THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Alessio OZANNE

Le 1 juillet 2025

essais en économie financière et en économie de l'information

École doctorale : **TSE Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur Matthieu BOUVARD et M. Alexander GUEMBEL*

Composition du jury

Rapporteur : M. Ansgar WALTHER
Rapporteure : Mme Jing ZENG
Directeur de thèse : M. Matthieu BOUVARD
Co-directeur de thèse : M. Alexander GUEMBEL

Essays in Financial and Information Economics

by
Alessio Ozanne

Submitted in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy in Economics

Toulouse School of Economics
Toulouse Capitole University
Toulouse, France

May 2025

Examining Committee Membership

Supervisors:

Matthieu Bouvard
Professor
Toulouse School of Economics, Toulouse School of Management
Toulouse Capitole University

Alexander Guembel
Professor
Toulouse School of Economics, Toulouse School of Management
Toulouse Capitole University

External Members:

Ansgar Walther
Professor
Department of Finance
Imperial College Business School

Jing Zeng
Professor
Department of Economics
University of Bonn

Abstract

This dissertation explores how the design of information and incentives can mitigate inefficiencies in financial markets. It consists of three essays at the intersection of Financial and Information Economics. The first chapter studies the optimal transparency of credit scoring algorithms, showing that opacity can improve data sharing and credit access by reducing borrowers' strategic behavior. The second chapter analyzes the design of bank stress tests when market discipline complements supervision, demonstrating that coarse and lenient tests optimally leverage private information embedded in market prices. The third chapter examines risk governance under preemptive competition among financial firms, proposing incentive schemes that align firms' trading and compliance activities with efficient risk management. Methodologically, the first two chapters employ tools from information design and Bayesian persuasion, while the third draws on mechanism design and contract theory. Together, these essays highlight the central role of information structures and incentive mechanisms in shaping financial market outcomes and provide policy-relevant insights for the regulation of credit markets, banking supervision, and financial risk management.

Table of Contents

Chapter 1

Introduction

Financial markets are often plagued by inefficiencies driven by information asymmetries and distorted incentives. Unequal access to information often results in suboptimal outcomes, either because uninformed participants misjudge economic conditions or because better-informed agents alter their behavior to preserve or exploit their informational advantage. Similarly, when individual incentives diverge from social welfare and actions cannot be effectively monitored, market participants may prioritize private gains at the expense of collective efficiency. These frictions can disrupt price discovery, misallocate capital, encourage excessive risk-taking, and ultimately lead to welfare losses, market failures, and systemic instability. In such contexts, the structure of information flows and incentive schemes plays a central role in shaping individual behavior and, by extension, overall market performance. Understanding how regulators and institutions can correct these distortions without creating new ones remains one of the central challenges in financial economics.

This dissertation investigates how the strategic design of disclosure policies and incentive schemes can address key frictions in financial markets and better align individual behavior with socially optimal outcomes. The research employs both modern and classical tools from information economics to tackle three topical questions in finance: Should credit scoring algorithms be *explainable*? How transparent should stress tests be? How should traders be compensated? Recent developments in information design and Bayesian persuasion provide a natural framework for the first two questions, while classical mechanism design and contract theory inform the third. As Bergemann and Morris (2016, 2019) highlight, mechanism design selects the rules of the game based on a given information structure to achieve desired outcomes, while information design optimizes the information structure while keeping the rules fixed. This unified approach to economic design is especially relevant to finance, where regulators have the ability to shape both information flows and incentive structures, influencing the behavior of strategic agents and enhancing market efficiency.

The thesis comprises three independent papers, each addressing one of the questions outlined above. Chapter 2, titled **Black-Box Credit Scoring and Data Sharing**, investigates the optimal level of transparency in algorithmic credit markets. Advances in digital technology have the potential to mitigate adverse selection by enabling lenders to process vast amounts of personal data using complex machine learning models. However, the opaque, black-box nature of these algorithms has sparked significant public concerns about privacy and fairness, prompting regulatory scrutiny. This paper explores whether lenders should disclose their predictive algorithms or keep them opaque, taking into account borrowers' strategic data-sharing behavior. It finds that opacity can induce borrowers to withhold information as a safeguard against unpredictable algorithmic decisions, thereby depriving the technology of its vital fuel and weakening its ability to address asymmetric information problems. Conversely, transparency incentivizes borrowers to game the system by selectively hiding negative information, potentially impairing the screening process and the resulting credit allocation. The paper characterizes conditions under which either regime is optimal for lenders and socially efficient, identifying a wedge between the two. Contrary to conventional wisdom, algorithmic opacity can improve welfare by mitigating negative externalities on creditworthy yet privacy-conscious borrowers, typically operating through the lender's equilibrium inference. Methodologically, the analysis models the algorithmic disclosure problem as a Bayesian persuasion game, where the lender commits to either revealing or concealing the algorithm's parameters, and the borrower strategically decides whether to share data. Conceptually, the paper's main contribution is to show how information design, through algorithmic opacity, can mitigate adverse selection in credit markets.

Chapter 3, titled **Market Information in Banking Supervision: the Role of Stress Test Design**, is coauthored with Haina Ding and Alexander Guembel, and recently received a revise-and-resubmit decision from the *The Review of Financial Studies*. The paper examines bank stress tests and market discipline — the monitoring of financial institutions by financial markets. Since the 2008 financial crisis, stress tests have become a central supervisory tool for evaluating banks' resilience under adverse financial scenarios. At the same time, the Basel Committee on Banking Supervision acknowledges that financial markets can generate information that regulators do not have. This chapter explores how a supervisor should design and disclose stress-test results when banks' security prices provide additional market feedback, jointly influencing the bank creditors' rollover decisions. The analysis adopts an information design framework in which the regulator selects what to reveal about banks' stress resilience from a broad class of possible disclosure mechanisms. The key finding is that the optimal stress test is deliberately coarse and lenient, often taking the form of a simple pass/fail test. Introducing ambiguity generates profit opportunities for private investors to acquire information and identify institutions that may have been misclassified by the test. Thus, as residual market uncertainty increases, both trading profits and the accuracy of market signals improve, thereby justifying the optimality of a coarse test. We also show that stock markets are better at reducing Type II errors (failing to identify weak banks) but less effective at Type I errors (flagging healthy banks). Failing banks are less likely to receive funding and remain small, limiting trading incentives. In contrast, passing banks attract funding and grow larger, increasing the value of market scrutiny. This endogenously creates a motive for supervisory leniency, deliberately allowing some weak banks to pass in order to spur market scrutiny. The chapter makes two main contributions. First, it offers a tractable stress test design framework emphasizing the complementarity between public supervision and private market monitoring, treating both banks' values and market prices as endogenous to the information revealed by both sources. Second, it enriches the literature on stress testing by showing — paradoxically — that regulatory opacity can enhance market outcomes and capital allocation by optimally harnessing market discipline.

Chapter 4, **Agents under Pressure: Risk Governance in a Rat Race** coauthored with Matthieu Bouvard and Samuel Lee, examines incentive structures for risk governance in financial intermediaries. While efficient risk-taking is vital for effective prudential regulation, poorly designed compensation schemes within firms can undermine sound risk management. In addition, in today's fast-paced financial markets, preemptive competition tends to reward speed over caution, further eroding incentives for deliberate, prudent risk assessment. This chapter frames compensation design for financial traders as a systemic issue, where competitive pressure influences firm-level risk governance, and vice versa. It develops a principal-agent model in which traders, employed by competing firms, exert costly effort along two dimensions: searching for trading opportunities and assessing their risk. Importantly, the latter task involves delays, introducing an endogenous cost of (incentivizing) risk management, stemming from non-execution risk due to preemption by competitors. In the partial equilibrium analysis, we show that incentivizing both tasks creates countervailing demands on the contract shape: rewarding search requires linking pay to trading profits to encourage execution, whereas incentivizing compliance calls for decoupling pay from profits to dampen the urge to execute discovered opportunities. The optimal contract must balance these conflicting forces via pay-for-performance bonuses paired with claw-backs contingent on detected cheating. Crucially, preemption risk increases the firm's agency cost of incentivizing risk management, leading firms to weaken risk governance as competition intensifies. In general equilibrium, this dynamic creates a contractual externality: as some firms relax controls, the resulting increase in time pressure raises the cost of risk governance for others, triggering a race to the bottom that can trap the industry in a low-control regime, even when robust risk assessment is collectively optimal. The chapter thus uncovers a novel self-reinforcing channel through which agency frictions and competition jointly erode risk management, offering insights for the regulation of traders' compensations and their calibration on firms' trading environment.

Chapter 2

Black Box Credit Scoring and Data Sharing

Black Box Credit Scoring and Data Sharing

Alessio Ozanne

Abstract

Should credit scoring algorithms be transparent or opaque? I study this question in a model in which a lender uses data shared by borrowers for allocating and pricing credit, and is privately informed about how the algorithm maps data to allocations. I show that revealing the algorithm's parameters exposes it to gaming in the form of strategic withholding of unfavorable information. Under opacity, data withholding emerges as a prudent strategy against the unpredictability of the black box and the risk of credit rationing. The lender's optimal transparency regime maximizes data collection and is socially inefficient as it results in excessive credit rationing. Algorithmic opacity can be welfare-improving as it reduces the stigma around data withholding, thereby expanding credit access for privacy-concerned borrowers. I analyze the distributional impacts of recent algorithmic transparency regulations and offer policy recommendations.

Keywords: FinTechs, Data, Disclosure, Algorithms, Opacity.
JEL Classification: D82, G14, G21, G38.

2.1 Introduction

Motivation Recent advances in digital technology transformed how firms store, transmit, and analyze information. Notably, a growing number of financial institutions have embraced automated machine learning (ML) algorithms for the task of credit scoring and underwriting,[1] leveraging vast datasets that encompass both traditional financial metrics and alternative data sources.[2] ML algorithms, particularly when fed with alternative data, often reveal surprising relations between credit risk and seemingly unrelated variables, making it challenging to understand which specific factors influence algorithmic decisions,

[1]A prominent example is FinTech's "3-1-0 model": three minutes to apply, one second to approve, and zero human intervention. Traditional banks — such as J.P. Morgan and Bank of America — and credit scoring systems — such as the FICO and Vantage Score — also started to employ automated algorithms. Automated credit underwriting is mainly applied to (unsecured) consumer and small business loans, instead of mortgages.

[2]Alternative sources of data include, but are not limited to, social media data, digital footprints, web searches, mobile phone usage patterns, e-commerce transactions, payment information, gig economy income, utility bill payments, childcare payments, and data from IoT devices. To put it in the industry's words, "all data is credit data" (Douglas Mirrell, CEO of ZestCash to the New York Times). Alternative data have demonstrated superior performance compared to conventional FICO scores (see Iyer et al. (2016), Bazarbash (2019), and Berg et al. (2020)).

and how. This phenomenon is known as the "black box" problem of artificial intelligence (AI).[3]

Algorithmic opacity has raised significant public concern, fostering uncertainty and mistrust in the lending process even with accusations of discriminatory practices,[4] and prompted individuals to conceal their online and offline behaviors, thereby restricting the information available to lenders.[5] On the other hand, making the algorithms' inner working transparent could undermine their predictive ability, enabling borrowers to game the system through strategic information sharing.[6] Both regimes impact borrowers' willingness to share data — a crucial asset in a market plagued by asymmetric information — and may have a first-order impact on the allocative efficiency of credit markets. Should credit risk algorithms be transparent or opaque when we account for strategic information sharing? Regulators around the globe have decided to unpack the black box mandating transparency,[7] citing principles of truth, fairness, and equity, but possibly overlooking the importance of market efficiency.

Overview This paper explores the lender-optimal and the socially efficient level of transparency for credit risk algorithms in light of the information that borrowers are willing to share with a financial institution. The model features a lender and a continuum of borrowers in need of cash to finance a project. Borrowers are, on average, creditworthy but differ in credit risk, with some having negative NPV projects. Additionally, borrowers possess data that can predict their probability of default. According to recent data regulations — such as Open Banking initiatives, the General Data Protection Regulation (GDPR), and the Payment Services Directives (PSD, PSD2) — this data remains private unless borrowers voluntarily share it with the lender. Furthermore, some borrowers are assumed to prioritize privacy and consistently withhold information, regardless of what it might reveal about them.

The black box nature of the screening process arises from how the lender infers credit risk from data. The data-generating process (DGP) is unknown, while the lender can estimate its parameters from a previously collected training dataset, leveraging proprietary statistical technology. The allocation algorithm is calibrated on these privately known estimates, so the lender has exclusive knowledge of the mapping from disclosed data to

[3] See Pasquale (2015) for an investigation into our "Black Box Society".

[4] See Bartlett et al. (2022) for empirical evidence on discriminatory lending practices by FinTech lenders, and for a recent controversy see Neil Vigdor, "Apple Card Investigated After Gender Discrimination Complaints", The New York Times, November 11, 2019. See Glikson and Woolley (2020) for a study on human trust in AI.

[5] Sebastian Siemiatkowski, CEO of Klarna AB, once said: "Facebook is only as valuable as the information a consumer is willing to share and whether that consumer is willing to connect the financial service to their Facebook data". (See Evelyn M. Rusli, "Bad Credit? Start Tweeting", Wall Street Journal, April 1, 2013.))

[6] In a 2014 letter to the Consumer Financial Protection Bureau (CFPB), Freddie Mac raised concerns that disclosing their automated decision-making algorithm could make it vulnerable to being reverse-engineered (letter available here). See also Suzanne Woolley, "How More Americans Are Getting a Perfect Credit Score", Bloomberg, August 14, 2017.

[7] See the AI Act, Digital Services Act, Digital Markets Act of the European Commission; the Fair Credit Reporting Act and Equal Credit Opportunity Act of the Federal Trade Commission.

allocations. The DGP is assumed to be characterized by a single parameter, representing the correlation between credit risk and an explanatory variable. Hence, evidence can be represented on a normalized interval of the real line, and while borrowers may not discern good from bad news, they can identify extreme evidence based on its distance to the mean.[8] Within this framework, the paper investigates borrowers' voluntary data-sharing decisions under two scenarios: when the algorithm's parameter is kept secret (opacity) and publicly known (transparency). Additionally, it explores the welfare and redistributive implications of algorithmic opacity and transparency, including the regime that maximizes the lender's profits, borrowers' surplus, and overall social welfare.

Preview of the Results The algorithm allocates credit in an intuitive way: high-risk borrowers are denied credit, while interest rates rise with credit risk. These rates reflect the competitive market rate, with an additional markup that grows with the lender's bargaining power. Still, the algorithm also converts evidence in credit risk estimates, and thus, the map from data to allocations depends on the correlation between data and credit risk. If a borrower's characteristics positively correlate with credit quality, higher values of that variable lead to lower interest rates, while values below a certain threshold are denied credit. A symmetric reasoning applies when correlation is negative: higher values of the explanatory variable are progressively stronger signals of credit risk, and may lead to credit rationing. The algorithm is influenced by the strength of the correlation, not just its sign. As the correlation decreases in absolute value, the algorithm rations credit to a narrower range of data realizations, and the interest rate schedule flattens.

When the algorithm is transparent, borrowers understand the correlation between data and credit risk, allowing them to withhold bad news while selectively sharing good news. This behavior, enabled by transparency, will be referred to as *gaming*. Gaming occurs when privacy-concerned borrowers are sufficiently numerous. In this case, high-risk borrowers can mimic their behavior and withhold data to obtain more favorable contractual terms. This phenomenon happens because the stigma associated with data withholding is slight, as withholding is most likely non-strategic and, therefore, not a strong indicator of poor credit quality. As a result, credit-unworthy borrowers obtain credit by withholding information, while marginally creditworthy borrowers can secure lower interest rates. Conversely, borrowers share positive evidence, because understanding the lender's statistical technology, they anticipate being classified as low risks. The number of privacy-concerned borrowers reflects the extent of strategic borrowers' gaming ability: as this number increases, more borrowers can secure better contractual terms by withholding information. Gaming does not occur if borrowers are primarily strategic and data sufficiently informative. In this case, data withholding becomes a strong indicator of credit risk and leads to credit denial. As a result, only (strategic) observationally creditworthy borrowers are funded and interest rates reflect their observable credit risk.

[8]Examples of alternative data that exhibit these characteristics include the number of social media connections, number of posts and their frequency, engagement rates, total hours spent online per day, browsing diversity, average call duration, text message volume, evening vs daytime call ratio, number of apps used regularly, percentage of purchases made at unusual times, frequency of brand loyalty in purchases, travel frequency (see, e.g. the FICO or the EagleAlpha websites).

When the algorithm is opaque, borrowers do not know the correlation between data and credit risk, and data-withholding emerges as a prudent strategy against the algorithm's unpredictability. Although borrowers are risk-neutral, I term this strategic behavior as *hedging*, because it reflects their attempt to minimize exposure to uncertainty. Hedging is complete — meaning that all evidence is withheld — when the lender's bargaining power is sufficiently strong. In this scenario, the lender offers loans to all borrowers at a uniform rate, as the overall pool of borrowers is, on average, creditworthy. If data were shared, borrowers would expose themselves to the risk of credit rationing — with extreme borrowers being the most exposed — and conditionally on providing credit, the lender would capture most of their surplus through high personalized interest rates. Conversely, when borrowers have greater bargaining power, extreme evidence is disclosed in equilibrium, with the extent of disclosure increasing in the borrower's bargaining power. Although extreme borrowers are most at risk of credit rationing, they also stand to gain the most from favorable interest rates if they manage to secure credit — a benefit that grows with their bargaining power. Consequently, extreme borrowers take the risk of sharing information. Borrowers with less conclusive evidence lack this upside potential and thus stick with the hedge of no disclosure. Overall, the borrowers' relative bargaining power modulates their risk-taking behavior when sharing data with the black box, shifting their decisions from withholding all information to disclosing the most conclusive evidence.

Interestingly, privacy-concerned borrowers receive markedly different treatment depending on whether the algorithm is transparent or opaque. Under transparency, they experience negative externalities due to the disclosure decisions of strategic borrowers. When strategic borrowers are numerous, privacy-concerned borrowers are excluded from credit due to the stigma surrounding data withholding, despite being, on average, creditworthy. This situation does not occur when the algorithm is opaque, as no stigma is associated with data withholding. Since borrowers lack insight into what signals good or bad news, the lender cannot make adverse inferences based on a borrower's choice to withhold information. As a result, privacy-concerned borrowers, along with strategic non-disclosing borrowers, can secure credit.

Both transparency regimes lead to a data loss — either through strategic gaming or hedging behavior — which can undermine the algorithm's predictive accuracy and lead to misallocations of credit. Consequently, the transparency regime significantly influences welfare measures, including the lender's profits, borrowers' welfare, and overall social surplus. The lender's transparency choices maximize data collection, enhancing the algorithm's ability to ration credit effectively.[9] Credit misallocations increase with the data loss and thus amplify with borrowers' gaming ability under transparency (measured by the fraction of privacy-concerned users) and their hedging motives under opacity (captured by the lender's bargaining power). Therefore, the model suggests that in competitive credit markets where privacy concerns are intense, financial institutions should favor opaque algorithms, as hedging motives are limited while the likelihood of gaming is higher. Conversely,

[9] The potential benefits of personalized risk pricing are neglected. This effect is due to the lender's risk neutrality: detailed credit quality information reduces variance but does not impact the expected profits the lender gains from creditworthy borrowers.

transparency should be the optimal choice for lenders in concentrated credit markets with mild privacy concerns.

The lender's transparency choices are socially inefficient and opacity can lead to welfare improvements. Private and social preferences over credit inclusion differ as the lender only partially internalizes the surplus generated by credit provision, resulting in excessive rationing from a social perspective (even with complete information). When the misalignment between the lender's and planner's preferences is moderate, data collection enhances welfare by helping exclude strategic borrowers with negative surplus projects. Still, opacity provides a social benefit that is only partially internalized by the lender: credit inclusion of privacy-concerned yet credit-worthy borrowers. By suppressing the stigma typically associated with data withholding, opacity allows these borrowers to obtain credit and, therefore, is socially efficient. When the misalignment of preferences over credit provision increases, data collection starts to undermine welfare, as the borrowers that the lender seeks to ration hold positive-surplus projects. In this scenario, the welfare-optimal regime minimizes data extraction, making the lender's choices entirely inefficient.

Contrary to common beliefs, algorithmic opacity can be desirable from a societal perspective. This insight challenges the prevailing views in the ongoing policy debate on algorithmic transparency — e.g., the EU's AI Act — by highlighting a social benefit of opacity that regulators may have overlooked. By mitigating the stigma associated with data withholding, algorithmic opacity can promote regulatory goals such as financial inclusion. In contrast, algorithmic transparency regulations may unintentionally harm privacy-concerned borrowers, undermining the privacy protections regulators seek to uphold (e.g., in the EU's GDPR). The model points to allocative efficiency as a critical social objective regulators should pursue alongside ethical and safety considerations.

Outline The remainder of the paper proceeds as follows. Section 2.2 discusses the related literature. Section 2.3 describes the model. Section 2.4 derives the lender's allocation rule. Sections 2.5.1 and 2.5.2 present the data-sharing equilibrium under transparent and opaque algorithms. Section 2.6 compares welfare across transparency regimes. Sections 2.7 and 2.8 discuss respectively empirical and policy implications. All the proofs are relegated to Appendix A.

2.2 Related Literature

FinTech This paper connects to the literature on digital disruption in banking (see Vives (2019) for a review). Studies show that alternative data can outperform traditional metrics in predicting loan defaults (see Iyer et al. (2016), Bazarbash (2019), and Berg et al. (2020)), and expand credit access to individuals without a credit history (see Agarwal et al. (2019) and Gambacorta et al. (2022)). However, the ML methods that process these data can produce unintended distributional effects by capturing structural relationships among variables such as race, income, and gender (see Fuster et al. (2022)). I contribute to this literature by identifying a new channel through which these statistical technologies affect distribution and welfare: the transparency regime they adopt. While some studies have theoretically explored the optimal design of these technologies (see, e.g., Huang (2023), or He et al. (2024), or Blickle et al. (2024)), I examine whether they should be transparent or opaque.

Voluntary Disclosure The paper is related to the extensive literature on voluntary information disclosure. The classical "unraveling argument" put forward by Grossman (1981) and Milgrom (1981) shows that, in equilibrium, a privately informed sender will disclose all the available information to avoid the receiver's averse equilibrium inference. The subsequent literature tried to explain the commonly observed lack of disclosure by introducing disclosure frictions.[10] Disclosure frictions in my model resemble those in Dye (1985) (uncertainty about evidence, or privacy cost in my model) and Bond and Zeng (2022) (uncertainty about audience, or statistical technology in my model). My paper contributes to the literature by comparing these frictions in the specific context of the credit market and conducting a comprehensive welfare analysis. In Bond and Zeng (2022), hedging behavior stems from the sender's risk aversion. In my model, the sender is risk-neutral, while risk-averse or risk-loving behaviors arise endogenously due to the specific market context considered. Unlike Bond and Zeng (2022), I also explore the welfare implications of hedging behavior.

Economics of Data and Privacy This work belongs to a growing literature on data, privacy, and their implications for markets (see Acquisti et al. (2016) and Bergemann and Bonatti (2019) for surveys). A portion of this literature explores the welfare impact of data regulations granting customers rights over sharing personal data, such as the GDPR and Open Banking initiatives. For instance, Ali et al. (2023) study price discrimination under the GDPR, while He et al. (2023) study competition in the lending market under Open Banking, both yielding contrasting results regarding consumer welfare. My work complements these studies by highlighting how data regulation interacts with algorithm transparency regulation, suggesting that regulators should assess both jointly.

Gaming and Opacity The paper is related to the machine learning, economics, and finance literature studying the manipulation of transparent scoring systems.[11] This literature typically considers types of data that can be distorted by agents at some cost. In contrast, I consider data as hard information that can be shared or withheld but not misreported. A common finding in this literature is that an optimal scoring rule should underutilize data to deter manipulation. I show that opacity can emerge as an alternative strategy to data under-utilization.
A different strand of the literature has studied the role of opacity in softening strategic

[10] These frictions include disclosure costs (Verrecchia (1983)), uncertainty about the evidence available to the sender (Dye (1985) and Jung and Kwon (1988)), information processing costs (Fishman and Hagerty (2003)), receiver's outside information (Harbaugh and To (2020) and Quigley and Walther (2024)) and uncertainty about audience's preferences (Bond and Zeng (2022)).

[11] See Hardt et al. (2016) for a classic paper in machine learning, Frankel and Kartik (2019), Frankel and Kartik (2021), Björkegren and Knight (2021), Ball (2025), Perez-Richet and Skreta (2022), and Ekmekci et al. (2022) for recent papers in economics, and Cohn et al. (2024) and Gamba and Hennessy (2024) for recent finance applications.

agents' gaming behavior in various contexts, including incentives schemes (Ederer et al. (2018)), stress testing (Leitner and Williams (2023)), and algorithmic decision-making (Wang et al. (2023) and Sun (2021)). The paper closest to mine is Sun (2021), which examines efficient algorithmic disclosure in a competitive lending market where borrowers can modify a binary feature at a cost. My paper is different in several aspects. First, I include interest rates and show that data-driven pricing exacerbates borrowers' gaming behavior. Second, I introduce the lender's bargaining power, leading to new hedging behavior in borrowers' data provision and enabling an analysis of the lender-optimal transparency regime. Third, I consider voluntary data-sharing, in line with current open banking regulations. Finally, while Sun (2021) indicates that transparency harms the credit market, I highlight its potential benefits in certain contexts, aligning with the current policy debate.

Economics of Algorithms A nascent literature in economics studies how to regulate algorithmic decision-making when private and social objectives diverge (see Korinek and Balwit (2022) for a study of incentives misalignments in AI). A common scenario involves the algorithm aiming to maximize predictive accuracy, while the regulator seeks to balance accuracy with fairness,[12] typically by imposing restrictions on inputs or the prediction function (see, e.g., Kleinberg et al. (2018), Carleton Athey et al. (2020), Rambachan et al. (2021), Blattner et al. (2024), Liang et al. (2024)). My work contributes to this literature by considering a scenario in which the regulator prioritizes both accuracy and credit inclusion, placing a greater weight on the latter than the lender does. I consider the algorithm's transparency regime to be the only policy tool available to the regulator, in light of recent regulation on algorithmic transparency.

2.3 The Model

Economy The economy is composed of a lender and a continuum of penniless borrowers in need of cash to finance a project. All players are risk-neutral, and the borrowers are protected by limited liability. Each borrower has a project that requires a unit of cash and generates a return $X \in \mathbb{R}_+$ with probability $\theta \in [0, 1]$ and generates nothing otherwise. The projects' probability of success is heterogeneous across borrowers and uniformly distributed, $\theta \sim U[0, 1]$. I sometimes refer to θ as the borrower's credit quality and to $1 - \theta$ as the borrower's credit risk. If financed, the project generates a private benefit $b \in \mathbb{R}_+$ to the borrower, regardless of whether the project succeeds. One can think of b as being cash flows that cannot be pledged to the lender, for example, because of agency frictions, and are thus non-contractible (see Holmström and Tirole (1997a)).

Assumption 2.1 (Positive NPV)**.** $\mathbb{E}(\theta)X - 1 > 0$.

Assumption 1 posits that the pool of borrowers is, on average, creditworthy (even excluding their private benefit, b). This implies that an uninformed lender finds it optimal to provide credit. However, since the return in the case of success, X, is finite, the lender aims to

[12]See Cowgill and Tucker (2020) for an economic perspective on algorithmic fairness.

exclude from credit a subset of borrowers with high credit risk. This, in turn, implies that information is valuable for the lender, as it allows for more effective borrower screening.

Data The borrower's quality θ is unknown to both the lender and the borrower at the time of contracting and can be predicted using data $z \in [0, 1]$. The data-generating process (DGP) is an extension of the truth-or-noise information structure in Lewis and Sappington (1994), allowing for a negative correlation between the signal z and the state θ. Specifically, the DGP is such that

$$
\begin{aligned}
z &= \begin{cases} \theta & \text{with pr.} \quad \lambda \\ \varepsilon & \text{with pr.} \ 1 - \lambda \end{cases} \qquad \text{if } \lambda \geq 0, \\
z &= \begin{cases} 1 - \theta & \text{with pr.} \quad |\lambda| \\ \varepsilon & \text{with pr.} \ 1 - |\lambda| \end{cases} \qquad \text{if } \lambda < 0,
\end{aligned}
\tag{2.1}
$$

where $\varepsilon \sim U[0, 1]$ is independent of θ and $\lambda \in [-1, 1]$. Data perfectly reveals the borrower's credit quality with probability $|\lambda|$ and is noise, unrelated to quality, with the residual probability. Moreover, when $\lambda > 0$ (resp., $\lambda < 0$), data is positively (resp., negatively) correlated with credit quality. Hence, $|\lambda|$ will be referred to as the predictive power of data, while λ will be referred to as the correlation between data and credit quality. The predicted credit quality conditional on data z is a weighted average of signal and noise with weights proportional to λ (see the proof of Lemma 1 for details):

$$
\mu_\lambda(z) \equiv \lambda \, z + (1 - \lambda) \, \tfrac{1}{2} = \mathbb{E}(\theta | z). \tag{2.2}
$$

Importantly, I assume that $\lambda \sim U[-1, 1]$ and the lender privately knows its realization. In other words, the lender, having access to statistical technology, has superior knowledge of the DGP and thus privately knows how to convert data into credit risk estimates. This, in turn, implies that the lender is privately informed about the map from data to allocations (see below). Note also that knowing a data realization z, without knowing the actual credit quality θ, does not convey any information about λ to the borrower, as the unconditional distribution of z is independent of λ.

Data Sharing To capture current policies mandating consumers and borrowers control over their data,[13] the data realization z is assumed to be private information of the borrower unless disclosed to the lender. Specifically, knowing z, the borrower chooses a message $m \in \{\varnothing, z\}$, where $m = z$ means sharing data with the lender, while $m = \varnothing$ means withholding data (as in He et al. (2023) or Ali et al. (2023) and following the voluntary disclosure literature, e.g. Milgrom (1981), Grossman (1981)). The binary nature of the data-sharing technology captures the fact that data is hard information: it can be transferred or withheld

[13] Such policies include European Union's General Data Protection Regulation (GDPR) and Payment Service Directives (PSD, PSD2), California's Consumer Privacy Act (CCPA), China's Personal Information Protection Law (PIPL), and other globally adopted Open Banking Initiatives.

but not modified or manipulated.[14] Building on Dye (1985) and Jung and Kwon (1988), a fraction $\pi \in (0,1)$ of borrowers is assumed to withhold data for privacy-related reasons,[15,16] and the lender cannot distinguish privacy-concerned from strategic borrowers.[17]

Credit Allocation Given the knowledge of the DGP, i.e. λ, and the borrower's disclosure decision $m \in \{z, \varnothing\}$, the lender estimates the borrower's credit quality $\mu_\lambda(m)$. Given this estimate, the lender decides whether to provide credit or not, and conditional on providing credit, the lender also requires an interest rate from the borrower. Let $\ell \in \{0,1\}$ denote the credit rationing decision, with $\ell = 1$ denoting credit provision, and $x \in \mathbb{R}$ denote the gross interest payment. To capture the automated nature of credit underwriting, the credit allocation process is rule-based and data-driven.

Definition 2.1 (Algorithm)**.** *A credit allocation algorithm is* $a_\lambda(m) \equiv \big(\ell_\lambda(m), x_\lambda(m)\big)_{m\in\{z,\varnothing\}}$, *where*

$$\begin{aligned} x_\lambda(m) &\equiv \underset{x\in\mathbb{R}}{\arg\max}\ \Big[\mu_\lambda(m)\,x - 1\Big]^{\phi}\Big[\mu_\lambda(m)\big(X - x\big)\Big]^{1-\phi}, \\ \ell_\lambda(m) &\equiv \mathbb{1}\Big\{\mu_\lambda(m)x_\lambda(m) - 1 > 0\Big\}, \end{aligned} \tag{2.3}$$

where $\phi \in (0,1)$ *is the lender's bargaining power.*

Akin to Nash bargaining, the interest rate x is set to maximize the Nash product in the first line of Equation (2.3), a weighted geometric average of the lender's and borrower's contractible surplus with weights proportional to their bargaining power, ϕ and $(1-\phi)$, respectively.[18] Moreover, the algorithm provides credit only if the lender's participation

[14] This assumption differentiates this paper from previous studies (see Section 2, Gaming and Opacity), where data can be manipulated at some cost, and is more in line with current Open Banking initiatives where information is usually stored in a relationship bank (Sharpe (1990) and Rajan (1992)), and can be transferred to a competitor at the customer's will, through an API, at no monetary cost. Similarly, the GDPR empowers individuals to choose whether to share their cookies or digital footprints with an online firm (possibly a lender) or to keep them private.

[15] Alternatively, these agents can be thought of withholding information for other non-strategic reasons: lacking a credit history, being unbanked, technology-averse, or not possessing a device. Studies show that agents' unwillingness to share information online depends on cultural factors (attitudes towards privacy, trust in institutions), demographic factors (age, education level, income level), comfort with technology, user experience (ease of sharing data), trust in the lender/firm (reputation, credibility, brand recognition), types of data shared (breadth, depth, sensitivity), etc. See, e.g., Acquisti et al. (2016) for a general discussion and Morey et al. (2015) for a survey.

[16] Lin (2022) empirically separate two components of consumers' privacy preferences: an intrinsic component (an exogenous taste for privacy, i.e., privacy concerns strictly speaking) and an instrumental component (an endogenous economic loss from revealing private information, i.e., strategic concerns). Goldfarb and Tucker (2012) show that intrinsic privacy concerns in digital markets have increased in recent years.

[17] One way of microfounding this is by assuming that borrowers are privately informed about their disclosure cost, $c \in \{0, \infty\}$, while the lender believes that $\Pr(c = \infty) = \pi$. The main insight of the paper extends to the case where c is finite.

[18] This axiomatic (or cooperative) solution has a strategic microfoundation under symmetric information, as it coincides with the sequential equilibrium of a (non-cooperative) bargaining game with alternating offers (see Rubinstein (1985) and Binmore (1987)). I abstract from the specific microfoundations of bargaining to focus on the central research question of the paper: how algorithmic transparency impacts data sharing.

constraint at that interest rate is satisfied.[19] The parameter ϕ can be interpreted as reflecting the lender's market power,[20] arising from a more concentrated or segmented market structure.[21] Note that the borrower's private benefit b does not enter the Nash product, as these cash flows cannot be pledged to the lender and are thus not contractible.

I refer to $a \equiv (\ell, x)$ as a credit allocation algorithm as it maps data (or the lack of it) to allocations and is tailored to the specific DGP (described by λ). Hence, λ can be broadly interpreted as a parameter that governs the lender's algorithm and is privately known to the lender.

Algorithmic Transparency and Opacity In line with the Bayesian persuasion literature (see, e.g., Kamenica and Gentzkow (2011), Bergemann and Morris (2019), or Dworczak and Martini (2019)), the lender commits to a disclosure policy about the parameters that govern the allocation algorithm. Specifically, before learning λ, the lender commits to reveal it or to conceal it.[22]

Definition 2.2 (Transparency and Opacity)**.** *The lender's algorithm $a_\lambda(m)$ is:*

- *transparent, if λ is public information,*
- *opaque, if λ is the lender's private information.*

Payoffs When the borrower's probability of success is θ and the lender employs an algorithm a, the lender's and the borrower's payoffs are, respectively:

$$\begin{aligned} V^B(\theta, a) &= \ell\left[\theta\big(X - x\big) + b\right], \\ V^L(\theta, a) &= \ell\left[\theta\, x - 1\right], \end{aligned} \tag{2.4}$$

[19] I implicitly assume that when the lender makes nil profits, she does not provide credit. This assumption allows me to conveniently write the no disclosure set (analyzed in Section 5) as a closed set, thus improving readability. The assumption is otherwise inconsequential.

[20] More broadly, ϕ can also encompass other factors, independent of credit risk, that influence the lender's ability to charge higher interest rates, such as macroeconomic conditions (e.g., tight monetary policy) and regulations (e.g., looser capital requirements or consumer protection laws).

[21] As $\phi \to 1$ the market structure approaches a monopoly, while as $\phi \to 0$ it approaches perfect competition. This reduced-form approach to competition enables a focus on the effects of algorithmic disclosure on borrowers' data-sharing decisions while abstracting from its potential strategic implications within firms. Moreover, modeling imperfect competition in the credit market with endogenous information acquisition is challenging, and the literature often focuses on simple algorithms with binary signal realizations (see, e.g., Broecker (1990)), with few exceptions (such as Blickle et al. (2024)). The reduced-form approach facilitates the study of more complex algorithms, aligning with advancements in ML technologies.

[22] The Bayesian persuasion literature typically studies optimal disclosure among the set of all the possible disclosure policies (i.e., state-contingent distributions over signal realizations). Here, I focus on a fully revealing and a fully concealing policy, in line with the ongoing debate regarding algorithmic transparency and opacity. It can be shown that among symmetric monotone disclosure policies (i.e., symmetric convex partitions of the state space $[-1, 1]$) it is without loss to restrict attention to policies that fully reveal extreme values $\lambda \in [-1, -\lambda^\circ) \cup (\lambda^\circ, 1]$ and pool intermediate ones $\lambda \in [-\lambda^\circ, \lambda^\circ]$ for some $0 \leq \lambda^\circ \leq 1$. I study the cases $\lambda^\circ \in \{0, 1\}$, deferring a formal treatment of general disclosure policies to future work.

while the social surplus (or the egalitarian social welfare) is:

$$W(\theta, a) = \ell \left[\theta X - 1 + b\right]. \tag{2.5}$$

Timing and Equilibrium Concept The timing of the game is as follows:

$t = 0$ the lender commits to transparency or opacity, i.e., to reveal or conceal λ;

$t = 1$ (λ, θ, z) realizes, λ is observed by the lender (resp., everyone) under opacity (resp. transparency), z is observed by the borrower;

$t = 2$ the borrower shares or withholds data, i.e. he chooses $m(z) \in \{z, \varnothing\}$;

$t = 3$ the lender allocates credit according to $a_\lambda(m) = \big(\ell_\lambda(m), x_\lambda(m)\big)_{m \in \{z, \varnothing\}}$;

$t = 4$ if the project is financed, the project's returns are realized.

The equilibrium concept is Perfect Bayesian Equilibrium.

Discussion of Model Assumptions A few comments about the model are in order.

- **Positive NPV** The lender extends credit based on prior beliefs, reflecting the common practice of incorporating alternative data sources in a secondary round of the screening process (see, e.g., Nam (2024)), a practice referred to as *multistage screening*. Typically, borrowers who initially qualify for credit by submitting compulsory traditional metrics - such as credit scores and financial history - are then given the option to share additional information - such as payment data, social media activity, utility payments, or online behavior. This voluntary sharing can potentially enhance or impair their credit terms.

- **Unknown Credit Risk** Borrowers are often unaware of their credit risk, a common assumption in the literature. In addition, this excludes the lender's ability to extract the borrower's private information through contractual terms, particularly collateral requirements.[23] In fintech credit, alternative data sources and advanced statistical technologies typically substitute traditional collateral.[24] Given that this paper focuses on the latter topic, I abstract from the possibility of designing truth-telling contracts.

[23]An established body of literature has extensively examined the optimal design of screening contracts. See, e.g., Bester (1985), Freixas and Laffont (1990), and Besanko and Thakor (1987).

[24]Recent studies have explored the effectiveness of these alternative practices in addressing agency problems and expanding access to finance. See, e.g., Agarwal et al. (2019) and Gambacorta et al. (2022).

- **Opaque Statistical Technology** The framework links algorithmic opacity to the underlying statistical technology. One can think of the DGP as being a priori unknown to everyone, while the lender, having access to a statistical technology and a training dataset, can obtain an estimate $\widehat{\lambda}$ of the DGP. Based on this estimate, the lender can tailor an algorithm, which can be described by the parameter $\widehat{\lambda}$. Private knowledge of the estimate of the DGP then coincides with private knowledge of the algorithm's parameters. The model assumes that such an estimate is unbiased, i.e., $\widehat{\lambda} = \lambda$. Essential to my model is the borrowers' uncertainty over the estimate $\widehat{\lambda}$, while the underlying λ can be known or unknown to borrowers and may even differ from $\widehat{\lambda}$. Such uncertainty reflects public mistrust in opaque ML technologies, originating, for example, from concerns about algorithmic fairness and bias.

- **Uniform Beliefs** Borrowers have uniform and symmetric beliefs over the employed statistical technology. This assumption captures the black-box nature of ML algorithms that utilize alternative data sources, as borrowers typically do not understand how these data points influence the lender's assessment of their credit quality. However, the assumption is mainly made to streamline results and can be relaxed, for instance, assuming $\lambda \sim [\underline{\kappa}, \overline{\kappa}]$ where $-1 \leq \underline{\kappa} < 0 < \overline{\kappa} \leq 1$. Essential to borrowers' equilibrium hedging behavior against opacity is that they face some uncertainty about whether the algorithm treats data as good or bad news.

- **Commitment to Disclosure Policy** Committing to opacity is straightforward: the lender keeps the algorithm secret, regardless of its complexity. Commitment to transparency is more nuanced and may depend on the algorithm's *interpretability*. When using interpretable algorithms, such as logistic regression, the lender can directly disclose model parameters — akin to the λ in my model. However, when employing more complex ML techniques — such as support vector machines, tree-based models, gradient boosting machines, or deep learning methods — the lender can be, at least, *explainable*. In practice, this might include disclosing feature importance metrics, SHAP values, LIME explanations, partial dependence plots, and permutation importance to provide borrowers insights into how different features influence predictions (see, e.g., Molnar (2023)).

- **Single-Dimensional Data** The model considers a single explanatory variable, simplifying interpretation and algebra, and facilitating welfare analysis. This framework can be extended to include multidimensional characteristics with an all-or-nothing disclosure technology. In this setup, the probability of success θ is a weighted average of n latent characteristics $t = (t_1, \ldots, t_n)$, i.e. $\theta = \sum_{i=1}^{n} \alpha_i t_i$ where $t_i \sim U[0,1]$ and $\alpha_i \in [0,1]$ with $\sum_{i=1}^{n} \alpha_i = 1$. Each latent characteristic t_i can be estimated from an observable z_i generated by a DGP analogous to Equation (2.1), resulting in predicted credit quality being equal to the hyperplane $\mu_\lambda(z) = \sum_{i=1}^{n} \alpha_i \mu_{\lambda_i}(z_i)$, where $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $z = (z_1, \ldots, z_n)$. The disclosure technology allows borrowers to share all observables or none, i.e., $m(z) \in \{z, \varnothing\}$. Intuitively, under transparency, borrowers can identify which direction of the hyperplane constitutes bad news, allowing them to selectively withhold such information (gaming). Conversely, under opacity, borrowers know that evidence near the boundaries of the z space constitutes an extreme outcome, enabling them to withhold this evidence (hedging).

2.4 Data-Driven Credit Underwriting

This section describes the credit allocation algorithm when the lender has access to the borrower's data z. This is also the allocation the borrower receives when choosing to share data voluntarily.

Lemma 2.1 (Data-Driven Credit Underwriting)**.** *If the lender observes the borrower's data z, the credit allocation algorithm $a_\lambda(z)$ is:*

$$\begin{aligned} x_\lambda(z) &= \frac{1}{\mu_\lambda(z)} + \phi\left(X - \frac{1}{\mu_\lambda(z)}\right), \\ \ell_\lambda(z) &= \mathbb{1}\left\{\mu_\lambda(z) > \frac{1}{X}\right\}. \end{aligned} \tag{2.6}$$

Proof. See Appendix A. ■

The allocation rule is rather intuitive and depends on predicted credit quality as follows. First, the interest rate is decreasing in expected quality $\mu_\lambda(z)$: less risky borrowers get lower interest rates. Moreover, the interest rate equals the competitive rate $1/\mu_\lambda(z)$ plus a mark-up that is increasing in the lender's bargaining power ϕ, the return of the project X and expected quality $\mu_\lambda(z)$. Second, the credit rationing decision is a cutoff rule: borrowers with credit quality below the threshold $1/X$ are credit rationed.

Importantly, the algorithm depends on the correlation λ between data and credit quality through the lender's forecast of quality $\mu_\lambda(z)$. When λ is positive (negative), higher data realizations are good news (bad news), and the interest rate decreases (increases) in z, while it is flat and independent of data when $\lambda = 0$. Moreover, the interest rate is steeper the higher the predictive power of data $|\lambda|$. Similarly, the allocation rule rations credit to different sets of data points depending on the value of λ. One can rewrite the credit rationing rule in Equation (2.6) as

$$\ell_\lambda(z) = \begin{cases} \mathbb{1}\{z > r(\lambda)\} & \text{if} \quad \lambda \in [\overline{\lambda}, 1] \\ 1 & \text{if} \quad \lambda \in (\underline{\lambda}, \overline{\lambda}) \\ \mathbb{1}\{z < r(\lambda)\} & \text{if} \quad \lambda \in [-1, \underline{\lambda}], \end{cases} \tag{2.7}$$

where

$$r(\lambda) = \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right), \tag{2.8}$$

will be refereed to as the rationing threshold, while $\overline{\lambda} \equiv 1 - 2/X \in (0,1)$ and $\underline{\lambda} = -\overline{\lambda}$. Note that when the correlation is sufficiently high, $\lambda \in [\overline{\lambda}, 1]$, the rationing threshold is positive, $r(\lambda) \in [0, 1/X]$, and some borrowers with low data points are credit rationed.

Conversely, borrowers with high data points do not get credit when the correlation is sufficiently negative, $\lambda \in \big[-1, \underline{\lambda}\big]$ so that $r(\lambda) \in [1 - 1/X, 1]$. Finally, when the predictive power of data is low $|\lambda| < \overline{\lambda}$, there is no data-driven credit ratio; still, data will be optimally used to price credit (as long as $|\lambda| > 0$).

Given the allocation rule, the lender's and the borrowers' expected profits given a data realization z are, respectively:

$$\begin{aligned} V_\lambda^L(z) &= \ell_\lambda(z)\Big[\phi\Big(\mu_\lambda(z)X - 1\Big)\Big], \\ V_\lambda^B(z) &= \ell_\lambda(z)\Big[(1-\phi)\Big(\mu_\lambda(z)X - 1\Big) + b\Big]. \end{aligned} \tag{2.9}$$

Conditional on credit provision, the lender and the borrower obtain a fraction of the (contractible) social surplus equal to their bargaining power, while the borrower also enjoys the private benefit. Moreover, greater statistical accuracy (higher $|\lambda|$) has a redistributive impact, creating winners and losers (as in Fuster et al. (2022)). Better-than-average borrowers enjoy lower interest rates and are thus better off due to greater predictive power. In contrast, riskier borrowers are worse off, facing higher interest rates and potential credit rationing due to increased precision.

2.5 Voluntary Data Sharing

This section studies borrowers' voluntary data-sharing behavior. In this framework, the lender draws inferences from the data provided and the borrower's disclosure decisions. The analysis considers, in turn, transparent and opaque allocation algorithms, as defined in Definition 2.

Before studying the borrowers' data-sharing decision, consider the lender's equilibrium inference following data withholding. Let $Q \equiv \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that strategically withhold data, the lender's inference of the borrower's credit quality is $\mu_\lambda(\varnothing) = \mu_\lambda(z(\pi, Q))$ where $\mu_\lambda(z)$ is defined in Equation (2.2), while

$$\begin{aligned} z(\pi, Q) &\equiv \omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q), \\ \omega(\pi, Q) &\equiv \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)}. \end{aligned} \tag{2.10}$$

Data withholding can either be strategic, stemming from a borrower with data realization $z \in Q$, or be non-strategic. Thus, following withholding, the lender assigns an inferred data realization $z(\pi, Q)$, a weighted average of the expected data realization withheld by strategic and non-strategic borrowers. The weight $\omega(\pi, Q)$ is the posterior probability of non-strategic withholding. It increases with π, the fraction of non-strategic users, and decreases with the mass of strategically withheld data realizations. Note that the inferred data realization $z(\pi, Q)$ is also weighted by λ, reflecting the likelihood of data being informative rather than noise. Following withholding, the algorithm maps the inferred data quality to allocations as in Lemma 2.1 replacing $z(\pi, Q)$ to z (see Lemma A.1 in Appendix A for a formal treatment).

2.5.1 Gaming under Transparency

This section shows that transparency makes the algorithm vulnerable to gaming in the form of strategic withholding of unfavorable information. When the algorithm's parameters are publicly known, the borrower's data-sharing strategy depends on this knowledge and is formalized in the following proposition.

Proposition 2.1 (Data-Sharing to a Transparent Algorithm)**.** *When the lender uses a transparent algorithm, the set of strategic borrowers that optimally withhold data is:*

$$\mathcal{G}(\lambda,\pi)=\begin{cases}\Big[0,\max\big\{r(\lambda);\gamma(\pi)\big\}\Big] & \textit{if } \lambda\in(0,1],\\ \Big[0,1\Big] & \textit{if } \lambda=0,\\ \Big[\min\big\{r(\lambda);1-\gamma(\pi)\big\},1\Big] & \textit{if } \lambda\in[-1,0),\end{cases} \tag{2.11}$$

where

$$\gamma(\pi)\equiv\frac{\sqrt{\pi}(1-\sqrt{\pi})}{1-\pi}\in\left(0,\frac{1}{2}\right) \tag{2.12}$$

is increasing.

Proof. See Appendix A. ■

Consider the case where the correlation is positive and low, such that the algorithm uses data for pricing but not to ration credit, i.e. $\lambda\in(0,\overline{\lambda})$ (see Equation (2.7) and the discussion that follows). The algorithm is known to assign lower interest rates to borrowers with higher z. Risky borrowers — with low z — would get pretty high interest rates by sharing information and thus opt to withhold data as this grants them better contractual terms. Suppose the set of borrowers withholding data is $Q=[0,q]$ for some $q\in(0,1)$. By pooling with non-strategic borrowers with an average credit quality of $1/2$, borrowers with $z\in Q$ are perceived as having lower credit risks than they actually are and thus receive lower interest rates. However, data withholding is perceived by the lender as a signal of credit risk, and better borrowers prefer to separate from the risky no-disclosure pool by sharing information.

The threshold q can be identified as the borrower that receives the same interest rates by sharing and withholding data. For this to occur, the inferred data realization following data withholding must match the borrower's actual data realization. The indifference condition can be expressed as

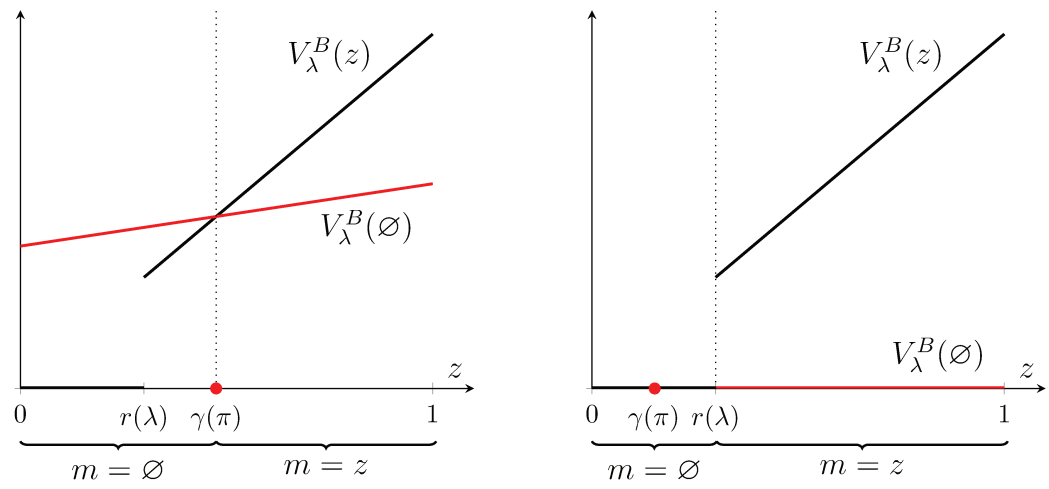
$$q=z(\pi,q),$$

where $z(\pi,q)$ is determined by substituting $Q=[0,q]$ into both lines of Equation (2.10). Its unique (positive) solution is $q=\gamma(\pi)$ given by Equation (2.12). Note that this threshold is independent of the predictive power of data λ. This is because the lender's equilibrium inference, $\mu_\lambda(\cdot)$, scales by the same weight λ both the disclosed data realization z and the data realization $z(\pi,q)$ inferred from data withholding. Hence, λ is irrelevant to the borrower's indifference condition; all that matters is the comparison between the borrower's actual data realization and the inferred data realization when data is withheld. On the other hand, the indifference condition depends on the fraction of non-strategic borrowers π. The indifference threshold $\gamma(\pi)$ is increasing and strictly between 0 and 1/2. When borrowers are primarily strategic ($\pi \to 0$), data withholding is a strong indicator of low credit quality. Consequently, the algorithm significantly penalizes withholding by assigning very high interest rates. As a result, even the riskiest borrowers choose to separate, causing the no-disclosure set to shrink, and $\gamma(\pi)$ approaches zero. This essentially aligns with the unraveling result of by Grossman (1981) and Milgrom (1981), which states that all available information is disclosed to prevent an adverse equilibrium inference by the receiver. When the fraction of privacy-concerned borrowers increases, unraveling forces are milder because withholding data becomes a less reliable indicator of low credit quality. Consequently, a growing number of risky borrowers prefer to withhold information, as this strategy yields increasingly favorable interest rates as π rises. In the opposite limit, where borrowers are primarily non-strategic ($\pi \to 1$), withholding data provides no information, and the algorithm's equilibrium inference aligns with the prior. As a result, only above-average borrowers choose to separate, and $\gamma(\pi)$ approaches 1/2.

Consider now more data-intensive allocation rules such that $\lambda \geq \overline{\lambda}$. Disclosed data is used not only for pricing but also for rationing credit. In particular, borrowers with $z \in [0, r(\lambda)]$, with $r(\lambda) > 0$ given by Equation (2.8), do not receive credit if they share their data. However, withholding information may allow them to obtain credit if the inferred data realization $z(\pi, \gamma(\pi)) = \gamma(\pi)$ exceeds the rationing threshold $r(\lambda)$. This happens when unraveling forces are sufficiently mild that the lender's equilibrium inference is not too averse. Specifically, this occurs when non-strategic borrowers are relatively numerous, i.e., $\pi > \gamma^{-1}(r(\lambda)) = \big(r(\lambda)/(1 - r(\lambda))\big)^2 \in (0, 1)$ or, equivalently, when the predictive power of data is sufficiently low, i.e. $\lambda < r^{-1}(\gamma(\pi)) = \big(\frac{1}{2} - \frac{1}{X}\big) / \big(\frac{1}{2} - \gamma(\pi)\big) \in (\overline{\lambda}, 1)$. The equilibrium is depicted in Figure 2.1, Panel (a). When these conditions are not met, unraveling forces are stronger, and the lender's equilibrium inference is so averse that withholding data induces credit rationing. Consequently, borrowers with data below the rationing threshold $r(\lambda)$ are indifferent between sharing and withholding as they do not receive credit anyway. Only observationally credit-worthy borrowers share information and get funded at an interest rate reflecting their expected credit risk. The equilibrium is depicted in Figure 2.1, Panel (b).

When $\lambda < 0$, symmetric reasoning applies, and the right tail of the data distribution is withheld in equilibrium. When $\lambda = 0$, data is uncorrelated with quality, and the algorithm extends credit, ignoring any disclosed information. As a consequence, borrowers are indifferent between sharing and withholding data.

The equilibrium withholding of information can be considered a form of gaming — enabled by transparent allocation rules — by which strategic borrowers secure better contractual terms by selectively concealing bad news while sharing good news. The misallocation of credit resulting from gaming has welfare consequences for both the lender and the borrowers and will be analyzed in Section 6. Gaming can lower interest rates for risky but



(a) $r(\lambda) < \gamma(\pi)$.

(b) $\gamma(\pi) \leq r(\lambda)$

Figure 2.1: **Gaming under Transparency.** The figure depicts the borrower's surplus from data sharing, $V_\lambda^B(z)$ (in black), and from data withholding, $V_\lambda^B(\varnothing)$ (in red), when $\lambda > \overline{\lambda} > 0$. The red dot corresponds to the borrower's data realization inferred from data withholding and equals $\gamma(\pi)$ defined in Equation (2.12). $r(\lambda)$ denotes the rationing threshold and is defined in Equation (2.8). Panel (a) shows the case where privacy-concerned borrowers are sufficiently numerous, i.e., $\pi > \gamma^{-1}(r(\lambda))$, and gaming occurs in equilibrium. Panel (b) shows the case where privacy-concerned borrowers are few, i.e., $\pi \leq \gamma^{-1}(r(\lambda))$, and gaming does not occur in equilibrium.

credit-worthy borrowers and even induce credit provision to credit-unworthy borrowers. In both cases, the extent of gaming increases with the fraction of non-strategic users, which thus can be seen as a measure of gaming ability. Moreover, gaming is more pronounced when data is more informative, and the allocation rule relies more strongly on it.[25] This result is related to several papers examining the allocation rules manipulation by strategic users, finding that gaming incentives increase when the rule heavily relies on data (see the paragraph on Gaming and Opacity in Section 2). These papers suggest that an optimal scoring rule should underutilize data to mitigate gaming. Section 5.2 explores an alternative strategy lenders might employ to deter gaming: making the rule opaque without reducing its data sensitivity.

2.5.2 Hedging under Opacity

This section examines borrowers' data-sharing behavior when the algorithm is opaque, meaning the parameters governing it are kept secret from borrowers. While opacity can sometimes mitigate gaming, it also induces borrowers to hedge against the unpredictability of the algorithm by withholding information. The equilibrium data-sharing strategy is described in the following proposition.

[25] When $\lambda \in [\overline{\lambda}, \widehat{\lambda}(\gamma(\pi)))$, the mass of credit-unworthy borrowers that get credit by withholding information increases with the predictive power of data.

Proposition 2.2 (Data-Sharing to an Opaque Algorithm)**.** *When the lender uses an opaque algorithm, the set of strategic borrowers that optimally withhold data is:*

$$\mathcal{H}(b, \phi) \equiv [\eta(b, \phi), 1 - \eta(b, \phi)] \tag{2.13}$$

where

$$\eta(b, \phi) \equiv \max\left\{0, \frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right)\right\} \in \left[0, \frac{1}{X}\right) \tag{2.14}$$

is weakly decreasing in ϕ and b.

Proof. See Appendix A. ■

Before providing the intuition behind this result, it is helpful to understand why data-withholding constitutes a safe hedging strategy against algorithmic opacity. Consider *any* symmetric data-withholding set Q.[26] For any of these sets, the lender's inference of credit quality following data-withholding is

$$\mu_\lambda(\varnothing) = \mu_\lambda\left(\tfrac{1}{2}\right) = \tfrac{1}{2},$$

and is thus independent of i) the lender's statistical technology, and ii) the actual data-withholding set (see Equation (2.10)). No matter the true link between data and credit quality (i.e., λ) and the size of the no-disclosure pool, the data-withholding set contains borrowers with mixed positive and negative evidence, so the lender does not infer much from data withholding. Consequently, as long as the no-disclosure pool is symmetric, there is no stigma attached to data withholding, and disclosing borrowers impose no externalities on withholding borrowers. Moreover, since the overall pool of borrowers is credit-worthy, the lender provides credit following no-disclosure. Similarly, the ex-post interest rate a borrower gets from withholding is independent of λ. It follows that if the data-withholding set is symmetric, data-withholding is a safe strategy against algorithmic opacity: it insures the borrower against the risk of credit rationing and the variability in interest rates.

Consider now what the borrowers expect from data sharing. We can distinguish two types of borrowers depending on their data realizations:

- Borrowers with extreme data realizations $z \in [0, 1/X] \cup [1 - 1/X, 1]$ face a positive probability of being denied credit when they share information, and this probability increases as their data realization becomes more extreme — i.e., when z moves further from the center of the distribution. This is intuitive, as increasingly extreme data realizations are deemed credit-unworthy for progressively weaker correlation values λ, resulting in a denial of credit by a wider range of allocation rules that the lender

[26] A set $Q \subset [0, 1]$ is symmetric about $\frac{1}{2}$ if $\forall z \in Q$, $1 - z \in Q$. Some examples are: the entire interval $Q = [0, 1]$, a subinterval of the form $Q = [q, 1 - q]$ with $q \in [0, 1/2)$, disjoint subintervals of the form $Q = [0, q] \cup [1 - q, 1]$ with $q \in (0, 1/2]$, etc.

may employ.[27]
Conditional on getting credit, the interest rates are also uncertain — as these depend on λ through the lender's opaque statistical inference from data — and the expected interest rate decreases as data becomes more extreme. This is also intuitive, as, conditional on receiving credit, increasingly extreme data realizations are seen as progressively stronger signals of credit quality.[28]

- Borrowers with central data realizations $z \in (1/X, 1-1/X)$ face no risk of credit rationing, and they always obtain credit by sharing information. Their data realizations are close to average and represent a weak indicator of credit risk regardless of whether the data is positively or negatively correlated with credit quality. Since the average borrower is creditworthy, they are granted credit by every allocation rule the lender might employ.
However, these borrowers still face algorithmic opacity because the interest rate they receive is uncertain. Given the borrowers' risk neutrality and the symmetric treatment of data across rules, this uncertainty perfectly balances out, and the expected interest rate is independent of z.

The equilibrium data-sharing behavior can now be better understood. When the lender's bargaining power is relatively high (i.e., $\phi \geq 1-2b$) or, equivalently, when the borrower's private benefit is substantial (i.e., $b \geq (1-\phi)/2$), the equilibrium involves no data-sharing, i.e., $Q=[0,1]$. In this scenario, borrowers prioritize securing credit and enjoying their private benefits b over obtaining marginally lower interest rates, as interest rates extract most of their surplus once the credit is extended. Consequently, the utility of data sharing exhibits a symmetric inverted U-shape: borrowers with more extreme data realizations face an increasing risk of credit rationing and would like to appear closer to the average than they truly are to mitigate this risk. This can be achieved by withholding information. The strategy labels them as average borrowers, fully insulates them against the risk of credit rationing, and is, therefore, strictly optimal. In contrast, borrowers with central data realizations are indifferent between the two strategies, as they obtain credit anyway and are risk-neutral conditional on receiving credit. No borrower has an incentive to separate from the no-disclosure pool. Hence, data-withholding emerges as a safe hedging strategy against the unpredictability of the algorithm: it insures the borrower against the

[27] Consider $z \in [0, 1/X]$. From Equation (2.7) and Equation (2.8), a data realization in this set is denied credit when $z \leq r(\lambda)$, or equivalently when the correlation between data and credit quality is sufficiently high, i.e., when

$$\lambda \geq r^{-1}(z) \equiv \widehat{\lambda}(z) = \frac{\frac{1}{2}-\frac{1}{X}}{\frac{1}{2}-z} \in (\overline{\lambda}, 1],$$

where $r^{-1}(\cdot)$ denotes the inverse function of $r(\cdot)$. The probability of getting credit is thus $p(z) \equiv \Pr(\lambda < \widehat{\lambda}(z)) = (\widehat{\lambda}(z)+1)/2$, so that the probability of credit rationing, $1-p(z)$, is decreasing in z for $z \in [0, 1/X]$. Symmetric reasoning applies for $z \in [1-1/X, 1]$.

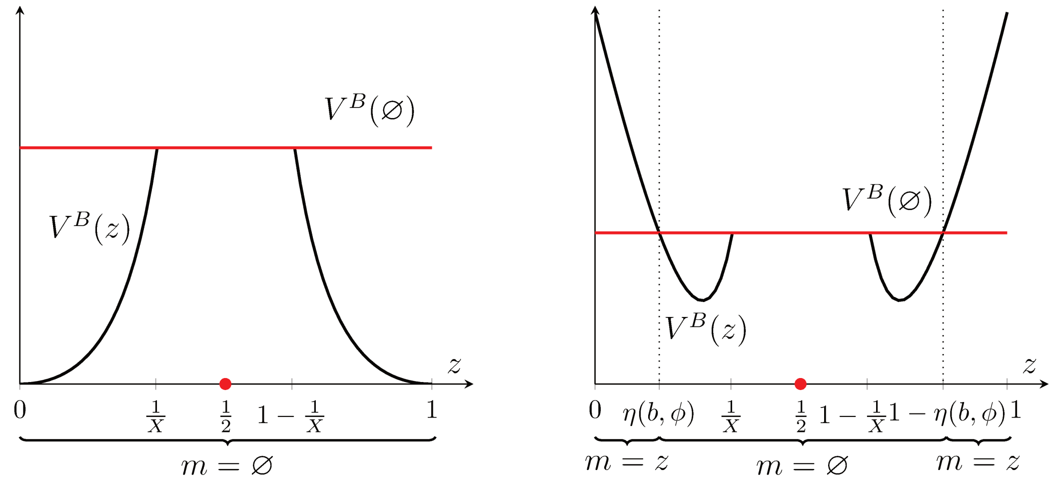
[28] Consider $z \in [0, 1/X]$. The inferred credit quality across all the allocation rules that provide the borrower credit is $\mu_-(z) = \mathbb{E}(\mu_\lambda(z) | \lambda < \widehat{\lambda}(z))$ and is decreasing in z for $z \in [0, 1/X]$. This is because the borrow receives credit when $\lambda < \widehat{\lambda}(z)$, i.e. when data is predominately negatively correlated with credit quality, so when lower data realizations are viewed as more favorable. Therefore, conditional on receiving credit, the expected interest rate increases with z. Symmetric reasoning applies for $z \in [1-1/X, 1]$.

risk of credit rationing and the variability in interest rates.[29] The equilibrium is depicted in Figure 2.2, Panel (a).

Conversely, when the borrower has relatively more bargaining power (i.e. $\phi < 1 - 2b$) or a lower private benefit ($b < (1-\phi)/2$), the most extreme evidence is disclosed in equilibrium, i.e. $Q = [\eta(b,\phi), 1-\eta(b,\phi)]$ with $\eta(b,\phi) \in (0, 1/X)$. In this case, the potential for better interest rates drives extreme borrowers' decision to share information. These are still the most likely to face credit rationing if they disclose, but they also stand to gain the best rates if they obtain credit. Consequently, when the potential upside reward outweighs the guaranteed benefit of obtaining credit, these borrowers forgo the safety of non-disclosure and take the risk of sharing information. However, less extreme borrowers have less exposure to the potential upside and stick to the safe hedge. We can thus write the no-disclosure set as $Q = [q, 1-q]$ and identify the threshold q as the borrower for whom these forces perfectly balance out, being indifferent between sharing and withholding data. The indifferent borrower is given by $q = \frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right) \in (0, 1/X)$. This threshold increases when ϕ and b decrease: as borrowers gain bargaining power or derive lower benefits from securing credit, hedging motives decrease, and they share more information in equilibrium. In the limit when b approaches 0, hedging motives are absent: no matter the lender's bargaining power, every borrower facing the threat of credit rationing strictly prefers to disclose information, hoping for a better interest rate. The equilibrium is depicted in Figure 2.2, Panel (b).

In summary, algorithmic opacity prompts some borrowers to hedge against its unpredictability by withholding information. Hedging motives are amplified when the lender can extract borrowers' surplus more effectively, potentially preventing data sharing from occurring in equilibrium. However, leveraging borrowers' risk-taking behavior, opacity may also lead to the extraction of extreme and highly conclusive evidence when borrowers possess greater bargaining power. Thus, algorithmic opacity presents both benefits and costs in terms of data collection and the resulting reduction of misallocation of credit. These factors will be studied in Section 2.6.

[29]Since the value from data-sharing is symmetric around $z = 1/2$, the set of borrowers withholding data, Q, is also symmetric in equilibrium. Note that the hedge is perfect because of the symmetry of Q, but is only partial when the conjectured Q is asymmetric. Still, the intuition extends to asymmetric no-disclosure sets, and perfect hedging will be optimal in equilibrium. Similarly, a symmetric equilibrium data-withholding set is due to the symmetry in algorithmic uncertainty, which drives the symmetry of the value from data-sharing around $z = 1/2$. One can show that hedging type of behavior (i.e. pooling at the extremes) emerges in equilibrium when $\lambda \sim [\underline{\kappa}, \overline{\kappa}]$ where $-1 \leq \underline{\kappa} < 0 < \overline{\kappa} \leq 1$ and $\underline{\kappa}$ is sufficiently low.



(a) $\phi \geq 1 - 2b$.

(b) $\phi < 1 - 2b$

Figure 2.2: **Hedging under Opacity.** The figure depicts the borrower's expected surplus from data sharing, $V^B(z)$ (in black), and from data withholding, $V^B(\varnothing)$ (in red). The red dot corresponds to the borrower's data realization inferred from data withholding and equals $\frac{1}{2}$. $\eta(b, \phi)$ denotes the indifferent borrower and is defined in Equation (2.14). Panel (a) shows the case where the lender's bargaining power is sufficiently high, and hedging is complete in equilibrium. Panel (b) shows the case where the lender's bargaining power is sufficiently low, and hedging is only partial in equilibrium.

2.5.3 Externalities and Privacy Concerns

Interestingly, privacy-concerned borrowers are affected differently depending on whether the lender's algorithm is transparent or opaque.

Corollary 2.1 (Externalities)**.** *When the lender uses a transparent algorithm, non-strategic borrowers are denied credit when they are few, i.e., when $\pi \leq \gamma^{-1}(r(\lambda)) \in (0, 1)$. Instead, when the lender uses an opaque algorithm, non-strategic borrowers always obtain credit.*

Under transparency, when privacy-concerned borrowers are few, no disclosure is a strong signal of low credit quality and leads to credit denial. Hence, strategic borrowers impose a negative externality on privacy-concerned borrowers through their data-sharing decisions, ultimately pushing them out of the market despite being creditworthy on average. This operates through the classic unraveling logic and stems from the lender's inability to differentiate between strategic and non-strategic borrowers. On the contrary, unraveling forces are absent under opacity. The no-disclosure pool contains both observationally good and bad borrowers; hence, no stigma is attached to no-disclosure, regardless of the actual statistical link between data and credit quality. As a result, borrowers who disclose information impose no externalities on privacy-concerned borrowers, allowing them to always receive credit.[30]

[30]Note that similar effects would arise in a smoother specification where privacy-concerned borrowers face a finite disclosure cost. Under transparency, when unraveling forces come into play, privacy-concerned

2.6 Welfare Analysis

The previous discussion highlighted that both transparency and opacity lead to information losses through either gaming or hedging behavior of strategic borrowers. This section studies the resulting misallocation of credit and mispricings. It demonstrates that transparency regimes have significant welfare implications, affecting the lender's profits, borrowers' welfare, and overall social surplus.

2.6.1 Lender's Profits

We begin by analyzing the transparency regime that maximizes the lender's ex-ante profits. Since data is valuable for the lender, the optimal regime minimizes the information loss resulting from borrowers' strategic behavior. The lender's transparency choices are described in the following proposition and illustrated in Figure 2.3.

Proposition 2.3 (Lender-Optimal Transparency Regime)**.** *Let*

$$\pi_L(b,\phi) \equiv \min\left\{\widehat{\pi}(b,\phi), \widehat{\pi}(X)\right\} \in (0, \widehat{\pi}(X)]\,, \tag{2.15}$$

where $\widehat{\pi}(X) \equiv \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ *while* $\widehat{\pi}(b,\phi)$ *is defined in Equation (A.27) and is increasing in* b *and* ϕ*. If the fraction of non-strategic borrowers is sufficiently high, i.e.,* $\pi \geq \pi_L(b,\phi)$ *(resp. sufficiently low, i.e.,* $\pi < \pi_L(b,\phi)$*), an opaque (resp. transparent) algorithm maximizes the lender's profits.*

Proof. See Appendix A. ■

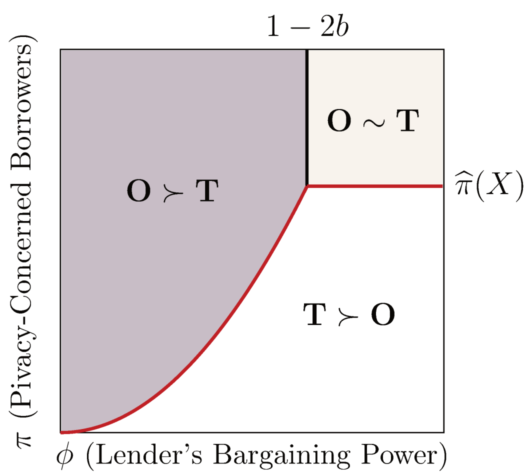


Figure 2.3: **Lender-Optimal Transparency Regime.** The figure depicts the lender's preferences between a transparent (**T**) and an opaque (**O**) algorithm as defined in Definition 2. The symbol ≻ denotes a strict preference, while ∼ denotes indifference. The red line represents $\pi_L(b,\phi)$ as defined in Equation (2.15).

borrowers are compelled to incur the disclosure cost to distinguish themselves from the no-disclosure pool and secure credit. In contrast, under opacity, unraveling forces are absent, allowing privacy-concerned borrowers to avoid the disclosure cost while still obtaining credit.

The lender's optimal transparency regime maximizes the algorithm's efficacy in rationing credit, foregoing the potential benefits of price discrimination. This happens because the Blackwell (1951) value of information is nil above the rationing threshold, as the lender is risk-neutral in this region. While detailed credit quality information can reduce the variance of the lender's returns from credit-worthy borrowers, it does not affect their expected value since potential mispricings offset each other. Therefore, from an ex-ante perspective, the lender gains no additional profits from acquiring more granular information than necessary for optimal credit rationing. It follows that the algorithm's transparency regime is irrelevant in the region where the algorithm prescribes no credit rationing, i.e., when $\lambda \in (\underline{\lambda}, \overline{\lambda})$.

Building on this result, we can examine credit misallocations resulting from gaming when $\lambda \geq \overline{\lambda}$ (the analysis is symmetric when $\lambda \leq \underline{\lambda}$). Under transparency, when $\lambda \in \left[\overline{\lambda}, \widehat{\lambda}(\gamma(\pi))\right)$, strategic high-risk borrowers (with $z \in [0, r(\lambda)]$) evade credit rationing by withholding bad news. These borrowers represent a negative NPV investment for the lender and thus induce a loss in expectation. The magnitude of this loss increases with λ as a greater number of high-risk borrowers slip through the lender's screening process and obtain credit. In addition, some marginally credit-worthy borrowers (with $z \in (r(\lambda), \gamma(\pi))$) receive more favorable interest rates than what their actual credit risk would suggest. Because of risk neutrality, this second group of borrowers does not impact the lender's expected profits. For higher correlation levels, $\lambda \in \left[\widehat{\lambda}(\gamma(\pi)), 1\right]$, gaming does not occur. Only observationally credit-worthy borrowers receive credit, while non-disclosing borrowers are credit rationed. Within the group of strategic borrowers, the lender achieves profits comparable to those with full data observability. However, the lender suboptimally refuses credit to non-strategic borrowers who are, on average, credit-worthy.

Overall, the profits the lender yields employing a transparent algorithm weakly decrease in the number of behavioral borrowers π. This happens for two reasons. First, as π increases, gaming behavior spreads across a broader range of allocation rules, exacerbating misallocations and resultant losses. Second, in the region where gaming is absent due to effective unraveling forces, an increasing number of non-strategic credit-worthy borrowers are erroneously denied credit. At a certain point, when $\pi \geq \widehat{\pi}(X) \in (0,1)$ i.e. $\widehat{\lambda}(\gamma(\pi)) \geq 1$, gaming extends throughout the entire range of λ and the lender's profits equal the no-information level.

Consider now credit misallocations resulting from an opaque algorithm. When the lender's bargaining power is sufficiently high (i.e., $\phi \geq 1 - 2b$), causing borrowers' hedging motives to be strong enough to prevent any data sharing in equilibrium, the lender is unable to perform either credit rationing or price discrimination, regardless of the true predictive power of the data. Consequently, the lender earns the lowest possible profits, equivalent to the no-information profits. Instead, when borrowers have relatively more bargaining power (i.e., $\phi < 1 - 2b$), they share extreme evidence hoping for favorable interest rates but may be excluded from credit ex-post. When correlation is mild, i.e. when $\lambda \in \left[\overline{\lambda}, \widehat{\lambda}(\eta(b, \phi))\right]$ or equivalently $r(\lambda) \leq \eta(b, \phi)$, high-risk borrowers with $z \leq r(\lambda) \leq \eta(b, \phi)$ share data, hoping to be classified as low risks, instead the algorithm uses the disclosed information to ration credit. Hence, only observationally credit-worthy borrowers obtain funding, and the lender achieves profits comparable to those with full data observability. When correlation is higher, $\lambda \in \left(\widehat{\lambda}(\eta(b,\phi)), 1\right]$ or equivalently $\eta(b,\phi) < r(\lambda)$, rationing occurs ex-post but is only partially optimal. In fact, hedging behavior precludes the lender from rationing credit to those high risks that withhold information, i.e., those with $z \in [\eta(b,\phi), r(\lambda)]$. Finally, non-strategic borrowers always obtain credit, and this is optimal as they are, on average, creditworthy.

In summary, the lender's bargaining power ϕ and the borrower's private benefit, b, strengthen the borrower's hedging behavior, limiting the data available to the lender. As a result, credit rationing occurs for a narrower range of allocation rules, thereby reducing the lender's profits.

Transparency regimes can now be easily compared. When the borrower's hedging motives are strong, and gaming ability is high (i.e., $\phi \geq 1 - 2b$ and $\pi \geq \widehat{\pi}(X)$), the lender is unable to implement credit rationing, regardless of the chosen transparency regime. As a result, the lender is indifferent between the two regimes. For the remaining parameter values, the lender faces a trade-off. When gaming ability outweighs hedging motives, an opaque algorithm maximizes the lender's profits by extracting more information, thus enabling credit rationing for those high-risk borrowers who would otherwise game a transparent system. Conversely, when gaming ability is weaker than hedging motives, transparency becomes optimal by leveraging unraveling forces while eliminating hedging behavior. The threshold $\widehat{\pi}(b,\phi) \in \widehat{\pi}(X)\big)$ at which the lender switches from transparency to opacity is increasing in the lender's bargaining power ϕ and the borrower's private benefit b: as hedging motives strengthen, transparency becomes optimal for a broader range of parameters. When the borrower's private benefit exceeds $\frac{1}{2}$, hedging motives are so strong that an opaque algorithm results in no data-sharing irrespectively of the lender's bargaining power. Consequently, transparency benefits the lender by enhancing data extraction and the algorithm's credit rationing capabilities.

2.6.2 Social Welfare

We now examine the socially optimal transparency regime — the one that maximizes egalitarian social welfare (or total surplus) — and show that the lender's transparency choices are often socially inefficient.

It is useful to first consider a benchmark in which the social planner can segment the data available to the lender, who then optimally allocates credit based on this segmentation (the approach is similar to Bergemann et al. (2015)). In this benchmark, the planner observes the lender's statistical technology, λ, and commits to a segmentation of observable credit risk, i.e., a monotone partition of the borrower's characteristics. Specifically, a credit risk segmentation is a partition $Z = \{z_0, z_1, \ldots, z_{n-1}, z_n\}$ of the borrower's data space $[0,1]$ with $0 = z_0 < z_1 < \cdots < z_{n-1} < z_n = 1$, where the lender only observes in which risk bucket r_i a borrower belongs, where $r_i \equiv \{z \in [z_{i-1}, z_i)\}$ for $i \in \{1, \ldots, n\}$.

Every risk bucket r_i induces a lender-optimal allocation, that is, a credit provision decision $a_\lambda(r_i)$ and an interest rate $x_\lambda(r_i)$, and thus can be seen as a recommendation about credit allocations from the planner to the lender. Hence, the welfare-optimal credit risk segmentation — the set of all these recommendations — has to i) maximize social welfare and ii) be incentive-compatible (or obedient) for the lender.

Lemma 2.2 (Welfare-Optimal Credit Risk Segmentation)**.** *The welfare-optimal credit risk segmentation contains at most two risk buckets ($n \leq 1$) with cutoff*

$$z_1 = r^\star(\lambda, b) \equiv \begin{cases} \max\left\{r(\lambda) - \frac{b}{\lambda X}, 0\right\} & \text{if } \lambda \in [0,1], \\ \min\left\{r(\lambda) - \frac{b}{\lambda X}, 1\right\} & \text{if } \lambda \in [-1, 0), \end{cases} \tag{2.16}$$

where $r(\lambda)$ is defined in Equation (2.8).

Proof. See Appendix A. ■

The lemma shows that the market solution entails excessive rationing from a social perspective, and the misalignment of incentives between the lender and the planner increases with the borrower's private benefit, b. To see this, note that the credit provision induced by the welfare-optimal segmentation (as a function of the underlying characteristic z) is

$$\ell^\star_\lambda(z) = \begin{cases} \mathbb{1}\left\{z > r(\lambda) - \frac{b}{\lambda X}\right\} & \text{if} \quad \lambda \in \left[\overline{\lambda}(b), 1\right] \\ 1 & \text{if} \quad \lambda \in \left(\underline{\lambda}(b), \overline{\lambda}(b)\right) \\ \mathbb{1}\left\{z < r(\lambda) - \frac{b}{\lambda X}\right\} & \text{if} \quad \lambda \in \left[-1, \underline{\lambda}(b)\right], \end{cases} \tag{2.17}$$

where $\overline{\lambda}(b) = \overline{\lambda} + \frac{2b}{X}$ and $\underline{\lambda}(b) = -\overline{\lambda}(b)$. Comparing this with the lender-optimal credit allocation in Equation (2.7), we find that the market solution results in insufficient provision of credit, as $\ell_\lambda(z) \leq \ell^\star_\lambda(z)$. Specifically, when $\lambda \geq \overline{\lambda}(b)$, there exists a subset of borrowers with $z \in (r(\lambda) - \frac{b}{\lambda X}, r(\lambda)]$ that the lender would ration despite having a positive surplus project. The size of this set increases with b, as a higher private benefit for the borrower increases the overall surplus without changing the lender's profits. Nevertheless, in this region, the planner would deny credit to certain high-risk borrowers seeking to fund negative surplus projects. When $\lambda \in [\overline{\lambda}, \overline{\lambda}(b))$, the planner would extend credit to all borrowers, while the lender would still ration some with sufficiently low realizations $z \leq r(\lambda)$. The range of allocation rules for which this happens also increases with b. Only when b tends to 0 does the lender implement the socially efficient level of credit provision.

Note also that the welfare-optimal segmentation entails, at most, two coarse risk buckets. This happens for two reasons. First, interest rates determine how the surplus is divided between the lender and the borrower without altering the total surplus created. Therefore, the planner's only concern is to ensure welfare-optimal credit provision. Since credit provision is a binary decision, the segmentation contains at most two buckets, as all the potential sub-buckets inducing the same decision can be pooled in a unique category. This is due to the revelation principle, for which messages (risk buckets) can be thought of as incentive-compatible action recommendations. Second, the risk buckets have to be sufficiently coarse to be incentive-compatible. All the borrowers for which there is a misalignment of incentives between the lender and the planner have to be lumped in a unique risk category that the lender is willing to fund. For this to occur, the bucket has to contain a sufficient number of low-risk borrowers alongside high-risk borrowers, and thus has to be sufficiently coarse.

The following proposition and Figure 2.4 describe the transparency regime that maximizes social welfare.

Proposition 2.4 (Welfare-Optimal Transparency Regime)**.** *Let*

$$\pi_W(b,\phi) \equiv \max\left\{0, \min\left\{\pi^\circ(b,\phi), \pi^\circ(b)\right\}\right\} < \pi_L(b,\phi), \tag{2.18}$$

where $\pi^\circ(b,\phi)$ and $\pi^\circ(b)$ are defined in Equation (A.45) and Equation (A.46), respectively, and $\min\left\{\pi^\circ(b,\phi), \pi^\circ(b)\right\}$ is decreasing in b and weakly increasing in ϕ. If the fraction of non-strategic borrowers is sufficiently high, i.e., $\pi \geq \pi_W(b,\phi)$ (resp. sufficiently low, i.e., $\pi < \pi_W(b,\phi)$), an opaque (resp. transparent) algorithm maximizes social welfare.

Proof. See Appendix A. ■

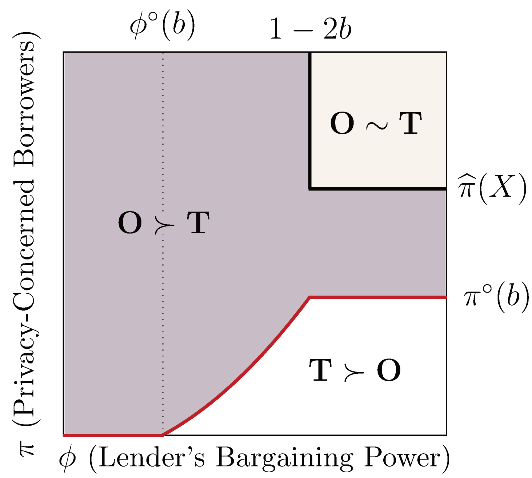


Figure 2.4: **Welfare-Optimal Transparency Regime.** The figure depicts the social planner's preferences between a transparent (**T**) and an opaque (**O**) algorithm as defined in Definition 2. The symbol $\succ$ denotes a strict preference, while $\sim$ denotes indifference. The red line represents $\pi_W(b,\phi)$ as defined in Equation (2.18), $\phi^\circ(b)$ is defined in Equation (A.39), while $\widehat{\pi}(X) \equiv \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$.

The welfare-optimal transparency regime depends on the borrower's private benefit through two key channels. First, as established in Lemma 2, the private benefit b creates a gap between socially efficient and lender-optimal credit provision. Second, combined with the lender's bargaining power ϕ, the private benefit reinforces the borrower's hedging motives under an opaque algorithm. This, in turn, impacts the extent of information available to the lender and the resulting — potentially inefficient — credit rationing decisions.

When credit provision is socially efficient (i.e., when b approaches 0), the lender's transparency choices also are. In this case, an opaque algorithm is efficient (and optimal) for two reasons. First, opacity allows non-strategic credit-worthy borrowers to withhold data without being denied credit. Second, it maximizes data extraction from strategic borrowers — since hedging motives are absent and every credit-unworthy borrower shares data — and the lender uses the disclosed information efficiently. On the other hand, under transparency, some negative-surplus projects would be funded because of gaming, while non-strategic borrowers would occasionally face rationing. It follows that no matter the lender's bargaining power and the mass of non-strategic borrowers, opacity is welfare-improving.

However, the lender's transparency decisions are socially inefficient when preferences for credit inclusion deviate, even slightly, from those of the social planner. Moreover, the welfare-optimal regime favors more opacity than the market-driven solution, i.e., $\pi_W(b, \phi) < \pi_L(b, \phi)$. The social benefit of opacity arises because it eliminates the stigma associated with no disclosure in a transparent regime, thereby enabling privacy-concerned borrowers to fund positive-surplus projects. This benefit is only partially internalized by the lender, who captures just part of the surplus generated by these projects, as b is privately enjoyed by the borrower. As a result, transparency in the credit market is excessive. This highlights a social benefit of opacity often overlooked by algorithmic transparency advocates, who usually focus on equity, fairness, and truth as the primary social objectives. When the goal is improving credit market allocative efficiency, algorithmic opacity can perform surprisingly well and promote important values such as financial inclusion and borrowers' privacy. Section 2.8 provides further insights into the policy implications of these findings.

To delve deeper into this mechanism, consider the case where the misalignment of preferences over credit inclusion is moderate, i.e. $b \in (0, b^{\circ}(X))$ where $b^{\circ}(X) \in (\frac{1}{4}, \frac{1}{3})$, meaning that data collection holds some social value by excluding strategic borrowers with negative-surplus projects. For simplicity, also suppose that opacity results in full hedging and no credit rationing ($\phi \geq 1-2b$) while transparency induces some credit rationing ($\pi < \widehat{\pi}(X)$).[31] In this scenario, the planner faces a trade-off. On the one hand, transparency maximizes the lender's ability to extract information from strategic borrowers, allowing the exclusion of negative surplus projects. On the other hand, transparency comes with a stigma around data-withholding and leads to the exclusion of those privacy-concerned but creditworthy borrowers, who would secure credit under opacity. Transparency is socially efficient when the first effect dominates, that is, when strategic borrowers are numerous enough, i.e., when $\pi < \pi^{\circ}(b)$. When $\pi \geq \pi^{\circ}(b)$, opacity is socially efficient. In this region, algorithms are overly transparent from a social standpoint, since the lender opts for transparency despite opacity being the welfare-maximizing choice (cfr. Figure 2.3). The inefficiency arises because the lender is more inclined to ration credit compared to the planner and thus only partially internalizes the benefits of expanded credit access that comes with opacity.

The lender's decisions become increasingly inefficient as the misalignment over credit provision preferences widens. When b exceeds the threshold $b^{\circ}(X)$, the cutoff $\pi^{\circ}(b)$ approaches 0 and opacity unambiguously maximizes welfare. Beyond this point, projects that are rationed under transparency generate a positive social surplus from an ex-ante perspective, so transparency has no social benefit. When $b > \frac{1}{2}$, the misalignment between the lender and the planner reaches its peak, leading to complete disagreement on the appropriate transparency regime to implement. Beyond this point, data extraction destroys

[31] Similar effects are at play when both regimes induce some credit rationing ($\phi < 1 - 2b$ and $\pi < \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$).

surplus, and the welfare optimal regime is the one that minimizes data-sharing, that is, opacity.

2.6.3 Borrower's Surplus and Redistributive Effects

We now analyze borrowers' surplus across different transparency regimes, distinguishing between privacy-concerned and strategic borrowers. Strategic borrowers' surplus depends on their observable credit risk, which influences their allocations through their data-sharing decisions. In contrast, the surplus of privacy-concerned borrowers is independent of data as they consistently withhold information.

Proposition 2.5 (Redistributive Effects)**.** *The surplus of privacy-concerned borrowers is strictly lower when the lender employs a transparent algorithm. The surplus of strategic borrowers is strictly higher when the lender employs a transparent algorithm only if their data realization is extreme, i.e.,* $z \in [0, z^\star(\pi)) \cup (1 - z^\star(\pi), 1]$ *where* $z^\star(\pi) \in [0, \frac{1}{2})$ *is weakly increasing in* π *and defined in Equation (A.68).*

Proof. See Appendix A. ∎

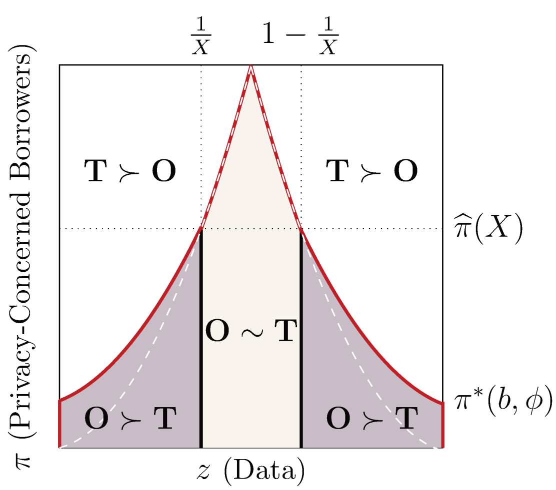


Figure 2.5: **Transparency Regime Preferred by Strategic Borrowers.** The figure depicts the strategic borrowers' preferences between a transparent ($\mathbf{T}$) and an opaque ($\mathbf{O}$) algorithm as defined in Definition 2, when $\phi \geq 1 - 2b$. The symbol $\succ$ denotes a strict preference, while $\sim$ denotes indifference. The red lines represent $z^\star(\pi)$ and $1 - z^\star(\pi)$ where $z^\star(\pi)$ is defined in Equation (A.68), while $\widehat{\pi}(X)$ and $\pi^*(b, \phi)$ are defined in Equations (A.55) and (A.62), respectively. The white dashed lines represent $\gamma(\pi)$ and $1 - \gamma(\pi)$ where $\gamma(\pi)$ is defined in Equation (2.12).

Privacy-concerned borrowers are better off under opacity. As noted in Corollary 1, a transparent algorithm imposes a negative externality on them through the stigma associated with no disclosure. Under transparency, they face higher interest rates compared to opacity and may even be excluded from credit.

Strategic borrowers favor transparency when their data is extreme and privacy-concerned borrowers sufficiently numerous. This is because transparency provides borrowers the option to withhold bad news while sharing good news, and this option value is the greatest for extreme borrowers if bad news can be easily concealed. When they share good news, they stand to gain the most, as their data provides a strong, highly indicative signal of credit quality due to its extreme nature. When they withhold bad news (and still obtain credit), they achieve the most remarkable improvement in contractual terms by being grouped with borrowers of significantly higher credit quality than their actual, observable risk level.

To elaborate on this further, consider the case where opacity leads to complete data withholding — i.e., when the lender's bargaining power is sufficiently high, $\phi \geq 1-2b$ — and the lender offers credit to all borrowers at a single interest rate.[32] This situation is reported in Figure 2.5. When privacy-concerned borrowers are sufficiently numerous, i.e., $\pi \geq \widehat{\pi}(X)$, the stigma around data-withholding is mild, and every borrower gets credit even under transparency. Central borrowers — those with $z \in (z^\star(\pi), 1 - z^\star(\pi)) = (\gamma(\pi), 1 - \gamma(\pi))$ — always share information, and because data is treated symmetrically across allocation rules, the average interest rate they receive is the same under both transparent and opaque regimes, leaving them indifferent between the two. In contrast, by withholding data, extreme borrowers obtain rates significantly lower than their observable risk would suggest, resulting in lower average rates under transparency. The improvement in interest rates grows as the gap between true and inferred credit risk widens, hence when data is more extreme and privacy-concerned borrowers are more numerous. When the stigma of no-disclosure is moderate, i.e. $\pi < [\pi^*(b, \phi), \widehat{\pi}(X))$, withholding data may lead to credit denial. The option value of data withholding reduces and central borrowers — those with $z \in (z^\star(\pi), 1 - z^\star(\pi))$ with $z^\star(\pi) < \gamma(\pi)$ — weakly prefer the security of credit provided by opacity. When unraveling forces increase even further, i.e., $\pi < \pi^\star(b, \phi)$, every borrower prefers the safe hedge of opacity.

Aggregating over credit risk and privacy types, we obtain the aggregate borrower's surplus and the following result.

Proposition 2.6 (Borrower-Optimal Transparency Regime)**.** *An opaque algorithm maximizes the ex-ante borrower's surplus.*

Proof. See Appendix A. ■

Although strategic borrowers may sometimes prefer a transparent algorithm, the overall borrower's surplus is higher when the lender employs an opaque algorithm. This happens because strategic borrowers favor transparency precisely when they are few and thus have a smaller weight in aggregate surplus. Ultimately, the security of credit provision offered by opacity outweighs the advantages of selective disclosure enabled by transparency, increasing aggregate borrower surplus.

[32] Similar effects are at play when $\phi < 1 - 2b$. See the proof of Proposition 2.5 for a formal treatment.

2.7 Empirical Implications

The above analysis offers several testable implications, with some findings supported by recent studies and others presenting opportunities for further research.

Recent papers show that information has distributional effects in credit markets (see, e.g., Liberman et al. (2019), Dobbie et al. (2020), Nelson (2023), and Jansen et al. (2024)). My paper suggests that this holds even when borrowers have control over information sharing - regardless of the algorithm's transparency regime - and data sharing may reduce credit provision compared to a scenario where no information is shared.[33] While some studies find that Open Banking enhances credit provision (see, e.g., Nam (2024)), others observe that the policy creates both winners and losers, in line with my findings. For instance, Doerr et al. (2023) show that the California Consumer Privacy Act — giving borrowers more control over their data — led to increased data-sharing with FinTech lenders, which in turn resulted in higher denial rates and greater dispersion in interest rates, despite an overall reduction in the average interest rate. Babina et al. (2025) find that borrowers their relationship banks denied credit did not see improved access to credit after the implementation of the Commercial Credit Data Sharing in the UK, as the shared data marked them as high credit risk and reduced their ability to establish new lending relationships. Similarly, Rishabh (2024) find that payment data, central to Open Banking initiatives, benefits most borrowers but disadvantages high credit risks.

In line with my differentiation between non-strategic and strategic agents, Lin (2022) empirically separates two components of consumers' privacy preferences: an intrinsic component (an exogenous taste for privacy) and an instrumental component (an endogenous economic loss from revealing private information). Critically, the study was conducted in a controlled experimental setting, where "how the instrumental incentive depends on a consumer's type is straightforward and explicitly explained" to participants, akin to a transparent regime in my model. The paper's empirical findings are consistent with my model. In particular, consumers self select into data sharing depending on the respective magnitude of the two components and exhibit gaming type of behavior: high-type consumers are more willing to share data, while low- types tend to withhold data, as intrinsic motives attenuate the receiver's equilibrium inference following no disclosure. The study could be extended to an opaque regime to investigate whether the receiver's equilibrium inference is less averse compared to a transparent setting and whether senders adopt hedging behaviors in response to opacity.

Some research has shown that intrinsic motives for privacy are particularly pronounced in credit markets. Tang (2019) studies loan applications to a Chinese Fintech lender where the disclosure of personal data is a pre-condition for loans, thus allowing her to isolate borrowers' intrinsic preferences for privacy.[34] The paper shows that stricter disclosure requirements reduce loan application completion rates, suggesting that borrowers' intrin-

[33] Note that in my model, by Assumption 2.1, the lender extends credit to the entire pool of borrowers in the absence of information.

[34] The paper shows that the disclosure group is not statistically less risky than the no-disclosure group, supporting the causal impact of intrinsic (rather than instrumental) motive for privacy on application withdrawal.

sic privacy concerns are strong enough to outweigh the potential benefits of obtaining a loan.[35] This finding aligns with my model's assumption that privacy-concerned borrowers consistently withhold information. It also suggests that unraveling forces may drive such borrowers out of the market in a transparent regime. Whether the data support this finding remains an open question for future empirical investigation.

My analysis provides fresh insights into earlier studies on borrowers' data-sharing behavior. Notably, Nam (2024) examines borrowers' voluntary data-sharing decisions with a major German FinTech lender. The study takes place under a relatively opaque regime, as the platform discloses that shared information ("behavioral data, web data, and experience data") could potentially lead to credit denial or higher interest rates but offers little explanation of the underlying decision logic. First, the study finds that the average rate of data sharing was relatively low, at 8% across the entire sample period. Second, it finds no evidence of lender's equilibrium inference in response to data withholding, as an increase in data sharing results in a negligible negative impact on loan approval for those who choose not to disclose. While these findings are inconsistent with the standard unraveling logic (e.g., Grossman (1981) and Milgrom (1981)), they align with the equilibrium proposed in Proposition 2, where borrowers hedge against opacity by withholding information, and the lender does not penalize the absence of disclosure.

Recent papers provide suggestive evidence for the primary mechanism in my model, which is that firms leverage their transparency regime to maximize data collection (Proposition 3). First, in light of my theory, firms with greater bargaining power should be more transparent to mitigate borrowers' hedging motives. Relatedly, Ramadorai et al. (2021) show that larger firms are more likely to have a privacy policy, to display it visibly, and to write longer and more sophisticated policies (conditional on having one). These firms also engage in more extensive data extraction, utilizing a greater number of cookies to track consumer behavior. These findings support the notion that transparency is a strategic tool to reassure users and enhance data extraction. Second, my model suggests that firms facing more privacy-concerned borrowers should be more opaque. Relatedly, Bian et al. (2021) show that mandated disclosure of privacy labels in the AppStore negatively impacted firms' profits, with this effect being more pronounced in countries with stronger privacy concerns. In other words, the lack of transparency in firm data collection practices led consumers to share excessive data. This suggests that firms have stronger incentives to be opaque in markets with stronger privacy concerns. Although these findings center on input data transparency rather than model transparency, they indicate that firms strategically optimize transparency to maximize data collection.

My model could inform further empirical investigation on this mechanism. According to it, financial institutions with greater bargaining power should reveal finer details about the inner workings of their algorithm — not only about the input data they use — to mitigate mistrust and foster information sharing. Conversely, financial institutions facing borrowers with stronger privacy concerns should reveal less about their algorithm's operating logic

[35] As the author acknowledges, all Chinese platforms commonly ask for non-standard personal information from borrowers, as there is no official credit score system and there are limited resources available to verify borrower credentials and documents. Hence, Chinese borrowers have limited ability to substitute for other, less privacy-intrusive platforms.

to prevent gaming behavior. In addition, my theory offers several testable implications on the link between algorithmic transparency, data sharing, and credit market outcomes that have not yet been explored by empirical literature. In Appendix A, I present closed-form computations of key observables in my model (such as the amount of data shared, levels of credit provision, and interest rates) to guide hypothesis testing.

The above discussion raises the question of how to measure primitive variables in my model, particularly an algorithm's transparency regime. Computer scientists have developed quantitative indicators of AI transparency based on publicly available information. For example, the Foundation Model Transparency Index (see Bommasani et al. (2023)) codifies AI transparency by incorporating various aspects, including data sources, model development, and decision-making processes, and reveals significant variation across firms. A similar index could be tailored specifically for credit scoring models. More qualitative measures of transparency could include explainability scores, which assess how well the algorithm's decisions can be understood, combined with disclosure levels, that is, the extent to which information about the algorithm — including data used, methodologies, and underlying logic — is made publicly available on a lender's website. The staggered implementation of the AI Act (see Section 2.8) could introduce an additional source of variation. The second variable of interest is borrowers' privacy concerns, for which the literature provides various proxies. Research has demonstrated that consumers' privacy concerns in online environments correlate with several factors, including cultural influences (such as trust in institutions), demographic variables (like age, education level, and income), trust in the firm (encompassing reputation, credibility, and brand recognition), and the nature of the data shared (including breadth, depth, and sensitivity).[36] In the credit market, privacy concerns vary across countries, types of data, demographic and income groups (see Tang (2019), Babina et al. (2025), Doerr et al. (2023) and Nam (2024)). Lastly, a lender's bargaining power can be effectively measured using traditional indicators of market power and concentration, such as market share or the Herfindahl-Hirschman Index.

2.8 Policy Discussion

The EU Artificial Intelligence (AI) Act leads global efforts to regulate the development, deployment, and use of artificial intelligence technologies. The regulation entered into force on August 1st 2024 and shall apply from August 2nd 2026 (see Article 113). According to the Act, credit scoring and credit underwriting models are classified as high-risk AI systems (see Article 6(2) and Annex III, 5(b)). As such, they are subject to various transparency requirements aimed at ensuring accountability and clarity in their operations. In addition to mandating transparency towards deployers and the Commission,[37] the Act imposes both

[36] See Acquisti et al. (2016) for a general discussion and Morey et al. (2015), or Armantier et al. (2021), or Prince and Wallsten (2022) for surveys.

[37] Credit scoring models should be explainable, that is, they "should be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately" (see Article 13). Moreover, models should be auditable, as the Commission can require firms to provide technical documentation, as well as a general description of the AI model - including the architecture, the number of parameters, the modality and format of inputs

ex-post (Article 86) and ex-ante (Article 71) disclosure requirements towards the general public. Article 71 specifies that the Commission will collect information from developers to set up and maintain a database containing, among other things, a "description of the information used by the system (data, inputs) and its operating logic" (see Annex VIII, Section A(6)). This information should "accessible and publicly available in a user-friendly manner".

My model suggests that the transparency requirements of the AI Act may impair the credit market's allocative efficiency in some cases, enabling gaming behavior and disproportionately harming privacy-concerned borrowers, somehow at odds with the regulator's stated commitment to privacy in frameworks like the GDPR. This finding enriches the current policy debate surrounding the AI Act by offering some high-level insights. First, it suggests that the allocative efficiency of the impacted markets should not be overlooked. Currently, the regulation's core guiding principles — human oversight, accountability, safety, fairness, and non-discrimination — focus primarily on ethical and safety considerations, ignoring implications for market efficiency. Second, an industry-based approach should complement the existing risk-based, technology-neutral framework. In the finance industry, lenders' preferences for financial inclusion often conflict with social efficiency, leading to inefficient transparency choices by financial institutions. The gap between social and private goals may be narrower in other industries, so an industry-based approach could help regulators better allocate limited resources for more effective regulation. Third, regulations on algorithmic transparency should be implemented in tandem with data privacy regulations, such as the GDPR. Algorithmic transparency directly influences individuals' privacy choices, making it essential to coordinate both areas to ensure comprehensive protection for users.

My model also offers concrete recommendations for regulators, helping them establish more explicit guidance on compliance expectations within the financial services industry.

- **Selective Transparency** Algorithmic transparency should only be mandated when a lender's market power is significant enough to prompt borrowers' hedging behavior against opacity (e.g., for BigTechs) and for information for which borrowers have minimal privacy concerns (e.g., basic demographic information). When a lender's market power is limited or privacy concerns are heightened, preserving opacity enhances market efficiency.

- **Operating Logic** Regulators should clarify what is meant by an algorithm's *operating logic* in Article 71. My model suggests that disclosing the *direction* in which data influences predictions is an effective and practical approach to transparency. It can be shown that revealing only the sign of λ to borrowers results in the same credit allocation as disclosing its exact value. Intuitively, knowing whether a variable's higher values are seen as good or bad news is sufficient to create a stigma

and outputs - and a detailed description of the elements of the model - including the design specifications, training process, training methodologies, training techniques, design choices, assumptions made, what the model is designed to optimize for, training, testing and validation datasets, their scope and main characteristics (see Article 91, Article 53, and Annex XI).

around non-disclosure, triggering unraveling forces. Providing more granular information about λ is sometimes impractical, especially with complex machine learning models, while direction-based disclosure is feasible thanks to recent explainable AI techniques. This laxer requirement would enable financial institutions to use high-performing algorithms without compromising their screening ability. On the other hand, direction-based disclosure is essential for achieving transparency. Leaving uncertainty about whether data signals good or bad news prevents unraveling forces and activates hedging behavior, ultimately preserving opacity.

- **Access to the Database** Regulators should require financial institutions to provide borrowers direct access to the database mentioned in Article 71 during loan applications or, ideally, have lenders disclose the necessary information on their websites. In light of my model, transparency can be more effectively achieved when borrowers understand the algorithm's operating logic at the time of the data-sharing decision and when lenders are aware that borrowers possess this information. Establishing the database without ensuring borrowers are informed about it could lead to situations where only some borrowers are aware of its existence or where only a few are willing to bear the costs to access the information. This would potentially create unequal access to the database, leading some borrowers to operate under transparency while others under opacity and complicating the lender's equilibrium inference.

- **Price for Data** In addition to promoting transparency, regulators should encourage financial institutions to implement measures for screening borrowers' privacy types. This perspective aligns with the ongoing debate about introducing a price for personal data (see, e.g., Acquisti et al. (2016)). In my model, within the transparency regime, identifying whether a borrower is privacy-concerned or strategic would enable the lender to extend credit to creditworthy privacy-concerned borrowers who might otherwise be excluded from the market due to negative externalities (see Corollary 1). On the other hand, unraveling forces would induce optimal, but sometimes inefficient, credit rationing of strategic users. Hence, screening always enhances the lenders' profits (net of agency rents) and, when the first effect dominates, also improves social welfare. Some financial institutions are already taking steps in this direction. For instance, Auxmoney, a prominent German FinTech lender, features what appears to be a menu of contractual terms on its website designed to encourage disclosure from borrowers who are less concerned about privacy. The lender showcases that sharing personal data with the platform may result in an *average* discount of €390 on a €5.000 loan, even though interest rates could increase or applications might still be declined (see Nam (2024)).

2.9 Conclusion

This paper studies whether credit risk algorithms should be transparent or opaque, considering their impact on borrowers' data-sharing decisions. Transparency exposes the lender's model to gaming through strategic withholding of unfavorable information. Opacity mitigates gaming but leads borrowers to withhold information as a hedge against the unpredictability of the black box. The lender's transparency choices are aimed at data extraction and are thus influenced by the lender's bargaining power, which affects borrowers' hedging motives, as well as borrowers' privacy concerns that enhance their ability to game the system. These choices often lead to inefficiencies due to the lender's excessive inclination toward credit rationing. Surprisingly, algorithmic opacity is often socially efficient because it removes the stigma associated with non-disclosure, thereby promoting credit access for privacy-concerned, creditworthy borrowers.

Several questions remain open for further research. First, the optimal transparency regime may lie between full transparency and full opacity. Second, the analysis may be extended to data used in the monitoring process – rather than screening – or to more complex financial contracts. Third, competition may influence the lender's strategic choices of opacity. Lastly, algorithmic opacity could be examined in other financial contexts, such as algorithmic trading and portfolio management, as well as in other markets, including online ranking, targeted marketing, health care, demand forecasting, and fraud detection.

Chapter 3

Market Information in Banking Supervision: the Role of Stress Test Design

Market Information in Banking Supervision: the Role of Stress Test Design

Haina Ding, Alexander Guembel, Alessio Ozanne

Abstract

The Basel committee views market discipline as complementing banking supervision. This paper studies how supervisors should design stress tests when markets discipline banks via price signals their traded securities provide to bank creditors. We show that the optimal stress test is coarse and lenient. Speculators have incentives to identify bad banks that erroneously passed the test, which makes markets useful at reducing the type-2, but not the type-1, error of a stress test. Our results hold even when the supervisor can intervene directly based on private information. In the limit of costless supervisory interventions, the optimal stress test is uninformative.

Keywords: Feedback, Market Discipline, Information Design.
JEL Classification: G14, G28.

3.1 Introduction

The Basel Committee on Banking Supervision elevates *market discipline* to one of its three pillars of the prudential regulation of banks.[1] In this lies an acknowledgement that markets may achieve things that regulators cannot. Market discipline is often thought of as having two broad roles (see, e.g., Kwan (2002)): a direct role by restricting undeserving banks' ability to access capital, and an indirect, informational role.[2] In both cases, supervisors can presumably not easily replicate what markets achieve. A supervisor may find it costly to impose directly the penalties corresponding to the cost of modified funding conditions, and markets can provide information that supervisors cannot.[3] Market prices

[1] The other two pillars are, loosely speaking, capital requirements (see Flannery (2014) and Ngambou Djatche (2022) for overviews) and supervisory monitoring (see, among others, Colliard (2019), and Carletti et al. (2021)).

[2] The distinction between direct and indirect market discipline is related to the *monitoring* and *influence* functions of markets, identified by Bliss and Flannery (2002). Market discipline can affect banks in a variety of other ways. See Flannery and Bliss (2019) for a detailed discussion and overview of research on market discipline.

[3] Acharya et al. (2014) and Haldane (2011) provide evidence that simple market capitalization based measures of bank health were better than regulatory measures at identifying banks that eventually experienced distress. Berger et al. (2000) show that both stock and bond prices are more accurate than supervisory assessments. As Flannery and Bliss (2019) argue: "We believe that market discipline can, potentially, complement and support official oversight of risky financial institutions, [...] by providing market signals that supervisors can use to motivate their own actions..."

are free and forward-looking, generated by speculators with monetary incentives that are hard to replicate by supervisors facing up to increasingly complex banks (see Goldstein (2023), and the citations therein, e.g., Stern (2001)). Unfortunately, the Basel Framework provides little help in identifying how supervision can be designed to leverage whatever it is that markets are good at. Pillar 3 largely reduces to a recommendation to improve disclosure, so as to facilitate direct market discipline. By this logic, more disclosure of supervisory information, including the disclosure of stress test results, fosters direct market discipline.[4] This leaves open the question of how the design and disclosure of stress test results affects jointly the direct *and* indirect role of market discipline. This paper explores the informational spillovers from supervisory information production, notably in the form of stress tests, to market discipline in its direct and indirect roles.

We provide a model in which banks can be of two, privately known types. Both types try to raise funds to invest in a risky loan portfolio, but it is only efficient for a high-type bank to do so. The supervisor is concerned not to allow low-type banks to engage in wasteful risk taking.[5] A supervisor can generate noisy information about the bank type by studying its resilience to a stress test. The supervisor commits to a set of stress scenarios to which the bank is subjected and then publicly discloses the test results. This is followed by direct and indirect market discipline. Indirect market discipline, operating via informative price signals, is arguably most effectively exercised by traded claims, such as equity due to the ready availability of data and the liquidity of the underlying markets.[6] We model this by having a speculator decide how much costly information to produce, after having observed the stress test result. He then trades in the bank's shares which generates noisy price signals. Direct market discipline is mostly exercised by capital providers, such as short-term creditors, who need to renew their funding commitment frequently (see Kwan (2002)). We assume that the supervisor finds it too costly to intervene directly in funding or shutting down banks, an assumption we relax in an extension. She therefore relies on capital providers, such as uninsured depositors, to deny funding to low-type banks and provide funding to high type ones. Capital providers decide on the funding terms based on the information contained in the stress test result and the subsequent stock price signals. To summarize, we allow markets to complement banking supervision in two ways: indirectly, by producing information about banks and directly, by withdrawing funding from potentially undeserving banks, conditional on the available information.

We first study a setting in which the supervisor's cost of intervention is high, so she relies exclusively on market discipline to avoid low-type banks from getting access to funds that would then be squandered. We show that the optimal stress test design is a coarse *pass*

[4]For example, Bernanke (2013) argues "[...] the disclosure of stress test results and assessments provides valuable information to market participants and the public [...] and promotes market discipline." As Flannery and Bliss (2019), however, point out "The appropriate relationship between market and regulatory discipline has never been fully developed - at least not in official documents."

[5]This way of modelling a low-type bank is consistent with regulators' fear that some banks will engage in zombie lending if they can raise funds. See Acharya et al. (2019) for empirical evidence on the prevalence of zombie lending.

[6]The literature has also highlighted the potential role of sub-ordinated bonds (for an excellent discussion of this point, see Flannery and Bliss (2019). Our model is sufficiently stylized to make this distinction moot.

/ *fail* test that exhibits leniency. We demonstrate that stock markets are not equally good at identifying type-1 and type-2 errors of the stress test. Markets are better at identifying bad banks that were erroneously classified as good by the stress test (type-2 error), but they are less useful at identifying good banks that the stress test has mistakenly classified as bad (type-1 error). This asymmetric reaction is the result of the speculator's financial incentives to acquire and trade on information, depending on whether a bank did well or poorly in the stress test. Banks that fail a test are less likely to obtain funding, which reduces the scale of their operations, making information production less valuable. The opposite is true for banks that do well in the stress test: they are likely to get funding and have relatively larger operations, increasing the speculator's potential trading profits. A lenient stress test design makes it more likely that a bank passes the test. This has the advantage of improving the quality of the price signal upon which direct market discipline is based. However, distorting the test towards leniency also has a cost. Since the price signal is noisy, the positive stress test outcome will sometimes allow undeserving banks to obtain funding. Due to this trade-off, the optimal design features leniency, but not to the point of rendering the test uninformative. Other papers, discussed in more detail below, have shown that supervisors may employ lenient stress tests, either because they suffer from ex post forbearance, or are driven by a concern about inefficient bank runs. Our set-up features neither of these elements and is instead based on the interaction between the supervisor's direct information supply and the information environment in which market discipline operates.

We also show that a *pass/fail* structure of the test is optimal, although the supervisor could have chosen an arbitrarily granular test design. First, the optimal test, by virtue of being lenient, must be somewhat coarse. That is, for a *pass* test to be credible, banks with low(ish) resilience that would correspond to a marginal *fail* need to be lumped into the same test outcome as significantly more resilient banks. Second, a more granular information design reduces the average amount of information produced. Suppose the stress test had, in addition to a *fail* outcome, two *pass* levels, a moderate and a strong *pass*. Following a moderate *pass*, uninsured depositors would demand a fairly high interest rate. This dilutes equity, reduces trading profits and thereby dulls the speculator's information production incentives. It is therefore better to enlarge the moderate *pass* to include more resilient banks, which reduces the expected interest rate that banks in that category have to pay, thereby improving information production incentives.

Some papers have argued that stress tests should be lenient or kept confidential so as to prevent runs on banks (see Williams (2017) and Bouvard et al. (2015)).[7] At the same time, recent crisis episodes have shown that supervisors are also worried that sharp declines in a bank's stock price may itself trigger a run. Regulators have therefore at times prohibited short sales of bank stocks during crisis periods, somewhat in contradiction to their stated commitment to market discipline.[8] The question of how to design a stress test

[7] In these papers the supervisor's objective is for all banks to be funded, i.e., there is no role for market discipline.

[8] The Securities and Exchange Commission (SEC) imposed restrictions on short sales of bank stocks during the 2007-09 financial crisis, as did several European regulators during the 2011 sovereign debt crisis. Beber and Pagano (2013) show that short-sales bans slowed down price discovery in these markets.

remains open, when the supervisor must, on the one hand, worry about potential crises in which fundamentally sound banks risk not being able to roll over their debt, and, on the other hand, wish to allow market discipline to operate. We extend our model to capture the possibility of a bank run or debt overhang. We do so by making creditors potentially deny funding to banks that the supervisor would like to see funded. In other words, for banks to obtain funding, creditors must have more positive beliefs about its type than the belief threshold the supervisor would apply. We show that, in a benchmark where there is no scope for information production by the speculator, debt overhang makes the optimal stress test more lenient.

Compared to this benchmark, the optimal stress test can be more lenient or more conservative when endogenous speculator information production is introduced. The argument for more leniency remains the same as before: it increases the likelihood of a *pass* with a resulting positive effect on information production. However, there is now also a counter-veiling effect. Since the benchmark *pass* grade is already quite lenient, and funds are provided by the market, the funding cost of a bank that passes the test is quite high. This dilutes equity holders and reduces information production, as mentioned above. In order to encourage information production, the supervisor may optimally apply a test that is more conservative than the benchmark to ensure that banks' funding costs are moderate with a correspondingly positive impact on market information and discipline. We show that the supervisor optimally sets a test more conservative than the benchmark when the gulf between social preferences and creditor preferences is particularly wide.[9]

We also extend our model to allow for direct intervention by the supervisor, when the latter has noisy private information beyond the stress test result. A direct intervention can take two forms. Either the supervisor shuts down a bank, even though markets would be willing to fund it, or the supervisor provides public funding to a bank that is denied access to funding from private markets. We assume that intervention is costly so that, for a high enough cost, our extension nests the baseline model in which the supervisor never intervenes directly. When the cost of intervention is zero, direct market discipline loses its purpose and the optimal stress test degenerates to become completely uninformative. However, indirect market discipline remains useful: The supervisor, whose private information is noisy, can still learn from price signals. Since the supervisor's intervention decision is no longer directly related to the stress test outcome, passing the test ceases to be a pre-condition for stimulating information production by a speculator. In this case, the supervisor wants to make as little information available as possible, since any public information merely crowds out the speculator's private information. When the banking system is in good shape, such that creditors are willing to roll over debt in the absence of any information, the optimal stress test is informative and lenient for any strictly positive intervention cost, no matter how small. Intuitively, the supervisor wishes to enlist market discipline to save

A ban on short sales can be justified by their potentially manipulative nature, as shown by Goldstein and Guembel (2008), Brunnermeier and Oehmke (2014) or Gao et al. (2024). We do not analyse this possibility here. See also Acharya et al. (2011) for a detailed discussion of how rollover risks can lead to market breakdown.

[9] It is conceivable that the supervisor would like to apply a harsher funding rule than credtiors, for example if there are default externalities. We analyse this case in Section 3.7.

on her intervention cost. Since price signals are noisy, having some information from the stress test allows for more effective market discipline. When the banking system is in worse shape, such that creditors' default decision would be not to roll over debt in the absence of any information, a completely uninformative stress test is optimal for small, but strictly positive intervention costs and becomes lenient for higher intervention costs. The supervisor is more reluctant to design a highly lenient stress test, because under such a test, and given the poor health of the banking system, a *pass* grade is not sufficient for a bank to roll over its debt. To make it useful, the supervisor would have to raise the stress test's *pass* threshold to a high level. A *pass* is then so informative that it leaves little scope for an informational advantage to the speculator. At the same time, an uninformative test induces speculator information production at low levels of intervention costs, since the supervisor will often intervene and fund the bank. Hence, when the intervention cost is low, it is better to induce speculator information production with a completely uninformative test, while an informative and lenient design becomes optimal when the intervention cost increases.

The remainder of the paper proceeds as follows. We provide a review of the literature in Section 3.2. Section 3.3 provides the description of the model, which is solved in Section 3.4 for the benchmark case without informational feedback from the stock market. Section 3.5 presents the main results on stress test design with market feedback. We extend the model to allow for debt overhang and default externalities, respectively, in Section 3.6 and Section 3.7, and introduce supervisor private information in Section 3.8. Section 3.9 concludes. All the proofs are relegated to Appendix B

3.2 Related Literature

There has been considerable interest in recent years in the question how information conveyed by prices in secondary financial markets feeds back into real decisions (see Bond et al. (2012a) and Goldstein (2023) for surveys). One application of that literature points to the importance of stock price information in guiding intervention decisions of regulators, for example, a supervisor who needs to decide whether to intervene in a troubled bank (Bond et al. (2010), and Bond and Goldstein (2015)). The papers closest to ours are Bond and Goldstein (2015) and Siemroth (2019) who study the interaction of a regulator's information (including a decision to disclose such information) with information revealed by share prices, when that information is in turn used by the regulator. They show that more public information may crowd out private information as it reduces the informational advantage of speculators.[10] This effect is balanced by a crowding-in effect, as public information reduces the riskiness of speculators' trades, inducing them to take larger positions. Also related is Goldstein and Yang (2019) who study the interaction between public disclosure and market-based information in a context where the decision maker learns from both the public signal and market prices (unlike in Bond and Goldstein (2015) where the regulator has information regardless of whether or not it is made public). Goldstein and

[10]Recent empirical evidence by Heitz and Wheeler (2023) supports the notion that the information contained in stress tests does indeed crowd out information production by financial markets.

Yang (2019) focus on two dimensions of uncertainty and explore how disclosure affects the weight that traders put on one of the two private signals they possess. They show that when information is disclosed about the dimension of uncertainty that is relevant for the real decision, then this will reduce the weight that traders put on that dimension of their private signals. By crowding out information aggregation on the "useful" dimension, more public disclosure may reduce the overall amount of information relevant to the real decision.

Our focus is different in several respects. First, we focus on endogenous information production and not on the aggregation of an exogenous information endowment by speculators. Second, we model feedback from prices via a bank's access to funding. This is important because the bank's expected funding cost affects incentives for information production. Finally, we study information design in a way that allows us to identify leniency and coarseness as decision variables. The papers by Bond and Goldstein (2015), Goldstein and Yang (2019) and Siemroth (2019), share their focus on the *intensity* with which speculators trade on their private information. They, like many others, use variants of the Grossman-Stiglitz framework that assume normal distributions and thereby preserve the quasi-linearity of trades, which is a key property for tractability. That framework, however, has a very specific property: Residual uncertainty from the speculator's perspective is independent of the *realization* of the public signal. In this context, information design reduces to choosing the standard deviation of signal noise. This property makes the framework arguably less well suited to studying trade in non-linear claims such as highly leveraged bank equity. Quite plausibly, residual uncertainty is smaller for lower signal realizations, i.e., when the expected equity payoff is nearer the default region. This induces very different incentives to acquire information, depending on which part of the payoff distribution a speculator expects to navigate.[11] In our paper, the information production decision therefore depends sensitively on the *realization* of the public signal, with less information being produced following a negative public signal than following a positive one.

The effect that trading profits differ, depending on whether the outlook is positive or negative, is related to Dow et al. (2017) who show that speculators' information production may break down when firms' investment prospects are unfavorable. Such firms are unlikely to invest, which therefore undermines the incentive for speculators to produce information about those prospects. Deng and Shapiro (2024) identify feedback via consumer learning as a further channel that can affect the information sensitivity of a firm's shares, including a degenerate case where firm profits become independent of the underlying state of the world. In this paper we focus on the *ex ante* information design problem when information production in financial markets depends on the trader's belief about fundamentals. Moreover, the information environment is designed by a planner who cares about *ex ante* bank value, while market information is produced by stock traders. Since equity claims are protected by limited liability and diluted by the bank's fund-raising, this introduces a wedge between the payoffs that are relevant for the planner and the information producer. As such, stock traders have little incentive to produce information about banks with resilience levels close to and below the threshold where they can obtain funding. Those are,

[11] The cost of moving away from the normal, quasi-linear framework, is that we cannot study the speculator's trading intensity, which is the main focus of the above papers.

however, precisely the banks that the planner would most like to learn about. While we are by no means the first to point out that private incentives for information production differ from social value (see Paul (1992), or Lenkey and Song (2017), for a more recent example), we identify a new wedge between the two.

Davis and Gondhi (2024) analyse risk shifting with informational feedback from the stock market. Their focus is different from ours in that they explore how an agency conflict between debt and equity holders interacts with the endogenous information available in the stock market. They show that the relationship depends crucially on whether investment distortions are of a risk shifting or a debt-overhang type. As risk shifting in their model increases speculators' incentives to produce information, the feedback mechanism mitigates the inefficiency caused by the agency problem.

There are a number of papers that have studied whether stress test results should be disclosed, e.g., Bouvard et al. (2015), and Leitner and Williams (2023) (see also Goldstein and Sapra (2014) and Goldstein and Yang (2017) for a more general discussion and review). Disclosure matters, as it may affect market discipline, the functioning of the inter-bank market, financial stability, bank lending behaviour and risk sharing. Our model can be re-interpreted as a disclosure choice: Since the supervisor relies on markets to discipline banks, all that matters is publicly available information. Our results thus suggest that partial disclosure dominates full disclosure of stress test results. In an extension we allow the supervisor to act, at a cost, on private information. This explicitly introduces a meaningful difference between what the supervisor knows (a bank's precise resilience level) and what she discloses (a stress test result which corresponds to a region in which the resilience level lies).

Some papers have modelled the disclosure choice as a Bayesian persuasion problem, that is, a supervisor chooses an information design to which she commits. One common theme among those papers is a supervisor's concern to design a stress test in a way to prevent bank runs (see Faria-e Castro et al. (2017), or Williams (2017)). In line with Kamenica and Gentzkow (2011), this pushes the optimal experiment to be of a *pass/fail* nature, featuring a maximum of *pass* grades consistent with avoiding a run. The optimal stress test is lenient, in the sense that it admits type-2 errors (some low-type banks passing the test), but no type-1 errors (no high-type banks failing the test). Some papers extend the basic Bayesian-persuasion-*cum*-global-games approach, while remaining in relatively abstract settings that are not specifically geared towards modelling bank stress tests. Quigley and Walther (2023) study how a publicly disclosed stress test affects a bank's incentives to privately disclose verifiable information at a cost. They show that private disclosure may lead to unravelling, which the stress test can preempt by applying a richer message space than a simple *pass/fail*. Inostroza and Pavan (2023) look into robust information design in a global games framework with privately informed agents. The optimal policy coordinates all market participants on the same course of action, but without fully revealing the state. Under some conditions, the optimal policy is a *pass/fail* stress test.

Other papers add flesh to the Bayesian persuasion approach by modelling the details of financial frictions faced by banks. Faria-e Castro et al. (2017) show that the opacity implied by the test can generate an adverse selection cost at the fundraising stage. They investigate the optimal test design as a function of a country's fiscal capacity, when the regulator can trade off more transparency against the fiscal costs of guarantees that prevent bank runs. Goldstein and Leitner (2018) show that a more informative stress test may destroy insurance opportunities among banks. This potentially renders no disclosure of stress test results optimal. Inostroza (2023) studies stress testing with multiple audiences, such as short-term creditors and shareholders. He shows that the optimal policy is opaque when the bank has high-quality assets, and transparent when the bank has poor-quality assets. Full transparency is optimal because the complementarity in incentives to provide funds between the two types of capital providers generates a convexity in bank value as a function of the underlying fundamentals. Orlov et al. (2023) show the optimality of *pass/fail* tests failing all weak and some strong banks in order to limit the stigma of failure. The optimal test is not fully informative, because banks are subject to a convex cost of distress, which renders bank value concave in its initial wealth. Fuchs et al. (2024) study the interaction between *ex ante* rules and *ex post* disclosure. They show that regulation helps *ex ante* incentive provision, while *ex post* disclosure serves to provide insurance. Overall, the conclusions from these papers are quite nuanced as optimal stress test design depends sensitively on the precise financing frictions faced by banks.

Although we share with the above papers the feature that a supervisor chooses an information design and then commits to it, our focus is on optimal learning. As such, we assume that the supervisor is limited to noisy experiments, as in Parlatore and Philippon (2022). The choice of scenario adversity then affects the trade-off between type-1 and 2 errors of the experiment. A more adverse scenario increases the probability of a type-1 error (mistakenly classifying a good bank as bad) and reduces that of a type-2 error (mistakenly classifying a bad bank as good).

Shapiro and Zeng (2024) study the reputational implications of stress test design for a supervisor. A supervisor may design either a lenient or a tough stress test, depending on whether she wishes to build a reputation for being soft or tough. This approach is closer in spirit to Bouvard et al. (2015) and Parlasca (2024) where, in contrast to Bayesian persuasion, the supervisor chooses information revelation strategically, *after* having become privately informed herself.

Our paper is also related to the literature on banking regulation which regards a moral hazard problem at the bank level as a central friction to address by regulation, for example, Bhattacharya (1982), Rochet (1992), Hellmann et al. (2000a), Gorton and Huang (2004), Morrison and White (2005), Calzolari and Lóránth (2011), Calzolari et al. (2019) or Fecht et al. (2022). In Carletti et al. (2021) banks take too much risk in a laissez-faire equilibrium and supervision is designed to reduce their risk exposure. The supervisor monitors and learns about the amount of a bank's capital (and its portfolio) and can then intervene so as to reduce risk exposure. When an intervention occurs, shareholders are expropriated. Our model is similar in spirit, except that the supervisor finds it costly to intervene directly and therefore wishes to enlist market discipline. High leverage associated with a poorly capitalized bank could also lead to debt overhang, a problem addressed by Philippon and Schnabl (2013) who analyze the efficient design of a recapitalization when the regulator does not know the bank's type. We extend our model to analyse the possibility of debt overhang. Our central point on designing the supervisor's information and its interaction with market-based information and discipline is new to this literature.

Also somewhat related are models on the design of credit rating agencies' evaluation scheme. Goldstein and Huang (2020) predict that CRAs inflate ratings in a model where creditors' heterogeneous beliefs affect credit market conditions, which in turn generates a feedback loop from the CRA to the firm's actual investment decisions.[12] Apart from the difference in focus, Goldstein and Huang (2020) have in mind a CRA without commitment power over its rating announcements, so ratings are subject to ex-post opportunism by the CRA. Moreover, in their paper the issue of information production by speculators or other market participants does not arise, as creditors have an exogenous information endowment. Piccolo and Shapiro (2022) look at a CRA who is subject to a moral hazard problem in information production. Informative stock prices serve to mitigate the agency problem. Higher ratings precision reduces information production.[13]

3.3 The Model

We begin with a brief overview of the model. There are five dates $t = 0, ..., 4$. At the initial date $t = 0$, a banking supervisor designs a stress test. The outcome of the stress test is publicly observable at date $t = 1$. Afterwards, a speculator decides how much effort to expend on information acquisition. At date $t = 2$ the speculator can trade in the bank's shares and prices are publicly observed. Then, at $t = 3$, providers of capital, such as uninsured depositors, choose whether and at what terms to roll over credit to the bank. If the bank can roll over the credit, it invests in a risky loan portfolio. Payoffs are realized at the final date $t = 4$.

We now describe the full model. There is a state of the world ω, which can take the values l and h with equal probability. The state ω realizes at $t = 0$, is unobservable and determines whether the bank is worth funding ($\omega = h$) or not ($\omega = l$). We can think of the model as applying to a single bank, or to many ex-ante identical banks. In the latter case, ω should be interpreted as specific to banks, i.e., we do not model learning about an aggregate shock from conducting stress tests across many banks (see Parlatore and Philippon (2022), or Parlasca (2024), for learning about aggregate shocks). Although the bank's true type ω is not directly observed, there is a learnable characteristic that is correlated with the bank's type ω, which we call the bank's resilience $s \in [0, 1]$. For analytical tractability, we assume

$$\begin{aligned} f(s|\omega = h) &= 2s, \\ f(s|\omega = l) &= 2(1 - s), \end{aligned} \tag{3.1}$$

[12] Terovitis (2020) models a similar feedback loop from credit rating to project financing where managers have private information about the project quality.

[13] Note that private information in the loan market may also be transmitted through interest rates, which act to coordinate banks' actions in supplying credit to the real economy (see Shen (2021)).

with corresponding cumulative distributions $F_\omega(s) \equiv F(s|\omega)$. The supervisor designs a stress test at $t = 0$, which reveals information about the bank's resilience s.[14] Both the test design and its outcome are publicly observed. A stress test is formally defined as follows.

Definition 3.1 (Stress Test and Outcome)**.** *A stress test is a partition $S = \{s_0, s_1, \ldots, s_n\}$ of the stress resilience space $[0, 1]$ with $0 = s_0 < s_1 < \cdots < s_n = 1$. An outcome of the stress test S is a public signal m_i for $s \in [s_{i-1}, s_i)$, $i \in \{1, \ldots, n\}$.*

Suppose, for example, that the supervisor chooses a partition $\{0, s_1, s_2, 1\}$ where $0 < s_1 < s_2 < 1$. This stress test can be interpreted as consisting of two scenarios s_1 and s_2. The supervisor can first apply the more benign stress scenario s_1 which will result in either a *pass* or a *fail*.[15] If the bank fails scenario s_1 the supervisor knows that the bank's underlying resilience is quite weak ($s < s_1$). If the bank passes scenario s_1 the supervisor can apply the more adverse scenario $s_2 > s_1$. If the bank passes scenario s_1 but fails scenario s_2 the supervisor knows that the bank's resilience s is in the interval $[s_1, s_2)$. If, however, the bank passes both scenarios, the supervisor knows that resilience is $s \geq s_2$. Note that from (3.1) it follows that all banks, even $\omega = l$ types, pass the most lenient stress scenario 0 and all banks, including the $\omega = h$ types, fail the most adverse scenario, given by 1.

Note that our definition of stress test S is quite flexible. In particular, since the supervisor can run as many scenarios as she wishes at no cost, we allow for complete learning of s, i.e. fully granular grades ($n \to \infty$). Alternatively, the stress test may provide no information about s ($n = 1$), or learning s noisily (n finite).[16] Nevertheless, we impose two notable restrictions. First, we require the stress test to be monotone, i.e. two disjoint intervals cannot produce the same test outcome. This assumption is motivated by the fact that the stress test consists of a sequential application of stress scenarios. This way of learning is plausible and rules out that an observer may believe that resilience can be high or low, but not in the middle. Second, in some of the Bayesian persuasion literature, the supervisor can condition the public signal m_i directly on the state of the world ω, such that the full revelation of the true state is possible. We rule this out by assuming that a bank's resilience s is itself only a noisy signal of ω. This is intended to capture real world limitations to how informative stress tests can be.[17]

[14]Since the supervisor does not intervene directly, it does not matter whether she observes s or learns about s from the stress test. In Section 3.8 we allow the supervisor to intervene at a cost and observe s independently of the stress test design.

[15]The *pass/fail* nature of the response to an individual scenario is only for illustrative purposes. Since a stress test is an arbitrarily granular partition, the overall stress test can be much more nuanced than a simple *pass/fail*.

[16]Strictly speaking, with n being an integer, the stress test cannot fully reveal s which is a real number. However, since the limiting case is not materially affected by this distinction, we prefer to avoid complicating the notation in a way that would be required to formally take on board this point.

[17]As Leitner and Yilmaz (2019) argue, more intense monitoring by the supervisor may lead to a reduction in the informativeness of the bank's internal model. This puts a limit on how much a supervisor can learn, even if the supervisor's monitoring technology could be arbitrarily precise. Parlatore and Philippon (2022) study the design of stress test scenarios as an optimal learning problem when a supervisor receives noisy signals from multiple banks in response to the application of a stress scenario.

Using (3.1) it is easy to show that a stress test S induces a distribution of posterior beliefs

$$\begin{aligned} \nu_i &\equiv \Pr(m_i) = s_i - s_{i-1}, \\ \mu_i &\equiv \Pr(\omega = h | m_i) = \frac{s_{i-1} + s_i}{2}, \end{aligned} \tag{3.2}$$

satisfying Bayes-plausibility, i.e. $\sum_{i=1}^{n} \nu_i \mu_i = \Pr(\omega = h) = 1/2$, with $\sum_{i=1}^{n} \nu_i = 1$.

At date 1 the outcome of the stress test m_i is publicly observed. The speculator then chooses how much private information to acquire about the underlying state ω.[18] Information acquisition generates a signal $z \in \{l, h, \varnothing\}$. The signal is fully informative ($z = \omega$) with probability σ and uninformative ($z = \varnothing$) otherwise. The speculator can choose σ, i.e., how much information to acquire, subject to a cost $\frac{1}{2}\tau\sigma^2$ (with $\tau > 0$) this incurs. We assume throughout that τ is large enough to ensure that the optimal $\sigma \leq 1$.

At $t = 2$ the speculator can trade. His order is denoted $x_I \in \mathbb{R}$. In addition to the speculator, there is a noise trader who either buys or sells with equal probability a number of units that we normalize to one. The noise trader's order is denoted $x_U \in \{-1, 1\}$. The speculator and the noise trader both submit their market order to a market maker, who can observe each order, but not its originator, i.e., orders are anonymous. Formally, we let order flow X be random, taking either the value $X = (x_I, x_U)$ or $X = (x_U, x_I)$ with equal probability. The market maker sets a price that allows him to make zero profits in expectation on any trades he makes out of his inventory. That is, like in a standard Kyle (1985) model, the market maker sets the price equal to the expected value of a share, conditional on the information contained in the order flow.

At date 3 the bank can make an investment of 1 in a risky loan portfolio. For a type $\omega = h$ bank the loan portfolio returns, at date 4, R with probability p and 0 otherwise. Assume $pR > 1$. A type l bank has returns R with probability p_l and zero otherwise. For simplicity, we set $p_l = 0$. Suppose investing generates a (small) private benefit for the banker and the social planner does not care about the banker's private benefit. The bank thus invests whenever it can, regardless of its type.[19] If the bank does not invest, it has a value that we normalize to 0.[20]

In order to capture capital market discipline, assume that the bank has internal funds normalized to 1 and short-term creditors who have a total claim of 1 coming to maturity at date 3. If the creditors do not roll over their loans, the bank has to pay out its internal funds and cannot invest. If the creditors roll over their loans, the bank can use its internal funds for investment in risky lending. Note that it would make no difference if we assumed

[18] Although we do not explicitly endogenize the timing of information acquisition, it is clearly optimal for the speculator to wait until after he observes the stress test result. Doing so allows him to condition the amount of costly information acquisition on the information contained in the stress test.

[19] This is a simple way of modeling excessive risk taking or over-investment. For our purposes it does not matter whether the bank knows its own type as long as the low type banker cannot be prevented contractually from engaging in excessive risk-taking.

[20] An alternative interpretation is that all banks have a brick-and-mortar line of activities which has zero net present value, but only high type banks have access to an additional, positive NPV project. Under this interpretation, even low type banks deserve to operate the brick-and-mortar business, but they should be prevented from expanding into further activities.

instead that the bank does not have any of its own funds and needs to raise 1 from outside providers of capital. The two are equivalent, because the bank could pay off the old creditors using internal funds and then raise funds from fresh creditors for the investment. Assume that the bank can invest if and only if it secures private funding. Hence, we rule out any direct capital injections by the supervisor. We relax this assumption in Section 3.8.

After the stress test and the bank's share price have been observed, the bank can make a take-it-or-leave-it offer asking creditors to roll over their loans at a gross interest rate r.[21] If creditors do not roll over their loans, the bank is forced to pay out 1 to creditors and does not invest.

The timing is summarized in Figure 3.1.

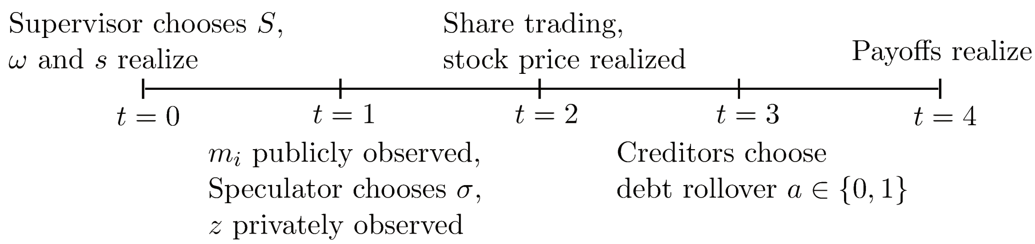


Figure 3.1: Timing

The supervisor cares about the total value created from the bank's activity. We denote by V the bank's expected payoffs at date 4, net of any amount invested at date 3.[22] This payoff depends on both the state of the world ω and an action $a \in \{0, 1\}$, denoting the creditors' rollover decision. We set $a = 1$ if creditors roll over their debt and $a = 0$ if they do not. We can therefore write the supervisor's payoff V_ω^a as follows:

$$\begin{aligned} V_h^1 &= pR - 1, \\ V_l^1 &= -1, \\ V_h^0 &= V_l^0 = 0. \end{aligned} \tag{3.3}$$

Since $V_h^1 > V_h^0$ and $V_l^0 > V_l^1$, it is socially optimal to allow a high type bank to invest, while it is optimal not to provide funds to a low type bank. The same two inequalities imply that $V_h^1 - V_h^0 > 0 > V_l^1 - V_l^0$ which in turn implies $V_h^1 - V_l^1 > V_h^0 - V_l^0$. That is, the action $a = 0$ reduces the variability of bank value compared to $a = 1$. As such, we can think of $a = 0$ more broadly as an action that leads to the reduction in risk, be it downsizing / failing to expand the bank's operations, or a direct intervention to reduce the bank's risk exposure (see also Carletti et al. (2021) or the discussion of *market influence* in Flannery and Bliss (2019)).

[21]We do not introduce any frictions in negotiations with multiple creditors, so we can think of the bank rolling over debt with a single creditor who is subject to a break-even constraint.

[22]Since capital providers break even in expectation, the supervisor's objective is the same as when maximizing the joint payoffs of the bank and its providers of capital.

3.4 Optimal Stress Test without Price Signals

This section develops the benchmark in which stock markets provide no information, so the capital providers can only condition their roll-over decision on the stress test result. Denote by μ the belief that the state is $\omega = h$, conditional on all publicly available information. In this section, public information is limited to the stress test result, while the next section allows for an additional endogenous signal stemming from noisy stock prices. It will be useful to define belief thresholds for which the supervisor prefers for the bank to be able to invest. The supervisor prefers the bank to continue if

$$\mu V_h^1 + (1-\mu) V_l^1 \geq \mu V_h^0 + (1-\mu) V_l^0. \tag{3.4}$$

Defining $\Delta V_h \equiv V_h^1 - V_h^0$ and $\Delta V_l \equiv V_l^0 - V_l^1$, this can be re-written as

$$\mu \geq \frac{\Delta V_l}{\Delta V_l + \Delta V_h}. \tag{3.5}$$

or, using the definitions of ΔV_h and ΔV_l, as

$$\mu \geq \mu^* \equiv \frac{1}{pR}. \tag{3.6}$$

Consider next the bank's funding problem. Good and bad banks will try to roll over debt and invest, but only good banks repay with probability p. Creditors' participation constraint therefore depends on their belief μ as follows:

$$\mu p r \geq 1. \tag{3.7}$$

Since the bank can make a take-it-or-leave-it offer, the interest rate is set at

$$r = \frac{1}{\mu p}. \tag{3.8}$$

For a debt roll-over to be feasible, we also require

$$r \leq R. \tag{3.9}$$

The constraints (3.8) and (3.9) together imply that a debt roll-over is only feasible if creditors are sufficiently optimistic they are lending to a high-type bank:

$$\mu \geq \widehat{\mu} \equiv \frac{1}{pR}. \tag{3.10}$$

Note that $\mu^* = \widehat{\mu}$. Hence, the supervisor's preferred action is also the one implemented via market discipline. We relax this assumption in Section 3.6 and Section 3.7.[23]

[23] In Section 3.6 we allow for debt overhang, so that some bank types are unable to raise funds, although it would be socially optimal to do so, i.e., $\mu^* < \hat{\mu}$. In Section 3.7 we allow for default externalities, which implies that the supervisor would prefer to apply a harsher continuation threshold than that of capital providers ($\mu^* > \hat{\mu}$). In both cases, we show the robustness of our main result, namely an optimal distortion toward leniency.

The supervisor's stress test design problem is potentially quite complex as the partition describing the test can be arbitrarily granular. In the benchmark, the problem simplifies considerably, because the stress test induces a single action a which can take only two values, 0 or 1. From the second line of (3.2) we know that beliefs μ_i, induced by stress test outcome m_i are increasing in i. We can thus collect all partitions that induce action $a = 0$ into one message, and all those that induce $a = 1$ into another message. The stress test can thus be described by a single cut-off s_1 such that for a test result m_1 (i.e., $s < s_1$), we have $\mu_1 < \widehat{\mu}$ and $a = 0$. If the test result is m_2 (i.e., $s \geq s_1$), then $\mu_2 \geq \widehat{\mu}$ and $a = 1$.[24] The supervisor's objective function is thus given by

$$\begin{aligned} v(s_1) &\equiv \frac{1}{2}\Big(F_h(s_1)V_h^0 + F_l(s_1)V_l^0 + \big(1 - F_h(s_1)\big)V_h^1 + \big(1 - F_l(s_1)\big)V_l^1\Big) \\ &= \frac{1}{2}\Big(s_1^2 V_h^0 + (2s_1 - s_1^2)V_l^0 + (1 - s_1^2)V_h^1 + \big(2(1 - s_1) - (1 - s_1^2)\big)V_l^1\Big), \end{aligned} \tag{3.11}$$

and the stress test design problem is

$$\begin{aligned} &\max_{s_1}\ v(s_1) \\ &\text{s.t. } \mu_1 < \widehat{\mu}, \\ &\qquad \mu_2 \geq \widehat{\mu}, \end{aligned} \tag{3.12}$$

where from (3.2), $\mu_1 = \frac{s_1}{2}$ and $\mu_2 = \frac{1+s_1}{2}$. Denote by s_N the solution to (3.12), that is, the optimal cut-off in the no-feedback benchmark.

Lemma 3.1 (Benchmark)**.** *Without information acquisition by the speculator, the optimal stress test is a binary partition with a passing threshold $s_N = \mu^*$.*

Proof. See Appendix B. ■

The optimal stress test simplifies to a *pass/fail* experiment. Failing the test (message m_1), shows that the bank's resilience level s is below $s_N = \widehat{\mu} = \mu^*$, which means that market discipline bites and the bank cannot roll over its debt. The opposite happens if the bank passes the test (message m_2). Note that the optimal cut-off s_N depends on the relative costs of making type-1 and type-2 errors. Define a type-1 error as denying funds to a good bank. The benefit of avoiding this error is ΔV_h.[25] Under a type-2 error a bad bank can roll over debt and invest. Avoiding a type-2 error has a benefit of ΔV_l. If a type-2 error is relatively more costly, then μ^* increases (see (3.5)). This in turn corresponds to an increase in the optimal *pass* threshold s_N: The test becomes more conservative.

In what follows we will use $s_N = \mu^*$ as the reference level of how lenient / conservative a stress test should be and we will be interested in how the possibility of generating informative price signals affects the optimal stress test design compared to this benchmark.

[24]Since μ_1 tends to zero for s_1 close to 0 and μ_2 tends to 1 for s_1 close to 1, it is always possible to choose an s_1 that induces actions that are contingent on the stress test result. It is also optimal to do so because a stress test that never affects a would throw away useful information.

[25]Denying funding to a high type bank leads to a reduction in credit supply to the real economy. See Acharya et al. (2018) for the declined lending by stress-tested banks in the US and Ahmed and Calice (2023) for the UK banks that failed the stress tests.

3.5 Optimal Stress Test with Feedback from Stock Prices

Before fully characterizing the stress test design problem with an active speculator, we need to calculate the profits the speculator can reap from acquiring and trading on private information. For this, we need to determine the fundamental value of the shares, which depends on the underlying state ω and on the bank's access to capital, captured by a. If the bank rolls over its debt, equity value will depend on the interest rate r, which is a function of beliefs μ. We denote the underlying equity value by $E^a_\omega(\mu)$.

If the bank raises funds ($a = 1$) the required repayment is $r = \frac{1}{\mu p}$ (see (3.8)). The expected equity value of a high type bank which can roll over its debt is therefore

$$E^1_h(\mu) = p(R - r) = \frac{1}{\widehat{\mu}} - \frac{1}{\mu}. \tag{3.13}$$

Note that $E^1_h(\mu)$ is increasing in μ, because equity is more valuable when the bank can roll over debt at a lower interest rate. This in turn happens when the creditors have more positive beliefs (higher μ) about the bank's type. The equity value is zero in all other states, either because the bank fails to roll over its debt ($E^0_h = E^0_l = 0$), or because the bank rolls over debt but wastes the funds on a bad investment ($E^1_l = 0$).[26] We can now state the speculator's trading strategy and resulting profits, $\pi(\mu)$, conditional on the belief μ induced by a stress test.

Lemma 3.2 (Trading Strategy and Profits)**.** *The speculator's optimal trading strategy is*

$$x(z) = \begin{cases} 1 & \text{if } z = h \\ 0 & \text{if } z = \varnothing \\ -1 & \text{if } z = l. \end{cases}$$

For a given amount σ of information produced, the speculator's trading profits $\pi(\mu)$ are

$$\pi(\mu) = \begin{cases} \sigma\, \mu(1-\mu) \left(\frac{1}{\widehat{\mu}} - \frac{1}{\mu} \right) & \text{if } \mu \geq \widehat{\mu} \\ 0 & \text{if } \mu < \widehat{\mu}. \end{cases} \tag{3.14}$$

Proof. See Appendix B. ■

[26] The assumption that equity is wiped out following a bank's failure to secure funding is stronger than strictly necessary, but significantly simplifies the exposition. What is crucial for our mechanism to work, is that the traded claim (be it equity or subordinated bonds) becomes less responsive to the true state of the world when the bank fails to raise funds ($a = 0$). Note that this property may hold more generally since $a = 0$ is defined as a risk-reducing action ($V^1_h - V^1_l > V^0_h - V^0_l$). Here we make the simplifying assumption that security payoffs following $a = 0$ do not depend on ω at all. Note that this can be the case in practice, even if equity retains a positive value. For example, if $a = 0$ corresponded to a liquidation or a forced takeover by another bank one could have $E^0_h = E^0_l > 0$. In that case, security payoffs do not depend on ω following $a = 0$ and a speculator cannot benefit from acquiring private information about ω.

Note that the speculator's trading profits are zero if the stress test induces a belief $\mu < \widehat{\mu}$. To see why, consider possible order flows and associated trading profits. If the speculator acquires information and trades on it, order flow can either reveal or hide his direction of trade. When order flow is $(-1,-1)$ or $(1,1)$, the speculator's direction of trade is fully revealed. Since the speculator only buys if $\omega = h$ and only sells when $\omega = l$, these orders fully reveal ω. The market maker then sets a price that fully reflects ω and the speculator therefore cannot make a trading profit. An order flow of $(-1,1)$ or $(1,-1)$ does not reveal the speculator's order. The market maker learns nothing and sets a correspondingly uninformative price. Hence, the roll-over creditors do not learn anything from market prices, and stick to the belief μ induced by the outcome of the stress test. When that outcome is so negative as to deny the bank access to funds ($\mu < \widehat{\mu}$), the bank's equity value drops to zero, regardless of the bank's true type. Since the equity valuations no longer depend on the true state ω, the speculator cannot benefit from learning and trading on knowledge of ω.

When the belief μ induced by the stress test is high enough to allow the bank to roll over its debt ($\mu \geq \widehat{\mu}$), trading profits are hump-shaped. As μ becomes very large, everyone, including the market maker, is confident that the bank is of a high type. This leaves little scope for the speculator to benefit from acquiring private information, which, with a high likelihood, will simply confirm the public belief. The speculator can benefit most from acquiring private information, when doing so confers a significant informational advantage. This is the case when the stress test is least conclusive, i.e., when possible resilience levels are intermediate (μ close to $\frac{1}{2}$). Moreover, as μ drops, the bank will have to roll over debt at less favorable terms, leaving less value for equity holders. This makes it less attractive to speculate on the bank's stock. When $\mu = \widehat{\mu}$, trading profits drop to zero, because rolling over debt is so expensive as to reduce equity value to zero. Overall, trading profits are maximized at $\mu = \frac{1+\widehat{\mu}}{2} \in \left(\frac{1}{2}, 1\right)$.

From the expression for trading profits (3.14) we can take the first-order condition with respect to σ to find the optimal amount of information acquired by the speculator:

$$\sigma(\mu) = \begin{cases} \frac{1}{\tau}\,\mu(1-\mu)\left(\frac{1}{\widehat{\mu}} - \frac{1}{\mu}\right) & \text{if } \mu \geq \widehat{\mu} \\ 0 & \text{if } \mu < \widehat{\mu}. \end{cases} \tag{3.15}$$

In what follows, we assume

$$\tau > \frac{1}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^2, \tag{3.16}$$

which ensures that information acquisition in (3.15) is a non degenerate probability.

We can now express the supervisor's problem in a simplified manner.

Lemma 3.3 (Stress Test Design Problem)**.** *The supervisor solves the following stress test design problem:*

$$\max_S V(S) = v(s_1) + \Sigma(S)$$
$$\text{s.t. } \mu_1 < \widehat{\mu} \qquad (3.17)$$
$$\mu_2 \geq \widehat{\mu},$$

where $v(s_1)$ is defined in (3.11) and

$$\Sigma(S) \equiv \frac{1}{2\tau} \sum_{i=2}^{n} \nu_i \, \mu_i (1-\mu_i)^2 \left(\frac{1}{\widehat{\mu}} - \frac{1}{\mu_i} \right). \qquad (3.18)$$

Proof. See Appendix B. ■

Lemma 3.3 states that the supervisor's objective function can be decomposed into two parts. The first part, $v(s_1)$ consists of the supervisor's expected payoff, when the only source of information is the stress test. This corresponds to the payoff under the benchmark in Section 3.4. The second part, $\Sigma(S)$, consists of the additional payoff from an improved allocation of capital when the stock market provides useful information. Note that the choice of s_1 also enters this second part via its effect on ν_2 and μ_2 (see 3.2)).

We can now state one of our key results.

Proposition 3.1 (Optimal Stress Test)**.** *With information acquisition by the speculator, the optimal stress test is a binary partition with a single passing threshold, s_F, where s_F is the unique s_1 that solves*

$$s_1 = \mu^* - \frac{1}{4\tau} (1-s_1)^2 \left(s_1 - \frac{3\mu^* - 1}{2} \right). \qquad (3.19)$$

The test is lenient ($s_F < \mu^*$) and informative ($s_F > 0$).

Proof. See Appendix B. ■

When market information matters, the stress test is optimally distorted toward leniency ($s_F < s_N = \mu^*$). That is, it awards *pass* grades to some banks that would have failed the test in the benchmark without market feedback.

Why does the supervisor wish to apply a more lenient *pass* threshold? The answer is that, by virtue of being lenient, the stress test generates more *pass* grades. Since a *pass* grade is a precondition for the speculator to acquire information, leniency encourages the production of market information. This information helps creditors in their roll-over decision. In particular, some banks that would be marginal *fail* under the benchmark (banks with resilience levels $s \in [s_F, \widehat{\mu})$) are worth investigating further before denying them access to capital. That way, some of them will be identified as high type banks who were unlucky to have a marginally sub-standard resilience level. When their stock price holds up after the stress test announcement, creditors will be willing to roll over their debt, which is efficient. Of course, some of them will see their stock price drop and be denied funding. While this is also efficient, it does not constitute an improvement compared to the situation where the stress test was not distorted towards leniency: those banks would have failed the benchmark stress test and thus also have been denied funding.

But leniency also has a cost: it allows some banks to roll over debt and invest, although this is socially inefficient. This is the case of a low type bank that marginally passes the lenient test, for which the stock price fails to adjust downwards. This allows such a bank to inefficiently roll over debt, which would have been prevented under a less lenient stress test. The trade-off between the direct information value of the stress test, and its role in encouraging information production by the speculator implies that the optimal cut-off is determined as an internal solution, $s_F \in (0, \mu^*)$.

Note that our model encompasses the corner solutions of a completely uninformative stress test ($s_1 \in \{0, 1\}$), which is just like not conducting a stress test at all. If the test was uninformative and $\hat{\mu} > \frac{1}{2}$, there would never be any information production by the speculator. An informative test then leads to crowding-in of speculator information. In this case it is obviously optimal to have an informative test. When $\hat{\mu} \leq \frac{1}{2}$, and the stress test is uninformative, the speculator would always produce some information. This is, however, not efficient, because costly information gets produced even about banks that have a resilience level that pins down their type with high precision. Instead, it is better to have the speculator produce a lot of information about some banks and little (or no) information about others, rather than an intermediate amount of information about all banks. By rendering the stress test informative, the supervisor can boost the speculator's information production for the banks that pass the test. This is because the bank's debt roll-over is less dilutive when creditors are more optimistic about the bank's type. An informative stress test changes the allocation of information production incentives across bank types, but in general does not neatly map into crowding out (or crowding in) as in some of the literature (see Goldstein and Yang (2017)).

As a thought experiment it is instructive to consider a modification of our model whereby it is easy to induce speculator information production. Suppose τ was very small, but positive such that the speculator would always produce the maximum amount of information ($\sigma = 1$) as long as he anticipates the bank to roll over its debt following an uninformative stock price. An informative test would then clearly lead to crowding out when $\mu^* \leq \frac{1}{2}$: with an uninformative test, the speculator would produce a maximum amount of information on all banks. Following an informative test, the speculator would produce no information for banks that failed the test, and (still) produce the maximum amount for those that pass it. That is, there is an overall reduction in information production. An uninformative test, however, would not be optimal even in this case. That is because the speculator's information gets impounded into the price with noise. Even if an uninformative test were to lead to maximum information production about all banks, the capital providers will roll over debt for many undeserving banks: When the stock price is uninformative and the bank has a low resilience level, which remains unknown since the stress test is uninformative, a bank gets funded, but should not. It is therefore optimal to reveal very low resilience levels, i.e., make the test informative and lenient, even if it is very easy to induce information production by speculators.

Another result from Proposition 3.1 is that the optimal test retains its *pass/fail* nature. This no longer follows immediately from the binary nature of the capital providers' decision ($a \in \{0, 1\}$), because the stress test also affects information production which is a continuous choice variable. The mechanism described before points to a robust reason why the test must be coarse on a sub-interval around μ^*: Leniency aims to encourage information production for some banks with resilience levels below μ^*. But for the capital market to fund such banks, they must be lumped into the same category as banks with a resilience level $s > \widehat{\mu} = \mu^*$. In other words, for a *pass* category to contain sub-standard banks ($s < \widehat{\mu}$) *and* be credible, it must also contain a sufficient number of above-standard banks.

Note also that it is sub-optimal to introduce further sub-categories of a *pass* test. With a single *pass* category and a lenient cut-off $s_F < \mu^*$, the induced belief μ_2^F is below that which generates maximum information production ($\mu_2^F < \tilde{\mu} = \frac{\mu^*+1}{2}$). Suppose now that s_F was kept unchanged, but a second *pass* grade with a cut-off $s_2 \in (\mu^*, 1)$ was introduced. There would thus be two possible *pass* grades m_2 and m_3 with corresponding induced beliefs which we denote by μ'_2 and μ'_3, respectively. By construction, we have $\mu'_2 < \mu_2^F < \tilde{\mu}$, and $\mu'_3 > \tilde{\mu}$. As a result, information production following either *pass* grade m_2 or m_3 would drop. Intuitively, after a moderate *pass* (m_2) the interest rate of roll-over debt increases, which dulls information production incentives. After a strong *pass* m_3 there is little uncertainty over the bank's true type. This reduces the speculator's potential advantage from becoming privately informed, decreasing information acquisition. Hence, splitting a single *pass* grade in two will reduce information production and is therefore sub-optimal.

Going back to the broader question of the role of market discipline in supporting banking supervision, our analysis reveals the following insight. Markets, via the information they provide, can help reduce type-2 errors. That is, if a bad bank slips through the supervisor's net (by passing a stress test), it will be subject to market scrutiny and possibly "disciplined" by being denied funding. This provides a reason to allow more banks to slip through the net, i.e., to be lenient. On the other hand, markets are not good at reducing type-1 errors. Banks that are caught in the supervisor's net (by failing the stress test), will not be subject to market scrutiny. Any mistake made in the supervisory process will thus not be corrected. Given the asymmetry in the way that (indirect) market discipline operates, a lenient stress test design is optimal. This contrasts with the literature, which has mainly associated leniency with a misalignment of the supervisor's objective and the objective of the recipients of the stress test result. For example, in Bouvard et al. (2015), Williams (2017), Goldstein and Leitner (2018), Parlasca (2024) and Shapiro and Zeng (2024), the supervisor may wish to hide information from markets to avoid a bank run.

In addition to identifying a new channel via which stress tests matter, our theory also provides new policy implications. The improvement in the information environment from stress test leniency only accrues to banks whose shares are publicly traded. Our theory therefore implies that publicly listed banks should be subject to more lenient stress tests than otherwise equivalent privately held banks. To the extent that regulation does not explicitly distinguish between publicly listed and privately held banks in the stress test design, this would lead to sub-optimal supervision. Ignoring the impact of stress test design on the quality of price signals, would lead to stress tests that are too adverse, reducing the information in the price signals available to creditors in banks who do poorly in the test.

The following proposition describes how the optimal degree of leniency is affected by model parameters.

Proposition 3.2 (Degree of Leniency)**.** *At the optimum, the stress test is more lenient, i.e. s_F decreases, when:*

- *the high type bank's expected returns are higher (p or R increase),*
- *information acquisition is less expensive (τ decreases).*

Proof. See Appendix B. ■

In general, the optimal *pass* threshold s_F is directly affected by changes in μ^* and indirectly by the optimal extent of distorting s_1 away from μ^*. In developing the intuition for the comparative statics, we will make use of this distinction.

Consider a reduction in the cost of information acquisition, τ. Note that τ has no direct effect on μ^*. It does, however, have an effect on s_F: When financial markets can cheaply acquire information about the bank's fundamentals, private information becomes more precise, and the supervisor obtains more benefit from distorting the stress test towards more leniency.

By contrast, changes in p and R have a direct effect on the relative costs of type-1 and type-2 errors and thereby on μ^*. In addition, they have an indirect effect on the optimal extent of distorting s_1 away from μ^*. As p and R increase, the cost of a type-1 error increases, as denying funds to a good bank becomes more costly. A bank is therefore allowed to continue for a lower resilience level, i.e., μ^* decreases and so does s_F. In addition, there is an indirect effect. Higher financial returns of the good bank increase the information sensitivity of the equity claim and thus the speculator's incentives to acquire information. As a consequence, market information becomes more precise, increasing the benefit of distorting the stress test towards more leniency. The direct effect of an increase in p and R on μ^* and the indirect effect thus work in the same direction towards more leniency.

3.5.1 Social versus Private Value of Information

It is instructive to ask how much information the supervisor would acquire if she had access to the speculator's information technology and could make the collected information public so as to allow market discipline to be based on it. We do not consider this possibility throughout the paper, because we want to capture the notion, advanced by regulators, that financial markets can generate information that supervisors cannot. However, since the speculator's incentives are not aligned with the supervisor's, this raises the question of what distortion the misalignment may create. In particular, one may ask how the belief μ affects the supervisor's value of producing additional information. This depends on whether a bank can roll over its debt when $z = \varnothing$, which in turn depends on whether μ is above or below μ^*. When $\mu < \mu^*$, and $z = \varnothing$, creditors do not roll over debt ($a = 0$). Hence, the supervisor's expected payoff is

$$E(V) = \mu(\sigma V_h^1 + (1-\sigma)V_h^0) + (1-\mu)V_l^0 - \frac{\tau}{2}\sigma^2. \tag{3.20}$$

When $\mu \geq \mu^*$, the expected payoff is instead

$$E(V) = \mu V_h^1 + (1-\mu)(\sigma V_l^0 + (1-\sigma)V_l^1 - \frac{\tau}{2}\sigma^2. \tag{3.21}$$

Taking the first-order condition in each of the two regions gives us the supervisor's optimal amount of information production, as a function of belief μ, denoted by $\sigma_S(\mu)$:

$$\sigma_S(\mu) = \begin{cases} \frac{1}{\tau}(1-\mu)\Delta V_l & \text{if } \mu \geq \mu^* \\ \frac{1}{\tau}\mu \Delta V_h & \text{if } \mu < \mu^*. \end{cases} \tag{3.22}$$

In the above expression, $\sigma_S(\mu)$ is continuous and maximized at $\mu = \mu^*$. This is intuitive. At μ^*, the expected cost of making a type-1 or a type-2 error is the same. That is precisely when additional information is most valuable. In sharp contrast, the speculator's incentives to acquire information are zero at the belief $\mu = \mu^*$. This is because at the corner $\mu = \mu^*$ the bank needs to raise funds at the least favorable conditions: creditors only roll over their debt at such a high interest rate that nothing is left for equity holders. The speculator therefore cannot make a trading profit, even if the bank can roll over its debt. There is therefore a wedge between the social and the private value of information. Viewing the problem from this angle gives us a further understanding of the main result on the optimal leniency of the stress test: The social value of information is highest for banks with levels of resilience around the threshold μ^*. However, these are precisely the banks for which the speculator's private value of information is particularly low. The supervisor therefore induces information production about these banks by lumping them into the same stress test result as those banks for which the speculator has a strong information production incentive.

3.5.2 Empirical Implications

While our analysis is largely normative in nature, our theory has a number of empirical implications. In particular, our theory predicts that the amount of informed trading in a bank's shares depends on how a bank performs in the stress test. We would expect there to be less informed trading following the announcement that a bank failed a stress test, compared to when it passes it. Some papers have shown that there is abnormal trading volume after the announcement of stress test results (see Flannery et al. (2017) among others). This is indicative of informed trading activity, although not proof of it. More direct measures of informed trading have been developed in the market micro-structure literature, for example, bid-ask spreads, the probability of informed trading (PIN) (see Easley et al. (1997)) or Multimarket Information Asymmetry (MIA) (see Johnson and So (2018)). It would thus be possible to estimate such microstucture-based measures for banks after the announcement of stress tests and check whether they are lower for banks that did poorly in the test compared to those that did well. There is no empirical research to date that conducts such an analysis.

A few papers have looked into abnormal returns following stress test announcements. If the stress test conveys information, one would expect prices to increase if the stress test result is better than expected and decrease if the opposite is true. Petrella and Resti (2013) and Morgan et al. (2014) provide evidence supporting this hypothesis using stock price reactions to the first stress tests after the 2008-09 crisis. The findings remain similar for the more recent, regular stress tests implemented by both the ECB and the US Federal Bank (See Ahnert et al. (2020), for example). Since the event date (announcement of stress test results) is known in advance, one would, however, expect there to be no abnormal returns on average after the announcement. As pointed out by Flannery et al. (2017), this is a direct implication of market efficiency. It should thus be true in our model, but is not specific to it. As Flannery et al. (2017) argue, a more relevant metric is |CAR|, the absolute value of cumulative abnormal returns, which should be higher after the announcement date for banks that are subject to a stress test compared to banks that are not. Flannery et al. (2017) do not condition specifically on the stress test outcome and find that |CAR| is indeed higher for the sample of tested banks compared to non-tested banks.[27] We can compute a comparable metric implied by our model, by calculating the ex-ante expected |CAR| as a function of the stress test design. One can think of the banks that were not stress tested as the limit case in our model where the stress test is uninformative. Figure 3.2 depicts the difference between |CAR| for tested and non-tested institutions, $\chi(\tau, \mu^*)$, using the optimal design for tested institutions and an uninformative design for non-tested institutions (see Appendix B for detailed computations). This speaks directly to the additional price changes induced by information production following the stress test announcement, which is positive for reasonable parameter values, as found by Flannery et al. (2017).

3.6 Recapitalization under Debt Overhang ($\mu^* < \widehat{\mu}$)

So far, we have analysed the case where the supervisor's preferred course of action coincided with that implemented by capital providers ($\mu^* = \widehat{\mu}$). In practice, supervisors often worry about disclosing negative news for fear of tightening financial constraints for sound banks, including the extreme case of provoking a bank run. Similarly, banks may be unable to access private funding due to a debt overhang problem (see Philippon and Schnabl (2013)). In our framework, this corresponds to the case $\mu^* < \widehat{\mu}$. Banks with observed resilience levels $s \in [\mu^*, \widehat{\mu})$ will not be able to fund their activities, although the supervisor would like them to.

[27] Note that from a theoretical perspective it is not entirely clear whether |CAR| should increase for tested banks. The test has two implications for |CAR|. First, the information contained in the announcement has a direct effect on prices. Second, information production and trade depend on the public information revealed (or not) through the test. Hence, non-tested banks could have a higher |CAR| if markets produced significantly more private information about them.

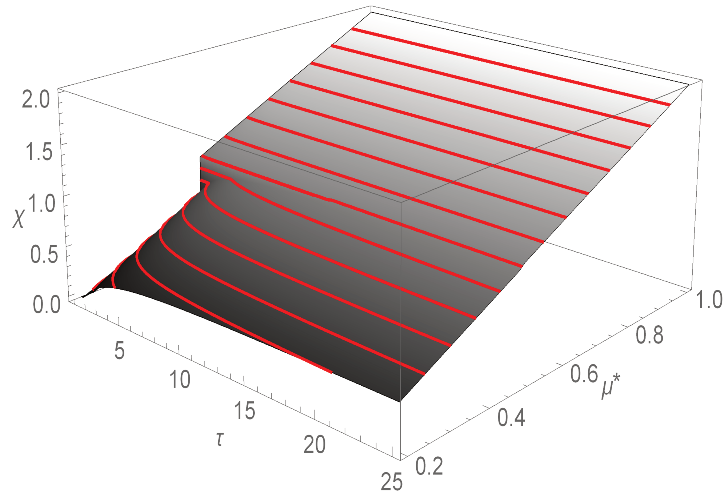


Figure 3.2: Difference between |CAR| for Tested and Non-Tested Institutions: $\chi(\tau, \mu^*)$

One way, among others, to illustrate this case is by modifying our micro-foundation to allow for debt overhang, which drives a wedge between the efficient investment choice and that actually implemented by the capital market.[28] To adapt our model, suppose banks have an existing $t = 1$ level of senior debt $D < R$ outstanding, which cannot be renegotiated. D could capture the amount of insured deposits, for example. The level of pre-existing debt does not affect the desirability of investing, so μ^* remains unchanged. In order to abstract away from complications arising with multiple classes of pre-existing creditors, we no longer interpret the funding stage as a roll-over of existing debt. Suppose instead that the bank needs to raise 1 from new providers of capital who are junior to depositors.[29] The participation constraint of capital providers, such as uninsured creditors, changes to

$$\mu p(r - D) \geq 1. \tag{3.23}$$

Together with the feasibility constraint $r \leq R$ this pins down a new belief threshold given by

$$\widehat{\mu} = \frac{1}{p(R - D)} > \mu^*. \tag{3.24}$$

[28] An alternative would be to micro-found the belief threshold based on a coordination problem among multiple uninsured depositors who may run on the bank, as in Bouvard et al. (2015). Applying a global game refinement generates a unique belief cut-off, much like what we have. One shortcoming of the bank-run approach as typically implemented is that uninsured depositors' claims are assumed to be fixed. Hence, equityholders' payoffs depend only on whether a run occurs, but not on the beliefs when a run does not occur. In our setting creditors demand a higher interest rate when they extend credit at more pessimistic beliefs.

[29] With debt overhang, a junior creditor who needs to roll over debt faces a different outside option than an outside provider of capital, since the former depends on the bank's liquidation value. In order to keep the treatment simple, we do not introduce this complication here. In general, if the bank has a large liquidation value, it may be the case that markets are willing to fund it, even though the planner would prefer a liquidation, i.e., $\mu^* > \widehat{\mu}$. We provide an analysis of this case in Section 3.7.

In order not to burden the notation, we continue to refer to the threshold defined in (3.24) simply as $\widehat{\mu}$. Taking the special case $D = 0$ gets us back to the definition of $\widehat{\mu}$ from Section 3.4. We begin by clarifying how the benchmark is affected by debt overhang.

Lemma 3.4. *Without information acquisition by the speculator, the optimal stress test is a binary partition with passing threshold* $s_N^D \equiv \max\{2\widehat{\mu} - 1, \mu^*\}$.

Proof. See Appendix B. ∎

Lemma 3.4 shows that the stress test has a cut-off at μ^*, just like before. Since $\mu^* < \widehat{\mu}$ this test is lenient from the capital provider's point of view: The supervisor labels banks as a *pass* when she thinks they should have access to capital, not when capital markets would like them to. From the perspective of capital providers, the supervisor is too generous with *pass* grades. Note that a cut-off at $s_N^D = \mu^*$ only implements the supervisor's preferred outcome, if it induces a sufficiently positive belief such that the capital market is willing to provide funds. When μ^* is too low, a *pass* grade may no longer allow the bank to raise funds. In this case, the supervisor chooses a threshold s_N^D at the lowest level that still ensures that a *pass* grade allows the corresponding banks to access capital markets.[30]

Proposition 3.3. *With information acquisition by the speculator, the optimal stress test is a binary partition with passing threshold*

$$s_F^D \equiv \max\{2\widehat{\mu} - 1, \widetilde{s}\} \tag{3.25}$$

where $\widetilde{s}$ *is the unique* s_1 *that solves*

$$s_1 = \mu^* - \frac{\mu^*}{\widehat{\mu}} \frac{1}{4\tau} (1 - s_1)^2 \left(s_1 - \frac{3\widehat{\mu} - 1}{2} \right). \tag{3.26}$$

The stress test is:

- *neutral, i.e.* $s_F^D = s_N^D$, *when* $\mu^* \in (0, \mu^\circ]$ *or* $\mu^* = \frac{3\widehat{\mu}-1}{2}$,
- *conservative, i.e.* $s_F^D > s_N^D$, *when* $\mu^* \in \left(\mu^\circ, \frac{3\widehat{\mu}-1}{2}\right)$,
- *lenient, i.e.* $s_F^D < s_N^D$, *when* $\mu^* \in \left(\frac{3\widehat{\mu}-1}{2}, \widehat{\mu}\right)$.

where

$$\mu^\circ \equiv \frac{2\widehat{\mu} - 1}{1 + \frac{1}{2\tau} \frac{(1-\widehat{\mu})^3}{\widehat{\mu}}}. \tag{3.27}$$

Proof. See Appendix B. ∎

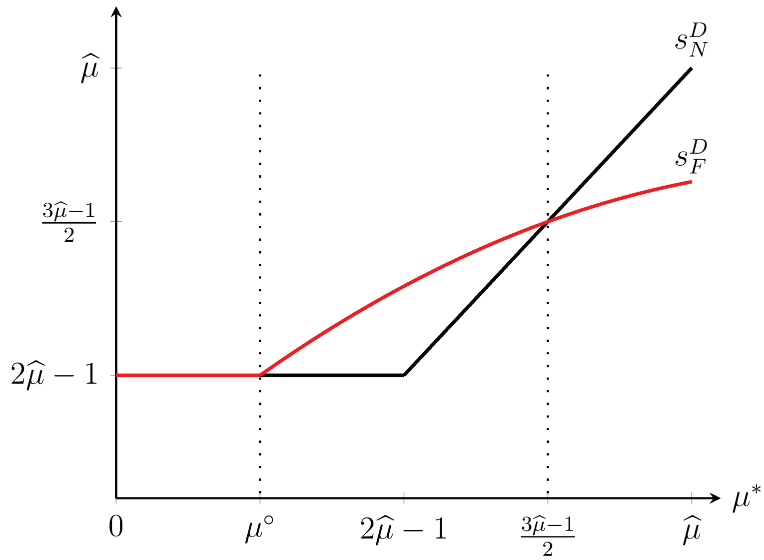


Figure 3.3: Passing Threshold of the Optimal Stress when $\mu^* < \widehat{\mu}$.
Benchmark s_N^D (in black) and Model with Feedback from Stock Prices s_F^D (in red).

Figure 3.3 summarizes Proposition 3.3 by depicting the cutoff of the optimal stress test with feedback from stock prices and in the benchmark without such feedback. The key takeaways from Proposition 3.3 are as follows. When μ^* is close to $\widehat{\mu}$, the optimal stress test is distorted towards leniency ($s_F^D < \mu^*$). This is just an extension of the result in Proposition 3.1, where we have already shown the optimality of leniency. As μ^* drops, distorting towards leniency becomes increasingly unattractive as the cut-off s_1 moves further and further below the cut-off $\widehat{\mu}$, making the recapitalization ever more expensive. This in turn reduces the trading profits a speculator can make. In order to provide sufficient incentives for the speculator to acquire information, the recapitalization must not be too expensive, requiring a high enough threshold s_F^D to make the *pass* grade convey sufficiently good news. Hence, as μ^* drops, the supervisor lowers s_F^D less than one for one and eventually the test becomes conservative ($s_N^D < s_F^D$). In other words, conservatism serves to make the financing terms at the recapitalization stage adequately attractive to preserve incentives for the speculator to acquire information about stock values. This result is reminiscent of Orlov et al. (2023) who show that the optimal stress test admits false-negatives. In their paper, banks that fail the test must recapitalize. By making the test more conservative, failing it conveys less negative information, allowing such banks to recapitalize at more favorable terms. In our case, a more conservative test conveys more positive information for a bank that passes it, allowing the latter to raise funds at a lower cost. One key difference is that in Orlov et al. (2023) a more dilutive recapitalization constitutes an inefficient allocation

[30] This observation is akin to the finding in Williams (2017), or Bouvard et al. (2015). In their model, the supervisor would like all banks to be able to access capital markets, corresponding to the case $\mu^* = 0$. When $\widehat{\mu} \leq \frac{1}{2}$, then $s_N^D = \mu^* = 0$, is optimal as a completely uninformative test allows all banks to access capital markets. An uninformative test is like no information disclosure in Bouvard et al. (2015). When $\widehat{\mu} > \frac{1}{2}$, an uninformative test would result in no bank being able to access capital markets. In that case, it is better to raise the threshold just enough to allow the banks that receive a *pass* grade to get funding. That threshold is given by s_1 such that $\widehat{\mu} = \frac{s_1+1}{2}$.

of capital and thus reduces welfare. In contrast, in our model, any dilution at the capital-raising stage redistributes wealth from high to low type banks, but is not inefficient per se. Nevertheless, dilution matters because it affects the quality of the price signal.

Finally, as μ^* drops even further, s_F^D will reach the lower bound given by $2\widehat{\mu} - 1$. At this point, passing the test becomes such a weak signal that the recapitalization will be so expensive as to undermine the speculator's information acquisition incentives. Although the market signal disappears at this point, the stress test still generates market discipline by directly providing information to capital providers. This improves resource allocation just like in the benchmark of Lemma 3.4.

The following comparative statics hold.

Proposition 3.4. *At the optimum, the passing threshold s_F^D decreases, when:*

- *the high type bank's financial returns are higher (p and R increase),*
- *the level of outstanding debt is lower (D decreases).*

When the stress test is lenient (conservative) the optimal passing threshold s_F^D increases (decreases) in τ. When the stress test is neutral the passing threshold s_F^D is independent of τ.

Proof. See Appendix B. ■

First, consider the case where a *pass* grade conveys sufficiently positive news to render the debt rollover constraint non-binding ($s_F^D > 2\widehat{\mu} - 1$). An increase in p and R reduces the optimal pass threshold as in the model without debt overhang (see Proposition 1): higher financial returns for the good bank increase the cost of a type-1 error (direct effect) and increase the speculator's incentives to acquire information (indirect effect). By contrast, an increase in the debt exposure of the bank, D, has no direct effect on μ^* but reduces the optimal pass threshold s_F^D through its indirect effect on the speculator's incentives to acquire information. A higher level of outstanding debt destroys value to equity holders only in the state where the bank is able to repay such debt ($\omega = h$). As a consequence, the equity claim becomes less information sensitive, depressing the speculator's profits. This reduces information acquisition and the information content of the price signal, which weakens the supervisor's motive to distort the stress test towards leniency. Interestingly, a reduction in the cost of information acquisition, τ, affects the optimal *pass* threshold differently, depending on whether the test is lenient or conservative. This happens because a reduction in τ makes market information more precise, and thus amplifies the supervisor's motives toward leniency/conservatism.

Finally, consider the case where μ^* and therefore s_F^D are so low that any further reduction in the cut-off would undermine the possibility to recapitalize the bank. In this case, the *pass* threshold is pinned down by the belief threshold that induces funding provision, $\widehat{\mu}$. Nevertheless, the qualitative effects of changes in p, R, and D remain the same: higher financial returns of the good bank and lower debt exposure relax the funding constraint reducing $\widehat{\mu}$ and hence s_F^D. Since the speculator makes no profits and does not acquire information, the *pass* threshold is independent of the cost of information acquisition, τ.

3.7 Default Externalities ($\mu^* > \widehat{\mu}$)

So far, we considered the case where the supervisor would finance more banks than the capital providers are inclined to support ($\mu^* \leq \widehat{\mu}$). In principle, it could be the case that the social value of liquidating a bank is higher than internalized by the bank's capital providers and the supervisor is more inclined than the market to cease the bank's operations. This might be the case, for example, if a bank's default generates negative externalities, either on other banks or the real economy. This corresponds to the case $\mu^* > \widehat{\mu}$: banks with resilience levels $s \in [\widehat{\mu}, \mu^*)$ would be able to raise funds, although the supervisor would like to prevent that.

To capture this situation, we modify our baseline micro-foundation in Section 3.4 by introducing a social cost c of defaulting. A default occurs when the bank raises funds and invests, but then generates a zero cash flow from the investment. Capital providers do not internalize the default externality so that $\widehat{\mu}$ remains unchanged. For a given belief μ the supervisor prefers for a bank to invest ($a = 1$) if

$$\mu(p(R-1) + (1-p)(-1-c)) + (1-\mu)(-1-c) \geq 0. \tag{3.28}$$

Hence, the supervisor prefers $a = 1$ for beliefs

$$\mu \geq \mu^* \equiv \frac{1+c}{p(R+c)} > \widehat{\mu}. \tag{3.29}$$

We start by providing the optimal stress test in the absence of stock market signals.

Lemma 3.5. *Without information acquisition by the speculator, the optimal stress test is a binary partition with passing threshold* $s_N^L \equiv \min\{2\widehat{\mu}, \mu^*\}$.

Proof. See Appendix B. ∎

The test is a *pass/fail* experiment just like in previous cases. Whenever the market implements the supervisor's preferred course of action, the optimal cut-off is set at $s_N^L = \mu^*$ to minimize type-1 and type-2 errors. When μ^* is too high, even a *fail* grade conveys sufficiently positive news for capital providers to extend credit to the bank. Therefore, the pass threshold is optimally adjusted downward to $s_N^L = 2\widehat{\mu}$, ensuring the credibility of the *fail* grade and dissuading the market from extending credit.

The following proposition describes the optimal stress test design with feedback from stock prices.

Proposition 3.5. *With information acquisition by the speculator, the optimal stress test is as follows:*

- *if* $\mu^* \in (\widehat{\mu}, \underline{\mu})$*, where* $\underline{\mu}$ *is defined in (B.21), it is a binary partition with passing threshold* $s_F^L = \widetilde{s}$*, where* $\widetilde{s}$ *is the unique* s_1 *that solves (3.26);*

- *if* $\mu^* \in [\underline{\mu}, 1)$ *and* $\widehat{\mu} < 2/5$ *it contains two coarse buckets followed by granular grades for resilience levels above the buckets, where the threshold that separates the two buckets is*
$$s_{1,F}^L \equiv \min\left\{2\widehat{\mu}, s^\dagger\right\}, \tag{3.30}$$
where $s^\dagger$ *is the unique* s_1 *that solves (B.23), while the threshold that separates the upper bucket from the granular grades is*
$$s_{2,F}^L \equiv 1 - \frac{1}{2}\left(s_{1,F}^L - \widehat{\mu}\right); \tag{3.31}$$

- *if* $\mu^* \in [\underline{\mu}, 1)$ *and* $\widehat{\mu} \geq 2/5$*, there exists a threshold* $\overline{\mu}$ *(defined in (B.25)) such that if* $\mu^* < \overline{\mu}$ *the optimal stress test is as described above and if* $\mu^* \geq \overline{\mu}$ *the test contains only one coarse bucket followed by granular grades, where the threshold that separates the coarse bucket from the granular grades is*
$$s_{F,g}^L \equiv \min\left\{2\widehat{\mu}, s^\ddagger\right\}, \tag{3.32}$$
where $s^\ddagger$ *is the unique* s_1 *that solves (B.26).*

Proof. See Appendix B. ∎

The stress test is a lenient binary partition when the supervisor's and the capital providers' preferences about debt roll-over decisions are sufficiently aligned, i.e. when μ^* is close to $\widehat{\mu}$. The advantages of leniency are once again rooted in the additional information generated by the speculator, which in turn guides funding decisions more effectively. As μ^* rises, the optimal passing threshold follows suit to minimize the test's statistical errors, but remains lenient to incentivize information production by the speculator. As the passing threshold rises, the *pass* grade progressively conveys better news to speculators by encompassing banks of increasingly superior quality. Consequently, residual uncertainty following a *pass* diminishes, reaching a point where speculators' incentives to acquire information begin to decline. This happens when s_1 surpasses $\widehat{\mu}$, as the *pass* grade induces beliefs $\mu_2 = \frac{s_1+1}{2}$ while information acquisition is maximized at $\mu = \frac{\widehat{\mu}+1}{2}$. At this point, it is optimal to exclude highly sound banks from the *pass* grade by setting s_2 below 1. This maintains residual uncertainty following a *pass* without the need to lower s_1 far below the benchmark s_F^L. By doing so, the supervisor improves both direct and indirect market discipline by maintaining the speculator's incentives to gather information while optimizing resource allocation in the absence of market signals. Surprisingly, resilience levels above s_2 are fully disclosed. Why does the supervisor not lump them into a unique, coarse grade m_3? Refining the coarse grade m_3 by splitting it into two grades $m_3^{'}$ and $m_3^{''}$ increases information production following $m_3^{'}$ and reduces it following $m_3^{''}$, as the latter message constitutes more conclusive news. Since information production has value only when a poorly capitalized bank passes the test, an event that is less and less likely for higher resilience levels, refining the grade increases the ex-ante level of information acquisition.

As μ^* continues to increase, the passing threshold will eventually reach its upper bound defined by $2\widehat{\mu}$. At this point, the *fail* grade represents excessively positive news, to the extent that any further increase in the passing threshold would prompt capital providers to finance the bank after a *fail*. To ensure direct market discipline and efficient resource allocation, it is optimal to maintain the passing threshold at a constant level. Figure 3.4 depicts the cutoffs of the optimal stress test with feedback from stock prices and in the benchmark.

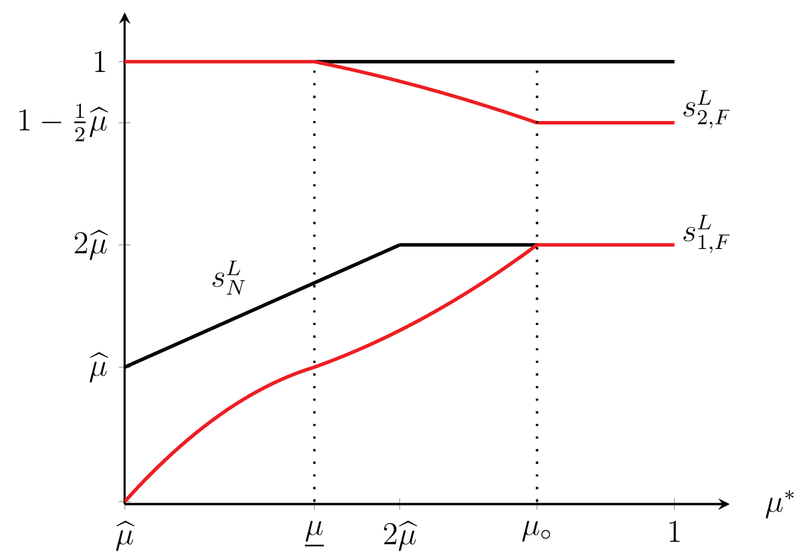


Figure 3.4: Thresholds of the Optimal Stress when $\mu^* > \widehat{\mu}$, and $\widehat{\mu} < 2/5$. Benchmark $(s^L_N, 1)$ (in black), and Model with Feedback from Stock Prices $(s^L_{1,F}, s^L_{2,F})$ (in red).

3.8 Supervisor Private Information

So far, we have assumed that the supervisor does not directly intervene in banks, be it to fund those unable to roll over their debt, or to restrict the activities of banks that have access to private funding. Whether banks are able to invest in a risky project depends entirely on their ability to roll over their debt, i.e., on market discipline. Going back to the baseline model without debt overhang ($\mu^* = \widehat{\mu}$), the lack of direct intervention by the supervisor can be justified in several ways. First, when $\mu^* = \widehat{\mu}$, creditors implement the supervisor's preferred action. If they have the same information as the supervisor, there is no need for the latter to intervene. Second, even if the supervisor had private information, she might not intervene if there is a sufficiently high cost of doing so. The appeal to market discipline as a pillar in the Basel framework is an implicit acknowledgement of the practical relevance of such costs. For example, a supervisor may find it costly to inject public funds into banks that are unable to roll over their debt, because of tax distortions (e.g., White and Yorulmazer (2014), Faria-e Castro et al. (2017), or Shapiro and Zeng (2024)). Moreover, the supervisor may suffer from forbearance or be reluctant to restrict the activities of banks that can access credit markets (Martynova et al. (2022)).

In this section we extend our analysis to the case where the supervisor has private information and can intervene at a cost, including the corner of a zero cost. Suppose parameter values are such that $\mu^* = \widehat{\mu}$, i.e., we are back to the base-line model.[31] The timing is as follows. Like before, assume that the supervisor designs a stress test at date 0. Subsequently, the stress test result is publicly observed. The supervisor simultaneously and privately observes the bank's true resilience level s. Since the supervisor's information is independent of the stress test design, the latter is just a disclosure strategy to which the supervisor commits.[32] At date 1, after having observed the stress test result, the speculator decides whether to become informed and trade. After a trade has taken place and prices are observed, the supervisor can take an action $a_s \in \{\varnothing, 0, 1\}$. If the supervisor does nothing ($a_s = \varnothing$), the bank approaches creditors and tries to roll over its debt. Assume that creditors observe a_s. When the supervisor chooses not to intervene, creditors update their beliefs and funding conditions are set accordingly. If the supervisor intervenes ($a_s \in \{0, 1\}$), she incurs a cost $\delta \geq 0$ and can either provide public funding to the bank ($a_s = 1$),[33] or shut it down ($a_s = 0$). As before, the intervention $a_s = 0$ should be thought of more broadly as an action that reduces the bank's risk exposure, but for simplicity we just refer to it as shutting down of the bank. If the supervisor provides public funding ($a_s = 1$), we assume that she does so at an interest rate that allows her to break even, conditional on her private signal s.[34] To keep the treatment simpler, we restrict the stress test design problem to a *pass/fail* test, i.e., there is a single cut-off $s_1 \in [0, 1]$.

The bank's expected value depends on the cut-off s_1 via the beliefs and actions that the stress test induces. The speculator needs to form beliefs about the likelihood and direction of the supervisor's future direct interventions. In particular, it may happen that under a lenient test design, the supervisor chooses to unwind a bank, even though it passed a test. Conversely, under a conservative test, the supervisor may fund a bank that failed the test. For which resilience levels s a supervisor intervenes depends on the cost δ of such an action. We can thus distinguish the following regions.

Region a(i): $s_1 < 2\mu^* - 1$. The cut-off is so low, i.e., the stress test so lenient, that in the absence of a price signal, creditors are not willing to roll over debt, even if the bank

[31] Studying the case $\mu^* = \widehat{\mu}$ allows us to focus on what we view as the core focus of this paper, namely the information spillovers from stress tests to market discipline. If we had $\mu^* < \widehat{\mu}$, the supervisor might want to intervene even without private information, simply because market discipline is too tough. This case (without stock price feedback) has been studied by Faria-e Castro et al. (2017) among others. For analytical clarity and tractability, we focus only on the first mechanism.

[32] This distinguishes our approach from Bouvard et al. (2015) and Parlasca (2024) who analyze the signalling game where the supervisor has private information *before* deciding a disclosure strategy or stress test design.

[33] One may associate this to the practice of the Emergency Liquidity Assistance (ELA) that the ECB may offer to a financially distressed bank, such as in the case of Greek banks in 2014.

[34] In principle the supervisor could provide funding at a subsidized rate. This may be undesirable if the cost of public funds is high. Moreover, a supervisor may be reluctant, for political reasons, to undercut private providers of capital for the benefit of leaving a rent to the bank. Finally, the assumption enables a clearer comparison to the main model since the only margin is *when* a bank can raise funds, but not whether the funding is subsidized. The supervisor cannot provide funds at a higher rate than the competitive one, since her willingness to do so signals to the market that $s \geq \widehat{\mu}$, in which case creditors will choose to roll over debt at competitive terms.

passes the test. This happens when $\mu_2 = \frac{s_1+1}{2} < \mu^*$, i.e., when $s_1 < 2\mu^* - 1$. This region is empty if $\mu^* < \frac{1}{2}$. The bank can roll over its debt following a positive signal from the stock price. Following an uninformative stock price, creditors would not roll over their claims, but the supervisor would intervene and provide funding if

$$sV_h^1 + (1-s)V_l^1 - \delta \geq sV_h^0 + (1-s)V_l^0,$$

i.e., when

$$s \geq \bar{s} \equiv \frac{\Delta V_l + \delta}{\Delta V_l + \Delta V_h} = \mu^*(1+\delta). \tag{3.33}$$

Region a(ii): $s_1 \in [2\mu^* - 1, \mu^*(1-\delta))$. In this region, the stress test is sufficiently informative to allow a bank to roll over debt (in the absence of a price signal) if it passes the test, but not if it fails it, i.e., $\mu_1 < \mu^* \leq \mu_2$. Since $s_1 < \mu^*$, the supervisor never wants to fund a bank that failed the stress test. However, the supervisor may wish to intervene by shutting down the bank, even though it passed the test. Following message m_2, and an uninformative stock price, the creditors would roll over their debt, while, the supervisor, knowing s, would prefer to intervene by shutting down the bank ($a_s = 0$), if

$$sV_h^0 + (1-s)V_l^0 - \delta > sV_h^1 + (1-s)V_l^1.$$

The supervisor would thus intervene and shut down the bank when

$$s < \underline{s} \equiv \frac{\Delta V_l - \delta}{\Delta V_l + \Delta V_h} = \mu^*(1-\delta). \tag{3.34}$$

Region p: $s_1 \in [\mu^*(1-\delta), \mu^*(1+\delta)]$. In this region, the supervisor is always passive since the intervention cost is higher than the expected benefit of intervening.

Region b(i): $s_1 \in (\mu^*(1+\delta), 2\mu^*)$. The test is conservative. Absent a price signal, the creditors are willing to roll over debt for a bank that passed the test, but not for a bank that failed it, i.e., $\mu_1 < \mu^* \leq \mu_2$. The supervisor is willing to intervene and fund a bank that has failed the test if the resilience level is $s > \bar{s}$. The supervisor never shuts down a bank that passed the test since $s_1 > \mu^*$.

Region b(ii): $s_1 \in [2\mu^*, 1]$. In this region s_1 is so high, i.e., the test so conservative, that even a bank that failed the test would be able to roll over its debt. This region is empty if $\mu^* > \frac{1}{2}$. The supervisor intervenes to unwind a bank if $s < \underline{s}$.

Note that the regions a(ii) and b(i), in which the supervisor intervenes with a positive probability, are non-empty only if $\delta \leq \min\{1, \frac{1}{\mu^*} - 1\}$. For higher values of δ the supervisor never intervenes and the analysis of Section 3.5 is directly applicable.

The supervisor's objective function depends on the region in which s_1 is located. Lemma B.1 in Appendix B provides the full details and Figure B.1, also in Appendix B, depicts the objective function for some parameter values. We denote by s_F^P the optimal cut-off in the supervisor private-information case, when there is feedback from stock prices.

Proposition 3.6. *When the supervisor can intervene according to her private information s, the optimal stress test is uninformative (i.e. $s_F^P = 0$) when intervention is costless ($\delta = 0$). When intervention is costly ($\delta > 0$) and $\mu^* \leq \frac{1}{2}$, the optimal stress test is informative and lenient with passing threshold*

$$s_F^P = \begin{cases} s_1^{a(ii)} & \text{if } s_1^{a(ii)} < \mu^*(1-\delta) \\ s_F & \text{if } s_1^{a(ii)} \geq \mu^*(1-\delta), \end{cases} \tag{3.35}$$

where $s_1^{a(ii)} \equiv \delta \frac{1-\frac{\sigma_2}{2}}{\frac{\sigma_2}{2}(pR-1)}$ with σ_2 defined in (B.33), and s_F solves (3.19). When $\mu^* > \frac{1}{2}$ and δ is high enough, the optimal stress test is informative and lenient with passing threshold $s_1 = s_F$.

Proof. See Appendix B. ■

Proposition 3.6 shows robustness of leniency when a privately informed supervisor can directly intervene in banks. Leniency only breaks down in the limit case when the supervisor's intervention cost is $\delta = 0$, in which case the optimal stress test is completely uninformative. When $\delta = 0$, the supervisor always intervenes, based on her private information, and direct market discipline becomes superfluous. Effectively, feedback from prices to real decisions now only acts via direct learning by the supervisor (as in Bond and Goldstein (2015)). The bank's ability to continue and its funding conditions therefore do not depend on the stress test outcome and a negative test result no longer undermines the speculator's information production incentives.[35] These are now maximized when the market maker remains completely in the dark about the bank's type, that is, when the stress test is uninformative ($s_1 = 0$).

Consider now what happens when the intervention cost is positive but very small. We need to distinguish between the cases $\mu^* \leq \frac{1}{2}$ and $\mu^* > \frac{1}{2}$. When $\mu^* \leq \frac{1}{2}$ the default is for banks about which nothing is known to be funded by the market. In this case, it is optimal to set $s_1 > 0$ for any $\delta > 0$. For a small positive s_1, banks with very low resilience will fail the stress test and not get funded by the market. This allows the supervisor to save on the intervention cost for banks with very low resilience levels, i.e., those for which the supervisor is in any case very confident that they should be shut down.[36] For the large majority of banks that pass the test, there will be market information available. In those cases where the market price is uninformative, but the supervisor observes a resilience level $s \in (s_1, \underline{s})$, she can intervene by shutting down the bank. In setting the stress test cut-off s_1 the supervisor trades off the cost of direct interventions against the loss of market information. The optimum is thus a lenient yet informative test for any $\delta > 0$.

When $\mu^* > \frac{1}{2}$ banks do not get funded in the absence of further information. The supervisor thus needs to intervene by funding banks. When moving from $s_1 = 0$ to a small

[35]Note that $\delta = 0$ also eliminates the dilution effect that a more lenient test may have at the roll-over stage. This is because the bank is recapitalized by the supervisor at conditions that entirely depend on her private observation of the bank's resilience, decoupling funding conditions from the stress test result.

[36]Note that a bank with resilience level $s = 0$ is certain to be a low type.

$s_1 > 0$ the market still does not fund a bank even if it passes the test. The supervisor therefore does not benefit from a reduction in intervention costs. Setting s_1 close to 1 would achieve that as banks that pass the test would now get funded, even in the absence of a price signal. It would, however, undermine the speculator's information production incentives: for the few banks that pass the test, the speculator stands to gain little informational advantage, while most of the banks that fail will end up being shut down (only those with $s \in (\bar{s}, s_1)$ will be rescued by the supervisor). It is therefore better to leave the test uninformative for small but positive values of δ, and then jump to the usual lenient test, once δ becomes large enough so that the supervisor prefers to refrain from direct interventions.

To summarize, by allowing the supervisor to learn from stock prices and intervene at a cost, we bridge the gap between two sets of models. First, there are papers that focus on a supervisor learning from stock prices (e.g., Bond and Goldstein (2015) and Siemroth (2019)). Others study the implication of information design on access to capital (e.g., Inostroza (2023), Orlov et al. (2023) or Fuchs et al. (2024)). In this extension we allow for an interplay between these mechanisms: stress tests directly affect funding conditions, but to the extent that the supervisor can intervene at a cost, supervisory learning from stock prices becomes a more important determinant of the information design problem. While the precise stress test design depends on model parameters, we identify leniency as a robust feature.

3.9 Conclusion

This paper models the link between bank stress test design and market discipline. It allows markets to play an indirect role by generating useful price signals about bank fundamentals, and a direct role by providing funding at terms that are sensitive to the stress test results as well as stock price signals. We show that markets are useful at reducing type-2 errors, that is, identifying bad banks that did well in a stress test. Markets are less good at reducing type-1 errors, i.e., providing funds to good banks that did poorly in the stress test. The supervisor optimally distorts the test to be lenient, because this improves market discipline. We extend the model to allow for non-trivial interplay between direct intervention by the supervisor and market discipline. The supervisor prefers to save on direct interventions as they are costly, and therefore enlists market discipline. The optimal stress test is still distorted towards leniency, and it may degenerate into a completely uninformative test when the intervention cost becomes negligible.

Although our model is set up to address the design of bank stress tests, we believe the underlying information design problem is pertinent in other contexts. For example, a credit rating agency needs to decide on a rating system, keeping in mind that this may have an impact on the information that speculators subsequently produce about the issuing firm. Similarly, there is a degree of freedom in setting up accounting rules so that a firm's financial health can appear better or worse (e.g., marking to market versus historical value rules, loan-loss accounting rules etc.). Little is understood about how such rules interact with other sources of information, particularly those contained in stock prices. Our paper proposes a tractable model that can be used in future research to address these questions.

Finally, our model makes empirical predictions concerning the information content of bank stock prices, depending on the stress test design and outcome. These remain to be tested in future research.

Chapter 4

Agents under Pressure: Risk Governance in a Rat Race

Agents under Pressure: Risk Governance in a Rat Race

Matthieu Bouvard, Samuel Lee, Alessio Ozanne

Abstract

We study agency problems in a financial market with preemptive competition. Firms hire agents who search for rival trading opportunities that can, but need not, be beneficial for their respective firms. In each firm, there is tension between incentivizing initiative and compliance with risk governance rules that create latency in trading. Time pressure not only increases the opportunity cost of risk management (that is lost opportunities) but also amplifies the agency rent required to maintain governance. Because each firm's contract choice affects market-wide time pressure, these "contractual externalities" can trigger a race to the bottom culminating in a constrained-inefficient market equilibrium with low risk controls. As a result, there is scope for regulation that i) targets both firms and individual traders, and ii) blends mandates on controls and incentive compensation.

Keywords: Risk Management, Preemptive Competition, Incentive Contracts.
JEL Classification: G20, G32, G01, D86.

4.1 Introduction

There is growing recognition among regulators and financial institutions that the efficacy of risk management cannot be separated from the broader risk culture of the finance industry. The view is that risk-control protocols have limited impact in an environment that does not support and incentivize their effective implementation. The 'London Whale' scandal is one of many rogue trading incidents demonstrating that the controls can be circumvented even in institutions that devote significant resources to risk management systems (McConnell, 2014). In the aftermath of the Great Financial Crisis, the Financial Stability Board published guidance to supervisors on how to assess risk culture in financial institutions.[1] Moving beyond risk modeling and information systems, its approach includes indicators for the commitment from top management to promoting risk management practices, employee accountability, and both explicit and implicit incentives. In addition, it has been argued that risk cultures in finance are shaped not just by firm-level factors but, more

[1] https://www.fsb.org/2014/04/guidance-on-supervisory-interaction-with-financial-institutions-on-risk-culture-a-framework-for-assessing-risk-culture-2/. See also the special issue of the Economic Policy Review (volume 22, issue 1) edited by the New York Fed for an academic treatment of the role of culture, governance and reporting on risk management in financial firms.

broadly, by industry-wide norms or market pressures (Power et al., 2013). This supports an approach in which risk culture is treated as a systemic phenomenon.

In this paper, we propose that one force that shapes financial institutions' attitudes toward risk is time competition. In a number of markets, preemption is a crucial aspect of competition due to first-mover advantages.[2] While there is a large literature that analyzes how this affects strategic interactions *between* firms (e.g., Bouvard and Lee, 2020 for an application to financial firms), less attention has been paid to the impact on governance processes *inside* those firms. It stands to reason that time pressure puts a strain on the governance of decisions and activities that are vital to preempting competitors: finding opportunities before others do and deciding whether to seize opportunities early, even if the risks are not yet fully understood. We formalize this idea to show that time pressure in the market makes it more difficult for firms to incentivize prudent behavior. That is, such market forces create an agency cost of risk management within firms. Moreover, because market forces are endogenous to the behavior of market participants, it holds in turn that firms' risk management incentive structures determine the time pressure in the market. This two-way feedback implies that firms' risk culture—the way incentives within firms shape risk-taking—cannot be separated from their effect and dependence on market-wide forces. The associated externalities also highlight that a "front office culture" equilibrium may not only harm social welfare but also the profitability of the financial industry; a system-wide shift in the risk culture can be Pareto-improving. We discuss merits and drawbacks of various regulatory instruments.

In our model, firms compete for latent and fleeting trade opportunities in a stylized financial market. Identifying these opportunities requires search effort which is delegated to traders and unobservable to the firms' management. A second dimension of moral hazard is that the traders choose to what extent to comply with risk management protocols that can improve the private assessment of a trade value to the firm, but delays execution (pre-trade controls). Firms design compensation contracts that specify wages contingent on trading profits and periodical reviews of trading activity that imperfectly detect non-compliance (post-trade controls).

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

At the firm level, this multi-task conflict generates agency costs. This "partial equilibrium" effect, all else equal, already discourages risk management. In addition, there is an equilibrium effect at the market level. When the speed at which trading opportunities are preempted rises, traders are more strongly tempted to breach risk management protocols to boost trading profit. The tension between search and compliance incentives thus causes the agency rents *to increase with time pressure*, making it optimal for each firm to incen-

[2] Lieberman and Montgomery (1988) identify three types of first-mover advantages: technological leadership, preemption of assets or switching costs of customers.

tivize less compliance under higher time pressure. The firms disregard that, through this channel, the contractual incentives they implement for their own traders—which govern their trading speed—affect the agency cost of risk management and hence the (shape of the) optimal contract used by other firms.

These contractual externalities reinforce the strategic complementarities and thus the scope for a race to the bottom already present in a model of preemptive competition without agency. In the model with agency, time pressure undermines risk management along two dimensions: an "opportunity cost" effect decreases its value to a firm as a whole, while the "agency rent" effect shifts its value increasingly to the firm's traders. Because both of these effects feed back into time pressure and thereby reinforce each other, it is moot to debate in this setting whether risk management failures are to be blamed on the (deliberate) negligence of firms or the moral hazard of its traders—the answer is both.[3] Consequently, legal and regulatory penalties aimed at deterring such failures should target firms and individuals to address both effects.

However, the agency problem is not merely an amplification mechanism. The contractual externalities can create coordination failures among firms *even if the time delay caused by risk management processes goes to zero* in which case the opportunity cost effect disappears. Time pressure in the market depends on compliance even in the absence of the time delay because risk management makes firms trade more *selectively*. Less selective trading strategies cause trade opportunities to disappear faster and so increase time pressure in the market. Since the agency cost of incentivizing compliance rises with time pressure, this selectivity channel alone can render the emergence of "front office cultures"—incentive structures which prioritize trade execution over risk management—a self-enforcing market outcome.

Moreover, the agency problem could cause the outcome to be inefficient even if firms were able to coordinate their incentive contracts. Agency rents amount to a redistribution of surplus from the firms' principals to the traders, and in sufficiently fast or competitive markets, could make risk management too costly for each principal irrespective of the other principals' choices. These different sources and types of inefficiencies, and their interactions, suggest that multiple policy levers may be needed for effective regulation.

If incentive contracts play a central role in the effective implementation of risk management, compensation regulation is a salient policy instrument. Regulation that jointly addresses risk management processes and trader compensation can in principle achieve the first-best outcome in our model. However, the information demands for identifying the optimal rule are daunting. In practice, it has been suggested that caps on salary-incentive ratios—to make compensation less "high-powered"—would promote compliance. But our results caution against the inference that high-powered contracts (with low base wage relative to bonuses) indicate weak compliance incentives. We show that the equilibrium

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

relationship between salary-incentive ratios and risk management quality is ambiguous. Because the severity of the agency problem is endogenous to overall compliance (via time pressure), traders' compensation can become higher or lower-powered when shocks to the trading environment decrease the equilibrium level of compliance. Therefore one cannot infer from high-powered contracts whether a firm finds it too expensive or, on the contrary, cheap to induce compliance, at least not without information about other parameters that determine the degree of time pressure in the market. Indiscriminate limits on salary-incentive ratios across the board can overshoot the target and even backfire.

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

The rest of the paper is organized as follows. Section 4.2 discusses the related literature. Section 4.3 presents the model of delegated trading and risk management, while Section 4.4 derives equilibrium predictions. Section 4.5 contains a regulatory discussion. Section 4.6 concludes. All the proofs are relegated to Appendix C.

4.2 Related Literature

Smith and Stulz (1985) and Froot et al. (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 hedging contracts reduces a firm's capacity to enter such commitments to finance current investment. We do not (explicitly) model hedging decisions; instead, we focus on firms' choices to establish governance systems for risk monitoring. In our model, the resource firms commit to risk management is time and opportunity costs of risk management arise from preemption in financial markets (as in Bouvard and Lee, 2020). Crucially, we also account for the additional agency costs firms incur to ensure traders' compliance with established risk governance systems.

Preemption is similar to the first-come-first-served rule in bank run models (Bryant,

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

1980; Diamond and Dybvig, 1983), from which our model departs in two noteworthy ways.[5] First, absent agency frictions, risk management design is the outcome of long-run decisions that do not coincide with on-the-spot preemptive actions: such organizational choices precede individual trades. Yet since we model trade as randomly staggered through time as a result of independent search processes, preemption motives pass via "time pressure" to risk management choices, which through this medium inherit the strategic complementarities known from bank runs.

Second, we further introduce agency problems that firms must address to implement their chosen risk management protocols. This allows us to study the interaction of agency problems across firms subject to "bank run" externalities. Costly monitoring theories are common in financial intermediation, internal capital markets, and corporate governance (Diamond, 1984; Gertner et al., 1994; Holmström and Tirole, 1997b; Burkart et al., 1997). We are, however, unaware of existing work where the cost of monitoring is time, or depends on competition or on others' monitoring choices.

Our theory provides a rationale for risk management regulation – as distinct from capital or liquidity regulation – similar to theories that justify corporate governance regulation based on externalities. This literature has focused on pecuniary externalities (Acharya and Volpin, 2010; Dicks, 2012) and learning externalities (Raff, 2011; Cheng, 2011; Acharya et al., 2016) in the context of managerial labor markets, we highlight contractual externalities that arise from the impact of competitive pressure on agency rents.

Links between competition and risk taking have been studied in the banking literature with a focus on the effect of competition on bank franchise values (Keeley, 1990; Hellmann et al., 2000b; Boyd and De Nicolò, 2005; Martinez-Miera and Repullo, 2010) and the returns to screening (Ruckes, 2004; Dell'Ariccia and Marquez, 2004).[6] The key mechanism in our paper is that competition raises the (opportunity and agency) costs of screening. Our analysis of agency is reminiscent of Heider and Inderst (2012), who examine a multi-task conflict between screening and loan "prospecting," which, despite lacking an explicit time dimension, bears similarity to preemptive competition.

A literature in industrial organization studies more broadly how competition interacts with agency and has identified a variety of effects operating through information revelation, marginal returns to managerial effort, and total firm income.[7] The overall effect is generally ambiguous, qualifying the "Hicks conjecture" that product market competition curbs managerial slack. These papers typically study one-dimensional moral hazard in oligopolistic models. Our results rely on multi-dimensional moral hazard, and in this respect, are closer to Benabou and Tirole (2014) who study how labor market competition skews contractual incentives across different types of tasks.

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

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

Time-based competition is essential to the sizable literature on innovation and patent races. Most of this literature uses sequential games or real options models in which strategic choices coincide with the acts of preemption.[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 et al. (2006) in which firms that compete on innovation choose ex ante between "mechanistic" and "organistic" organizational designs that differ in production efficiency and "time-to-market."

Recently, time-based competition has become the focus of research on high-frequency (low latency) trading in financial markets. Apart from showing that the race to reduce latency spurs overinvestment in technology, these papers trace out the impact on market liquidity, asset prices, and trading volume (Budish et al., 2015; Biais et al., 2015; Dugast and Foucault, 2018).[9] Our analysis is not specific to high-frequency trading but shares a similar view, and adds to the list of concerns that competing on speed may impair the risk allocation in financial markets by undermining governance processes inside firms. This connects our paper to the literature on the allocative role of secondary capital markets. Most existing results in this literature revolve around (efficient) prices as a source of information that can destroy risk-sharing opportunities, guide investment decisions, enhance incentive contracts, and frustrate takeovers.[10] In our model, allocation is driven not by information revealed through market prices but by processes inside firms, which are, however, affected by the market's speed.

Our focus on a risk management link between markets and organizations combines the perspectives of two earlier papers. Gârleanu and Pedersen (2007) study the role of risk management protocols in creating liquidity feedback loops in the market, whereas Landier et al. (2013) focus on the "dissent" function of risk management inside an organization and when this function may be compromised. In our model, market interactions and organizational choices are jointly determined by a trade-off between immediacy in the market and "dissent" in organizations.

4.3 The Model

4.3.1 Model Description

We present a model in which firms face preemptive competition over trading opportunities, delegate the search for trading opportunities to agents, rely on agents' compliance to conduct risk management, and can use compensation and ex-post monitoring to incentivize the agents. We will derive our results in a setting where infinitely many atomistic traders can continuously adjust their actions. To properly specify that model, we present it here as

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

[9]A notable exception is Pagnotta and Philippon (2018) who, instead of focusing on preemption among traders, study the incentives of securities exchanges to offer trading platforms of different speed to heterogenous traders.

[10]See, e.g., Bond et al. (2012b) for a survey of this literature.

the limit of a model where n traders of size $s(n) \equiv 1/n$ can adjust their action at discrete time intervals.

Trading Time is continuous but divided into periods of length Δ each indexed by the time t at which it starts, i.e., period t represents the interval of time $[t, t+\Delta)$. Exogenous shocks create trade opportunities that cannot be exploited immediately due to informational frictions.[11] A number $i\Delta$ of these opportunities are born at the beginning of each period, and each stays alive until a trade has been executed with respect to it.

Traders are employed in different firms and seek to exploit trading opportunities. At the beginning of each period t, trader $k \in \{1, ..., n\}$ chooses his search intensity $e_{kt} \in [0, \lambda^{-1}]$ for the period. A trader who searches with intensity e_{kt} in period $[t, t+\Delta)$ bears a cost $ce_{kt}s(n)\Delta$ but, in return, finds each live trading opportunity with a probability $e_{kt}s(n)\Delta$ over the period. Note that through its dependence on size $s(n)$, the probability a trader finds a given opportunity in a given period tends to 0 as n increases. This is the sense in which traders become atomistic as $n \to +\infty$.[12]

Two intuitive properties of this model are that search exhibits a scale effect and opportunities grow stale: the more opportunities are alive, the more are found for given effort, and opportunities of any given vintage become rarer over time as more of them will have been exploited. The scale effect is similar to the constant-returns-to-scale property often assumed for the matching function in search-matching models.

Risk Management Trading on an opportunity yields π, which is a common value for all firms, and a firm-specific private value α_{kt}.[13] We view π as the direct dollar profit from a trade. By contrast, α_{kt} does not represent a trait of the trade *per se* but the "fit" of the trade with the risk profile of firm k.[14]

There is uncertainty about the private value of a discovered trade opportunity. At the time of discovery, trader k knows only that

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

[11]The shocks could change fundamental asset values, create discrepancies in the prices of correlated assets, or shift (latent) demand and supply for liquidity. Thus, trade opportunities could be based on information about fundamental values, identifying (discrepancies between) affected assets, locating potential counterparties, or conceiving of creative ways to overcome limits to arbitrage.

[12]If the instantaneous probability of discovering a live trading opportunity was not going to 0 as $n \to +\infty$, then opportunities would be instantaneously found as $n \to \infty$. This assumption captures the idea that trading opportunities are latent.

[13]We index the private values only by trader identity k and discovery time t, even though it could also depend directly on the trading opportunity. Since trading opportunities are found sequentially and all α_{kt} are independently distributed, indexing by trading opportunity would be redundant.

[14]For example, the trade may exacerbate imbalances in the firm's risk profile that raise potential distress costs or lower collateral pledgeability (See Froot et al., 1993 and Rampini and Viswanathan, 2010). Such changes in shadow costs are reflected in α_{kt}. Viewing α_{kt} as a private value is consistent with canonical theory (e.g. Froot et al., 1993), where the rationale for risk management is a private value (firm-specific benefits of hedging idiosyncratic risks) of writing financial contracts traded at common value (market prices of hedging contracts).

Private value realizations are independent across firms and trade opportunities (and time), and have mean zero. We assume $-\alpha_- > \pi > -\alpha_+$ and $\pi > 0$, that is, a trade is desirable for firm k if and only if $\alpha_{kt} = \alpha_+$. Together with the zero mean assumption, this implies that

$$0 < \pi < \frac{\rho}{1-\rho}\alpha_+. \tag{A.0}$$

Firms can produce information on α_{kt} in order to condition trading decisions on it; we refer to this process of managing trade positions based on assessed risk "fits" as risk management. Risk management has the information benefit that a requested trade is executed only if $\alpha_{kt} = \alpha_+$. However, the process has two limitations. First, it requires a deterministic time $\iota > 0$, and hence delays trade execution. Second, it requires the cooperation of the trader who discovered the opportunity. Specifically, at the beginning of each period t, the trader can choose a probability $m_t \in [0, 1]$ with which trade opportunities found in $[t, t + \Delta)$ bypass risk management. We interpret this as violating *pre-trade* protocols or deceiving a "back office," and in this spirit, refer to m_t as the manipulation rate.

Risk Governance The trader's tasks are subject to moral hazard. The first task is to search. Instead of exerting costly effort to search for trade opportunities with common value π, a trader can submit frivolous trade ideas, which require no effort and generate zero profit. Search effort and thus the (ex ante) nature of a trade idea are unobservable. The scope for frivolous trading makes it suboptimal to reward trade volume.[15] Firms hence incentivize search based on trade performance. Direct trading profits, i.e., realized common values π are publicly observed and contractible. By contrast, private values α_{kt} are not publicly observed and non-contractible.

The second task is to comply with pre-trade protocol. However, manipulation is unobservable. Without contractible signals about the private values, incentivizing compliance requires a contractible signal about manipulation. We assume that firms address this problem through *post-trade* reviews. Specifically, if a trader's average manipulation rate over some history is $\bar{m}$, a review detects verifiable evidence of manipulation, which contracts can be conditioned on, with probability $\gamma\bar{m}$ where $0 < \gamma < 1$. The assumption that this probability is linear in $\bar{m}$ is not crucial; our results hold for $\gamma'(\bar{m}) \geq 0$ and $\gamma''(\bar{m}) \geq -[\gamma'(\bar{m})]^2/[1 - \gamma(\bar{m})]$.

While the firms have infinite lives, traders are employed for a finite number D/Δ of periods, i.e., for a time D, and are replaced at no cost. We assume, for simplicity, that every trader undergoes post-trade reviews once, after their employment ends. Under this assumption, the length of employment D is immaterial for our equilibrium analysis; it can be arbitrarily long or short, or differ across firms and traders.

[15]The reason is not that agents can engage in limitless frivolous trading; the mass of trades that could credibly be requested is bounded from above by the maximum rate at which one can find valuable opportunities in equilibrium. The reason is that offering any reward for trade volume—and thereby for frivolous trades—makes it more expensive for the firm to incentivize genuine search effort.

In sum, to incentivize a trader's two tasks, a firm has as many governance instruments at its disposal. It can condition a trader's wage on his realized trading profits, denoted $\{\Pi_{kt}\}$, and on a binary signal indicating whether there is evidence of manipulation. In Section 4.3.2, we motivate these assumptions and discuss which ones are crucial for our results.

Additional Assumptions We exclude certain parameter constellations to condense the analysis. To ensure preemptive competition, we assume that the mass of firms ρ for whom a given trade is a good "fit" exceeds the size i of any vintage (of trade opportunities), i.e.,

$$\frac{i}{\rho} < 1. \tag{A.1}$$

We also bound the marginal cost of search from above:

$$c < \pi i \lambda. \tag{A.2}$$

This ensures that firms find it worthwhile to incentivize some search effort. It also implies that, in the absence of agency problems, search intensity would be set to its upper bound λ^{-1}. These restrictions suppress cases in which time is not a concern or firms are inactive in the market. Last, we assume that everyone is risk-neutral and has a discount factor of 1.

4.3.2 Discussion of Model Assumptions

Firms in our model firms acquire information about investment opportunities. We model this as a two-stage process for the following reasons. First, once information gathered up to a point suggests that a trade opportunity is attractive ("search"), there is still a choice between trading immediately and waiting for more information ("risk assessment"). Second, we think of traders searching in *market* data while risk management also aggregates *firm* data. A trader has some knowledge about the quality of his trades but less about how his trades interact with other exposures of the firm. It is also unlikely that there are publicly verifiable ex-post signals of "fit" assessments across time-varying risk exposures from decentralized trades across a firm. This motivates our assumption that the private values are hidden information for which there are no contractible signals; we treat them like "private benefits"—to the principal.

Third, this goes hand in hand with the assumption that the firms use *pre-trade approval* to make trade decisions contingent on information about α_{kt} and rely on *post-trade monitoring* to punish non-compliance with that requirement. This dual control process has a role to play to the extent that compensation cannot make traders condition on α_{kt}. If firms could condition traders' wages on the realized sums $\pi + \alpha_{kt}$, the controls would be redundant, and information acquisition would de facto become a one-stage process (of evaluating the sum).[16]

[16]The *complete* absence of contractible signals for the private values α_{kt} is not crucial. As long as the firms use control processes in addition to *noisy* signals of α_{kt}, our results remain qualitatively robust. There is limited value in expanding the model in this direction, while economizing on this aspect seems realistic.

In practice, pre-trade approvals are implicit in *risk limits* that require traders or business units to stay below preset risk thresholds. Examples of risk metrics used in setting limits include notional value ("position limits"), market risk measures (e.g., volatility, duration, or convexity), correlations ("concentration limits"), losses ("stop-loss"), and value-at-risk. The purpose of these limits is to provide a check on traders' performance and to trigger reviews of any positions that breach the limit because of concerns about the increased exposure of the firm. Risk management sets these limits and handles exceptions. Suppose a trader requires approval to trade more than l units of an asset since it would breach a limit. Pre-trade approval in our model proxies for such exceptions, with l normalized to 0. Managing exceptions and monitoring compliance depends in part on information provided by traders. Because this leaves room for manipulation, firms conduct ex post reviews of their traders' valuation assumptions, a process known as *valuation control*.[17] But this process is imperfect, as reflected in the model assumption that manipulation sometimes goes undetected.[18]

Finally, the constant birth of trading opportunities serves to allow for *steady states*, where calendar time is uninformative about latent market conditions, improving the tractability of the model.

4.4 Analysis

4.4.1 Continuous Time and Atomistic Traders Limit

A key feature of our model is that it provides analytical tractability when the time increment Δ tends to 0 and the number of traders n tends to ∞. In that limit, we can characterize a steady state of the model where search effort e_t and manipulation rate m_t are time invariant.

To see this, suppose that all traders expend effort e in period t. The probability that a trading opportunity alive at t is found in that period is $1 - (1 - \frac{e}{n}\Delta)^n$. Therefore, in the continuous-time limit, the probability that a specific trade opportunity, still alive, is found instantaneously is then

$$\lim_{\Delta \to 0} \frac{1 - (1 - \frac{e}{n}\Delta)^n}{\Delta} = e. \tag{4.1}$$

In other words, if all traders search at intensity e, discoveries of a trading opportunity follow a Poisson process with rate e. Hence, continuous-time decision-making generates a

[17] "Rogue traders" are the most infamous examples of traders manipulating data to undercut risk controls, if only because their actions culminated in spectacular losses for the affected banks. A short list includes Jerome Kerviel (Societe Generale), Kweku Adoboli (UBS), Nick Leeson (Barings Bank), John Rusnack (Allied Irish Bank), Yasuo Hamanaka (Sumitomo Corporation), and Toshihide Iguchi (Daiwa Bank).

[18] Valuation controls are also known as *post-trade audits*, and especially relevant for complex trades, derivatives and structured products, over-the-counter or illiquid markets, and generally all settings where publicly observable market prices are unavailable, unreliable, or imprecise. In its settlement with the Securities Exchange Commission over the London Whale trading scandal, J.P. Morgan Chase acknowledged that the valuation control unit failed to properly oversee those trading activities (Trindle, 2013).

simple and analytically convenient discovery process, in addition to avoiding some cases of limited economic interest.[19]

In addition, as discussed already, when the number of traders n tends to ∞, the probability that a trader finds a given opportunity becomes negligible. The continuum of traders along with the i.i.d. random variables implies that, in effect, there is only *strategic uncertainty*: for traders (and firms) to infer the state of the market in every instant merely requires conjectures about each other's governance choices.

4.4.2 Steady-State Market Conditions

Suppose now that all traders use the same constant search intensity e and manipulation rate m and consider the distribution of the lifetime of a trade opportunity. Precisely, what is the probability that an opportunity born at t is alive at $t+T$? If $T < \iota$, this requires that all discoveries of the opportunity between t and $t+T$ are run through risk assessment. The probability of this is the cumulative Poisson probability over all the possible numbers of times $j \in \mathbb{N}$ the opportunity is found in a time interval of length T, each multiplied with the probability that every discovery is referred to risk assessment:

$$\exp\left[-eT\right]\sum_{j=0}^{+\infty}\frac{(eT)^j}{j!}(1-m)^j = \exp\left[-eT\right]\exp\left[eT(1-m)\right] = \exp\left[-emT\right]. \tag{4.2}$$

If $T > \iota$, an opportunity born at t is alive at $t+T$ if (a) all discoveries between t and $t+T-\iota$ are assessed and found to be of negative private value and (b) all discoveries between $t+T-\iota$ and $t+T$ are assessed. The probability of the intersection of these events is

$$\begin{aligned}
& \overbrace{\exp\left(-e(T-\iota)\right)\sum_{j=0}^{+\infty}\frac{(e(T-\iota))^j}{j!}\left[(1-m)(1-\rho)\right]^j}^{\text{Part a}} \times \overbrace{\exp\left(-e\iota\right)\sum_{j=0}^{+\infty}\frac{(e\iota)^j}{j!}(1-m)^j}^{\text{Part b}} \\
= \;& \exp\left(-e(T-\iota)\right)\exp\left(e(T-\iota)(1-m)(1-\rho)\right)\times\exp\left(-e\iota\right)\exp\left(e\iota(1-m)\right) \\
= \;& \exp\left[-(em)T - e(T-\iota)\rho(1-m)\right].
\end{aligned} \tag{4.3}$$

It follows that

$$Z_{e,m}(T) \equiv \min\{\exp\left[-(em)T\right], \exp\left[-(em)T - e(T-\iota)\rho(1-m)\right]\}$$

is the probability that the survival time of a trade opportunity exceeds the span T.

The total mass of trade opportunities that are alive at t and were born after $t-T$ is the sum of all the opportunities with a birth date $\tau \in [t-T, t]$ times the probability that such an opportunity survives for a time span $t-\tau$.

$$\int_{t-T}^{t} iZ_{e,m}(t-\tau)d\tau = \int_{0}^{T} iZ_{e,m}(\tau)d\tau. \tag{4.4}$$

[19] For instance, in continuous time, the probability that multiple traders find the same opportunity at the same time is zero, which avoids devising tie-breaker rules.

Now letting T tend to ∞ yields the total mass of trading opportunities alive at t:

$$\begin{aligned} I(e,m) &\equiv \int_0^{+\infty} iZ_{e,m}(\tau)d\tau \\ &= \frac{i}{e}\left[\frac{1}{m} - \exp\left(-e\iota m\right)\left(\frac{1}{m} - \frac{1}{m+\rho(1-m)}\right)\right]. \end{aligned} \tag{4.5}$$

The probability that an opportunity randomly picked at t was born after $t-T$ is the ratio of (4.4) to (4.5),

$$\mathcal{G}_{e,m}(T) \equiv \frac{\int_0^T Z_{e,m}(\tau)d\tau}{\int_0^{+\infty} Z_{e,m}(\tau)d\tau}. \tag{4.6}$$

$\mathcal{G}_{e,m}(T)$ defines the cumulative distribution function of the age of trade opportunities. It is time-invariant if traders play a time-invariant strategy (e,m). In that case, if $\mathcal{G}_{e,m}(T)$ describes the age distribution of trade opportunities at t, it also describes the age distribution of trade opportunities at all $t' > t$.

When $\mathcal{G}_{e,m}(T)$ is stationary, the preemption risk for a trader who submits a trade to risk assessment is stationary. Consider the probability that a trade idea referred to risk assessment will no longer be available once the assessment is completed. By stationarity, the mass of trade opportunities alive at t that disappears between t and $t+\iota$ is equal to the mass of opportunities born between t and $t+\iota$ and still alive at $t+\iota$. The latter is

$$\int_t^{t+\iota} iZ_{e,m}(t+\iota-\tau)d\tau = \frac{i}{em}\left[1-\exp\left(-em\iota\right)\right].$$

Dividing this by the total stationary mass of trading opportunities alive defined in (4.5) yields the probability that a trade opportunity referred to risk assessment in t disappears by $t+\iota$:

$$p(e,m) \equiv \left(1 + \frac{m}{\rho+m(1-\rho)}\frac{\exp\left(-e\iota m\right)}{1-\exp\left(-e\iota m\right)}\right)^{-1}. \tag{4.7}$$

We will refer to $p(e,m)$ as the *preemption risk* and to $I(e,m)$ as the *opportunity stock* in steady state.

Note that the market asymptotically tends to the steady-state distribution $\mathcal{G}_{e,m}(T)$ as time passes and traders' strategies are kept fixed. If at time t the distribution of trading opportunities is exactly $\mathcal{G}_{e,m}(T)$ and traders play the strategy (e,m), then the distribution $\mathcal{G}_{e,m}(T)$ is stationary. In what follows we focus on steady-state equilibria, that is equilibria such that the equilibrium strategy (e,m) is constant over time and optimal given the distribution $\mathcal{G}_{e,m}(T)$. A full characterization of trade opportunities in the market requires the age distribution $\mathcal{G}_{e,m}(T)$, but the market conditions $I(e,m)$ and $p(e,m)$ will be sufficient statistics for firms' and traders' decisions in a steady state.[20]

[20]Note that p is not conditional on the age a (time since birth) of a trade opportunity. Under the assumption that the manipulation rate for $[t, t+dt)$ is set "ex ante" at t, traders only care about average preemption probabilities. This assumption is not innocuous, even though the Poisson process is memoryless, because risk management delays make the depletion rate of an opportunity dependent on its age. In

Lemma 4.1 (Steady-State Market Conditions)**.** *For any time-invariant strategy* (e, m)*, there exists a stationary distribution of trading opportunities age distribution* $\mathcal{G}_{e,m}(T)$*. This distribution is characterized by an opportunity stock* I *and preemption risk* p*, such that, if* $e > 0$ *and* $m < 1$*, we have that*

- *opportunity stock* I *decreases in* e*,* m*, and* ρ*, and increases in* ι *and* i*;*
- *preemption risk* p *increases in* e*,* m*,* ρ*, and* ι*.*

Proof. See Appendix C. ■

The comparative statics are intuitive: (i) Any increase in the speed or frequency of trading—be it due to more search e, more manipulation m, higher trade approval rates ρ, or shorter risk management delays ι—raises the exit rate of opportunities, which, for a given birth rate, lowers the steady-state opportunity stock; (ii) a higher exit rate also raises preemption risk, except when it is caused by shorter delays ι: faster risk management leads to fewer foregone opportunities.

4.4.3 Steady-State Optimal Contracts

We now study a single firm-trader pair and derive an optimal contract taking market conditions as given. We focus on contracts where incentive-compatibility constraints are binding (as usual in principal-agent models) by giving all bargaining power to the firm and assuming limited liability for the trader.

4.4.3.1 Path-Independent Contracts

To construct a steady-state equilibrium, we must show that, under stationary market conditions, there exists an optimal contract that implements a time-invariant trader strategy. The firm offers a contract to hire a trader for the period $[t_0, t_0 + D]$. We normalize the trader's reservation wage to 0, and without loss of generality, set $t_0 = 0$. A generic contract is a pair of functions

$$\mathcal{W} = \{\underline{W}(\{\Pi_t\}_D), \overline{W}(\{\Pi_t\}_D)\}$$

which specify the trader's wage as a function of the entire profit path $\{\Pi_t\}_D$, respectively, when manipulation is detected ex post and when none is detected. Subscript D is short for $t \in [0, D]$. Π_t is the total common-value profit from all opportunities discovered in t and eventually executed (with or without risk assessment). Under stationary market conditions (I, p), this is

$$\Pi_t = I e_t \left[m_t + (1 - m_t)(1 - p)\rho\right] \pi \equiv \Pi(e_t, m_t). \tag{4.8}$$

fact, with deterministic delay ι, the preemption probability as a function of age has a kink at $a = \iota$. These features make the analysis considerably more complex. However, it should still be possible to construct a steady state, in which $m(a)$ is time-independent and uniform across traders and the distribution of a is stationary.

Recall that this is not equal to the value of trading to the firm, which includes the private values. Gross of wages, that value is

$$V_t = Ie_t[m_t\pi + (1-m_t)(1-p)\rho(\pi + \alpha_+)] \equiv V(e_t, m_t). \tag{4.9}$$

Given that each trader is infinitesimal and there is no aggregate exogenous uncertainty, no trader directly conditions actions on those of other traders and the market evolves deterministically. A trader's strategy can hence be written as $\{\tilde{e}_t, \tilde{m}_t\}_D$, with tildes indicating possible randomization, and the trader's strategy choice can be formulated as the ex-ante problem

$$\underset{\{\tilde{e}_t,\tilde{m}_t\}_D}{\text{maximize}} \quad \mathbb{E}\left[\tilde{w}(\{\tilde{e}_t, \tilde{m}_t\}_D, \mathcal{W}) - \int_0^D c\tilde{e}_t dt\right] \tag{4.10}$$

where $\tilde{w}(\{e_t, m_t\}_D, \mathcal{W})$ is his wage given a contract $\mathcal{W}$ and an ex-post realized strategy $\{e_t, m_t\}_D$, and the expectation is taken at the time of hire. The firm's contracting problem is then

$$\begin{aligned} &\underset{\mathcal{W}}{\text{maximize}} \quad \mathbb{E}\left[\int_0^D V_t dt - \tilde{w}(\{\tilde{e}_t, \tilde{m}_t\}_D, \mathcal{W})\right] \\ &\text{subject to} \quad \overline{W}(\{\Pi_t\}_D), \underline{W}(\{\Pi_t\}_D) \ \geq 0, \\ &\qquad\qquad\quad \text{and } \{e_t, m_t\}_D \text{ solves (4.10).} \end{aligned} \tag{4.11}$$

The key insight in the construction of the stationary equilibrium is that, in solving the contracting problem in a stationary market environment, i.e., (4.11), firms can limit attention to contracts that are contingent on a trader's *average* common-value profit

$$\bar{\Pi} \equiv \frac{1}{D}\int_0^D \Pi_t dt = I\bar{e}[\bar{m} + (1-\bar{m})(1-p)\rho]\pi = \Pi(\bar{e}, \bar{m}), \tag{4.12}$$

or total profit $\bar{\Pi}D$, instead of the entire profit path $\{\Pi_t\}_D$, where

$$\bar{e} \equiv \frac{1}{D}\int_0^D e_t dt \qquad \text{and} \qquad \bar{m} \equiv \frac{\int_0^D m_t e_t dt}{\int_0^D e_t dt}; \tag{4.13}$$

are a trader's average search intensity and manipulation rate. In turn, given a stationary market environment and given contracts only contingent on average profit, agents can limit attention to time-invariant strategies to produce whichever average profit they choose to achieve.

Lemma 4.2 (Path-Independent Contracts)**.** *Under stationary market conditions* (I, p), *there exists an optimal contract such that*

1. *if the trader generates average profit* $\bar{\Pi}$, *he is paid* $\overline{W}(\bar{\Pi})$ *when no manipulation is detected, and* $\underline{W}(\bar{\Pi})$ *when manipulation is detected,*

2. *the trader exerts constant search and manipulation, i.e. for every* t, $e_t = e$ *and* $m_t = m$.

Proof. See Appendix C. ∎

To sketch the intuition for the result, note first that the total value of a given path of trading *to a firm* is $\bar{V}D$ where

$$\bar{V} \equiv \frac{1}{D}\int_0^D V_t dt = \Pi(\bar{e},\bar{m}) + (1-\bar{m})(1-p)\rho\alpha_+, \tag{4.14}$$

comprises the average common-value profit plus the average gain from (partly manipulated) risk management. We see in (4.14) that, in a stationary market environment and gross of wages, the firm only cares about a trader's *average* search intensity and manipulation rate $(\bar{e},\bar{m})$. The question is whether incentivizing some average strategy $(\bar{e},\bar{m})$ is more effectively achieved by contracting on the entire profit path $\{\Pi_t\}_D$ than by contracting merely on the average profit $\bar{\Pi}$. The basic issue is that, in the trader's strategy choice, search and manipulation are partial substitutes in generating trades and thereby common-value profits. But, crucially, because this substitutability applies just as much at the level of instantaneous profits Π_t as at the level of average profits $\bar{\Pi}$, it turns out that contracting on the former restricts the trader's strategy no more than contracting on the latter.

For a cursory illustration, let $\Omega_{\hat{\Pi}} \equiv \{\{\Pi_t\}_D : \bar{\Pi} = \hat{\Pi}\}$ denote the set of all feasible profit paths that generate average profit $\hat{\Pi}$. Seeing in (4.13) that average profits only depend on average search and manipulation, let $\Sigma_{\hat{\Pi}} \equiv \{(\bar{e},\bar{m}) : \Pi(\bar{e},\bar{m}) = \hat{\Pi}\}$ denote the set of all feasible combinations of $\bar{e}$ and $\bar{m}$ that generate $\hat{\Pi}$. Arbitrarily pair *any* $\{\Pi'_t\}_D \in \Omega_{\hat{\Pi}}$ with *any* $(\bar{e}',\bar{m}') \in \Sigma_{\hat{\Pi}}$. Construct the dynamic trader strategy

$$e_t = \frac{\Pi'_t}{\hat{\Pi}}\bar{e}' \qquad \text{and} \qquad m_t = \bar{m}'. \tag{4.15}$$

By (4.8), this strategy generates the instantaneous common-value profits

$$\Pi_t = \frac{\Pi'_t}{\hat{\Pi}} I\bar{e}'\left[\bar{m}' + (1-\bar{m}')(1-p)\rho\right]\pi = \frac{\Pi'_t}{\hat{\Pi}}\hat{\Pi} = \Pi'_t,$$

where the second equality uses (4.12). By (4.13), the average search intensity and manipulation rate under this strategy are

$$\bar{e} = \frac{1}{D}\int_0^D \frac{\Pi'_t}{\hat{\Pi}}\bar{e}' dt = \bar{e}'\frac{\hat{\Pi}}{\hat{\Pi}} = \bar{e}' \qquad \text{and} \qquad \bar{m} = \frac{\int_0^D \bar{m}' e_t dt}{\int_0^D e_t dt} = \bar{m}'.$$

That is, the strategy constructed in (4.15) generates the chosen profit path $\{\Pi'_t\}_D$ with the chosen average search intensity and manipulation rate $(\bar{e}',\bar{m}')$. This argument can be applied to any pair of elements from $\Omega_{\hat{\Pi}}$ and $\Sigma_{\hat{\Pi}}$, which means that any path $\{\Pi_t\}_D$ that yields some average profit $\hat{\Pi}$ can be implemented with any pair $(\bar{e},\bar{m})$ that yields that average profit. Hence, conditioning the contract on elements in $\Omega_{\hat{\Pi}}$ restricts the trader's choice of elements from $\Sigma_{\hat{\Pi}}$ no more than merely conditioning the contract on $\hat{\Pi}$.

By Lemma 4.2, the program (4.10)-(4.11) can be reduced to a quasi-static problem: We can analyze it as if contracts can only be contingent on average (or total) common-value profits and the post-trade review,

$$\mathcal{W} = \{\underline{W}(\bar{\Pi}), \overline{W}(\bar{\Pi})\},$$

and as if a trader chooses a single search effort e and manipulation rate m for her whole tenure to maximize

$$(1 - \gamma m)\overline{W}(\Pi(e, m)) + \gamma m \underline{W}(\Pi(e, m)) - ceD. \tag{4.16}$$

Here e and m carry no time index t to indicate that they are time-invariant.

4.4.3.2 Cost-Minimal Contracts

Suppose the firm designs the contract $\mathcal{W}$ to incentivize the trader to choose a specific pair (e, m). A contract that achieves this while minimizing payments to the trader takes the following form.

Lemma 4.3 (Cost-Minimal Contract)**.** *A cost-minimal contract that incentivizes search effort e and manipulation m exhibits the following two features:*

- *a* **bonus***: $\overline{W}(\bar{\Pi}) > 0$ and $\overline{W}'(\bar{\Pi}) \geq 0$ if $\bar{\Pi} \geq \Pi(e, m)$ and $\overline{W}(\bar{\Pi}) = 0$ otherwise;*
- *a* **clawback***: if $m < 1$, $\underline{W}(\bar{\Pi}) < \overline{W}(\bar{\Pi})$ for $\bar{\Pi} \geq \Pi(e, m)$; moreover, if $m < m_0(p)$, $\underline{W}(\bar{\Pi}) = 0$ for any $\bar{\Pi}$, where*

$$m_0(p) \equiv \max\left\{0, \min\left\{\frac{1}{2}\left[\frac{1}{\gamma} - \frac{(1-p)\rho}{1-(1-p)\rho}\right], 1\right\}\right\}.$$

Under a cost-minimal contract, the expected payment to the trader per unit of time is

$$\mathbb{E}(w_t) = \begin{cases} ce & \text{if } m \geq m_0, \\ ce + r(e, m) & \text{if } m < m_0, \end{cases}$$

where

$$r(e, m) \equiv \frac{2(1-(1-p)\rho)}{(1-p)\rho + (1-(1-p)\rho)m}(m_0(p) - m)ce$$

represents an agency rent.

Proof. See Appendix C. ∎

The bonus provision specifies that the trader receives a strictly positive payment if and only if the average (common-value) trading profit meets (or exceeds) the threshold $\Pi(e, m)$. The clawback provision specifies that part—or if the firm implements a manipulation rate m below m_0, all—of the trader's wage is clawed back if the post-trade reviews detect manipulation. If the firm is willing to tolerate a manipulation rate above m_0, the trader merely recoups her effort costs. By contrast, if the firm wants to incentivize a manipulation rate below m_0, it must pay the trader an agency rent $r(e, m)D$ over and above compensating her for search effort.

The agency rent reflects an endogenous tension between the two tasks, search and compliance. To see this, consider either task in isolation. Suppose m is fixed. The firm can implement any e with a contract under which $\mathbb{E}(\tilde{w}) = (1-\gamma m)\overline{W}(\bar{\Pi}) + \gamma m \underline{W}(\bar{\Pi})$ equals ceD for all $\bar{\Pi} \geq \Pi(e,m)$ and zero otherwise.[21] Since m is fixed, the simplest contract sets $\overline{W} = \underline{W}$ and hence—regardless of post-trade reviews—simply pays the trader a bonus ceD if her performance reaches the target profit $\Pi(e,m)$; the trader just breaks even. Now suppose e is fixed, so the contract only affects m. We write the difference quotient for the trader's objective (4.16) with respect to changes in m as

$$\frac{\partial \mathbb{E}(\tilde{w})}{\partial \Pi}\frac{\partial \Pi}{\partial m} - \gamma(\overline{W} - \underline{W}). \tag{4.17}$$

Setting $\overline{W} - \underline{W} > 0$ makes manipulation costly. Since manipulation boosts trade profit, $\frac{\partial \Pi}{\partial m} > 0$, the trader may still be inclined to manipulate if his expected wage increases with trade profit. But the firm can shut this down by decoupling wages from trade profit: $\frac{\Delta \mathbb{E}(\tilde{w})}{\Delta \Pi} = 0$. A simple contract is $\overline{W} = ceD$ and $\underline{W} = 0$: the firm pays a fixed wage that covers the trader's effort cost, but claws back the wage if evidence of manipulation is found. Under this contract, (4.17) collapses to $-\gamma ceD$, so the trader chooses $m = 0$ and breaks even. In fact, if e is fixed, the firm can also elicit $m = 0$ by paying a *fixed* wage $\overline{W} = \underline{W} = ceD$ such that the trader is indifferent regarding m. Thus neither task *per se* generates agency rents in our model,[22] but incentivizing both imposes countervailing demands on contract shape: Incentivizing search requires that pay be sensitive to trading profit, inducing an incentive to trade. In direct contrast, compliance is best incentivized by decoupling pay from trade profit to mute the incentive to (convert a discovered opportunity to a) trade.

However, the key aspect of the model is not merely that search and compliance create countervailing incentives – a feature common to multitasking problems (see, e.g., Holmström and Milgrom (1991)) – but rather that the intensity of the resulting agency problem depends on time pressure in the market. To see this, derive from (4.8) the iso-profits curve for a target profit level of $\bar{\Pi}$:

$$m(e) = \frac{1}{1-(1-p)\rho}\left(\frac{\bar{\Pi}}{eI\pi} - (1-p)\rho\right).$$

That is, $m(e)$ is the level of manipulation required to reach common-value profit $\bar{\Pi}$ given a search effort e. The marginal rate of substitution between search e and compliance m is then how much manipulation needs to increase (at the margin) to make up for a marginal decrease in effort and keep the profit on target $\bar{\Pi}$. This is given by the negative of the slope of $m(e)$:

$$MRS_{e,m} = -\frac{\partial m}{\partial e} = \frac{\bar{\Pi}}{(1-(1-p)\rho)I\pi e^2}.$$

[21] The trader is then indifferent between effort e and no effort; any other effort level is strictly dominated.

[22] In the case of search effort, this is because, conditional on a manipulation rate m, average trade profit perfectly reveals average search intensity. Given our continuum and i.i.d. assumptions, this would hold even if the profit from each trade were random. Adding exogenous shocks that render average profit $\bar{\Pi}$ a noisy signal of search effort exacerbates agency costs. As is, our model abstracts from agency costs of search, conveniently isolating those that arise purely from risk governance.

Note that this marginal rate of substitution is decreasing in the preemption risk p,

$$\frac{\partial MRS_{e,m}}{\partial p} = -\frac{\bar{\Pi}\rho}{(1-(1-p)\rho)^2 I\pi e^2} < 0.$$

So, as time pressure intensifies and given fixed costs of effort and manipulation, it becomes increasingly tempting for the agent to substitute manipulation for search effort. Intuitively, when preemption risk intensifies, an increase in manipulation has a larger impact on profits because an increasingly high fraction of the discovered opportunities are lost during risk assessment due to competition. This, in turn, implies that the cost for the firm of maintaining a given level of search and compliance, (e, m), increases with time pressure p. We formally state this partial equilibrium result in the following proposition.

Proposition 4.1 (Cost of Risk Management)**.** *If $m < m_0$, the firm's marginal cost of incentivizing risk management, i.e. $-\frac{\partial r(e,m)}{\partial m}$, is increasing in the preemption risk p.*

Proof. See Appendix C. ■

Note also that the highest level of compliance that can be induced without leaving the trader a rent, m_0, is itself increasing with preemption risk. In other words, heightened time pressure not only causes the agency rent to increase for compliance levels where it is already strictly positive, but it also widens the range of compliance levels requiring a strictly positive agency rent.

4.4.4 Equilibrium Risk Management

The previous section treats preemption risk p as a parameter that affects firms' cost of incentivizing search and risk management. We move now from this partial equilibrium within a firm to the analysis of the general equilibrium where p becomes the endogenous product of firms' collective decisions to set up incentive structures for traders.

We start the analysis with a result that simplifies one dimension of firms' best responses.

Lemma 4.4 (Optimal Search)**.** *In an optimal contract, the firm implements maximal effort, i.e. $e^* = \lambda^{-1}$.*

Proof. See Appendix C. ■

To understand this result, note that given Assumption A.2, incentivizing search with no risk management generates strictly positive profits. It follows that at the optimal level of compliance m^* the marginal benefit of incentivizing search is strictly positive. Then given that effort cost is linear, firms choose the maximal level $e^* = \lambda^{-1}$. The linearity of the effort cost provides tractability because the corner solution we obtain for search reduces the analysis of firms' best response to the single-dimensional choice of (induced) compliance m. This also implies that firms respond to increased time pressure by reducing risk management, when a more convex specification of the search cost would generate a reduction in both risk management and search effort. To the extent that the source of inefficiency in the model comes from the level of risk management, this specification also focuses the analysis on the main friction.

To isolate the effect of the agency problem, it is useful to start from a benchmark equilibrium where this friction vanishes, i.e., the case where each firm can directly choose a level of search effort and of risk management.

Proposition 4.2 (No Agency Benchmark)**.** *In the setting with no agency, search effort is maximal, (i.e., $e^\star = \lambda^{-1}$) and there exist two thresholds $0 < \underline{\iota} < \bar{\iota}$ such that:*

- *for $\iota < \underline{\iota}$, there is full risk management (i.e. $m^\star = 0$);*
- *for $\iota > \bar{\iota}$, there is no risk management (i.e. $m^\star = 1$);*
- *for $\iota \in [\underline{\iota}, \bar{\iota}]$, both equilibria coexist.*

Proof. See Appendix C. ■

An analogous result is shown in Bouvard and Lee (2020) where coordination failures can lead to inefficiently low levels of risk management, even absent agency frictions. Intuitively, the latency introduced by risk management in the trading process creates an opportunity cost that is sensitive to time pressure. As preemption risk increases, the probability that a trading opportunity assessed by risk management is lost to competitors increases. Conversely, as firms give up risk management, preemption risk increases both because firms act more quickly when bypassing risk controls but also because they trade more indiscriminately, that is, any time they find a trading opportunity, rather than only when the private value α is high. When the risk management latency ι is large enough, this feedback loop supports equilibria where risk management is inefficiently low. Indeed, there is no social benefit to speed in this model, therefore running every trading opportunity through risk management maximizes total surplus. Absent agency frictions, equilibria with risk management Pareto-dominate equilibria without risk management when both exist. Furthermore, when only no risk management can be sustained in equilibrium (when $\iota > \bar{\iota}$), firms would be strictly better off if they could commit to full risk management.

Turn now to the model where firms need to incentivize traders to understand how agency frictions affect the equilibrium level of risk management relative to the benchmark in Proposition 4.2. The first observation is that the agency rent distorts firms' response to time pressure.

Lemma 4.5 (Optimal Manipulation)**.** *There exists three ordered thresholds $(\underline{p}(I), \overline{p}(I), p^*)$ where $0 \leq \underline{p}(I) \leq \overline{p}(I) \leq p^* \leq 1$, such that the best response manipulation level is:*

$$m^*(p, I) = \begin{cases} 0 & \text{if} \quad p < \underline{p}(I) \\ m^\circ(p, I) & \text{if} \quad \underline{p}(I) \leq p < \overline{p}(I) \\ m_0(p) & \text{if} \quad \overline{p}(I) \leq p < p^* \\ [m_0(p^*), 1] & \text{if} \quad p = p^* \\ 1 & \text{if} \quad p^* < p, \end{cases} \tag{4.18}$$

and is weakly increasing in p and weakly decreasing in I.

Proof. See Appendix C. ■

Time pressure now affects firms' choice of risk management through two channels. As in the benchmark without agency, time pressure increases the opportunity cost of risk management (the preemption risk). In addition, time pressure increases the cost of incentivizing risk management (the agency rent), as stated in Proposition 4.1. When time pressure is high ($p > p^*$), firms choose to forgo risk management and would do the same absent agency friction for the same reason as in the no agency case. However, in that benchmark, firms would switch to full risk management ($m^* = 0$) as p falls below p^*. This does not happen with agency frictions because an agency cost of incentivizing compliance has to be paid as soon as m falls below $m_0(p)$. As a result, when time pressure eases just below p^*, the firm implements $m_0(p)$, the highest level of risk management that can be implemented for free. Note that in the region $(\overline{p}(I), p^*)$, as preemption risk decreases, risk management improves because $m_0(p)$ decreases with p. As preemption risk diminishes further, it becomes worthwhile to offer the trader a rent. This occurs both because the opportunity cost of risk management declines with p (as in the benchmark), and because incentivizing a given level of risk management becomes cheaper as p falls. Eventually, compliance reaches its upper bound, $m^* = 0$, once preemption risk falls below $\underline{p}(I)$.

Given this discussion, we should expect the level of risk management to decrease in the model with agency relative to the no-agency benchmark, which is confirmed in our next result.

Proposition 4.3. *For any given set of parameters, the highest and lowest risk management quality that can be supported as an equilibrium outcome is weakly, and sometimes strictly, lower in the setting with agency than in the setting without agency.*

Proof. See Appendix C. ■

While the result in Proposition 4.3 – that agency frictions lead to a deterioration of risk management – is generally intuitive, this effect combines two distinct sources of inefficiencies with different implications. The first one is that agency frictions create an additional cost of risk management beyond its opportunity cost, directly depressing the extent to which firms engage in it. Although agency rents represent a private cost to the firm rather than a social cost – being mere transfers – they lead to insufficient levels of risk management, thereby generating a social inefficiency. The second, more subtle effect is a *contractual externality* operating through time pressure and its impact on agency rents. As more firms give up (incentivizing) risk management, preemption risk increases, making it more expensive for others to sustain such incentives. Unlike the direct cost of agency rents, which results from partial equilibrium frictions within a single firm, this externality emerges from strategic interactions among firms, constituting a general equilibrium phenomenon that further exacerbates inefficient coordination failures. If firms could jointly commit to stronger risk management practices, they would reduce the cost of doing so for each other. Note also that this feedback loop is distinct from the one in the benchmark without agency, as it operates through a contractual channel instead of an opportunity cost channel. Because both externalities are at play here, we focus below on a particular case that shuts down the latter.

Proposition 4.4 (Contractual Externalities)**.** *Even for $\iota = 0$, the equilibrium level of risk management is weakly lower than under full coordination, and when $\pi \in (\underline{\pi}, \overline{\pi}) \neq \varnothing$ this inequality is strict.*

Proof. See Appendix C. ■

Setting the risk management latency to zero shuts down the opportunity cost of risk management, as the risk of preemption vanishes. This implies, in particular, that in the benchmark without agency risk management is maximal ($m^\star = 0$) and any coordination issue between firms disappears. However, and perhaps surprisingly, coordination failures can still arise in the setting with agency, even when risk management latency disappears. This happens for two reasons. The first reason is that even if $\iota = 0$, risk management affects the speed at which an opportunity disappears from the market. This is because risk management not only affects how fast firms can execute (when $\iota > 0$) but also the selectivity of their trade. Absent risk management, firms execute any trade they identify, while they only act on those with a high private value ($\alpha_{kt} = \alpha_+$) when risk management is in place. This, in turn, affects the steady state stock I of latent trading opportunities that traders can discover at each instant t. The second ingredient is that the agency's cost of risk management is instead independent of I, that is, it is fixed relative to the size of the pool of trading opportunities. This is because the benefit of search scales (one for one) with I: as the number of latent opportunities grows, the rewards from search efforts become more substantial. Furthermore, the probability of detecting manipulation γ depends on the relative mass of trades that bypass risk management. This implies that the agency problem is invariant to I. Suppose now risk management is strict; then the equilibrium stock of trading opportunity is high, and so is the benefit of risk management (relative to its cost). In contrast, firms can be trapped in an efficient equilibrium where fast execution with little risk control makes the trading pool shallow, which in turn discourages risk management. Overall, Proposition 4.4 shows that agency frictions are not only an additional cost of risk management but also an additional channel through which (contractual) externalities cause coordination failures between firms.

4.5 Traders Compensation and Regulation

One motivation for introducing agency in a model of risk management is that it speaks to one observable key dimension of financial intermediaries' organization that is compensation design. Efficient risk-taking is the central issue of prudential regulation and incentive structures have attracted regulators' attention. Consider for example, the view taken by the U.S. Fed:[23]

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

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

The guidelines recommend supervising compensation practices (e.g., deferred pay and clawbacks) and associated risk control and governance processes together (36397).[24]

Our approach is consistent at a basic level with the view that incentives are essential to the implementation of risk management. Our model also links compensation structures that can be observed, the equilibrium level of risk management, and fundamental features of financial intermediaries' competitive environment. These linkages can be useful to address one frequent criticism of pay structure in financial intermediaries, namely that the "bonus culture" inherently leads to excessive risk-taking.[25] In the aftermath of the 2008 financial crisis, the European Parliament introduced a limitation of the ratio between the variable and fixed components of remuneration to 100% with the explicit objective of limiting incentives for risk-taking induced by profit-sensitive remunerations within banks.[26]

In the model, when the trader perceives a rent ($m^\star < m_0$), the binding incentive constraint is the one that governs the level of compliance. The incentive constraint that governs search and imposes that pay be sensitive to profit is then slack so that part of the trader's remuneration could be paid as a fixed wage.[27] In the optimal contract with the highest fixed component, we derive the ratio of fixed pay to total pay (i.e., fixed pay plus bonus). In practice, this measure is sometimes referred to as the "pay mix" or the "salary/incentive ratio," and is higher if the contract is lower-powered (less sensitive to profit). It takes a simple analytical form in our model:

$$\text{S/I-ratio} = 2\gamma \max\left\{m_0(p^\star) - m^\star, 0\right\}. \tag{4.19}$$

From a casual inspection of Equation (4.19), one might conclude that high risk management (low $m^\star$) is indeed associated with a high S/I-ratio (low-powered incentives). However, we show next that equilibrium effects make the mapping between incentives and risk management quality difficult to interpret.

Proposition 4.5. *In the Pareto-superior equilibrium,*

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

[25] "The bonus system, which focused on short-term profits made over the course of the year, encouraged risk taking and excessive leverage on a massive scale." Nouriel Roubini in Crisis Economics: A Crash Course in the Future of Finance (2011).

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

[27] See the proof of Proposition 4.5 in the Appendix.

- *when competition intensifies (i decreases), risk management deteriorates while the traders' expected compensation and the S/I-ratio decrease;*
- *when search becomes more effective (λ decreases), risk management deteriorates while the traders' expected compensation (per unit of search effort) and the S/I-ratio first increase and then decrease.*

Proof. See Appendix C. ∎

Going back to Equation (4.19), the reason the S/I-ratio may be difficult to interpret is because it is driven by the level of risk management $m^\star$ and by the cost of incentivizing risk management that is captured by $m_0(p^\star)$. When time pressure goes up, either because competition intensifies (fewer opportunities are born per trader) or because search becomes more efficient, time pressure goes up which creates opposing forces. Time pressure makes risk management more costly for the firm so that $m^\star$ goes up and the S/I-ratio goes down (the incentive structure looks lower-powered). Time pressure also increases the rent that needs to be paid to the trader for a given level of risk management, so that $m_0(p^\star)$ increases, driving up the S/I-ratio goes (the incentive structure looks higher-powered). Proposition 4.5 shows that the ability of the S/I-ratio to proxy, even directionally, for the level or risk management depends on the sources of time pressure. For example, when comparing two markets with different search frictions, it could be that the market with higher-powered incentives has a lower m, that is, better risk management.

Overall, Equation (4.19) suggests that properly calibrating the regulatory constraint on traders' bonuses to achieve proper risk management requires a careful evaluation of firms' trading environment. Our model supports deferred pay and clawbacks but provides a more nuanced view on "bonus cultures." In particular, it suggests that an exclusive reliance on bonuses in trader compensation need not imply weak compliance incentives, provided the bonuses are deferred and can be clawed back. One concern is that a blanket requirement to use lower-powered contracts would then impose unnecessary agency costs on those trading activities with unintended side effects (such as discouraging the activity or investment in other risk controls). Instead, our analysis suggests confining regulatory intervention to "problem areas" that display high-powered compensation in conjunction with other characteristics, such as a high degree of time pressure, reliance on soft information, slow pre-trade controls, and weak post-trade controls. In October 2023, the Bank of England and the Financial Conduct Authority announced they were scrapping the banker bonus cap introduced by the EU a decade earlier. Among the reasons mentioned by the UK regulators was that the rule made it more difficult for UK banks to compete for talent with financial intermediaries outside of Europe.[28]

Our approach also speaks to the different functions that prudential regulation can play and to the political economy of that regulation. As discussed in the previous section, agency frictions introduce two sources of inefficiency, an agency cost of regulation and (through that agency cost) a contractual externality. Depending on which of these two inefficiencies

[28]See for example "UK financial regulators scrap cap on bankers' bonuses," The Guardian, October 24, 2023.

are affected by regulatory intervention, the financial sector may profit or lose from stricter risk management requirements. To the extent that the risk management mandate reduces the externality that firms impose on each other, it can be in the collective interest of intermediaries to see that such regulations are implemented (Proposition 4.4). In that scenario the role of regulation is to help firms coordinate on a higher level of risk management when individual incentives to reach this level may not be sufficient in an unconstrained equilibrium. On the other hand, a regulator may want to push risk management up to levels where the required agency rents hurt the financial sector's profitability, even though it still creates social value, in that case captured by traders. Note that from Proposition 4.5, higher trader compensation may be required to obtain stricter risk management. We should then expect the financial sector to push against such requirements. While our model shows that risk regulation can have these conflicting effects on the profits of the financial sector, it also highlights that disentangling these forces and identifying the degree to which the interests of financial intermediaries and social welfare coincide should be challenging in practice. The model also suggests that a regulator who aligns itself with the interests of the financial industry may still want to regulate risk-taking, although to a lesser extent than if it maximizes total welfare.

4.6 Conclusion

This paper has shown that time pressure in financial markets does far more than accelerate trading—it fundamentally reshapes firms' internal governance by raising the agency cost of risk compliance. As fleeting opportunities heighten the premium on speed, traders' bonus-driven incentives clash ever more sharply with pre-trade controls. The result is a self-reinforcing cycle: greater time pressure makes it more expensive for firms to induce compliance, which in turn erodes risk governance and further intensifies the race to preempt competitors. This mechanism explains why even the most well-resourced institutions can slide into a "front office culture" that prioritizes execution over prudence. It also suggests that there may be scope for risk management regulation not warranted by the capital structure of financial firms but rather the type of markets they compete in.

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

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

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

Bibliography

Viral Acharya and Paolo Volpin. Corporate governance externalities. *Review of Finance*, 14:1–33, 2010. 80

Viral Acharya, Douglas Gale, and Tanju Yorulmazer. Rollover risk and market freezes. *Journal of Finance*, 66(4):1177–1209, 2011. 44

Viral Acharya, Robert Engle, and Diane Pierret. Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics*, 65:36–53, 2014. 41

Viral Acharya, Marco Pagano, and Paolo Volpin. Seeking alpha: Excess risk taking and competition for managerial talent. *The Review of Financial Studies*, 29(10):2565–2599, 06 2016. 80

Viral Acharya, Allen N. Berger, and Raluca A. Roman. Lending implications of us bank stress tests: Costs or benefits? *Journal of Financial Intermediation*, 34:58–90, 2018. 54

Viral Acharya, Tim Eisert, Christian Eufinger, and Christian Hirsch. Whatever it takes: The real effects of unconventional monetary policy. *Review of Financial Studies*, 32(9): 3366–3411, 2019. 42

Alessandro Acquisti, Curtis Taylor, and Liad Wagman. The economics of privacy. *Journal of Economic Literature*, 54(2):442–492, 2016. 10, 13, 36, 38

Sumit Agarwal, Shashwat Alok, Pulak Ghosh, and Sudip Gupta. Financial inclusion and alternate credit scoring: Role of big data and machine learning in fintech. Working Paper, 2019. 9, 15

Felipe Aguerrevere. Real options, product market competition, and asset returns. *Journal of Finance*, 64:957–983, 2009. 81

Kasim Ahmed and Giovanni Calice. The effects of supervisory stress testing on bank lending: examining large uk banks. *Journal of Banking Regulation*, 24(2):228, 2023. 54

Lukas Ahnert, Pascal Vogt, Volker Vonhoff, and Florian Weigert. Regulatory stress testing and bank performance. *European Financial Management*, 26(5):1449–1488, 2020. 62

S Nageeb Ali, Greg Lewis, and Shoshana Vasserman. Voluntary disclosure and personalized pricing. *Review of Economic Studies*, 90(2):538–571, 2023. 10, 12

Franklin Allen and Douglas Gale. Optimal financial crises. *Journal of Finance*, 53:1245–1284, 1998. 80

Olivier Armantier, Sebastian Doerr, Jon Frost, Andreas Fuster, and Kelly Shue. Whom do consumers trust with their data? US survey evidence. Bank for International Settlements, May 2021. 36

Philippe Askenazy, David Thesmar, and Mathis Thoenig. On the relation between organisational practices and new technologies: the role of (time based) competition. *Economic Journal*, 115:128–154, 2006. 81

Tania Babina, Saleem Bahaj, Greg Buchak, Filippo De Marco, Angus Foulis, Will Gornall, Francesco Mazzola, and Tong Yu. Customer data access and fintech entry: Early evidence from open banking. *Journal of Financial Economics*, 169:103950, 2025. 34, 36

Ian Ball. Scoring strategic agents. *American Economic Journal: Microeconomics*, 17(1): 97–129, February 2025. 10

Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace. Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56, 2022. 6

Majid Bazarbash. Fintech in financial inclusion: Machine learning applications in assessing credit risk. IMF Working Paper, 2019. 5, 9

Alessandro Beber and Marco Pagano. Short-selling bans around the world: Evidence from the 2007–09 crisis. *Journal of Finance*, 68(1):343–381, 2013. 43

Roland Benabou and Jean Tirole. Bonus culture: Competitive pay, screening, and multitasking. *Journal of Political Economy*, 2014. forthcoming. 80

Katrin Bennhold and Nicole Clark. Report says french bank trader's managers were negligent. New York Times, August 2008. 78

Tobias Berg, Valentin Burg, Ana Gombović, and Manju Puri. On the rise of fintechs: Credit scoring using digital footprints. *Review of Financial Studies*, 33(7):2845–2897, 2020. 5, 9

Dirk Bergemann and Alessandro Bonatti. Markets for information: An introduction. *Annual Review of Economics*, 11:85–107, 2019. 10

Dirk Bergemann and Stephen Morris. Information design, bayesian persuasion, and bayes correlated equilibrium. *American Economic Review*, 106(5):586–91, May 2016. 1

Dirk Bergemann and Stephen Morris. Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019. 1, 14

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. The limits of price discrimination. *American Economic Review*, 105:921–957, 2015. 28

Allen Berger, Sally Davies, and Mark J. Flannery. Comparing market and supervisory assessments of bank performance: Who knows what when? *Journal of Money, Credit and Banking*, 32(2):641–667, 2000. 41

Ben Bernanke. Stress testing banks: What have we learned? Technical report, Speech at Financial Markets Conference, Federal Reserve Bank of Atlanta, 2013. 42

David Besanko and Anjan Thakor. Collateral and rationing: Sorting equilibria in monopolistic and competitive credit markets. *International Economic Review*, 28(3):671–689, 1987. 15

Helmut Bester. Screening vs. rationing in credit markets with imperfect information. *American Economic Review*, 75(4):850–855, 1985. 15

Sudipto Bhattacharya. Aspects of monetary and banking theory and moral hazard. *Journal of Finance*, 37(2):371–384, 1982. 48

Bruno Biais, Thierry Foucault, and Sophie Moinas. Equilibrium fast trading. *Journal of Financial Economics*, 116(2):292–313, 2015. 81

Bo Bian, Xinchen Ma, and Huang Tang. The supply and demand for data privacy: Evidence from mobile apps. Working Paper, 2021. 35

Ken Binmore. Perfect equilibria in bargaining models. In *The Economics of Bargaining*. Basil Blackwell, Oxford, 1987. 13

Joshua E. Björkegren, Danieland Blumenstock and Samsun Knight. Manipulation-proof machine learning. Working Paper, 2021. 10

David Blackwell. Comparison of experiments. In J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, 1951. 27

Laura Blattner, Scott Nelson, and Jann Spiess. Unpacking the black box: Regulating algorithmic decisions. Working Paper, 2024. 11

Kristian Blickle, Zhiguo He, Jing Huang, and Cecilia Parlatore. Information-based pricing in specialized lending. National Bureau of Economic Research, February 2024. 10, 14

Robert Bliss and Mark Flannery. Market discipline in the governance of us bank holding companies: Monitoring vs. influencing. *Review of Finance*, 6(3):361–396, 2002. 41

Rishi Bommasani, Kevin Klyman, Sayash Kapoor, Shayne Longpre, Betty Xiong, Nestor Maslej, and Percy Liang. The foundation model transparency index. Working Paper, 2023. 36

Philip Bond and Itay Goldstein. Government intervention and information aggregation by prices. *Journal of Finance*, 70(6):2777–2812, 2015. 45, 46, 72, 73

Philip Bond and Yao Zeng. Silence is safest: Information disclosure when the audience's preferences are uncertain. *Journal of Financial Economics*, 145(1):178–193, 2022. 10

Philip Bond, Itay Goldstein, and Edward Simpson Prescott. Market-based corrective actions. *The Review of Financial Studies*, 23(2):781–820, 02 2010. 45

Philip Bond, Alex Edmans, and Itay Goldstein. The Real Effects of Financial Markets. *Annual Review of Financial Economics*, 4(1):339–360, October 2012a. 45

Philip Bond, Alex Edmans, and Itay Goldstein. The real effects of financial markets. *Annual Review of Financial Economics*, 4:339–360, 2012b. 81

Matthieu Bouvard and Samuel Lee. Risk management failures. *The Review of Financial Studies*, 33(6):2468–2505, 06 2020. 77, 79, 94

Matthieu Bouvard, Pierre Chaigneau, and Adolfo de Motta. Transparency in the financial system: Rollover risk and crises. *Journal of Finance*, 70(4):1805–1837, 2015. 43, 47, 48, 59, 63, 65, 70

John Boyd and Gianni De Nicolò. The theory of bank risk taking and competition revisited. *Journal of Finance*, 60:1329–1343, 2005. 80

Thorsten Broecker. Credit-worthiness tests and interbank competition. *Econometrica*, 58 (2):429–452, 1990. 14

Markus Brunnermeier and Martin Oehmke. Predatory short-selling. *Review of Finance*, 18(6):2153–2195, 2014. 44

John Bryant. A model of reserves, bank runs and deposit insurance. *Journal of Banking and Finance*, 4:335–344, 1980. 79

Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 07 2015. 81

Mike Burkart, Denis Gromb, and Fausto Panunzi. Large shareholders, monitoring, and the value of the firm. *Quarterly Journal of Economics*, 112:693–728, 1997. 80

Ricardo Caballero and Robert Pindyck. Uncertainty, investment, and industry evolution. *International Economic Review*, 37:641–662, 1996. 81

Giacomo Calzolari and Gyöngyi Lóránth. Regulation of multinational banks: A theoretical inquiry. *Journal of Financial Intermediation*, 20(2):178–198, 2011. 48

Giacomo Calzolari, Jean-Edouard Colliard, and Gyöngyi Lóránth. Multinational banks and supranational supervision. *Review of Financial Studies*, 32(8):2997–3035, 2019. 48

Susan Carleton Athey, Kevin Bryan, and Joshua S. Gans. The allocation of decision authority to human and artificial intelligence. *AEA Papers and Proceedings*, 110:80–84, 2020. 11

Elena Carletti, Giovanni Dell'Ariccia, and Robert Marquez. Supervisory incentives in a banking union. *Management Science*, 67(1):455–470, 2021. 41, 48, 52

V.V. Chari and Ravi Jagannathan. Banking panics, information, and rational expectations equilibrium. *Journal of Finance*, 43:749–761, 1988. 80

Yehning Chen. Banking panics: The role of the first-come, first-served rule and information externalities. *Journal of Political Economy*, 107:946–968, 1999. 80

Ing-Haw Cheng. Corporate governance spillovers. Working Paper, 2011. URL http://ssrn.com/abstract=1299652. 80

Jonathan B Cohn, Uday Rajan, and Günter Strobl. Credit ratings: strategic issuer disclosure and optimal screening. *Review of Finance*, 29(1):169–199, 09 2024. 10

Jean-Edouard Colliard. Strategic selection of risk models and bank capital regulation. *Management Science*, 67(6):2591–2606, 2019. 41

Bo Cowgill and Catherine Tucker. Algorithmic fairness and economics. Columbia Business School Research Paper, 2020. 11

Jesse Davis and Naveen Gondhi. Learning in financial markets: Implications for debt equity conflicts. *Review of Financial Studies*, 37(5):1584–1639, 2024. 47

Giovanni Dell'Ariccia and Robert Marquez. Information and bank credit allocation. *Journal of Financial Economics*, 72:185–214, 2004. 80

Ken Deng and Joel Shapiro. Feedback between the financial market and the product market. Working Paper, 2024. 46

Douglas Diamond. Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51:393–414, 1984. 80

Douglas Diamond and Philip Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91:401–419, 1983. 80

David L. Dicks. Executive compensation and the role for corporate governance regulation. *Review of Financial Studies*, 25:1971–2004, 2012. 80

Will Dobbie, Paul Goldsmith-Pinkham, Neale Mahoney, and Jae Song. Bad credit, no problem? credit and labor market consequences of bad credit reports. *The Journal of Finance*, 75(5):2377–2419, 2020. 34

Sebastian Doerr, Leonardo Gambacorta, Luigi Guiso, and Marina Sanchez del Villar. Privacy regulation and fintech lending. BIS Working Paper, 2023. 34, 36

James Dow, Itay Goldstein, and Alexander Guembel. Incentives for information production in markets where prices affect real investment. *Journal of the European Economic Association*, 15(4):877–909, 2017. 46

Jérôme Dugast and Thierry Foucault. Data abundance and asset price informativeness. *Journal of Financial Economics*, 130(2):367–391, 2018. 81

Piotr Dworczak and Giorgio Martini. The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048, 2019. 14

Ronald A. Dye. Disclosure of nonproprietary information. *Journal of Accounting Research*, 23(1):123–145, 1985. 10, 13

David Easley, Nicolas Kiefer, and Maureen O'Hara. The information content of the trading process. *Journal of Empirical Finance*, 4(2-3):159–186, 1997. 61

Florian Ederer, Richard Holden, and Margaret Meyer. Gaming and strategic opacity in incentive provision. *RAND Journal of Economics*, 49(4):819–854, 2018. 11

Mehmet Ekmekci, Leandro Gorno, Lucas Maestri, Jian Sun, and Dong Wei. Learning from manipulable signals. *American Economic Review*, 112(12):3995–4040, 2022. 10

Miguel Faria-e Castro, Joseba Martinez, and Thomas Philippon. Runs versus lemons: Information disclosure and fiscal capacity. *Review of Economic Studies*, 84(4):1683–1707, 2017. 47, 69, 70

Falko Fecht, Roman Inderst, and Sebastian Pfeil. A theory of the boundaries of banks with implications for financial integration and regulation. CEPR Discussion Paper, 2022. 48

Michael J. Fishman and Kathleen M. Hagerty. Mandatory versus voluntary disclosure in markets with informed and uninformed customers. *The Journal of Law, Economics, and Organization*, 19(1):45–63, 2003. 10

Mark J. Flannery. Maintaining adequate bank capital. *Journal of Money, Credit and Banking*, 46(s1):157–180, 2014. 41

Mark J. Flannery and Robert R. Bliss. Market discipline in regulation: Pre- and post-crisis. In *Oxford Handbook of Banking*, chapter 23, pages 736–775. Oxford University Press, 2019. 41, 42, 52

Mark J. Flannery, Beverly Hirtle, and Anna Kovner. Evaluating the information in the federal reserve stress tests. *Journal of Financial Intermediation*, 29:1–18, 2017. 61, 62

Alex Frankel and Navin Kartik. Muddled information. *Journal of Political Economy*, 127 (4):1739–1776, 2019. 10

Alex Frankel and Navin Kartik. Improving information from manipulable data. *Journal of the European Economic Association*, 20(1):79–115, 2021. 10

Xavier Freixas and Jean Jacques Laffont. Optimal banking contracts. In *Essays in Honor of Edmond Malinvaud*, volume 2, pages 33–61. MIT Press, Cambridge, MA, 1990. 15

Kenneth A. Froot, David S. Scharfstein, and Jeremy C. Stein. Risk management: Coordinating corporate investment and financing policies. *Journal of Finance*, 48:1629–1658, 1993. 79, 82

William Fuchs, Satoshi Fukuda, and Daniel Neuhann. Rules versus disclosure: Prudential regulation and market discipline. Working Paper, 2024. 48, 73

Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *Journal of Finance*, 77:5–47, 2022. 9, 18

Andrea Gamba and Christopher Hennessy. Manipulable data, goodhart's law, and credit risk prediction. Working Paper, 2024. 10

Leonardo Gambacorta, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen. Data versus collateral. *Review of Finance*, 27(2):369–398, 04 2022. 9, 15

Pingyang Gao, Xu Jiang, and Jinzhi Lu. Manipulation, panic runs, and the short selling ban. Working Paper, 2024. 44

Robert H. Gertner, David S. Scharfstein, and Jeremy C. Stein. Internal versus external capital markets. *Quarterly Journal of Economics*, 109:1211–1230, 1994. 80

Ella Glikson and Anita Williams Woolley. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2):627–660, 2020. 6

Limor Golan, Christine A. Parlour, and Uday Rajan. Competition, managerial slack, and corporate governance. *Review of Corporate Finance Studies*, 2014. forthcoming. 80

Avi Goldfarb and Catherine Tucker. Shifts in privacy concerns. *American Economic Review: Papers & Proceedings*, 102(3):349–353, 2012. 13

Itay Goldstein. Information in financial markets and its real effects. *Review of Finance*, 27(1):1–32, 2023. 42, 45

Itay Goldstein and Alexander Guembel. Manipulation and the allocational role of prices. *Review of Economic Studies*, 75(1):133–164, 2008. 44

Itay Goldstein and Chong Huang. Credit rating inflation and firms' investments. *Journal of Finance*, 75(6):2929–2972, 2020. 49

Itay Goldstein and Yaron Leitner. Stress tests and information disclosure. *Journal of Economic Theory*, 177:34–69, 2018. 48, 59

Itay Goldstein and Haresh Sapra. Should banks' stress test results be disclosed? an analysis of the costs and benefits. *Foundations and Trends in Finance*, 8(1):1–54, 2014. 47

Itay Goldstein and Liyan Yang. Information disclosure in financial markets. *Annual Review of Financial Economics*, 9:101–125, 2017. 47, 58

Itay Goldstein and Liyan Yang. Good disclosure, bad disclosure. *Journal of Financial Economics*, 131:118–138, 2019. 45, 46

Gary Gorton and Lixin Huang. Liquidity, efficiency, and bank bailouts. *American Economic Review*, 94(3):455–483, 2004. 48

Gary Gorton and Andrew Winton. Financial intermediation. In George M. Constantinides, Milton Harris, and René M. Stulz, editors, *Corporate Finance*, volume 1 of *Handbook of the Economics of Finance*, pages 431–552. Elsevier, 2003. 80

Steven Grenadier. Option exercise games: An application to the equilibrium investment strategies of firms. *Review of Financial Studies*, 15:691–721, 2002. 81

Sanford J. Grossman. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24(3):461–483, 1981. 10, 12, 20, 35

Nicolae Gârleanu and Lasse Heje Pedersen. Liquidity and risk management. *American Economic Review*, 97:193–197, 2007. 81

Andrew Haldane. Capital discipline. Technical report, Speech to the American Economic Association, 2011. 41

Richmond Harbaugh and Theodore To. False modesty: When disclosing good news looks bad. *Journal of Mathematical Economics*, 87:43–55, 2020. 10

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wooters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 111–122. Association for Computing Machinery, 2016. 10

Zhiguo He, Jing Huang, and Jidong Zhou. Open banking: Credit market competition when borrowers own the data. *Journal of Financial Economics*, 147(2):449–474, 2023. 10, 12

Zhiguo He, Jing Huang, and Cecilia Parlatore. Information span in credit market competition. Working Paper, 2024. 10

Florian Heider and Roman Inderst. Loan prospecting. *Review of Financial Studies*, 25: 2381–2415, 2012. 80

Amanda Heitz and Barrett Wheeler. Bank stress test disclosures, private information production, and price informativeness. *Private Information Production, and Price Informativeness*, 2023. (June 7, 2023). 45

Thomas Hellmann, Kevin Murdock, and Joseph Stiglitz. Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 90(1):147–165, 2000a. 48

Thomas Hellmann, Kevin Murdock, and Joseph Stiglitz. Liberalization, moral hazard in banking, and prudential regulation: Are capital requirements enough? *American Economic Review*, 90:147–165, 2000b. 80

Benjamin Hermalin. The effects of competition on executive behavior. *Rand Journal of Economics*, 23:350–365, 1992. 80

Bengt Holmström and Paul Milgrom. Multitask principal–agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7(1):24–52, 1991. doi: 10.1093/jleo/7.1.24. 92

Bengt Holmström and Jean Tirole. Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics*, 112(3):663–691, 1997a. 11

Bengt Holmström and Jean Tirole. Financial intermediation, loanable funds and the real sector. *Quarterly Journal of Economics*, 112:663–691, 1997b. 80

Jing Huang. Fintech expansion. Working Paper, 2023. 10

Nicolas Inostroza. Persuading multiple audiences: Strategic complementarities and (robust) regulatory disclosures. Rotman School of Management Working Paper, 2023. 48, 73

Nicolas Inostroza and Alessandro Pavan. Adversarial coordination and public information design. *Theoretical Economics*, 2023. forthcoming. 47

Rajkamal Iyer, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue. Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577, 2016. 5, 9

Charles J. Jacklin and Sudipto Bhattacharya. Distinguishing panics and information-based bank runs: Welfare and policy implications. *Journal of Political Economy*, 96:568–592, 1988. 80

Mark Jansen, Fabian Nagel, Constantine Yannelis, and Anthony Lee Zhang. Data and welfare in credit markets. Working Paper, 2024. 34

Travis Johnson and Eric So. A simple multimarket measure of information asymmetry. *Management Science*, 64(3):1055–1080, 2018. 61

Woon-Oh Jung and Young K. Kwon. Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research*, 26(1):146–153, 1988. 10, 13

Emir Kamenica and Matthew Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011. 14, 47

Michael C. Keeley. Deposit insurance, risk, and market power in banking. *American Economic Review*, 80:1183–1200, 1990. 80

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. *AEA Papers and Proceedings*, 108:22–27, 2018. 11

Anton Korinek and Avital Balwit. Aligned with whom? direct and social goals for ai systems. NBER Working Paper, 2022. 11

Simon Kwan. The promise and limits of market discipline in banking. Federal Reserve Bank of San Francisco Economic Letter, 2002. 41, 42

Albert Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985. 51

Augustin Landier, David Sraer, and David Thesmar. Financial risk management: When does independence fail? *American Economic Review*, 99:454–458, 2013. 81

John Leahy. Investment in competitive equilibrium: The optimality of myopic behavior. *Quarterly Journal of Economics*, 108:1105–1133, 1993. 81

Yaron Leitner and Basil Williams. Model secrecy and stress tests. *Journal of Finance*, 78 (2):1055–1095, 2023. 11, 47

Yaron Leitner and Bilge Yilmaz. Regulating a model. *Journal of Financial Economics*, 131(2):251–268, 2019. 50

Stephen Lenkey and Fenghua Song. Feedback, investment, and social value of financial expertise. Working Paper, 2017. 47

Tracy Lewis and David Sappington. Supplying information to facilitate price discrimination. *International Economic Review*, 35(2):309–327, 1994. 12

Annie Liang, Jay Lu, Xiaosheng Mu, and Kyohei Okumura. Algorithm design: A fairness-accuracy frontier. *Journal of Political Economy*, 2024. forthcoming. 11

Andres Liberman, Christopher Neilson, Luis Opazo, and Seth Zimmerman. The equilibrium effects of information deletion: Evidence from consumer credit markets. Chicago Booth Research Paper #19-15; Fama-Miller Working Paper, 2019. 34

Marvin B. Lieberman and David B. Montgomery. First-mover advantages. *Strategic Management Journal*, 9(Special Issue: Strategy Content Research):41–58, 1988. 77

Tesary Lin. Valuing intrinsic and instrumental preferences for privacy. *Marketing Science*, 41(4), 2022. 13, 34

Stephen Martin. Endogenous firm efficiency in a cournot principal-agent model. *Journal of Economic Theory*, 59:445–450, 1993. 80

David Martinez-Miera and Rafael Repullo. Does competition reduce the risk of bank failure? *Review of Financial Studies*, 23:3638–3664, 2010. 80

Natalya Martynova, Enrico Perotti, and Javier Suarez. Capital forbearance in the bank recovery and resolution game. *Journal of Financial Economics*, 146(3):884–904, 2022. 69

Patrick McConnell. Dissecting the jpmorgan whale: A post-mortem. *The Journal of Operational Risk*, 9(2):59, 2014. 76

Paul Milgrom. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics*, 12(2):380–391, 1981. 10, 12, 20, 35

Christoph Molnar. *Interpretable Machine Learning: a Guide for Making Black Box Models Explainable*. Leanpub, 2nd edition, 2023. 16

Timothy Morey, Theodore "Theo" Forbath, and Allison Schoop. Customer data: Designing for transparency and trust. *Harvard Business Review*, 93(5):96–105, 2015. 13, 36

Donald P. Morgan, Stavros Peristiani, and Vanessa Savino. The information value of the stress test. *Journal of Money, Credit and Banking*, 46(7):1479–1500, 2014. 62

Alan D. Morrison and Lucy White. Crises and capital requirements in banking. *American Economic Review*, 95(5):1548–1572, 2005. 48

Barry J. Nalebuff and Joseph Stiglitz. Information, competition, and markets. *American Economic Review*, 73:278–283, 1983. 80

Rachel J. Nam. Open banking and customer data sharing: Implications for fintech borrowers. Working Paper, 2024. 15, 34, 35, 36, 38

Scott Nelson. Private information and price regulation in the us credit card market. *Econometrica*, 2023. forthcoming. 34

Melchisedek Joslem Ngambou Djatche. Monetary policy, prudential policy and bank's risk-taking: A literature review. *Journal of Economic Surveys*, 36(5):1559–1590, 2022. 41

Dmitry Orlov, Pavel Zryumov, and Andrzej Skrzypacz. Design of macro-prudential stress tests. *Review of Financial Studies*, 36(11):4460–4501, 2023. 48, 65, 73

Emiliano S. Pagnotta and Thomas Philippon. Competing on speed. *Econometrica*, 86(3):1067–1115, 2018. 81

Markus Parlasca. Stress tests by an informed regulator. Working Paper, 2024. 48, 49, 59, 70

Cecilia Parlatore and Thomas Philippon. Designing stress scenarios. NBER Working Paper, 2022. 48, 49, 50

Christine A. Parlour and Uday Rajan. Competition in loan contracts. *American Economic Review*, 91:1311–1328, 2001. 80

Frank Pasquale. *The Black Box Society: the Secret Algorithms that Control Money and Information*. Harvard University Press, 2015. 6

Jonathan Paul. On the efficiency of stock-based compensation. *Review of Financial Studies*, 5(3):471–502, 1992. 47

James Peck and Karl Shell. Equilibrium bank runs. *Journal of Political Economy*, 111: 103–123, 2003. 80

Eduardo Perez-Richet and Vasiliki Skreta. Test design under falsification. *Econometrica*, 90(3):1109–1142, 2022. 10

Giovanni Petrella and Andrea Resti. Supervisors as information producers: Do stress tests reduce bank opaqueness? *Journal of Banking and Finance*, 37:5406–5420, 2013. 62

Thomas Philippon and Philip Schnabl. Efficient recapitalization. *Journal of Finance*, 68 (1):1–42, 2013. 48, 62

Alessio Piccolo and Joel Shapiro. Credit ratings and market information. *Review of Financial Studies*, 35(10):4425–4473, 2022. 49

Andrew Postlewaite and Xavier Vives. Bank runs as an equilibrium phenomenon. *The Journal of Political Economy*, 95:485–491, 1987. 80

Michael Power, Simon Ashby, and Tommaso Palermo. *Risk culture in financial organisations: A research report*. CARR-Analysis of Risk and Regulation, 2013. 77

Jeffrey T. Prince and Scott Wallsten. How much is privacy worth around the world and across platforms? *Journal of Economics & Management Strategy*, 31(4):841–861, 2022. 36

Daniel Quigley and Ansgar Walther. Inside and outside information. *Journal of Finance*, 79(4):2667–2714, 2023. 47

Daniel Quigley and Ansgar Walther. Inside and outside information. *Journal of Finance*, 79(4):2667–2714, 2024. 10

Konrad Raff. Information externalities in corporate governance. Working Paper, 2011. URL http://ssrn.com/abstract=1787130. 80

Micahel Raith. Competition, risk, and managerial incentives. *American Economic Review*, 93:1425–1436, 2003. 80

Raghuram G. Rajan. Insiders and outsiders: the choice between informed and arm's-length debt. *Journal of Finance*, 47(4):1367–1400, 1992. 13

Tarun Ramadorai, Antoine Uettwiller, and Ansgar Walther. Privacy policies and consumer data extraction: Evidence from u.s. firms. Working Paper, 2021. 35

Ashesh Rambachan, Jon Kleinberg, Sendhil Mullainathan, and Jens Ludwig. An economic approach to regulating algorithms. NBER Working Paper Series, 2021. 11

Adriano Rampini and S. Viswanathan. Collateral, risk management, and the distribution of debt capacity. *Journal of Finance*, 65:2293–2322, 2010. 79, 82

Kumar Rishabh. Beyond the bureau: Interoperable payment data for loan screening and monitoring. Working Paper, 2024. 34

Jean-Charles Rochet. Capital requirements and the behaviour of commercial banks. *European Economic Review*, 36(5):1137–1170, 1992. 48

Ariel Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50:97–110, 1985. 13

Martin Ruckes. Bank competition and credit standards. *Review of Financial Studies*, 17: 1073–1102, 2004. 80

David Scharfstein. Product-market competition and managerial slack. *Rand Journal of Economics*, 19:147–155, 1988. 80

Klaus M. Schmidt. Managerial incentives and product market competition. *Review of Economic Studies*, 64:191–213, 1997. 80

Joel Shapiro and Jing Zeng. Stress testing and bank lending. *Review of Financial Studies*, 37(4):1265–1314, 2024. 48, 59, 69

Steven A. Sharpe. Asymmetric information, bank lending, and implicit contracts: a stylized model of customer relationships. *Journal of Finance*, 45(4):1069–1087, 1990. 13

Lin Shen. Capital flows in the financial system and supply of credit. Working Paper, 2021. 49

Christoph Siemroth. The informational content of prices when policy makers react to financial markets. *Journal of Economic Theory*, 179:240–274, 2019. 45, 46, 73

Clifford W. Smith and René M. Stulz. The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis*, 20:391–406, 1985. 79

Gary Stern. Taking market data seriously. Minneapolis Fed, 2001. 42

René M. Stulz. Optimal hedging policies. *Journal of Financial and Quantitative Analysis*, 19:127–140, 1984. 79

Jian Sun. Algorithmic transparency. Working Paper, 2021. 11

Huan Tang. The value of privacy: Evidence from online borrowers. Working Paper, 2019. 34, 36

Spyros Terovitis. Information disclosure and the feedback effect in capital markets. *Journal of Financial Intermediation*, 49:100897, 2020. 49

Jamila Trindle. A whale of a trade: one man to check the books. Wall Street Journal, September 2013. 85

Robert E. Verrecchia. Discretionary disclosure. *Journal of Accounting and Economics*, 5 (1):179–194, 1983. 10

Xavier Vives. Digital disruption in banking. *Annual Review of Financial Economics*, 11: 243–272, 2019. 9

Qiaochu Wang, Yan Huang, Stefanus Jasin, and Param Vir Singh. Algorithmic transparency with strategic users. *Management Science*, 69(4):1935–2545, 2023. 11

Phoebe White and Tanju Yorulmazer. Bank resolution concepts, trade-offs, and changes in practices. *Economic Policy Review*, pages 153–173, December 2014. 69

Basil Williams. Stress tests and bank portfolio choice. Working Paper, 2017. 43, 47, 59, 65

APPENDICES

Chapter A

Appendix for Chapter 2

A.1 Proofs

Proof of Lemma 2.1. We proceed in 2 steps: Step 1 derives the lender's estimate of the borrower's quality given data z, while Step 2 finds the credit allocation given the lender's estimate.

Step 1. Given the DGP in Equation (2.1), the conditional pdf is

$$f_\lambda(\theta|z) = \begin{cases} \lambda\,\delta(\theta - z) + (1-\lambda) & \text{if } \lambda \geq 0 \\ |\lambda|\,\delta(1-\theta-z) + (1-|\lambda|) & \text{if } \lambda < 0, \end{cases}$$

where

$$\delta(t-z) \triangleq \begin{cases} \infty & \text{if } t - z = 0 \\ 0 & \text{if } t - z \neq 0, \end{cases} \tag{A.1}$$

is the Dirac's delta function, satisfying the following property:

$$\int_{s-a}^{s+a} f(t)\,\delta(t-z)\,\mathrm{d}t = f(z) \quad \forall a > 0. \tag{A.2}$$

When $\lambda \geq 0$ we have

$$\begin{aligned} \mathbb{E}(\theta|z,\lambda) &= \int_0^1 \theta f_\lambda(\theta|z)\,\mathrm{d}\theta \\ &= \lambda \int_0^1 \theta\,\delta(\theta - z)\,\mathrm{d}\theta + (1-\lambda)\int_0^1 \theta\,\mathrm{d}\theta \\ &= \lambda z + (1-\lambda)\,\tfrac{1}{2}. \end{aligned}$$

When $\lambda < 0$, performing similar computations, we get

$$\begin{aligned}\mathbb{E}(\theta|z,\lambda) &= |\lambda|(1-z) + (1-|\lambda|)\,\tfrac{1}{2}\\ &= -\lambda\,(1-z) + (1+\lambda)\,\tfrac{1}{2}\\ &= \lambda\,z + (1-\lambda)\,\tfrac{1}{2}.\end{aligned}$$

Step 2. The first order conditions of the problem in the first line of Equation (2.3) are

$$\phi\Big[\mu_\lambda(z)x-1\Big]^{\phi-1}\mu_\lambda(z)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{1-\phi}-\Big[\mu_\lambda(z)x-1\Big]^{\phi}(1-\phi)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{-\phi}\mu_\lambda(z)=0.$$

Moving the second addend on the right-hand side and dividing both sides by

$$(1-\phi)\Big[\mu_\lambda(z)x-1\Big]^{\phi-1}\mu_\lambda(z)\Big[\mu_\lambda(z)\big(X-x\big)\Big]^{1-\phi}$$

we get

$$\frac{\phi}{1-\phi}=\frac{\mu_\lambda(z)x-1}{\mu_\lambda(z)\big(X-x\big)},$$

or equivalently

$$\phi=\frac{\mu_\lambda(z)x-1}{\mu_\lambda(z)X-1}.$$

Solving for x we get the interest rate

$$x_\lambda(z)=\frac{1}{\mu_\lambda(z)}+\phi\left(X-\frac{1}{\mu_\lambda(z)}\right).$$

The lender provides credit only if her expected profits are strictly positive, i.e. only if

$$\mu_\lambda(z)\,x_\lambda(z)-1=\phi\Big(\mu_\lambda(z)X-1\Big)>0,$$

so the credit rationing rule is

$$\ell_\lambda(z)=\mathbb{1}\left\{\mu_\lambda(z)>\frac{1}{X}\right\}.$$

For future reference, I compute the lender's expected profits, the borrower's expected profits and expected social surplus (plugging the allocation in Equation (2.4) and Equation (2.5) and taking expectation)

$$\begin{aligned}V^L_\lambda(z)&=\ell_\lambda(z)\Big[\phi\Big(\mu_\lambda(z)X-1\Big)\Big],\\ V^B_\lambda(z)&=\ell_\lambda(z)\Big[(1-\phi)\Big(\mu_\lambda(z)X-1\Big)+b\Big],\\ W_\lambda(z)&=\ell_\lambda(z)\Big[\mu_\lambda(z)X-1\Big].\end{aligned}\tag{A.3}$$

■

Lemma A.1 (Credit Allocation following Data Withholding)**.** *Let* $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ *be the set of borrowers that withhold data, the credit allocation algorithm is*

$$\begin{aligned} x_\lambda(\varnothing) &= \frac{1}{\mu_\lambda(\varnothing)} + \phi\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right), \\ \ell_\lambda(\varnothing) &= \mathbb{1}\left\{\mu_\lambda(\varnothing) > \frac{1}{X}\right\}, \end{aligned} \tag{A.4}$$

where

$$\begin{aligned} \mu_\lambda(\varnothing) &\triangleq \lambda\, z(\pi, Q) + (1-\lambda)\tfrac{1}{2}, \\ z(\pi, Q) &\triangleq \omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q), \\ \omega(\pi, Q) &\triangleq \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)}. \end{aligned} \tag{A.5}$$

Proof of Lemma A.1. We proceed in 2 steps: Step 1 derives the lender's inference of the borrower's quality following data-withholding, Step 2 finds the optimal credit allocation.

Step 1. Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data. Consider $\lambda \geq 0$, and let

$$\omega(\pi, Q) \triangleq \frac{\pi}{\pi + (1-\pi)\Pr(z \in Q)},$$

be the posterior probability that data-withholding is non-strategic. The lender's posterior beliefs over θ given $m = \varnothing$ are:

$$f_\lambda(\theta|\varnothing) = \omega(\pi, Q) f(\theta) + \big(1 - \omega(\pi, Q)\big) f_\lambda(\theta|z \in Q)$$

where $f(\theta) = 1$ is the prior distribution of θ, while

$$\begin{aligned} f_\lambda(\theta|z \in Q) &= \frac{\Pr\big(z \in Q|\theta\big) f(\theta)}{\Pr\big(z \in Q\big)}, \\ &= \frac{\displaystyle\int_Q \lambda\delta(\theta - z) + (1-\lambda)\ \mathrm{d}z}{\Pr\big(z \in Q\big)}, \\ &= \lambda \frac{\mathbb{1}\{\theta \in Q\}}{\Pr(\theta \in Q)} + (1-\lambda) \end{aligned}$$

is the posterior distribution given that data withholding is strategic ($\delta(\theta - z)$ is the Dirac's delta function defined in Equation (A.1) and satisfying property in Equation (A.2)). The expected θ given data withholding is

$$\begin{aligned} \mathbb{E}_\lambda(\theta|\varnothing) =& \int_0^1 \theta f_\lambda(\theta|\varnothing)\mathrm{d}\theta, \\ =& \,\omega(\pi, Q)\int_0^1 \theta\,\mathrm{d}\theta + \big(1 - \omega(\pi, Q)\big)\left[\lambda \int_Q \theta \frac{1}{\Pr(\theta \in Q)}\,\mathrm{d}\theta + (1-\lambda)\int_0^1 \theta\,\mathrm{d}\theta\right], \\ =& \,\omega(\pi, Q)\mathbb{E}(\theta) + \big(1 - \omega(\pi, Q)\big)\Big[\lambda\,\mathbb{E}(\theta|\theta \in Q) + (1-\lambda)\,\mathbb{E}(\theta)\Big], \\ =& \,\lambda\Big[\omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q)\Big] + (1-\lambda)\tfrac{1}{2} \triangleq \mu_\lambda(\varnothing). \end{aligned}$$

Consider now $\lambda < 0$. For some $Q \subseteq [0,1]$ and some $y \in Q$, let $Q_r \triangleq \{1-y|y \in Q\}$ be the reflection of Q over the axis $y = \frac{1}{2}$. The lender's posterior beliefs over θ given $m = \varnothing$ are:

$$f_\lambda(\theta|\varnothing) = \omega(\pi, Q) f(\theta) + \big(1 - \omega(\pi, Q)\big) f_\lambda(\theta|z \in Q)$$

where

$$\begin{aligned} f_\lambda(\theta|z \in Q) &= |\lambda| \frac{\mathbb{1}\{1-\theta \in Q\}}{\Pr(z \in Q)} + (1 - |\lambda|), \\ &= |\lambda| \frac{\mathbb{1}\{\theta \in Q_r\}}{\Pr(\theta \in Q_r)} + (1 - |\lambda|), \end{aligned}$$

so that the expected θ given data withholding is

$$\begin{aligned} \mathbb{E}(\theta|\varnothing) =&\ |\lambda| \Big[\omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q_r)\Big] + (1 - |\lambda|)\tfrac{1}{2}, \\ =& -\lambda \Big[\omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\big(1 - \mathbb{E}(\theta|\theta \in Q)\big)\Big] + (1 + \lambda)\tfrac{1}{2}, \\ =&\ \lambda \Big[\omega(\pi, Q)\tfrac{1}{2} + \big(1 - \omega(\pi, Q)\big)\mathbb{E}(\theta|\theta \in Q)\Big] + (1 - \lambda)\tfrac{1}{2} = \mu_\lambda(\varnothing). \end{aligned}$$

Step 2. Suppose that after data withholding the lender provides credit, the interest rate solves:

$$\max_{x \in \mathbb{R}} \quad \Big[\mu_\lambda(\varnothing)\, x - 1\Big]^{\phi} \Big[\mu_\lambda(\varnothing)\big(X - x\big)\Big]^{1-\phi}.$$

Proceeding as in the proof of Lemma 2.1 we get the optimal interest rate:

$$x_\lambda(\varnothing) \triangleq \frac{1}{\mu_\lambda(\varnothing)} + \phi\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right).$$

The lender provides credit only if her expected profits are strictly positive, i.e. only if:

$$\ell_\lambda(\varnothing) = \mathbb{1}\left\{\mu_\lambda(\varnothing) > \frac{1}{X}\right\}.$$

For future reference, I compute the lender's and borrower's expected profits and expected social welfare, conditioning on $m = \varnothing$,

$$\begin{aligned} V_\lambda^L(\varnothing) &= \ell_\lambda(\varnothing)\Big[\phi\Big(\mu_\lambda(\varnothing)X - 1\Big)\Big], \\ V_\lambda^B(\varnothing) &= \ell_\lambda(\varnothing)\Big[(1-\phi)\Big(\mu_\lambda(\varnothing)X - 1\Big) + b\Big], \\ W_\lambda(\varnothing) &= \ell_\lambda(\varnothing)\Big[\mu_\lambda(\varnothing)X - 1\Big], \end{aligned} \tag{A.6}$$

as well as the borrower's expected profits conditional on z

$$V_\lambda^B(\varnothing, z) = \ell_\lambda(\varnothing)\left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]. \tag{A.7}$$

■

Proof of Proposition 2.1. Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data, the borrower's utility from data sharing and withholding are respectively $V_\lambda^B(z)$ and $V_\lambda^B(\varnothing, z)$ from Equation (A.3) and Equation (A.7).

Consider first the case where $\lambda = 0$ and the allocation rule is independent of data. No matter the data sharing strategy of the borrower there is credit provision at a flat interest rate, so that borrowers are indifferent between data sharing and withholding. Assuming that indifferent borrowers withholds data, the set of borrowers that withhold data is $Q = [0, 1]$.

Consider now $\lambda > 0$ and suppose that the set of borrowers that withhold data is $Q = [0, q]$ with $q \in (0, 1)$. From Lemma A.1 we can compute the lender's estimate of the borrower's quality after data-withholding, $\mu_\lambda(z(\pi, q))$, where

$$z(\pi, q) = \frac{\pi}{\pi + (1-\pi)q}\frac{1}{2} + \left(1 - \frac{\pi}{\pi + (1-\pi)q}\right)\frac{q}{2},$$

as well as the optimal credit allocation. We want to find the borrower q that is indifferent between sharing and withholding data. Suppose that q is such that the lender provides credit both after data sharing and data withholding. If $\lambda \in (0, \overline{\lambda})$ with $\overline{\lambda} = 1 - 2/X$, this is the case for every z and $z(\pi, q)$ (see Lemma 1 and the discussion that follows). If $\lambda \in [\overline{\lambda}, 1]$, this requires that $q > r(\lambda)$ and $z(\pi, q) > r(\lambda)$ (see Lemma 1 and Lemma A.1) where $r(\lambda)$ is defined in Equation (2.8). Suppose that these conditions are satisfied, the indifferent borrower has to get the same surplus, i.e.

$$(1-\phi)\Big(\mu_\lambda(q)X - 1\Big) + b = (1-\phi)\mu_\lambda(q)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b,$$

which simplifies to

$$q = z(\pi, q), \tag{A.8}$$

and whose unique positive solution is

$$q = \frac{\sqrt{\pi}(1-\sqrt{\pi})}{1-\pi} \triangleq \gamma(\pi) \in \left(0, \tfrac{1}{2}\right). \tag{A.9}$$

Hence, when $\gamma(\pi) > r(\lambda)$ the set of borrowers that withhold data is $Q = [0, \gamma(\pi)]$. When $\gamma(\pi) \leq r(\lambda)$, withholding data leads to credit rationing, so that all the borrowers with $z \in [0, r(\lambda)]$ are indifferent between data sharing and withholding. Assuming that indifferent borrowers withhold data, the set of borrowers that withhold data is $Q = [0, r(\lambda)]$.

When $\lambda \in [-1, 0)$, by a symmetric argument we have that when $1 - \gamma(\pi) < r(\lambda)$ the set of borrowers that withhold is $Q = [1 - \gamma(\pi), 1]$, while when $1 - \gamma(\pi) \geq r(\lambda)$ it is $Q = [r(\lambda), 1]$.

In summary, the equilibrium is described by the following data withholding set:

$$Q = \mathcal{G}(\lambda, \pi) \triangleq \begin{cases} \Big[0, \max\big\{r(\lambda); \gamma(\pi)\big\}\Big] & \text{if } \lambda \in (0, 1], \\ \Big[0, 1\Big] & \text{if } \lambda = 0, \\ \Big[\min\big\{r(\lambda); 1 - \gamma(\pi)\big\}, 1\Big] & \text{if } \lambda \in [-1, 0), \end{cases}$$

where $\gamma(\pi)$ and $r(\lambda)$ are given respectively by Equation (A.9) and Equation (2.8). ■

Proof of Proposition 2.2. We proceed in 2 steps. Step 1 builds the borrower's expected utilities from data sharing and data withholding. Step 2 finds the equilibrium disclosure strategy.

Step 1. The λ-lender's allocation rule is described by Lemma 2.1, so the borrower's expected utility form data sharing is

$$V^B(z) = \int_{-1}^{1} V_\lambda^B(z)\,\frac{1}{2}\,\mathrm{d}\lambda$$

where $V_\lambda^B(z)$ is defined in Equation (A.3). Let

$$\widehat{\lambda}(z) \triangleq r^{-1}(z) = \frac{\frac{1}{2}-\frac{1}{X}}{\frac{1}{2}-z}, \tag{A.10}$$

be the inverse function of the rationing threshold defined in Equation (2.8). From Equation (2.7), we can distinguish 3 sets of borrowers: i) borrowers with $z \in [0, 1/X]$ are credit rationed whenever $\lambda \geq \widehat{\lambda}(z)$; ii) borrowers with $z \in (1/X, 1-1/X)$ get credit for every $\lambda \in [-1, 1]$; and iii) borrowers with $z \in [1-1/X, 1]$ are credit rationed whenever $\lambda \leq \widehat{\lambda}(z)$. Thus, the borrower's expected surplus from data sharing can be rewritten as

$$V^B(z) = \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z)} \Big[(1-\phi)\Big(\mu_\lambda(z)X-1\Big)+b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[0, \dfrac{1}{X}\right] \\ \displaystyle\int_{-1}^{1} \Big[(1-\phi)\Big(\mu_\lambda(z)X-1\Big)+b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left(\dfrac{1}{X}, 1-\dfrac{1}{X}\right) \\ \displaystyle\int_{\widehat{\lambda}(z)}^{1} \Big[(1-\phi)\Big(\mu_\lambda(z)X-1\Big)+b\Big]\,\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[1-\dfrac{1}{X}, 1\right], \end{cases} \tag{A.11}$$

or, equivalently, as

$$V^B(z) = \begin{cases} p(z)\Big[(1-\phi)\Big(\mu_-(z)X-1\Big)+b\Big] & \text{if } z \in \left[0, \dfrac{1}{X}\right] \\ (1-\phi)\left(\frac{1}{2}X-1\right)+b & \text{if } z \in \left(\dfrac{1}{X}, 1-\dfrac{1}{X}\right) \\ \Big(1-p(z)\Big)\Big[(1-\phi)\Big(\mu_+(z)X-1\Big)+b\Big] & \text{if } z \in \left[1-\dfrac{1}{X}, 1\right], \end{cases} \tag{A.12}$$

where

$$p(z) \triangleq \Pr\Big(\lambda \leq \widehat{\lambda}(z)\Big) = \frac{\widehat{\lambda}(z)+1}{2},$$

$$\mu_-(z) \triangleq \mathbb{E}\Big(\mu_\lambda(z)\Big|\lambda \leq \widehat{\lambda}(z)\Big) = \frac{-1+\widehat{\lambda}(z)}{2}\,z + \left(1-\frac{-1+\widehat{\lambda}(z)}{2}\right)\frac{1}{2},$$

$$\mu_+(z) \triangleq \mathbb{E}\Big(\mu_\lambda(z)\Big|\lambda \geq \widehat{\lambda}(z)\Big) = \frac{\widehat{\lambda}(z)+1}{2}\,z + \left(1-\frac{\widehat{\lambda}(z)+1}{2}\right)\frac{1}{2}.$$

Let $Q \triangleq \{z \in [0,1] : m(z) = \varnothing\}$ be the set of borrowers that withhold data, the λ-lender's estimate of the borrower's quality $\mu_\lambda(\varnothing)$ and the optimal allocation rule $\big(\ell_\lambda(\varnothing), x_\lambda(\varnothing)\big)$ is given by Lemma A.1. The borrower's expected utility form data withholding is

$$V^B(\varnothing, z) = \int_{-1}^{1} V_\lambda^B(\varnothing, z)\,\frac{1}{2}\,\mathrm{d}\lambda,$$

where $V_\lambda^B(\varnothing, z)$ is defined in Equation (A.7). Proceeding as above we get the borrower's expected surplus from data withholding

$$V^B(\varnothing, z) = \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z(\pi,Q))} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left[0, \frac{1}{X}\right] \\ \displaystyle\int_{-1}^{1} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left(\frac{1}{X}, 1 - \frac{1}{X}\right) \\ \displaystyle\int_{\widehat{\lambda}(z(\pi,Q))}^{1} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(\varnothing)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z(\pi,Q) \in \left[1 - \frac{1}{X}, 1\right]. \end{cases}$$

Step 2. First, consider the case $\phi \geq 1 - 2b$. Suppose $Q = [0,1]$, from Lemma A.1 (Equation (A.5)) we can compute the λ-lender's estimate of the borrower's quality following data withholding to get

$$\mu_\lambda(\varnothing) = z(\pi, Q) = \frac{1}{2} \in \left(\frac{1}{X}, 1 - \frac{1}{X}\right) \tag{A.13}$$

for all $\lambda \in [-1, 1]$. It follows that the borrower's utility from data withholding simplifies to (suppressing the dependency of z)

$$V^B(\varnothing) = (1-\phi)(\tfrac{1}{2}X - 1) + b \tag{A.14}$$

for all $z \in [0,1]$.

We want to show that $V^B(\varnothing) \geq V^B(z)$ for all $z \in [0,1]$. Note that $V^B(z)$ is continuous, symmetric around $z = 1/2$, constant and equal to $V^B(\varnothing)$ for $z \in (1/X, 1 - 1/X)$, and convex for $z \in [0, 1/X] \cup [1 - 1/X, 1]$. To show the convexity of $V^B(z)$ over $z \in [0, 1/X]$ note that its first derivative in this region is

$$\begin{aligned} \frac{\partial V^B(z)}{\partial z} &= \widehat{\lambda}'(z)\Big[(1-\phi)\Big(\mu_{\widehat{\lambda}(z)}(z)X - 1\Big) + b\Big]\frac{1}{2} + (1-\phi)X\frac{1}{2}\int_{-1}^{\widehat{\lambda}(z)} \lambda\,\mathrm{d}\lambda, \\ &= \widehat{\lambda}'(z)b\frac{1}{2} + (1-\phi)X\frac{1}{2}\int_{-1}^{\widehat{\lambda}(z)} \lambda\,\mathrm{d}\lambda, \end{aligned}$$

where the first line uses Leibnitz integral rule, and the second line follows from the fact that

$$\mu_{\widehat{\lambda}(z)}(z) = \frac{1}{2} + \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - z}\left(z - \frac{1}{2}\right) = \frac{1}{X};$$

while the second derivative is

$$\frac{\partial^2 V^B(z)}{\partial z^2} = \widehat{\lambda}''(z)b\frac{1}{2} + (1-\phi)X\frac{1}{2}\widehat{\lambda}'(z)\widehat{\lambda}(z) > 0,$$

since

$$\begin{aligned}
\widehat{\lambda}(z) &> 0,\\
\widehat{\lambda}'(z) &= \frac{\frac{1}{2}-\frac{1}{X}}{\left(\frac{1}{2}-z\right)^2} = \frac{1}{\frac{1}{2}-z}\widehat{\lambda}(z) > 0,\\
\widehat{\lambda}''(z) &= 2\frac{\frac{1}{2}-\frac{1}{X}}{\left(\frac{1}{2}-z\right)^3} = 2\frac{1}{\frac{1}{2}-z}\widehat{\lambda}'(z) > 0,
\end{aligned}$$

because every $z \in [0, \frac{1}{X}]$ is strictly below $\frac{1}{2}$ since by Assumption 2.1 we have $\frac{1}{X} < \frac{1}{2}$. By symmetry, $V^B(z)$ is also convex over $[1-\frac{1}{X}, 1]$.
It follows that $Q = [0,1]$ is the equilibrium data-withholding set if borrowers with extreme $z \in 0, 1$ weakly prefers to withhold data. By symmetry of $V^B(z)$ it is sufficient that

$$V^B(\varnothing) = (1-\phi)\left(\frac{1}{2}X - 1\right) + b \geq \left(1 - \frac{1}{X}\right)\left[(1-\phi)\left(\left(1+\frac{1}{X}\right)\frac{1}{2}X - 1\right) + b\right] = V^B(0).$$

This inequality simplifies to $\phi \geq 1 - 2b$, the parameters space we are considering.
Note also that any other putative equilibrium data withholding set of the form $Q = [0, q] \cup [1-q, 1]$ with $q \in (0, 1/2)$ would lead to the estimated borrower's quality in Equation (A.13) and yield the profits in Equation (A.14) to a borrower withholding data, so that every borrower with $z \in (q, 1-q)$ would deviate to data withholding.

Consider now $\phi < 1 - 2b$ and suppose $Q = [q, 1-q]$ with $q \in (0, 1/X)$. From Lemma A.1 (Equation (A.5)) the λ-lender's estimate of the borrower's quality and the borrower's profits following data withholding are still given by Equation (A.13) and Equation (A.14), but for $\phi < 1 - 2b$ the borrower with $z = 0$ (and $z = 1$) strictly prefers to share data, as Equation (A.1) is not satisfied. The indifferent borrower q has to obtain the same expected surplus from both strategies, i.e.

$$V^B(\varnothing) = (1-\phi)\left(\frac{1}{2}X - 1\right) + b = p(q)\Big[(1-\phi)\Big(\mu_-(q)X - 1\Big) + b\Big] = V^B(q),$$

whose unique solution lower than $1/X$ is

$$q = \frac{1}{X}\left(1 - \frac{2b}{(1-\phi)}\right) \in \left(0, \tfrac{1}{X}\right). \tag{A.15}$$

By the symmetry of $V^B(z)$, we also have $V^B(\varnothing) = V^B(1-q)$.

Since $q > 0$ if and only if $\phi < 1 - 2b$, the equilibrium can be compactly described by the data withholding set:

$$Q = \mathcal{H}(b, \phi) \triangleq [\eta(b, \phi), 1 - \eta(b, \phi)],$$

where

$$\eta(b, \phi) \triangleq \max\left\{0, \frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right)\right\}. \tag{A.16}$$

■

Proof of Proposition 2.3. We proceed in 4 steps: Step 1 and Step 2 derive the lender's expected profits when the lender employs a transparent and an opaque algorithm, respectively; Step 3 studies the graph of the lender's profits as a function of π; Step 4 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda, \pi), \mathcal{H}(b, \phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda, \pi)$ and $\mathcal{H}(b, \phi)$ are defined in Equation (2.12) and Equation (2.13), respectively. The lender's equilibrium profits from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q are:

$$\begin{aligned} V_\tau^L(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} V_Q^L(\pi, \lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 V_Q^L(\pi, \lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{A.17}$$

where

$$V_Q^L(\pi, \lambda) \triangleq \pi V_\lambda^L(\varnothing) + (1-\pi)\int_0^1 V_\lambda^L\big(m_Q(z)\big)\mathrm{d}z, \tag{A.18}$$

$V_\lambda^L(z)$ and $V_\lambda^L(\varnothing)$ are defined in Equation (A.3) and Equation (A.6), while the second line of Equation (A.17) follows from the lender's profit function being even in λ.

Step 1. Consider the lender's profits with a transparent algorithm.

If $\lambda = 0$ data is not used by the allocation rule and borrowers withhold data, i.e. $\mathcal{G}(\lambda, \pi) = [0, 1]$, thus the lender's expected profits are equal to

$$V_{\mathcal{G}}^L(\pi, 0) = \phi\left(\tfrac{1}{2}X - 1\right).$$

Now consider $\lambda \in \big(0, \widehat{\lambda}\big(\gamma(\pi)\big)\big)$ where $\widehat{\lambda}(\cdot)$ is defined in Equation (A.10). For these parameter values we either have that the allocation rule does not use data to ration credit (when $\lambda \in (0, \overline{\lambda})$), or the allocation rule rations credit but borrowers escape rationing by withholding data since $r(\lambda) < \gamma(\pi)$ (i.e. $\lambda \in \big[\overline{\lambda}, \widehat{\lambda}(\gamma(\pi))\big)$). In both cases $\mathcal{G}(\lambda, \pi) = [0, \gamma(\pi)]$ (see Proposition 2.1). The lender's profits are:

$$\begin{aligned} V_{\mathcal{G}}^L(\pi, \lambda) &= \Big(\pi + (1-\pi)\gamma(\pi)\Big)\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + (1-\pi)\int_{\gamma(\pi)}^1 \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z, \\ &= \Big(\pi + (1-\pi)\gamma(\pi)\Big)\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + \\ &+ (1-\pi)(1-\gamma(\pi))\phi\left(\mu_\lambda\left(\tfrac{\gamma(\pi)+1}{2}\right)X - 1\right), \\ &= \phi\left(\tfrac{1}{2}X - 1\right), \end{aligned}$$

where the last line follows from the fact that, in equilibrium, $\mu_\lambda(\varnothing) = \lambda\gamma(\pi) + (1-\lambda)\frac{1}{2}$ (see Proposition 2.1, Equation (A.8) and Equation (A.9)), and

$$\big(\pi + (1-\pi)\gamma(\pi)\big)\gamma(\pi) + (1-\pi)(1-\gamma(\pi))\frac{\gamma(\pi)+1}{2} = \frac{1}{2}.$$

Lastly, consider $\lambda \in \left[\widehat{\lambda}(\gamma(\pi)), 1\right]$. We have $r(\lambda) \geq \gamma(\pi)$ and $\mathcal{G}(\lambda, \pi) = [r(\lambda), 1]$, since borrowers that withhold data are credit rationed $\ell_\lambda(\varnothing) = 0$. The lender's profits are:

$$V^L_{\mathcal{G}}(\pi, \lambda) = (1-\pi)\int_{r(\lambda)}^1 \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z.$$

Note that $\widehat{\lambda}\big(\gamma(\pi)\big) < 1$ if and only if $\gamma(\pi) < 1/X$, i.e. if $\pi < \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2 \in (0,1)$, hence letting

$$\widetilde{\lambda}(\pi) \triangleq \min\left\{\widehat{\lambda}\big(\gamma(\pi)\big), 1\right\},$$

and integrating $V_{\mathcal{G}}(\pi, \lambda)$ over λ we get the ex-ante expected profits from a transparent algorithm:

$$V^L_T(\pi) = \widetilde{\lambda}(\pi)\,\phi\left(\tfrac{1}{2}X - 1\right) + (1-\pi)\int_{\widetilde{\lambda}(\pi)}^1 \int_{r(\lambda)}^1 \phi\Big(\mu_\lambda(z)X - 1\Big)\;\mathrm{d}z\,\mathrm{d}\lambda. \tag{A.19}$$

Step 2. Consider the lender's profits when with an opaque algorithm when $\phi < 1-2b$, i.e. $\eta(b,\phi) > 0$.

If $\lambda \in [0, \overline{\lambda})$, where $\overline{\lambda} = 1 - 2/X$, data is not used to ration credit, hence the lender's profits conditional on λ are:

$$\begin{aligned}
V^L_{\mathcal{H}}(\pi, \lambda) &= \Big(\pi + (1-\pi)\big(1 - 2\eta(b,\phi)\big)\Big)\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + \\
&+ (1-\pi)\left[\int_0^{\eta(b,\phi)} \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z + \int_{1-\eta(b,\phi)}^1 \phi\Big(\mu_\lambda(z)X - 1\Big)\mathrm{d}z\right], \\
&= \Big(\pi + (1-\pi)\big(1 - 2\eta(b,\phi)\big)\Big)\phi\Big(\mu_\lambda(\varnothing)X - 1\Big) + \\
&+ (1-\pi)\eta(b,\phi)\phi\left[\mu_\lambda\left(\tfrac{\eta(b,\phi)}{2}\right)X - 1 + \mu_\lambda\left(\tfrac{2-\eta(b,\phi)}{2}\right)X - 1\right], \\
&= \phi\left(\tfrac{1}{2}X - 1\right),
\end{aligned}$$

where the last line follows from the fact that in equilibrium $\mu_\lambda(\varnothing) = \frac{1}{2}$ (see Proposition 2.2, Equation (A.13)) and

$$\mu_\lambda\left(\tfrac{\eta(b,\phi)}{2}\right)X - 1 + \mu_\lambda\left(\tfrac{2-\eta(b,\phi)}{2}\right)X - 1 = 2\left(\tfrac{1}{2}X - 1\right).$$

Consider next $\lambda \in [\overline{\lambda}, \widehat{\lambda}(\eta(b,\phi))]$ where $\widehat{\lambda}(\cdot)$ is defined in Equation (A.10). For these parameters we have $r(\lambda) \leq \eta(b,\phi)$, so that some of the borrowers sharing data are credit rationed. The lender's profits conditional on λ are:

$$\begin{aligned}
V^L_{\mathcal{H}}(\pi,\lambda) &= \pi\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+\\
&+(1-\pi)\int_0^{\eta(b,\phi)}\mathbb{1}\{z>r(\lambda)\}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z+\\
&+(1-\pi)\Big[\int_{\eta(b,\phi)}^{1-\eta(b,\phi)}\phi\Big(\mu_\lambda(\varnothing)X-1\Big)\mathrm{d}z+\int_{1-\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z\Big],\\
&=\pi\phi\Big(\tfrac{1}{2}X-1\Big)+(1-\pi)\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z,
\end{aligned}$$

where the fourth line follows from the fact that:

$$\begin{aligned}
\int_{\eta(b,\phi)}^{1-\eta(b,\phi)}\phi\Big(\mu_\lambda(\varnothing)X-1\Big)\mathrm{d}z &= \int_{\eta(b,\phi)}^{1-\eta(b,\phi)}\phi\left(\tfrac{1}{2}X-1\right)\mathrm{d}z\\
&=\int_{\eta(b,\phi)}^{1-\eta(b,\phi)}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z.
\end{aligned} \tag{A.20}$$

Lastly, consider $\lambda\in(\widehat{\lambda}(\eta(b,\phi)),1]$, that is parameters such that $r(\lambda)>\eta(b,\phi)$ and all the borrowers that share data $z<\eta(b,\phi)$ are credit rationed. The lender's profits are:

$$\begin{aligned}
V^L_{\mathcal{H}}(\pi,\lambda) &= \pi\phi\Big(\mu_\lambda(\varnothing)X-1\Big)+\\
&+(1-\pi)\Big[\int_{\eta(b,\phi)}^{1-\eta(b,\phi)}\phi\Big(\mu_\lambda(\varnothing)X-1\Big)\mathrm{d}z+\int_{1-\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z\Big],\\
&=\pi\phi\left(\tfrac{1}{2}X-1\right)+(1-\pi)\int_{\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\mathrm{d}z,
\end{aligned}$$

where the third line follows from Equation (A.20).

Integrating $V^L_{\mathcal{H}}(b,\phi)$ over λ, we get the lender's ex-ante expected profits from an opaque algorithm are:

$$\begin{aligned}
V^L_O(\pi) &= \overline{\lambda}\,\phi\left(\tfrac{1}{2}X-1\right)+\\
&+\int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))}\pi\phi\left(\tfrac{1}{2}X-1\right)+(1-\pi)\int_{r(\lambda)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda+\\
&+\int_{\widehat{\lambda}(\eta(b,\phi))}^{1}\pi\phi\left(\tfrac{1}{2}X-1\right)+(1-\pi)\int_{\eta(b,\phi)}^{1}\phi\Big(\mu_\lambda(z)X-1\Big)\,\mathrm{d}z\,\mathrm{d}\lambda.
\end{aligned} \tag{A.21}$$

Note that when $\phi\geq 1-2b$ we have $\eta(b,\phi)=0$ and $\widehat{\lambda}(\eta(b,\phi))=\overline{\lambda}$, hence the lender's profits from Equation (A.21) simplifies to $V^L_O(\pi)=\phi\left(\tfrac{1}{2}X-1\right)$.

Step 3. We now study the graph of the lender's profits in Equation (A.19) and Equation (A.21). For an arbitrary function $y:(0,1)\to\mathbb{R}$, we will use the simplified notation

$y(0) = \lim_{\pi \to 0} y(\pi)$ and $y(1) = \lim_{\pi \to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1) = 1$ and $\widetilde{\lambda}(0) = \overline{\lambda}$, we have that

$$V_O^L(1) = V_T^L(1) = \phi\left(\tfrac{1}{2}X - 1\right),$$

and

$$V_T^L(0) - V_O^L(0) = -\int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \phi\Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z\, \mathrm{d}\lambda > 0,$$

where this uses the fact that

$$\left(\tfrac{1}{2}X - 1\right) = \int_0^1 \Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z, \tag{A.22}$$

and the inequality follows from the fact that the integrand is negative for $z \leq r(\lambda)$.

Second, $V_O^L(\pi)$ is flat and equal to $\phi\left(\frac{1}{2}X - 1\right) = V_L^O(1)$ if $\phi \geq 1 - 2b$ (i.e. when $\eta(b,\phi) = 0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and is strictly decreasing if $\phi < 1 - 2b$ (i.e. when $\eta(b,\phi) > 0$ and $\widehat{\lambda}(\eta(b,\phi)) > \overline{\lambda}$), since

$$\begin{aligned}
\frac{\partial V_O^L}{\partial \pi} &= \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \phi\left(\tfrac{1}{2}X - 1\right) - \int_{r(\lambda)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z\, \mathrm{d}\lambda + \\
&\quad + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \phi\left(\tfrac{1}{2}X - 1\right) - \int_{\eta(b,\phi)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z\, \mathrm{d}\lambda, \\
&= \int_{\overline{\lambda}}^{1} \int_0^{\min\{r(\lambda),\eta(b,\phi)\}} \phi\Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z\, \mathrm{d}\lambda < 0,
\end{aligned}$$

where the third line follows from Equation (A.22) and the fact that $r(\lambda) \geq \eta(b,\phi)$ if and only if $\lambda \geq \widehat{\lambda}(\eta(b,\phi))$, while the inequality follows from the integrand being negative for $z \leq r(\lambda)$.

Third, $V_T^L(\pi)$ is flat and equal to $\phi\left(\frac{1}{2}X - 1\right) = V_T^L(1)$ if $\pi \geq \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2 \in (0,1)$ (i.e. when $\widetilde{\lambda}(\pi) = 1$) and is strictly decreasing for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$). To prove that $V_T^L(\pi)$ is decreasing for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$, rewrite it as

$$V_T^L(\pi) = \widehat{\lambda}(\gamma(\pi))\, \phi\left(\tfrac{1}{2}X - 1\right) + (1 - \pi) \int_{\widehat{\lambda}(\gamma(\pi))}^{1} K(\lambda)\, \mathrm{d}\lambda$$

where

$$\begin{aligned}
K(\lambda) &\triangleq \int_{r(\lambda)}^{1} \phi\Big(\mu_\lambda(z)X - 1\Big)\, \mathrm{d}z, \\
&= (1 - r(\lambda))\phi\left(\mu_\lambda\left(\tfrac{r(\lambda)+1}{2}\right)X - 1\right),
\end{aligned} \tag{A.23}$$

and $\mu_\lambda(\cdot)$ is defined in Equation (2.2). The first derivative of $V_T^L(\pi)$ is:

$$\begin{aligned}\frac{\partial V_T^L}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[\phi\left(\tfrac{1}{2}X-1\right)-(1-\pi)K\left(\widehat{\lambda}(\gamma(\pi))\right)\right]-\int_{\widehat{\lambda}(\gamma(\pi))}^{1}K(\lambda)\mathrm{d}\lambda, \\ &= -\int_{\widehat{\lambda}(\gamma(\pi))}^{1}K(\lambda)\mathrm{d}\lambda<0,\end{aligned} \tag{A.24}$$

where the first line uses Leibnitz integral rule, the second line follows from the term in square brackets being nil, while the inequality follows from the fact that the integrand Equation (A.23) is strictly positive for $z > r(\lambda)$. To see that the term in square brackets in the first line of Equation (A.24) is nil, substitute $\widehat{\lambda}(\gamma(\pi))$ in Equation (A.23) and note that

$$r\left(\widehat{\lambda}(\gamma(\pi))\right)=\gamma(\pi), \tag{A.25}$$

since $\widehat{\lambda}(\cdot)=r^{-1}(\cdot)$ by definition (see Equation (A.10)), implying that

$$\begin{aligned}\mu_{\widehat{\lambda}(\gamma(\pi))}\left(\tfrac{r\left(\widehat{\lambda}(\gamma(\pi))\right)+1}{2}\right)X-1 &= \left(\widehat{\lambda}(\gamma(\pi))\frac{\gamma(\pi)+1}{2}+\left(1-\widehat{\lambda}(\gamma(\pi))\right)\frac{1}{2}\right)X-1, \\ &= \left(\tfrac{1}{2}X-1\right)+\frac{1}{2}\widehat{\lambda}(\gamma(\pi))\gamma(\pi)X, \\ &= \left(\tfrac{1}{2}X-1\right)+\frac{\sqrt{\pi}}{1-\sqrt{\pi}}\left(\tfrac{1}{2}X-1\right)>0;\end{aligned} \tag{A.26}$$

hence the term in square brackets in Equation (A.24) simplifies to

$$\phi\left[\left(\tfrac{1}{2}X-1\right)-(1-\pi)(1-\gamma(\pi))\left(\left(\tfrac{1}{2}X-1\right)+\frac{\sqrt{\pi}}{1-\sqrt{\pi}}\left(\tfrac{1}{2}X-1\right)\right)\right]=0,$$

since $(1-\pi)(1-\gamma(\pi))=1-\sqrt{\pi}$.

Step 4. Let

$$\widehat{\pi}(b,\phi)\triangleq\left\{\pi\in\left(0,\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right)\middle| V_T^L(\pi)=V_O^L(\pi)\ \text{ when}\,\phi<1-2b\right\} \tag{A.27}$$

be the π that solves $V_T^L(\pi) = V_O^L(\pi)$ where $V_T^L(\pi)$ is given by Equation (A.19) with $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi))$, while $V_O^L(\pi)$ is given by Equation (A.21) with $\eta(b, \phi) > 0$. From the shapes of $V_T^L(\pi)$ and $V_O^L(\pi)$ described in Step 3, it follows that:

- if $\phi \geq 1-2b$, we have
  - $V_T^L(\pi) > V_O^L(\pi)$ for $\pi < \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,
  - $V_T^L(\pi) = V_O^L(\pi)$ for $\pi \geq \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,
- if $\phi < 1-2b$, we have
  - $V_T^L(\pi) > V_O^L(\pi)$ for $\pi < \widehat{\pi}(b,\phi)$,

– $V_T^L(\pi) = V_O^L(\pi)$ for $\pi = \widehat{\pi}(b, \phi)$,
– $V_T^L(\pi) < V_O^L(\pi)$ for $\pi > \widehat{\pi}(b, \phi)$.

■

Proof of Lemma 2.2. Given a risk bucket r_i, the lender allocates credit as described in Lemma 2.1. In particular, expected credit quality is

$$\mathbb{E}(\theta|r_i) = \lambda \frac{z_{i-1} + z_i}{2} + (1-\lambda)\tfrac{1}{2} \triangleq \mu_\lambda(r_i),$$

while the allocation rule is:

$$x_\lambda(r_i) = \frac{1}{\mu_\lambda(r_i)} + \phi\left(X - \frac{1}{\mu_\lambda(r_i)}\right),$$
$$\ell_\lambda(r_i) = \mathbb{1}\left\{\mu_\lambda(r_i) > \frac{1}{X}\right\}.$$

Suppose $\lambda > 0$, the welfare-optimal data segmentation solves:

$$\max_Z \quad \sum_{i=1}^{n} (z_i - z_{i-1})\, \ell_\lambda(r_i)\big[\mu_\lambda(r_i)\, X - 1 + b\big].$$

By the revelation principle, we can pool all the potential risk buckets r_i inducing the same credit provision decision in a unique signal and interpret signals as incentive-compatible action recommendations. The planner's problem simplifies to:

$$\begin{aligned} \max_{z_1 \in [0,1]} \quad & (1 - z_1)\left[\mu_\lambda\left(\tfrac{z_1+1}{2}\right) X - 1 + b\right] \\ \text{s.t.} \quad & \mu_\lambda\left(\tfrac{z_1+1}{2}\right) > \frac{1}{X} \\ & \mu_\lambda\left(\tfrac{0+z_1}{2}\right) \le \frac{1}{X}. \end{aligned} \tag{A.28}$$

The objective function is concave and the first order condition is solved for

$$z_1 = \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right) - \frac{b}{\lambda X} = r(\lambda) - \frac{b}{\lambda X},$$

which is positive as long as $b \le 1 - \frac{1}{2}(1-\lambda)X$. One can easily show that the constraints are satisfied for

$$z_1 = \max\left\{r(\lambda) - \frac{b}{\lambda X}, 0\right\},$$

hence this is the solution of the problem in Equation (A.28) for $\lambda > 0$.

For $\lambda < 0$, relabeling action recommendations, we get

$$z_1 = \min\left\{r(\lambda) - \frac{b}{\lambda X}, 1\right\}.$$

When $\lambda = 0$, data segmentation is irrelevant for both the social planner and the lender.

■

Proof of Proposition 2.4. We proceed in 3 steps: Step 1 derives the social welfare when the lender employs a transparent and an opaque algorithm; Step 2 studies the graph of social welfare in the two regimes as a function of π; Step 3 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Equation (2.12) and Equation (2.13), respectively. The equilibrium social welfare from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} W_\tau(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} W_Q(\pi,\lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 W_Q(\pi,\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{A.29}$$

where

$$W_Q(\pi,\lambda) \triangleq \pi W_\lambda(\varnothing) + (1-\pi)\int_0^1 W_\lambda\big(m_Q(z)\big)\mathrm{d}z, \tag{A.30}$$

$W_\lambda(z)$ and $W_\lambda(\varnothing)$ are defined in Equation (A.3) and Equation (A.6), while the second line of Equation (A.29) follows from the social surplus function being even in λ.

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 2.3, we get the social welfare from a transparent and an opaque algorithm. These are, respectively

$$W_T(\pi) = \widetilde{\lambda}(\pi)\left(\frac{1}{2}X - 1 + b\right) + (1-\pi)\int_{\widetilde{\lambda}(\pi)}^{1}\int_{r(\lambda)}^{1}\Big(\mu_\lambda(z)X - 1 + b\Big)\,\mathrm{d}z\,\mathrm{d}\lambda, \tag{A.31}$$

and

$$\begin{aligned} W_O(\pi) = \overline{\lambda}&\left(\frac{1}{2}X - 1 + b\right) + \\ &+ \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))}\left[\pi\left(\frac{1}{2}X - 1 + b\right) + (1-\pi)\int_{r(\lambda)}^{1}\Big(\mu_\lambda(z)X - 1 + b\Big)\,\mathrm{d}z\right]\mathrm{d}\lambda + \\ &+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1}\left[\pi\left(\frac{1}{2}X - 1 + b\right) + (1-\pi)\int_{\eta(b,\phi)}^{1}\Big(\mu_\lambda(z)X - 1 + b\Big)\,\mathrm{d}z\right]\mathrm{d}\lambda. \end{aligned} \tag{A.32}$$

Step 2. We now study the graph of the lender's profits in Equation (A.31) and Equation (A.32).

Step 2a. We first evaluate $W_T(\pi)$ and $W_O(\pi)$ when π approaches 0 and 1. For an arbitrary function $y : (0,1) \to \mathbb{R}$, we will use the simplified notation $y(0) = \lim_{\pi\to 0} y(\pi)$

and $y(1) = \lim_{\pi \to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1) = 1$ and $\widetilde{\lambda}(0) = \overline{\lambda}$, we have that

$$W_O(1) = W_T(1) = \tfrac{1}{2}X - 1 + b,$$

and, using Equation (A.22)

$$W_O(0) - W_T(0) = \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mu_\lambda(z)X - 1 + b \;\mathrm{d}z\,\mathrm{d}\lambda \triangleq Z(b,\phi). \tag{A.33}$$

The sign of $Z(b,\phi)$ depends on its arguments. Rewrite this function as

$$Z(b,\phi) = \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} Y(b,\phi,\lambda)\,\mathrm{d}\lambda$$

where

$$Y(b,\phi,\lambda) \triangleq \int_{\eta(b,\phi)}^{r(\lambda)} \mu_\lambda(z)X - 1 + b \;\mathrm{d}z. \tag{A.34}$$

Note that $Z(b,\phi)$ is continuous, constant in ϕ for $\phi \geq 1 - 2b$ (i.e. when $\eta(b,\phi) = 0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and equal to

$$Z_-(b) \triangleq \int_{\overline{\lambda}}^{1} \int_{0}^{r(\lambda)} \mu_\lambda(z)X - 1 + b\,\mathrm{d}z\,\mathrm{d}\lambda, \tag{A.35}$$

while it is strictly decreasing in ϕ for $\phi < 1 - 2b$, since

$$\begin{aligned}
\frac{\partial Z}{\partial \phi} &= -\frac{\partial \widehat{\lambda}}{\partial \eta}\frac{\partial \eta}{\partial \phi} Y\big(b,\phi,\widehat{\lambda}\big(\eta(b,\phi)\big)\big) + \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \frac{\partial Y}{\partial \phi}\,\mathrm{d}\lambda, \\
&= \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \frac{\partial Y}{\partial \phi}\,\mathrm{d}\lambda, \\
&= \frac{2b}{X(1-\phi)^2} \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \mu_\lambda\big(\eta(b,\phi)\big)X - 1 + b\,\mathrm{d}\lambda, \\
&= \frac{2b}{X(1-\phi)^2}\big(1 - \widehat{\lambda}(\eta(b,\phi))\big)\left[\left(\tfrac{\widehat{\lambda}(\eta(b,\phi))+1}{2}\eta(b,\phi) + \left(1 - \tfrac{\widehat{\lambda}(\eta(b,\phi))+1}{2}\right)\frac{1}{2}\right)X - 1 + b\right], \\
&= \frac{2b}{X(1-\phi)^2}\big(1 - \widehat{\lambda}(\eta(b,\phi))\big)\left[\left(\frac{1}{X} - \frac{b}{(1-\phi)X}\right)X - 1 + b\right] < 0,
\end{aligned} \tag{A.36}$$

where the first line uses the Leibnitz integral rule, the second lines follows from the fact that $Y\big(b,\phi,\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = 0$ since $r\big(\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = \eta(b,\phi)$ because of Equation (A.25), while the inequality follows from the fact that the term in square brackets is strictly negative.

Moreover, $Z(b,\phi)$ is strictly increasing in b. This is obvious when $\phi \geq 1-2b$ and $Z(b,\phi) = Z_-(b)$, while for $\phi < 1-2b$ we have

$$\begin{aligned}
\frac{\partial Z}{\partial b} &= -\frac{\partial \widehat{\lambda}}{\partial \eta}\frac{\partial \eta}{\partial b} Y\big(b,\phi,\widehat{\lambda}(\eta(b,\phi))\big) + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[-\frac{\partial \eta}{\partial b}\big(\mu_\lambda(\eta(b,\phi))X - 1 + b\big) + \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z\right] \mathrm{d}\lambda, \\
&= \frac{2}{X(1-\phi)} \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \Big[\mu_\lambda(\eta(b,\phi))X - 1 + b\Big] \mathrm{d}\lambda + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z\, \mathrm{d}\lambda, \\
&= \frac{2}{X(1-\phi)} \left[\Big(1 - \widehat{\lambda}(\eta(b,\phi))\Big)\left(\tfrac{1}{2}X - 1 + b + \frac{\widehat{\lambda}(\eta(b,\phi)) + 1}{2}\eta(b,\phi)\right)\right] + \\
&\quad + \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \mathrm{d}z\, \mathrm{d}\lambda > 0,
\end{aligned} \tag{A.37}$$

where the first lines uses Leibnitz integral rule, the second lines follows from $Y\big(\phi,\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = 0$ since $r\big(\widehat{\lambda}\big(\eta(b,\phi)\big)\big) = \eta(b,\phi)$ because of Equation (A.25), while the inequality in the last line follows from the following facts: i) the term in square brackets is positive because $\widehat{\lambda}(\eta(b,\phi)) \in (0,1)$ and $\frac{1}{2}X - 1 > 0$ by Assumption 2.1; and ii) the double integral is positive since for $\lambda > \widehat{\lambda}(\eta(b,\phi))$ we have $r(\lambda) > \eta(b,\phi)$.

Moreover, note that $Z(0,\phi) = 0$ since $\widehat{\lambda}(\eta(0,\phi)) = 1$, and that $Z_-(b) > 0$ if and only if $b > b^\circ(X)$ where

$$b^\circ(X) \triangleq -\frac{\int_{\overline{\lambda}}^{1}\int_{0}^{r(\lambda)} \mu_\lambda(z)X - 1\, \mathrm{d}z\, \mathrm{d}\lambda}{\int_{\overline{\lambda}}^{1}\int_{0}^{r(\lambda)} \mathrm{d}z\, \mathrm{d}\lambda} \in \left(\tfrac{1}{4}, \tfrac{1}{3}\right). \tag{A.38}$$

Let

$$\phi^\circ(b) \triangleq \left\{\phi \in (0, 1-2b) \;\middle|\; Z(b,\phi) = 0\right\} \tag{A.39}$$

be the ϕ that solves $Z(b,\phi) = 0$ where $Z(b,\phi)$ is given by Equation (A.33) with $\eta(b,\phi) > 0$.

It follows that:

- if $b > b^\circ(X)$, then $Z(b,\phi) > 0$ for every $\phi \in (0,1)$;
- if $b = b^\circ(X)$, then $Z(b,\phi) > 0$ for $\phi \in (0, 1-2b)$ and $Z(b,\phi) = 0$ for $\phi \in [1-2b, 1)$;
- if $b < b^\circ(X)$, then $Z(b,\phi) > 0$ if $\phi \in (0, \phi^\circ(b))$, $Z(b,\phi) = 0$ if $\phi = \phi^\circ(b)$, and $Z(b,\phi) < 0$ if $\phi \in (\phi^\circ(b), 1)$.

Step 2b. We now study the graph of $W_T(\pi)$ and $W_O(\pi)$ over the interior of the interval $(0,1)$.

First, note that $W_O(\pi)$ is flat and equal to $\frac{1}{2}X - 1 + b = W_O(1)$ if $\phi \geq 1 - 2b$ (i.e. when $\eta(b,\phi) = 0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and is strictly decreasing if $\phi < 1 - 2b$ (i.e. when $\eta(b,\phi) > 0$ and $\widehat{\lambda}(\eta(b,\phi)) > \overline{\lambda}$) since

$$\begin{aligned}
\frac{\partial W_O}{\partial \pi} &= \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \left[\left(\tfrac{1}{2}X - 1 + b\right) - \int_{r(\lambda)}^{1} \mu_\lambda(z)X - 1 + b \ \mathrm{d}z \right] \mathrm{d}\lambda + \\
&+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[\left(\tfrac{1}{2}X - 1 + b\right) - \int_{\eta(b,\phi)}^{1} \mu_\lambda(z)X - 1 + b \ \mathrm{d}z \right] \mathrm{d}\lambda, \\
&= \int_{\overline{\lambda}}^{1} \int_{0}^{\min\{r(\lambda),\eta(b,\phi)\}} \left[\mu_\lambda(z)X - 1 + b\right] \mathrm{d}z \, \mathrm{d}\lambda, \\
&< \int_{\overline{\lambda}}^{1} \int_{0}^{\eta(b,\phi)} \left[\mu_\lambda(z)X - 1 + b\right] \mathrm{d}z \, \mathrm{d}\lambda, \\
&= (1 - \overline{\lambda})\eta(b,\phi) \left[\left(\frac{\overline{\lambda} + 1}{2} \frac{\eta(b,\phi)}{2} + \left(1 - \frac{\overline{\lambda} + 1}{2}\right) \frac{1}{2} \right) X - 1 + b \right] < 0
\end{aligned} \tag{A.40}$$

where the third line follows from Equation (A.22) and the fact that $r(\lambda) \geq \eta(b,\phi)$ if and only if $\lambda \geq \widehat{\lambda}(\eta(b,\phi))$, while the last inequality follows from the fact that the term in square brackets is strictly negative. To see this note that the term in square brackets is linear in b; it is increasing in b for $\phi < 1/X$ and equal to $-1/2 + (1-\phi)/2 < 0$ for the highest value of b among the parameters considered, i.e. $b = (1-\phi)/2$; it is decreasing in b for $\phi > 1/X$ and equal to $-1/(2X)$ for the lowest value of b, i.e. $b = 0$; it is constant in b for $\phi = 1/X$ and equal to $-1/(2X) < 0$.

Second, note that $W_T(\pi)$ is flat and equal to $\frac{1}{2}X - 1 + b = W_T(1)$ if $\pi \geq \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2 \in (0,1)$ (i.e. when $\widetilde{\lambda}(\pi) = 1$) and is strictly convex for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$) approaching the flat part from below. To prove that $W_T(\pi)$ is convex for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$ and approaches the flat part from below, rewrite it as

$$W_T(\pi) = \widehat{\lambda}(\gamma(\pi)) \ \left[\tfrac{1}{2}X - 1 + b\right] + (1 - \pi) \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda$$

where

$$\begin{aligned}
H(\lambda) &\triangleq \int_{r(\lambda)}^{1} \left[\mu_\lambda(z)X - 1 + b\right] \ \mathrm{d}z, \\
&= (1 - r(\lambda)) \left[\mu_\lambda \left(\tfrac{r(\lambda)+1}{2}\right) X - 1 + b\right].
\end{aligned} \tag{A.41}$$

The first derivative of $W_T(\pi)$ is :

$$\begin{aligned}
\frac{\partial W_T}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) \left[\left(\tfrac{1}{2}X - 1 + b\right) - (1 - \pi) \, H\big(\widehat{\lambda}(\gamma(\pi))\big) \right] - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda, \\
&= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) \Big[1 - (1 - \pi)(1 - \gamma(\pi))\Big] b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda, \\
&= \frac{\overline{\lambda}}{(1 - \sqrt{\pi})^2} \, b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} H(\lambda) \, \mathrm{d}\lambda,
\end{aligned}$$

where the first line uses the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (A.41) and using Equation (A.25) and Equation (A.26) and the fact that $(1-\pi)(1-\gamma(\pi)) = 1-\sqrt{\pi}$, while the last line follows from the fact that

$$\begin{aligned} \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2} > 0, \\ (1-\pi)(1-\gamma(\pi)) &= 1-\sqrt{\pi}. \end{aligned} \tag{A.42}$$

Note that the left derivative of $W_T(\pi)$ at $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ is

$$\frac{\partial W_T}{\partial \pi}\bigg|_{\pi=\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} = \frac{\overline{\lambda}}{\left(1-\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} b > 0, \tag{A.43}$$

since $\widehat{\lambda}(\gamma(\pi)) = 1$ for $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$.

The second derivative of $W_T(\pi)$ is

$$\begin{aligned} \frac{\partial^2 W_T}{\partial \pi^2} &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^3} b + \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\, H\big(\widehat{\lambda}(\gamma(\pi))\big), \\ &= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2}\left[b\left(\frac{1}{1-\sqrt{\pi}} + 1 - \gamma(\pi)\right) + (1-\gamma(\pi))\left(\tfrac{1}{2}X - 1\right)\left(1 + \frac{\sqrt{\pi}}{1-\sqrt{\pi}}\right)\right] > 0, \end{aligned} \tag{A.44}$$

where the first line follows from the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (A.41) and using Equation (A.25), Equation (A.26) and Equation (A.42), while the inequality follows from the fact that each of the term within the square brackets is strictly positive.

Step 3. Let

$$\pi^{\circ}(b,\phi) \triangleq \left\{ \pi \in \left(0, \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right) \middle| W_T(\pi) = W_O(\pi) \ \text{ when}\, \phi < 1-2b \right\} \tag{A.45}$$

be the π that solves $W_T(\pi) = W_O(\pi)$ where $W_T(\pi)$ is given by Equation (A.31) with $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi))$, while $W_O(\pi)$ is given by Equation (A.32) with $\eta(b,\phi) > 0$. $\pi^{\circ}(b,\phi)$ is increasing in ϕ and decreasing in b.
Let

$$\pi^{\circ}(b) \triangleq \left\{ \pi \in \left(0, \left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2\right) \middle| W_T(\pi) = W_O(\pi) \ \text{ when}\, \phi \geq 1-2b \right\} \tag{A.46}$$

be the π that solves $W_T(\pi) = W_O(\pi)$ where $W_T(\pi)$ is given by Equation (A.31) with $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi))$, while $W_O(\pi)$ is given by Equation (A.32) with $\eta(b,\phi) = 0$. $\pi^{\circ}(b)$ is decreasing in b.

From the shapes of $W_T(\pi)$ and $W_O(\pi)$ described in Step 2, it follows that:

- if $b \leq b^{\circ}(X)$ where $b^{\circ}(X)$ is defined in Equation (A.38), we have that
  - if $\phi \in (0, \phi^{\circ}(b))$ with $\phi^{\circ}(b)$ defined in Equation (A.39), we have $W_O(\pi) > W_T(\pi)$ for every $\pi \in (0,1)$,
  - if $\phi \in [\phi^{\circ}(b), 1-2b)$, we have that
    - $W_O(\pi) > W_T(\pi)$ for $\pi \in (\pi^{\circ}(b,\phi), 1)$, where $\pi^{\circ}(b,\phi)$ is defined in Equation (A.45),
    - $W_O(\pi) \leq W_T(\pi)$ for $\pi \in (0, \pi^{\circ}(b,\phi)]$,
  - if $\phi \in [1-2b, 1)$, we have that
    - $W_O(\pi) = W_T(\pi)$ if $\pi \in \Big[\big(\frac{1}{X}/(1-\frac{1}{X})\big)^2, 1\Big)$,
    - $W_O(\pi) > W_T(\pi)$ if $\pi \in \Big(\pi^{\circ}(b), \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2\Big)$, where $\pi^{\circ}(b)$ is defined in Equation (A.46),
    - $W_O(\pi) \leq W_T(\pi)$ if $\pi \in (0, \pi^{\circ}(b)(b)]$,
- if $b > b^{\circ}(X)$, we have that
  - if $\phi \in (0, 1-2b)$ we have $W_O(\pi) > W_T(\pi)$ for every $\pi \in (0,1)$,
  - if $\phi \in [1-2b, 1)$ we have that
    - $W_O(\pi) = W_T(\pi)$ if $\pi \in \Big[\big(\frac{1}{X}/(1-\frac{1}{X})\big)^2, 1\Big)$,
    - $W_O(\pi) > W_T(\pi)$ if $\pi \in \Big(0, \big(\frac{1}{X}/(1-\frac{1}{X})\big)^2\Big)$.

■

Proof of Proposition 2.5. We proceed in 4 steps: Step 1 derives the privacy-concerned borrowers' ex-ante surplus when the lender employs a transparent and an opaque algorithm, Step 2 compares their surplus across regimes, Step 3 derives the strategic borrowers' ex-ante surplus when the lender employs a transparent and an opaque algorithm, Step 4 compares their surplus across regimes.

Preliminaries. Consider the strategic borrower's data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Equation (2.12) and Equation (2.13), respectively.
The privacy-concerned borrower's ex-ante equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} V_{\tau}^{B,p}(z) &\triangleq \frac{1}{2}\int_{-1}^{1} V_{\lambda}^{B}(\varnothing, z)\,\mathrm{d}\lambda, \\ &= \int_{0}^{1} V_{\lambda}^{B}(\varnothing, z)\,\mathrm{d}\lambda, \end{aligned} \tag{A.47}$$

where $V_\lambda^B(\varnothing, z)$ is defined in Equation (A.7), while the second line of Equation (A.47) follows from the borrower's profit function being even in λ.
The strategic borrower's ex-ante equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$V_\tau^{B,s}(z) \triangleq \frac{1}{2} \int_{-1}^{1} V_\lambda^B(m_Q(z), z)\, \mathrm{d}\lambda, \tag{A.48}$$

$V_\lambda^B(z, z) = V_\lambda^B(z)$ is defined in Equation (A.3) while $V_\lambda^B(\varnothing, z)$ is defined in Equation (A.7).

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 2.3, we get the privacy-concerned borrowers' surpluses from a transparent and an opaque algorithm. These are, respectively:

$$V_T^{B,p}(z) = \int_0^{\widetilde{\lambda}(\pi)} (1-\phi)\mu_\lambda(z) \left(X - \frac{1}{\mu_\lambda(\gamma(\pi))} \right) + b \, \mathrm{d}\lambda \tag{A.49}$$

and

$$V_O^{B,p}(z) = \int_0^{1} (1-\phi)\mu_\lambda(z) \left(X - \frac{1}{\mu_\lambda(\frac{1}{2})} \right) + b \, \mathrm{d}\lambda \tag{A.50}$$

Step 2. Note that the integrand in Equation (A.50) is strictly bigger than the integrand in Equation (A.49) since when $\lambda > 0$ we have $\mu_\lambda(\frac{1}{2}) > \mu_\lambda(\gamma(\pi))$ as $\gamma(\pi) < \frac{1}{2}$. Moreover, both integrands are strictly positive and $\widetilde{\lambda}(\pi) \leq 1$. It follows that $V_O^{B,p}(z) > V_T^{B,p}(z)$.

Step 3. We now compute the ex-ante surplus of strategic borrowers, conditional on z.

When the lender uses a transparent algorithm, the set of borrowers that withhold data is $Q = \mathcal{G}(\lambda, \pi)$, given in Proposition 1. We can distinguish three types of borrowers:

- Borrowers with $z \in [0, \gamma(\pi)]$ withhold information when $\lambda > 0$ and obtain credit only if $r(\lambda) < \gamma(\pi)$, i.e. if $\lambda < \widehat{\lambda}(\gamma(\pi))$, while they are denied credit when $\lambda \geq \widehat{\lambda}(\gamma(\pi))$. If $\gamma(\pi) > \frac{1}{X}$ they always obtain credit, i.e. $\widehat{\lambda}(\gamma(\pi)) > 1$. Moreover, they disclose information and obtain credit whenever $\lambda < 0$. It follows that their surplus conditional on λ is

$$V_\lambda^B(m_Q(z), z) = \begin{cases} (1-\phi)\mu_\lambda(z) \left(X - \dfrac{1}{\mu_\lambda(z)} \right) + b & \text{if } \lambda \in [-1, 0) \\ \mathbb{1}\{\lambda < \widetilde{\lambda}(\pi)\} \left[(1-\phi)\mu_\lambda(z) \left(X - \dfrac{1}{\mu_\lambda(\gamma(\pi))} \right) + b \right] & \text{if } \lambda \in [0, 1], \end{cases}$$

  where $\widetilde{\lambda}(\pi) = \min\{\widehat{\lambda}(\gamma(\pi)), 1\}$.

- Borrowers with $z \in [1-\gamma(\pi), 1]$ withhold information when $\lambda < 0$ and obtain credit only if $1 - \gamma(\pi) < r(\lambda)$, i.e. if $\lambda > -\widehat{\lambda}(\gamma(\pi))$, while they are denied credit when $\lambda \leq -\widehat{\lambda}(\gamma(\pi))$. If $1 - \gamma(\pi) < 1 - \frac{1}{X}$ they always obtain credit and $-\widehat{\lambda}(\gamma(\pi)) < -1$.

Moreover, they disclose information and obtain credit whenever $\lambda > 0$. It follows that their surplus conditional on λ is

$$V_\lambda^B(m_Q(z), z) = \begin{cases} \mathbb{1}\left\{\lambda > -\widetilde{\lambda}(\pi)\right\} \left[(1-\phi)\mu_\lambda(z)\left(X - \dfrac{1}{\mu_\lambda(1-\gamma(\pi))}\right) + b\right] & \text{if } \lambda \in [-1, 0) \\ (1-\phi)\mu_\lambda(z)\left(X - \dfrac{1}{\mu_\lambda(z)}\right) + b & \text{if } \lambda \in [0, 1]. \end{cases}$$

- Borrowers with $z \in (\gamma(\pi), 1-\gamma(\pi))$ disclose information and obtain credit only if $z > r(\lambda)$ when $\lambda > 0$ and when $z < r(\lambda)$ when $\lambda < 0$, that is when $\lambda \in (\lambda_-(z), \lambda_+(z))$, where

$$\lambda_-(z) \triangleq \begin{cases} -1 & \text{if } z \in \left[0, 1 - \dfrac{1}{X}\right] \\ \widehat{\lambda}(z) & \text{if } z \in \left(1 - \dfrac{1}{X}, 1\right], \end{cases}$$

$$\lambda_+(z) \triangleq \begin{cases} \widehat{\lambda}(z) & \text{if } z \in \left[0, \dfrac{1}{X}\right) \\ 1 & \text{if } z \in \left[\dfrac{1}{X}, 1\right]. \end{cases} \tag{A.51}$$

It follows that their surplus conditional on λ is

$$V_\lambda^B(m_Q(z), z) = \mathbb{1}\left\{\lambda \in (\lambda_-(z), \lambda_+(z))\right\} \left[(1-\phi)\mu_\lambda(z)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right].$$

Integrating over λ as in Equation (A.48), we can write the strategic borrowers' ex-ante surplus under transparency as

$$V_T^{B,s}(z) = \begin{cases} \displaystyle\int_{-1}^{0} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda + \\ \quad \displaystyle + \int_{0}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \gamma(\pi)\big] \\ \displaystyle\int_{\lambda_-(z)}^{\lambda_+(z)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big(\gamma(\pi), 1-\gamma(\pi)\big) \\ \displaystyle\int_{-\widetilde{\lambda}(\pi)}^{0} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(1-\gamma(\pi))}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda + \\ \quad \displaystyle + \int_{0}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right] \frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[1-\gamma(\pi), 1\big], \end{cases}$$

$$= V^B(z) + O^B(z, \pi) \geq V^B(z), \tag{A.52}$$

where $V^B(z)$ is the expected borrower's surplus from full disclosure and is defined in Equation (A.11) while

$$O^B(z,\pi) \triangleq \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda + \\ \quad + \displaystyle\int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \gamma(\pi)\big] \\ 0 & \text{if } z \in \big(\gamma(\pi), 1-\gamma(\pi)\big) \\ \displaystyle\int_{-\widetilde{\lambda}(\pi)}^{\widehat{\lambda}(z)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(1-\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda + \\ \quad + \displaystyle\int_{\widehat{\lambda}(z)}^{0} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(1-\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[1-\gamma(\pi), 1\big], \end{cases} \tag{A.53}$$

with $O^B(z,\pi) > 0$ for $z \in [0, \gamma(\pi)) \cup (1-\gamma(\pi), 1]$ since every integrand is positive as we have that i) $\mu_\lambda(\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda > 0$, which also implies that $\mu_\lambda(\gamma(\pi))X - 1 > \mu_\lambda(z)X - 1 > 0$ for $\lambda > \widehat{\lambda}(z)$, and ii) $\mu_\lambda(1-\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda < 1$, which also implies that $\mu_\lambda(1-\gamma(\pi))X - 1 > \mu_\lambda(z)X - 1 > 0$ for $\lambda < -\widehat{\lambda}(z)$.

When the lender uses an opaque algorithm, the set of borrowers that withhold data is $Q = \mathcal{H}(b\,\phi)$, given in Proposition 2.2. Borrowers with $z \in [0, \eta(b,\phi))$ share data and obtain credit only if $\lambda < \widehat{\lambda}(z)$, borrowers with $z \in (1-\eta(b,\phi), 1]$ share data and obtain credit only if $\lambda > \widehat{\lambda}(z)$, instead borrowers with $z \in [\eta(b,\phi), 1-\eta(b,\phi)]$ withhold data and always obtain credit. Integrating over λ as in Equation (A.48), we can write the strategic borrowers' surplus under opacity as

$$\begin{aligned} V_O^{B,s}(z) &= \begin{cases} \displaystyle\int_{-1}^{\widehat{\lambda}(z)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[0, \eta(b,\phi)\big) \\ \displaystyle\int_{-1}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\frac{1}{2}}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big[\eta(b,\phi), 1-\eta(b,\phi)\big] \\ \displaystyle\int_{\widehat{\lambda}(z)}^{1} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \big(1-\eta(b,\phi), 1\big], \end{cases} \\ &= \max\left\{V^B(z), V^B(\varnothing)\right\} \end{aligned} \tag{A.54}$$

where $V^B(z)$ is the expected borrower's surplus from full disclosure defined in Equation (A.11) while $V^B(\varnothing)$ is the expected borrower's surplus from no disclosure defined in Equation (A.14).

Step 4. Let

$$\Delta^{B,s}(z) \triangleq V_T^{B,s}(z) - V_O^{B,s}(z)$$

be the difference between the strategic borrower's surplus with a transparent and an opaque algorithm. We now study the sign of $\Delta^{B,s}(z)$, for $z \leq \frac{1}{2}$. The analysis is symmetric for $z \geq \frac{1}{2}$. We distinguish 2 cases.

Case 1 ($\phi \geq 1 - 2b$). For $\phi \geq 1 - 2b$ we have $\eta(b, \phi) = 0$.

Consider first $\pi \geq \widehat{\pi}(X)$ where

$$\widehat{\pi}(X) \triangleq \left(\frac{\frac{1}{X}}{1 - \frac{1}{X}} \right)^2 \in (0, 1). \tag{A.55}$$

For these parameter values we have $\gamma(\pi) \geq \frac{1}{X}$, so $\widetilde{\lambda}(\pi) = 1$ and $\lambda_-(z) = -1$ for all $z > \gamma(\pi)$ and $\lambda_+(z) = 1$ for all $z < 1 - \gamma(\pi)$. Noting that the second piece of Equation (A.54) can be rewritten as

$$\int_{-1}^{1} \left[\mu_\lambda(z)(1 - \phi) \left(X - \frac{1}{\frac{1}{2}} \right) + b \right] \frac{1}{2} \, \mathrm{d}\lambda = \int_{-1}^{1} \left[\mu_\lambda(z)(1 - \phi) \left(X - \frac{1}{\mu_\lambda(z)} \right) + b \right] \frac{1}{2} \, \mathrm{d}\lambda \tag{A.56}$$

we have

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_{0}^{1} \mu_\lambda(z)(1 - \phi) \left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))} \right) \frac{1}{2} \, \mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \frac{1}{2}]. \end{cases} \tag{A.57}$$

This is weakly positive and strictly so for $z \in [0, \gamma(\pi))$ since $\mu_\lambda(\gamma(\pi)) > \mu_\lambda(z)$ for $\lambda > 0$.

Consider now $\pi < \widehat{\pi}(X)$, that is $\gamma(\pi) < \frac{1}{X}$ so that $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$ and $\lambda_+(z) \leq 1$ and $\lambda_-(z) \geq -1$. Using again Equation (A.56) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_{0}^{\widehat{\lambda}(\gamma(\pi))} \left[\mu_\lambda(z)(1 - \phi) \left(X - \frac{1}{\mu_\lambda(\gamma(\pi))} \right) + b \right] \frac{1}{2} \, \mathrm{d}\lambda \; + & \\ \qquad - \displaystyle\int_{0}^{1} \left[\mu_\lambda(z)(1 - \phi) \left(X - \frac{1}{\mu_\lambda(z)} \right) + b \right] \frac{1}{2} \, \mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ - \displaystyle\int_{\widehat{\lambda}(z)}^{1} \left[\mu_\lambda(z)(1 - \phi) \left(X - \frac{1}{\mu_\lambda(z)} \right) + b \right] \frac{1}{2} \, \mathrm{d}\lambda & \text{if } z \in \left(\gamma(\pi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{A.58}$$

The sign of $\Delta^{B,s}(z)$ depends on the value of π and z.

First, for $z \in (\gamma(\pi), \frac{1}{X})$ we have $\Delta^{B,s}(z) < 0$. To see this note that for $z \in (\gamma(\pi), \frac{1}{X})$ we have $V_T^{B,s}(z) = V^B(z)$ where $V^B(z)$ is the borrowers' expected utility from data sharing under an opaque algorithm and is defined in Equation (A.11) while $V_O^{B,s}(z) = V^B(\varnothing)$ where $V^B(\varnothing)$ is the borrowers' utility from data withholding under an opaque algorithm defined in Equation (A.14). It follows that $\Delta^{B,s}(z) = V^B(z) - V^B(\varnothing) < 0$ since for $\phi \geq 1 - 2b$ it is strictly optimal to withhold data for $z \in [0, \frac{1}{X})$ as shown in Proposition 2.2.

Second, for $z \in [0, \gamma(\pi)]$ we have that $\Delta^{B,s}(z)$ is strictly decreasing in z since its first-order derivative on that interval is

$$\frac{\partial \Delta^{B,s}}{\partial z} = \frac{1}{2}(1-\phi)X\left[\left(\int_0^{\widehat{\lambda}(\gamma(\pi))} \lambda \mathrm{d}\lambda - \int_0^1 \lambda \mathrm{d}\lambda\right) - \int_0^{\widehat{\lambda}(\gamma(\pi))} \lambda \frac{1}{X}\frac{1}{\mu_\lambda(\gamma(\pi))}\mathrm{d}\lambda\right] < 0, \tag{A.59}$$

since the term in brackets is negative as $\widehat{\lambda}(\gamma(\pi)) < 1$ while the integrand in the second addend is strictly positive.

Third, for $z \in [0, \gamma(\pi)]$ we have that $\Delta^{B,s}(z)$ is strictly increasing in π since

$$\begin{aligned}\frac{\partial \Delta^{B,s}}{\partial \pi} &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[\mu_{\widehat{\lambda}(\gamma(\pi))}(z)(1-\phi)\left(X - \frac{1}{\mu_{\widehat{\lambda}(\gamma(\pi))}}\right) + b\right]\frac{1}{2} + \\ &\quad + \int_0^{\widehat{\lambda}(\gamma(\pi))} (1-\phi)\frac{\mu_\lambda(z)}{\big(\mu_\lambda(\gamma(\pi))\big)^2}\lambda\gamma'(\pi)\,\mathrm{d}\lambda \\ &= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\frac{1}{2}b + \int_0^{\widehat{\lambda}(\gamma(\pi))} (1-\phi)\frac{\mu_\lambda(z)}{\big(\mu_\lambda(\gamma(\pi))\big)^2}\lambda\gamma'(\pi)\,\mathrm{d}\lambda > 0,\end{aligned} \tag{A.60}$$

where the first and second lines use Leibnitz integral rule, the last line follows from the fact that

$$\mu_{\widehat{\lambda}(\gamma(\pi))} = \frac{1}{2} + \frac{\frac{1}{2} - \frac{1}{X}}{\frac{1}{2} - \gamma(\pi)}\left(\gamma(\pi) - \frac{1}{2}\right) = \frac{1}{X},$$

while the inequality follows from Equation (A.42) and the fact that the integrand is positive since $\gamma'(\pi) > 0$.

Forth, since we are considering $\pi \in (0, \widehat{\pi}(X))$ we can evaluate $\Delta^{B,s}(0)$ at its extreme values of π to get

$$\begin{aligned}\lim_{\pi \to 0} \Delta^{B,s}(0) &= -\int_{\overline{\lambda}}^1 \Big[(1-\phi)\big(\mu_\lambda(0)X - 1\big) + b\Big]\frac{1}{2}\,\mathrm{d}\lambda < 0 \\ \lim_{\pi \to \left(\frac{1}{X}/\left(1-\frac{1}{X}\right)\right)^2} \Delta^{B,s}(0) &= \int_0^1 \mu_\lambda(0)(1-\phi)\left(\frac{1}{\mu_\lambda(0)} - \frac{1}{\mu_\lambda(\frac{1}{X})}\right)\frac{1}{2}\,\mathrm{d}\lambda > 0.\end{aligned} \tag{A.61}$$

Since $\Delta^{B,s}(0)$ is strictly increasing in π (see Equation (A.60)), there exists a threshold $\pi^*(b,\phi) \in (0, \widehat{\pi}(X))$ such that $\Delta^{B,s}(0) < 0$ if and only if $\pi < \pi^*(b,\phi)$. Let this threshold be defined as

$$\pi^*(b,\phi) \triangleq \left\{\pi \in \big(0, \widehat{\pi}(X)\big) \,\middle|\, \Delta^{B,s}(0) = 0\right\}. \tag{A.62}$$

Since $\Delta^{B,s}(z)$ is strictly decreasing in z (see Equation (A.59)), if $\pi < \pi^*(b,\phi)$ then $\Delta^{B,s}(z) < 0$ for every $z \in [0, \gamma(\pi)]$. Instead, since $\Delta^{B,s}(z)$ is strictly decreasing in z

for $z \in [0, \gamma(\pi)]$ and strictly increasing in π, when $\pi \geq \pi^*(b, \phi)$, there exists an increasing threshold $z^*(\pi)$ such that $\Delta^{B,s}(z) < 0$ if and only if $z > z^*(\pi)$. Let this threshold be defined as

$$z^*(\pi) \triangleq \left\{ z \in (0, \gamma(\pi)) \,\middle|\, \Delta^{B,s}(z) = 0 \text{ for } z \in [0, \gamma(\pi)] \right\}. \tag{A.63}$$

In summary, by symmetry of $\Delta^{B,s}(z)$ around $z = \frac{1}{2}$, when $\phi \geq 1 - 2b$ we have that:

- if $\pi \in (0, \pi^*(b, \phi))$, where $\pi^*(b, \phi)$ is defined in Equation (A.62), we have
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in [0, \frac{1}{X}) \cup (1 - \frac{1}{X}, 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1 - \frac{1}{X}]$;
- if $\pi \in [\pi^*(b, \phi), \widehat{\pi}(X))$, where $\widehat{\pi}(X)$ is defined in Equation (A.55), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, z^*(\pi)) \cup (1 - z^*(\pi), 1]$, where $z^*(\pi)$ is defined in Equation (A.63), and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in (z^*(\pi), \frac{1}{X}) \cup (1 - z^*(\pi), 1 - \frac{1}{X})$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1 - \frac{1}{X}] \cup \{z^*(\pi), 1 - z^*(\pi)\}$,
- if $\pi \in [\widehat{\pi}(X), 1)$ we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1 - \gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), 1 - \gamma(\pi)]$.

Case 2 ($\phi < 1 - 2b$). For $\phi < 1 - 2b$ we have $\eta(b, \phi) > 0$.

Consider $\pi \geq \widehat{\pi}(X)$ where $\widehat{\pi}(X)$ is defined in Equation (A.55), so that we have $\gamma(\pi) \geq \frac{1}{X}$ and $\widetilde{\lambda}(\pi) = 1$ and $\lambda_-(z) = -1$ for all $z > \gamma(\pi)$ and $\lambda_+(z) = 1$ for all $z < 1 - \gamma(\pi)$. Using Equation (A.56), we have

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda \, + & \\ \qquad \displaystyle + \int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \eta(b, \phi)) \\ \displaystyle\int_0^1 \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [\eta(b, \phi), \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \frac{1}{2}). \end{cases} \tag{A.64}$$

This is weakly positive and strictly so for $z \in [0, \gamma(\pi))$. To see this note that this $\Delta^{B,s}(z)$ is the same as in Equation (A.57) for $z \geq \eta(b, \phi)$, and we already prove that $\Delta^{B,s}(z) > 0$ for $z \in [\eta(b, \phi), \gamma(\pi))$ after Equation (A.57). For $z \in [0, \eta(b, \phi))$ note that $\Delta^{B,s}(z) = O^{B,s}(z) > 0$, where $O^{B,s}(z)$ is defined in Equation (A.53).

Consider now $\pi \in [\pi^{**}(b,\phi), \widehat{\pi}(X))$ where

$$\pi^{**}(b,\phi) \triangleq \left(\frac{\eta(b,\phi)}{1-\eta(b,\phi)}\right)^2 \in (0, \widehat{\pi}(X)). \tag{A.65}$$

For these parameter values we have $\eta(b,\phi) \leq \gamma(\pi) < \frac{1}{X}$ so that $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$ and $\lambda_+(z) \leq 1$ and $\lambda_-(z) \geq -1$. Using again Equation (A.56) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda \, + \\ \displaystyle\qquad + \int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \eta(b,\phi)) \\ \displaystyle\int_0^{\widehat{\lambda}(\gamma(\pi))} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda \, + \\ \displaystyle\qquad - \int_0^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [\eta(b,\phi), \gamma(\pi)] \\ \displaystyle -\int_{\widehat{\lambda}(z)}^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left(\gamma(\pi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{A.66}$$

Note that this $\Delta^{B,s}(z)$ is the same as in Equation (A.58) for $z \geq \eta(b,\phi)$. We already proved that $\Delta^{B,s}(z) < 0$ for $z \in (\gamma(\pi), \frac{1}{X})$ (see the discussion following Equation (A.58)). For $z \in [0, \eta(b,\phi))$ we have $\Delta^{B,s}(z) = O^{B,s}(z) > 0$ where $O^{B,s}(z)$ is defined in Equation (A.53). Since $\Delta^{B,s}(z)$ is continuous (as it is the difference of two continuous functions), decreasing in z for $z \in [\eta(b,\phi), \gamma(\pi)]$ (see Equation (A.59) for $z \in [0, \gamma(\pi)]$) and increasing in π for $z \in [\eta(b,\phi), \gamma(\pi)]$ (see Equation (A.60)) there exists an increasing $z^*(\pi) \in [\eta(b,\phi), \gamma(\pi)]$ defined in Equation (A.63) such that $\Delta^{B,s}(z) < 0$ if and only if $z > z^*(\pi)$.

Consider now $\pi < \pi^{**}(b,\phi)$ so that $\gamma(\pi) < \eta(b,\phi)$. Using again Equation (A.56) we get

$$\Delta^{B,s}(z) = \begin{cases} \displaystyle\int_0^{\widehat{\lambda}(z)} \mu_\lambda(z)(1-\phi)\left(\frac{1}{\mu_\lambda(z)} - \frac{1}{\mu_\lambda(\gamma(\pi))}\right)\frac{1}{2}\,\mathrm{d}\lambda \, + \\ \displaystyle\qquad + \int_{\widehat{\lambda}(z)}^{\widetilde{\lambda}(\pi)} \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(\gamma(\pi))}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in [0, \gamma(\pi)] \\ 0 & \text{if } z \in (\gamma(\pi), \eta(b,\phi)) \\ \displaystyle -\int_{\widehat{\lambda}(z)}^1 \left[\mu_\lambda(z)(1-\phi)\left(X - \frac{1}{\mu_\lambda(z)}\right) + b\right]\frac{1}{2}\,\mathrm{d}\lambda & \text{if } z \in \left[\eta(b,\phi), \frac{1}{X}\right) \\ 0 & \text{if } z \in \left[\frac{1}{X}, \frac{1}{2}\right]. \end{cases} \tag{A.67}$$

Once again $\Delta^{B,s}(z) = O^{B,s}(z) > 0$ for $z \in [0, \gamma(\pi)]$, where $O^{B,s}(z)$ is defined in Equation (A.53). Moreover, $\Delta^{B,s}(z) < 0$ for $z \in [\eta(b,\phi), \frac{1}{X})$, since in this interval we have $V_T^{B,s}(z) =$

$V^B(z)$ where $V^B(z)$ is the borrowers' expected utility from data sharing under an opaque algorithm and is defined in Equation (A.11) while $V_O^{B,s}(z) = V^B(\varnothing)$ where $V^B(\varnothing)$ is the borrowers' utility from data withholding under an opaque algorithm defined in Equation (A.14). It follows that $\Delta^{B,s}(z) = V^B(z) - V^B(\varnothing) < 0$ since for $z \in [\eta(b,\phi), \frac{1}{X})$ it is strictly optimal to withhold data as shown in Proposition 2.2.

In summary, by symmetry of $\Delta^{B,s}(z)$ around $z = \frac{1}{2}$, when $\phi < 1 - 2b$ we have that:

- if $\pi \in (0, \pi^{**}(b,\phi))$, where $\pi^{**}(b,\phi)$ is defined in Equation (A.65), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1-\gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), \eta(b,\phi)) \cup [\frac{1}{X}, 1 - \frac{1}{X}] \cup (1-\eta(b,\phi), 1-\gamma(\pi)]$, and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in [\eta(b,\phi), \frac{1}{X}) \cup (1-\frac{1}{X}, 1-\eta(b,\phi)]$;
- if $\pi \in [\pi^{**}(b,\phi), \widehat{\pi}(X))$, where $\widehat{\pi}(X)$ is defined in Equation (A.55), we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, z^*(\pi)) \cup (1-z^*(\pi), 1]$, where $z^*(\pi)$ is defined in Equation (A.63), and
  - $V_T^{B,s}(z) < V_O^{B,s}(z)$ for $z \in (z^*(\pi), \frac{1}{X}) \cup (1-z^*(\pi), 1-\frac{1}{X})$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\frac{1}{X}, 1-\frac{1}{X}] \cup \{z^*(\pi), 1-z^*(\pi)\}$,
- if $\pi \in [\widehat{\pi}(X), 1)$ we have
  - $V_T^{B,s}(z) > V_O^{B,s}(z)$ for $z \in [0, \gamma(\pi)) \cup (1-\gamma(\pi), 1]$, and
  - $V_T^{B,s}(z) = V_O^{B,s}(z)$ for $z \in [\gamma(\pi), 1-\gamma(\pi)]$.

Considering both Case 1 ($\phi \geq 1-2b$) and Case 2 ($\phi < 1-2b$) we can determine a threshold $z^\star(\pi)$ such that the surplus of a strategic borrower with data z is strictly higher under transparency if and only if $z \in [0, z^\star(\pi)) \cup (1-z^\star(\pi), 1]$. This threshold is

$$z^\star(\pi) \triangleq \begin{cases} \begin{cases} 0 & \text{if } \pi \in \big(0, \pi^*(b,\phi)\big) \\ z^*(\pi) & \text{if } \pi \in \big[\pi^*(b,\phi), \widehat{\pi}(X)\big) \\ \gamma(\pi) & \text{if } \pi \in \big[\widehat{\pi}(X), 1\big) \end{cases} & \text{if } \phi \geq 1-2b, \\ \begin{cases} \gamma(\pi) & \text{if } \big(0, \pi^{**}(b,\phi)\big) \\ z^*(\pi) & \text{if } \big[\pi^{**}(b,\phi), \widehat{\pi}(X)\big) \\ \gamma(\pi) & \text{if } \big[\widehat{\pi}(X), 1\big) \end{cases} & \text{if } \phi < 1-2b, \end{cases} \tag{A.68}$$

where $\pi^*(b,\phi)$, $\widehat{\pi}(X)$ and $\pi^{**}(b,\phi)$ are defined in Equations (A.62), (A.55) and (A.65), respectively, while $z^*(\pi)$ and $\gamma(\pi)$ are defined in Equations (A.63) and (2.12). ■

Proof of Proposition 2.6. We proceed in 3 steps: Step 1 derives the borrower's surplus when the lender employs a transparent and an opaque algorithm, Step 2 studies the graph of the borrower's surplus as a function of π; Step 3 compares the graphs.

Preliminaries. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Equation (2.12) and Equation (2.13), respectively. The borrower's equilibrium surplus from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} V_\tau^B(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} V_Q^B(\pi,\lambda)\,\mathrm{d}\lambda, \\ &= \int_0^1 V_Q^B(\pi,\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{A.69}$$

where

$$V_Q^B(\pi,\lambda) \triangleq \pi V_\lambda^B(\varnothing) + (1-\pi)\int_0^1 V_\lambda^B\big(m_Q(z)\big)\mathrm{d}z, \tag{A.70}$$

$V_\lambda^B(z)$ and $V_\lambda^B(\varnothing)$ are defined in Equation (A.3) and Equation (A.6), while the second line of Equation (A.69) follows from the borrower's profit function being even in λ.

Step 1. Proceeding as in Step 1 and Step 2 of Proposition 2.3, we get the borrower's surpluses from a transparent and an opaque algorithm. These are, respectively:

$$\begin{aligned} V_T^B(\pi) &= \widetilde{\lambda}(\pi)\ \left[(1-\phi)\left(\tfrac{1}{2}X-1\right)+b\right]+ \\ &+ (1-\pi)\int_{\widetilde{\lambda}(\pi)}^1 \int_{r(\lambda)}^1 (1-\phi)\big(\mu_\lambda(z)X-1\big)+b\ \ \mathrm{d}z\,\mathrm{d}\lambda \end{aligned} \tag{A.71}$$

and

$$\begin{aligned} V_O^B(\pi) &= \overline{\lambda}\ \left[(1-\phi)\left(\tfrac{1}{2}X-1\right)+b\right]+ \\ &+ \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \pi\left[(1-\phi)\left(\tfrac{1}{2}X-1\right)+b\right]+(1-\pi)\int_{r(\lambda)}^1 (1-\phi)\big(\mu_\lambda(z)X-1\big)+b\ \ \mathrm{d}z\,\mathrm{d}\lambda+ \\ &+ \int_{\widehat{\lambda}(\eta(b,\phi))}^1 \pi\left[(1-\phi)\left(\tfrac{1}{2}X-1\right)+b\right]+(1-\pi)\int_{\eta(b,\phi)}^1 (1-\phi)\big(\mu_\lambda(z)X-1\big)+b\ \ \mathrm{d}z\,\mathrm{d}\lambda. \end{aligned} \tag{A.72}$$

Step 2. We now study the graphs of the borrower's surplus in Equation (A.71) and Equation (A.72). For an arbitrary function $y : (0,1) \to \mathbb{R}$, we will use the simplified notation $y(0) = \lim_{\pi\to 0} y(\pi)$ and $y(1) = \lim_{\pi\to 1} y(\pi)$ to denote the limiting values of $y(\pi)$ as π approaches 0 and 1, respectively.

First, since $\widetilde{\lambda}(1) = 1$ and $\widetilde{\lambda}(0) = \overline{\lambda}$, we have that

$$V_O^B(1) = V_T^B(1) = (1-\phi)\left(\tfrac{1}{2}X-1\right)+b,$$

and

$$V_O^B(0) - V_T^B(0) = \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \int_{\eta(b,\phi)}^{r(\lambda)} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\, \mathrm{d}\lambda,$$
$$= \int_{\widehat{\lambda}\left(\eta(b,\phi)\right)}^{1} \left(r(\lambda) - \eta(b,\phi)\right) \left[(1-\phi)\left(\mu_\lambda\left(\tfrac{\eta(b,\phi)+r(\lambda)}{2}\right) X - 1\right) + b\right] \mathrm{d}\lambda > 0, \tag{A.73}$$

where $\mu_\lambda(\cdot)$ is defined in Equation (2.2) and the first lines uses Equation (A.22). To prove the inequality in Equation (A.73), notice that $\widehat{\lambda}\big(\eta(b,\phi)\big) < 1$ for $b > 0$ and the integrand in the last line is positive since $r(\lambda) \geq \eta(b,\phi)$ for $\lambda \geq \widehat{\lambda}\big(\eta(b,\phi)\big)$ and the term in square brackets is positive if

$$\frac{\eta(b,\phi) + r(\lambda)}{2} \geq \frac{1}{2} - \frac{1}{\lambda}\left(\frac{1}{2} - \frac{1}{X}\right) - \frac{1}{\lambda}\frac{1}{X}\frac{b}{1-\phi} = r(\lambda) - \frac{1}{\lambda}\frac{1}{X}\frac{b}{1-\phi},$$

which reduces to $\lambda \leq 1$ when $\eta(b,\phi) > 0$ (i.e. when $\frac{1}{X}\big(1 - 2b/(1-\phi)\big) > 0$) and to $\lambda \leq 1 - 2\frac{1}{X}\left(1 - \frac{2b}{1-\phi}\right)$ when $\eta(b,\phi) = 0$ (i.e. when $\frac{1}{X}\big(1 - 2b/(1-\phi)\big) \leq 0$), and are both satisfied.

Second, notice that $V_O^B(\pi)$ is flat and equal to $(1-\phi)\left(\frac{1}{2}X - 1\right) + b = V_O^B(1)$ if $\phi \geq 1 - 2b$ (i.e. when $\eta(b,\phi) = 0$ and $\widehat{\lambda}(\eta(b,\phi)) = \overline{\lambda}$) and is strictly decreasing if $\phi < 1 - 2b$ (i.e. when $\eta(b,\phi) > 0$ and $\widehat{\lambda}(\eta(b,\phi)) > \overline{\lambda}$), since

$$\frac{\partial V_O^B}{\partial \pi} = \int_{\overline{\lambda}}^{\widehat{\lambda}(\eta(b,\phi))} \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] - \int_{r(\lambda)}^{1} (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b \;\mathrm{d}z\,\mathrm{d}\lambda +$$
$$+ \int_{\widehat{\lambda}(\eta(b,\phi))}^{1} \left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] - \int_{\eta(b,\phi)}^{1} (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b \;\mathrm{d}z\,\mathrm{d}\lambda,$$
$$= \int_{\overline{\lambda}}^{1} \int_{0}^{\min\{r(\lambda),\eta(b,\phi)\}} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\,\mathrm{d}\lambda,$$
$$< \int_{\overline{\lambda}}^{1} \int_{0}^{\eta(b,\phi)} \left[(1-\phi)\left(\mu_\lambda(z)X - 1\right) + b\right] \mathrm{d}z\,\mathrm{d}\lambda,$$
$$= (1-\overline{\lambda})\eta(b,\phi)\left[(1-\phi)\left(\left(\frac{\overline{\lambda}+1}{2}\frac{\eta(b,\phi)}{2} + \left(1 - \frac{\overline{\lambda}+1}{2}\right)\frac{1}{2}\right)X - 1\right) + b\right] < 0, \tag{A.74}$$

where the third line follows from Equation (A.22) and the fact that $r(\lambda) \geq \eta(b,\phi)$ if and only if $\lambda \geq \widehat{\lambda}(\eta(b,\phi))$, while the last inequality reduces to $\phi < 1 - 2b$, the parameter values considered.

Third, $V_T^B(\pi)$ is flat and equal to $(1-\phi)\left(\frac{1}{2}X - 1\right) + b = V_T^B(1)$ if $\pi \geq \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2 \in (0,1)$ (i.e. when $\widetilde{\lambda}(\pi) = 1$) and is strictly convex for $\pi < \left(\frac{1}{X}/(1 - \frac{1}{X})\right)^2$ (i.e. when $\widetilde{\lambda}(\pi) = \widehat{\lambda}(\gamma(\pi)) < 1$) approaching the flat part from below. To prove that $V_T^B(\pi)$ is convex and approaches the flat part from below, rewrite it as

$$V_T^B(\pi) = \widehat{\lambda}(\gamma(\pi))\;\left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b\right] + (1-\pi)\int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda$$

where

$$G(\lambda) \triangleq \int_{r(\lambda)}^{1} (1-\phi)\big(\mu_\lambda(z)X - 1\big) + b \ \mathrm{d}z, \\ = (1-r(\lambda))\left[(1-\phi)\left(\mu_\lambda\left(\tfrac{r(\lambda)+1}{2}\right)X - 1\right) + b\right]. \tag{A.75}$$

The first derivative of $V_T^B(\pi)$ is :

$$\frac{\partial V_T^B}{\partial \pi} = \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\left[(1-\phi)\left(\tfrac{1}{2}X - 1\right) + b - (1-\pi)\,G\big(\widehat{\lambda}(\gamma(\pi))\big)\right] - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda,$$
$$= \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\Big[1-(1-\pi)(1-\gamma(\pi))\Big]b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda,$$
$$= \frac{\overline{\lambda}}{(1-\sqrt{\pi})^2}\,b - \int_{\widehat{\lambda}(\gamma(\pi))}^{1} G(\lambda)\,\mathrm{d}\lambda,$$

where the first line uses the Leibnitz integral rule, the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (A.75) and using Equation (A.25) and Equation (A.26), while the last line follows from the fact that

$$\widehat{\lambda}'(\gamma(\pi))\gamma'(\pi) = \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2} > 0, \\ (1-\pi)(1-\gamma(\pi)) = 1-\sqrt{\pi}. \tag{A.76}$$

Note that the left derivative of $V_T^B(\pi)$ at $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$ is

$$\frac{\partial V_T^B}{\partial \pi}\Big|_{\pi=\left(\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2} = \frac{\overline{\lambda}}{\left(1-\frac{\frac{1}{X}}{1-\frac{1}{X}}\right)^2}b > 0, \tag{A.77}$$

since $\widehat{\lambda}(\gamma(\pi)) = 1$ for $\pi = \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$.
The second derivative of $V_T^B(\pi)$ is

$$\frac{\partial^2 V_T^B}{\partial \pi^2} = \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^3}b + \widehat{\lambda}'(\gamma(\pi))\gamma'(\pi)\,G\big(\widehat{\lambda}(\gamma(\pi))\big),$$
$$= \frac{\overline{\lambda}}{\sqrt{\pi}(1-\sqrt{\pi})^2}\left[b\left(\tfrac{1}{1-\sqrt{\pi}} + 1 - \gamma(\pi)\right) + (1-\phi)\left(\tfrac{1}{2}X - 1\right)\left(1 + \frac{\sqrt{\pi}}{1-\sqrt{\pi}}\right)\right] > 0, \tag{A.78}$$

where the first line follows from Leibnitz integral rule, while the second line follows from plugging $\widehat{\lambda}(\gamma(\pi))$ in Equation (A.75) and using Equation (A.25), (A.26) and Equation (A.76).

Step 3. From the shapes of $V_T^B(\pi)$ and $V_O^B(\pi)$ described above, it follows that:

- if $\phi \geq 1-2b$, we have
  - $V_O^B(\pi) > V_T^B(\pi)$ for $\pi < \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,

– $V_O^B(\pi) = V_T^B(\pi)$ for $\pi \geq \left(\frac{1}{X}/(1-\frac{1}{X})\right)^2$,

- if $\phi < 1 - 2b$ we have $V_O^B(\pi) > V_T^B(\pi)$ for every $\phi \in (0,1)$.

■

A.2 Computation of Observables

This section provides closed-form expressions for variables observable to econometricians, as relevant to Section 2.7, such as the amount of data shared and overall credit provision.

Proposition A.1 (Observables)**.** *The amount of data shared in equilibrium with a transparent and an opaque algorithm are, respectively:*

$$\begin{aligned} S_T(\pi, X) &= (1-\pi)\left[\widetilde{\lambda}(\pi)\big(1-\gamma(\pi)\big) + \int_{\widetilde{\lambda}(\pi)}^{1} \big(1-r(\lambda)\big)\mathrm{d}\lambda\right], \\ S_O(\pi, X, b, \phi) &= (1-\pi)2\eta(b,\phi); \end{aligned} \tag{A.79}$$

while the levels of credit provision under transparency and opacity are, respectively:

$$\begin{aligned} I_T(\pi, X) &= \widetilde{\lambda}(\pi) + (1-\pi)\int_{\widetilde{\lambda}(\pi)}^{1}\int_{r(\lambda)}^{1} \mathrm{d}z\,\mathrm{d}\lambda, \\ I_O(\pi, X, b, \phi) &= \pi + (1-\pi)\left[\overline{\lambda} + \int_{\overline{\lambda}}^{1}\int_{\min\{r(\lambda),\eta(b,\phi)\}}^{1} \mathrm{d}z\,\mathrm{d}\lambda\right]. \end{aligned} \tag{A.80}$$

Proof of Proposition A.1. Consider a data sharing strategy of the form

$$m_Q(z) = \begin{cases} \varnothing & \text{if } z \in Q, \\ z & \text{if } z \notin Q, \end{cases}$$

where $Q \in \{\mathcal{G}(\lambda,\pi), \mathcal{H}(b,\phi)\}$ is the data-withholding set induced by a specific transparency regime, and where $\mathcal{G}(\lambda,\pi)$ and $\mathcal{H}(b,\phi)$ are defined in Proposition 1 and Proposition 2, respectively.

The ex-ante level of equilibrium data-sharing from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} S_\tau(\pi) &\triangleq \frac{1}{2}\int_{-1}^{1} S_Q(\pi,\lambda)\,\mathrm{d}\lambda, \\ &= \int_{0}^{1} S_Q(\pi,\lambda)\,\mathrm{d}\lambda, \end{aligned} \tag{A.81}$$

where

$$S_Q(\pi,\lambda) \triangleq (1-\pi)\int_0^1 \mathbb{1}\left\{m_Q(z) = z\right\}\mathrm{d}z, \tag{A.82}$$

where the second line of Equation (A.81) follows from $S_Q(\pi, \lambda)$ being even in λ. Proceeding as in Proposition 3, after simple algebra we obtain the expression in Equation (A.79).

The ex-ante equilibrium level of credit provision from a transparency regime $\tau \in \{T, O\}$ inducing a data-withholding set Q is:

$$\begin{aligned} I_\tau(\pi) &\triangleq \frac{1}{2} \int_{-1}^{1} I_Q(\pi, \lambda) \, \mathrm{d}\lambda, \\ &= \int_{0}^{1} I_Q(\pi, \lambda) \, \mathrm{d}\lambda, \end{aligned} \tag{A.83}$$

where

$$I_Q(\pi, \lambda) \triangleq \pi \ell_\lambda(\varnothing) + (1 - \pi) \int_{0}^{1} \ell_\lambda\big(m_Q(z)\big) \mathrm{d}z, \tag{A.84}$$

where $\ell_\lambda(z)$ and $\ell_\lambda(\varnothing)$ are defined in Equation (2.6) and Equation (A.4), while the second line of Equation (A.83) follows from $I_Q(\pi, \lambda)$ being even in λ. Proceeding as in Proposition 3, after simple algebra we obtain the expression in Equation (A.80). ■

Chapter B

Appendix for Chapter 3

B.1 Proofs

Proof of Lemma 3.1. Ignoring the constraints in problem (3.12), the FOC for s_1 is :

$$v'(s_1) = 1 - \frac{s_1}{\mu^*} = 0,$$

and is solved for $s_1 = \mu^*$. Since $\widehat{\mu} = \mu^*$ the constraints are satisfied at the unconstrained optimum. ■

Proof of Lemma 3.2. Consider the trading game at date 2 and let μ be the public posterior beliefs (that $\omega = h$) after a generic stress test result. Consider the following speculator's equilibrium trading strategy:

$$x_I(z) = \begin{cases} 1 & \text{if } z = h \\ 0 & \text{if } z = \varnothing \\ -1 & \text{if } z = l. \end{cases}$$

First, we determine the price $p(X)$ chosen by the market maker as a function of the order X. Notice that any order $x_I \notin \{-1, 1\}$ fully reveals the speculator's order who can therefore not make a trading profit. When the speculator is active, we can restrict attention to orders of size $x_I \in \{-1, 1\}$. When $X = (1, 1)$ the market maker infers that the informed speculator submitted a buy order and hence must have received a private signal $z = h$. It follows that the market will fund the bank ($a = 1$) at date 3. The market maker sets a price that reflects the speculator's private information and the supervisor's intervention decision, i.e. $P(1, 1) = E_h^1(1)$. Similarly, when $X = (-1, -1)$ the low state is revealed, the market maker sets $P(-1, -1) = E_l^0 = 0$. When $X \in \{(1, -1), (-1, 1)\}$ the order flow allows no inference over the speculator's private information and the market maker's posterior belief therefore remains equal to the prior, μ. When $X \in \{(1,0),(-1,0)\}$ the market maker understands that the speculator received an uninformative signal $z = \varnothing$ and abstained from trading. Again, the market maker does not update from the prior. It follows that, for $X \in \{(1,-1),(-1,1),(1,0),(-1,0)\} \equiv X_{\varnothing}$, i) at date 3, the market will make the funding decision contingent on the outcome of the stress test alone, as the equity price reveals no additional information; and ii) the market maker sets a price $P(X) = \mu \max\{E_h^1(\mu), 0\}$ (since when $E_h^1(\mu) < 0$ the capital providers will not fund the bank). Thus, the price schedule is

$$P(X) = \begin{cases} E_h^1(1) & \text{if } X = (1,1) \\ \mu \max\{E_h^1(\mu), 0\} & \text{if } X \in X_{\varnothing} \\ 0 & \text{if } X = (-1,-1). \end{cases}$$

Next, we compute the speculator's profits from the proposed trading strategy. If $z = h$, the speculator submits a buy order ($x_I = 1$). With probability 1/2 the liquidity trader trades in the same direction ($x_U = 1$ and $X = (1,1)$) and the speculator's private information is revealed to the market maker (and the investors) who then sets a price equal to $E_h^1(1)$. The speculator's trading profits are nil, $E_h^1(1) - P(1,1) = 0$. With probability 1/2 the liquidity trader trades in the opposite direction ($x_U = -1$ and $X = (-1,1)$), the speculator retains his private information and makes profits equal to $\max\{E_h^1(\mu), 0\} - P(-1,1) = (1-\mu)\max\{E_h^1(\mu), 0\}$. It follows that, given $z = h$ the speculator's expected trading profits are $\frac{1}{2}(1-\mu)\max\{E_h^1(\mu), 0\}$. If $z = l$, the speculator submits a sell order ($x_I = -1$) and, by the same reasoning trading profits are $\frac{1}{2}\mu \max\{E_h^1(\mu), 0\}$. If $z = \varnothing$ the speculator abstains from trading and makes profits equal to 0. In summary, the expected trading profits are

$$\pi(z,\mu) = \begin{cases} \frac{1}{2}(1-\mu)\max\{E_h^1(\mu), 0\} & \text{if } z = h \\ 0 & \text{if } z = \varnothing \\ \frac{1}{2}\mu \max\{E_h^1(\mu), 0\} & \text{if } z = l \end{cases}$$

Taking expectations over z, we get the expected equilibrium trading profits after the realization of the stress test result:

$$\begin{aligned} \pi(\mu) &= \mu\,\sigma\left(\frac{1}{2}(1-\mu)\max\{E_h^1(\mu), 0\}\right) + (1-\mu)\sigma\left(\frac{1}{2}\mu \max\{E_h^1(\mu), 0\}\right) \\ &= \sigma\mu(1-\mu)\max\{E_h^1(\mu), 0\}. \end{aligned}$$

We show that the proposed trading strategy is indeed optimal. Consider trading after a test result inducing a belief $\mu \geq \widehat{\mu}$. Given $z = h$ it is optimal to buy: abstaining from trading yields profits equal to $0 < \frac{1}{2}(1-\mu)E_h^1(\mu)$ and selling yields profits $\frac{1}{2}(\mu E_h^1(\mu) - E_h^1(\mu)) + \frac{1}{2}(0-0) = -\frac{1}{2}(1-\mu)E_h^1(\mu) < 0$. Given $z = l$ it is optimal to sell: abstaining from trading yields profits equal to $0 < \frac{1}{2}\mu E_h^1(\mu)$. If the speculator buys instead, the order flow will be either $X = (1,1)$ or $X = (-1,1)$. In either case, the bank gets funded and expected trading profits are $\frac{1}{2}(0 - E_h^1(1)) + \frac{1}{2}(0 - \mu E_h^1(\mu)) < 0$. If $z = \varnothing$ and the speculator buys, the bank will be funded and trading profits will be $\frac{1}{2}(\mu E_h^1(1) - E_h^1(1)) + \frac{1}{2}(\mu E_h^1(\mu) - \mu E_h^1(\mu)) < 0$. If the speculator sells instead, the order flow can be $X = (-1,1)$, in which case the bank obtains funding and has equity value $\mu E_h^1(\mu)$. Since this is equal to the price paid in this state, profits are zero. Instead, order flow may be $X = (-1,-1)$. The price will now be zero, the bank will not be funded and its equity value zero, yielding again zero trading profits.

Consider trading after a test result inducing a belief $\mu < \widehat{\mu}$. If the speculator sells, he always gets a price of zero, and there will never be funding so the equity value will also be zero. Hence, selling yields zero profits. If the speculator buys, with probability $\frac{1}{2}$ order flow will be $X = (-1, 1)$ in which case the price is zero, there will be no funding, and equity value will also be zero. With equal probability, the order flow will be $X = (1, 1)$, and the price equals $E_h^1(1)$ while the equity value would be lower and equal to $\mu E_h^1(1)$. If the speculator deviated to purchasing information and learned $\omega = h$, he makes zero trading profits. Hence, the deviation generated a loss, net of the information acquisition cost. If the speculator deviated to buying without a positive signal, the expected value of equity is below $E_h^1(\mu)$ so the speculator makes a loss. ■

Proof of Lemma 3.3. Consider all the outcomes m_i that induce beliefs $\mu_i < \widehat{\mu}$. All these m_i result in no information production by the speculator ($\sigma(\mu_i) = 0$) and no funding provision by the market ($a(\mu_i) = 0$). Hence, we can pool all these potential signals in a unique signal m_1 for all $s \in [0, s_1)$ with $\mu_1 < \widehat{\mu}$. The corresponding expected value for the supervisor generated by the outcome m_1 is

$$\begin{aligned} g(s_1) &= \frac{1}{2}\Big(\Pr(m_1|\omega = h)V_h^0 + \Pr(m_1|\omega = l)V_l^0 \Big) \\ &= \frac{1}{2}\Big(s_1^2 V_h^0 + (2s_1 - s_1^2)V_l^0 \Big). \end{aligned}$$

Now, consider all the outcomes m_i for $i \in \{2, 3, \ldots, n\}$. These outcomes induce posterior beliefs $\mu_i \geq \widehat{\mu}$ (otherwise we could have pooled the signal m_i with m_1). If, at date 2, the order flow is uninformative the market invests ($a(\mu_i) = 1$) at date 3. However, since the outcome induces a positive level $\sigma(\mu_i)$ of information production by the speculator, if the order flow reveals that $\omega = l$, the market does not invest at date 3 and chooses $a = 0$. If the state is $\omega = l$, order flow reveals it with probability $\frac{1}{2}\sigma(\mu_i)$. The corresponding expected bank value generated by some outcome m_i is

$$\begin{aligned} f(s_{i-1}, s_i) =& \frac{1}{2}\left(\Pr(m_i|\omega = h)V_h^1 + \Pr(m_i|\omega = l)\left(V_l^1 + \frac{1}{2}\sigma(\mu_i)\Delta V_l \right) \right) \\ =& \frac{1}{2}\left((s_i^2 - s_{i-1}^2)V_h^1 + \big(2(s_i - s_{i-1}) - (s_i^2 - s_{i-1}^2)\big)\left(V_l^1 + \frac{1}{2}\sigma(\mu_i)\Delta V_l \right) \right). \end{aligned}$$

The ex-ante expected value of the bank for a given stress test S can be written as

$$V(S) = g(s_1) + \sum_{i=2}^{n} f(s_{i-1}, s_i). \tag{B.1}$$

Note that, the second term in (B.1) is a telescoping sum where

$$\sum_{i=2}^{n} (s_i^2 - s_{i-1}^2) = (1 - s_1^2),$$

$$\sum_{i=2}^{n} \big(2(s_i - s_{i-1}) - (s_i^2 - s_{i-1}^2)\big) = 2(1 - s_1) - (1 - s_1^2).$$

It follows that the objective function can be written as

$$\begin{aligned}V(S) =& \frac{1}{2}\Big(s_1^2 V_h^0 + (2s_1 - s_1^2)V_l^0 + (1-s_1^2)V_h^1 + \big(2(1-s_1) - (1-s_1^2)\big)V_l^1\Big) + \\ &+ \frac{1}{2}\sum_{i=2}^{n}\big(2(s_i - s_{i-1}) - (s_i^2 - s_{i-1}^2)\big)\frac{1}{2}\sigma(\mu_i)\Delta V_l \\ =& \, v(s_1) + \frac{1}{2}\sum_{i=2}^{n}(s_i - s_{i-1})\left(1 - \frac{s_{i-1}+s_i}{2}\right)\sigma(\mu_i)\Delta V_l,\end{aligned}$$

where $v(s_1)$ is defined in (3.11). By applying the definitions in equation (3.2) and (3.15) we obtain the objective function in (3.18). ■

Proof of Proposition 3.1. We first introduce some notation. Let

$$S(a,b) \equiv \{a = s_0, s_1, \ldots, s_{n-1}, s_n = b\}$$

be a partition of the interval $[a,b] \subset \mathbb{R}$ such that $a = s_0 < s_1 < s_2 \cdots < s_{n-1} < s_n = b$. In our application we will have $0 \leq a < b \leq 1$ so that $S(a,b)$ can be thought of as a partition of a subspace of $[0,1]$. Let $\mathcal{S}(a,b)$ be the set of all possible partitions $S(a,b)$ over the interval $[a,b]$. Let $\overline{S}(a,b)$ be the finest partition in $\mathcal{S}[a,b]$, i.e. such that $n \to \infty$; and let $\underline{S}(a,b)$ be the coarsest partition in $\mathcal{S}[a,b]$, i.e. such that $n = 1$. Lastly, for some function $f : [a,b] \to \mathbb{R}$, we define

$$R\big(f, S(a,b)\big) \equiv \sum_{i=1}^{n}(s_i - s_{i-1})f\left(\frac{s_{i-1}+s_i}{2}\right)$$

as the midpoint Riemann sum of f with respect to the partition $S(a,b)$. In what follows, we will use the following properties of the midpoint Riemann sum (see, e.g., Davis and Rabinowitz (1984) p. 54):

- if f is convex over $[a,b]$ then $R(f, \overline{S}(a,b)) \geq R(f, S(a,b))$, $\forall\, S(a,b) \in \mathcal{S}(a,b)$;
- if f is concave over $[a,b]$ then $R(f, \underline{S}(a,b)) \geq R(f, S(a,b))$, $\forall\, S(a,b) \in \mathcal{S}(a,b)$.

We proceed in 3 steps. Step 1 establishes the general structure of the stress test. Step 2 simplifies the objective function in problem (3.17) and writes it as a function of two thresholds (s_1, s_2). Finally, Step 3 determines the optimal thresholds.

Step 1 (General Structure). Fix the optimal s_1 and assume it is interior, i.e. $s_1 \in (0, s_2)$ (this will be true at the optimum), consider a subset $[s_1, 1] \subseteq [0,1]$ and define the function $\widehat{\Sigma}(s) : [s_1, 1] \to \mathbb{R}$ as

$$\widehat{\Sigma}(s) = \frac{1}{2\tau}s(1-s)^2\left(\frac{1}{\mu^*} - \frac{1}{s}\right) = \frac{1}{2\tau}(1-s)^2\left(\frac{s}{\mu^*} - 1\right).$$

Consider a partition $S(s_1,1) \equiv \{s_1, s_2, s_3, \dots, s_{n-2}, s_{n-1}, 1\}$ and note that the stress test design problem (3.17), with $\widehat{\mu} = \mu^*$, is equivalent to

$$\begin{aligned} \max_{S(s_1,1)} \; & R(\widehat{\Sigma}, S(s_1,1)) \\ \text{s.t. } & \frac{s_1+s_2}{2} \geq \mu^* \\ & \frac{0+s_1}{2} < \mu^*. \end{aligned} \tag{B.2}$$

The first and second derivative of $\widehat{\Sigma}(s)$ are, respectively:

$$\begin{aligned} \widehat{\Sigma}'(s) &= \frac{1}{2\tau}\left[-2(1-s)\left(\frac{s}{\mu^*}-1\right)+(1-s)^2\frac{1}{\mu^*}\right], \\ \widehat{\Sigma}''(s) &= \frac{1}{2\tau}\left[2\left(\frac{s}{\mu^*}-1\right)-4(1-s)\frac{1}{\mu^*}\right], \end{aligned}$$

so that $\widehat{\Sigma}(s)$ attains its maximum at $s = \frac{1}{3}+\frac{2}{3}\mu^*$ and it is concave over $\left[s_1, \frac{2}{3}+\frac{1}{3}\mu^*\right]$ and convex over $\left(\frac{2}{3}+\frac{1}{3}\mu^*, 1\right]$. For some j, we have $S(s_1,1) = S(s_1,s_j) \cup S(s_j,1)$ and $R(\widehat{\Sigma}, S(s_1,1)) = R(\widehat{\Sigma}, S(s_1,s_j)) + R(\widehat{\Sigma}, S(s_j,1))$. For the properties of the midpoint Riemann sum introduced above, the partition that maximizes $R(\widehat{\Sigma}, S(s_1,1))$ is $S(s_1,1) = \underline{S}(s_1,s_j) \cup \overline{S}(s_j,1)$ for some optimally chosen s_j, that is the partition that solves problem (B.2) has a unique coarse sub-interval $[s_1,s_j]$ with $s_1 < s_j$ and a collection of infinitesimally small sub-intervals over the interval $[s_j,1]$. Relabelling $s_j = s_2$, we have that the optimal stress test has two coarse messages ($m_0 = s \in [0,s_1)$ and $m_1 = s \in [s_1,s_2)$) and a set of granular grades for $s \in [s_2,1]$.

Step 2 (Simplifying the Objective Function). Step 1 allows us to write the objective $V(S)$ as a function of the thresholds (s_1,s_2) only. Note that if the stress test S has fully granular grades for $s \in [s_2,1]$ we have $\nu_i = ds$ and $\mu_i = s$ for $i \in \{3,\dots,n\}$ where $n \to \infty$. It follows that the objective function in problem (3.17) reduces to:

$$V(s_1,s_2) = v(s_1) + \frac{1}{2\tau}\left(\nu_2\mu_2(1-\mu_2)^2\left(\frac{1}{\mu^*}-\frac{1}{\mu_2}\right) + \int_{s_2}^{1} s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\mathrm{d}s\right).$$

Adding and subtracting $\int_{s_1}^{s_2} s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\mathrm{d}s$ on the right-hand side we get

$$V(s_1,s_2) = v(s_1) + \frac{1}{2\tau}\left(\int_{s_1}^{1} s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\mathrm{d}s + A(s_1,s_2)\right)$$

where

$$\begin{aligned}
A(s_1,s_2) &\equiv \nu_2\mu_2(1-\mu_2)^2\left(\frac{1}{\mu^*}-\frac{1}{\mu_2}\right)-\int_{s_1}^{s_2} s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\mathrm{d}s,\\
&= \int_{s_1}^{s_2}\left[\mu_2(1-\mu_2)^2\left(\frac{1}{\mu^*}-\frac{1}{\mu_2}\right)-s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\right]\mathrm{d}s,\\
&= \int_{s_1}^{s_2}\left[\frac{1}{\mu^*}\Big(\mu_2(1-\mu_2)^2-s(1-s)^2\Big)-\Big((1-\mu_2)^2-(1-s)^2\Big)\right]\mathrm{d}s,\\
&= \left(\frac{1}{\mu^*}\frac{1}{12}(s_2-s_1)^3(2-3\mu_2)\right)-\left(-\frac{1}{12}(s_2-s_1)^3\right).
\end{aligned}$$

It follows that the objective function simplifies to:

$$V(s_1,s_2)=v(s_1)+\frac{1}{2\tau}\left(\int_{s_1}^{1} s(1-s)^2\left(\frac{1}{\mu^*}-\frac{1}{s}\right)\mathrm{d}s+\frac{1}{12}(\nu_2)^3\left(\frac{1}{\mu^*}(2-3\mu_2)+1\right)\right), \tag{B.3}$$

where $v(s_1)$ is defined in (3.11), while $\nu_2 = s_2 - s_1$ and $\mu_2 = \frac{s_1+s_2}{2}$, as defined in (3.2), and the stress test design problem in (3.17) simplifies to:

$$\begin{aligned}
&\max_{s_1,s_2} V(s_1,s_2)\\
&s.t.\ \mu_1<\mu^*\\
&\qquad \mu_2\geq\mu^*.
\end{aligned} \tag{B.4}$$

Step 3 (Optimal Thresholds). Neglecting the constraints in problem (B.4), the FOCs are:

$$\begin{aligned}
\frac{\partial V}{\partial s_1} &= v'(s_1)+\frac{1}{2\tau}\left[-s_1(1-s_1)^2\left(\frac{1}{\mu^*}-\frac{1}{s_1}\right)-\frac{3}{12}(\nu_2)^2\left(\frac{1}{\mu^*}(2-3\mu_2)+1\right)-\frac{3}{24}(\nu_2)^3\frac{1}{\mu^*}\right]=0,\\
\frac{\partial V}{\partial s_2} &= \frac{1}{2\tau}\left[\frac{3}{12}(\nu_2)^2\left(\frac{1}{\mu^*}(2-3\mu_2)+1\right)-\frac{3}{24}(\nu_2)^3\frac{1}{\mu^*}\right]=0,
\end{aligned}$$

and after simple algebra they simplify to:

$$\begin{aligned}
\frac{\partial V}{\partial s_1} &= v'(s_1)+\frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\mu^*}-1\right)-(\nu_2)^2\frac{1-\frac{1}{2}(s_2-\mu^*)-s_1}{2\mu^*}\right]=0,\\
\frac{\partial V}{\partial s_2} &= \frac{1}{2\tau}(\nu_2)^2\frac{1-\frac{1}{2}(s_1-\mu^*)-s_2}{2\mu^*}=0.
\end{aligned} \tag{B.5}$$

First, we show that when the constraints in problem (B.4) are slack the optimal stress test is a binary partition, $s_2 = 1$; we show that the partition features leniency, $s_1 < \mu^*$; and we provide the equation that implicitly defines the optimal s_1. Note that the second equation in (B.5) has two solutions: $s_2 = s_1$ and $s_2 = 1 - \frac{1}{2}(s_1 - \mu^*)$. The first solution is a stationary point where $\frac{\partial V}{\partial s_1} = \frac{\partial^2 V}{\partial s_1^2} = 0$ but is not a maximum. The second solution, $s_2 = 1 - \frac{1}{2}(s_1 - \mu^*)$, is greater than 1, if $s_1 < \mu^*$.

Suppose the optimal $s_1 < \mu^*$ and therefore the optimal $s_2 = 1$. If $s_2 = 1$, the first-order condition with respect to s_1 becomes:

$$\begin{aligned}\left.\frac{\partial V}{\partial s_1}\right|_{s_2=1} &= v'(s_1) + \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\mu^*} - 1\right) - (1-s_1)^2 \frac{1 - \frac{1}{2}(1-\mu^*) - s_1}{2\mu^*}\right] \\ &= 1 - \frac{s_1}{\mu^*} + \frac{1}{2\tau}(1-s_1)^2 \frac{\frac{3\mu^*-1}{2} - s_1}{2\mu^*} = 0,\end{aligned} \tag{B.6}$$

which is equivalent to (3.19). The first-order condition (B.6) has a unique solution in $s_1 \in [0,1]$, since the left-hand side of (the second line of) equation (B.6) is strictly decreasing. To see this differentiate it with respect to s_1 to get:

$$-\frac{1}{\mu^*} - \frac{3}{4\mu^*\tau}(1-s_1)(\mu^* - s_1)$$

which is strictly negative if

$$\tau > \frac{3}{4}(1-s_1)(s_1 - \mu^*).$$

This inequality is satisfied even when the right-hand side takes the highest possible value, i.e. $\frac{3}{4}\left(\frac{1-\mu^*}{2}\right)^2$ (attained at $s_1 = (1+\mu^*)/2$), since we have

$$\tau > \frac{1}{\mu^*}\left(\frac{1-\mu^*}{2}\right)^2 > \frac{3}{4}\left(\frac{1-\mu^*}{2}\right)^2$$

by assumption (3.16). Finally, using (B.6) and substituting $s_1 = \mu^*$ it is immediate that the resulting expression is negative, hence the solution to (B.6) features $s_1 < \mu^*$. Similarly, substituting $s_1 = 0$ yields a positive expression, implying that the solution to (B.6) features $s_1 > 0$.

To rule out a maximum where $\mu^* < s_1 < s_2 < 1$, substitute the candidate interior optimum $s_2 = 1 - \frac{1}{2}(s_1 - \mu^*)$ into the first expression of (B.5) to get

$$\begin{aligned}\frac{\partial V}{\partial s_1} =& v'(s_1) + \\ &+ \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\mu^*} - 1\right) - \left(1 - \frac{1}{2}(s_1 - \mu^*) - s_1\right)^2 \frac{1 - \frac{1}{2}(1 - \frac{1}{2}(s_1-\mu^*) - \mu^*) - s_1}{2\mu^*}\right].\end{aligned} \tag{B.7}$$

Note that $v'(s_1) < 0$ for $s_1 > \mu^*$. It is therefore sufficient to show that the expression in square brackets is negative. We have

$$-(1-s_1)^2\left(\frac{s_1}{\mu^*} - 1\right) - \left(1 - \frac{1}{2}(s_1-\mu^*) - s_1\right)^2 \frac{1 - \frac{1}{2}(1 - \frac{1}{2}(s_1-\mu^*) - \mu^*) - s_1}{2\mu^*} < 0,$$

if and only if

$$2(1-s_1)^2(s_1-\mu^*)+\frac{1}{2}\left(1-s_1-\frac{1}{2}(s_1-\mu^*)\right)^3>0.$$

By our hypothesis that $s_1>\mu^*$ it follows that $2(1-s_1)^2(s_1-\mu^*)>0$. We now show that the second addend is also positive. The term is decreasing in s_1, so showing that it is positive for the highest admissible value for s_1 is sufficient. Note that the constraint $s_2>s_1$ with $s_2=1-\frac{1}{2}(s_1-\mu^*)$ implies $s_1<\frac{2+\mu^*}{3}$. For $s_1=\frac{2+\mu^*}{3}$, the term $\left(1-s_1-\frac{1}{2}(s_1-\mu^*)\right)$ is equal to

$$1-\frac{2+\mu^*}{3}-\frac{1}{2}\left(\frac{2+\mu^*}{3}-\mu^*\right)=0.$$

Hence, we have $\forall s_1\in\left(\mu^*,1-\frac{1}{2}(s_1-\mu^*)\right)$,

$$\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1>\mu^*\\ s_2=1-\frac{1}{2}(s_1-\mu^*)}}<0,$$

hence the optimal s_1 must be below μ^*.

Lastly, we show that the constraints in problem (3.17) are satisfied at the unconstrained optimum. The first constraint $\mu_1=(0+s_1)/2<\mu^*$ is satisfied since we know $s_1<\mu^*$. The second constraint, $\mu_2=(s_1+1)/2\geq\mu^*$, is satisfied if $s_1\geq 2\mu^*-1$. Consider the following derivative:

$$\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=2\mu^*-1\\ s_2=1}}=1-\frac{2\mu^*-1}{\mu^*}+\frac{1}{2\tau}\frac{(1-\mu^*)^3}{\mu^*}.$$

Note that $\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=2\mu^*-1\\ s_2=1}}>0$ (and so $s_1>2\mu^*-1$) since $\mu^*<1$. Thus, we have $\mu_2>\mu^*$ and all the constraints are satisfied at the optimum.
It follows that the optimal stress test is a binary partition with $s_1<\ \mu^*$. ∎

Proof of Proposition 3.2. The optimal cutoff s_F is implicitly defined as the solution to equation (3.19), which can be rearranged as

$$\mu^*-s_F+\frac{1}{4\tau}K(s_F,\mu^*)=0, \tag{B.8}$$

where

$$K(s_F,\mu^*)\equiv(1-s_F)^2\left(\frac{3\mu^*-1}{2}-s_F\right).$$

Since the s_F that solves equation (B.8) is in the interval $\left(\frac{3\mu^*-1}{2},\mu^*\right)$ we have that $K(s_F,\mu^*)<0$. In what follows, we will also use the fact that

$$\begin{aligned}k(s_F,\mu^*)\equiv\frac{\partial K}{\partial s_F}&=2(1-s_F)(-1)\left(\frac{3\mu^*-1}{2}-s_F\right)+(1-s_F)^2(-1)\\&=-3(1-s_F)(\mu^*-s_F)<0.\end{aligned}$$

Employing the implicit function theorem, we totally differentiate equation (B.8) with respect to τ and μ^* to get:

$$\frac{\partial s_F}{\partial \tau} = \frac{\frac{1}{4\tau^2} K(s_F, \mu^*)}{\frac{1}{4\tau} k(s_F, \mu^*) - 1} > 0$$

$$\frac{\partial s_F}{\partial \mu^*} = \frac{1 + \frac{1}{4\tau}\frac{\partial K}{\partial \mu^*}}{1 - \frac{1}{4\tau} k(s_F, \mu^*)} > 0,$$

where the last line follows from the fact that $\frac{\partial K}{\partial \mu^*} = (1 - s_F)^2 \frac{3}{2} > 0$. ■

Proof of Lemma 3.4. Neglecting the constraints in problem (3.12), the FOC for s_1 is :

$$v'(s_1) = 1 - \frac{s_1}{\mu^*} = 0,$$

and is solved for $s_1 = \mu^*$. The constraint $\mu_1 = s_1/2 < \widehat{\mu}$ is satisfied for $s_1 = \mu^*$ since $\mu^* < \widehat{\mu}$. The constraint $\mu_2 \geq \widehat{\mu}$ is satisfied at the unconstrained optimum when $(\mu^* + 1)/2 \geq \widehat{\mu}$ and is binding otherwise. When the constraint is slack, the optimum is $s_1 = \mu^*$. When it is binding, s_1 is chosen to satisfy the constraint, i.e. $s_1 = 2\widehat{\mu} - 1$. Thus, the optimal stress test is a binary partition with cutoff $s_1 = \max\{2\widehat{\mu} - 1, \mu^*\}$. ■

Proof of Proposition 3.3. The proof proceeds as in Proposition 3.1 until Step 3, and the stress test design problem can be written as:

$$\begin{aligned} &\max_{s_1, s_2} V(s_1, s_2) \\ &s.t.\ \mu_1 < \widehat{\mu} \\ &\qquad \mu_2 \geq \widehat{\mu}, \end{aligned} \tag{B.9}$$

where

$$V(s_1, s_2) = v(s_1) + \frac{1}{2\tau}\left(\int_{s_1}^{1} s(1-s)^2 \left(\frac{1}{\widehat{\mu}} - \frac{1}{s}\right) \mathrm{d}s + \frac{1}{12}(\nu_2)^3 \left(\frac{1}{\widehat{\mu}}(2 - 3\mu_2) + 1\right)\right), \tag{B.10}$$

and $v(s_1)$ is defined in (3.11), while $\nu_2 = s_2 - s_1$ and $\mu_2 = \frac{s_1 + s_2}{2}$, as defined in (3.2).

Neglecting the constraints in problem (B.9), the FOCS, after some algebra, simplify to:

$$\begin{aligned} \frac{\partial V}{\partial s_1} &= v'(s_1) + \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}} - 1\right) - (\nu_2)^2 \frac{1 - \frac{1}{2}(s_2 - \widehat{\mu}) - s_1}{2\widehat{\mu}}\right] = 0, \\ \frac{\partial V}{\partial s_2} &= \frac{1}{2\tau}(\nu_2)^2 \frac{1 - \frac{1}{2}(s_1 - \widehat{\mu}) - s_2}{2\widehat{\mu}} = 0. \end{aligned} \tag{B.11}$$

First, we show that when the constraints in problem (B.9) are slack the optimal stress test is a binary partition, $s_2 = 1$, and we provide the equation that implicitly defines the optimal s_1. Note that the second equation has two solutions: $s_2 = s_1$ and $s_2 = 1 - \frac{1}{2}(s_1 - \widehat{\mu})$. The first solution is a stationary point where $\frac{\partial V}{\partial s_1} = \frac{\partial^2 V}{\partial s_1^2} = 0$ but is not a maximum. The second solution, $s_2 = 1 - \frac{1}{2}(s_1 - \widehat{\mu})$, is greater than 1, since at the optimum we have $s_1 < \widehat{\mu}$.

Suppose the optimal $s_1 < \widehat{\mu}$ and therefore the optimal $s_2 = 1$. If $s_2 = 1$, the first-order condition with respect to s_1 becomes:

$$\begin{aligned}\left.\frac{\partial V}{\partial s_1}\right|_{s_2=1} &= v'(s_1) + \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right) - (1-s_1)^2\frac{1-\frac{1}{2}(1-\widehat{\mu})-s_1}{2\widehat{\mu}}\right] \\ &= 1 - \frac{s_1}{\mu^*} + \frac{1}{2\tau}(1-s_1)^2\frac{\frac{3\widehat{\mu}-1}{2}-s_1}{2\widehat{\mu}} = 0,\end{aligned} \tag{B.12}$$

which is equivalent to (3.26). The first-order condition (B.12) has a unique solution in $s_1 \in [0,1]$, since the left-hand side of (the second line of) equation (B.12) is strictly decreasing. To see this differentiate it with respect to s_1 to get:

$$-\frac{1}{\mu^*} - \frac{3}{4\widehat{\mu}\tau}(1-s_1)(\widehat{\mu}-s_1)$$

which is strictly negative if

$$\tau > \frac{3}{4}\frac{\mu^*}{\widehat{\mu}}(1-s_1)(s_1-\widehat{\mu}).$$

This inequality is satisfied even when the right-hand side takes the highest possible value, i.e. $\frac{3}{4}\frac{\mu^*}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^2$ (attained at $s_1 = (1+\widehat{\mu})/2$), since we have

$$\tau > \frac{1}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^2 > \frac{3}{4}\frac{\mu^*}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^2$$

by assumption (3.16). Finally, using (B.12) and substituting $s_1 = \widehat{\mu}$ it is immediate to show that the resulting expression is negative, hence the solution to (B.12) features $s_1 < \widehat{\mu}$. Similarly, substituting $s_1 = 0$ yields a positive expression, implying that the solution to (B.12) features $s_1 > 0$.

To rule out a maximum where $\widehat{\mu} < s_1 < s_2 < 1$, substitute the candidate interior optimum $s_2 = 1 - \frac{1}{2}(s_1 - \widehat{\mu})$ into the first expression of (B.12) to get

$$\begin{aligned}\frac{\partial V}{\partial s_1} =& v'(s_1) + \\ & \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right) - \left(1-\frac{1}{2}(s_1-\widehat{\mu})-s_1\right)^2\frac{1-\frac{1}{2}(1-\frac{1}{2}(s_1-\widehat{\mu})-\widehat{\mu})-s_1}{2\widehat{\mu}}\right].\end{aligned} \tag{B.13}$$

This expression is similar to (B.7) with the exception that $\widehat{\mu}$ replaces μ^* in the expression in square brackets. Hence, the proof proceeds as in Proposition 3.1, and this yields that the optimal s_1 must be below $\widehat{\mu}$.

Second, we show that the constraints in problem (B.9) are satisfied at the unconstrained optimum when $\mu^* \geq \mu^\circ$ where

$$\mu^\circ \equiv \frac{2\widehat{\mu}-1}{1+\frac{1}{2\tau}\frac{(1-\widehat{\mu})^3}{\widehat{\mu}}}. \tag{B.14}$$

The first constraint $\mu_1 = (0 + s_1)/2 < \widehat{\mu}$ is satisfied since we know $s_1 < \widehat{\mu}$. The second constraint, $\mu_2 = (s_1+1)/2 \geq \widehat{\mu}$, is satisfied if $s_1 \geq 2\widehat{\mu}-1$. Consider the following derivative:

$$\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=2\widehat{\mu}-1\\ s_2=1}} = 1 - \frac{2\widehat{\mu}-1}{\mu^*} + \frac{1}{2\tau}\frac{(1-\widehat{\mu})^3}{\widehat{\mu}}.$$

Note that $\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=2\widehat{\mu}-1\\ s_2=1}} \geq 0$ (and so $s_1 \geq 2\widehat{\mu}-1$) if and only if $\mu^* \geq \mu^\circ$. Thus, we have that the constraint $\mu_2 \geq \widehat{\mu}$ is satisfied when $\mu^* \geq \mu^\circ$ and binding otherwise.

Lastly, we show that when the constraint $\mu_2 \geq \widehat{\mu}$ is binding the optimal stress test is a binary partition with cutoff $s_1 = 2\widehat{\mu}-1$. When the constraint is binding we have $s_1 = 2\widehat{\mu} - s_2$. Plugging this in the first-order derivative of s_2 we get

$$\left.\frac{\partial V}{\partial s_2}\right|_{s_1=2\widehat{\mu}-s_2} = \frac{1}{2\tau}(s_2-\widehat{\mu})^2\frac{2-\widehat{\mu}-s_2}{\widehat{\mu}} \geq 0 \quad \text{for } s_2 \leq 2-\widehat{\mu}.$$

Hence the objective function $V(2\widehat{\mu}-s_2, s_2)$ is weakly increasing in s_2 over the entire domain so that the optimum is $s_2 = 1$.

It follows that the optimal stress test is a binary partition with cutoff given by (3.25).

■

Proof of Proposition 3.4. When the constraint is not binding, the optimal cutoff is $s_F^D = \widetilde{s}$ and is implicitly defined as the solution to equation (3.26), which can be rearranged as

$$\mu^* - \widetilde{s} + \frac{1}{4\tau}H(\widetilde{s}, \mu^*, \widehat{\mu}) = 0, \tag{B.15}$$

where

$$H(\widetilde{s}, \mu^*, \widehat{\mu}) \equiv (1-\widetilde{s})^2\frac{\mu^*}{\widehat{\mu}}\left(\frac{3\widehat{\mu}-1}{2} - \widetilde{s}\right).$$

The sign of $H(\widetilde{s}, \mu^*, \widehat{\mu})$ depends on whether the stress test is lenient or conservative, in particular, $H(\widetilde{s}, \mu^*, \widehat{\mu}) > 0$ if and only if $\mu^* < \frac{3\widehat{\mu}-1}{2}$. To see this note that from Proposition 3 we have that $\mu^* < \frac{3\widehat{\mu}-1}{2}$ implies that $\widetilde{s} > \mu^*$, hence the $\widetilde{s}$ that solves (B.15) has to be such that $H(\widetilde{s}, \mu^*, \widehat{\mu}) > 0$. Similarly, $\mu^* > \frac{3\widehat{\mu}-1}{2}$ implies that $\widetilde{s} < \mu^*$ and $H(\widetilde{s}, \mu^*, \widehat{\mu}) < 0$.
In what follows, we make use of the fact that:

$$\begin{aligned} h(\widetilde{s}, \mu^*, \widehat{\mu}) \equiv \frac{\partial H}{\partial \widetilde{s}} &= 2(1-\widetilde{s})(-1)\frac{\mu^*}{\widehat{\mu}}\left(\frac{3\widehat{\mu}-1}{2} - \widetilde{s}\right) + (1-\widetilde{s})^2\frac{\mu^*}{\widehat{\mu}}(-1) \\ &= -3(1-\widetilde{s})\frac{\mu^*}{\widehat{\mu}}(\widehat{\mu}-\widetilde{s}) < 0, \end{aligned} \tag{B.16}$$

since we have $\widetilde{s} < \widehat{\mu}$ from Proposition 3.3.
To do comparative statics, we express equation (B.15) in terms of the primitives (p, R, D) to get:

$$\frac{1}{pR} - \widetilde{s} + \frac{1}{4\tau}H(\widetilde{s}, p, R, D) = 0, \tag{B.17}$$

where

$$H(\widetilde{s}, p, R, D) = (1-\widetilde{s})^2 \left(\frac{3}{2}\frac{1}{pR} - \left(\frac{1}{2}+\widetilde{s}\right)\left(1-\frac{D}{R}\right)\right),$$

and

$$\begin{aligned} \frac{\partial H}{\partial p} &= -(1-\widetilde{s})^2 \frac{3}{2}\frac{1}{p^2 R} < 0, \\ \frac{\partial H}{\partial R} &= -(1-\widetilde{s})^2 \left(\frac{3}{2}\frac{1}{pR^2} + \left(\frac{1}{2}+\widetilde{s}\right)\frac{D}{R^2}\right) < 0, \\ \frac{\partial H}{\partial D} &= (1-\widetilde{s})^2 \left(\frac{1}{2}+\widetilde{s}\right)\frac{1}{R} > 0. \end{aligned} \tag{B.18}$$

Employing the implicit function theorem, we totally differentiate equation (B.17) with respect to (p, R, D) to get:

$$\begin{aligned} \frac{\partial \widetilde{s}}{\partial p} &= \frac{\frac{1}{4\tau}\frac{\partial H}{\partial p} - \frac{1}{p^2 R}}{1 - \frac{1}{4\tau}h(\widetilde{s}, \mu^*, \widehat{\mu})} < 0, \\ \frac{\partial \widetilde{s}}{\partial R} &= \frac{\frac{1}{4\tau}\frac{\partial H}{\partial R} - \frac{1}{pR^2}}{1 - \frac{1}{4\tau}h(\widetilde{s}, \mu^*, \widehat{\mu})} < 0, \\ \frac{\partial \widetilde{s}}{\partial D} &= \frac{\frac{1}{4\tau}\frac{\partial H}{\partial D}}{1 - \frac{1}{4\tau}h(\widetilde{s}, \mu^*, \widehat{\mu})} > 0, \end{aligned}$$

where the inequalities follow from (B.16) and (B.18). Doing the same for τ we get:

$$\frac{\partial \widetilde{s}}{\partial \tau} = -\frac{\frac{1}{4\tau^2}H(\widetilde{s}, \mu^*, \widehat{\mu})}{1 - \frac{1}{4\tau}h(\widetilde{s}, \mu^*, \widehat{\mu})},$$

where $\frac{\partial \widetilde{s}}{\partial \tau} < 0$ if and only if $\mu^* < \frac{3\widehat{\mu}-1}{2}$ (as this implies $H(\widetilde{s}, \mu^*, \widehat{\mu}) > 0$), since $h(\widetilde{s}, \mu^*, \widehat{\mu}) < 0$ by (B.16).
When the constraint is binding, so that $s_F^D = 2\widehat{\mu} - 1$, the comparative statics are driven by $\widehat{\mu} = 1/p(R-D)$, and are thus the same as above. ■

Proof of Lemma 3.5. The stress test design problem is given by (3.12) but, to ensure that an optimum exists, the constrained set changes to $\mu_1 \leq \widehat{\mu}$ and $\mu_2 > \widehat{\mu}$, that is we want to solve:

$$\begin{aligned} &\max_{s_1} \, v(s_1) \\ &\text{s.t. } \mu_1 \leq \widehat{\mu}, \\ &\qquad \mu_2 > \widehat{\mu}, \end{aligned} \tag{B.19}$$

where $v(s_1)$ is defined in (3.11). Neglecting the constraints in problem (B.19), the FOC for s_1 is :

$$v'(s_1) = 1 - \frac{s_1}{\mu^*} = 0,$$

and is solved for $s_1 = \mu^*$. The constraint $\mu_2 = (s_1+1)/2 > \widehat{\mu}$ is satisfied for $s_1 = \mu^*$ since $\widehat{\mu} < \mu^*$. The constraint $\mu_1 = s_1/2 \leq \widehat{\mu}$ is satisfied at the unconstrained optimum when $\mu^*/2 \leq \widehat{\mu}$ and is binding otherwise. When the constraint is slack the optimum is $s_1 = \mu^*$, when it is binding, s_1 is chosen to satisfy the constraint, i.e. $s_1 = 2\widehat{\mu}$. Thus, the optimal stress test is a binary partition with cutoff $s_1 = \min\{2\widehat{\mu}, \mu^*\}$. ■

Proof of Proposition 3.5. When $\mu^* > \widehat{\mu}$, the stress test design problem is given by (3.17) but, to ensure that an optimum exists, the constrain set changes to $\mu_1 \leq \widehat{\mu}$ and $\mu_2 > \widehat{\mu}$. Hence, the stress test design problem is

$$\begin{aligned} &\max_{s_1,s_2} V(s_1,s_2) \\ &s.t.\ \mu_1 < \widehat{\mu} \\ &\qquad \mu_2 \geq \widehat{\mu}, \end{aligned} \tag{B.20}$$

where $V(s_1,s_2)$ is defined in (B.10), $v(s_1)$ is defined in (3.11), while $\nu_2 = s_2 - s_1$ and $\mu_2 = \frac{s_1+s_2}{2}$, as defined in (3.2).

The proof proceeds as in Proposition 3.1 until Step 3, and the FOCs are the ones in (B.11). We replicate the FOCs here for ease of exposition:

$$\begin{aligned} \frac{\partial V}{\partial s_1} &= v'(s_1) + \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right) - (\nu_2)^2\frac{1-\frac{1}{2}(s_2-\widehat{\mu})-s_1}{2\widehat{\mu}}\right] = 0, \\ \frac{\partial V}{\partial s_2} &= \frac{1}{2\tau}(\nu_2)^2\frac{1-\frac{1}{2}(s_1-\widehat{\mu})-s_2}{2\widehat{\mu}} = 0. \end{aligned}$$

First, we show that when $\mu^* < \underline{\mu}$, where

$$\underline{\mu} \equiv \frac{\widehat{\mu}}{1-\frac{1}{\tau}\frac{1}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^3}, \tag{B.21}$$

the optimal stress test is a binary partition, $s_2 = 1$, and we provide the equation that implicitly defines the optimal s_1. Suppose $s_2 > s_1$, the second equation is solved for $s_2 = 1 - \frac{1}{2}(s_1 - \widehat{\mu})$. If at the optimum we have $s_1 < \widehat{\mu}$, then it is optimal to set $s_2 = 1$. To determine when this is the case, we evaluate the sign of the following derivative:

$$\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=\widehat{\mu} \\ s_2=1-\frac{1}{2}(s_1-\widehat{\mu})}} = 1 - \frac{\widehat{\mu}}{\mu^*} - \frac{1}{2\tau}\frac{(1-\widehat{\mu})^3}{4\widehat{\mu}}.$$

We have that $\frac{\partial V}{\partial s_1}\Big|_{\substack{s_1=\widehat{\mu} \\ s_2=1-\frac{1}{2}(s_1-\widehat{\mu})}} < 0$ when $\mu^* < \underline{\mu}$. It follows that for these parameter values, we have $s_1 < \widehat{\mu}$ and $s_2 = 1$. The optimal s_1 solves:

$$\begin{aligned} \frac{\partial V}{\partial s_1}\Big|_{s_2=1} =& 1 - \frac{s_1}{\mu^*} + \frac{1}{2\tau}\left[-(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right) - (1-s_1)^2\frac{1-\frac{1}{2}(1-\widehat{\mu})-s_1}{2\widehat{\mu}}\right] \\ =& 1 - \frac{s_1}{\mu^*} + \frac{1}{2\tau}(1-s_1)^2\frac{\frac{3\widehat{\mu}-1}{2}-s_1}{2\widehat{\mu}} = 0, \end{aligned} \tag{B.22}$$

and thus is equal to $\widetilde{s}$, the unique solution to (3.26). The constraints are satisfied. The constraint $\mu_1 = (0+s_1)/2 \leq \widehat{\mu}$ is satisfied since at the optimum we have $s_1 < \widehat{\mu}$. The constraint $\mu_2 = (s_1+s_2) > \widehat{\mu}$ is also satisfied since at the optimum we have $s_1 > 2\widehat{\mu}-1$. To see this note that:

$$\frac{\partial V}{\partial s_1}\Big|_{\substack{s_1=2\widehat{\mu}-1 \\ s_2=1}} = 1 - \frac{2\widehat{\mu}-1}{\mu^*} + \frac{1}{2\tau}\frac{(1-\widehat{\mu})^3}{\widehat{\mu}} > 0,$$

since we have that $1-(2\widehat{\mu}-1)/\mu^* > 0$ as $\mu^* > \widehat{\mu} > 2\widehat{\mu}-1$, and the other term is also positive.

We now distinguish two cases when $\mu^* > \underline{\mu}$: Case A, where $\widehat{\mu} < 2/5$, and Case B, $\widehat{\mu} \geq 2/5$.
Consider Case A. We show that when $\mu^* > \underline{\mu}$ the optimal stress test contains two buckets with fully granular grades for resilience levels above the buckets, i.e. $s_2 < 1$. From the discussion above we already know that for these parameter values, we have $s_1 > \widehat{\mu}$ and $s_2 = 1 - \frac{1}{2}(s_1 - \widehat{\mu})$.
When the constraints are satisfied, the optimal optimal s_1 solves:

$$\frac{\partial V}{\partial s_1}\Big|_{s_2=1-\frac{1}{2}(s_1-\widehat{\mu})} = 1 - \frac{s_1}{\mu^*} - \frac{1}{2\tau}\left[(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right) + \frac{(1-\frac{3}{2}s_1+\frac{1}{2}\widehat{\mu})^3}{4\widehat{\mu}}\right] = 0. \tag{B.23}$$

The constraints are indeed satisfied when

$$\mu^* \leq \frac{2\widehat{\mu}}{1-\frac{1}{2\tau}\left((1-2\widehat{\mu})^2 + \frac{(1-\frac{5}{2}\widehat{\mu})^3}{4\widehat{\mu}}\right)} \equiv \mu_\circ. \tag{B.24}$$

To see this note that the constraint $\mu_2 = (s_1+s_2)/2 > \widehat{\mu}$ is satisfied since $s_2 \geq s_1 > \widehat{\mu}$, while the constraint $\mu_1 = s_1/2 \leq \widehat{\mu}$ is satisfied if $s_1 \leq 2\widehat{\mu}$ and this is the case when

$$\frac{\partial V}{\partial s_1}\Big|_{\substack{s_1=2\widehat{\mu} \\ s_2=1-\frac{1}{2}(s_1-\widehat{\mu})}} = 1 - \frac{2\widehat{\mu}}{\mu^*} + \frac{1}{2\tau}\left((1-2\widehat{\mu})^2 + \frac{(1-\frac{5}{2}\widehat{\mu})^3}{4\widehat{\mu}}\right) \leq 0,$$

which is equivalent to (B.24).
We can also verify that at the optimum $s_2 = 1 - \frac{1}{2}(s_1-\widehat{\mu}) > s_1$. This happens when $s_1 < \frac{2}{3} + \frac{1}{3}\widehat{\mu}$, which is the case when

$$\left.\frac{\partial V}{\partial s_1}\right|_{\substack{s_1=\frac{2}{3}+\frac{1}{3}\widehat{\mu} \\ s_2=1-\frac{1}{2}(s_1-\widehat{\mu})}} = 1-\frac{\frac{2}{3}+\frac{1}{3}\widehat{\mu}}{\mu^*}-\frac{1}{2\tau}\frac{2}{27}\frac{(1-\widehat{\mu})^3}{\widehat{\mu}}<0,$$

or equivalently, when

$$\mu^* < \frac{\dfrac{2}{3}+\dfrac{1}{3}\widehat{\mu}}{1-\dfrac{1}{\tau}\dfrac{1}{\widehat{\mu}}\left(\dfrac{1-\widehat{\mu}}{3}\right)^3} \equiv \overline{\mu}. \tag{B.25}$$

One can show that this condition is milder than (B.24) when $\widehat{\mu}<2/5$ and $\tau>\frac{1}{\widehat{\mu}}\left(\frac{1-\widehat{\mu}}{2}\right)^2$, as by assumption (3.16).
When the constraint $\mu_1 \leq \widehat{\mu}$ is binding, i.e. (B.24) is not satisfied, it is optimal to choose s_1 to satisfy the constraint, $s_1 = 2\widehat{\mu}$, and $s_2 = 1-\frac{1}{2}\widehat{\mu}$. Note that, once again, $s_2 > s_1$ if $\widehat{\mu} < 2/5$.

Consider now Case B. From the discussion above, we have that $s_1 < s_2$ when (B.25) is satisfied and the optimal stress test is as discussed above. When (B.25) is not satisfied, a coarse pass grade is not optimal anymore and thus we have $s_2 = s_1$. Plugging $s_2 = s_1$ in the FOC for s_1 we get the equation that implicitly defines the optimal s_1:

$$\left.\frac{\partial V}{\partial s_1}\right|_{s_2=s_1} = 1-\frac{s_1}{\mu^*}-\frac{1}{2\tau}(1-s_1)^2\left(\frac{s_1}{\widehat{\mu}}-1\right)=0. \tag{B.26}$$

The constraint $\mu_1 = s_1/2 \leq \widehat{\mu}$ is satisfied when

$$\mu^* \leq \frac{2\widehat{\mu}}{1-\frac{1}{2\tau}(1-2\widehat{\mu})^2} \tag{B.27}$$

since this implies that

$$\left.\frac{\partial V}{\partial s_1}\right|_{s_2=s_1=2\widehat{\mu}} = 1-\frac{2\widehat{\mu}}{\mu^*}-\frac{1}{2\tau}(1-2\widehat{\mu})^2 \leq 0. \tag{B.28}$$

When the constraint is binding, it is optimal to choose s_1 and s_2 to satisfy the constraint, $s_1 = s_2 = 2\widehat{\mu}$. ■

Lemma B.1. *When the supervisor can intervene according to her private information s, she solves the following stress test design problem:*

$$\max_{s_1\in[0,1]} V(s_1)$$

where

$$V(s_1) = \begin{cases} V_{a(i)}(s_1) & \text{if } s_1 \in \big[0, 2\mu^*-1\big), \\ V_{a(ii)}(s_1) & \text{if } s_1 \in \big[2\mu^*-1, \mu^*(1-\delta)\big), \\ V_p(s_1) & \text{if } s_1 \in \big[\mu^*(1-\delta), \mu^*(1+\delta)\big], \\ V_{b(i)}(s_1) & \text{if } s_1 \in \big(\mu^*(1+\delta), 2\mu^*\big), \\ V_{b(ii)}(s_1) & \text{if } s_1 \in \big[2\mu^*, 1\big]. \end{cases} \tag{B.29}$$

and $V_{a(i)}$, $V_{a(ii)}$, V_p, $V_{b(i)}$, $V_{b(ii)}$ are defined in (B.30)-(B.41).

Proof of Lemma B.1. The supervisor's value function depends on i) the default action taken by the market without a stock market signal, and, ii) whether the supervisor intervenes or not to revert the market's action. Accordingly, we distinguish several cases and first state the relevant functions $V(s_1)$ for each region. The proof of the expression (B.30)-(B.41) follows further below.

- region a(i), $s_1 \in [0, 2\mu^* - 1)$. Note that this region is non-empty only if $\mu^* > \frac{1}{2}$. Expected bank value is

$$\begin{aligned} V_{a(i)}(s_1) =& \frac{V_h^1 + V_l^1}{2} - \frac{1}{2}\Delta V_h \left(\frac{\sigma_2}{2}s_1^2 + \left(1 - \frac{\sigma_2}{2}\right)\bar{s}^2\right) \\ &+ \frac{1}{2}\Delta V_l \left(\frac{\sigma_2}{2} + \left(1 - \frac{\sigma_2}{2}\right)(2\bar{s} - \bar{s}^2)\right) - \delta\left(1 - \frac{\sigma_2}{2}\right)(1 - \bar{s}), \end{aligned} \tag{B.30}$$

  where σ_2 is the information produced by the speculator if the bank passes the test, given by

$$\sigma_2 = \frac{1}{4\tau}\left(1 - \bar{s}^2\right)\left(pR - \frac{1}{\frac{1+\bar{s}}{2}}\right). \tag{B.31}$$

  The speculator does not produce information if the bank fails the test.

- region a(ii), $s_1 \in [2\mu^* - 1, \mu^*(1-\delta))$.

$$\begin{aligned} V_{a(ii)}(s_1) =& \frac{V_h^1 + V_l^1}{2} - \frac{1}{2}\Delta V_h \left(\frac{\sigma_2}{2}s_1^2 + \left(1 - \frac{\sigma_2}{2}\right)\underline{s}^2\right) \\ &+ \frac{1}{2}\Delta V_l \left(\frac{\sigma_2}{2} + \left(1 - \frac{\sigma_2}{2}\right)(2\underline{s} - \underline{s}^2)\right) - \delta\left(1 - \frac{\sigma_2}{2}\right)(\underline{s} - s_1). \end{aligned} \tag{B.32}$$

  There is no information acquisition following a *fail* result, and information acquisition following a *pass* is given by

$$\sigma_2 = \frac{1}{4\tau}\left(1 - \underline{s}^2\right)\left(pR - \frac{1}{\frac{1+\underline{s}}{2}}\right). \tag{B.33}$$

  Note that σ_2 from (B.33) and (B.31) are the same. This can be shown using the expressions for $\underline{s}, \bar{s}$ and $\mu^* = \frac{1}{pR}$.

- region p, $s_1 \in [\mu^*(1-\delta), \mu^*(1+\delta))$;

$$\begin{aligned} V_p(s_1) =& \frac{V_h^1 + V_l^1}{2} - \frac{1}{2}s_1^2\Delta V_h \\ &+ \frac{1}{2}\Delta V_l \left(\frac{\sigma_2}{2} + \left(1 - \frac{\sigma_2}{2}\right)(2s_1 - s_1^2)\right). \end{aligned} \tag{B.34}$$

  There is no information acquisition following a *fail* result, and information acquisition following a *pass* is given by

$$\sigma_2 = \frac{1}{4\tau}\left(1 - s_1^2\right)\left(pR - \frac{1}{\frac{1+s_1}{2}}\right). \tag{B.35}$$

- region b(i), $s_1 \in [\mu^*(1+\delta), 2\mu^*)$.

$$\begin{aligned} V_{b(i)}(s_1) =& \frac{V_h^1 + V_l^1}{2} - \frac{1}{2}\Delta V_h \left(1 - \frac{\sigma_1}{2}\right) \bar{s}^2 \\ &+ \frac{1}{2}\Delta V_l \left(\frac{\sigma_1}{2}(2s_1 - s_1^2) + \left(1 - \frac{\sigma_1}{2}\right)(2\bar{s} - \bar{s}^2)\right) + \frac{1}{2}\Delta V_l \frac{\sigma_2}{2}(1 - s_1)^2 \\ &- \delta \left(1 - \frac{\sigma_1}{2}\right)(s_1 - \bar{s}). \end{aligned} \tag{B.36}$$

Since the bank is sometimes funded even when it fails the test, the speculator acquires information σ_1 following m_1 (and σ_2 following m_2):

$$\sigma_1 = \frac{1}{2\tau}\left(s_1^2 - \bar{s}^2\right)\frac{1 - \frac{s_1}{2}}{s_1}\left(pR - \frac{1}{\frac{s_1 + \bar{s}}{2}}\right), \tag{B.37}$$

$$\sigma_2 = \frac{1}{4\tau}\left(1 - s_1^2\right)\left(pR - \frac{1}{\frac{1+s_1}{2}}\right). \tag{B.38}$$

- region b(ii), $s_1 \in [2\mu^*, 1]$. Note that this region is non-empty only if $\mu^* < \frac{1}{2}$. We again get information acquisition following both a *fail* and a *pass* result.

$$\begin{aligned} V_{b(ii)}(s_1) =& \frac{V_h^1 + V_l^1}{2} - \frac{1}{2}\Delta V_h \left(1 - \frac{\sigma_1}{2}\right) \underline{s}^2 \\ &+ \frac{1}{2}\Delta V_l \left(\frac{\sigma_1}{2}(2s_1 - s_1^2) + \left(1 - \frac{\sigma_1}{2}\right)(2\underline{s} - \underline{s}^2)\right) + \frac{1}{2}\Delta V_l \frac{\sigma_2}{2}(1 - s_1)^2 \\ &- \delta \left(1 - \frac{\sigma_1}{2}\right)\underline{s}, \end{aligned} \tag{B.39}$$

where

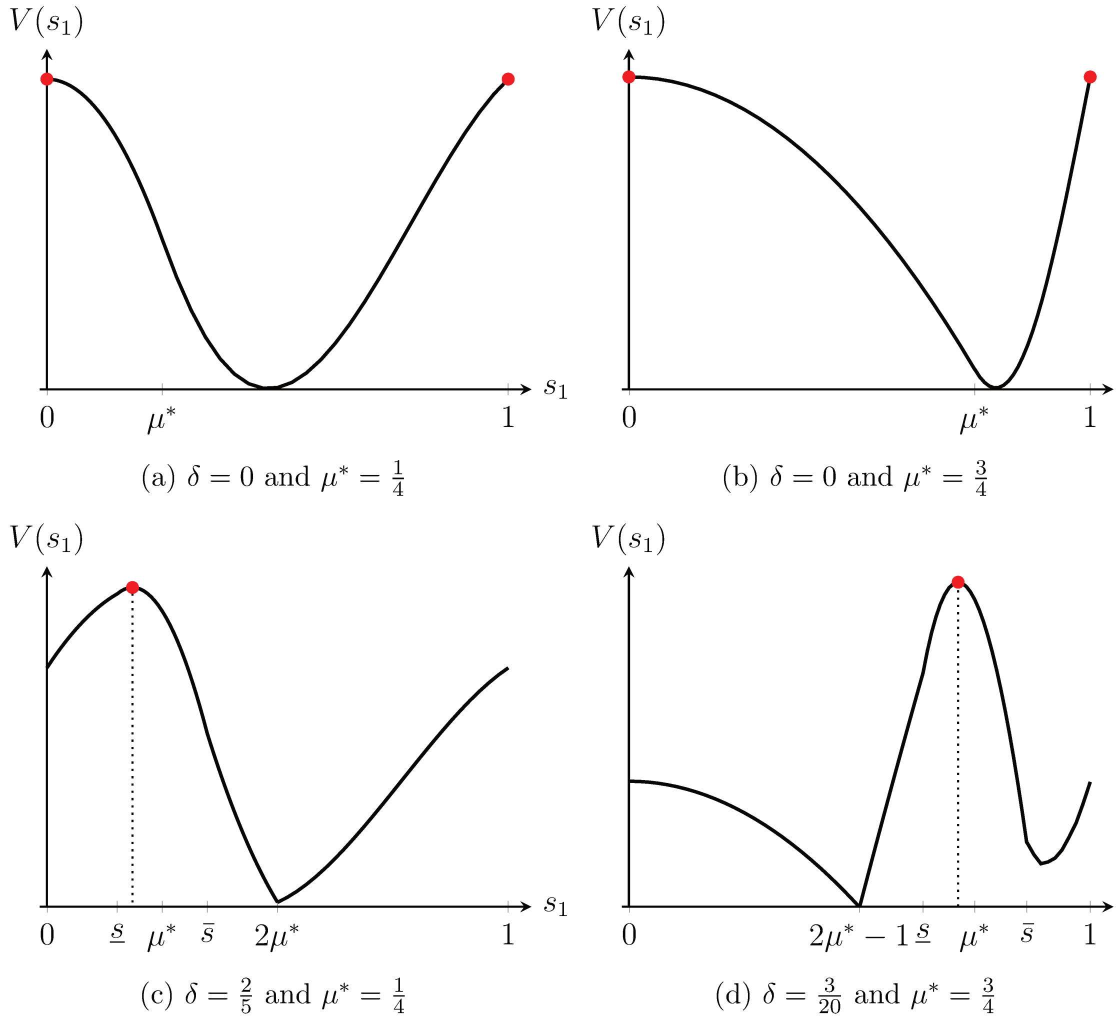
$$\sigma_1 = \frac{1}{2\tau}\left(s_1^2 - \underline{s}^2\right)\frac{1 - \frac{s_1}{2}}{s_1}\left(pR - \frac{1}{\frac{s_1 + \underline{s}}{2}}\right), \tag{B.40}$$

$$\sigma_2 = \frac{1}{4\tau}\left(1 - s_1^2\right)\left(pR - \frac{1}{\frac{1+s_1}{2}}\right). \tag{B.41}$$

Note that σ_1 from (B.40) and (B.37) are identical. This can be seen by using the expressions for $\underline{s}, \bar{s}$ and μ^*. Moreover, σ_2 in (B.41) and (B.38) are also identical. Finally, note also that on the interval $s_1 \in [\bar{s}, 1]$, σ_1 is maximized at $s_1 = 1$, σ_2 is maximized at $s_1 = \bar{s}$ and that $\sigma_1(s_1 = 1) = \sigma_2(s_1 = \bar{s})$. We define $\sigma \equiv \sigma_1(s_1 = 1)$, given by

$$\sigma = \frac{1}{2\tau}(1 - \bar{s})\left(\frac{1 + \bar{s}}{2\mu^*} - 1\right). \tag{B.42}$$

It is straightforward to verify that $V(s_1)$ is continuous over the entire interval $s_1 \in [0, 1]$ and that $V(s_1 = 0) = V(s_1 = 1)$.



(a) $\delta = 0$ and $\mu^* = \frac{1}{4}$

(b) $\delta = 0$ and $\mu^* = \frac{3}{4}$

(c) $\delta = \frac{2}{5}$ and $\mu^* = \frac{1}{4}$

(d) $\delta = \frac{3}{20}$ and $\mu^* = \frac{3}{4}$

Figure B.1: Objective Function $V(s_1)$ and $\max_{s_1} V(s_1)$ (in red).

Figure B.1 depicts the objective function $V(s_1)$ for some parameter values. We now prove the expressions (B.30)-(B.41).

Region a(i): $s_1 \in [0, 2\mu^* - 1)$, non-empty only if $2\mu^* - 1 > 0$. Since $s_1 < 2\mu^* - 1$, we have $\mu_1 < \mu_2 < \mu^*$. Without information from the share price, neither the regulator nor the capital provider will act to fund the bank if the stress test generates message m_1. There is hence no information production by the speculator. Creditors do not roll over their debt even after a *pass* result, unless the stock price reveals $\omega = h$. Following a test outcome m_2, the supervisor funds the bank if the stock price is uninformative and $s > \bar{s}$. This then provides incentives to the speculator to acquire information and trade following m_2.

Given the funding rate $r = \frac{1}{ps}$ the equity value given $\omega = h$ and $a = 1$ is

$$E_h^1(s) = pR - \frac{1}{s}.$$

We have $\Pr(\omega = h, s|m_2) = \Pr(\omega = h|s) f(s|m_2) = \frac{s}{1-s_1}$ and hence, the market maker sets the price when the order flow is uninformative following m_2 at

$$P(m_2) = \int_{\bar{s}}^{1} \left(pR - \frac{1}{s}\right) \frac{s}{1-s_1} \mathrm{d}s, \\ = \frac{1}{2}\frac{1-\bar{s}^2}{1-s_1} E_h^1,$$

where

$$E_h^1 \equiv pR - \frac{1}{\frac{\bar{s}+1}{2}}.$$

The trading profit is therefore

$$E(\pi) = \frac{\sigma_2}{2}\left[\int_{\bar{s}}^{1} \frac{s}{1-s_1}\left(E_h^1(s) - P(m_2)\right)\mathrm{d}s + \frac{1}{2}\frac{\bar{s}^2 - s_1^2}{1-s_1}(-P(m_2)) + \frac{1}{2}\frac{(1-s_1)^2}{1-s_1}P(m_2)\right] - \frac{\tau}{2}\sigma_2^2, \\ = \frac{\sigma_2}{2}\left[\frac{1}{2}\frac{1-\bar{s}^2}{1-s_1}\left(E_h^1 - P(m_2)\right) + \frac{1}{2}\frac{\bar{s}^2 - s_1^2}{1-s_1}(-P(m_2)) + \frac{1}{2}\frac{(1-s_1)^2}{1-s_1}P(m_2)\right] - \frac{\tau}{2}\sigma_2^2,$$

which can be simplified to

$$E(\pi) = \frac{\sigma_2}{4}(1-\bar{s}^2)E_h^1 - \frac{\tau}{2}\sigma^2$$

Taking the first order condition, we get the amount of information acquisition at the optimum, following message m_2, given by (B.31). The supervisor's expected payoff is thus

$$V_{a(i)}(s_1) = \frac{1}{2}(1-\bar{s}^2)\left[\frac{\sigma_2}{2}V_h^1 + \left(1 - \frac{\sigma_2}{2}\right)\left(V_h^1 - \delta\right)\right] + \frac{1}{2}(1-\bar{s})^2\left(\frac{\sigma_2}{2}V_l^0 + \left(1 - \frac{\sigma_2}{2}\right)\left(V_l^1 - \delta\right)\right) \\ + \frac{1}{2}\left[\bar{s}(2-\bar{s}) - s_1(2-s_1)\right]V_l^0 + \frac{1}{2}(\bar{s}^2 - s_1^2)\left(\frac{\sigma_2}{2}V_h^1 + \left(1-\frac{\sigma_2}{2}\right)V_h^0\right) \\ + \frac{1}{2}\left[\left(2s_1 - s_1^2\right)V_l^0 + s_1^2 V_h^0\right].$$

After some calculations, we get (B.30).

Region a(ii): $s_1 \in [\max\{0, 2\mu^* - 1\}, \underline{s})$, where $\underline{s} = \mu^*(1-\delta)$. We have $\mu_1 < \mu^* \leq \mu_2$ so that, in the absence of an informative stock price, creditors roll over their debt following a *pass* but not a *fail* result. If $s < s_1$ there is no information production and the bank will fail to raise funds from private markets or the supervisor. Following a *pass* result, capital providers are willing to fund the bank, unless the stock price signal reveals $\omega = l$. The supervisor does not intervene, unless the bank passes the test, the stock price is uninformative and $s \in [s_1, \underline{s})$. In that case, the supervisor shuts down the bank ($a_s = 0$). Overall, this yields the following expected payoff to the supervisor:

$$V_{a(ii)}(s_1) = \frac{1}{2}\left[(1-\underline{s}^2)V_h^1 + (\underline{s}^2 - s_1^2)\left[\frac{\sigma_2}{2}V_h^1 + (1-\frac{\sigma_2}{2})(V_h^0 - \delta)\right] + s_1^2 V_h^0\right] \\ + \frac{1}{2}\left[1 - (2\underline{s} - \underline{s}^2)\right]\left[\frac{\sigma_2}{2}V_l^0 + (1-\frac{\sigma_2}{2})V_l^1)\right]] \\ + \frac{1}{2}\left[\left[2\underline{s} - \underline{s}^2 - (2s_1 - s_1^2)\right]\left[\frac{\sigma_2}{2}V_l^0 + (1-\frac{\sigma_2}{2})(V_l^0 - \delta)\right] + (2s_1 - s_1^2)V_l^0\right].$$

After some calculations, we get the expression in (B.32).

Next, we need to determine prices and trading profits in order to determine the relation between s_1 and the amount of information σ_2 produced by the speculator. When m_2 is observed and the order is uninformative, the market maker does not know whether the bank will be able to continue. This depends on whether s is above or below $\underline{s}$. Denote by $\mu_\omega^a(m_2) = Pr(\omega, a|m_2)$. Here a denotes whether the firm is able to continue ($a = 1$) or not ($a = 0$). To ease the exposition we do not introduce extra notation to distinguish between the case where the supervisor is decisive (unwinds before the bank can try to raise capital), or the providers of capital are not willing to fund the bank. For the sake of brevity, we drop the function argument m_2 from $\mu_\omega^a(m_2)$. Beliefs following m_1 do not require a similar distinction since the supervisor and capital providers agree that the optimal action in this case is $a = 0$ and the speculator produces no information.

$$
\begin{aligned}
\mu_h^1 &= \frac{1}{2}\frac{1-\underline{s}^2}{1-s_1}, \\
\mu_h^0 &= \frac{1}{2}\frac{\underline{s}^2 - s_1^2}{1-s_1}, \\
\mu_l^1 &= \frac{1}{2}\frac{1-(2\underline{s}-\underline{s}^2)}{1-s_1}, \\
\mu_l^0 &= \frac{1}{2}\frac{2\underline{s}-\underline{s}^2-(2s_1-s_1^2)}{1-s_1}.
\end{aligned}
$$

The speculator's expected trading profits, after having observed m_2 and acquired information σ_2 are

$$E(\pi) = \frac{\sigma_2}{2}\left[\mu_h^1(E_h^1 - P) + \mu_h^0(-P) + (\mu_l^1 + \mu_l^0)P\right] - \frac{\tau}{2}\sigma_2^2.$$

Since the market maker learns nothing if the order flow is uninformative, and since $E_h^0 = E_l^1 = E_l^0 = 0$, the price is

$$P(m_2) = \mu_h^1 E_h^1$$

where E_h^1 is computed below. The expression for trading profits can be simplified to

$$E(\pi) = \sigma_2 E_h^1 \mu_h^1(1 - \mu_h^1 - \mu_h^0) - \frac{\tau}{2}\sigma_2^2.$$

To calculate E_h^1 note that the bank only gets to approach capital markets if the supervisor has allowed it to go ahead. This reveals to the capital providers that $s \geq \underline{s}$. Hence, at this point, the capital providers believe $Pr(\omega = h) = \frac{1+\underline{s}}{2}$. Since the interest rate is set as a function of the belief μ at $r = \frac{1}{p\mu}$, we get

$$E_h^1 = pR - \frac{1}{\frac{1+\underline{s}}{2}} = pR - \frac{2}{1+\underline{s}}.$$

From this, we can calculate overall trading profits as

$$E(\pi) = \frac{\sigma_2}{2}\left(pR - \frac{2}{1+\underline{s}}\right)\left(1-\underline{s}^2\right)\frac{1-\frac{1+s_1}{2}}{1-s_1} - \frac{\tau}{2}\sigma_2^2,$$

which can be simplified to

$$E(\pi) = \frac{\sigma_2}{4}\left(pR - \frac{2}{1+\underline{s}}\right)\left(1-\underline{s}^2\right) - \frac{\tau}{2}\sigma_2^2.$$

Taking the first order condition with respect to σ_2 we get the expression in (B.33).

Region p: $s_1 \in [\underline{s}, \bar{s})$, where $\underline{s} = \mu^*(1-\delta)), \bar{s} = \mu^*(1+\delta)$. We have $\mu_1 < \mu^* \leq \mu_2$ so that creditors roll over their debt following a *pass* but not a *fail* result. The supervisor is always passive: Following a *pass* result, $s > \underline{s}$ so the supervisor does not want to intervene with $a_s = 0$. Following a *fail* result, we always have $s < \bar{s}$ and hence the supervisor does not want to fund the bank. Since the supervisor is completely passive, the game proceeds as in the baseline model, and the supervisor's value function is given by Lemma 3. By setting $s_2 = 1$ and after some simplifications we get the value function in (B.34) and the amount of information acquisition in (B.35).

Region b(i): $s_1 \in [\bar{s}, \min\{2\mu^*, 1\})$. We have $\mu_1 < \mu^* < \mu_2$. If the bank fails the test, but the supervisor observes $s \in [\bar{s}, s_1]$, she funds the bank at rate $r = \frac{1}{ps}$. Since the bank gets funded with positive probability following a test outcome m_1, the speculator may now acquire information even after such a comparatively unfavorable test result, i.e., in spite of an induced belief $\mu_1 < \mu^*$. Denote by σ_1, the amount of information produced by the speculator, following m_1, and similarly, σ_2 denotes the information acquired following m_2.

We proceed to calculate trading profits following m_1. As before, the equity value is only non-zero when $\omega = h$ and $a = 1$. Given the funding costs $r = \frac{1}{ps}$, the equity value is

$$E_h^1(s) = pR - \frac{1}{s}.$$

We have $\Pr(\omega = h, s|m_1) = f(s|m_1)\Pr(\omega = h|s) = \frac{s}{s_1}$ and hence, the market maker sets the price when order flow is uninformative following m_1 at

$$\begin{aligned} P(m_1) &= \int_{\bar{s}}^{s_1}\left(pR - \frac{1}{s}\right)\frac{s}{s_1}\mathrm{d}s, \\ &= \frac{1}{2}\frac{s_1^2 - \bar{s}^2}{s_1}E_h^1, \end{aligned}$$

where

$$E_h^1 \equiv pR - \frac{1}{\frac{\bar{s}+s_1}{2}}.$$

From this we can calculate trading profits:

$$\begin{aligned} E(\pi) &= \frac{\sigma_1}{2}\left[\int_{\bar{s}}^{s_1}\frac{s}{s_1}\left(E_h^1(s) - P(m_1)\right)\mathrm{d}s + \frac{1}{2}\frac{\bar{s}^2}{s_1}(-P(m_1)) + \frac{1}{2}\frac{s_1(2-s_1)}{s_1}P(m_1)\right] - \frac{\tau}{2}\sigma_1^2, \\ &= \frac{\sigma_1}{2}\left[\frac{1}{2}\frac{s_1^2 - \bar{s}^2}{s_1}\left(E_h^1 - P(m_1)\right) + \frac{1}{2}\frac{\bar{s}^2}{s_1}(-P(m_1)) + \frac{1}{2}\frac{s_1(2-s_1)}{s_1}P(m_1)\right] - \frac{\tau}{2}\sigma_1^2. \end{aligned}$$

After some simplifications and taking the first-order condition with respect to σ_1, we find the expression in (B.37).

Trading profits and therefore information acquisition following m_2 can be calculated as in the baseline model. Using the function (3.14) derived previously and setting $\mu = \mu_2 = \frac{1+s_1}{2}$ we get the expression in (B.38). The supervisor's expected payoff is given by

$$\begin{aligned} V_{b(i)}(s_1) = \frac{1}{2}\Big[(1-s_1^2)V_h^1 + (s_1^2 - \bar{s}^2)\left(V_h^1 - (1-\frac{\sigma_1}{2})\delta\right) + \bar{s}^2\left(\frac{\sigma_1}{2}V_h^1 + \left(1-\frac{\sigma_1}{2}\right)V_h^0\right)\Big] \\ + \frac{1}{2}\Big[(1 - s_1(2-s_1))\left(\frac{\sigma_2}{2}V_l^0 + (1-\frac{\sigma_2}{2})V_l^1\right)\Big] \\ + \frac{1}{2}\Big[(s_1(2-s_1) - \bar{s}(2-\bar{s}))\left(\frac{\sigma_1}{2}V_l^0 + (1-\frac{\sigma_1}{2})(V_l^1 - \delta)\right) + \bar{s}(2-\bar{s})V_l^0\Big]. \end{aligned}$$

and after some simplification, this can be written as the expression in (B.36).

Region b(ii): $s_1 \in [2\mu^*, 1]$, non-empty only if $2\mu^* < 1$. In this region $\mu^* \leq \mu_1 < \mu_2$ so that capital providers are willing to fund the bank even if it fails the test (unless the stock price reveals $\omega = l$). The supervisor intervenes and sets $a_s = 0$ if she observes $s < \underline{s}$. After the message m_2 the game proceeds as in the baseline model: the speculator's information acquisition after m_2 is given by (B.41) and in the absence of a revealing stock price the market invests.

Following the message m_1, if the stock price reveals nothing, the supervisor may intervene depending on the realization of her private signal. When the market sees no intervention it realizes that $s \in [\underline{s}, s_1)$ and chooses the interest rate accordingly, leading to an equity value of

$$E_h^1 = pR - \frac{1}{\frac{\underline{s}+s_1}{2}}.$$

We can determine the speculator's and market maker's beliefs over the state and the expected supervisor's decision, $\mu_\omega^{a_s} = \Pr(\omega, a_s|m_1) = \Pr(\omega|a_s)\Pr(a_s|m_1)$. These are:

$$\begin{aligned} \mu_h^1 &= \frac{\underline{s}+s_1}{2}\frac{s_1-\underline{s}}{s_1} \\ \mu_h^0 &= \frac{\underline{s}}{2}\frac{\underline{s}}{s_1} \\ \mu_l^1 &= \left(1-\frac{\underline{s}+s_1}{2}\right)\frac{s_1-\underline{s}}{s_1} \\ \mu_l^0 &= \left(1-\frac{\underline{s}}{2}\right)\frac{\underline{s}}{s_1}. \end{aligned}$$

The market maker chooses a price

$$P = \mu_h^1 E_h^1,$$

and the speculator's expected profits are

$$\begin{aligned} E(\pi) &= \frac{\sigma_1}{2}\left[\mu_h^1(E_h^1 - P) + \mu_h^0(-P) + (\mu_l^1 + \mu_l^0)P\right] - \frac{1}{2}\tau\sigma_1^2 \\ &= \sigma_1\mu_h^1(1-\mu_h^1-\mu_h^0)E_h^1 - \frac{1}{2}\tau\sigma_1^2. \end{aligned}$$

By taking the first-order condition with respect to σ_1 we get the expression in (B.40). The supervisor's objective function is:

$$\begin{aligned}
V_{b(ii)}(s_1) =& \frac{1}{2}\left[\underline{s}^2\left(\frac{\sigma_1}{2}V_h^1 + \left(1-\frac{\sigma_1}{2}\right)(V_h^0-\delta)\right) + \left((s_1^2-\underline{s}^2)+(1-s_1^2)\right)V_h^1\right] \\
&+ \frac{1}{2}\left[(2\underline{s}-\underline{s}^2)\left(\frac{\sigma_1}{2}V_l^0 + \left(1-\frac{\sigma_1}{2}\right)(V_l^0-\delta)\right)\right] \\
&+ \frac{1}{2}\left[\left(2(s_1-\underline{s})-(s_1^2-\underline{s}^2)\right)\left(\frac{\sigma_1}{2}V_l^0 + \left(1-\frac{\sigma_1}{2}\right)V_l^1\right)\right] \\
&+ \frac{1}{2}\left[\left(2(1-s_1)-(1-s_1^2)\right)\left(\frac{\sigma_2}{2}V_l^0 + \left(1-\frac{\sigma_2}{2}\right)V_l^1\right)\right].
\end{aligned}$$

This can be simplified to get (B.39). ■

Lemma B.2. *The function $V(s_1)$ defined in (B.29) attains a local maximum on the interval $s_1 \in [0,\bar{s}]$ at*

$$s_1 = s_F^P \equiv \begin{cases} s_1^{a(ii)} & \text{if } s_1^{a(ii)} < \mu^*(1-\delta) \\ s_F & \text{if } s_1^{a(ii)} \geq \mu^*(1-\delta), \end{cases} \in [0,\mu^*), \tag{B.43}$$

where $s_1^{a(ii)} \equiv \delta\frac{1-\frac{\sigma_2}{2}}{\frac{\sigma_2}{2}(pR-1)}$ with σ_2 defined in (B.33), and s_F solves (3.19).

Proof of Lemma B.2. When $\mu^* \leq \frac{1}{2}$, region a(i) is empty. For s_1 in region a(ii), the supervisor's objective is given by (B.32). It is straightforward to show that

$$\frac{\partial V_{a(ii)}}{\partial s_1} = \delta\left(1-\frac{\sigma_2}{2}\right) - s_1\frac{\sigma_2}{2}\Delta V_h. \tag{B.44}$$

Note that $\frac{\partial V_{a(ii)}}{\partial s_1}$ is strictly positive at $s_1 = 0$ and $\delta > 0$. Moreover, $V_{a(ii)}(0) = V_{b(ii)}(1)$. Hence, for any $\delta > 0$ the optimal $s_1 \in (0,1)$. If s_1 increases so that we are in region p, we know from Proposition 3.1 that $V_p(s_1)$ attains a maximum at $s_1 < \mu^*$, as we have $\frac{\partial V_p}{\partial s_1} < 0$ at $s_1 = \mu^*$. Hence, $V(s_1)$ must attain a local maximum between 0 and μ^*.

The following can be said about whether the local maximum lies in region a(ii) or region p. The maximum of $V_{a(ii)}(s_1)$ is reached at

$$s_1 = \delta\frac{1-\frac{\sigma_2}{2}}{\frac{\sigma_2}{2}(pR-1)} \equiv s_1^{a(ii)}, \tag{B.45}$$

where σ_2 is defined in (B.33). If $s_1^{a(ii)} < \underline{s} = \mu^*(1-\delta)$, then $V(s_1)$ reaches a local maximum in region a(ii). Moreover, using the expressions for $V_{a(ii)}(s_1)$ and $V_p(s_1)$, it can be shown that at the point $s_1 = \underline{s}$ we have $\frac{\partial V_{a(ii)}}{\partial s_1} < \frac{\partial V_p}{\partial s_1}$. Hence, if $\frac{\partial V_p}{\partial s_1} < 0$ at the corner $s_1 = \underline{s}$ of region p, then the local maximum is in region a(ii). If $s_1^{a(ii)} > \underline{s}$ we have that $V_{a(ii)(s_1)}$ is strictly increasing on the entire region a(ii) and the local maximum lies in region p. There can be cases where $\frac{\partial V_{a(ii)}}{\partial s_1} < 0$ for the corner $s_1 = \underline{s}$ in region a(ii), but $\frac{\partial V_p(s_1)}{\partial s_1} > 0$ for the corner $s_1 = \underline{s}$ in region p. In that case, $V(s_1)$ has two local maxima on $[0,\mu^*]$.

When $\mu^* > \frac{1}{2}$, the relevant regions are a(i) to b(i). $V_{a(i)}(s_1)$ is decreasing on the entire region a(i), as can be seen by calculating its derivative:

$$\frac{\partial V_{a(i)}(s_1)}{\partial s_1} = -s_1 \frac{\sigma_2}{2} \Delta V_h, \tag{B.46}$$

which is negative. Moreover, we know from before that for s_1 in regions a(ii)-p, the local maximum is below μ^*. It then follows that the local maximum over regions a(i)-p is either at the corner $s_1 = 0$, or is interior and given by s_F^P.

■

Lemma B.3. *For $\mu^* \leq \frac{1}{2}$, s_F^P is a global maximum.*

Proof of Lemma B.3. The proof of the lemma follows from the following two claims.

Claim 1. *On regions b(i)-b(ii), the objective function $V(s_1)$ attains a maximum at one of its corners $s_1 \in \{\bar{s}, 1\}$.*

Claim 2. *For $\delta > 0$ we have $V(s_F^P) > \max\left\{V\left(\bar{s}\right), V(1)\right\}$. For $\delta = 0$ we have $s_F^P = 0$ and $V(0) = V(1) > V(\mu^*)$.*

Proof of Claim 1: The proof proceeds by constructing a straight line connecting the extreme points $(\bar{s}, V(\bar{s}))$ and $(1, V(1))$ of the value function $V(s_1)$ and then showing that $V(s_1)$ lies below that line for all $s_1 \in (\bar{s}, 1)$. First, consider region b(i), i.e. $s_1 \in [\bar{s}, 2\mu^*)$. We want to show that the straight line connecting $(\bar{s}, V_{b(i)}(\bar{s}))$ to $(1, V_{b(ii)}(1))$ lies above $V_{b(i)}(s_1)$ for all $s_1 \in [\bar{s}, 2\mu^*)$. Note that $V_{b(i)}(s_1) > V_{b(ii)}(s_1)$ if and only if $s_1 < 2\mu^*$ so that the line connecting $(\bar{s}, V_{b(i)}(\bar{s}))$ to $(1, V_{b(ii)}(1))$ lies above the line $\widehat{V}_{b(i)}(s_1)$ connecting $(\bar{s}, V_{b(i)}(\bar{s}))$ to $(1, V_{b(i)}(1))$. We prove the stronger claim that the line $\widehat{V}_{b(i)}(s_1)$ lies above $V_{b(i)}(s_1)$ for all $s_1 \in [\bar{s}, 2\mu^*)$. The line $\widehat{V}_{b(i)}(s_1)$ is defined as

$$\widehat{V}_{b(i)}(s_1) = V_{b(i)}(\bar{s}) + \frac{V_{b(i)}(1) - V_{b(i)}(\bar{s})}{1 - \bar{s}}(s_1 - \bar{s}).$$

This can be re-written as

$$\widehat{V}_{b(i)}(s_1) = \frac{1 - s_1}{1 - \bar{s}} V_{b(i)}(\bar{s}) + \frac{s_1 - \bar{s}}{1 - \bar{s}} V_{b(i)}(1).$$

We then get $\widehat{V}_{b(i)}(s_1) \geq V_{b(i)}(s_1)$ if

$$\Delta V_h \left(\sigma \frac{s_1 - \bar{s}}{1 - \bar{s}} - \sigma_1(s_1)\right) \bar{s}^2 + 2\delta(\sigma - \sigma_1(s_1))(s_1 - \bar{s}) - \Delta V_l \sigma_2(s_1)(1 - s_1)^2 + \Delta V_l \left(\sigma(1 - \bar{s})^2 + \sigma_1(s_1)(1 - s_1)^2 - \sigma_1(s_1)(1 - \bar{s})^2\right) \geq 0,$$

where σ is given by (B.42) and $\sigma_1(s_1)$ and $\sigma_2(s_1)$ by (B.37) and (B.38), respectively.

Using

$$(1-\bar{s})^2 = (1-\bar{s})^2 - (1-s_1)^2 + (1-s_1)^2,$$

this can be re-written as

$$\Delta V_h \left(\sigma\frac{s_1-\bar{s}}{1-\bar{s}} - \sigma_1(s_1)\right)\bar{s}^2 + 2\delta(\sigma-\sigma_1(s_1))(s_1-\bar{s})$$
$$+ \Delta V_l \left((\sigma-\sigma_1(s_1))\left[(1-\bar{s})^2-(1-s_1)^2\right] + (\sigma-\sigma_2(s_1))(1-s_1)^2\right) \geq 0. \quad \text{(B.47)}$$

We then make use of $\Delta V_l = 1$, $\Delta V_h = \frac{1}{\mu^*} - 1$ and

$$\sigma - \sigma_1(s_1) = \frac{1}{2\tau}\frac{1-s_1}{s_1}\left(\frac{s_1(1-s_1)}{2\mu^*} + s_1 - \mu^*(1-\delta^2)\right),$$
$$\sigma\frac{s_1-\bar{s}}{1-\bar{s}} - \sigma_1(s_1) = -\frac{1}{2\tau}\frac{(s_1-\bar{s})(1-s_1)}{s_1}\frac{s_1-2\mu^*(1-\delta)}{2\mu^*}, \quad \text{(B.48)}$$
$$\sigma - \sigma_2(s_1) = \frac{1}{2\tau}(s_1-\bar{s})\left(\frac{s_1+\bar{s}}{2\mu^*} - 1\right).$$

Since $s_1 \geq \bar{s}$ and $\sigma \geq \sigma_1(s_1)$ and $\sigma \geq \sigma_2(s_1)$, the last line of (B.47) is positive. A sufficient condition then is that the first line is also positive.

Using (B.48) the first line is positive if

$$\frac{(s_1-\bar{s})(1-s_1)}{s_1}\left(\mu^*(1-\mu^*)(1+\delta)^2\frac{2\mu^*(1-\delta)-s_1}{2\mu^*} + 2\delta\left[\frac{s_1(1-s_1)}{2\mu^*} + s_1 - \mu^*(1-\delta^2)\right]\right) \geq 0.$$

The factor $\frac{(s_1-\bar{s})(1-s_1)}{s_1} \geq 0$. The remainder of the expression is a negative quadratic function. It is therefore sufficient to check that at its borders $s_1 = \bar{s}$ and $s_1 = 2\mu^*$, the terms in brackets is positive. At $s_1 = \bar{s}$, the expression becomes

$$\frac{1}{2}\mu^*(1-\mu^*)(1+\delta)^2(1-\delta) - \mu^*(1-\mu^*)\delta(1+\delta)^2 + \delta(1+\delta) - \mu^*\delta(1-\delta^2),$$

which can also be written as

$$\frac{1}{2}\mu^*(1-\mu^*)(1+\delta)^2(1-\delta) + \delta(1+\delta)\left[1-\mu^*(1-\mu^*)(1+\delta) - \mu^*(1-\delta)\right]. \quad \text{(B.49)}$$

Since $\mu^* \leq \frac{1}{2}$ we have $\mu^*(1-\mu^*) \leq \frac{1}{4}$ and $1+\delta \leq 2$ so that $\mu^*(1-\mu^*)(1+\delta) \leq \frac{1}{2}$. Moreover, $\mu^*(1-\delta) \leq \frac{1}{2}$ and hence (B.49) is positive.

At $s_1 = 2\mu^*$ we get

$$\delta(2 - 2\mu^*(1-\delta^2) - \mu^*(1-\mu^*)(1+\delta)^2),$$

which is also positive. It follows that the line connecting $V_{b(i)}(\bar{s})$ to $V_{b(i)}(1)$, lies above $V_{b(i)}(s_1)$ for all $s_1 \in [\bar{s}, 2\mu^*]$, implying that the line connecting $V_{b(i)}(\bar{s})$ to $V_{b(ii)}(1)$, also lies above $V_{b(i)}(s_1)$ for all $s_1 \in [\bar{s}, 2\mu^*]$.

Now, consider region b(ii), i.e. $s_1 \in [2\mu^*, 1]$. We want to show that $V_{b(ii)}(1) \geq V_{b(ii)}(s_1)$ for all $s_1 \in [2\mu^*, 1]$. Since, $V(s_1)$ is continuous across regions, this is enough to conclude the proof of the claim.

Using (B.39) we can write

$$V_{b(ii)}(1) = \frac{V_h^1 + V_l^1}{2} - \delta\left(1 - \frac{\sigma}{2}\right)\underline{s} - \frac{1}{2}\Delta V_h\left(1 - \frac{\sigma}{2}\right)\underline{s}^2 + \frac{1}{2}\Delta V_l\left(\frac{\sigma}{2} + \left(1 - \frac{\sigma}{2}\right)\left(2\underline{s} - \underline{s}^2\right)\right).$$

We can then write the condition $V_{b(ii)}(1) - V_{b(ii)}(s_1) \geq 0$ as follows:

$$\frac{1}{4}\Delta V_h \underline{s}^2\left(\sigma - \sigma_1(s_1)\right) + \frac{1}{2}\delta\underline{s}(\sigma - \sigma_1(s_1)) - \frac{1}{4}\Delta V_l \sigma_2(s_1)(1 - s_1)^2 + \frac{1}{4}\Delta V_l\left(\sigma(1-\underline{s})^2 + \sigma_1(s_1)\left[(1-s_1)^2 - (1-\underline{s})^2\right]\right) \geq 0,$$

where σ is given by (B.42) and $\sigma_1(s_1)$ and $\sigma_2(s_1)$ by (B.40) and (B.41), respectively.

Using $(1-\underline{s})^2 = (1-\underline{s})^2 - (1-s_1)^2 + (1-s_1)^2$, this can be re-written as

$$\left(\Delta V_h \underline{s}^2 + 2\delta\underline{s}\right)(\sigma - \sigma_1(s_1)) + \Delta V_l(\sigma - \sigma_2(s_1)) + \Delta V_l(\sigma - \sigma_1(s_1))\left((1-\underline{s})^2 - (1-s_1)^2\right) \geq 0.$$

Note that $\sigma \geq \sigma_1(s_1)$ and $\sigma \geq \sigma_2(s_1)$. Moreover, $(1-\underline{s})^2 \geq (1-s_1)^2$ since $s_1 \geq 2\mu^* > \underline{s}$. Hence, the inequality holds. It follows that on regions b(i)-b(ii), $V(s_1)$ attains a maximum at one of its corners.

Proof of Claim 2: For $\delta > 0$, we have that $V(s_F^P) > V(0)$ since we know from Lemma 7 that $V(s_1)$ is increasing at $s_1 = 0$. Since $V(0) = V(1)$, it follows that $V(s_F^P) > V(1)$. Moreover, we know that $V(s_F^P) > V(\mu^*) > V\big(\mu^*(1+\delta)\big)$, where the last inequality follows from $V_p(s_1)$ being decreasing for $s_1 \geq \mu^*$. Hence, for $\delta > 0$, s_F^P is the global maximum. For $\delta = 0$, we have $s_F^P = 0$, region p disappears, and $V(0) > V(\mu^*)$ since $V_{a(ii)}(s_1)$ is decreasing over region a(ii) (see (B.44)). Hence, $s_F^P = 0$ is the global maximum when $\delta = 0$. ■

Lemma B.4. *For $\mu^* > \frac{1}{2}$, when $\delta = 0$ the global maximum is $s_1 = 0$ and there exist a $\widetilde{\delta}$ such that $s_1 = s_F^P$ is a global maximum for $\delta \geq \widetilde{\delta}$.*

Proof of Lemma B.4. The following two claims establish that the local maximum is also the global maximum.

Claim 3. *On region b(i), the objective function $V(s_1)$ attains a maximum at one of its corners $s_1 \in \{\bar{s}, 1\}$.*

Claim 4. *There exists a $\widetilde{\delta}$, such that $V(\mu^*) > V(1)$ for $\delta \geq \widetilde{\delta}$. For $\delta = 0$, we have $V(0) > V(\mu^*)$.*

Since $\arg\max V_p(s_1) < \mu^*$, and $V(s_F^P) > V(\mu^*)$ claims 3 and 4 imply that for $\delta \geq \widetilde{\delta}$, s_F^P is the global maximum.

Proof of Claim 3: We want to show that the straight line passing from $V_{b(i)}(\mu^*(1+\delta)) = V_{b(i)}(\bar{s})$ and $V_{b(i)}(1)$ lies above $V_{b(i)}(s_1)$ for all $s_1 \in [\bar{s}, 1]$. Define the straight line by $\widehat{V}_{b(i)}(s_1)$. The maximum of $\widehat{V}_{b(i)}(s_1)$ is obviously attained at either of the two corners $s_1 = \bar{s}$ or $s_1 = 1$. Moreover, if $\widehat{V}_{b(i)}(s_1) \geq V_{b(i)}(s_1)$ for $s_1 \in [\bar{s}, 1]$, knowing that by construction $\widehat{V}_{b(i)}(s_1) = V(s_1)$ at the corners $s_1 = \bar{s}$ and $s_1 = 1$, then $V_{b(i)}(s_1)$ must also attain its maximum at either of the two corners.

First, construct the function $\widehat{V}_{b(i)}(s_1)$. Denoting by

$$T \equiv \frac{V_1^h + V_1^l}{2} - \frac{\bar{s}^2}{2}\Delta V_h + \frac{\bar{s}(2-\bar{s})}{2}\Delta V_l,$$

we can write

$$V_{b(i)}(s_1) = T + \frac{\sigma_1(s_1)}{4}\left(\bar{s}^2\Delta V_h + (1-\bar{s})^2\Delta V_l\right) + \frac{\sigma_2(s_1) - \sigma_1(s_1)}{4}(1-s_1)^2\Delta V_l - \delta\left(1 - \frac{\sigma_1(s_1)}{2}\right)(s_1 - \bar{s}),$$

where $\sigma_1(s_1)$ and $\sigma_2(s_1)$ are defined in (B.37) and(B.38), respectively. The linear function is given by

$$\widehat{V}_{b(i)}(s_1) = V_{b(i)}(\bar{s}) + \frac{V_{b(i)}(1) - V_{b(i)}(\bar{s})}{1-\bar{s}}(s_1 - \bar{s}).$$

It is useful to note that $\sigma_1(s_1 = \bar{s}) = 0$ and $\sigma_2(s_1 = \bar{s}) = \sigma$, where σ is given by (B.42). Moreover, $\sigma_1(s_1 = 1) = \sigma$ and $\sigma_2(s_1 = 1) = 0$. We can then write

$$\widehat{V}_{b(i)}(s_1) = T + \frac{\sigma}{4}(1-\bar{s})^2\Delta V_l + \left[\frac{\sigma}{4}\frac{\bar{s}^2}{1-\bar{s}}\Delta V_h - \delta\left(1 - \frac{\sigma}{2}\right)\right](s_1 - \bar{s}).$$

The inequality $\widehat{V}_{b(i)}(s_1) \geq V_{b(i)}(s_1)$ can be written as:

$$(\sigma - \sigma_1(s_1))(1-\bar{s})^2\Delta V_l - (\sigma_2(s_1) - \sigma_1(s_1))(1-s_1)^2\Delta V_l + \left(\sigma\frac{s_1 - \bar{s}}{1-\bar{s}} - \sigma_1(s_1)\right)\bar{s}^2\Delta V_h + 2(\sigma - \sigma_1(s_1))(s_1 - \bar{s})\delta \geq 0. \quad \text{(B.50)}$$

We can re-write the first line as follows

$$(\sigma - \sigma_1(s_1))(1-\bar{s})^2\Delta V_l - (\sigma_2(s_1) - \sigma_1(s_1))(1-s_1)^2\Delta V_l = (\sigma - \sigma_2(s_1))(1-s_1)^2\Delta V_l + (\sigma - \sigma_1(s_1))(s_1 - \bar{s})\left(2(1-s_1) + (s_1 - \bar{s})\right)\Delta V_l.$$

Substituting this expression into (B.50) and simplifying yields

$$\begin{aligned}
&\frac{1}{2\tau}\frac{(1-s_1)(s_1-\bar{s})}{s_1}\left[\frac{s_1-\mu^*(1-\delta)}{2\mu^*}s_1(1-s_1)\right] \\
&\qquad\qquad + \frac{1}{2\tau}\frac{(1-s_1)(s_1-\bar{s})}{s_1}\left[-\mu^*(1-\mu^*)(1+\delta)^2\frac{s_1-2\mu^*(1-\delta)}{2\mu^*}\right] \\
&\qquad\qquad + \frac{1}{2\tau}\frac{(1-s_1)(s_1-\bar{s})}{s_1}\left[2(1-s_1)\left(\frac{s_1(1-s_1)}{2\mu^*}+s_1-\mu^*(1-\delta^2)\right)\right] \\
&+\frac{1}{2\tau}\frac{(1-s_1)(s_1-\bar{s})}{s_1}\left[(s_1-\bar{s}+2\delta)\frac{s_1(1-s_1)}{2\mu^*}+(s_1-\bar{s}+2\delta)(s_1-\mu^*(1-\delta^2))\right]\geq 0.
\end{aligned}$$

This can be re-written as

$$\begin{aligned}
&s_1(1-s_1)\frac{s_1-\mu^*(1-\delta)}{2\mu^*}+2(1-s_1)\left[\frac{s_1(1-s_1)}{2\mu^*}+s_1-\mu^*(1-\delta^2)\right]+\frac{s_1(1-s_1)}{2\mu^*}(s_1-\bar{s}+2\delta) \\
&\qquad + \frac{1}{2}(1-\mu^*)(1+\delta)^2(2\mu^*(1-\delta)-s_1)+(s_1-\bar{s}+2\delta)(s_1-\mu^*(1-\delta^2))\geq 0.
\end{aligned}$$

Since in region b(i), $s_1 \geq \mu^*(1+\delta) = \bar{s}$, we know that the first line is non-negative. A sufficient condition is thus that the second line is also non-negative. Re-writing the second line, this reduces to the quadratic equation

$$(s_1-\bar{s})^2 + b(s_1-\bar{s}) + c \geq 0,$$

where

$$\begin{aligned}
b &\equiv \frac{1}{2}\left(\mu^*(1+\delta)(1+3\delta)-(1-\delta)^2\right) \\
c &\equiv \frac{1}{2}\mu^*(1+\delta)\left((1-\delta)^2+\mu^*(1+\delta)(3\delta-1)\right),
\end{aligned}$$

which is always non-negative if $b^2-4c \leq 0$. This last inequality, after simple algebra, can be re-written as:

$$\Delta(\mu^*) \equiv \frac{(1-\delta)^2}{4}\left(\left(3(1+\delta)\mu^*\right)^2-(1+\delta)(10+6\delta)\mu^*+(1-\delta)^2\right)\leq 0. \qquad \text{(B.51)}$$

Note that $\Delta(\mu^*)$ is convex so that if (B.51) is satisfied at either extremes of μ^*, then it also holds for any interior μ^*. We know that $\mu^* \in [1/2, 1/(1+\delta))$ since for region b(ii) to exists we also need $\bar{s} = \mu^*(1+\delta) < 1$. Consider first the lower bound and note that:

$$\begin{aligned}
\Delta\left(\frac{1}{2}\right) &= \frac{(1-\delta)^2}{4}\left(\frac{9}{4}(1+\delta)^2-(1+\delta)(5+3\delta)+(1-\delta)^2\right) \\
&= \frac{(1-\delta)^2}{4}\left(-(1+\delta)\frac{3}{4}-2+\frac{(1-\delta)^2}{1+\delta}\right)<0,
\end{aligned}$$

since $(1-\delta)^2/(1+\delta) < 1 < 2$. Consider next the upper bound:

$$\begin{aligned}
\Delta\left(\frac{1}{1+\delta}\right) &= \frac{(1-\delta)^2}{4}\left(9-(10+6\delta)+(1-\delta)^2\right) \\
&= \frac{(1-\delta)^2}{4}\left(\delta(\delta-8)\right)<0,
\end{aligned}$$

since $\delta \leq 1$. It follows that (B.51) is satisfied, implying that $\widehat{V}_{b(i)}(s_1) \geq V_{b(i)}(s_1)$ for every $s_1 \in [\mu^*(1+\delta), 1]$, i.e. that on region b(i), the objective $V(s_1)$ attains a maximum at one of its corners.

Proof of Claim 4: Using the corresponding expressions, the condition $V_p(\mu^*) > V_{b(i)}(1)$ can be written as

$$-\frac{1}{2}\Delta V_h(\mu^*)^2 + \frac{1}{2}\Delta V_h(1-\frac{\sigma}{2})\bar{s}^2 + \frac{1}{2}\Delta V_l\left(\frac{\hat{\sigma}_2}{2} + (1-\frac{\hat{\sigma}_2}{2})(2\mu^* - (\mu^*)^2)\right)$$
$$-\frac{1}{2}\Delta V_l\left(\frac{\sigma}{2} + (1-\frac{\sigma}{2})(2\bar{s} - \bar{s}^2)\right) + \delta(1-\frac{\sigma}{2})(1-\bar{s}) > 0,$$

where $\sigma = \sigma_1(s_1 = 1)$ given in (B.42) and $\hat{\sigma}_2 = \sigma_2(s_1 = \mu^*) \geq \sigma$. The inequality can be rewritten as

$$\frac{1}{2}\left(1-\frac{\sigma}{2}\right)h(\delta) > \frac{1}{2}\Delta V_h\frac{\sigma}{2}(\mu^*)^2 + \frac{1}{2}\Delta V_l(1-\mu^*)^2\left(1-\frac{\hat{\sigma}_2}{2}\right), \tag{B.52}$$

where

$$h(\delta) \equiv \Delta V_h(\bar{s}^2 - (\mu^*)^2) + \Delta V_l(1-\bar{s})^2 + 2\delta(1-\bar{s}).$$

Note that σ is a decreasing function of δ while $\hat{\sigma}_2$ is independent of δ. Hence the right-hand side of inequality (B.52) is decreasing in δ. Moreover, $h(\delta)$ can be simplified to yield

$$h(\delta) = 1 - \mu^*(1-\mu^*) + 2\delta - \mu^*(1+\delta)^2.$$

Note that $h(\delta)$ reaches its maximum at $\delta = \frac{1}{\mu^*} - 1$ and is increasing for lower values of δ. Furthermore, since we require $\bar{s} = \mu^*(1+\delta) \leq 1$, the maximum admissible value of δ when $\mu^* > \frac{1}{2}$, is just $\delta = \frac{1}{\mu^*} - 1$. Hence, the left-hand side of inequality (B.52) is increasing in δ.

For the special case $\delta = 0$, we have $\sigma = \hat{\sigma}_2$ and inequality (B.52) reduces to

$$\frac{1}{2}\left(1-\frac{\hat{\sigma}_2}{2}\right)(1-\mu^*)^2 > \frac{1}{2}\mu^*(1-\mu^*)\frac{\hat{\sigma}_2}{2} + \frac{1}{2}\left(1-\frac{\hat{\sigma}_2}{2}\right)(1-\mu^*)^2,$$

which is violated. At the upper bound of δ we have $\bar{s} = 1$ or, equivalently $\delta = \frac{1}{\mu^*} - 1$. Using the observation that $\sigma(\bar{s} = 1) = 0$, inequality (B.52) can be written as

$$\frac{1}{2}\left(1 - \mu^*(1-\mu^*) + 2\left(\frac{1}{\mu^*} - 1\right) - \frac{1}{\mu^*}\right) > \frac{1}{2}(1-\mu^*)^2\left(1-\frac{\hat{\sigma}_2}{2}\right),$$

or, equivalently, as

$$\frac{1}{2}(1-\mu^*)^2\frac{1+\mu^*}{\mu^*} > \frac{1}{2}(1-\mu^*)^2\left(1-\frac{\hat{\sigma}_2}{2}\right),$$

which always holds as $\frac{1+\mu^*}{\mu^*} > 1 > 1 - \frac{\hat{\sigma}_2}{2}$. Hence, there is a threshold value of δ, denoted $\widetilde{\delta}$, such that for any δ above the threshold $V(\mu^*) > V(1)$.

Hence, for δ above $\widetilde{\delta}$ we have $V(s_F^P) > V(\mu^*) > V(1) = V(0)$, where the first inequality follows from Lemma 7. Moreover, we have that $V(s_F^P) > V(\mu^*) > V(\mu^*(1+\delta))$ where the second inequality follows from $V_p(s_1)$ being decreasing for $s_1 \geq \mu^*$. For $\delta = 0$, region p disappears, and $V(0) > V(\mu^*)$ since $V_{a(i)}(s_1)$ is decreasing over region a(i) (see B.46). ■

Proof of Proposition 3.6. The proof of the proposition follows from Lemma B.1, Lemma B.2, Lemma B.3 and Lemma B.4. ■

B.2 Computation of |CAR|

In this section, we derive the absolute value of the cumulative abnormal returns implied by our model, for use in Section 3.5.2. Denote by r_{t+1} the returns around the stress test, and by p_t and p_{t+1} the bank's equity prices before and after the stress test. The bank's absolute cumulative abnormal returns when the bank is subject to an informative stress test with passing threshold s_1 is:

$$|\text{CAR}|(s_1) \equiv \mathbb{E}\big(\big|r_{t+1} - \mathbb{E}(r_{t+1})\big|\big) = \mathbb{E}\left(\left|\frac{p_{t+1} - \mathbb{E}(p_{t+1})}{p_t}\right|\right) = \mathbb{E}\left(\left|\frac{p_{t+1} - \mathbb{E}(p_{t+1})}{\mathbb{E}(p_{t+1})}\right|\right),$$

where the last equality follows from the efficient-market hypothesis, i.e. the pre-test price incorporates future information, $p_t = \mathbb{E}(p_{t+1})$.

For a bank subject to a stress test with passing threshold s_1 the bank's expected equity price is:

$$\mathbb{E}(p_{t+1}) = (1 - s_1)\left[\mu_2\sigma(\mu_2)\frac{1}{2}E_h^1(1) + \left(1 - \frac{1}{2}\sigma(\mu_2)\right)\mu_2 E_h^1(\mu_2)\right] \equiv p(s_1). \tag{B.53}$$

With probability s_1 the bank fails the test, there is no information acquisition and no funding provision, so that the price is equal to 0. With the residual probability $(1 - s_1)$, the bank passes the test. The realized price is $E_h^1(1)$ whenever the bank is financially sound and the speculator's positive signal is revealed to the market maker. Conversely, when the bank is insolvent and a negative signal is revealed to the market maker, the price is equal to 0. Finally, when the order flow is uninformative, or when the speculator observes no signal, the price is $\mu_2 E_h^1(\mu_2)$. Note that if the stress test is informative ($\mu_2 \geq \widehat{\mu}$), information acquisition $\sigma(\mu_2)$ and the expected price $p(s_1)$ are positive numbers. Following a similar reasoning we can compute $|\text{CAR}|(s_1)$:

$$\begin{aligned} |\text{CAR}|(s_1) = {} & (1 - s_1)\mu_2\sigma(\mu_2)\frac{1}{2}\left|\frac{E_h^1(1) - p(s_1)}{p(s_1)}\right| + \\ & + (1 - s_1)\left(1 - \frac{1}{2}\sigma(\mu_2)\right)\left|\frac{\mu_2 E_h^1(\mu_2) - p(s_1)}{p(s_1)}\right| + \\ & + \left((1 - s_1)(1 - \mu_2)\sigma(\mu_2)\frac{1}{2} + s_1\right)\left|\frac{0 - p(s_1)}{p(s_1)}\right|. \end{aligned} \tag{B.54}$$

For institutions that are subject to a stress test, we compute the |CAR| at the optimal passing threshold $s_F(\tau, \mu^*)$ derived in Proposition 1 (equation (3.19)), while for non-tested institutions, we consider the |CAR| implied by an uninformative stress test. In the latter case, the price and its expectation differ depending on whether the speculator acquires information at the prior beliefs (in which case the expected price is $\mathbb{E}(p_{t+1}) = p(0)$), or not ($\mathbb{E}(p_{t+1}) = 0$). The difference between the |CAR| for tested and non-tested institution is:

$$\chi(\tau, \mu^*) = \begin{cases} |\text{CAR}|\big(s_F(\tau, \mu^*)\big) - 0 & \text{if } \frac{1}{2} < \mu^* \\ |\text{CAR}|\big(s_F(\tau, \mu^*)\big) - |\text{CAR}|\big(0\big) & \text{if } \frac{1}{2} \geq \mu^*. \end{cases} \tag{B.55}$$

Chapter C

Appendix for Chapter 4

C.1 Proofs

Proof of Lemma 4.1. The opportunity stock in (4.5) can be rewritten as

$$I(e,m) = i\left[\int_0^{\iota}\exp[-em\tau]d\tau + \int_{\iota}^{\infty}\exp[-em\tau - e(\tau-\iota)\rho(1-m)]d\tau\right].$$

By Leibniz integral rule, for $m \in [0,1)$, we have

$$\frac{\partial I}{\partial e} = -i\left[\int_0^{\iota}\exp[-em\tau](m\tau)d\tau + \int_{\iota}^{\infty}\exp[-em\tau - e(\tau-\iota)\rho(1-m)][m\tau + (\tau-\iota)\rho(1-m)]d\tau\right] < 0,$$

$$\frac{\partial I}{\partial m} = -i\left[\int_0^{\iota}\exp[-em\tau](e\tau)d\tau + \int_{\iota}^{\infty}\exp[-em\tau - e(\tau-\iota)\rho(1-m)][e\tau - e(\tau-\iota)\rho]d\tau\right] < 0,$$

$$\frac{\partial I}{\partial \rho} = -i\left[\int_{\iota}^{\infty}\exp[-em\tau - e(\tau-\iota)\rho(1-m)][e(\tau-\iota)(1-m)]d\tau\right] < 0,$$

$$\frac{\partial I}{\partial \iota} = i\left[\int_{\iota}^{\infty}\exp[-em\tau - e(\tau-\iota)\rho(1-m)][e\rho(1-m)]d\tau\right] > 0,$$

where the inequalities follow from the fact that the integrands are strictly positive for $m \in [0,1)$. The preemption risk in (4.7) can be rewritten as

$$p(e,m) = [1 + k(m)]^{-1}$$

where

$$k(m) \equiv \frac{m}{\rho + m(1-\rho)}\frac{\exp(-e\iota m)}{1-\exp(-e\iota m)}.$$

For $m \in (0,1)$ we have

$$\frac{\partial p}{\partial e} = p(e,m)^2 k(m) \frac{\iota m}{1-\exp(-e\iota m)} > 0,$$

$$\frac{\partial p}{\partial m} = p(e,m)^2 k(m) \left[\frac{e\iota}{1-\exp(-e\iota m)} - \frac{\rho}{m[\rho + m(1-\rho)]}\right] > 0,$$

$$\frac{\partial p}{\partial \rho} = p(e,m)^2 k(m) \frac{(1-m)}{\rho + m(1-\rho)} > 0,$$

$$\frac{\partial p}{\partial \iota} = p(e,m)^2 k(m) \frac{em}{1-\exp(-e\iota m)} > 0,$$

where the second inequality follows from the fact that $\frac{e\iota m}{1-\exp(-e\iota m)} > 1 > \frac{\rho}{\rho+m(1-\rho)}$ for $m > 0$, while the other inequalities are trivial. The same inequalities are satisfied when m tends to 0, since by L'Hôpital's rule we have

$$\lim_{m\to 0} k(m) = \frac{1}{e\iota\rho},$$

$$\lim_{m\to 0} \frac{m}{1-\exp(-e\iota m)} = \frac{1}{e\iota},$$

$$\lim_{m\to 0} \left[\frac{e\iota}{1-\exp(-e\iota m)} - \frac{\rho}{m[\rho + m(1-\rho)]}\right] = \frac{1-\rho}{\rho} + \frac{e\iota}{2},$$

so that

$$\lim_{m\to 0} \frac{\partial p}{\partial e} = \frac{\iota\rho}{(1+e\iota\rho)^2} > 0,$$

$$\lim_{m\to 0} \frac{\partial p}{\partial m} = \frac{e\iota\rho}{(1+e\iota\rho)^2}\left(\frac{1-\rho}{\rho} + \frac{e\iota}{2}\right) > 0,$$

$$\lim_{m\to 0} \frac{\partial p}{\partial \rho} = \frac{e\iota}{(1+e\iota\rho)^2} > 0,$$

$$\lim_{m\to 0} \frac{\partial p}{\partial \iota} = \frac{e\rho}{(1+e\iota\rho)^2} > 0.$$

■

Proof of Lemma 4.2. We showed in the main text that the firm's total profit only depends on average search e and average manipulation m. To complete the proof, we show now that for any contract contingent on specific profit paths $\{\Pi_t\}_{t\in[0,D]}$, there exists an *equivalent* contract that is contingent on average profit $\hat{\Pi}$ only: Under both contracts, the trader's *average* search and manipulation, and payments are the same in equilibrium.

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

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

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

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

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

Average profit is

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

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

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

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

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

and

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

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

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

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

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

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

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

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

Consider the strategy

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

By construction (and using (4.12))

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

and

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

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

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

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

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

Now consider a trader's strategy choice when contracts are made contingent only on average profits and post-trade reviews, that is,

$$\mathcal{W} = \{\underline{W}(\bar{\Pi}), \overline{W}(\bar{\Pi})\}.$$

Given such a contract, a trader chooses a trading strategy that maximizes

$$(1 - \gamma\bar{m})\overline{W}(\bar{\Pi}(\{e_t, m_t\}_D)) + \gamma\bar{m}\underline{W}(\bar{\Pi}(\{e_t, m_t\}_D) - \int_0^D ce_t dt.$$

where $\bar{\Pi}(\{e_t, m_t\}_D)$ denotes the average profit generated by the dynamic trading strategy $\{e_t, m_t\}_D$. Recall from (4.12) that average profit ultimately depends only on average search and manipulation $(\bar{e}, \bar{m})$, so the above objective function be rewritten as

$$(1 - \gamma\bar{m})\overline{W}(\Pi(\bar{e}, \bar{m}) + \gamma\bar{m}\underline{W}(\Pi(\bar{e}, \bar{m}) - Dc\bar{e},$$

which depends solely on the average search intensity and average manipulation rate. Hence, there exists a (weakly) optimal time-invariant trader strategy. ■

Proof of Lemma 4.3. Consider a contract $\{\overline{W}(.), \underline{W}(.)\}$. The trader chooses a search effort e and a manipulation level m to maximize

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

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

$$\overline{W}'(\Pi) \equiv \begin{cases} \overline{W}(\Pi) & \text{if } \Pi = \Pi(e^*, m^*), \\ 0 & \text{otherwise.} \end{cases} \tag{C.1}$$

and

$$\underline{W}'(\Pi) \equiv \begin{cases} \underline{W}(\Pi) & \text{if } \Pi = \Pi(e^*, m^*), \\ 0 & \text{otherwise.} \end{cases} \tag{C.2}$$

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

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

which implies

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

and hence, (e^*, m^*) weakly dominates no effort. As the two contracts elicit the same effort and manipulation levels for the same equilibrium payment, they are equivalent from the point of view of the firm, and we therefore restrict attention to contracts that have the form described in (C.1) and (C.2).

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

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

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

$$(1 - \gamma m^*)\overline{W}[\Pi(e^*, m^*)] + \gamma m^* \underline{W}[\Pi(e^*, m^*)] \geq cDe^*, \tag{C.3}$$

$$\gamma \left\{\overline{W}[\Pi(e^*, m^*)] - \underline{W}[\Pi(e^*, m^*)]\right\} \geq -cD\frac{\delta e}{\delta m}[m^*, \Pi(e^*, m^*)], \tag{C.4}$$

where (C.4) holds with equality if $m^* \in (0, 1)$. (C.3) states that the trader has to be compensated at minimum for his cost of effort. (C.4) states that within the set of effort-manipulation couples (e, m) that generate the same profit $\Pi(e^*, m^*)$, (e^*, m^*) has to be optimal.

We start by deriving levels of compliance that can be achieved "for free", that is, without leaving a rent to the trader. If the trader has no rent then (C.3) must hold with equality, so that the expected payment to the trader per unit of time is

$$\mathbb{E}(w_t) = \frac{1}{D}\Big[(1 - \gamma m^*)\overline{W}[\Pi(e^*, m^*)] + \gamma m^* \underline{W}[\Pi(e^*, m^*)]\Big] = ce^*.$$

Note that this leaves us with one degree freedom to set $\overline{W}[\Pi(e^*, m^*)]$ relative to $\underline{W}[\Pi(e^*, m^*)]$. We use (C.3) to make substitutions in (C.4), we also use

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

and obtain, after rearranging,

$$\left[\frac{\gamma}{1-\gamma m^*} - \frac{1-(1-p)\rho}{m^* + (1-m^*)(1-p)\rho}\right] cDe^* - \frac{\gamma}{1-\gamma m^*}\underline{W}[\Pi(e^*, m^*)] \geq 0. \tag{C.5}$$

Maximum manipulation $m^* = 1$ can be obtained by setting $\overline{W}[\Pi(e^*, m^*)] = \underline{W}[\Pi(e^*, m^*)]$. The lowest m^* that can be obtained with no agency rent obtains from (C.5) taking $\underline{W}[\Pi(e^*, m^*)] = 0$. This lower bound was defined in the main text as

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

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

It follows from the above analysis that any contract that incentivizes a level of manipulation m^* strictly lower than m_0 must leave the trader with a rent. In that case, (C.3) is slack, and in the cost-minimal contract, (C.4) must hold with equality. In addition, $\underline{W}[\Pi(e^*, m^*)] = 0$. Indeed, if $\underline{W}[\Pi(e^*, m^*)]$ was strictly positive, decreasing it while keeping $\overline{W}[\Pi(e^*, m^*)] - \underline{W}[\Pi(e^*, m^*)]$ constant would not affect (C.4), but would reduce payments to the trader. (Since (C.3) is slack, there exists some room to lower payments while keeping the search constraint satisfied.) (C.4) holding with equality pins down the minimum payment $\overline{W}[\Pi(e^*, m^*)]$ to the trader, conditional on no manipulation detection, and in turn, his agency rent. All in all, for $m < m_0$, the cost-minimal contract is such that

$$\begin{aligned}
&\underline{W}[\Pi(e^*, m^*)] = 0,\\
&\overline{W}[\Pi(e^*, m^*)] = \frac{1}{\gamma}\frac{1-(1-p)\rho}{(1-p)\rho + m^*(1-(1-p)\rho)}cDe^*,
\end{aligned}$$

and the expected payment to the trader per unit of time is

$$\begin{aligned}
\mathbb{E}(w_t) &= \frac{1}{D}\Big[(1-\gamma m^*)\overline{W}[\Pi(e^*, m^*)] + \gamma m^* \underline{W}[\Pi(e^*, m^*)]\Big],\\
&= (1-\gamma m^*)\frac{1}{\gamma}\frac{1-(1-p)\rho}{(1-p)\rho + m^*(1-(1-p)\rho)}ce^*,\\
&= ce^* + \frac{2(1-(1-p)\rho)}{(1-p)\rho + m^*(1-(1-p)\rho)}\left[\frac{1}{2}\left[\frac{1}{\gamma} - \frac{(1-p)\rho}{1-(1-p)\rho}\right] - m^*\right] ce^*.
\end{aligned}$$

Finally, note that one can make the payment monotonic in the profit $\overline{\Pi}$: for $\overline{\Pi} > \Pi(e^*, m^*)$, setting $\overline{W}(\overline{\Pi}) = \overline{W}[\Pi(e^*, m^*)]$ rather than $\overline{W}(\overline{\Pi}) = 0$ (and respectively for

$\underline{W}(\overline{\Pi})$) is still incentive compatible. Indeed, suppose the trader chooses (e, m) such that $\Pi(e, m) > \Pi(e^*, m^*)$. Then by reducing e, the trader saves on the cost of effort while keeping the expected payment unchanged. It follows that the trader's relevant choice set reduces to (e, m) such that $\Pi(e, m) = \Pi(e^*, m^*)$. ■

Proof of Proposition 4.1. Rewrite the agency rent as

$$r(e, m) = K(m, p)(m_0 - m)ce,$$

where

$$K(m, p) \equiv \frac{2}{m + \frac{(1-p)\rho}{1-(1-p)\rho}} > 0.$$

Since

$$\frac{\partial K}{\partial m} = -\frac{1}{2}K(m, p)^2,$$

the marginal cost of incentivizing risk management when $m_0 \in (0, 1)$ and $m < m_0$ is

$$\begin{aligned} -\frac{\partial r}{\partial m} &= \left[\frac{1}{2}K(m, p)^2(m_0 - m) + K(m, p)\right] ce, \\ &= \frac{\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}}{\left(m + \frac{(1-p)\rho}{1-(1-p)\rho}\right)^2} ce, \\ &= B(x(p, \rho))ce, \end{aligned} \tag{C.6}$$

where $B(x) \equiv \frac{\frac{1}{\gamma}+x}{(m+x)^2}$ and $x(p, \rho) \equiv \frac{(1-p)\rho}{1-(1-p)\rho}$. Note that

$$\begin{aligned} \frac{\partial B}{\partial x} &= -\frac{\frac{2}{\gamma} + x - m}{(m + x)^3} < 0, \\ \frac{\partial x}{\partial p} &= -\frac{\rho}{(1 - (1 - p)\rho)^2} < 0, \\ \frac{\partial x}{\partial \rho} &= \frac{1 - p}{(1 - (1 - p)\rho)^2} > 0, \\ \frac{\partial B}{\partial \gamma} &= -\frac{1}{(m + x(p, \rho))^2}\frac{1}{\gamma^2} < 0, \end{aligned} \tag{C.7}$$

where the first line follows from the fact that $m < 1 < 2/\gamma$ and $x \geq 0$ while the third line from the fact that the highest value of p is $p(e, 1) = 1 - \exp(-e\iota) < 1$. We have the following comparative statics:

$$\begin{aligned} \frac{\partial}{\partial p}\left(-\frac{\partial r}{\partial m}\right) &= \frac{\partial B}{\partial x}\frac{\partial x}{\partial p}ce > 0, \\ \frac{\partial}{\partial \rho}\left(-\frac{\partial r}{\partial m}\right) &= \frac{\partial B}{\partial x}\frac{\partial x}{\partial \rho}ce < 0, \\ \frac{\partial}{\partial \gamma}\left(-\frac{\partial r}{\partial m}\right) &= \frac{\partial B}{\partial \gamma}ce < 0. \end{aligned} \tag{C.8}$$

■

Proof of Lemma 4.4. Assuming optimal manipulation $m = m^*$, the firm's optimal search effort e^* solves:

$$\max_{e \in [0, \lambda^{-1}]} V(e, m^*) \tag{C.9}$$

where

$$V(e, m^*) \equiv Ie\big[m^*\pi + (1 - m^*)(1 - p)\rho(\pi + \alpha_+)\big] - ce - \max\{0, r(e, m^*)\}. \tag{C.10}$$

Note that $V(e, m^*)$ is linear in e, hence $e^* \in \{0, \lambda^{-1}\}$. Moreover, $V(e, m^*) \geq V(e, 1) = Ie\pi - ce$, since optimal manipulation is weakly preferred to full manipulation and $r(e, 1) = 0$. It follows, that if $I\lambda^{-1}\pi - c\lambda^{-1} = V(\lambda^{-1}, 1) > V(0, 1) = 0$, then $V(\lambda^{-1}, m^*) > V(0, m^*)$. Under assumption (A.2), the last condition is satisfied even for the lowest value of I, that is $I = i\lambda$, achieved when $e = \lambda^{-1}$, $\iota = 0$ and $\rho = 1$. It follows that it is optimal to implement $e^* = \lambda^{-1}$.

■

Proof of Proposition 4.2. First, we show that the best response level of search effort and manipulation are, respectively, $e^* = \lambda^{-1}$ and

$$m^*(p) = \begin{cases} 0 & \text{if} \quad p < p^* \\ [0, 1] & \text{if} \quad p = p^* \\ 1 & \text{if} \quad p^* < p \end{cases} \tag{C.11}$$

where $p^* = 1 - \frac{\pi}{(\pi + \alpha_+)\rho} \in (0, 1)$ was defined in (C.17).

Assuming optimal manipulation $m = m^*$, the firm's optimal search effort e^* solves:

$$\max_{e \in [0, \lambda^{-1}]} V(e, m^*) \tag{C.12}$$

where

$$V(e, m^*) \equiv Ie\big[m^*\pi + (1 - m^*)(1 - p)\rho(\pi + \alpha_+)\big] - ce. \tag{C.13}$$

Note that $V(e, m^*)$ is linear in e, hence $e^* \in \{0, \lambda^{-1}\}$. Moreover, $V(e, m^*) \geq V(e, 1) = Ie\pi - ce$, since optimal manipulation is weakly preferred to full manipulation. It follows, that if $I\lambda^{-1}\pi - c\lambda^{-1} = V(\lambda^{-1}, 1) > V(0, 1) = 0$, then $V(\lambda^{-1}, m^*) > V(0, m^*)$. Under assumption (A.2), the above condition, is satisfied even for the lowest value of I, that is $I = i\lambda$, achieved when $e = \lambda^{-1}$, $\iota = 0$ and $\rho = 1$. It follows that it is optimal to choose $e^* = \lambda^{-1}$.

Given the optimal search effort, the best response manipulation level $m^*(p)$ solves

$$\max_{m \in [0, 1]} I\lambda^{-1}\big[m\pi + (1 - m)(1 - p)\rho(\pi + \alpha_+)\big] - c\lambda^{-1}. \tag{C.14}$$

If $p > p^*$, we have $\pi - (1 - p)\rho(\pi + \alpha_+) > 0$ so that the objective is strictly increasing, and the optimum is $m^* = 1$. If $p = p^*$ the objective is flat, hence the firm is indifferent among all the $m \in [0, 1]$. If $p < p^*$, the objective is strictly decreasing, and the optimum is $m^* = 0$.

Let $p(m) \equiv p(\lambda^{-1}, m)$ and $\widehat{m}(p) \equiv p^{-1}(p)$ be its inverse function. An equilibrium is an intersection of $m^*(p)$ and $\widehat{m}(p)$. Note that $\widehat{m}'(p) = 1/p'(m)$ and $\frac{\partial \widehat{m}}{\partial \iota} = -\frac{\partial p}{\partial \iota}/p'(m)$, hence by Lemma 4.1 we have that $\widehat{m}(p)$ is strictly increasing in p and strictly decreasing in ι. Moreover, $p(0) = \iota/(\iota + \lambda/\rho) > 0$ and $p(1) = 1 - \exp(-\iota/\lambda) < 1$, hence $\widehat{m}(p) = 0$ for $p = p_*(\iota) > 0$ and $\widehat{m}(p) = 1$ for $p = p_{**}(\iota) \in (p_*(\iota), 1)$ where both $p_*(\iota)$ and $p_{**}(\iota)$ are strictly increasing, approach 0 when ι tends to 0 and approach 1 when ι tends to infinity. It follows that there exists a threshold $\underline{\iota}$ such that, for all $\iota < \underline{\iota}$, we have $p_{**}(\iota) < p^*$, and $\widehat{m}(p)$ intersects $m^\star(p)$ only once, at the point $(p, m) = (p_*(\iota), 0)$. Moreover, there exists a threshold $\bar{\iota} > \underline{\iota}$ such that, for all $\iota > \bar{\iota}$, we have $p_*(\iota) > p^*$, and $\widehat{m}(p)$ intersects $m^\star(p)$ only once, at the point $(p, m) = (p_{**}(\iota), 1)$. When $\iota \in (\underline{\iota}, \bar{\iota})$, the function $\widehat{m}(p)$ intersects $m^\star(p)$ at exactly three points:

$$(p, m) \in \left\{ (p_*(\iota), 0)\,,\ (p_{**}(\iota), 1)\,,\ (p^*, \widehat{m}(p^*)) \right\},$$

where $\widehat{m}(p^*) \in (0, 1)$. ■

Proof of Lemma 4.5. By Lemma 4.4, we have $e^* = \lambda^{-1}$. Hence, the firm's best response $m^*(p)$ solves:

$$\max_{m \in [0,1]} V(m)$$

where

$$V(m) \equiv V(\lambda^{-1}, m) = I\lambda^{-1}\big[m\pi + (1-m)(1-p)\rho(\pi + \alpha_+)\big] - c\lambda^{-1} - \max\{0, r(m)\},$$

$$r(m) \equiv r(\lambda^{-1}, m) = \frac{2(1 - (1-p)\rho)}{(1-p)\rho + (1 - (1-p)\rho)m}(m_0(p) - m)c\lambda^{-1}. \tag{C.15}$$

First, note that the objective function is continuous in m and that

$$V'(m) = \begin{cases} I\lambda^{-1}[\pi - (1-p)\rho(\pi + \alpha_+)] + \dfrac{\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}}{\left(m + \frac{(1-p)\rho}{1-(1-p)\rho}\right)^2}c\lambda^{-1} & \text{if } m < m_0 \\ I\lambda^{-1}[\pi - (1-p)\rho(\pi + \alpha_+)] & \text{if } m \geq m_0, \end{cases}$$

where the first line uses Equation (C.6). The objective is concave for $m < m_0$, and linear otherwise. Moreover, we have that when $0 < m_0 < 1$ and m approaches m_0 from the left, the derivative takes value

$$V'(m_0(p)^-) = I\lambda^{-1}[\pi - (1-p)\rho(\pi + \alpha_+)] + \frac{4}{\left(\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}\right)}c\lambda^{-1} \equiv a(p, I), \tag{C.16}$$

where the second addend is strictly positive.

Let

$$p^* \equiv 1 - \frac{\pi}{(\pi + \alpha_+)\rho} \in (0, 1). \tag{C.17}$$

If $p > p^*$, we have $\pi - (1-p)\rho(\pi + \alpha_+) > 0$ so that the derivative of the linear part of the objective is strictly positive and the left derivative at m_0 is bigger. Hence, the objective is globally increasing and the optimum is $m = 1$.

If $p = p^*$ the objective is flat for $m \geq m_0(p)$ and strictly increasing for $m < m_0(p)$. Hence the firm is indifferent among all the $m \in [m_0(p^*), 1]$.

If $p < p^*$, the linear part of the objective is decreasing, hence the optimum must be in $[0, m_0]$. Suppose $m_0 > 0$. The maximum is $m = m_0$ when $a(p, I)$ defined in (C.16) is positive (i.e. the concave part is increasing over $[0, m_0(p))$), which is the case when $\max\{0, \overline{p}(I)\} \leq p < p^*$ where

$$\overline{p}(I) \equiv \Big\{ p < p^* : a(p, I) = 0 \Big\} \tag{C.18}$$

is strictly increasing. To see this note that $a(p, I)$ is strictly increasing in p (by the second line of (C.7)), strictly decreasing in I (since $p < p^*$), and

$$\begin{aligned} a(0, I) &= I\lambda^{-1}[\pi - \rho(\pi + \alpha_+)] + \frac{4}{\left(\frac{1}{\gamma} + \frac{\rho}{1-\rho}\right)} c\lambda^{-1}, \\ a(p^*, I) &= \frac{4}{\left(\frac{1}{\gamma} + \max\left\{0, \frac{\pi}{\alpha_+}\right\}\right)} c\lambda^{-1} > 0, \end{aligned}$$

where $a(0, I)$ can be either positive or negative. The threshold $\overline{p}(I)$ is strictly positive if $a(0, I) < 0$, which is the case for

$$I > \frac{4}{\frac{1}{\gamma} + \frac{\rho}{1-\rho}} \; \frac{1}{\rho(\pi + \alpha_+) - \pi} c \equiv \underline{I}. \tag{C.19}$$

For $p < \overline{p}(I)$, the optimum must be strictly below $m_0(p)$ and solve the first-order condition

$$I[\pi - (1-p)\rho(\pi + \alpha_+)] + \frac{\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}}{\left(m + \frac{(1-p)\rho}{1-(1-p)\rho}\right)^2} c = 0, \tag{C.20}$$

which after rearranging yields

$$m = \sqrt{\frac{\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}}{[(1-p)\rho(\pi + \alpha_+) - \pi]} \frac{c}{I}} - \frac{(1-p)\rho}{1-(1-p)\rho} \equiv m^\circ(p, I) \tag{C.21}$$

which is strictly increasing in p and strictly decreasing in I.

The optimum $m^\circ(p, I)$ eventually reaches its lower bound when the left-hand side of (C.20) evaluated at $m = 0$, that is

$$b(p, I) \equiv I[\pi - (1-p)\rho(\pi + \alpha_+)] + \frac{\frac{1}{\gamma} + \frac{(1-p)\rho}{1-(1-p)\rho}}{\left(\frac{(1-p)\rho}{1-(1-p)\rho}\right)^2} c,$$

is weakly negative, which is the case when $p \leq \underline{p}(I)$ and $\underline{p}(I) \geq 0$ where

$$\underline{p}(I) \equiv \Big\{p < \overline{p}(I) : b(p, I) = 0\Big\}, \tag{C.22}$$

is strictly increasing. To see this note that $b(p, I)$ is strictly increasing in p and strictly decreasing in I, and

$$\begin{aligned} b(0, I) =& I[\pi - \rho(\pi + \alpha_+)] + \frac{\frac{1}{\gamma} + \frac{\rho}{1-\rho}}{\left(\frac{\rho}{1-\rho}\right)^2} c \\ b(\overline{p}, I) >& 0, \end{aligned}$$

where $b(0, I)$ can be either positive or negative. The threshold $\underline{p}(I)$ is strictly positive if $b(0, I) < 0$, which is the case for

$$I > \frac{\frac{1}{\gamma} + \frac{\rho}{1-\rho}}{\left(\frac{\rho}{1-\rho}\right)^2} \frac{1}{\rho(\pi + \alpha_+) - \pi} c \equiv \overline{I} > \underline{I}. \tag{C.23}$$

Lastly, note that $m^*(p, I)$ is weakly increasing in p, since $m_0(p)$ is weakly increasing in p and $m^\circ(p, I)$ is strictly increasing in p. Second $m^*(p, I)$ is weakly decreasing in I, since $m^\circ(p, I)$ is strictly decreasing in I and $m_0(p)$ is independent of I. ■

Proof of Proposition 4.3. Consider the best response level of manipulation given by Lemma 4.5 and Lemma 4.2. Let $\mathcal{S}_m(p)$ and $\mathcal{S}_m^{ag}(p, I)$ denote the set of m that can be supported in equilibrium, respectively, in the setting without and with agency:

$$\begin{aligned} \mathcal{S}_m(p) &= \begin{cases} 0 & \text{if} \quad p < p^* \\ [0, 1] & \text{if} \quad p = p^* \\ 1 & \text{if} \quad p > p^*, \end{cases} \\ \mathcal{S}_m^{ag}(p) &= \begin{cases} \{0, m^\circ(p), m_0(p)\} & \text{if} \quad p < p^* \\ [m_0(p^*), 1] & \text{if} \quad p = p^* \\ 1 & \text{if} \quad p > p^*. \end{cases} \end{aligned} \tag{C.24}$$

Note that

$$\begin{aligned} \min \mathcal{S}_m(p) &\begin{cases} = 0 & \text{if} \quad p \leq p^* \\ = 1 & \text{if} \quad p > p^*, \end{cases} \\ \min \mathcal{S}_m^{ag}(p) &\begin{cases} \geq 0 & \text{if} \quad p \leq p^* \\ = 1 & \text{if} \quad p > p^*, \end{cases} \end{aligned} \tag{C.25}$$

and that

$$\max \mathcal{S}_m(p) \begin{cases} = 0 & \text{if} \quad p < p^* \\ = 1 & \text{if} \quad p \geq p^*, \end{cases}$$
$$\max \mathcal{S}_m^{ag}(p) \begin{cases} \geq 0 & \text{if} \quad p < p^* \\ = 1 & \text{if} \quad p \geq p^*. \end{cases} \tag{C.26}$$

■

Proof of Proposition 4.4. When $\iota = 0$ we have $p = 0$. The best response level of manipulation simplifies to (suppressing dependency on p):

$$m^*(I) = \begin{cases} m_0 & \text{if} \quad I \leq \underline{I} \\ m^\circ(I) & \text{if} \quad \underline{I} < I < \overline{I} \\ 0 & \text{if} \quad \overline{I} \leq I \end{cases} \tag{C.27}$$

where $m^*(I)$ is continuous, $m^\circ(I)$ is strictly decreasing, while $\underline{I}$ and $\overline{I}$ are defined in (C.19) and (C.23), respectively.

Let

$$\widehat{m}(I) \equiv I^{-1}(I) = -\frac{\rho}{1-\rho} + \frac{\lambda i}{I(1-\rho)} \tag{C.28}$$

be the inverse of the mass of trading opportunities $I(m) \equiv I(\lambda^{-1}, m)$ where $I(e, m)$ is defined in (4.5). An equilibrium is an intersection of $\widehat{m}(I)$ and $m^*(I)$. Note that

$$m^{*\prime}(I) = \begin{cases} 0 & \text{if} \quad I \leq \underline{I} \\ -\frac{\frac{\rho}{1-\rho} + m^\circ(I)}{2I} & \text{if} \quad \underline{I} < I < \overline{I} \\ 0 & \text{if} \quad \overline{I} \leq I \end{cases},$$
$$\widehat{m}'(I) = -\frac{\frac{\rho}{1-\rho} + \widehat{m}(I)}{I},$$

hence $m^{*\prime}(I) \leq 0$ and $\widehat{m}'(I) < 0$. Moreover, if an equilibrium exists (so that $\widehat{m}(I) = m^*(I)$), we have $\widehat{m}'(I) < m^{*\prime}(I)$ implying that $\widehat{m}(I)$ crosses $m^*(I)$ from above. This, together with the continuity of both $m^*(I)$ and $\widehat{m}(I)$, implies that there exists at most one equilibrium. An equilibrium always exists as $\widehat{m}(I(1)) = 1 \geq m^*(I(1))$ and $\widehat{m}(I(0)) = 0 \leq m^*(I(0))$.

Note that m^* is increasing in π (since both $\underline{I}$ and $\overline{I}$ are increasing and continuous in π, and so is $m^\circ(p, I)$) while $\widehat{m}$ is independent of π. It follows that the equilibrium level of manipulation $m^\star$ is weakly increasing in π, and there exist two ordered thresholds $\underline{\pi} < \overline{\pi}$ such that $m^\star \in (0, m_0)$ for $\pi \in (\underline{\pi}, \overline{\pi})$ and is strictly increasing in π.

The equilibrium profits for the firm are

$$I(m^\star)(m^\star \pi + (1 - m^\star)\rho(\pi + \alpha_+)) - r(m^\star).$$

Consider the level of manipulation arising from full coordination, that is the m that solves

$$\max_{m\in[0,1]} I(m)\left[m\pi + (1-m)\rho(\pi+\alpha_+)\right] - r(m).$$

The first-order derivative of the objective function is

$$I'(m)\left[m\pi + (1-m)\rho(\pi+\alpha_+)\right] + I(m)\left[\pi - \rho(\pi+\alpha_+)\right] - r'(m). \tag{C.29}$$

Suppose $\pi \in (\underline{\pi}, \overline{\pi})$ so that $m^\star \in (0, m_0)$. This derivative is strictly negative at $m = m^\star$, implying that the optimal level of manipulation under full coordination is strictly lower than in equilibrium. To see this note that $m^\star$ solves the first order condition (C.20) so that the sum of the last two addends in (C.29) is nil. Moreover, $I'(m) < 0$ by Lemma 4.1 and $[m\pi + (1-m)\rho(\pi+\alpha_+)] > 0$ (otherwise there will be no search in equilibrium). Using similar arguments, one shows that if $\pi \leq \underline{\pi}$, $m^\star = 0$ Pareto dominates $m > 0$, and if $\pi \geq \overline{\pi}$, $m^\star = m_0$ is either Pareto optimal or Pareto-dominated by $m < m_0$. ■

Proof of Proposition 4.5. Defining $I(m) \equiv I(\lambda^{-1}, m)$ and $p(m) \equiv p(\lambda^{-1}, m)$ and $X(m) \equiv m^*(p(m), I(m))$, an equilibrium level of manipulation $m^\star$ is a solution to

$$m^\star = X(m^\star),$$

while equilibrium market conditions are $p^\star = p(m^\star)$ and $I^\star = I(m^\star)$.

Pareto-Superior Equilibrium First, if several equilibria coexist, the one with the lowest $m^\star$ Pareto dominates the other ones. Indeed, suppose $m_2^\star > m_1^\star$ are equilibria. $m_1^\star$ being a best response to $I(m_1^\star)$ implies

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

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

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

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

Second, $X(m)$ crosses the 45-degrees line from above at $m^\star$, that is, $X'(m^\star) < 1$. To see this, note that $X(0) \geq 0$ and $X(1) \leq 1$, and that $X(m)$ is weakly increasing since

$$X'(m) = m_p^*\big(p(m), I(m)\big)p'(m) + m_I^*\big(p(m), I(m)\big)I'(m) \geq 0,$$

as both terms are weakly positive by Lemma 4.1 and Lemma 4.5. Since $X(m)$ is continuous and weakly increasing, and $X(0) \geq 0$, $X(1) \leq 1$, there exists a smallest fixed point $m^\star \in [0,1]$ such that $X(m^\star) = m^\star$. For any $m < m^\star$, we must have $X(m) > m$, otherwise we would contradict the minimality of $m^\star$. Therefore, $X(m)$ crosses the 45-degree line from above at $m^\star$.

Trader's Compensation and Market Conditions We now study how the trader's compensation changes with equilibrium market conditions, $p^\star$ and $I^\star$. The trader's rent is nil whenever $m^\star \geq m_0$ and positive otherwise. When $m^\star = 0$ it takes value

$$r(m^\star) = \frac{2(1-(1-p^\star)\rho)}{(1-p^\star)\rho} m_0(p^\star) c\lambda^{-1}, \tag{C.30}$$

which is strictly increasing in $p^\star$ and independent of $I^\star$. When $m^\star = m^\circ(p(m^\star), I(m^\star))$, rewriting the trader's rent as

$$r(m) = \frac{2(m_0(p) - m)}{\frac{(1-p)\rho}{1-(1-p)\rho} + m} c\lambda^{-1}$$

and plugging $m^\star = m^\circ(p^\star, I^\star)$ into $r(m)$ yields

$$r(m^\star) = \left[\sqrt{\left(\frac{1}{\gamma} + \frac{(1-p^\star)\rho}{1-(1-p^\star)\rho}\right)\frac{[(1-p^\star)\rho(\pi+\alpha_+) - \pi]I^\star}{c}} - 2\right] c\lambda^{-1}, \tag{C.31}$$

which is decreasing in $p^\star$ and increasing in $I^\star$.

Consider now the S/I ratio. The derivation of the S/I-ratio is as follows. Let F denote the fixed component of trader's wage. The incentive compatibility constraints respectively for search and compliance become:

$$\begin{aligned}(1-\gamma)\left[\overline{W}[\Pi(\lambda, m)] + F\right] &\geq F + cD\lambda, \\ \gamma\left[\overline{W}[\Pi(\lambda, m)] + F\right] &\geq \frac{1-(1-p)\rho}{m + (1-m)(1-p)\rho} cD\lambda.\end{aligned}$$

1. If $m \geq m_0$, the incentive compatibility constraint for search is binding when $F = 0$, which is therefore the highest possible fixed wage in an optimal contract.

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

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

Consider now equilibrium behavior of the S/I ratio when it is positive (i.e. $m^\star < m_0$). When $m^\star = 0$, the ratio is proportional to $m_0(p^\star)$ which is increasing in $p^\star$ and independent of $I^\star$. When $m^\star = m^\circ(p^\star, I^\star)$ we have

$$2\gamma(m_0(p^\star) - m^\circ(p^\star, I^\star)) = \sqrt{\frac{1}{\gamma} + x(p^\star)}\left[\frac{1}{2}\sqrt{\frac{1}{\gamma} + x(p^\star)} - \sqrt{\frac{1}{[(1-p^\star)\rho(\pi+\alpha_+) - \pi]}\frac{c}{I^\star}}\right], \tag{C.32}$$

which is increasing in $I^\star$ and decreasing in $p^\star$ since $x(p) \equiv \frac{(1-p)\rho}{1-(1-p)\rho}$ is decreasing.

Comparative Statics We now study how the equilibrium level of risk management and trader's compensation changes with i and λ. Equilibrium manipulation $m^\star$ weakly decreases with i. In fact, in equilibrium we have $m^\star = m^*(p(m^\star), I(m^\star))$ and by the implicit function theorem, noting that $I(m)$ directly depends on i, we have

$$\frac{\partial m^\star}{\partial i} = \frac{m_I^*\big(p(m^\star), I(m^\star)\big)\frac{\partial I(m)}{\partial i}}{1 - X'(m^\star)} \leq 0,$$

since the numerator is weakly negative by Lemma 4.1 and Lemma 4.5. The inequality is strict when $m_I^* < 0$, which is the case when the best response is determined by $m^\circ(p(m^\star), I(m^\star))$. Note also that $m^\star(i)$ exhibits discrete downward jumps as i increases and a lower fixed point is selected. The trader's equilibrium rent, as well as the rent per unit of effort, weakly increase with i. If $m^\star = 1$ or $m^\star = m_0(p(m^\star))$, then $r(m^\star) = 0$ is independent of i. If $m^\star = m^\circ(p(m^\star), I(m^\star))$ then the rent is given by Equation (C.31) and is increasing in i since by Lemma (4.1) we have

$$\frac{\delta p^\star}{\delta i} = p'(m^\star)\frac{\delta m^\star}{\delta i} < 0$$
$$\frac{\delta I^\star}{\delta i} = I'(m^\star)\frac{\delta m^\star}{\delta i} + \frac{\delta I}{\delta i} > 0,$$

and $r^\star$ decreases with $p^\star$ and increases with $I^\star$, as shown before. When $m^\star = 0$, the rent is given by Equation (C.30) and is independent of i since $p^\star = p(0) = \left(\iota + \frac{\lambda}{\rho}\right)^{-1} \iota$ is independent of i. Using similar arguments one show that the S/I ratio exhibits a similar behavior and weakly increases with i.

Consider now the comparative static with respect to λ. By the implicit function theorem, noting that both $p(m)$ and $I(m)$ directly depend on $e^\star = \lambda^{-1}$, we have

$$\frac{\partial m^\star}{\partial \lambda} = \frac{m_p^*\big(p(m^\star), I(m^\star)\big)\frac{\partial p(m)}{\partial \lambda} + m_I^*\big(p(m^\star), I(m^\star)\big)\frac{\partial I(m)}{\partial \lambda}}{1 - X'(m^\star)} \leq 0,$$

since the numerator is weakly negative by Lemma 4.1 and Lemma 4.5. The inequality is strict whenever $m^\star \notin \{0, 1\}$ and the best response is determined by either $m_0(p(m^\star))$ (in which case $m_p^* > 0$ and $m_I^* = 0$) or $m^\circ(p(m^\star), I(m^\star))$ (in which case $m_p^* > 0$ and $m_I^* < 0$). Note also that $m^\star(\lambda)$ exhibits discrete downward jumps as λ increases and a lower fixed point is selected. The trader's rent per unit of effort $r(m^\star)/e^\star$ exhibits an inverted U-shaped behavior as λ increases. If $m^\star = 1$ or $m^\star = m_0(p(m^\star))$, then $r(m^\star) = 0$ is independent of λ. If $m^\star = m^\circ(p(m^\star), I(m^\star))$ then the rent is given by Equation (C.31) and the normalized rent is increasing in λ since by Lemma (4.1) we have

$$\frac{\delta p^\star}{\delta \lambda} = p'(m^\star)\frac{\delta m^\star}{\delta \lambda} + \frac{\delta p}{\delta \lambda} < 0$$
$$\frac{\delta I^\star}{\delta \lambda} = I'(m^\star)\frac{\delta m^\star}{\delta \lambda} + \frac{\delta I}{\delta \lambda} > 0,$$

and $r^\star$ decreases with $p^\star$ and increases with $I^\star$, as shown above. If $m^\star = 0$ then $r(m^\star)$ is given by Equation (C.30), hence the (normalized) rent is increasing in $p^\star = p(0) =$

$\left(\iota + \frac{\lambda}{\rho}\right)^{-1} \iota$ and thus decreasing in λ. Using similar arguments, one shows that the S/I ratio exhibits an inverted U-shaped behavior when λ increases. ■