefficient and, for example, high-powered contracts may indicate such inefficiency.

Online Appendix: Additional proofs and extensions

Proof of Lemma 1

Lemma 1 follows from the properties of $\Delta(T)$, which we formally state below:

1. If $\pi > 0$, there exists $T^*(\pi)$ such that $\Delta(T) > 0$ if $T < T^*(\pi)$ and $\Delta(T) < 0$ if $T < T^*(\pi)$.
2. If $\pi > 0$, $\Delta(.)$ is strictly decreasing on $(0, \iota)$ and strictly increasing on $(\iota, +\infty)$, and $T^*(\pi) > \iota$.
3. If $\pi \leq 0$, $\Delta(.)$ is strictly positive and strictly increasing on $(0, +\infty)$.
4. For any $T > 0$, $\Delta(T)$ is strictly decreasing in π.

Proof. As a reminder,

$$\begin{aligned} \Delta(T) &= \rho(\pi + \alpha_+)p_d(T) - \pi p_h(T) \\ &= \rho(\pi + \alpha_+) \max\left\{0, 1 - e^{-(T-\iota)/\lambda}\right\} - \pi\left(1 - e^{-T/\lambda}\right) \end{aligned} \tag{OA.1}$$

Suppose $\pi > 0$. Check that (a) $\lim_{T \to 0} \Delta(T) = 0$; (b) $\lim_{T \to +\infty} \Delta(T) = \rho(\pi + \alpha_+) - \pi = -(1-\rho)(\pi + \alpha_-) > 0$; (c) $\Delta'(T) < 0$ if $T \in (0, \iota)$ and $\Delta'(T) > 0$ if $T \in (\iota, +\infty)$. Together, these facts prove points 1 and 2.

Suppose $\pi \leq 0$. We have (a) $\lim_{T \to 0} \Delta(T) = 0$; (b) $\pi + \alpha_+ > 0$ (by assumption); (c) $p_d(T)$ is weakly increasing and $p_h(T)$ is strictly increasing. Altogether, these prove point 3.

Finally, for any T, $p_h(T) > \rho p_d(T)$, which, from (OA.1) implies $\frac{\partial \Delta}{\partial \pi} < 0$. □

It remains to show that T^* is an increasing function of π when $\pi > 0$. Using the implicit function theorem,

$$\Delta'(T^*)\frac{\partial T^*}{\partial \pi} + \frac{\partial \Delta}{\partial \pi}(T^*) = 0.$$

From point 2, $\Delta'(T^*) > 0$. Furthermore, from point 4, $\frac{\partial \Delta}{\partial \pi} < 0$. Therefore, $\frac{\partial T^*}{\partial \pi} > 0$.

Proof of Proposition 2, Step 2: Any equilibrium is a threshold equilibrium.

As mentioned in main text, our setup has the same properties (one-sided strategic complementarities) as Goldstein and Pauzner (2005). The proof follows their strategy, and we provide here a simplified demonstration that restricts attention to symmetric pure-strategy equilibria but follows the same steps as the complete proof. We refer the reader to the aforementioned paper for a proof that allows for any possible strategy.

Suppose all traders play an equilibrium strategy that maps their signal s_k into a trading behaviour that can be hasty or deliberate. Given this strategy, for each realization of π, a mass $q(\pi)$ of traders are hasty, which maps one-to-one into a deadline $\tau(\pi)$ at which the opportunity is exhausted. [The additional complexity in the proof in Goldstein and Pauzner (2005) comes from the possibility that $q(\pi)$ could be random when allowing a larger set of strategies.] Adapting notation we let

$$u[s, \tau(.)] \equiv \frac{1}{2\varepsilon} \int_{s-\varepsilon}^{s+\varepsilon} \Delta[\pi, \tau(\pi)] d\pi$$

denote the net benefit of being deliberate for a trader with a signal s in this equilibrium.

Let s_A denote the signal below which traders are always deliberate, that is

$$s_A \equiv \inf\{s : u[s, \tau(.)] \leq 0\}.$$

The existence of dominance regions guarantees the existence of s_A. Note also that $u[s, \tau(.)]$ is continuous in s, which implies $u[s_A, \tau(.)] = 0$.

Suppose that traders do not follow a threshold strategy. Then, there exists signals $s > s_A$ such that $u[s, \tau(.)] \geq 0$. Let s_B be their infimum:

$$s_B \equiv \inf\{s > s_A : u[s, \tau(.)] \geq 0\}.$$

By continuity again, $u[s_B, \tau(.)] = 0$, and therefore $u[s_A, \tau(.)] = u[s_B, \tau(.)]$, that is,

$$\frac{1}{2\varepsilon} \int_{s_A-\varepsilon}^{s_A+\varepsilon} \Delta[\pi, \tau(\pi)] d\pi = \frac{1}{2\varepsilon} \int_{s_B-\varepsilon}^{s_B+\varepsilon} \Delta[\pi, \tau(\pi)] d\pi. \tag{OA.2}$$

By definition, for any $s < s_A$, $u[s, \tau(.)] > 0$ and for any $s \in (s_A, s_B)$, $u[s, \tau(.)] < 0$. For $s > s_B$, the sign of $u[s, \tau(.)]$ is indeterminate. The proof consists in showing that (OA.2) cannot hold.

Let $\overline{s}_A \equiv \min\{s_A + \varepsilon, s_B - \varepsilon\}$ and $\underline{s}_B \equiv \max\{s_A + \varepsilon\, s_B - \varepsilon\}$. Cancelling out the (potentially empty) region $[s_B - \varepsilon, s_A + \varepsilon]$ in (OA.2), one obtains

$$\overline{u}[s_A, \tau(.)] \equiv \frac{1}{2\varepsilon} \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi, \tau(\pi)] d\pi = \frac{1}{2\varepsilon} \int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi, \tau(\pi)] d\pi \equiv \overline{u}[s_B, \tau(.)]. \tag{OA.3}$$

Note that the two integrals have the same length: $\overline{s}_A - s_A + \varepsilon = s_B + \varepsilon - \underline{s}_B \equiv d$.

Notice next that since Δ is monotonically decreasing in π, so is the function

$$v_A[\pi, \tau(.)] \equiv \frac{1}{2\varepsilon} \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi, \tau(\tilde{\pi})] d\tilde{\pi},$$

and $v_A[\overline{s}_A, \tau(.)] < u[s_A, \tau(.)] < v_A[s_A - \varepsilon, \tau(.)]$. Therefore, there exists $\pi_A \in [s_A - \varepsilon, \overline{s}_A]$, such that $v_A[\pi_A, \tau(.)] = \overline{u}[s_A, \tau(.)]$. Similarly, there exists $\pi_B \in [\underline{s}_B, s_B + \varepsilon]$, such that

$$v_B[\pi_B, \tau(.)] \equiv \frac{1}{2\varepsilon} \int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi_B, \tau(\pi)] d\pi = \overline{u}[s_B, \tau(.)].$$

Using again the strict monotonicity of Δ in π and $\pi_A < \pi_B$, we get that

$$\overline{u}[s_B, \tau(.)] = v_B[\pi_B, \tau(.)] < v_B[\pi_A, \tau(.)]. \tag{OA.4}$$

The end of the proof consists in showing that $v_B[\pi_A, \tau(.)] \leq v_A[\pi_A, \tau(.)] = \overline{u}[s_A, \tau(.)]$ which, together with (OA.4) contradicts (OA.3) and hence, (OA.2).

Let $\overleftarrow{\tau}(\pi) \equiv \tau(\overline{s}_A + s_A - \varepsilon - \pi)$. $\overleftarrow{\tau}(\pi)$ is the mirror image of $\tau(\pi)$ over $[s_A - \varepsilon, \overline{s}_A]$, that is, when π increases from $s_A - \varepsilon$ to $\overline{s}_A$, $\tau(\pi)$ follows the same path as $\overleftarrow{\tau}(\pi)$ when π decreases from $\overline{s}_A$ to $s_A - \varepsilon$. Hence,

$$v_A[\pi_A, \tau(.)] = \frac{1}{2\varepsilon}\int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi_A, \tau(\pi)]d\pi = \frac{1}{2\varepsilon}\int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi_A, \overleftarrow{\tau}(\pi)]d\pi$$

Claim 4. *$\overleftarrow{\tau}(.)$ is strictly increasing on $[s_A - \varepsilon, \overline{s}_A]$, and $\tau(.)$ is weakly increasing on $[\underline{s}_B, s_B + \varepsilon]$. Furthermore, $\overleftarrow{\tau}(.)$ increases at a faster rate on $[s_A-\varepsilon, \overline{s}_A]$ than $\tau(.)$ on $[\underline{s}_B, s_B+\varepsilon]$. That is, if $(\pi_a, \pi_b) \in [s_A-\varepsilon, \overline{s}_A] \times [\underline{s}_B, s_B+\varepsilon]$ and $\overleftarrow{\tau}(\pi_a) = \tau(\pi_b)$, then $\overleftarrow{\tau}(\pi_a + \delta) \geq \tau(\pi_b + \delta)$ for $\delta > 0$.*

Proof. Suppose that $\pi = s_A - \varepsilon$, then agents receive signals in $[s_A - 2\varepsilon, s_A]$ and therefore for all agents (except at s_A), $u[s, \tau(.)] > 0$. It follows that almost all agents are deliberate and therefore $\tau(s_A - \varepsilon) = \overleftarrow{\tau}(\overline{s}_A) = T_d$. Suppose that π increases by δ, then agents with signals in $[s_A - 2\varepsilon, s_A - 2\varepsilon + \delta)$ are replaced one for one with agents with signals in $(s_A, s_A + \delta]$. That is, one substitutes agents for whom $u[s, \tau(.)] > 0$ with agents for whom $u[s, \tau(.)] < 0$. As a result $\tau(.)$ (resp. $\overleftarrow{\tau}(.)$) decreases (resp. increases) at the fastest possible rate. Symmetrically, $\tau(.)$ increases on $[\underline{s}_B, s_B + \varepsilon]$, but at a (weakly) slower rate: there can be values of s in $[s_B, s_B + \varepsilon]$ such that $u[s, \tau(.)] < 0$ in which case, as π increases, one substitute hasty agents with other hasty agents, leaving $\tau(.)$ unchanged. □

Claim 5. *For any $\delta \in [0, d]$, $\tau(\underline{s}_B + \delta) \leq \overleftarrow{\tau}(s_A - \varepsilon + \delta)$.*

Proof. Notice that $(s_A, \overline{s}_A]$ and $[\underline{s}_B, s_B)$ have the same measure and $u[s, \tau(.)]$ is always strictly negative on these two segments. By contrast, $[s_A-\varepsilon, s_A)$ and $(s_B, s_B+\varepsilon]$ have the same measure but while $u[s, \tau(.)]$ is always strictly positive on the first segment, it can change signs on the second one. This implies $\tau(\underline{s}_B) \leq \tau(\overline{s}_A) = \overleftarrow{\tau}(s_A - \varepsilon)$. Claim 4 completes the proof. □

Note that if $\Delta(\pi, .)$ was monotonically increasing (that is, under global strategic complementarities), Claim 5 would directly imply that $v_A[\pi_A, \tau(.)] \geq v_B[\pi_A, \tau(.)]$.

Claim 6. $v_A[\pi_A, \tau(.)] \geq 0$.

Proof. Note first that the monotonicity result in Claim 4 can be extended: $\tau(.)$ is weakly decreasing on $[\overline{s}_A, s_A+\varepsilon]$. Indeed, as π increases in this interval, one substitutes deliberate traders with deliberate or hasty traders (the latter can be deliberate if $s_A+\varepsilon > s_B$). Let $\hat{\pi}_A \equiv \inf\{\pi \in [s_A-\varepsilon, s_A+\varepsilon] : \Delta[\pi_A, \tau(\pi_A)] \leq 0\}$, which is well defined since $u[s_A, \tau(.)] = 0$. Using the single-crossing property of Δ together with the monotonicity of $\tau(.)$ on $[s_A - \varepsilon, s_A + \varepsilon]$ and the fact that $\frac{\partial \Delta}{\partial \pi} < 0$, we get that for any $\pi \in (\hat{\pi}_A, s_A + \varepsilon]$, $\Delta[\pi, \tau(\pi)] < 0$ and for any $\pi \in [s_A - \varepsilon, \hat{\pi}_A)$, $\Delta[\pi, \tau(\pi)] > 0$. In words the integrand in $u[s_A, \tau(.)]$ is positive below a threshold and negative above it, which together with the fact that

$$u[s_A, \tau(.)] = \int_{s_A-\varepsilon}^{s_A+\varepsilon} \Delta[\pi, \tau(\pi)]d\pi = 0$$

implies that

$$\overline{u}[s_A, \tau(.)] = \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi, \tau(\pi)]d\pi \geq 0,$$

which is equivalent to $v_A[\pi_A, \tau(.)] \geq 0$. □

We want to show

$$v_A[\pi_A, \tau(.)] - v_B[\pi_A, \tau(.)] = \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi_A, \overleftarrow{\tau}(\pi)]d\pi - \int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi_A, \tau(\pi)]d\pi \geq 0 \tag{OA.5}$$

Suppose that $\tau(s_B + \varepsilon) < T^*(\pi_A)$. Then, since $\tau(.)$ is weakly increasing on $[\underline{s}_B, s_B + \varepsilon]$, for any $\pi \in [\underline{s}_B, s_B + \varepsilon]$, $\tau(\pi) < T^*(\pi_A)$ and therefore $\Delta[\pi_A, \tau(\pi)] < 0$. Hence, $v_B[\pi_A, \tau(.)] < 0$, and since, from Claim 6, $v_A[\pi_A, \tau(.)] \geq 0$, (OA.5) holds.

Suppose that $\tau(\underline{s}_B) \geq T^*(\pi_A)$.[OA1] Then $\Delta(\pi_A, .)$ strictly increasing for $T > T^*(\pi_A)$ and Claim 5 imply that (OA.5) holds.

Finally, suppose that $\tau(\underline{s}_B) < T^*(\pi_A) \leq \tau(s_B + \varepsilon)$. Let $\overleftarrow{\pi}(T) \equiv \overleftarrow{\tau}^{-1}(T)$ and

$$\pi(T) \equiv \begin{cases} \underline{s}_B & \text{if } T = \tau(\underline{s}_B) \\ \max\{\pi \in (\underline{s}_B, s_B + \varepsilon] : \tau(\pi) = T\} & \text{if } \tau(\underline{s}_B) < T \leq \tau(s_B + \varepsilon) \end{cases}$$

In words, $\overleftarrow{\pi}(.)$ and $\pi(.)$ are inverse functions of $\overleftarrow{\tau}(.)$ and $\tau(.)$. Therefore from Claim 4, $\overleftarrow{\pi}(.)$ and $\pi(.)$ are increasing and $\pi(.)$ increases faster than $\overleftarrow{\pi}(.)$. Using this notation and Claim 5, rewrite (OA.5):

$$\begin{aligned} & v_A[\pi_A, \tau(.)] - v_B[\pi_A, \tau(.)] = \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi_A, \overleftarrow{\tau}(\pi)]d\pi - \int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi_A, \tau(\pi)]d\pi \\ = \quad & -\int_{\tau(\underline{s}_B)}^{\tau(s_A-\varepsilon)} \Delta[\pi_A, T]d\pi(T) && \text{(OA.6)} \\ & + \int_{\tau(s_A-\varepsilon)}^{T^*(\pi_A)} \Delta[\pi_A, T]d\overleftarrow{\pi}(T) - \int_{\tau(s_A-\varepsilon)}^{T^*(\pi_A)} \Delta[\pi_A, T]d\pi(T) && \text{(OA.7)} \\ & + \int_{\overleftarrow{\pi}[T^*(\pi_A)]}^{\overline{s}_A} \Delta[\pi_A, \overleftarrow{\tau}(\pi)]d\pi - \int_{\pi[T^*(\pi_A)]}^{s_B+\varepsilon} \Delta[\pi_A, \tau(\pi)]d\pi. && \text{(OA.8)} \end{aligned}$$

This equation is a decomposition of the two integrals in (OA.5) along the interval $[\tau(\underline{s}_B), \tau(\overline{s}_A)]$. Note that this decomposition assumes $\tau(s_A - \varepsilon) \leq T^*(\pi_A)$, the demonstration would be a fortiori true if $\tau(s_A - \varepsilon) > T^*(\pi_A)$.

At the bottom, (OA.6) corresponds to the part of the integral in v_B with values of T below the lowest T in v_A, that is, $\tau(s_A - \varepsilon)$. These values are below $T^*(\pi_A)$, therefore this part of v_B is negative and (OA.6) is strictly positive. In the interval $[\tau(s_A - \varepsilon), T^*(\pi_A))$ integrands in both v_A and v_B are negative as T is still below $T^*(\pi_A)$. However, $d\overleftarrow{\pi}(T) \leq d\pi(T)$, implies that (OA.7) is positive (Intuitively, v_B visits any negative values of Δ that v_A takes but "stays longer" at each of them.) Finally, (OA.8) corresponds to values of T above T^*. Note first that $\tau(s_A-\varepsilon) \geq \tau(\underline{s}_B)$ (Claim 5) and $\overleftarrow{\pi}(.)$ increasing more slowly than $\pi(.)$ imply $\overleftarrow{\pi}[T^*(\pi_A)] - s_A + \varepsilon \leq \pi[T^*(\pi_A)] - \underline{s}_B$, and therefore, $\overline{s}_A - \overleftarrow{\pi}[T^*(\pi_A)] \geq s_B - \varepsilon - \pi[T^*(\pi_A)]$. (i.e., the LHS integral takes a larger range of values of π than the RHS integral.) In addition, from Claim 4, for $\delta > 0$, $\overleftarrow{\tau}\{\overleftarrow{\pi}[T^*(\pi_A)] + \delta\} \geq \tau\{\pi[T^*(\pi_A)] + \delta\} \geq 0$, and $\Delta(\pi_A, .)$ is strictly increasing for $T \geq 0$, (Lemma 1). Hence, (OA.8) is positive (intuitively, $\overleftarrow{\tau}(.)$ shifts more weight towards high values of Δ than $\tau(.)$.)

[OA1] Note that we implicitly assume $\pi_A > 0$. If $\pi_A \leq 0$, then $\Delta(\pi_A, .)$ is monotonically increasing and the result is immediate.

This eventually shows $v_B[\pi_A, \tau(.)] \leq v_A[\pi_A, \tau(.)] = \overline{u}[s_A, \tau(.)]$, which together with $\overline{u}[s_B, \tau(.)] = v_B[\pi_B, \tau(.)] < v_B[\pi_A, \tau(.)]$ shows $\overline{u}[s_B, \tau(.)] < \overline{u}[s_A, \tau(.)]$, a contradiction. $\square$

Proof of Lemma 2

We showed in the main text that the firm's total profit only depends on average search e and average manipulation m. To complete the proof, we show now that for any contract contingent on specific profit paths $\{\Pi_t\}_{t\in[0,D]}$, there exists an *equivalent* contract that is contingent on average profit $\hat{\Pi}$ only: Under both contracts, the trader's *average* search and manipulation, and payments are the same in equilibrium.

The gist of the proof relies on the following argument: if two profit paths have the same average profit, then any *average* search and manipulation that the trader can produce while generating the first path, can also be produced while generating the second path.

Assume contracting takes place at $t = 0$. A generic contract is a pair of functions

$$W = \{\overline{W}(\{\Pi_t\}_{t\in[0,D]}), \underline{W}(\{\Pi_t\}_{t\in[0,D]})\}$$

that specify the trader's payment as a function of the entire profit path $\{\Pi_t\}_{t\in[0,D]}$, when, respectively, no manipulation is detected ex post, and manipulation is detected. As in the main text, average effort and average manipulation are defined as

$$e(\{e_t\}_t) = D^{-1}\int_0^D e_t dt \qquad \text{and} \qquad m(\{e_t, m_t\}_t) = \frac{\int_0^D m_t e_t dt}{\int_0^D e_t dt}.$$

Average profit is

$$\hat{\Pi}(\{\Pi_t\}_t) = D^{-1}\int_0^D \Pi_t dt.$$

Consider an arbitrary contract W. The trader's expected utility, given the contract W and a strategy $\{e_t, m_t\}_t$ is

$$U(\{e_t, m_t\}_t, W) = [1 - \gamma m(\{e_t, m_t\}_t)]\overline{W}(\{\Pi(e_t, m_t\}_t) + \gamma m(\{e_t, m_t\}_t)\underline{W}(\{\Pi(e_t, m_t\}_t) - \kappa D e(\{e_t\}_t)$$

Suppose there exist two profit paths, $\{\Pi_t^a\}_t$ and $\{\Pi_t^b\}_t$ that generate *the same average profit*, $\hat{\Pi}(\{\Pi_t^a\}_t) = \hat{\Pi}(\{\Pi_t^b\}_t)$, but *different payments*, $\overline{W}(\{\Pi_t^a\}_t) \neq \overline{W}(\{\Pi_t^b\}_t)$ or $\underline{W}(\{\Pi_t^a\}_t) \neq \underline{W}(\{\Pi_t^b\}_t)$. Let $\{e_t^a, m_t^a\}_t$ and $\{e_t^b, m_t^b\}_t$ maximize the trader's utility under the constraint of generating, respectively, $\{\Pi_t^a\}_t$ and $\{\Pi_t^b\}_t$ i.e.,

$$\{e_t^a, m_t^a\}_t = \underset{\{\tilde{e}_t, \tilde{m}_t\}_t \,:\, \forall t, \Pi(\tilde{e}_t, \tilde{m}_t) = \Pi_t^a}{\text{argmax}} U(\tilde{e}_t, \tilde{m}_t, W)$$

and

$$\{e_t^b, m_t^b\}_t = \underset{\{\tilde{e}_t, \tilde{m}_t\}_t \,:\, \forall t, \Pi(\tilde{e}_t, \tilde{m}_t) = \Pi_t^b}{\text{argmax}} U(\{\tilde{e}_t, \tilde{m}_t\}_t, W).$$

Suppose $\{e_t^*, m_t^*\}_t$ is an optimal strategy under W, and without loss of generality, suppose $\{e_t^*, m_t^*\}_t \neq \{e_t^b, m_t^b\}_t$. Consider the alternative contract W' *that only differ from* W *for the specific profit path* $\{\Pi_t^b\}_t$. Specifically, for any $\{\Pi_t\}_t \neq \{\Pi_t^b\}_t$, $\overline{W}'(\{\Pi_t\}_t) = \overline{W}(\{\Pi_t\}_t)$ and $\underline{W}'(\{\Pi_t\}_t) = \underline{W}(\{\Pi_t\}_t)$, and

$$\overline{W}'(\{\Pi_t^b\}_t) = \overline{W}(\{\Pi_t^a\}_t) \text{ and } \underline{W}'(\{\Pi_t^b\}_t) = \underline{W}(\{\Pi_t^a\}_t).$$

We claim that $\{e_t^*, m_t^*\}_t$ is an optimal strategy under W'.

Notice first that $\{e_t^b, m_t^b\}_t \neq \{e_t^*, m_t^*\}_t$ implies

$$U(\{e_t^*, m_t^*\}_t, W) = U(\{e_t^*, m_t^*\}_t, W').$$

Next, suppose $\{e_t^*, m_t^*\}_t$ is *not* optimal under W', Since W and W' only differ for $\{\Pi_t^b\}_t$, there exists $\{e_t', m_t'\}_t$ such that

$$U(\{e_t', m_t'\}_t, W') > U(\{e_t^*, m_t^*\}_t, W') \text{ and for any } t,\ \Pi(e_t', m_t') = \Pi_t^b.$$

Consider the strategy

$$e_t^{**} = e(\{e_t'\})\frac{\Pi_t^a}{\hat{\Pi}(\{\Pi_t^b\}_t)} \text{ and } m_t^{**} = m(\{e_t', m_t'\}) = \Pi_t^a$$

By construction (and using (11))

$$\Pi_t(e_t^{**}, m_t^{**}) = \frac{e(\{e_t'\})[m(\{e_t', m_t'\}) + (1 - m(\{e_t', m_t'\}))p\rho]\pi}{\hat{\Pi}(\{\Pi_t^b\}_t)}\Pi_t^a = \Pi_t^a,$$

and

$$e(\{e_t^{**}\}_t) = e(\{e_t'\})\frac{\hat{\Pi}(\{\Pi_t^a\}_t)}{\hat{\Pi}(\{\Pi_t^b\}_t)} = e(\{e_t'\})$$

It follows that $\{e_t^{**}, m_t^{**}\}_t$ generates profit path $\{\Pi_t^a\}_t$ and the same average search and manipulation as $\{e_t', m_t'\}_t$. Hence,

$$U(\{e_t^*, m_t^*\}_t, W) = U(\{e_t^*, m_t^*\}_t, W') < U(\{e_t', m_t'\}_t, W') = U(\{e_t^{**}, m_t^{**}\}_t, W') = U(\{e_t^{**}, m_t^{**}\}_t, W)$$

Hence, $\{e_t^*, m_t^*\}_t$ is *not* optimal under W, a contradiction.

It follows that W and W' are equivalent: they generate the same optimal strategy and the same payments to the trader. By iterating the process, we can construct a contract W'', equivalent to W, such that if any two sequences $\{\Pi_t^a\}_t$ and $\{\Pi_t^b\}_t$ generate the same average profit, $\hat{\Pi}(\{\Pi_t^a\}_t) = \hat{\Pi}(\{\Pi_t^b\}_t)$, then they generate the same payments: $\overline{W}(\{\Pi_t^a\}_t) = \overline{W}(\{\Pi_t^b\}_t)$ and $\underline{W}(\{\Pi_t^a\}_t) = \underline{W}(\{\Pi_t^b\}_t)$. W'' is then by definition conditional on average profit $\hat{\Pi}$ only. This, in turn implies the existence of an optimal strategy for the trader that is time-independent. □

Proof of Lemma 3

Consider a contract $\{\overline{W}(.), \underline{W}(.)\}$. The trader chooses a search effort e and a manipulation level m to maximize

$$(1 - \gamma m)\overline{W}[\Pi(e, m)] + \gamma m\underline{W}[\Pi(e, m)] - \kappa e D.$$

Let e^* and m^* be search effort and manipulation levels that jointly maximize the above function, and consider an alternative contract $\{\overline{W}'(.), \underline{W}'(.)\}$ defined as follows:

$$\overline{W}'(\Pi) \equiv \begin{cases} \overline{W}(\Pi) & \text{if } \Pi = \Pi(e^*, m^*), \\ 0 & \text{otherwise.} \end{cases} \tag{OA.9}$$

and

$$\underline{W}'(\Pi) \equiv \begin{cases} \underline{W}(\Pi) & \text{if } \Pi = \Pi(e^*, m^*), \\ 0 & \text{otherwise.} \end{cases} \tag{OA.10}$$

Since conditional on delivering $\overline{\Pi} = \Pi(e^*, m^*)$, the agent receives the same payoff in the new contract as in the original contract, he must weekly prefer (e^*, m^*) to any other couple (e, m) such that $\Pi(e, m) = \Pi(e^*, m^*)$. Note also that any couple (e, m) that delivers a strictly positive profit $\Pi(e, m)$ different from $\Pi(e^*, m^*)$ is strictly dominated as it requires a strictly positive amount of costly effort and generates no payment. Finally, since the (e^*, m^*) was weekly preferred to exerting no effort under the original contract,

$$(1 - \gamma m^*)\overline{W}[\Pi(e^*, m^*)] + \gamma m^* \underline{W}[\Pi(e^*, m^*)] - \kappa e^* D \geq 0,$$

which implies

$$(1 - \gamma m^*)\overline{W}'[\Pi(e^*, m^*)] + \gamma m^* \underline{W}'[\Pi(e^*, m^*)] - \kappa e^* D \geq 0,$$

and hence, (e^*, m^*) weakly dominates no effort. As the two contracts elicit the same effort and manipulation levels for the same equilibrium payment, they are equivalent from the point of view of the firm, and we therefore restrict attention to contracts that have the form described in (OA.9) and (OA.10).

The level of effort level necessary to reach profit $\Pi(e^*, m^*)$ for a given manipulation m is

$$e[m, \Pi(e^*, m^*)] \equiv \frac{\Pi(e^*, m^*)}{eT[m + (1 - m)p\rho]\pi}$$

Note that $e[., \Pi(e^*, m^*)]$ is a strictly decreasing and convex function. It follows that necessary and sufficient conditions for a contract to incentivize effort e^* and manipulation m^* are

$$(1 - \gamma m^*)\overline{W}[\Pi(e^*, m^*)] + \gamma m^* \underline{W}[\Pi(e^*, m^*)] \geq \kappa D e^*, \tag{OA.11}$$

$$\gamma \left\{ \overline{W}[\Pi(e^*, m^*)] - \underline{W}[\Pi(e^*, m^*)] \right\} \geq -\kappa D \frac{\delta e}{\delta m}[m^*, \Pi(e^*, m^*)]. \tag{OA.12}$$

where (OA.12) holds with equality if $m^* \in (0, 1)$. (OA.11) states that the trader has to be compensated at minimum for his cost of effort. (OA.12) states that within the set of effort-manipulation couples (e, m) that generate the same profit $\Pi(e^*, m^*)$, (e^*, m^*) has to be optimal.

We start by deriving levels of compliance that can be achieved "for free," that is, without leaving a rent to the trader. If the trader has no rent then (OA.11) must hold with equality. Note that this leaves us with one degree freedom to set $\overline{W}[\Pi(e^*, m^*)]$ relative to $\underline{W}[\Pi(e^*, m^*)]$. We use (OA.11) to make substitutions in (OA.12), we also use $\frac{\delta e}{\delta m}[m^*, \Pi(e^*, m^*)] = -\frac{e^*(1-p\rho)}{m^*+(1-m^*)p\rho}$, and obtain, after rearranging,

$$\left[\frac{\gamma}{1-\gamma m^*} - \frac{1-p\rho}{m^*+(1-m^*)p\rho}\right]\kappa D e^* - \frac{\gamma}{1-\gamma m^*}\underline{W}[\Pi(e^*, m^*)] \geq 0. \tag{OA.13}$$

Maximum manipulation $m^* = 1$ can be obtained by setting $\overline{W}[\Pi(e^*, m^*)] = \underline{W}[\Pi(e^*, m^*)]$. The lowest m^* that can be obtained with no agency rent obtains from (OA.13) taking $\underline{W}[\Pi(e^*, m^*)] = 0$. This lower bound was defined in the main text as

$$m_0 \equiv \frac{1}{2}\left[\frac{1}{\gamma} - \frac{p\rho}{1-p\rho}\right].$$

Note that m_0 does not depend on the level of effort e^*.

It follows from the above analysis that any contract that incentivizes a level of manipulation m^* strictly lower than m_0 must leave the trader with a rent. In that case, (OA.11) is slack, and in the cost-minimal contract, (OA.12) must hold with equality. In addition, $\underline{W}[\Pi(e^*, m^*)] = 0$. Indeed, if $\underline{W}[\Pi(e^*, m^*)]$ was strictly positive, decreasing it while keeping $\overline{W}[\Pi(e^*, m^*)] - \underline{W}[\Pi(e^*, m^*)]$ constant would not affect (OA.12), but would reduce payments to the trader. (Since (OA.11) is slack, there exists some room to lower payments while keeping the search constraint satisfied.) (OA.12) holding with equality pins down the minimum payment $\overline{W}[\Pi(e^*, m^*)]$ to the trader, conditional on no manipulation detection, and in turn, his agency rent.

Finally, note that one can make the payment monotonic in the profit $\overline{\Pi}$: for $\overline{\Pi} > \Pi(e^*, m^*)$, setting $\overline{W}(\overline{\Pi}) = \overline{W}[\Pi(e^*, m^*)]$ rather than $\overline{W}(\overline{\Pi}) = 0$ (and respectively for $\underline{W}(\overline{\Pi})$) is still incentive compatible. Indeed, suppose the trader chooses (e, m) such that $\Pi(e, m) > \Pi(e^*, m^*)$. Then by reducing e, the trader saves on the cost of effort while keeping the expected payment unchanged. It follows that the trader's relevant choice set reduces to (e, m) such that $\Pi(e, m) = \Pi(e^*, m^*)$. $\square$

Exogenous deadline

Suppose that the trading opportunity disappears at the first of these two times: (1) an endogenous deadline T at which the mass of traders who have executed the trade reaches I and the opportunity is depleted; (2) an exogenous deadline X, exponentially distributed with intensity $1/\chi$.

Under the hasty strategy, the implementation probability becomes

$$\begin{aligned} p_h(T) &= \int_0^T e^{-\frac{t}{\chi}} e^{-\frac{t}{\lambda}} \frac{1}{\lambda} dt \\ &= \frac{\chi}{\lambda+\chi}\left[1-e^{-\left(\frac{1}{\lambda}+\frac{1}{\chi}\right)T}\right]. \end{aligned}$$

Under the hasty strategy, the implementation probability is 0 if $T \leq \iota$ and otherwise,

$$\begin{aligned} p_d(T) &= \int_0^{T-\iota} e^{-\frac{t+\iota}{\chi}} e^{-\frac{t}{\lambda}} \frac{1}{\lambda} dt \\ &= \frac{\chi e^{-\frac{\iota}{\chi}}}{\lambda+\chi}\left[1-e^{-\left(\frac{1}{\lambda}+\frac{1}{\chi}\right)(T-\iota)}\right]. \end{aligned}$$

The definition of the net (private) benefit of risk management is unchanged:

$$\Delta(T) = p_d(T)\rho(\pi+\alpha_+) - p_h(T)\pi.$$

Note also that if $T > \iota$,

$$\frac{\partial \Delta}{\partial T} = \frac{1}{\lambda} e^{-\left(\frac{1}{\lambda}+\frac{1}{\chi}\right)T}\left[e^{\frac{\iota}{\lambda}}\rho(\pi+\alpha_+) - \pi\right] > 0.$$

Since the shape of $\Delta(.)$ is unchanged, Lemma 1 still holds, and there exists a unique threshold T^* such that a trader chooses risk management if and only if $T > T^*$, and T^* is an increasing function of π. T^* can be expressed as a function of π. In particular, if $\pi < 0$, it is optimal to be deliberate, and $T^* = 0$. Conversely,

$$\lim_{T\to+\infty} \Delta(T) = \frac{\chi}{\lambda+\chi}\left[\pi - \rho e^{-\frac{\iota}{\chi}}(\pi+\alpha_+)\right], \tag{OA.14}$$

and hence, if the expression between brackets is positive, then RM is never profitable, even when T is arbitrarily large. That is, rearranging (OA.14), if $\pi > -\frac{1-\rho}{e^{\iota/\chi}-\rho}\alpha_-$, the hasty strategy is dominant and $T^* = +\infty$. In between these two bounds, T^* is a strictly increasing function of π.

The bounds $\underline{\pi}'$ and $\overline{\pi}'$ can be defined in the same way as in the main text, and one shows that the deliberate equilibrium exists iff $\pi < \overline{\pi}'$, while the hasty equilibrium exists if $\pi > \overline{\pi}'$. It is however no longer true that the deliberate strategy is collectively optimal for traders. Indeed, the difference in aggregate surplus between every trader being deliberate and every trader being hasty is

$$\begin{aligned} &\left[\int_\iota^{T_d} \frac{1}{\chi} e^{-\frac{t}{\chi}}\left(1-e^{-(t-\iota)/\lambda}\right)dt + e^{-\frac{T_d}{\chi}} I\right]\rho(\pi+\alpha_+) \\ &-\left[\int_0^{T_h} \frac{1}{\chi} e^{-\frac{t}{\chi}}\left(1-e^{-t/\lambda}\right)dt + e^{-\frac{T_h}{\chi}} I\right]\pi \end{aligned} \tag{OA.15}$$

For every realization of π, there exists ι sufficiently large relative χ that (OA.15) becomes negative. That is, while in the baseline case, the aggregate profits when all traders are deliberate arbitrarily was independent of ι, it can now be made arbitrarily small by taking ι sufficiently large.

However, the coordination failure remains. Indeed, when both equilibria co-exist, the deliberate one is Pareto-optimal. To see this, notice that a trader's profit in the deliberate equilibrium has to be higher under the deliberate strategy than under the hasty strategy keeping the (endogenous) deadline unchanged, that is, equal to T_d. The profit from a hasty strategy is, in turn, higher under the deliberate deadline T_d than under the hasty deadline T_h. It follows that every trader is better off in the deliberate equilibrium than in the hasty one.

Decision making in continuous time

We consider here a variation of the baseline model in Section 3 where the trader can adjust his risk management decision "on the fly" as the trading game unfold. We show that, as in the original version of the model, strategic complementarities generate socially inefficient risk management decisions. Interestingly, in this specification, a unique pure-strategy equilibrium obtains even without perturbing the model with dispersed information.

We introduce two modifications to the original model. First, traders, instead of being committed to a risk management decision, can execute the trade at any point in time after they have located the trading opportunity. That is, a trader k who identifies the potential trade and engages into risk management can decide to execute the trade before the procedure is completed and produces information on the private value α_k. Second, we assume that the risk management time frame, instead of being deterministic (equal to ι) follows an exponential distribution with intensity $\frac{1}{\iota}$. This second change introduces continuity through time in the expected payoff from risk management: at any t, conditional on having located the trading opportunity before t, trader k has a probability $\frac{1}{\iota}dt$ of learning α_k between t and $t + dt$.

Given these assumptions, a trader who finds the investment opportunity at time t enters into a continuous decision process: at any subsequent time, he needs to decide wether to execute the trade or extend the risk management process for another "small" period dt in the hope of learning his private value α_k. This decision process stops either with the trader executing the trade before learning his private value, or with the trader learning α_k and optimizing his trading decision, or with the trading opportunity disappearing before the trader had a chance to implement his strategy. To simplify the exposition, we assume that traders do not observe trades by other traders. In other words, we allow traders' strategy to depend only on time.[OA2] Since, conditional on learning α_k, trader k has a dominant option, a strategy only needs to specify the investigation period following the discovery, after which he executes the trade if he has not learnt α_k. Note that a trader's decision to execute a trade without learning α_k does not depend on the point in time at which the opportunity was located. Note also that if a trader is willing to execute a trade at time t without knowing α_k, he will a fortiori be willing to do so at any subsequent time. Hence, trader's k strategy can be captured in a single variable τ_k. If trader k finds the investment opportunity before τ_k but does not learn α_k before τ_k, he executes the trade at τ_k. If trader k finds the investment opportunity after τ_k, he executes the trade immediately.

T_h is defined as in the previous sections, but because the investigation time is now random, the definition of T_d (the time at which the trading opportunity is depleted if traders never execute without learning α_k) changes. Specifically, T_d solves

$$\rho\left(1 - \frac{\frac{1}{\iota}e^{-\lambda T_d} - \lambda e^{-\frac{1}{\iota}T_d}}{\frac{1}{\iota} - \lambda}\right) = i.$$

Note that, instead of being a binary decision as in the original specification, risk management is now a continuous variable: the higher τ_k, the more likely it is that trader k makes an informed trading decision. When $\tau_k = T_d$, trader k is fully deliberate: he never trades without knowing α_k. We show the following result.

Proposition OA.1. *There is a unique pure-strategy equilibrium. In this equilibrium $\tau_k = \tau^\star$ for all k, with $T_h < \tau^\star < T_d$ and $\tau^\star$ is a decreasing function of i.*

[OA2]One can in fact show that allowing for the trader's strategy to depend on the mass of executed up to time t does not change the results.

Proof. Notice first that if $t < T_h$, investigating is a dominating strategy: even if every firm follows the hasty strategy, the investment opportunity will not be fully exhausted until T_h. Hence we can delete all strategies $\tau_j < T_h$.

Given that any firm that finds the opportunity before T_h investigates until T_h, at T_h, the fraction of firms that have found the investment opportunity and do not know their α_k is

$$m(T_h) = 1 - e^{-T_h/\lambda} - \left(1 - \frac{\frac{1}{\iota}e^{-T_h/\lambda} - \lambda^{-1}e^{-\frac{1}{\iota}T_h}}{\frac{1}{\iota} - \lambda^{-1}}\right)$$

while the size i of the investment opportunity is

$$i(T_h) = i - \rho\left(1 - \frac{\frac{1}{\iota}e^{-T_h/\lambda} - \lambda^{-1}e^{-\frac{1}{\iota}T_h}}{\frac{1}{\iota} - \lambda^{-1}}\right).$$

Note that $m(T_h) < i(T_h)$. Intuitively, this reflects the fact that we derived T_h assuming that all firms would follow the hasty strategy, while firms will in fact follow a strategy where they investigate at least until T_h (if they do not learn α_k before T_h). We move now to the next round of deletion of dominated strategy.

Suppose that all firms play the hasty strategy from T_h on. Then the mass of investment between T_h (included) and $t > T_h$ is

$$m(T_h) + e^{-T_h/\lambda} - e^{-t/\lambda}.$$

Hence, it is a dominating strategy to choose $\tau \geq T_h^1$ where T_h^1 solves

$$m(T_h) + e^{-T_h/\lambda} - e^{-T_h^1/\lambda} = i(T_h) \Leftrightarrow 1 - e^{-(T_h^1 - T_h)/\lambda} = e^{T_h/\lambda}[i(T_h) - m(T_h)]$$

By continuing to iterate, we obtain an increasing sequence $(T_h^n)_n$ which converges. Hence the only strategies that survives iterated deletion of strictly dominated strategies are such that $\tau_j \geq T_h^\infty$, where $\tau_j \geq T_h^\infty$ is solution to $i(T) - m(T) = 0$, that is,

$$1 - e^{-T_h^\infty/\lambda} - \rho\left(1 - \frac{\frac{1}{\iota}e^{-T_h^\infty/\lambda} - \lambda^{-1}e^{-\frac{1}{\iota}T_h^\infty}}{\frac{1}{\iota} - \lambda^{-1}}\right) = s_0. \tag{OA.16}$$

(OA.16) has a unique solution if $i < \rho$, and one can show $T_h < T_h^\infty < T_d$.

This concludes the first part of the proof.

Conjecture an equilibrium where a strictly positive mass of firms choose to investigate at $t = T_h^\infty$. Let i_t denote the size of the investment opportunity at t. Let p_t denote the ratio of the size of the investment opportunity to the mass of firms that have found the opportunity but have not invested yet. By definition of T_h^∞, $p_t < 1$ if $t > T_h$ and $p_t = 1$ if $t = T_h^\infty$.

Finally, let T denote the time at which the investment opportunity is depleted. In equilibrium, the following condition must be true

$$T = \max_k \tau_k \tag{OA.17}$$

Suppose indeed that there exists τ_j such that $\tau_j > T$ (intuitively, firms keep deliberating until after the investment opportunity is depleted). For dt "small", the net benefit of waiting at $T - dt$ is

$$\frac{1}{\iota}dt\frac{\pi+1}{2} - \pi,$$

which becomes negative for dt small enough. Hence, $\tau_k > T$ cannot be an equilibrium strategy. Thus, for all j, $\tau_k \leq T$, which implies that all firms that have found the opportunity before T invest at T at the latest. Since $p_t \leq 1$ for $t \geq T_h^\infty$, it implies that the investment opportunity is fully depleted exactly at $\max_k \tau_k$.

The net benefit of waiting at $T - dt$ for the firm that plays T is

$$\frac{1}{\iota}dt\frac{\pi+1}{2} + (1 - \frac{1}{\iota}dt)\pi p_T - \pi,$$

which becomes negative for dt small enough unless $p_T = 1$. Thus the only possible equilibrium is $\tau_k = T_h^\infty$ for all k.

□

Proposition OA.1 is consistent with the conclusions of the original model. First, externalities between traders create a coordination failure. While it would be optimal for them too coordinate on a fully deliberate strategy, $\tau_k = T_d$, traders sometimes execute trades without learning α_k. Second, competitive pressure intensifies this inefficiency: when i goes down and the preemption motive becomes more stringent, traders spend less time deliberating in equilibrium. Third, in spite of externalities, there can still be some risk management, that is, $\tau_k > 0$ in equilibrium. To understand this, consider the case of a trader who locates the trading opportunity at the very beginning of the game ($t = 0$). Even if he anticipates that every other trader will execute the trade as soon as he finds it, he knows that the investment opportunity cannot be fully depleted before T_h. In other words, it is a strictly dominant strategy to be deliberate between 0 and T_h, and one can delete strategies τ_k smaller than T_h. But now, given that a fraction of traders who found the opportunity before T_h have realized that their private value was low by T_h and exited the market without trading, at T_h, the mass of traders who have identified the trading opportunity and have not executed the trade yet is strictly smaller than the size of the current investment opportunity. Hence, it is again a dominant strategy to deliberate a little longer. Reiterating this deletion of strictly dominated strategies, one can construct a series of thresholds that converge to $\tau^\star$. At $\tau^\star$, the mass of traders who have identified the trading opportunity and have not executed the trade yet is exactly equal to the size of the current size of the investment opportunity. The end of the proof consists in showing that there cannot be an equilibrium strategy strictly higher than this threshold.