Risk Management Failures

Matthieu Bouvard[*] Samuel Lee[†]

December 2014

Abstract

We present a theory in which deficiencies in risk management arise from a coordination failure. Firms choose privately optimal risk management regimes to be competitive in a market with short-lived trading opportunities but in aggregate can find themselves in a constrained inefficient "race to the bottom," with their best responses to time pressure exhibiting strategic complementarities reminiscent of bank runs. Comparative statics based on global games suggest that greater market access or faster search (or trading) technologies may improve certain aspects of liquidity but at the same time generate excessive trading that undermines the allocative function of markets. We identify two sources of market failure operating through opportunity costs and agency rents, and discuss approaches to regulating risk management as a governance problem or as a public goods problem.

[*]McGill University, Desautels Faculty of Management. E-mail: matthieu.bouvard@mcgill.ca
[†]New York University, Stern School of Business, ECGI, and SIEPR. E-mail: slee@stern.nyu.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

Risk management failures in financial institutions are a recurring phenomenon but not well understood. Any attempt at an explanation should address the following questions: What defines risk management failures? What causes them? Do they call for regulation, and if so, what type of regulation?

Risk management failures are clearly not blamed for every loss, as it is generally understood that optimal risk taking can end up in misfortune, but they are often confounded with excessive risk taking. How much risk a firm should take on is the prerogative of senior management, and excessive risk taking a matter of misaligned management *incentives*. The task of risk management, in contrast, is merely

> getting the right information, at the right time, to the right people, such that those people can make the most informed judgments possible. (Senior Supervisors Group, 2008, 70)

This is an *information* process, and risk management failures are deficiencies in this process.[1] These phenomena can be related: A management with excessive risk taking incentives may be less concerned about uncertainty and hence underinvest in information (Maćkowiak and Wiederholt, 2012). Repairing management incentives would in this case resolve the lack of information acquisition. But the converse is not true: A fully informed management might still, perhaps even more effectively, engage in excessive risk taking.

Risk management failures define a distinct issue, rather than just another manifestation of excessive risk taking, when senior management monitors risks insufficiently conditional on its risk appetite. The simplest frame for a theory along these lines is the following question: Can effective monitoring, while feasible, be so costly that management prefers to remain uninformed (despite enormous stakes)?[2] We propose an answer that may explain why risk management failures are endemic in the financial sector. Our key assumption is that risk monitoring delays investment decisions. For firms that face short-lived investment opportunities, this represents a cost that never diminishes in relative magnitude; quite the contrary, it inherently *scales up* with the size of the investments.

We develop a model of trading under time pressure to trace the implications of this premise. In the model, firms search for a trading opportunity that vanishes at some point in time. Before beginning its search, each firm decides whether to activate a risk management system. An active system assesses any found opportunity on its "fit" with the firm's risk profile, which improves trade decisions but takes

[1] The Senior Supervisors Group is composed of the bank regulators from France, Germany, Switzerland, the U.K. and the U.S. In the report quoted in the main text, the premise of risk management is that "information flowing in and out of risk monitoring processes can be distilled and compiled in a coherent and timely manner and made available." In Stulz (2008)'s classification of risk management failures, almost all pertain 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 risk metrics.

[2] To put this into perspective, buyers in mergers and acquisitions hire advisors to conduct due diligence before a deal is closed. In 2012, they paid advisors on average .85% of the deal value (Barba, 2013). For a $100 million deal (1/60 of the $6 billion J.P. Morgan Chase lost in the 2012 London Whale scandal), this amounts to $850,000.

time. This leads to a basic tension between trade execution and risk management: Opportunities can vanish during the assessment process.

We use this model to characterize a firm's risk management choice as a function of time pressure. Time pressure makes all trading strategies less profitable but, if search is a Poisson process, it always impacts slower ones more severely. This in turn implies monotonic best responses: Firms discard risk management if and only if time pressure exceeds a threshold. Thus, under sufficient time pressure, a firm maximizes its value by prioritizing trade execution over risk management, even though it becomes more vulnerable to "trading debacles."

We then show that these privately optimal choices can lead to a constrained inefficient equilibrium outcome when firms look for the same trading opportunities. The monotonicity of the firms' responses has the consequence that risk management choices are strategic complements: Firms that discard risk management to accelerate their trading raise the time pressure on other firms, who hence become more inclined to do the same. This can push the market into a competitive equilibrium where firms compel each other to abandon risk management against their collective interest – a "race to the bottom." Key to this market failure is that each firm, when striking its privately optimal trade-off, fails to internalize that its own speed contributes to the time pressure in the market. Furthermore, such failures are likely to befall "hot" markets, where trading opportunities are ex ante more attractive and accordingly the private opportunity cost of risk management to each individual firm larger.

To explore how the trade-off between risk management and trade execution interacts with agency problems, we embed a multi-task conflict in the model that subjects risk management to moral hazard: Traders now choose whether to invest effort into search. Furthermore, under active risk management, they must obtain approval for trades but choose to either comply or manipulate the approval process. Firms design compensation contracts and choose whether to activate risk management, which consists of the approval process and a post-trade verification process that detects manipulation, though only imperfectly.

Optimal contracts reflect how firms choose to resolve the multi-task problem: To incentivize search, they must pay traders bonuses tied to trading profits. To incentivize compliance, firms must pay wages even absent a trade that they can clawback if a manipulation is detected. These pay components have countervailing effects, raising the tension between risk management and "front offices," such that firms must leave traders with *agency rents of risk management* to simultaneously incentivize both tasks. Alternatively, firms can restrict pay to bonuses to incentivize search, thereby saving on agency rents but forgoing the benefits of risk management.

We show that time pressure raises the agency rents of risk management and thus affects contracts. Above some threshold, it renders the agency rents so large that firms "resolve" the multi-task problem by moving away from clawbacks and paying bonuses only. In parallel, this change in wage incentives makes it co-optimal to move away from risk management in favor of trade execution, thus yielding a mapping between the quality of risk management and the shape of compensation contracts. Causality also runs the other way: Firms may adopt a "bonus culture" as they discard risk management based purely on its opportunity costs (as in the model without agency). Whether to blame risk management failures on firms or (rogue) traders is hence, in general, a moot point.[3] Broadly speaking, we find that

[3]This is invariably the main point of contention when rogue traders stand trial. The trader at the heart of the 2008

time pressure devalues risk management on two fronts: The *opportunity cost* effect lowers a firm's total profit, while the *agency rent* effect shifts a increasing share of the profit to traders. Although distinct, these effects interact, which makes it complicated, if not futile, to pinpoint blame.

Indeed, we show that a race to the bottom can occur even without the opportunity cost effect (by letting the risk management delay converge to zero), which implies that the second-best contracts are *per se* constrained inefficient. The source of this market failure is a *contractual* externality: Each firm fails to internalize that incentives given to its own traders affect, by way of time pressure, the agency costs of reconciling the tension between risk management and "front office" activities in other firms. This, in and of itself, generates strategic complementarities that can propagate neglect of compliance – i.e., a "front-office culture" – through a market via competitive pressures that firms cannot withstand individually.

Our model also suggests that an improvement in certain measures of liquidity may not necessarily coincide with greater allocative efficiency. Using global games to select among equilibria, we study the comparative statics with respect to *competition* and an *external-internal speed ratio*, which measures the speed at which traders locate opportunities in the market relative to the speed at which information is processed inside their firms. Increasing either parameter reduces the time that opportunities remain unrealized, thus raising immediacy and trading volume. At the same time, the increased time pressure leads to lower levels of risk management in equilibrium, which worsens allocative efficiency. Or putting it differently, improved market access – in the form of entry or enabling firms to search markets faster – can lead to "excessive" trading (that is not just redundant but generates *worse* allocations). Key to this result are the strategic complementarities that aggregate firm-level tensions between risk management and "front offices" into a market-wide trade-off between risk allocation and trading activity.

Contrary to excessive risk taking incentives created by leverage or (implicit) bailouts, a race to the bottom in our model is impervious to capital or liquidity regulation. We discuss alternative (proposed and actual) interventions. Some have ambiguous effects as they may deter valuable as well as excessive trades, but two approaches seem reasonable: The first 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 views it as a public goods problem, subsidizes risk management investments, and provides risk controls through market processes and platforms (on which firms trade) to overcome the coordination failure.

2 Related literature

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.[4] 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

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.

[4]In reduced form, the benefit of risk management should be viewed as 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).

hedging contracts reduces a firm's capacity to enter such commitments to finance current investment. We do not model hedging decisions. Due to our interest in risk management *failures*, we focus instead on the firms' decisions to set up systems and processes to monitor risks (that they may want to hedge). In our model, the key resource 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.[5] First, risk management design is the outcome of long-run decisions that do not coincide with on-the-spot preemptive actions: Risk management design precedes individual trades. Yet, since we model trades 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 the strategic complementarities known from bank runs. Coordination failures thus assume the form of competitive responses to time pressure. Due to this structural similarity, we can also adapt the global games techniques used to select equilibria in bank run models (Goldstein and Pauzner, 2005) to risk management choices by dispersing the firms' expectations of time pressure.

Second, because risk management choices and traders' actions are distinct, we can further introduce agency problems that firms must address to implement their chosen risk management strategies. This produces a model of delegated trading cum risk management in a market with preemptive competition that lets us study the *interaction* of agency problems across firms subject to "bank run" externalities, which to our knowledge is novel. In particular, it lets us study how factors that heighten preemption, such as firm entry or search technologies, affect optimal contracts.

While there is a large literature on capital and liquidity regulation, our theory is, to our knowledge, the first to provide a distinct rationale for risk management regulation. But there is a conceptual link 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.

Our theory relates competition to risk taking. Research in banking has extensively studied effects of loan market competition on bank risk taking through bank franchise values (Keeley, 1990; Hellmann, Murdock and Stiglitz, 2000; Boyd and De Nicolo, 2005; Martinez-Miera and Repullo, 2010) and returns to screening (Ruckes, 2004; Dell'Ariccia and Marquez, 2004).[6] By contrast, the key mechanism in our paper is that competition raises the opportunity *costs* of screening. Our analysis is conceptually closer to Heider and Inderst (2012) who study a multi-task conflict between screening and loan "prospecting," which − despite lacking an explicit time dimension − bears similarity to preemptive competition.

There is also a large literature that studies, more broadly, how competition interacts with agency. It identifies a variety of effects operating through information revelation, marginal returns to managerial effort, and total firm income.[7] The overall effect is generally ambiguous, qualifying the Hicks conjecture

[5]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).

[6]A completely different perspective is taken in Gorton and He (2008) where variation in lending standards result from periodic breakdowns of collusive behavior in a repeated game between competitor banks.

[7]See, e.g., Holmstrom (1982), Nalebuff and Stiglitz (1983), Hart (1983), Scharfstein (1988), Martin (1993), Hermalin (1994), Schmidt (1997), Raith (2003), Golan, Parlour, and Rajan (2014), Dam and Robinson (2014).

that product market competition curbs managerial slack. These papers analyze one-dimensional moral hazard under (oligopolistic or reduced-form) price or quantity competition. Our model features multi-dimensional moral hazard that is exacerbated by time-based competition, and in this respect, is closer to Benabou and Tirole (2014) where firms skew incentives towards easily contractible tasks in response to labor market competition.

Time-based competition is a central theme in 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.[8] 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." This resemblance highlights that financial and innovation markets are alike in that speed is as important as information (innovation).

Time-based competition has become the focus of nascent research on high-frequency (low-latency) trading in financial markets. Apart from showing that the race to reduce latencies 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).[9] While our analysis is not specific to high-frequency trading, it shares the concerns about speed, and in this respect, points to a related issue: Competing on speed may impair the *risk allocation* in financial markets by altering governance processes inside firms.

This connects our paper to research on the allocative role – or real impact – of secondary capital markets. Existing work in this literature revolves around price efficiency as a source of information that can destroy risk-sharing opportunities, improve investment decisions, serve as benchmarks for incentive contracts, and frustrate takeovers.[10] In our model, allocation is driven not by the informational role of market prices but by information processes inside firms, which are, however, affected by the market ("speed").

Focusing on the linkage between markets and organizations combines the perspectives of two recent papers on risk management. Garleanu and Pedersen (2007) analyze a feedback loop between market liquidity and risk management in a search-matching model of financial trading. They adopt a *market* perspective but abstract from organizational aspects. Landier, Sraer, and Thesmar (2009) concentrate on the "dissent" function of risk management inside a firm and study when the independence required to perform this function may be compromised. Theirs is an *organizational* perspective that abstracts from market equilibrium interactions. In our framework, organizational design and market equilibrium are jointly determined as the result of a trade-off between "dissent" in organizations and preemption in the market.

The remainder of the paper proceeds as follows. Section 3 motivates the main premise of our theory

[8] See, e.g., Smit and Trigeorgis (2004) for an overview of this approach. More specifically, for work on real options in competitive environments, see Leahy (1993), Caballero and Pindyck (1996), Grenadier (2002), and Aguerrevere (2009), who examine how strategic interactions between firms affect their payoffs from exercising their options.

[9] A notable exception is Pagnotta and Philippon (2013) who, instead of focusing on the preemptive motives of traders, analyze the incentives of securities exchanges to offer trading platforms of different speed to heterogenous traders.

[10] For examples of these four mechanisms, see Hirshleifer (1971), Fishman and Hagerty (1992), Holmstrom and Tirole (1993), and Edmans, Goldstein, and Jiang (2012). Bond, Edmans, and Goldstein (2012) survey this extensive literature.

with stylized facts. Sections 4 and 5 analyze the model without agency. Section 6 introduces agency. Section 7 considers the robustness of our results to a number of extensions. The paper concludes with a regulatory discussion in Section 8 and final remarks in Section 9.

3 Motivating observations

3.1 Occupational characteristics of financial trading

Our theory builds on the assumption that financial traders operate (*i*) under significant time pressure (*ii*) in a highly competitive environment and (*iii*) make decisions with potentially material impact on their firms. We use data from the Department of Labor's Occupational Information Network (O*Net) to check whether this description is accurate. O*Net uses worker surveys to describe occupations along hundreds of characteristics. Most relevant to our analysis are the four "structural job characteristics" (from the category "work context" that are) displayed in Table 1. Each characteristic is measured on a 100-point scale, divided into five intervals of equal length that corresponds to qualitative assessments. A score in the highest interval (80-100) indicates that the respective characteristic is extremely salient in the occupation under consideration.

Table 1: O*Net characteristics.

O*Net characteristics	All 922 occupations			Sales Agents, Securities & Commodities
	Average	Median	Std. dev.	
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/

For these characteristics, financial traders are subsumed under the (group of) occupation(s) *Sales Agents, Securities and Commodities*. Table 1 displays the score of this occupation for each characteristic and, for comparison, the mean, median, and standard deviation of each characteristic across the 922 occupations that O*Net reports these characteristics for. *Sales Agents, Securities and Commodities* scores significantly above average in all four characteristics.

What the table does not convey is that it is rare for occupations to simultaneously score high on all of these characteristics. The three-dimensional scatter plots in Figure 1 illustrate this fact. In either

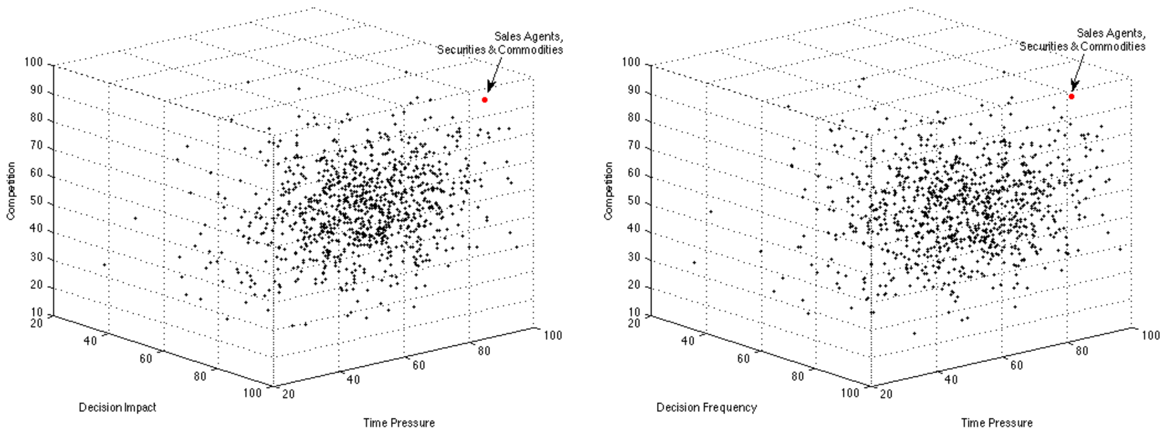

Figure 1: O*Net characteristics—scatter plots.

plot, 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, out of the 922 occupations, only two others score in the 80-100 interval simultaneously on time pressure, competition, and decision impact. Furthermore, *Sales Agents, Securities and Commodities* not only has the highest average across these characteristics but its average score is also a statistical outlier, as illustrated by the box plot in Figure 2.[11]

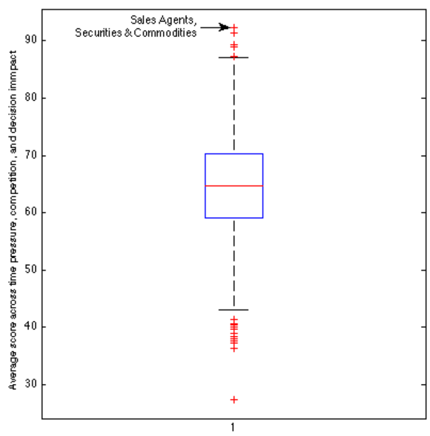

Figure 2: O*Net characteristics—box plot.

Of course, this does not imply any causal relationship between these characteristics. For example, while the correlation coefficient of .18 between time pressure and competition in the O*Net database indicates a positive relation, we do not know to what extent time pressure *arises from* competition.

[11] Adding a minimum score of 80 on decision frequency as a filter is redundant. Also, *Sales Agents, Securities and Commodities* remains an outlier when decision frequency is incorporated in the average.

3.2 Time-related concerns and tensions in risk management

Another key premise of our analysis is that the process of collecting, aggregating, and communicating risk information takes time and that this creates a tension between trading and risk management. In a recent survey, chief risk offers of major financial institutions identify their main challenges as (i) the conflict between trading desks ("front office") and risk management and (ii) the need for (investment in) information technology and data systems that can deliver "timely transparency" (Ernst and Young, 2013, 2-5 and 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 essence (of this and other passages in the survey) is that the delivery of higher-quality reports requires more time and that investment in technological and organizational infrastructure is necessary to reduce that time in order for risk management to perform its function properly. The most interesting aspect of these statements is that they silently presuppose that timeliness is *paramount*; the twin notion of slowing down the other operations to match the pace of risk management is absent.

Such a priority ranking can emerge in fast-moving markets because the value of risk management is a derivative of the ability to trade: A trade can be profitable even without risk management, whereas risk management is worthless if the option to execute the trade is forgone. In what follows, we describe some contexts in which (the need for fast) trading relegates risk management to a secondary role.

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

Value-at-risk (VAR) is the main tool banks use to evaluate the risk of losses due to changes in market prices (market risk). There are two basic choices in VAR design that shed light on what banks consider important in designing their risk management: whether to simulate scenarios using 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 trade-off inherent in these choices. Most of the interviewed 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)

As regards valuation methods, sensitivities reduce trading positions to a few parameters ("Greeks") that 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]... Across all banks, the survey found that average VAR run time ranges between 2 and 15 hours; in stressed environments, it can take much longer. (5)

The report also mentions the role of time in explaining the difficulty of aggregating data from the "front office" trading operations:

> 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* [emphasis added] has become a significant challenge.

The discussion of VAR methods suggests that a trade-off between accuracy and timeliness is central to risk management design. The trade-off also applies to the scope and frequency of stress tests, which are to provide early warning signals about a firm's total risk exposure (Ernst and Young, 2013, 37):

> 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... quick-turnaround results were not comprehensive.

3.2.2 Algorithmic trading: machine speed vs. human speed

Two fundamental reasons that financial markets have become so much faster over time are advances in quantitative finance modeling and the progress of computer technology. One outgrowth of these twin developments is algorithmic trading, especially its speed-oriented variant: high-frequency trading. In discussing the automatization of trading and its implications for financial 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 from the Federal Reserve Bank of Chicago voice similar concerns about the propensity of high-speed trading firms – being 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.[12] (2010, 3)

3.2.3 Operational risk: pre-trade vs. post-trade control

Operational risk in market-related activities is distinct from market risk and, by and large, subsumes various forms of moral hazard, such as rogue trading, unauthorized leverage, or pricing manipulation. Despite its inclusion in the Basel II framework and a slate of recent scandals, operational risk is still an area banks rate themselves least effective in (Deloitte, 2013, 24),[13] though not because they think it unimportant:

> One of the biggest risk challenges facing the industry right now is in the conduct, behavior, and operational risk areas of risk management. As a [chief risk officer], my focus is on instilling the right mix of culture and controls for the organization to help manage integrity and behavior. (25)

Operational risk, such as moral hazard, is often difficult to quantify as "hard" information. This reduces the usefulness of information technology and automatization and makes monitoring such risks more time-consuming, which is costly under time pressure. Notably, operational risk guidelines, such as those by the Committee of European Banking Supervisors (2010), often stress *incentive compensation* and *post-trade audits*, which are instruments that promote (deter) – rather than ex ante screen – good (bad) decisions without delaying the execution of trades.

3.2.4 Primacy of trading and the cost of disruption

As an illustration of the claim that firms may compromise on risk management to avoid disruptions to trading, consider the following anecdote about a bank that did not even tolerate a *transitory* disruption (Stulz, 2008, 63):

> The Union Bank of Switzerland was putting together risk management systems that would aggregate risks within its trading operations. One group of traders that focused on equity

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

[13] 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.

> derivatives was extremely successful. However, this group of traders was using different computers from the rest of the bank, so that integrating their systems into the bank's systems would have required them to change computers. Eventually, the bank decided, at the top level, that it was more important to let the traders make money than disrupt what they were doing through changes of computers. Soon thereafter, this group of traders lost a large amount of money for the bank. The loss was partly responsible for the bank having to merge with another Swiss bank.

Further circumstantial evidence comes from two sources. First, consistent with our claim, chief risk officers report that risk governance is more difficult – i.e., the inclination to sidestep risk management greater – when trading activity increases (Ernst and Young, 2013, 18):

> More than half of survey respondents cited market conditions, and 39% said the competitive environment affects risk appetite, indicating that risk appetite is often reined in during difficult times and expanded when markets pick up. This underscored banks' stress on the governance aspects of risk appetite, which become more important as market conditions improve.

The second observation, closely related to our point on operational risk, comes from survey answers regarding the extent to which firms interfere with – i.e., take decision power away from – "front offices" (Deloitte, 2013, 11):

> Most institutions reported that they followed a centralized approach to risk management. For example, roughly 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... However, there were several areas where institutions were more likely to report their business units played a leading role. For trading transaction approval, 54 percent of institutions said this was determined by their business units compared to 28 percent who cited independent risk management.

The shift towards decentralization of decision power in *trading*-related activities accords with our premise that risk oversight is disruptive where time is of the essence.

4 Framework

Consider a mass M of risk-neutral firms (traders), indexed by k, who compete 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 instance, shocks to the demands of agents operating in segmented markets can create discrepancies in the pricing of two assets with correlated cash flows, as in Shleifer and Vishny (1997) or Gromb and Vayanos (2002).

We introduce a time friction that interferes with the ability of traders to instantly take advantage of this mispricing. This friction could take two forms. First, discovery can take time: there can be delay between the appearance of the mispricing at $t = 0$ and the times at which each trader becomes aware of this trading opportunity. In addition, execution can take time: there can be a delay between a trader's final decision to trade and the effective execution. Because these two frictions are equivalent in our framework, we model only the first one.[14] Specifically, we assume that it takes each trader a random time $\tilde{t}_k$ to discover the trading opportunity. Discovery times are identically and independently distributed according to an exponential distribution:

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

The traders' discount factors are normalized to 1.

Upon locating an opportunity, a trader can request to trade one unit. Executing the trade yields a payoff $\pi + \alpha_k$, that is, the mispricing π, which is a common value across traders, plus a private value α_k. We interpret α_k as the fit between the trade and the risk profile of that particular trader, desk, or firm. For example, a financial intermediary may prefer trades that hedge rather than amplify its existing risk exposure.[15] α_k can also represent the shadow cost of mobilizing collateral to guarantee a trade position. More generally, α_k reflects any friction that amplifies the impact of cash flow shocks on a firm, thereby justifying risk management: bankruptcy costs, financial constraints, and coordination problems (see, e.g., Froot, Scharfstein, and Stein, 1993).

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

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

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 the fit between a trade and a trader's risk profile. 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.

Finally, we assume that the mispricing is sensitive to the pressure exerted by traders. Specifically, it disappears once the mass of trades that exploits 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).[16] The finite size of the trading opportunity creates a preemption motive for traders, and the intensity of this time-based competition is captured by the ratio of the size of the trading opportunity to the mass of traders, which we denote

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

[14] Specifically, a model in which all traders locate the trading opportunity instantaneoulsy and execution introduces latency is isomorphic to a model in which locating the opportunity introduces latency and execution is instantaneous.

[15] 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 and Young, 2013, 20).

[16] In Section 7, 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.

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, and we will later endogenize M by introducing an entry stage.[17]

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

First, risk management simply blocks trades that create undesirable firm-specific risks, captured by $\alpha_k = \alpha_-$. In practice, some risks, once identified, could be hedged, which would allow the firm to proceed with a trade even if $\alpha_k = \alpha_-$. This would not substantially change our analysis if hedges can be costly or imperfect. (Marking positions to market, for example, creates funding liquidity risk that is difficult to hedge.) More importantly, incorporating hedging does not change the gist of our analysis as long as the process of identifying and implementing hedges slows down other trading activities. Key to our model is that risk management consumes time that traders could use otherwise, but not what precisely it does with that time.

Second, the decision whether to activate risk management is made ex ante. Since this is the only strategic decision, trading afterwards unfolds mechanically. 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.[18] However, in Section 6, we allow agents that operate within a firm's chosen risk management framework to manipulate the controls. For completeness, in Section 7, we also 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 trader knows how many in total locate the opportunity over time and how many of those traders have positive private values. In fact, the only aggregate uncertainty is strategic: To infer how many trades are *executed* over time, traders must form beliefs about everyone else's risk management choices. We later show that introducing an exogenous source of aggregate uncertainty sharpens the predictions of the model (Section 5.3).

5 Risk management in equilibrium

5.1 Opportunity costs and privately optimal risk management

We start the analysis with a firm's private incentives to activate 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 $\pi < 0$:

[17] $i < \rho$ 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.

[18] The Basel Committee uses the term risk management "framework" to define systematic risk controls in firms.

It is only when trading without risk management is (on average) profitable that risk management has an opportunity cost.

By contrast, if the trader activates risk management, his expected profit is $\rho(\pi + \alpha_+)$ conditional on 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 trader's trade-off is captured by the difference between the unconditional expected profits of the two strategies, which we define for any $T > 0$ as

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

We refer to $\Delta(.)$ as the private value of risk management. On the bottom line, the first term represents the benefit of risk management: avoiding bad trades that would occur with probability $p_h(T)(1-\rho)$ under the hasty strategy. The second term represents the cost of risk management: failing to capture good opportunities with probability $[p_h(T) - p_d(T)]\rho$ that would have been executed under the hasty strategy. The benefit is contingent on the probability $p_h(.)$ of being able to execute a trade, indicating that the value of risk management is a *derivative* of trading. The opportunity cost of risk management depends on the *differential* ability to capture trades under the two strategies, $p_h(T) - p_d(T)$, which is driven by their relative speeds.

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 variation of $\Delta(.)$ is therefore ambiguous a priori and depends on the differential impact of time pressure across strategies. Clearly, 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. The interesting case is $T > \iota$. Here, the deliberate strategy is more sensitive to time pressure than the hasty one for reasons related to the benefits and costs of risk management:

(i) *Value per trade.* Conditional on implementation, the deliberate strategy pays off more than the hasty strategy: $\rho(\pi + \alpha_+) > \pi$. Hence, even if implementation probabilities $p_h(.)$ and $p_d(.)$ were equally sensitive to time pressure, relaxing time pressure would raise the unconditional expected profit on the margin more under the deliberate strategy.

(ii) *Decreasing marginal returns to time.* The implementation probability of the deliberate strategy, $p_d(.)$, is not only lower but more sensitive to time pressure than that of the hasty strategy, $p_h(.)$. Implementation requires finding the opportunity before T under the hasty strategy, and before

$T - \iota$ under the deliberate strategy. Because a hasty trader can search from 0 to T, he gains less from being able to search for an extra interval dT than a deliberate trader who can search only from 0 to $T - \iota$. Intuitively, additional time is more valuable to those who have less of it to start with.

From the above, it follows that $\Delta(.)$ is U-shaped and reaches a minimum at $T = \iota$ (see Figure 3).[19] 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.*

Proof. In the Appendix. □

This stands in stark contrast to explanations of risk management deficiencies based on risk shifting incentives. The propensity to take on undesirable risks in this model is not driven by differential claims held by various stakeholders (e.g., workers, managers, shareholders, and creditors), nor does it pit their interests against each other. On the contrary, under Lemma 1, a firm abandons risk management only to *maximize firm value* for all stakeholders by staying competitive under time pressure, i.e., for $T < T^*$.

Apart from points already discussed, Lemma 1 further proves that T^* increases in π, that is, risk management is less likely when the common value of trades is higher. The reasons are twofold and can be traced back to (1): Clearly, losses from bad trades, $\pi+\alpha_-$, become smaller, that is, the benefit of risk management decreases. Further, profits forgone by trading at a slower pace, $[p_h(T) - p_d(T)]\rho(\pi+\alpha_+)$, become larger, that is, the opportunity cost of risk management increases.

We conclude this subsection with two remarks. First, the threshold strategy implies that firms' best responses are monotonic: Their inclination towards risk management always (weakly) increases with T, which will be instrumental in generating strategic complementarity. Second, despite the monotonicity in strategies, the private value of risk management $\Delta(.)$ is decreasing in T, while remaining negative, as long as $T < \iota$. We will later discuss the implications of this non-monotonicity for the application of a global games refinement.

5.2 Strategic complements and constrained 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

[19] 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.

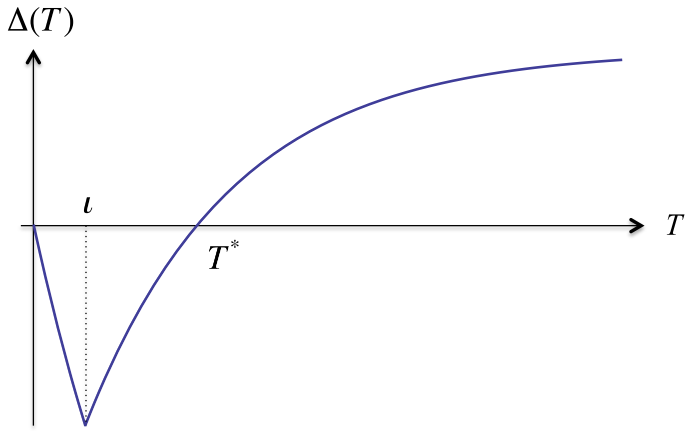

Figure 3: Time pressure and private value of risk management.

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

Since $p_h(T) > \rho p_d(T)$ and $p'_.(.) > 0$, it is easy to see that the deadline T implied by (2) decreases with q; quite naturally, time pressure increases with the fraction of hasty traders. This together with the traders' monotonic best responses to time pressure creates strategic complementarities.

It will be useful to define the shortest and the longest deadline. When all traders are hasty ($q = 1$), (2) becomes $p_h(T) = i$, the solution to which we denote by T_h. By time T_h, a fraction $p_h(T_h)$ of the hasty traders will have located the opportunity and executed a trade. In contrast, when all traders are deliberate ($q = 0$), (2) becomes $\rho p_d(T) = i$, the solution to which we denote by $T_d > T_h$. By time T_d, a fraction $p_d(T_d)$ of the deliberate traders will have located the opportunity and requested a trade, but only a fraction ρ of the trades is executed, the remaining $1 - \rho$ being blocked by risk management because the private value component is negative.

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. Suppose a trader believes everyone else is hasty. The conjectured deadline T_h is then smaller than $T^*(\pi)$ so that his best response is also to be hasty. Thus, there exists a hasty equilibrium. However, if the trader instead believes that everyone else is deliberate, the implied deadline is T_d is larger than $T^*(\pi)$ and supports a deliberate equilibrium. The source of this equilibrium multiplicity is that risk management choices are strategic complements. When more traders are deliberate, the trading opportunity is depleted at a slower pace because risk management delays execution and blocks executions when $\alpha_k = \alpha_-$. This reduces the time pressure on others, making the deliberate strategy more attractive to them. By the same token, more hastiness increases time pressure, which begets more hastiness.

Proposition 1. *There exists a unique interval $[\underline{\pi}, \overline{\pi}] \neq \emptyset$ such that the hasty equilibrium exists for all $\pi \geq \underline{\pi}$ and the deliberate equilibrium exists for all $\pi \leq \overline{\pi}$. The interval $[\underline{\pi}, \overline{\pi}]$ is empty iff $\iota = 0$.*

The existence of an intermediate region with multiple equilibria is reminiscent of other games of strategic complementarities, such as bank runs or currency attacks. Our model suggests that similar complementarities can also operate through time pressure and learning incentives, rather than direct payoff externalities: By choosing to be hasty, each trader increases the opportunity cost of acquiring information for every other trader.

The hasty equilibrium that the strategic complementarities can generate for $\pi \leq \overline{\pi}$ is a coordination failure – a "race to the bottom." Whether the equilibrium is hasty or deliberate, traders have the same probability of executing a trade. However, the value of an executed trade in the deliberate equilibrium, $\pi + \alpha_+$, is higher than in the hasty equilibrium, π. In fact, the difference in aggregate trader surplus between the two equilibria,

$$I[\pi + \rho\alpha_+ + (1-\rho)\alpha_-] - I[\pi + \alpha_+] = I(1-\rho)(\alpha_+ - \alpha_-), \tag{3}$$

does not depend on the common value π: Since the trading opportunity is depleted in both equilibria, the common value is always fully absorbed. The superiority of the deliberate equilibrium over the hasty equilibrium is that a fraction $1-\rho$ of hasty trades with negative private value is replaced by deliberate trades with positive private value. Thus, the deliberate equilibrium implements a more efficient risk allocation. Even if $\pi > \overline{\pi}$, when the deliberate equilibrium does not exist, a collective commitment to risk management would make all traders better off:

Corollary 1. *Every trader being deliberate Pareto-dominates any other strategy profile.*

The reason why traders can nonetheless be "trapped" in a hasty equilibrium is that their individual objective function fails to internalize the social value of risk management. To see this, consider trader k's net gain from deviating to the deliberate strategy when everyone else is the hasty:

$$\Delta(T_h) = p_d(T_h)(1-\rho)|\pi + \alpha_-| - [p_h(T_h) - p_d(T_h)]\pi. \tag{4}$$

The second term captures the preemption motive: By switching to the deliberate strategy, the trader is less likely to capture the common value π. This constitutes a private loss for k but not a social one since some trader will capture this common value in lieu of k. The first term corresponds to the private benefit of risk management: Under the deliberate strategy, when k implements his strategy and has a low private value, he avoids a loss of $|\pi + \alpha_-|$. However, the social benefit of risk management in this case also depends on the private value of the (hasty) trader who executes in lieu of k: This social gain is zero if that private value is low, but equal to $\alpha_+ - \alpha_-$ if that private value is high, which happens with probability ρ. Since $\rho(\alpha_+ - \alpha_-) > |\pi + \alpha_-|$, trader k does not fully internalize the allocative efficiency gain of risk management. In Section 7, we discuss an extension of the model in which every trader being deliberate, unlike in Corollary 1, need not be the (ex ante) collectively optimal outcome for all traders. However, we show that the qualitative implications of the coordination failure remain unchanged: There exist equilibria with too little risk management.

Note finally that the intermediate region $[\underline{\pi}, \overline{\pi}]$ also accommodates an equilibrium in which both strategies are used. This requires traders to be indifferent between the strategies, i.e., $\Delta(T) = 0$ or $T = T^*(\pi)$.[20] This mixed strategy equilibrium is unstable, unlike the equilibria in Proposition 1: Any small shock that shortens or extends the deadline T^* breaks the indifference condition and pushes all traders toward the same hasty or deliberate strategy. We will later comment on extensions of the model that may accommodate stable interior equilibria (Section 7).

5.3 Immediacy versus allocative efficiency

While equilibrium multiplicity highlights strategic complementarities, the indeterminacy generates ambiguity about the impact of (structural changes affecting) competition intensity or processing speed. A preliminary step for comparative statics consists therefore in refining the equilibrium predictions of Proposition 1.

This indeterminacy in games of strategic complementarities can typically be resolved with global games (Carlsson and van Damme, 1993; Morris and Shin, 1998). Standard global games techniques readily apply in models with *global* strategic complementarities (see, e.g., Morris and Shin, 2003). In our setting, global strategic complementarities require that raising the fraction of deliberate traders always increases the private value of risk management, $\Delta(.)$. This requirement is violated: While the deadline T is monotonically increasing in the fraction of deliberate traders, $\Delta(.)$ is not monotonic in T. That is, starting from a point where every trader is hasty and the deadline is tight (equal to T_h), converting some hasty traders into deliberate ones may *increase* the profitability of the hasty strategy relative to the deliberate one. However, our model satisfies a weaker form of strategic complementarity: By Lemma 1, $\Delta(.)$ crosses 0 once, and is monotonic whenever positive. These properties define *one-sided* strategic complementarities, 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 $(-\alpha_+, -\alpha_-)$.[21] 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, being more involved in the absence of global strategic complementarity, is relegated to the Appendix. Equilibrium strategies take the threshold form that is typical of global games: In the unique equilibrium, every trader is hasty if his signal s_k is below a common threshold $s^\star$, and deliberate otherwise. Note that because traders hold different beliefs about π, hasty traders may

[20] For a given π, the fraction $q^* \in (0, 1)$ of traders that (randomly) choose the hasty strategy must therefore satisfy $q^* p_h[T^*(\pi)] + (1 - q^*)\rho p_d[T^*(\pi)] = i$.

[21] 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 ($\pi < \underline{\pi}$)

coexist with deliberate traders. Note also that a higher realization of π lowers the mass of traders who receive a signal below the threshold $s^\star$, and hence the mass of traders who activate risk management. Finally, a well-known property of global games is that equilibrium uniqueness carries over even when the information structure moves 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 strategy if $\pi < \pi^*$. π^* is strictly positive and satisfies*

$$\int_0^1 \Delta[\tau(q), \pi^*]dq = 0, \tag{5}$$

where $\tau(q)$ solves

$$qp_h(\tau) + (1-q)\rho p_d(\tau) = i. \tag{6}$$

Proposition 2 is the counterpart of Proposition 1 in the richer environment of global games.[22] To understand the role of equations (5) and (6), consider the non-limit case where ε is bounded away from 0. The marginal trader who receives the threshold signal $s^\star$ believes that π is uniformly distributed between $[s^\star - \varepsilon, s^\star + \varepsilon]$. Since every trader plays the threshold strategy $s^\star$, and the errors in traders' signals are independently distributed, every π maps one-to-one into a proportion of traders who play a hasty strategy, which we denote q. 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 $\hat{\tau}(q)$ through (6). (5) then captures the indifference condition of the marginal trader under these deadlines when ε tends to 0, and his signal $s_k = s^\star$ becomes arbitrarily close to an underlying realization of the common value, $\pi^\star$.

Given the unique equilibrium prediction in Proposition 2, we can now study the impact of structural changes on the equilibrium level of risk management. Since preemption plays a key role in the model, we start by endogenizing the intensity of competition. To this end, we assume that prior to observing their signal s_k, 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

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

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

$$i\{\mathbb{E}(\pi) + \Pr[\pi < \pi^*(i)]\alpha_+\} = \chi. \tag{7}$$

[22] In equation (5), $\Delta(.,.)$ is defined as in (1), with the addition of a second argument explicitly recognizing the dependence on the common value π.

Note that the threshold π^* depends on i through (6) and (5). 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. Indeed, the marginal trader, the one receiving the signal $s^* = \pi^*$, as explained earlier, forms posterior beliefs about the distribution of deadlines $\tau(q)$, which spans $[T_h(i), T_d(i)]$ and includes T^*. By (6), when i decreases, $\tau(q)$ shifts down for all q; that is, the deadline becomes tighter for every realization of q. This shift increases the probability weight of the region below T^* in which the marginal trader prefers being hasty (i.e., $\Delta[\tau(q), \pi^*] < 0$), so that the integral in (5) turns negative and π^* must decrease for the indifference condition to continue holding.[23]

That $\pi^*(i)$ is increasing implies that the zero-profit condition in (7) 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 makes risk management less likely in equilibrium.*

Proof. In the Appendix. □

Proposition 3 suggests two effects of structural changes to the level of competition between traders. First, when barriers to entry go down and competition intensifies, the misallocation of assets worsens as traders are increasingly likely to give up risk management. At the same time, the increased proportion of hasty traders amplifies the impact of entry on execution speed. Indeed, entry has a mechanical effect on execution time: The larger the number of traders searching for an opportunity, the higher is the probability that one of them executes a trade in any time interval dt. Thus, holding the proportion of hasty and deliberate traders constant, the opportunity is depleted in a shorter amount of time. What Proposition 3 adds to this basic mechanism is an equilibrium effect: The rise in time pressure lowers the threshold for π above which traders abandon risk management, which amplifies the decrease in execution time. Note however that this amplification also counters entry by making profits fall faster as more traders enter the market. From the expression in (7), a trader's expected profit decreases in i because both the probability of implementation and the expected profit conditional on implementation go down.

We now turn to the impact of a technological shock to processing speed. Note that both search time and investigation (risk management) time are affected by changes in the speed at which information is processed. Let $\sigma = \frac{\iota}{\lambda}$ denote the external-internal speed ratio, that is, how fast search is compared to investigation.

Proposition 4. *An increase in the external-internal speed ratio σ makes risk management less likely, and reduces the mass of active traders.*

Proof. In the Appendix. □

To understand why only relative speed matters, consider how a increase in search speed impacts the marginal trader: On one hand, it raises the probability that he discovers the trading opportunity

[23] 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 Appendix).

before any given deadline $\hat{\tau}(q)$; on the other hand, all other traders also locate the opportunity faster, which shortens the deadlines $\{\hat{\tau}(q)\}_{q\in[0,1]}$. These two effects offset each other such that, under hasty strategies, no trader would gain or lose any advantage. That is, higher search speed does not advantage search *per se*. It does, however, raise the opportunity cost of risk management: A trading opportunity becomes more likely to vanish between the time t at which a trader discovers it and the time $t + \iota$ at which he may trade on it under the deliberate strategy. As the implementation probability of the deliberate strategy thus decreases relative to the hasty one, so must π^* for equilibrium condition (5) to hold. This highlights that key to our results is the latency that risk management imposes on trading *relative* to the time it takes traders to discover an opportunity.

Furthermore, by making the hasty strategy more prevalent, a relative increase in search (or trading) speed lowers the expected profit of every trader in the market so that, by the zero-profit condition (7), the mass of active traders must shrink. This suggests that technological shocks to speed influence not only the behavior but also the concentration of traders in the market, which creates a countervailing effect: Less entry slows down the market, which tends to make the deliberate equilibrium more likely. This effect is, however, of second-order magnitude.

6 Delegated trading and contractual externalities

Up to this point, we have ignored incentive problems inside firms: We have assumed that each trader acts in the best interest of his firm. In this section, we enrich the model to account for the fact that traders need to be incentivized to identify profitable trading opportunities and to properly apply risk management procedures. Our main insight is that the magnitude of this agency problem is related to the time pressure that competition creates in the market. This feature of the agency friction creates strategic complementarities between the firms' choices of incentive structures that persist even when the physical friction ι – the risk management latency – vanishes.

6.1 Multi-task problem: initiative and compliance

Each firm consists of a principal-agent pair. The principal delegates the task of locating opportunities to the trader. This task requires unobservable effort (*initiative*). We model this as a binary, ex ante choice: At $t = 0$, the trader can exert initiative at private cost c. Conditional on initiative, he locates a trading opportunity at time $\tilde{t}_k \sim \text{Exp}(\lambda^{-1})$. Without initiative, he finds nothing. As before, not every trade is in the principal's interest. To screen trades, the principal can activate risk management, which requires the trader to obtain approval for trades and selectively authorizes those with positive private values (*pre-trade control*). Identifying the private value α_k of a potential trade takes time ι as of submission of the request.[24]

While the formal decision to run risk management remains with the principal, we allow the trader to tamper with the process: Upon locating a trading opportunity, though without knowledge of α_k, he can manipulate pre-trade control or comply with it (*compliance*). Manipulation causes pre-trade control to rubber-stamp a trade regardless of whether the private value is α_+ or α_-. Because of this possibility, the risk management system investigates "bad" trades ex post and can prove manipulation

[24] Pre-trade controls can be interpreted as "position limits" or "risk limits."

with probability γ (*post-trade valuation control*).[25] The possibility of manipulation implies that not only initiative but also compliance must be incentivized.

We allow the following contract space in this model: As regards outcomes, only the common value π but not the private value α_k of a trade are contractible. As regards actions, initiative and compliance are unobservable, though evidence of manipulation found by valuation control is contractible (ex post verification); trades are observable insomuch as they must be authorized, but principals do not reward trade (requests) per se because it would invite frivolous trade (requests). Note that, although α_k is not contractible, the principal can provide risk management, or compliance, incentives indirectly through ex post verification.

The timeline is as follows: At $t = -2$, principals decide whether to activate risk management and offer contracts to traders. At $t = -1$, each trader either refuses the contract and exits the game or accepts it and makes an effort choice. From $t = 0$ onwards, traders discover the trading opportunity, conditional on effort, at random times. If risk management is active, they decide whether to manipulate pre-trade control when submitting trades and undergo valuation control after the trading opportunity is exhausted.

6.2 Privately optimal contracts

We begin by deriving optimal contracts between a principal and a trader assuming that the deadline T is exogenous. Let w^0 denote the trader's wage conditional on no trading. Without risk management, let w^+ denote his wage conditional on a trade delivering the common value π. With risk management, let w^+ denote his wage conditional on the trade further being validated by valuation control, and w^- the wage if valuation control detects manipulation.

First, suppose the principal chooses to incentivize trading without risk management. In that case, setting $w_h^0 = 0$ and w_h^+ such that $p_h(T) w_h^+ = c$, or

$$w_h^+ = \frac{c}{p_h(T)} \tag{8}$$

is optimal. Under this "hasty-strategy contract," the trader expects to recoup no more than his effort cost c. The absence of any agency rent here is due to our assumption that effort is strictly necessary to locate a trading opportunity. In a more general setup where search can also succeed without effort, the principal would have to leave the trader with a rent to incentivize effort.[26]

Now suppose the principal activates risk management and wants to incentivize both initiative and compliance. The trader's wage following the detection of manipulation is optimally set to $w_d^- = 0$. In addition, w_d^0 and w_d^+ must be set to satisfy two incentive compatibility constraints. The first constraint

[25] In investment banks, valuation control typically refers to the unit in charge of examining the valuation assumptions used by traders in their book. Valuation control is often imperfect. In the settlement between J.P. Morgan Chase and the SEC over the "London Whale" trading scandal, the bank acknowledged that its valuation control unit in London was understaffed and unable to properly oversee trading activities ("A whale of a trade: one man to check the books," The Wall Street Journal, September 19, 2013).

[26] Principals need to incentivize initiative irrespective of their choice of risk management. Hence, incentivizing initiative is important in our setup not so much because of the agency rent it could generate, but because of the structure it imposes on payments, which, as will become clear below, interferes with the provision of incentives for risk management.

ensures that the trader is willing to bear the effort cost: $p_d(T)\rho w_d^+ + [1-p_d(T)\rho]w_d^0 - c \geq w_d^0$, or

$$w_d^+ \geq \frac{1}{\rho}\left[w_d^0 + \frac{c}{p_d(T)}\right]. \tag{9}$$

The wage w_d^+ in (9) is higher than w_h^+ in (8) for several reasons. First, because risk management (i) imposes a time delay during which trading opportunities disappear with probability $p_h(T) - p_d(T)$ and (ii) blocks trades with probability ρ, the probability of trade is lower in (9), which in turn requires a larger wage conditional on trade for the trader to be willing to exert effort. Intuitively, (i) and (ii) are ways in which risk management "interferes" with trading and thereby undermines the trader's effort. In addition, w_d^+ is higher than w_h^+ because (iii) the principal may have to pay the trader $w_d^0 > 0$ even if there is no trade.

The second incentive compatibility constraint, which ensures that the trader does not manipulate the pre-trade control process once he finds a trading opportunity, is $\rho w_d^+ + (1-\rho)w_d^0 \geq (1-\gamma)w_d^+$, or

$$w_d^0 \geq \frac{1-\gamma-\rho}{1-\rho} w_d^+. \tag{10}$$

There are two cases to consider. If $1-\gamma \leq \rho$, (10) does not constrain the principal's choice of wages, in which case $w_d^0 = 0$ and w_d^+ is set via (9) to compensate the trader only for the effort cost c.[27] On the other hand, if $1-\gamma > \rho$, then $w_d^0 > 0$. Here, the trader must be given a strictly positive wage even if no trade occurs which – by virtue of being at risk after a manipulation – incentivizes compliance.

From this point on, we restrict attention to the more interesting case in which $1-\gamma > \rho$. This condition has an intuitive interpretation. When ρ is low, pre-trade control is likely to block a trade, which conflicts with the trader's endogenous incentive to maximize trading volume. When γ is low, valuation control is ineffective, which also raises the trader's incentive to manipulate. In the parlance of principal-agent theory, ρ and γ parametrize, respectively, the *congruence* between the principal and the agent and the principal's *monitoring* efficacy.

The optimal contract with risk management, or "deliberate-strategy contract," is the unique wage pair (w_d^+, w_d^0) for which both (9) and (10) are strictly binding. We summarize the qualitative features of optimal contracts below.

Lemma 2. *The optimal contract without risk management can be implemented with a bonus contingent on the common value of a trade. The optimal contract with risk management can be implemented with a base wage, a bonus contingent on the common value of a trade, and a clawback provision contingent on the ex post detection of a manipulation.*

The gist of this result is that when the implementation of risk management requires the (imperfectly observable) cooperation of traders, risk management protocols are futile unless they are accompanied by suitably designed incentive compensation schemes. Put differently, Lemma 2 suggests that it may be idle to think of risk management independently of incentive compensation. More specifically, it says that risk management frameworks must be coupled with clawback provisions (or compliance rewards) in wage contracts to be effective.

[27] Note that the contracts with and without risk management are still different as the former includes a (out-of-equilibrium) "punishment" in case of fraud, w_-.

6.3 Agency rents and time pressure

The agent is compensated for no more than his effort cost when only initiative is required, but extracts an agency rent when the principal also wants to incentivize compliance. By (9) and (10), this agency rent, or minimum expected cost of search and risk management to the principal, is

$$B(T) \equiv \frac{1-\gamma-\rho}{\rho p_d(T)\gamma} c. \tag{11}$$

The agency rent stems from a multi-task conflict: To comply with risk management, the trader must be given "skin in the game," $w_0 > 0$, that is clawed back if manipulation is detected, $w_- = 0$. But strictly positive, higher w_0 necessitate higher w_+ to incentivize initiative (via (9)), which in return necessitate even higher w_0 to deter manipulation (via (10)), and so on. That is, initiative and compliance call for conflicting incentives, which confers rents on the trader.

The principal may nonetheless prefer to incentivize compliance if this agency rent is small relative to the benefits of risk management. However, the rent is sensitive to time pressure:

Lemma 3. *The agency rent of risk management $B(T)$ is decreasing for $T > \iota$. Furthermore, $B(T) \to \infty$ as $T \to \iota^+$.*

Time pressure heightens the task conflict. A shorter deadline (smaller T) decreases the probability of trade (lower p_d), which requires an increase not only in w_+ to preserve initiative but, if the principal wants to maintain compliance as well, also in w_0. So, $B'(T) > 0$. Similarly, the more pre-trade control delays trading (larger ι), the more opportunities does the trader lose out on (lower p_d). This has the same qualitative effect on w_+ and w_0 as a decrease in T. That is, $\partial B(T)/\partial\iota > 0$. Intuitively, greater time pressure – be it due to external forces or internal frictions – raises the tensions between trading desks ("front office") and risk management, the resolution of which thus becomes more expensive for the firm. In the limit, as $T \to \iota$, it becomes prohibitively costly to avoid manipulation as $B(T) \to \infty$. This implies a natural mapping between the quality of risk management and the shape of compensation contracts as a function of time pressure: When time pressure rises, moving away from clawbacks to a pure "bonus culture" is co-optimal to relaxing risk controls.

We conclude this subsection with three remarks on modeling choices. First, the assumption that the realization of the private value α_k is not contractible is made both for convenience and because it may, in practice, be difficult to define a contractible measure of the fit between different, time-varying risk exposures. That said, even with α_k contractible, a trader has incentives to manipulate pre-trade control if trades that would be blocked have some (small) probability of generating α_+. Such a setting yields similar results, except that contract clauses contingent on α_k render valuation control (partly) obsolete.

Second, manipulation only occurs out of equilibrium in our current setting since the principal has no interest in risk management without adequate compliance incentives. However, suppose manipulation were imperfect in that it succeeds with a probability smaller than one. The principal could then find it optimal to activating risk management *without* providing traders with the incentives to comply. Under this regime, risk management would lose precision due to manipulation, yet still be informative, and the principal would save on traders' rent.[28]

[28] This argument is just a version of the standard rent-efficiency trade-off in agency problems.

Third, we assume that the trader cannot manipulate the duration of pre-trade control. If he could, the compliance incentive compatibility constraint (10) would also depend on the deadline T. Greater time pressure would then increase the cost of incentivizing compliance per se, thereby amplifying the effect time pressure already has on w_+ and w_0 through the task conflict and the initiative compatibility constraint (9) in our current setting.

6.4 Contractual risk management externalities

Consider now the principal's choice between a hasty and a deliberate strategy, for a given deadline T. As earlier, this decision depends on the net benefit of risk management to the principal,

$$\Delta_a(T) \equiv \Delta(T) - B(T). \tag{12}$$

Since $\Delta(.)$ is increasing in T for $T > \iota$ and $B(.)$ is decreasing, time pressure – a decrease in T – now attacks risk management on two fronts: Externally, time pressure renders the deliberate strategy less profitable to the whole firm relative to the hasty one (for $T > \iota$). Internally, time pressure forces the firm to allocate a larger share of the (shrinking) profit to its traders *as long as* it runs risk management. Both of these effects undermine risk management, which is reminiscent of the "attribution problem" in practice where the key controversy in the wake of risk management failures is invariably the question whether the firm should be held liable for lax controls or rogue traders for misconduct.

To stress this point, we isolate the effect of the agency problem by considering the limit where risk management latency vanishes, i.e., ι goes to 0. At this limit, risk management is a strictly dominant strategy in the absence of agency frictions, i.e., $\Delta(T) > 0$ for every $T > 0$. Yet, with agency, $\Delta_a(T)$ can be negative if the agency rent $B(T)$ is sufficiently large. That is, agency creates an additional cost of risk management for the principal. Furthermore, by Lemma 3, this cost increases with time pressure (i.e., lower T). Conversely, when principals give up on incentivizing risk management, time pressure increases (i.e., T decreases): Even when risk management is not intrinsically time-consuming ($\iota = 0$), trading opportunities deplete faster in the absence of risk management as hasty traders execute trades irrespective of their private value α_k, rather than only when α_k is positive. This feedback loop creates strategic complementarities that sustain multiple self-fulfilling equilibria as in Proposition 1.

Proposition 5. *If trading is delegated, there exists an interval $[\underline{\pi}_a, \overline{\pi}_a]$ such that the hasty equilibrium exists for all $\pi \geq \underline{\pi}_a$ and the deliberate equilibrium exists for all $\pi \leq \overline{\pi}_a$. The interval $[\underline{\pi}_a, \overline{\pi}_a]$ is non-empty even if $\iota = 0$.*

From Proposition 5, the coordination problem among firms with delegated trading persists even if risk management latency is arbitrarily small. Indeed, the deliberate equilibrium Pareto-dominates the hasty one provided that both co-exist: For any principal k, the efficiency gain from learning about α_k exceeds the cost of incentivizing risk management, given that trading in other firms is deliberate and hence execution slow enough. Still, principals can be trapped in the hasty equilibrium even if $\iota = 0$ because the agency problem introduces a new externality into the model: Each principal ignores the impact his risk management choice has on the agency rent that other principals must leave with their agents to provide incentives for risk management. In the general case where ι can be strictly positive, agency amplifies a race to the bottom not only through the "partial equilibrium" effect of making risk management costlier inside each firm but also through a "general equilibrium" effect of firms making it costlier *for each other*. While the former speaks to the shape of the "second-best" contract inside a firm, it is the latter that contributes to the constrained inefficiency of the market outcome, and which is reminiscent of claims of a *self-reinforcing* "front-office culture" in financial markets.

Several regulatory implications emerge from our analysis. First, in light of the distributional effects of the contractual externality, principals might not (prefer to) coordinate on risk management *even if they could* because of the agency rents. This increases the need for regulation. Second, the attribution problem extends to the "general equilibrium" effects in that the two externalities reinforce each other. This implies that targeting the opportunity costs and the agency rents of risk management in tandem potentiates the impact of either measure. Third, alternative market regimes (deliberate vs. hasty) are associated with markedly different incentive structures within firms. Processes inside firms thus offer targets for intervention, i.e., *risk management regulation.* To promote deliberate markets, regulators could supervise internal governance mechanisms, such as pre-trade controls (e.g., position limits), post-trade controls (e.g., valuation controls), and compensation (e.g., clawbacks).[29] We discuss regulatory approaches in more detail in Section 8.

7 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 be also 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 Appendix B).

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 the market. In fact, 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: There is now more trading in expectation in a hasty equilibrium than in a deliberate one. Interestingly, our model suggests that a higher trading volume can manifest a worse allocation of risks in the market. Put differently, there is a sense in which trading can be "excessive" in that it lowers welfare.

Alternative modeling of trading. Our model of trading is stylized but could accommodate more

[29] For example, following the "London Whale" scandal that revealed deep flaws in its risk management system, J.P. Morgan Chase has invested close to $1 billion in 2013 on strengthening internal controls and assigned more than 5,000 employees to compliance, arguably in response to the regulatory fine of (also) roughly $1 billion and regulatory pressure.

"realistic" features. 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: It allows efficient *re*-allocations, but it also lowers the private value of 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.

On-the-fly risk management choice. In our 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 Appendix C, 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 assets.

Continuous risk management choice. Also, the risk management choice in our model 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. Strategic complementarities, and hence constrained inefficient coordination failures, would continue to exist. The difference is that some of these equilibria could be stable and *interior*. If the firms' best responses to time pressure are unique (i.e., a function), all equilibria are symmetric in that firms choose identical strategies but some may feature an interior level of risk management quality. If best responses are non-unique (i.e., a correspondence), there may further be equilibria in which firms choose heterogenous levels of risk management quality.[30]

8 Risk management regulation

In our model banks must balance "business needs and risk appetite" (Ernst and Young, 2013, 12) to be competitive in the market. Yet, while a firm should not be criticized for adapting to competition

[30]Naturally, such equilibria can also exist in a model in which differences, e.g., in speed are exogenous.

and market conditions, if each bank freely strikes its optimal balance the aggregate outcome can be constrained inefficient. This makes it debatable whether the trade-off should be left to banks.

Posner and Weyl (2013a,b) argue that financial regulation should improve the risk-sharing function of markets and reduce speculative trading. In a similar vein, the objective of regulation in our model is to counter inefficient risk allocations and excessive trading. Below, we discuss a variety of approaches that have been part of the public debate on financial regulation or have been adopted by regulators: Pigouvian taxes, governance regulations, and interventions in market design.[31]

Pigouvian taxes. In the constrained inefficient outcome of our model, firms trade too much. One countermeasure is hence a transactions tax.[32] Suppose firms must pay a tax τ for every trade. This reduces 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 renders trades for which $\hat{\pi} + \alpha_+ < 0$ unprofitable. The tax is thus a double-edged sword: It can deter excessive as well as valuable trade.

Since the probability of the hasty equilibrium increases in the external-internal speed ratio $\sigma \equiv \iota/\lambda$, another approach is to "tax" (investment in) speed.[33] Suppose regulators can somehow lower external speed (increase λ). Again, the effect is ambiguous: While the decrease in time pressure promotes risk management, immediacy and trading volume decrease for valuable trades.

In contrast, subsidizing internal speed (decreasing ι) promotes both speed *and* risk management. Technological investment in infrastructure to automate information aggregation inside firms is indeed a top risk management priority (Ernst and 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... So there's a huge, multiyear, gazillion-dollar effort."

In our model, such investments (in practice largely a response to regulatory requirements) amount to contributions to a "public good," providing a rationale for public expenditure on risk management infrastructure.

Governance regulation. Our model identifies two distinct but mutually reinforcing sources of risk management failure: opportunity costs and agency costs. This suggests that a regulatory framework must address risk management protocols in conjunction with incentive compensation, which accords with the view taken by the Federal Reserve in its Guidance on Sound Incentive Compensation Policies:

> The final guidance recognizes that strong 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

[31] We do not discuss capital requirements or bailout policies, which have no effect in our model since the inefficiencies are not driven by excessive risk taking incentives due to "leverage" or limited liability.

[32] Tobin (1983) proposed such a tax based on the argument that excess currency speculation undermines the allocative role of exchange rates for hedging purposes, i.e., risk allocation.

[33] Budish, Cramton, and Shim (2013) provide examples of such arms races.

> 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. (Federal Register, 2010, 36401)

The guidelines further describe regulatory initiatives to supervise not only compensation practices (e.g., deferred pay, clawbacks, compliance rewards) but also the associated risk control and governance processes (36397).[34] Such qualitative inspections of risk management are now also central to regulatory assessments of bank capital adequacy ("stress tests"). For example, in 2014, four large banks saw their capital plans rejected on grounds of qualitative deficiencies in their risk management processes (Board of Governors of the Federal Reserve System, 2014).[35]

Our model also speaks to the question of whether institutions or individuals should be penalized for risk management violations. On one hand, individual liability fails to address the problem that firms may *want* lax controls, thereby inviting misbehavior, due to the opportunity costs of risk management. For example, in the London Whale scandal, J.P. Morgan Chase was fined nearly $1 billion for unsound risk protocols:

> 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, the complaint says, 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. (Dayen, 2013)

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

Institutional liability, on the other hand, does not address the agency costs of risk management, which can be prohibitive in competitive markets. Individual liability relaxes the incentive constraint for compliance, making it cheaper for firms to provide compliance and effort incentives simultaneously. In other words, it helps firms overcome the multi-dimensional moral hazard that is at the heart of the contractual externalities in our agency model.

Market design. Aside from supervising processes inside firms, regulators may also want to regulate market processes for two reasons. First, as time pressure rises, algorithmic trading becomes increas-

[34] See also SR letter 12-17, "Consolidated Supervision Framework for Large Financial Institutions," which the Federal Reserve issued in 2012.

[35] 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" (7).

ingly attractive, since machines are fast and require no agency rents to perform their function. Still, the decrease in human involvement does not preclude "underinvestment" in risk management; in fact, Kirilenko and Lo (2013, 70) offer arguments to the contrary and suggest *systemwide* regulation that should "be translated into computer code and executed by automated systems,... approach automated markets as complex systems... [and] encourage safeguards at multiple levels of the system." They also suggest that trading algorithms be available to regulators for inspection and testing.

Second, even if the trading algorithms were accessible to regulators, the degree to which computer codes sidestep (counterfactual) risk controls may be difficult to assess; all the while, firms may continue the "speed race." Interventions into *market* processes may hence be a useful complementary measure. Kirilenko and Lo (2013) and Budish et al. (2013) discuss the possibility of discretizing trading time to eliminate preemption at the microsecond level. Another intervention that recognizes the coordination problem is to build safeguards, such as trade limits and pre-trade risk controls, into the system at the level of intermediaries or central counterparties, such as dealer-brokers, clearing houses, or exchanges (Clark, 2012).[36]

To place the rationale for such interventions into context, consider Grossman and Miller (1988)'s framework in which the equilibrium market structure is the outcome of a 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.[37] In our setting, immediacy can invoke a race to the bottom with respect to risk management, which is why interventions in market design to reduce immediacy can be Pareto-improving.

Regulatory competition. Clearly, if competing firms can be regulated in different jurisdictions, regulators inherit the race-to-the-bottom incentives identified in our model insofar as that they care about "their" firms' competitiveness. Indeed, there is awareness about the need for regulatory coordination. For example, Clark (2010) writes,

> 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 Federal Reserve's Guidance on Sound Incentive Compensation Policies (Federal Register, 2010, 36399).

[36]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).

[37]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.

9 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 in markets characterized by time pressure due to preemptive competition.

Our framework could be extended in several directions. First, we have shown that 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 decisions, such as infrastructure or organizational choices, in a variety of contexts with time-based competition other than financial markets, thereby expanding the applicability of the global games apparatus that has been developed for models of panics.

Second, models with costly monitoring or state verification 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 rather general and clearly applicable beyond risk management. In particular, as we have shown, it can naturally create a trade-off between monitoring (by the principal) and initiative (by the agent), similar to tensions analyzed in the literature on delegation, which depends on time pressure. This offers several avenues for further research.

Finally, financial markets seem a natural context for the speed-information trade-off at the heart of our model. While this paper has focused on risk management, there are other questions our tractable framework could be modified to analyze. For example, market efficiency is a topic to which our model can possibly be extended to study how traders' endogenous choices between speed and learning affect competition, specialization, and information aggregation. One could also examine said trade-off from the perspective of other market participants, such as managers that learn from prices but also disclose information or securities exchanges that can affect the speed at which trading unfolds.

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.

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.

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

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

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.

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

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).

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

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.

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

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.

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.

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

Metha, 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).

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

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

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

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.

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

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

Appendix A: Proofs

Proof of Lemma 1

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

Claim 1.

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,

$$\begin{aligned}\Delta(T) &= \rho(\pi + \alpha_+)p_d(T) - \pi p_h(T) \\ &= \rho(\pi + \alpha_+) \max\left\{0, 1 - e^{-(T-\iota)/\lambda}\right\} - \pi\left(1 - e^{-T/\lambda}\right) \end{aligned} \tag{13}$$

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 of Claim 1.

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 of Claim 1.

Finally, for any T, $p_h(T) > \rho p_d(T)$, which, from (13) 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 around T^*,

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

From point 2 in Claim 1, $\Delta'(T^*) > 0$. Furthermore, from point 4, $\frac{\partial \Delta}{\partial \pi} < 0$. Therefore, $\frac{\partial T^*}{\partial \pi} > 0$.

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 start with a few definitions.

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

$$q(\pi, \hat{s}) \equiv \begin{cases} 0 & \text{if } \pi \leq \hat{s} - \varepsilon, \\ \dfrac{\pi + \varepsilon - \hat{s}}{2\varepsilon} & \text{if } \hat{s} - \varepsilon < \pi < \hat{s} + \varepsilon, \\ 1 & \text{if } \pi \geq \hat{s} + \varepsilon \end{cases} \tag{14}$$

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

$$u(s_k,\hat{s}) \equiv \mathbb{E}_\pi\{\Delta[\pi,\tau(\pi,\hat{s})]|s_k\} = \frac{1}{2\varepsilon}\int_{s_k-\varepsilon}^{s_k+\varepsilon}\Delta[\pi,\tau(\pi,\hat{s})]d\pi, \tag{15}$$

Step 1: Existence of a unique threshold equilibrium.

We start with a simple result

Claim 2. *$\tau(\pi,\hat{s})$ 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 (14), $q(\pi+a,\hat{s}+a)=q(\pi,\hat{s})$, which in turn implies $\tau(\pi+a,\hat{s}+a)=\tau(\pi,\hat{s})$. □

As mentioned in the main text, a necessary condition for a threshold strategy s^* to be an equilibrium is for the following claim to be true.

Claim 3. *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 $\pi\in[s-\varepsilon,s+\varepsilon]$, $T^*(\pi)<T_h\le\tau(\pi,s)$, 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 2 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 4. *$u(s,s^*)>0$ for $s<s^*$ and $u(s,s^*)<0$ for $s>s^*$.*

Proof. From (15), $u(s^*,s^*)=0$ 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, $\tau(\hat{\pi},s^*)=T^*(\hat{\pi})>0$.[38] Suppose $\pi<\hat{\pi}$, then $\tau(\pi,s^*)\ge T^*(\hat{\pi})$, and therefore using the single crossing property (Lemma

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

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.

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,

$$\frac{1}{2\varepsilon}\int_{s_A-\varepsilon}^{s_A+\varepsilon} \Delta[\pi, \tau(\pi)]d\pi = \frac{1}{2\varepsilon}\int_{s_B-\varepsilon}^{s_B+\varepsilon} \Delta[\pi, \tau(\pi)]d\pi. \tag{16}$$

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 (16) 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 (16), one obtains

$$\overline{u}[s_A, \tau(.)] \equiv \frac{1}{2\varepsilon}\int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi, \tau(\pi)]d\pi = \frac{1}{2\varepsilon}\int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi, \tau(\pi)]d\pi \equiv \overline{u}[s_B, \tau(.)]. \tag{17}$$

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

$$\overline{u}[s_B, \tau(.)] = v_B[\pi_B, \tau(.)] < v_B[\pi_A, \tau(.)]. \tag{18}$$

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 (18) contradicts (17) and hence, (16).

ï£ · 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 5. *$\overleftarrow{\tau}(.)$ 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. $\square$

Claim 6. *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 5 completes the proof. $\square$

Note that if $\Delta(\pi,.)$ was monotonically increasing (that is, under global strategic complementarities), Claim 6 would directly imply that $v_A[\pi_A,\tau(.)] \geq v_B[\pi_A,\tau(.)]$.

Claim 7. $v_A[\pi_A,\tau(.)] \geq 0$.

Proof. Note first that the monotonicity result in Claim 5 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 $v_A[\pi_A,\tau(.)] \geq 0$. $\square$

We want to show

$$v_A[\pi_A,\tau(.)] - v_B[\pi_A,\tau(.)] = \int_{s_A-\varepsilon}^{\overline{s}_A} \Delta[\pi_A,\overleftarrow{\tau}(\pi)]d\pi - \int_{\underline{s}_B}^{s_B+\varepsilon} \Delta[\pi_A,\tau(\pi)]d\pi \geq 0 \tag{19}$$

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 7, $v_A[\pi_A,\tau(.)] \geq 0$, (19) holds.

Suppose that $\tau(\underline{s}_B) \geq T^*(\pi_A)$.[39] Then $\Delta(\pi_A,.)$ strictly increasing for $T > T^*(\pi_A)$ and Claim 6 imply that (19) holds.

[39]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.

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 5, $\overleftarrow{\pi}(.)$ and $\pi(.)$ are increasing and $\pi(.)$ increases faster than $\overleftarrow{\pi}(.)$. Using this notation and Claim 6, rewrite (19):

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

$$= \quad - \int_{\tau(\underline{s}_B)}^{\tau(s_A-\varepsilon)} \Delta[\pi_A, T]d\pi(T) \tag{20}$$

$$+ \int_{\tau(s_A-\varepsilon)}^{T^*(\pi_A)} \Delta[\pi_A, T]d\overleftarrow{\pi}(T) - \int_{\tau(s_A-\varepsilon)}^{T^*(\pi_A)} \Delta[\pi_A, T]d\pi(T) \tag{21}$$

$$+ \int_{\overleftarrow{\pi}[T^*(\pi_A)]}^{\overline{s}_A} \Delta[\pi_A, \overleftarrow{\tau}(\pi)]d\pi - \int_{\pi[T^*(\pi_A)]}^{s_B+\varepsilon} \Delta[\pi_A, \tau(\pi)]d\pi. \tag{22}$$

This equation is a decomposition of the two integrals in (19) 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, (20) 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 (20) 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 (21) is positive (Intuitively, v_B visits any negative values of Δ that v_A takes but "stays longer" at each of them.) Finally, (22) corresponds to values of T above T^*. Note first that $\tau(s_A - \varepsilon) \geq \tau(\underline{s}_B)$ (Claim 6) 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 5, 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, (22) is positive (intuitively, $\overleftarrow{\tau}(.)$ shifts more weight towards high values of Δ than $\tau(.)$.)

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 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.

Let $\hat{\tau}(q, i)$ be defined as in (6), with the addition of the second argument explicitly recognizing its dependence on i. Let

$$U(\pi) \equiv \int_0^1 \Delta[\pi, \hat{\tau}(q, i)]dq.$$

Note that π^* solves $U(\pi) = 0$. Note also that $U'(.) < 0$, and from (6), $\hat{\tau}(q, i)$ is increasing in i. Consider two cases,

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

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

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

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

$$\hat{q}(i) \equiv \frac{i}{p_h(\iota)}. \tag{23}$$

$$\begin{aligned} U(\pi^*) &= \int_0^{\hat{q}(i)} \{\rho(\pi^* + \alpha_+) p_d[\hat{\tau}(q, i)] - \pi^* p_h[\hat{\tau}(q, i)]\} dq - \int_{\hat{q}(i)}^1 \pi^* p_h[\hat{\tau}(q, i)] dq \\ &= \hat{q}(i)\rho(\pi^* + \alpha_+) - \pi^* - \int_0^{\hat{q}(i)} \{\rho(\pi^* + \alpha_+)[1 - p_d[\hat{\tau}(q, i)]] - \pi^*[1 - p_h[\hat{\tau}(q, i)]]\} dq \\ &\quad + \int_{\hat{q}(i)}^1 \pi^*[1 - p_h[\hat{\tau}(q, i)]] dq \end{aligned} \tag{24}$$

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

$$1 - p_d[\hat{\tau}(q, i)] = e^{\iota/\lambda}[1 - p_h[\hat{\tau}(q, i)]] = e^{\iota/\lambda} \frac{q + (1 - q)\rho - i}{q + (1 - q)\rho e^{\iota/\lambda}}. \tag{25}$$

Hence, the first integral in (24) becomes

$$[\rho e^{\iota/\lambda}(\pi^* + \alpha_+) - \pi^*] \int_0^{\hat{q}(i)} \frac{q + (1 - q)\rho - i}{q + (1 - q)\rho e^{\iota/\lambda}} dq. \tag{26}$$

If $q > \hat{q}(i)$, then $p_d[\hat{\tau}(q, i)] = 0$. Using this in $qp_h[\hat{\tau}(q, i)] + (1 - q)\rho p_d[\hat{\tau}(q, i)] = i$, we get $1 - p_h[\hat{\tau}(q, i)] = \frac{q-i}{q}$. Hence, the second integral in (24) becomes

$$\pi^* \int_{\hat{q}(i)}^1 \frac{q - i}{q} dq \tag{27}$$

Consider the first line of equation (24). $U(\pi^*)$ depends on i both through the boundaries of the integrals (via $\hat{q}(i)$) and through the integrands (via $\hat{\tau}(q, i)$). However, since $p_d[\hat{\tau}(\hat{q}(i)), i] = 0$, the effect of a marginal change in i that goes through $\hat{q}(i)$ cancels out. Hence, using (25) and (26) to substitute into the second line of (24), we obtain

$$\frac{\partial U}{\partial i}(\pi^*) = [\rho e^{\iota/\lambda}(\pi^* + \alpha_+) - \pi^*] \int_0^{\hat{q}(i)} \frac{1}{q + (1-q)\rho e^{\iota/\lambda}} dq - \pi^* \int_{\hat{q}(i)}^1 \frac{1}{q} dq. \tag{28}$$

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

$$\frac{1}{q} = \frac{p_h[(\hat{\tau}(q)]}{i}, \tag{29}$$

In addition, rearranging (25),

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

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

$$\frac{p_d[\hat{\tau}(q)]}{e^{\iota/\lambda} i} < \frac{1}{q + (1-q)\rho e^{\iota/\lambda}} < \frac{p_h[\hat{\tau}(q)]}{i}. \tag{30}$$

Finally, using (28), (29) and (30),

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

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 $\sigma = \frac{\iota}{\lambda}$, which together with the discussion in the main text, proves Proposition 3.

As in (23), let

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

We have

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

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, $p_d[\hat{\tau}(\hat{q}(\iota/\lambda))] = 0$, 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 (24), (26) and (27), one obtains

$$\frac{\partial U(\pi^*)}{\partial x} = -\frac{\partial}{\partial x} \left\{ [\rho x(\pi^* + \alpha_+) - \pi^*] \int_0^{\min\{\hat{q}(\iota/\lambda),1\}} \frac{q + (1-q)\rho - i}{q + (1-q)\rho x} dq \right\}. \tag{31}$$

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. $\square$

Appendix B: Exogenous deadline

Suppose that the trading opportunity disappears at the first of these to 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:

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

Not 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,

$$\lim_{T\to+\infty} \Delta(T) = \frac{\chi}{\lambda+\chi}\left[\pi-\rho e^{-\frac{\iota}{\chi}}(\pi+\alpha_+)\right], \tag{32}$$

and hence, if the expression between brackets is positive, then RM is never profitable, even when T is arbitrarily large. That is, rearranging (32), 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

$$\begin{aligned} &\left[\int_\iota^{T_d} \frac{1}{\chi} e^{-\frac{t}{\chi}}\left(1-e^{-(t-\iota)/\lambda}\right)dt + e^{-\frac{T_d}{\chi}} I\right]\rho(\pi+\alpha_+) \\ &-\left[\int_0^{T_h} \frac{1}{\chi} e^{-\frac{t}{\chi}}\left(1-e^{-t/\lambda}\right)dt + e^{-\frac{T_h}{\chi}} I\right]\pi \end{aligned} \tag{33}$$

For every realization of π, there exists ι sufficiently large relative χ that (33) 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.

Appendix C: Decision making in continuous time

We consider here a variation of the baseline model in Section 4 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.[40] 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

[40] 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.

the following result.

Proposition 6. *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.*

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,

$$1 - e^{-T_h^\infty/\lambda} - \rho\left(1 - \frac{\frac{1}{\iota}e^{-T_h^\infty/\lambda} - \lambda^{-1}e^{-\frac{1}{\iota}T_h^\infty}}{\frac{1}{\iota} - \lambda^{-1}}\right) = s_0. \tag{34}$$

(34) 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

$$T = \max_k \tau_k \tag{35}$$

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 6 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.