Optimal Communication in Banking Supervision[*]

Jeong Ho (John) Kim[†] Kyungmin Kim[‡] Victoria Liu[§] Noam Tanner[¶]

March 13, 2024

Abstract

We present a model of banking supervision in which the supervisor first communicates her information to a privately informed bank and later decides whether to approve or disapprove the bank's investments. The supervisor's optimal communication strategy features "muddling" to introduce uncertainty into the bank's problem, thereby inducing the bank to act on its own signal. We demonstrate that the quality of supervision can deteriorate as the bank becomes more informed; in particular, the supervisor cannot utilize the bank's information when it is almost perfect. We propose a few ways to mitigate this problem within our framework.

JEL Classification: D82, D83, G21, G28.

Keywords: Banking Supervision, Communication Game, Information Design, Supervisory Guidance, Stress Tests.

[*] We are grateful for comments from Snehal Banerjee, Jean-Edouard Colliard, Divya Kirti, Mattia Landoni, Keeyoung Rhee, Larry Samuelson, Joel Shapiro, conference participants at the KIF-KAEA-KAFA Symposium and the EBA Policy Research Workshop, and seminar participants at UNSW Business School. The views expressed herein are those of the authors and do not necessarily represent those of the Federal Reserve Bank of Boston or the Federal Reserve system.

[†] Department of Finance, Florida State University, *Email:* johnkimjeongho@gmail.com

[‡] Department of Economics, Emory University, *Email:* kyungmin.kim@emory.edu

[§] Federal Reserve Bank of Boston, *Email:* victoria.liu@bos.frb.org

[¶] Federal Reserve Bank of Boston, *Email:* noam.tanner@bos.frb.org

1 Introduction

Banking institutions in the United States face supervisory regulations through two main channels: annual stress tests and periodic supervisory ratings. The Federal Reserve conducts stress tests during the second quarter of each year and bank examinations throughout the year to develop periodic supervisory ratings.[1] In both cases, supervisors leverage information produced with their own models, in addition to information provided by banks, to assess a bank's capital adequacy, and in turn, approve or disapprove of the bank's current investments.[2] To this end, supervisors also engage in supervisory communication prior to stress tests and bank examinations. Specifically, as part of the stress test cycle, the Federal Reserve publishes information about the stress test—hypothetical recession scenarios and some details about the models that will be used—during the first quarter of each year;[3] similarly, prior to a bank examination, supervisors write a first-day letter to tell the bank about the exam and its scope—what they will inspect—at least a few weeks before.[4] Importantly, these communications convey supervisors' views on the principal source of bank risks, i.e., supervisors' private information about the riskiness of investment opportunities.

Insofar as supervisory disapproval is costly to banks, it is readily apparent that supervisors' communication will play a key role in influencing banks' choice to engage in risky investments.[5] Despite its importance, research on the optimal design of supervisory communication to banks is relatively scare, with a few notable exceptions which we discuss in Section 2.[6]

In this paper, we study a model of supervisory communication that captures the essence of the banking supervision process outlined above.[7] Specifically, we consider a communication game between a supervisor and a bank. The supervisor prefers the bank to take high risks only when the

[1] The frequency with which a bank is examined and assigned supervisory ratings depends on its size and complexity. See Table 2 of How Federal Reserve Supervisors Do Their Jobs.

[2] We use the term "approve" to describe supervisors' action of passing a bank in its stress test or assigning the bank a satisfactory supervisory rating, in which case the bank is not required to change its current activities.

[3] See, for instance, 2023 Stress Test Scenarios and 2023 Supervisory Stress Test Methodology.

[4] See the *Bank Holding Company Supervision Manual* for a sample first-day letter.

[5] See, for example, Bräuning and Fillat (2020) for empirical evidence that large US banks have changed their portfolios in response to the Dodd-Frank Act Stress Test requirements.

[6] In contrast, there is quite a large theoretical literature on the optimal disclosure of stress test results to the public, which we also discuss in Section 2.

[7] While our model is designed to capture the banking supervision process in the US, it is more generally applicable. For example, in the United Kingdom, the Financial Services Act 2012 implemented a new regulatory framework for banks, effectively setting the bank's cost of supervisory disapproval (similar to the Dodd-Frank Act). In turn, regulators in the UK engage in ongoing dialogue with banks and conduct stress tests for banks (similar to the Fed).

state of the economy is good, whereas the bank always prefers to take high risks (conditional on approval) but incurs adjustment costs in case of disapproval. The supervisor first decides how to communicate her private information to the bank (supervisory guidance). Then, the bank decides whether to take high risks based on its private information and the supervisor's message. Finally, the supervisor approves or disapproves the bank's action (supervisory ruling).[8] We characterize the structure of optimal communication and also study how the supervisor's payoff—interpreted as social welfare—depends on the quality of private information in the hands of the bank versus the supervisor.

To understand the key issues in supervisory communication, note that the supervisor's first-best outcome is as follows: (i) when she is sufficiently optimistic or pessimistic about the state, she wants the bank to act regardless of its signal, always taking high risks or avoiding them, respectively; and (ii) only when the supervisor is moderate—neither optimistic nor pessimistic—does she want the bank to act on its signal, taking high risks if and only if its signal is good. If the supervisor employs a simple, honest strategy of fully revealing her information, then she can *never* achieve her first-best outcome. In particular, if the supervisor tells the bank to take high risks only when its signal is good and believes that the bank is following this guidance, the bank will have an incentive to always act as if its signal is good—taking high risks even when its signal is bad. This undermines the intended effects of the supervisor's transparency on the bank's behavior.

The supervisor can overcome, at least partially, the above problem due to the bank's incentives by deliberately muddling her communication.[9] Specifically, suppose the supervisor now tells the bank to act on its signal not only when she is moderate, but also when she is sufficiently pessimistic so that she would disapprove of the bank's high risk-taking even if the bank's signal were good. This introduces uncertainty into the bank's problem as it becomes vital for the bank to assess the likelihood that the supervisor is moderate rather than pessimistic. Crucially, the bank assigns a relatively higher probability to a moderate supervisor when its own signal is good than when it is bad. This is due to the fact that the supervisor's information and the bank's information are correlated: when the bank's signal is good, it is more likely that the state is indeed good and so the supervisor will also be more optimistic. The bank is willing to follow the supervisor's recommendation to act

[8] As will become clear shortly, it is crucial that the supervisor can learn (infer) about the bank's information only through its action. The supervisor's problem would become trivial if she can directly observe the bank's information.

[9] An example of such muddling is the recent move towards multiple scenarios in stress testing. This is an effort to "capture a wide range of outcomes for the banking system" and prevent "[stress] test[s] [from] becom[ing] predictable" as the Fed's Vice Chair Michael Barr said in his 2023 speech "Multiple Scenarios in Stress Testing."

on its signal when it assigns a high enough probability to a moderate supervisor when its signal is good and a low enough probability when its signal is bad.

We then use this insight to provide a general characterization of the supervisor's optimal communication strategy. We show that the muddling of communication allows the supervisor to obtain her first-best outcome when the bank's signal is sufficiently uninformative, whereas it is still not enough when the bank's signal is sufficiently informative. If the bank's signal is highly informative, the supervisor would base her ruling largely on the bank's signal (assuming it is available); in particular, the supervisor would disapprove based solely on her own information only when it makes her extremely pessimistic. This weakens the effectiveness of muddling to elicit information from the bank, as the supervisor cannot credibly feign enough pessimism to deter it from taking high risks when its signal is bad. In this case, the supervisor can only incentivize the bank to act on its signal for a smaller set of moderate beliefs than she would like.

Moreover, we establish a novel comparative statics result that the informativeness of the bank's signal has a non-monotonic effect on the supervisor's expected payoff. If the supervisor can obtain her first-best outcome (which we will see is the case when the bank's signal is largely uninformative), an increase in the bank signal informativeness is unambiguously good news; it simply enables the supervisor to make better-informed supervisory decisions. If the supervisor cannot obtain her first-best outcome (which we argued above is the case when the bank's signal is highly informative), however, an increase in the bank signal informativeness makes it even harder for the supervisor to incentivize the bank to act on its signal. Crucially, although the bank reveals better information, it does so less frequently, rendering the overall effect ambiguous. We show that, as the bank's signal becomes increasingly informative, the latter effect eventually dominates so that the supervisor's expected payoff declines. In the limit when the bank is perfectly informed, it reveals no information, ironically leaving the supervisor to rely only on her own (imperfect) information.

We propose two ways to improve the quality of banking supervision. First, raising the bank's cost in case of supervisory disapproval facilitates supervision by more effectively deterring the bank from blindly taking high risks.[10] Although this is straightforward in our model, it has an important policy implication because adjusting disapproval costs for banks generally involves passing legislation, which would be subject to political economy frictions, and so occurs only occasion-

[10] In practice, a bank that fails its stress test would be forced to take steps, such as cutting its dividend payouts and share buybacks, to preserve or build up its capital reserve, while a bank that has an unsatisfactory supervisory rating would be subject to more stringent standards for new business activities. Therefore, increasing the bank's cost of supervisory disapproval means strengthening its capital requirements and/or raising the bar for its new activities.

ally.[11] Our theoretical analysis then suggests that, given the occasional opportunity to change disapproval costs, it is better to err on the side of setting them too high than too low. If the costs are too high, supervisors can simply scale back how frequently they will disapprove after having told the bank to act on its signal. If the costs are set too low, however, the economy will be exposed to experiencing a welfare loss in case banks experience a sudden boost in their private information.[12]

Second, giving the supervisor commitment power over her supervisory ruling greatly improves banking supervision. In our model, the supervisor approves high risk-taking only when it is ex post optimal. If the supervisor can commit to disapproving even when doing so is ex post inefficient, she can reduce the likelihood of approval for the bank, thereby inducing it to act on its own signal more frequently. The supervisor still cannot obtain her first-best outcome, so commitment to supervisory ruling is certainly not a perfect solution. However, it renders the welfare effect of the bank's information unambiguous: with commitment over her supervisory ruling, the supervisor's expected payoff always increases as the bank's signal becomes more informative.

The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 introduces our formal model. Section 4 provides a characterization of the supervisor's first-best outcome and the condition for this to be achievable. Section 5 provides a general characterization of the supervisor's optimal communication strategy, particularly when she cannot achieve her first-best outcome. Section 6 analyzes how the supervisor's expected payoff depends on the bank's information and its cost of supervisory disapproval. Section 7 studies the case where the supervisor can communicate with the bank only through cheap talk, while Section 8 examines the case where the supervisor has commitment power over her supervisory ruling. Section 9 concludes.

2 Related Literature

A large body of research on banking regulation studies how much information about banks—which would be collected by regulators through stress testing in reality—should be disclosed to

[11] For example, the Dodd-Frank Act made all banks with assets above $50 billion subject to a much more aggressive supervisory regime, thus raising the cost for mid-sized banks; in 2018, Congress scaled back Dodd-Frank to raise the threshold for increased scrutiny of banks from $50 billion to $250 billion, thus reducing the cost for mid-sized banks.

[12] The recent collapse of Silicon Valley Bank provides an example in which the bank's cost in case of supervisory disapproval was too low: the Federal Reserve's banking supervisors knew about growing problems at SVB, but failed to take forceful action to address them due to the fact that they did not have enough power over banks that do not meet supervisory expectations (see "Fed Says It Failed to Act on Problems That Led to Silicon Valley Bank Collapse," *The Wall Street Journal*, Apr. 29, 2023). The result was the second-biggest bank failure in US history.

the public.[13] One common assumption in those papers is that the bank's portfolio ("bank type") is given. We join a number of recent papers that depart from this assumption by exploring how supervisory guidance can be used to persuade the bank to choose a particular portfolio (giving rise to an equilibrium "bank type" distribution).[14]

Most closely related among those papers is Leitner and Williams (2023), who also study optimal disclosure of the regulator's private information to the bank prior to its stress test. Specifically, they provide a characterization of the regulator's optimal communication when the bank itself is perfectly informed about the state. In contrast, our main focus is the comparative statics of the regulator's optimal communication strategy with respect to the bank's informational advantage over the regulator. This comparative-statics analysis is important since empirical evidence shows that there is significant variation in banks' ability to produce valuable information about borrowers through loan screening and monitoring (Petersen and Rajan, 1994; Berger and Udell, 1995; Sufi, 2007; Egan et al., 2022). Our contribution is to highlight that a bank's private information about its investment opportunities is not necessarily good news for the economy given the reality of banking supervision, as well as to propose ways to mitigate problems associated with the bank becoming "too" informed within our framework. Another difference is that Leitner and Williams (2023) assume that the regulator can commit to supervisory ruling, whereas we do not in our main model. This allows us to examine the welfare effect of communication separate from that of commitment over supervisory ruling.

Our modeling of communication follows the recent theoretical literature on information design;[15] this allows us to distill the strategic incentives of the supervisor without ad hoc structural assumptions. Most pertinent to our paper is Guo and Shmaya (2019), who study the sender's optimal disclosure (communication) when the receiver is also privately informed about the state.[16] They show that when the receiver's signal is binary, it is optimal for the sender to pool two disjoint

[13] A non-exhaustive list of earlier papers includes Bouvard et al. (2015), Faria-e-Castro et al. (2017), Williams (2017), Goldstein and Leitner (2018), and Orlov et al. (2023). See Goldstein and Leitner (2022) for a recent review on this earlier literature on stress test disclosure. More recent references include, among others, Huang (2021), Quigley and Walther (2023), Inostroza and Pavan (2023), Inostroza (2023), and Rhee and Dogra (2024).

[14] Other strategies that regulators can use to this effect include: likelihood of supervisory audits (Colliard, 2019; Leitner and Yilmaz, 2019), signaling of the severity of future stress test via current stress test design (Shapiro and Zeng, 2023), design of stress test scenarios (Parlatore and Philippon, 2023), as well as subsequent disclosure of stress test results to the public (Rhee and Dogra, 2024).

[15] See Bergemann and Morris (2019) and Kamenica (2019) for surveys of the literature.

[16] See Kolotilin et al. (2017) and Kolotilin (2018) for the case where the receiver's private information is independent of the sender's private information.

intervals, one consisting of the highest sender types, and the other of intermediate sender types, while separating the interval between them. This structure is similar to that of the supervisor's optimal communication strategy in our model, and is driven by the same mechanism—correlation between the sender's and the receiver's information—as in our model. However, our model differs from theirs in a few important respects. First, in Guo and Shmaya (2019), the sender intrinsically prefers the receiver to take a particular action regardless of the state, while the receiver wants to take a state-contingent optimal action, which is effectively the opposite of our model. Second, in our model, the supervisor (sender) takes a follow-up action to approve or disapprove after the bank (receiver) makes its decision, which influences the structure of optimal communication.

More broadly, this paper is related to a large empirical literature addressing the economic effects of policy uncertainty. For instance, Julio and Yook (2012) and Baker et al. (2016) find that political and policy uncertainty leads to reduced investment. More relevant to the issue at hand are Gissler et al. (2016) and Eckley et al. (2021), who offer evidence suggesting that higher regulatory uncertainty reduces bank lending and increases banks' voluntary capital surpluses, i.e., higher regulatory uncertainty induces lower risk-taking by banks. Despite the typically negative connotation associated with the effects of policy/regulatory uncertainty in this literature, we accentuate the fact that regulatory uncertainty can be an effective policy instrument in itself by characterizing when and how the supervisor ought to inject regulatory uncertainty via partial disclosure so as to make the bank behave more prudently and engage in greater risk-taking only when warranted.

3 The Model

We first present our formal model and then discuss a few key modeling assumptions.

3.1 Setup

Physical environment. There is a bank and a supervisor. The bank can choose one of two actions: take high risks ($a = h$) or take low risks ($a = \ell$). The payoff from high risk investments depends on the uncertain state of the economy ω, which can be either G (good) or B (bad). The bank's payoff is u_ω and the supervisor's payoff is v_ω. We assume that, without loss of generality, the payoff from low risk investments does not depend on ω and is normalized to zero for both the bank and the supervisor.

We maintain the following assumptions:

$$u_G = u_B = 1 > 0 \text{ and } v_G = 1 > 0 > v_B = -d.$$

These assumptions capture the following tension between the bank and the supervisor in the simplest way possible: when high risk investments turn sour (i.e., in the bad state of the economy), the bank may default, which in turn generates negative externality, say, due to spillovers (contagion) to other banks or to the rest of the economy. The bank, not internalizing this social loss, is naturally more willing to take high risks than the supervisor would prefer. Note that under our assumptions, (i) the bank prefers high risk investments to low risk investments in every state of the economy, but (ii) the supervisor prefers high risk investments only if the state of the economy is good.

After the bank decides on its risk level, the supervisor assesses the bank's risk-management practices, in which she observes the bank's risk level and decides whether to approve or disapprove of its activities. If the supervisor approves, the bank keeps the risk level intact as it chose. If the supervisor disapproves, the bank is forced to readjust its risk level to be low, incurring a rejection cost $c > 0$. The cost c can reflect the bank's cost due to a decline in its share price—stemming from the fact that it would be forced to take steps (e.g., cutting its dividend payouts and share buybacks) to preserve or build up its capital reserve—or due to its reputation loss. An implicit assumption here is that the supervisor never disapproves of a low-risk bank.

Information. Neither the bank nor the supervisor perfectly observes ω, but each of them receives a private signal about the state of the economy.[17] The prior probability that $\omega = G$ is $p_0 \in (0, 1)$. The bank's signal s is binary and can take either the value g or b. We let γ denote the probability that the bank receives a "correct" signal—$s = g$ (respectively, $s = b$) when $\omega = G$ (respectively, $\omega = B$). We assume that $\gamma \in (\frac{1}{2}, 1)$, so that the signal is indeed informative but noisy. The supervisor's signal $\widehat{s}$ is drawn from $[\underline{s}, \overline{s}]$ according to the distribution F_ω. We assume that each F_ω has a continuous and positive density f_ω, and the ratio $f_G(\widehat{s})/f_B(\widehat{s})$ is continuous and strictly increasing in $\widehat{s}$; as usual, the latter assumption implies that a higher observed value of the supervisor's signal indicates that the state is more likely to be good.

We employ the following convenient reformulation of the supervisor's signal. Each signal

[17] On the one hand, banks have a natural advantage over supervisors in acquiring superior granular information about their own portfolios. On the other hand, supervisors can aggregate information across multiple banks that enable them to identify systematic risks better and in turn assess capital adequacy of an individual bank better than the bank itself.

realization $\widehat{s}$ updates the supervisor's belief from p_0 to

$$t = \Pr\{\omega = G|\widehat{s}\} = \frac{p_0 f_G(\widehat{s})}{p_0 f_G(\widehat{s}) + (1 - p_0) f_B(\widehat{s})}.$$

By the assumed monotone likelihood ratio property, there is a one-to-one relationship between $\widehat{s} \in [\underline{s}, \overline{s}]$ and $t \in [0, 1]$. This means that we can interpret (re-label) each signal realization $\widehat{s}$ directly as the resulting posterior t. We let $F(t)$ denote the probability that the supervisor's posterior belief is less than t, i.e., $F(t) := p_0 F_G(t) + (1 - p_0) F_B(t)$. As we proceed, it will be apparent that the distribution F contains all relevant information about the supervisor's signal; thus, from now on, we refer to it as the supervisor's signal or information structure. Given F, we can also interpret the supervisor's posterior t as her "type."

Supervisory guidance. A key element of our model is that, before the bank decides on its risk level, the supervisor can disclose information about her type t. To study its full potential, we endow the supervisor with full flexibility in her communication with the bank. Formally, the supervisor can choose an arbitrary finite set M of messages and a Borel-measurable function $\pi : [0, 1] \to M$, where $\pi(t)$ denotes the message that is sent when the supervisor's type is t. The bank observes the signal structure (M, π) and updates its beliefs about the supervisor's type after observing $m \in M$. We let $F(\cdot|m)$ represent the bank's posterior belief distribution after observing $m \in M$and $a(m) \in \{h, \ell\}$ denote the bank's (observed) risk level following message $m \in M$.[18]

Timing. The timing of the game is as follows. The supervisor publicly commits to her communication strategy (M, π). Nature chooses ω, and the supervisor's type t is realized (conditional on ω). Then, the bank decides whether to take high risks ($a = h$) or low risks ($a = \ell$) after observing the realized message $\pi(t)$ and its own signal s. The supervisor assesses the bank's risk-management practices, i.e., if the bank takes high risks, she decides whether to approve or disapprove of its investments. Specifically, the supervisor approves high risk-taking only when it is optimal based on what he learns about the bank's signal (from conducting the stress test or the bank examination) and his own type t. Finally, the payoffs are realized.

[18] We assume that the bank always plays a pure strategy, choosing either h or ℓ with probability 1 after receiving any message m. See Appendix C for the analysis of the case where this assumption is relaxed.

3.2 Discussion

The bank's signal structure. One notable technical restriction of our model is that the bank's signal is binary. It allows us to derive our main economic implications in the simplest way possible and provide a sharp and clear characterization for the supervisor's optimal policy. As it becomes clear in the subsequent sections, our economic results do not crucially rely on the assumption and can be generalized beyond the tractable case we focus on. See Appendix B for a partial characterization of the case where the bank's signal is ternary.

It is important to note that the difference in the cardinality of the set of signal realizations does not imply that the supervisor's (continuous) signal is more informative about the state ω than the bank's (binary) signal. To be specific, suppose $F(\cdot)$ is uniform over $[0, 1]$. If $\gamma \leq 3/4$, then this distribution is more informative than the bank's signal in the sense of Blackwell (1951). If $\gamma \in (3/4, 1)$, however, the two signal structures cannot be clearly ranked; since $\gamma = 1$ would mean that the bank has perfect information about ω, one may argue that the bank's signal is more informative than the supervisor's if γ is sufficiently close to 1.

Commitment over communication. Our model assumes that the supervisor possesses commitment power over her communication strategy. As in the recent literature on information design, this allows us to focus on the supervisor's strategic incentives and identify her optimal communication that is not subject to any structural (ad hoc) constraint. This, however, is not merely a convenient theoretical assumption. The Federal Reserve Board has comprehensive institutional guidelines for supervisory communications which effectively amount to commitment; for example, development and communication of stress test scenarios need to follow the Board's Policy Statement on the Scenario Design Framework for Stress Testing ("Scenario Design Framework"),[19] while its *Bank Holding Company Supervision Manual* delineates a communication protocol that the supervisors effectively commit to.[20] Moreover, in Section 7, we consider the case where the supervisor, having no commitment over her communication, can use only cheap-talk messages. With cheap talk, we show that, whenever commitment power over communication matters, the supervisor dictates the bank's investment decision and the bank always follows. Therefore, the supervisor never disapproves, which is inconsistent with the fact that some banks fail stress tests or receive unsatisfactory supervisory ratings. Therefore, we believe that our assumption of the supervisor's commitment

[19] 12 C.F.R. pt. 252, Appendix A.

[20] A sample information request form is included with the manual.

over communication properly reflects the nature of her communication to the bank in reality.

Lack of commitment over supervisory ruling. As opposed to the above, our model assumes that the supervisor cannot commit to her supervisory ruling decision: the supervisor decides whether to approve or disapprove of the bank's risky investments *only* after the fact that the bank has chosen its risk level. As a result, the supervisor approves high risk-taking only when it is ex post optimal (see Section 4.1). It will be shown formally in Section 8 that the supervisor can significantly benefit from possessing commitment power over her subsequent supervisory ruling decision. In particular, the supervisor exercises this power to disapprove even when it is not ex post optimal to do so. This naturally raises the question of whether the supervisor can commit to a ruling in reality, which amounts to asking whether the supervisor can fail banks in stress tests or rate them unsatisfactory even though the evidence suggests that the banks are sufficiently capitalized or they are managing their risks well. Such ruling would be inconsistent with what stress test results and supervisory ratings are supposed to capture.[21] Therefore, we believe that our assumption of the supervisor's lack of commitment over supervisory ruling properly reflects the nature of supervisory ruling in practice, which precludes "punishing" a bank when it is not ex post optimal to do so.

4 Achieving the Supervisor's First-Best Outcome

4.1 Characterizing the First-Best Outcome

Let p denote the probability that the supervisor assigns to the event that $\omega = G$. The supervisor's expected payoff is $p - (1-p)d$ if she approves the bank's high risks and 0 if she disapproves. This means that the supervisor allows the bank's high risk investments if and only if[22]

$$p - (1-p)d \geq 0 \iff p \geq \widehat{t} := \frac{d}{1+d}.$$

[21] Moreover, banks may appeal against the Fed's supervisory ruling as described in SR-20-28/ CA-20-14, "Internal Appeals Process for Material Supervisory Determinations and Policy Statement Regarding the Ombudsman for the Federal Reserve System," or they might even contemplate legally challenging the Fed if they perceive the Fed to be subjective (see "Bank Groups Weigh Legal Challenge to Fed Stress Tests", *The Wall Street Journal*, Sept. 1, 2016).

[22] We assume that the supervisor allows high risk investments if her perceived probability of $\omega = G$ is exactly equal to $\widehat{t}$. This assumption is benign for our characterization below, as the event occurs with probability zero under the supervisor's optimal communication strategy.

In other words, in the last stage of the game, the supervisor allows high risk investments if and only if her belief that $\omega = G$ exceeds $\widehat{t}$.

The above value $\widehat{t}$ is the cutoff the supervisor will apply if she makes an approval decision based solely on her information. She will apply a different cutoff if she can also observe the bank's signal. To be formal, suppose the supervisor's type (belief) is t. If she learns that the bank's signal is g, then, by Bayes' rule, her belief updates to

$$\frac{\gamma t}{\gamma t + (1-\gamma)(1-t)} > t.$$

Similarly, if the supervisor learns that the bank's signal is b, then her belief becomes

$$\frac{(1-\gamma)t}{(1-\gamma)t + \gamma(1-t)} < t.$$

We define $\overline{t}$ to be the supervisor's belief (type) that would make her indifferent between approving and disapproving of high risk investments after learning that the bank received signal $s = b$:

$$\frac{(1-\gamma)\overline{t}}{(1-\gamma)\overline{t} + \gamma(1-\overline{t})} = \widehat{t} \iff \frac{\overline{t}}{1-\overline{t}} = \frac{\gamma}{1-\gamma}\frac{\widehat{t}}{1-\widehat{t}} = \frac{\gamma}{1-\gamma}d$$

Similarly, let $\underline{t}$ be the supervisor's type that would be indifferent between approving and disapproving of high risk investments after learning that the bank received signal $s = g$:

$$\frac{\gamma \underline{t}}{\gamma \underline{t} + (1-\gamma)(1-\underline{t})} = \widehat{t} \iff \frac{\underline{t}}{1-\underline{t}} = \frac{1-\gamma}{\gamma}\frac{\widehat{t}}{1-\widehat{t}} = \frac{1-\gamma}{\gamma}d$$

Clearly, $\underline{t} < \widehat{t} < \overline{t}$: (i) the supervisor of type $t \in \left[\underline{t}, \widehat{t}\right)$, who would disapprove of high risk investments a priori, is swayed to approve it by the bank's good news, and (ii) the supervisor of type $t \in \left[\widehat{t}, \overline{t}\right)$, who would approve of high risk investments a priori, is swayed to disapprove by the bank's bad news.

The supervisor's first-best outcome is as follows:

- if $t \geq \overline{t}$, then she wants the bank to take high risks regardless of its signal;
- if $t \in [\underline{t}, \overline{t})$, then she wants the bank to take high risks if and only if its signal is good; and
- if $t < \underline{t}$, then she wants the bank to take low risks regardless of its signal.

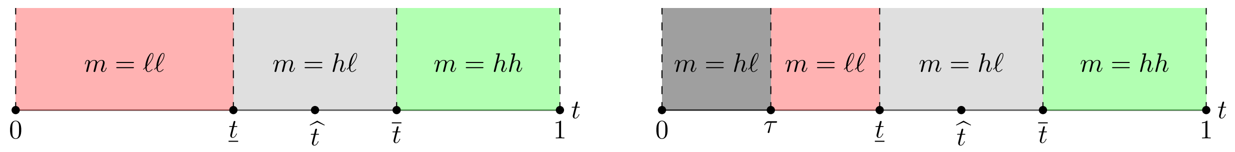


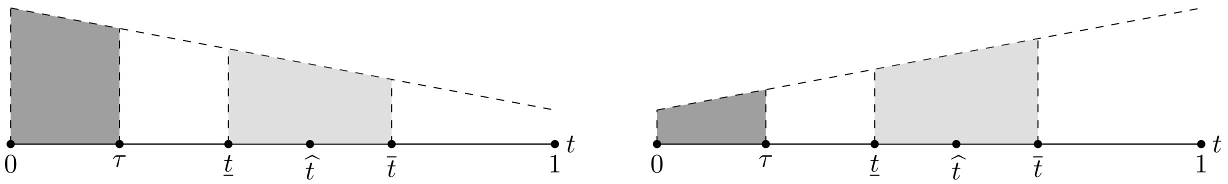
Figure 1: Two communication strategies of the supervisor.

4.2 Necessary and Sufficient Condition for Achieving the First-Best Outcome

To study when the supervisor can achieve her first-best outcome, first consider the following simple, honest communication strategy, whose structure is depicted in the left panel of Figure 1: the supervisor tells the bank to (i) take high risks whether its signal is g or b ($m = hh$) if $t \geq \bar{t}$, (ii) take high risks only when its signal is g ($m = h\ell$) if $t \in [\underline{t}, \bar{t})$, and (iii) never take high risks ($m = \ell\ell$) if $t < \underline{t}$. By construction, if the bank were to follow the supervisor's guidance, then the supervisor would achieve her first-best outcome. However, the bank will *not* follow the supervisor's guidance when the message is $h\ell$. To see this, note that the supervisor learns the bank's signal only through its action. If she thinks that the bank takes high risks only when $s = g$ and so approves whenever the bank takes high risks, then the bank will just take high risks whenever it receives the message $h\ell$, unraveling the desirable outcome.

In the current environment, the supervisor can provide an incentive for the bank only by modifying (muddling) her communication. In particular, we can use the fact that when $t < \underline{t}$, the bank taking high risks is not costly to the supervisor; she can simply disapprove, which is costly only to the bank. Specifically, consider the following modification of the previous communication strategy, described in the right panel of Figure 1: now, the supervisor sends the message $h\ell$ also when $t \leq \tau$ (for some $\tau \leq \underline{t}$) so that, following $m = h\ell$, the bank thinks that the supervisor's type (belief) could be either in $[\underline{t}, \bar{t})$ or in $[0, \tau)$.

To see why this structure can be used to provide a proper incentive for the bank, first recall that the bank's payoff is 1 if it takes high risks and the supervisor approves, $-c$ if it takes high risks but the supervisor disapproves, and 0 if it takes low risks. It then follows that the bank will take high risks if and only if the probability q that the supervisor will approve its high risk investments



Conditional on $s = b$

Conditional on $s = g$

Figure 2: The bank's belief on the supervisor's type t conditional on $m = h\ell$ and $s = b$ (left) or $s = g$ (right), when the underlying distribution is uniform over $[0, 1]$ and $T^{h\ell} = [0, \tau) \cup [\underline{t}, \overline{t})$.

satisfies

$$q - (1-q)c \geq 0 \iff q \geq \widehat{q} := \frac{c}{1+c}. \tag{1}$$

Next, suppose, after $m = h\ell$, the supervisor will approve the bank's high risks if and only if $t \geq \underline{t}$. This approval strategy, as explained above, is indeed optimal if the bank follows the supervisor's guidance. However, it injects uncertainty into the bank's problem and makes it crucial for the bank to assess the likelihood that the supervisor's type is in $[\underline{t}, \overline{t})$, relative to $[0, \tau)$. Crucially, the bank assigns a relatively higher probability to $[\underline{t}, \overline{t})$ when its signal is good than when its signal is bad (see Figure 2). This is due to the fact that the supervisor's information and the bank's information are correlated: when the bank's signal is good, it is more likely that the state is good and so the supervisor will be of a higher type. The bank is willing to follow the supervisor's guidance $h\ell$ when it assigns a higher probability than $\widehat{q}$ to $[\underline{t}, \overline{t})$ when its signal is good and a lower probability than $\widehat{q}$ when its signal is bad.

In principle, the set $T^{h\ell}$ of the supervisor's types that are associated with message $m = h\ell$ does not have to take the form described above; for the supervisor to obtain her first-best outcome, $T^{h\ell}$ should include $[\underline{t}, \overline{t})$, but it can take an arbitrary structure below $\underline{t}$. The following result, however, shows that restricting attention to the case where $T^{h\ell} = [0, \tau) \cup [\underline{t}, \overline{t})$ incurs no loss of generality.

Lemma 1 *If there exists $T^{h\ell} \subset [0, \overline{t})$ that enables the supervisor to obtain her first-best outcome, then there also exists $\mathcal{T}^{h\ell} = [0, \tau) \cup [\underline{t}, \overline{t})$ that allows the supervisor to do the same.*

By pooling her types that would approve high risk-taking (i.e., $[\underline{t}, \overline{t})$) with those that would disapprove (i.e., $T^{h\ell} \cap [0, \underline{t})$) in sending message $m = h\ell$, the supervisor induces the bank to act on its own private information, discouraging (encouraging) the bank to take high risks if its signal was bad (good). The main thrust of Lemma 1 is to show that the most efficient way to maintain the bank's incentives for different realizations of its signal is pooling $[\underline{t}, \bar{t})$ with the extreme types of the supervisor that would disapprove, i.e., those that are the most sure of the bad state of the economy, so $T^{h\ell} = [0, \tau) \cup [\underline{t}, \bar{t})$ for some $\tau \in [0, \underline{t}]$.

Suppose $T^{h\ell} = [0, \tau) \cup [\underline{t}, \bar{t})$ and the bank observes $m = h\ell$ and $s = g$. By Bayes' rule, the bank assigns the following probability to the event that $t \in [\underline{t}, \bar{t})$ and so the supervisor will approve its high risk-taking:

$$\Pr\{t \in [\underline{t}, \bar{t}) | m = h\ell, s = g\} = \frac{\Pr\{t \in [\underline{t}, \bar{t}), s = g\}}{\Pr\{t \in T^{h\ell}, s = g\}} = \frac{\int_{\underline{t}}^{\bar{t}} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t)}{\int_{T^{h\ell}} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t)}.$$

The bank is willing to take high risks if and only if this probability exceeds $\widehat{q} = c/(1+c)$, which reduces to

$$\int_{\underline{t}}^{\bar{t}} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t) \geq c \int_0^{\tau} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t). \tag{2}$$

Similarly, it is easy to show that the bank observing $s = b$ (instead of $s = g$) is not willing to take high risks if and only if

$$\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \leq c \int_0^{\tau} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t). \tag{3}$$

Clearly, (2) requires τ to be sufficiently small, while (3) holds only when τ is sufficiently large.

The following result establishes a necessary and sufficient condition for both (2) and (3) to hold.

Proposition 1 *The supervisor can obtain her first-best outcome if and only if* (3) *holds when* $\tau = \underline{t}$*, that is,*

$$\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \leq c \int_0^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t). \tag{4}$$

Proof. Let τ be the value that satisfies (3) with equality. Clearly, $\tau \leq \underline{t}$ if and only if (4) holds. If (3) holds with equality, however, (2) holds automatically, because $\gamma > 1/2$, so $(1-\gamma)t + \gamma(1-t)$ is strictly decreasing in t, while $\gamma t + (1-\gamma)(1-\gamma)$ is strictly decreasing in t (see Lemma 5 in

Appendix A for a complete proof). ■

Condition (4) necessarily holds if γ is sufficiently close to $\frac{1}{2}$ (i.e., the bank's signal is marginally informative), because both $\underline{t}$ and $\overline{t}$ would be close to $\widehat{t}$. On the contrary, it fails if γ is sufficiently close to 1 (i.e., the bank's information is highly informative), because $\underline{t}$ would be close to 0, while $\overline{t}$ close to 1. Given this, it is plausible that there exists a cutoff value of γ such that (4) holds if and only if γ is below the cutoff. Such cutoff property indeed holds with most "regular" distributions, but it does not hold in general.[23] This is because the integrand $(1-\gamma)t+\gamma(1-t)$ also depends on γ and may interfere with the (clear) effect through changes in $\underline{t}$ and $\overline{t}$.

5 Optimal Communication

In this section, we study the general problem of the supervisor. In particular, we provide a formal characterization of the supervisor's optimal communication strategy when she cannot achieve her first-best outcome (i.e., (4) fails to hold).

5.1 Supervisor's Problem

We begin by making the following useful observation.

Lemma 2 *For any given communication strategy (M,π), there exists a ternary signal (M',π'), with $M'=\{hh,h\ell,\ell\ell\}$, that induces the same outcome as (M,π).*

This result follows from the following two facts. First, given the supervisor's optimal behavior for supervisory ruling (to approve high risk-taking or not), the bank is always more likely to take high risks when its signal is good ($s=g$) than when its signal is bad ($s=b$). This implies that the supervisor can never induce the bank to take high risks only when its signal is bad ($s=b$), so there are at most three different outcomes after the bank receives a message from the supervisor. Second, multiple messages that trigger the same bank action can be combined into one, not affecting the bank's incentives.

Lemma 2 implies that the supervisor's optimal communication problem reduces to dividing her type space (i.e., $[0,1]$) into three subsets $\{T^m\}_{m\in\{h\ell,hh,\ell\ell\}}$ such that each T^m is the set of her

[23] For example, the distribution with density $f(t)=\frac{cos(5t)^2}{\int_0^1 cos(5t)^2dt}$ fails the cutoff property when $c=0.5$ and $d=2$.

types for which the supervisor sends message m. At the bank examination stage, the supervisor will approve the bank's high risk-taking if and only if (i) her type exceeds $\bar{t}$, (ii) her type exceeds $\widehat{t}$ and she does not know the bank's signal, or (iii) her type exceeds $\underline{t}$ and the bank's signal is $s = g$. This implies that given a partition $(T^{hh}, T^{h\ell}, T^{\ell\ell})$—and assuming that the bank follows the supervisor's guidance—the supervisor will approve the bank's high risk-taking either when $t \in T^{h\ell} \cap [\underline{t}, 1]$ (since, following $m = h\ell$, the bank would take high risks only if $s = g$) or when $t \in T^{hh} \cap [\widehat{t}, 1]$. This implies that the supervisor's expected payoff is given by

$$\int_{T^{h\ell}\cap[\underline{t},1]} [\gamma t - (1-\gamma)(1-t)d] \,\mathrm{d}F(t) + \int_{T^{hh}\cap[\widehat{t},1]} [t - (1-t)d] \,\mathrm{d}F(t). \tag{5}$$

Of course, the supervisor's communication strategy should ensure that the bank is willing to follow her guidance. In other words, for each message $m = \{hh, h\ell, \ell\ell\}$ and its own signal $s \in \{g, b\}$, the bank should have an incentive to follow the supervisor's recommendation (h or ℓ). Formally, the supervisor's strategy should satisfy the following six constraints:

$$\int_{T^{h\ell}\cap[\underline{t},1]} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t) \geq c \int_{T^{h\ell}\cap[0,\underline{t})} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t), \tag{$IC_g^{h\ell}$}$$

$$\int_{T^{h\ell}\cap[\underline{t},1]} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t) \leq c \int_{T^{h\ell}\cap[0,\underline{t})} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t), \tag{$IC_b^{h\ell}$}$$

$$\int_{T^{hh}\cap[\widehat{t},1]} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t) \geq c \int_{T^{hh}\cap[0,\widehat{t})} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t), \tag{IC_g^{hh}}$$

$$\int_{T^{hh}\cap[\widehat{t},1]} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t) \geq c \int_{T^{hh}\cap[0,\widehat{t})} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t), \tag{IC_b^{hh}}$$

$$\int_{T^{\ell\ell}\cap[\widehat{t},1]} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t) \leq c \int_{T^{\ell\ell}\cap[0,\widehat{t})} [\gamma t + (1-\gamma)(1-t)] \,\mathrm{d}F(t), \tag{$IC_g^{\ell\ell}$}$$

$$\int_{T^{\ell\ell}\cap[\widehat{t},1]} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t) \leq c \int_{T^{\ell\ell}\cap[0,\widehat{t})} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t). \tag{$IC_b^{\ell\ell}$}$$

To understand these IC constraints, suppose, for example, that the bank has received signal $s = b$ and message $m = h\ell$. The bank will follow the guidance and thus take low risks only if it assigns at most probability $\widehat{q}$ (defined in (1)) to the event that the supervisor will approve high risk-taking. Since, following $m = h\ell$, only the supervisor of type $t \geq \underline{t}$ approves high risk-taking, this condition is equivalent to

$$\Pr(t \geq \underline{t} | s = b, m = h\ell) = \frac{\int_{T^{h\ell} \cap [\underline{t},1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t)}{\int_{T^{h\ell}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t)} \leq \widehat{q} = \frac{c}{1+c}.$$

Rearranging the terms, we arrive at ($IC_b^{h\ell}$). One can analogously derive the other IC constraints.

The supervisor's problem is to choose $\{T^{hh}, T^{h\ell}, T^{\ell\ell}\}$ so as to maximize (5) subject to the above six constraints. Fortunately, the structure of our problem allows us to dramatically reduce the complexity of the problem. In particular, the following two lemmas establish that the only relevant constraint is ($IC_b^{h\ell}$).

Lemma 3 *(a)* (IC_b^{hh}) *implies* (IC_g^{hh}), *and* ($IC_g^{\ell\ell}$) *implies* ($IC_b^{\ell\ell}$).

(b) The supervisor's optimal communication strategy can always be modified so that $T^{hh} \subseteq [\widehat{t}, 1]$ *and* $T^{\ell\ell} \subseteq [0, \widehat{t})$, *for which* ($IC_b^{hh}$) *and* ($IC_g^{\ell\ell}$) *hold trivially.*

The first part in Lemma 3 holds simply because the bank is more willing to take high risks when its signal is $s = g$ than when its signal is $s = b$: if the bank would be willing to take high (low) risks even when its signal is b (g), it will certainly take high (low) risks when its signal is g (b). For the second part, note that if $m = hh$ or $m = \ell\ell$, the bank's action (risk level) does not provide any new information about the state ω. In that case, it is optimal for the supervisor to recommend the bank to take each action only if she prefers that action a priori, that is, based solely on her private information.

Lemma 4 *Suppose that* $T^{h\ell} \subset [0, 1]$ *satisfies the constraint* ($IC_b^{h\ell}$). *Then there exists* $\tau \in [0, \underline{t}]$ *such that* $\mathcal{T}^{h\ell} := [0, \tau) \cup (T^{h\ell} \setminus [0, \underline{t}))$ *satisfies both* ($IC_g^{h\ell}$) *and* ($IC_b^{h\ell}$).

This lemma is analogous to Lemma 1 in Section 4. The supervisor needs to introduce some strategic ambiguity into her communication to prevent the bank from deviating from the supervisory guidance. In particular, the supervisor should send $m = h\ell$ sometimes when $t < \underline{t}$, so that she may disapprove of the bank's high risk-taking after having sent $m = h\ell$. As in Section 4, the most efficient way to do so is pooling $T^{h\ell}_{\text{Appr.}} := T^{h\ell} \cap [\underline{t}, 1] \neq \emptyset$ with the extreme types of the supervisor that would disapprove, i.e., those that are the most sure of the bad state of the economy, so $T^{h\ell}_{\text{Disappr.}} := T^{h\ell} \cap [0, \underline{t}) = [0, \tau)$ for some $\tau \in [0, \underline{t}]$.

5.2 Solving the Problem

The above lemmas establish that the only potentially binding constraint is $(IC_b^{h\ell})$, so it suffices to consider the following simpler problem:

$$\max_{T^{h\ell}_{\text{Appr.}}\subset[\underline{t},1],T^{hh}\subset[\widehat{t},1],\tau\in[0,\underline{t}]} \int_{T^{h\ell}_{\text{Appr.}}} [\gamma t-(1-\gamma)(1-t)d]\,\mathrm{d}F(t)+\int_{T^{hh}}[t-(1-t)d]\,\mathrm{d}F(t) \tag{6}$$

subject to $T^{h\ell}_{\text{Appr.}}\cap T^{hh}=\emptyset$ and

$$\int_{T^{h\ell}_{\text{Appr.}}}[(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t)\le c\int_0^{\tau}[(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t). \tag{7}$$

Notice that the supervisor's first-best outcome in Section 4 is the *unconstrained* optimal solution of the problem (6) (i.e., when the constraint (7) is not binding), and condition (4) in Proposition 1 can be interpreted as the case when the supervisor's first-best outcome satisfies the constraint (7).

From now on, we focus on the case where the supervisor's first-best outcome is *not* feasible, i.e., condition (4) fails to hold, so the constraint (7) is binding. We begin by observing that it is optimal to set $\tau=\underline{t}$. In the current case, the supervisor would like to, but cannot, send the message $m=h\ell$ for every $t\in[\underline{t},\overline{t}]$ without violating the constraint. Increasing $\tau(<\underline{t})$ relaxes (7) by raising its right-hand side. However, it does not directly affect the objective (6). Therefore, this change strictly raises the supervisor's payoff by allowing her to send the message $m=h\ell$ for a wider range of $t\in[\underline{t},\overline{t})$ without violating the constraint.

Formally, the supervisor's problem is a linear programming problem and so can be analyzed using the standard Lagrangian method. The corresponding Lagrangian is given by

$$\begin{aligned}\mathcal{L}=&\int_{T^{h\ell}_{\text{Appr.}}}[\gamma t-(1-\gamma)(1-t)d]\,\mathrm{d}F(t)+\int_{T^{hh}}[t-(1-t)d]\,\mathrm{d}F(t)\\&+\lambda\left\{c\int_0^{\underline{t}}[(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t)-\int_{T^{h\ell}_{\text{Appr.}}}[(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \int_{T^{h\ell}_{\text{Appr.}}} \{\gamma t - (1-\gamma)(1-t)d - \lambda[(1-\gamma)t + \gamma(1-t)]\} \,\mathrm{d}F(t) \\
&\quad + \int_{T^{hh}} [t - (1-t)d] \,\mathrm{d}F(t) \\
&\quad + \lambda c \int_0^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t).
\end{aligned}
$$

Note that the last term affects $\mathcal{L}$ only through λ and so can be regarded as a constant.

Let $\underline{t}^*(\lambda)$ be the unique value of t that makes the first integrand equal to zero, that is,

$$
\gamma t - (1-\gamma)(1-t)d - \lambda[(1-\gamma)t + \gamma(1-t)] = 0 \iff \underline{t}^*(\lambda) := \frac{(1-\gamma)d + \lambda\gamma}{\gamma + (1-\gamma)d + \lambda(2\gamma - 1)}.
$$

Economically, $\underline{t}^*(\lambda)$ is the supervisor's type that is indifferent between conditionally approving and unconditionally disapproving of high risk investments, after accounting for the shadow cost of inducing the bank to reveal its signal. It is straightforward that $\underline{t}^*(\lambda)$ is equal to $\underline{t}$ when $\lambda = 0$. In addition, if $\overline{\lambda}$ is the value of λ such that $\underline{t}^*(\lambda) = \widehat{t}$, then $\underline{t}^*(\lambda)$ continuously and strictly increases as λ rises from 0 to $\overline{\lambda}$.

Similarly, let $\overline{t}^*(\lambda)$ be the unique value of t that makes the first two integrands equal, that is,

$$
\gamma t - (1-\gamma)(1-t)d - \lambda[(1-\gamma)t + \gamma(1-t)] = t - (1-t)d \iff \overline{t}^*(\lambda) := \frac{\gamma d - \lambda\gamma}{1 - \gamma + \gamma d - \lambda(2\gamma - 1)}.
$$

This is the supervisor's type that is indifferent between conditionally approving and unconditionally approving high risk investments, after accounting for the shadow cost of doing the former. As is intuitive, $\overline{t}^*(\lambda)$ continuously and strictly decreases from $\overline{t}$ to $\widehat{t}$ as λ rises from 0 to $\overline{\lambda}$.

By construction, $[\underline{t}^*(\lambda), \overline{t}^*(\lambda))$ is the optimal subinterval of $[\underline{t}, \overline{t})$ to send the message $m = h\ell$ given the marginal shadow cost λ.[24] Therefore, the optimal value of $\lambda \in [0, \overline{\lambda}]$, which we denote by λ^*, can be found from the fact that the constraint should be binding, that is,

$$
\int_{\underline{t}^*(\lambda^*)}^{\overline{t}^*(\lambda^*)} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t) = c \int_0^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \,\mathrm{d}F(t). \tag{8}
$$

[24] Note that $t \in [\underline{t}^*(\lambda), \overline{t}^*(\lambda))$ if and only if

$$
\gamma t - (1-\gamma)(1-t)d - \lambda[(1-\gamma)t + \gamma(1-t)] \geq \max\{0, t - (1-t)d\}.
$$

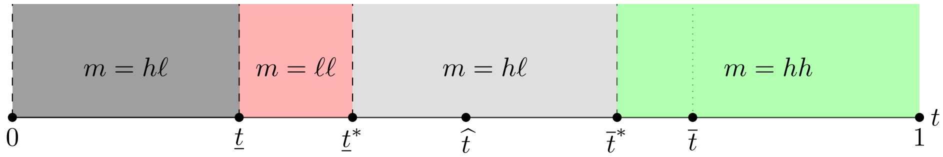


Figure 3: The structure of optimal communication strategy of the supervisor when she cannot achieve her first-best outcome.

The optimal value λ^* is well defined in $[0, \overline{\lambda}]$ since the left-hand side is larger (smaller) than the right-hand side if $\lambda = 0$ ($\lambda = \overline{\lambda}$) and continuously decreasing in λ.

The following proposition summarizes the preceding discussion, characterizing the supervisor's optimal communication strategy. This proposition, together with Proposition 1, constitutes our first main result.

Proposition 2 *Suppose that* (4) *fails to hold. Then the IC constraint* (7) *is binding, and the constrained optimal solution to* (6) *involves* $T^{h\ell}_{Appr.} = [\underline{t}^*(\lambda^*), \overline{t}^*(\lambda^*))$, $T^{hh} = [\overline{t}^*(\lambda^*), 1]$, *where* $\underline{t}^*(\lambda) \in [\underline{t}, \widehat{t}]$, $\overline{t}^*(\lambda) \in [\widehat{t}, \overline{t}]$, *and* $\lambda^* \in [0, \overline{\lambda}]$ *is the solution to* (8) *(so* $T^{h\ell} = [0, \underline{t}) \cup [\underline{t}^*(\lambda^*), \overline{t}^*(\lambda^*))$*).*

Figure 3 illustrates the resulting structure of the supervisor's optimal communication strategy when (4) fails. It has a structure similar to that of the communication strategy in Section 4 that enables the supervisor to achieve her first-best outcome. In particular, it again introduces strategic ambiguity into the message $h\ell$ by pooling the interval $[\underline{t}^*, \overline{t}^*)$ (which will ultimately approve high risk-taking) with $[0, \underline{t})$ (which will ultimately disapprove of high risk-taking). The difference is that, with the bank's incentive constraint following $m = h\ell$ now binding, the supervisor chooses to send $m = h\ell$ on a strict subset of $[\underline{t}, \overline{t}]$ and so fails to obtain her first-best outcome when $t \in (\underline{t}, \underline{t}^*) \cup (\overline{t}^*, \overline{t})$.

6 The Role of Bank's Information and Disapproval Cost

Using the characterization results so far, this section studies how the supervisor's indirect payoff—her expected payoff under the optimal communication strategy—depends on two key parameters of the model: γ (the informativeness of the bank's signal) and c (the cost of supervisory disapproval for the bank).

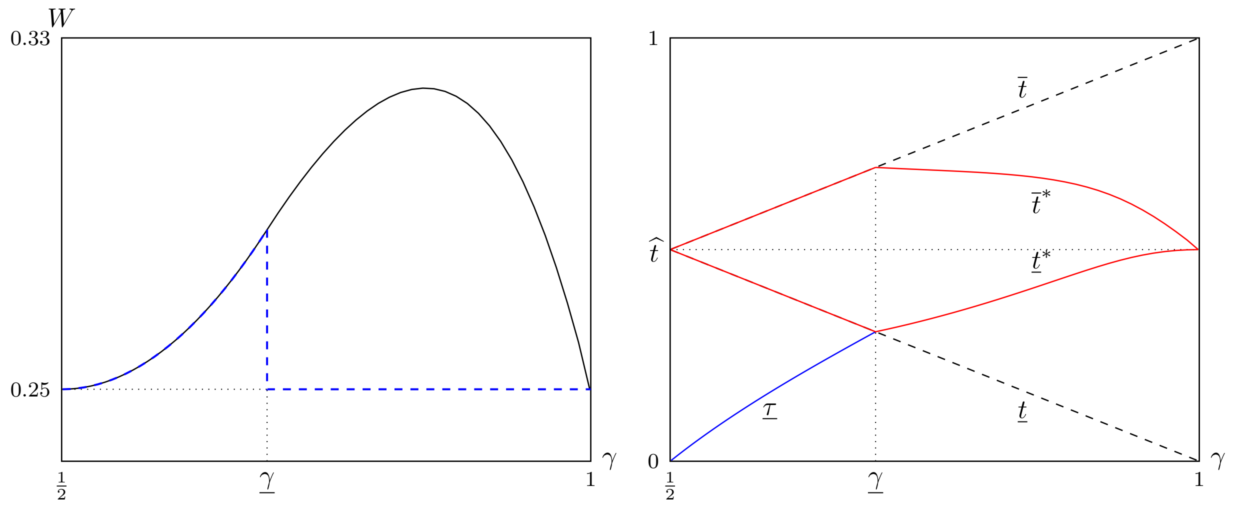


Figure 4: The supervisor's expected payoff (left) and the relevant cutoffs (right) as functions of γ. In the left panel, the blue dashed curve shows the supervisor's highest expected payoff with cheap talk. In both panels, $F = U[0,1]$ and $c = d = 1$.

6.1 Informativeness of the Bank's Signal

The left panel of Figure 4 shows how the supervisor's indirect payoff depends on γ when the supervisor's type is uniformly distributed over $[0,1]$. As γ rises, it increases until around $\gamma = 0.844$ and decreases afterward. Although the single-peakedness (quasi-concavity) relies on the nice regularity of the uniform distribution, the following result holds in general.

Proposition 3 *The supervisor's indirect payoff is minimized in the limit as γ approaches $\frac{1}{2}$ or 1.*

Importantly, Proposition 3 implies that the supervisor's indirect payoff is generally non-monotone, as it is necessarily increasing when γ is close to $\frac{1}{2}$, whereas it is decreasing when γ is close to 1. To understand this non-monotonicity and Proposition 3, let $\underline{t}^*(\gamma)$ and $\overline{t}^*(\gamma)$ denote the values such that $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$ coincides with the approving supervisor types in the optimal $T^{h\ell}$ for each γ; see the red solid curves in the right panel of Figure 4. Then, the supervisor's indirect payoff in terms of γ is given by

$$W(\gamma) := \int_{\underline{t}^*(\gamma)}^{\overline{t}^*(\gamma)} [\gamma t - (1-\gamma)(1-t)d] \,\mathrm{d}F(t) + \int_{\overline{t}^*(\gamma)}^{1} [t - (1-t)d] \,\mathrm{d}F(t).$$

On the one hand, whenever the supervisor can obtain her first-best outcome, $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$ coincides with $[\underline{t}, \overline{t})$; observe the coincidence of the red solid curves with the black dashed curves for $\gamma \leq \underline{\gamma}$ in the right panel of Figure 4. In this case, an increase in γ is unambiguously good news since it not only makes the bank's signal (action) more informative but also expands $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$—the region where the supervisor has access to and uses the bank's signal (in addition to her own signal). In other words, the bank reveals better information more frequently, enabling the supervisor to take better-informed supervisory actions. Together with the fact that (4) holds for γ sufficiently close to $\frac{1}{2}$, this explains why $W(\gamma)$ increases in γ around $\frac{1}{2}$.

On the other hand, when the supervisor cannot obtain her first-best outcome, $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$ is a proper subinterval of $[\underline{t}, \overline{t})$; observe that the red solid curves are strictly trapped between the black dashed curves for $\gamma > \underline{\gamma}$ in the right panel of Figure 4. In this case, the welfare effect of an increase in γ is more complicated. It still makes the bank's signal directly more informative. However, it contracts the interval $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$. In other words, the supervisor can utilize the bank's better information *less* frequently. To see this latter effect more clearly, recall that $\underline{t}$ is monotone decreasing in γ. This leads the bank to assign a lower probability to the event that the supervisor will disapprove of its high risk-taking. When its perceived probability is sufficiently low, the bank has strong incentives to take high risks regardless of its signal. To restrain the bank from doing so, the supervisor is forced to shrink $[\underline{t}^*(\gamma), \overline{t}^*(\gamma))$.

The tension between the positive direct effect of raising γ and its negative indirect effect renders $W(\gamma)$ non-monotone in general. When γ is sufficiently close to 1, however, the latter effect necessarily dominates, so $W(\gamma)$ falls: as γ approaches 1, both $\underline{t}^*(\gamma)$ and $\overline{t}^*(\gamma)$ converge to $\widehat{t}$, which means that the supervisor can rely only on her own information. Note that this is paradoxical since this is when the bank's signal is almost perfect and so can be particularly valuable. Intuitively, if the bank's signal is arbitrarily precise, the supervisor wants to utilize it as much as possible for supervisory ruling. But then, the bank faces an irresistible temptation to take high risks even with $s = b$, making it impossible for the supervisor to elicit *any* information from the bank.

6.2 Bank's Cost in Case of Supervisory Disapproval

The mechanism by which the supervisor elicits the bank's information relies on the fact that the bank incurs a cost c if its high risk investments get disapproved by the supervisor: after receiving $m = h\ell$ and $s = b$, the bank chooses to take low risks in order to avoid disapproval costs. Indeed, if $c = 0$, it is impossible to induce the bank to take high risks only when $s = g$. This suggests that the supervisor's guidance would be more effective, the higher c is. The following result confirms that it is indeed the case.

Proposition 4 *Let c^* be the value of c such that* (4) *holds with equality. Then, the supervisor's indirect payoff strictly increases in c whenever $c < c^*$ and stays constant if $c \geq c^*$.*

Intuitively, the bank worries not only about how frequently the supervisor will disapprove of its high risk-taking, but also about how costly those supervisory disapprovals would be. This implies that raising c can offset the adverse equilibrium effect of increased γ. This insight of Proposition 4 is particularly important because, in reality, the supervisor can be agile in her communication, but she cannot freely adjust the bank's cost in case of supervisory disapproval. In other words, the supervisor can easily adapt her message to the arrival of new information, but can change disapproval costs for the bank only occasionally by passing legislation to promote financial stability.[25] Proposition 4 suggests that, given the (infrequent) opportunity to adjust the bank's disapproval cost, it is better to err on the side of giving the supervisor too much power *in case she finds that the bank does not meet supervisory expectations*, i.e., err on the side of setting c too high. If c is too high (i.e., $c > c^*$), the supervisor could simply scale back how frequently she will disapprove of high risk-taking after having sent $m = h\ell$—$T^{h\ell}_{\text{Disappr.}} = [0, \underline{\tau}(c))$, where $\underline{\tau}(c) \in [0, \underline{t}]$ is such that (3) holds with equality. If c is too low (i.e., $c < c^*$), not only is the supervisor's unconstrained optimum infeasible (leaving welfare on the table), but the economy is exposed to experiencing a welfare loss in case the bank experiences a sudden boost in its private information.

7 Optimal Cheap-Talk Communication

Thus far, we have explored the full potential of supervisory guidance by allowing our supervisor to adopt any communication strategy. In particular, the supervisor has commitment power over her communication strategy. To understand the role of this commitment power, this section considers the case where the supervisor can only communicate via cheap talk à la Crawford and Sobel (1982). Specifically, we assume that the supervisor first observes her type t and then decides which

[25] For example, the Dodd-Frank Act made all banks with assets above $50 billion subject to a much more aggressive supervisory regime, effectively raising c for mid-sized banks; in 2018, Congress scaled back Dodd-Frank, raising the threshold for increased scrutiny of banks from $50 billion to $250 billion, effectively reducing c for mid-sized banks.

message m to send (rather than deciding which message to send before observing her type). All other aspects of the model remain the same.

With cheap talk, we can focus on the bank's (conditional) actions that can be induced; if there exists a cheap-talk message m that triggers a particular action by the bank, the supervisor should be able to induce the same action (by sending the message m) irrespective of her type. Let hh denote the bank's action of taking high risks regardless of its own signal and $h\ell$ the bank's action of taking high risks only with $s = g$.[26] Then, there are the following four possibilities: (i) both hh and $h\ell$ can be induced; (ii) only hh can be induced; (iii) only $h\ell$ can be induced; or (iv) neither hh nor $h\ell$ can be induced.

As argued in Section 4, the supervisor wants the bank to take high risks regardless of its own signal if $t \geq \bar{t}$ and only when $s = g$ if $t \in [\underline{t}, \bar{t})$. Hence, in case (i), the supervisor will choose to induce hh if $t \geq \bar{t}$ and $h\ell$ if $t \in [\underline{t}, \bar{t})$, thereby obtaining her first-best outcome. If this outcome can be supported as a cheap-talk equilibrium, it is certainly necessary that it can be achieved in the baseline model (which endows the supervisor with more power). Conversely, if the supervisor can achieve her first-best outcome with a particular communication strategy, she certainly has no incentive to deviate from it, so cheap talk is sufficient. Therefore, case (i) arises—and the supervisor obtains her first-best outcome with cheap talk—if and only if (4) holds.

In case (ii), the supervisor will induce hh whenever $t \geq \widehat{t}$. It is easy to see that this can always be supported as a (weak perfect Bayesian) equilibrium: it suffices to assume that the supervisor will send $m = hh$ if $t \geq \widehat{t}$ and $m = \ell\ell$ otherwise, and that the supervisor does not update her belief, thereby disapproving, if the bank deviates and takes high risks after receiving $\ell\ell$.

In case (iii), the supervisor will induce $h\ell$ whenever $t \geq \underline{t}$. Unlike in case (ii), this outcome can only be supported if the bank's incentive constraints are satisfied, ensuring that the bank should be willing to take high risks if $s = g$ and low risks if $s = b$. By the same logic used in Sections 4 and 5, such an equilibrium exists if and only if the following inequality—ensuring that the bank with $s = b$ has the right incentives—holds:

$$\int_{\underline{t}}^{1} \left[(1-\gamma)t + \gamma(1-t)\right] \, \mathrm{d}F(t) \leq c \int_{0}^{\underline{t}} \left[(1-\gamma)t + \gamma(1-t)\right] \, \mathrm{d}F(t).$$

Notice that this inequality is similar to (4), except that the integral on its left-hand side is over a

[26] It is not crucial whether the supervisor can induce the bank to take low risks regardless of its signal because she can obtain the same outcome by simply disapproving the bank's high risk-taking.

larger interval ($[\underline{t}, 1]$) than that of (4) ($[\underline{t}, \overline{t}]$). This immediately implies that this inequality is more stringent than (4). Therefore, case (iii) can arise only if (4) holds, so case (i) exists as well. But then, case (iii) cannot yield a higher payoff for the supervisor than case (i), which enables her to obtain her first-best outcome with cheap talk.

Case (iv), whether it exists or not, cannot yield a higher payoff for the supervisor than case (ii). Combining this observation with the discussion above leads to the following result.

Proposition 5 *If the supervisor can obtain her first-best outcome in our model (i.e.,* (4) *holds), then she can do so with cheap talk. Otherwise, the best that she can do with cheap talk is to implement (and approve) high risk-taking by the bank if and only if she is of type $t \geq \widehat{t}$.*

Intuitively, the fact that the supervisor is able to obtain her first-best outcome means that she obtains her optimal outcome for each supervisor type, so she simply has no incentive to deviate from the communication strategy. It is when the supervisor cannot obtain her first-best outcome that she needs to use her commitment power: by keeping some supervisor types that would like to ask the bank to act on its information from doing so, she enables some other types to be able to do so in an incentive-compatible manner. Commitment power is valuable exactly when the supervisor needs to sacrifice some supervisor types for higher ex ante welfare.

The left panel of Figure 4 shows how the supervisor's indirect payoff under cheap talk (blue dashed) compares to that in our baseline model (black solid). They coincide until $\gamma = \underline{\gamma}$, so the supervisor achieves her first-best outcome either way if and only if $\gamma \leq \underline{\gamma}$. For $\gamma > \underline{\gamma}$, the supervisor's indirect payoff under cheap talk stays constant at a level strictly below that in our baseline model; in this case, by Proposition 5, the supervisor can utilize only her own information, so her payoff is independent of γ.

It is noteworthy that the supervisor's indirect payoff in our baseline model converges to her maximal payoff with cheap talk as γ approaches 1. As discussed in Section 6.1, this is because the interval $[\underline{t}^*, \overline{t}^*]$—on which the supervisor has access to and uses the bank's information—collapses to $\{\widehat{t}\}$ as γ approaches 1. This suggests that superior information on the bank's side erodes away the welfare benefits of the supervisor's commitment power; in the limit as the bank's information approaches perfect, the welfare benefits of commitment on the supervisor's side are completely neutralized.

8 Commitment Over Supervisory Ruling

Our baseline model does not give the supervisor commitment power over her follow-up supervisory ruling. In particular, the supervisor approves high risk-taking only when it is ex post optimal; she will allow the bank's high risk-taking if and only if she is of type $t \geq \underline{t}$ ($t \geq \widehat{t}$) after having sent $m = h\ell$ ($m = hh$). This section examines the extent to which this lack of commitment hinders the effectiveness of supervision. Specifically, we consider the case where the supervisor can commit a priori to approving (or disapproving) high risk-taking even if it is ex post inefficient and compare the outcome with commitment over supervisory ruling to our baseline outcome.

Following similar steps as in the baseline model, it can be shown that we can still focus on the ternary message space, and the corresponding sets of supervisor types $\{T^m\}_{m\in\{h\ell,hh,\ell\ell\}}$ have the same structures and properties as those in the baseline case. The only crucial difference is that the supervisor can commit to disapproving even when $t > \underline{t}$ after having sent $m = h\ell$.[27] Intuitively, although it is not ex post optimal for the supervisor to disapprove when $t \in (\underline{t}, \tau)$ and $s = g$, doing so reduces the likelihood of approval from the bank's perspective, thereby inducing the bank to be more prudent conditional on $s = b$.

Formally, the supervisor's problem can be written as

$$\max_{T^{h\ell}_{\text{Appr.}},T^{h\ell}_{\text{Disappr.}},T^{hh}\subset[\widehat{t},1]} \int_{T^{h\ell}_{\text{Appr.}}} [\gamma t - (1-\gamma)(1-t)d]\,\mathrm{d}F(t) + \int_{T^{hh}} [t - (1-t)d]\,\mathrm{d}F(t)$$

subject to $T^{h\ell}_{\text{Appr.}} \cap T^{h\ell}_{\text{Disappr.}} = (T^{h\ell}_{\text{Appr.}} \cup T^{h\ell}_{\text{Disappr.}}) \cap T^{hh} = \emptyset$ and

$$\int_{T^{h\ell}_{\text{Appr.}}} [(1-\gamma)t + \gamma(1-t)]\,\mathrm{d}F(t) \leq c \int_{T^{h\ell}_{\text{Disappr.}}} [(1-\gamma)t + \gamma(1-t)]\,\mathrm{d}F(t) \tag{9}$$

Clearly, if the supervisor's unconstrained optimum is feasible in the baseline model, it is also feasible with this additional commitment power. Therefore, it suffices to consider the case where (4) fails to hold. Applying similar arguments to those in Section 5.2, we obtain the following result.

Proposition 6 *Suppose* (4) *fails and the supervisor can commit to her follow-up supervisory rul-*

[27] It is also possible that the supervisor chooses to approve when $t < \underline{t}$. But, the possibility does not materialize at the optimal solution because, as shown above, the relevant incentive constraint regards the bank's incentive to take high risks even with $s = b$. In other words, the supervisor has no incentive to further encourage the bank's risk-taking by approving more frequently.

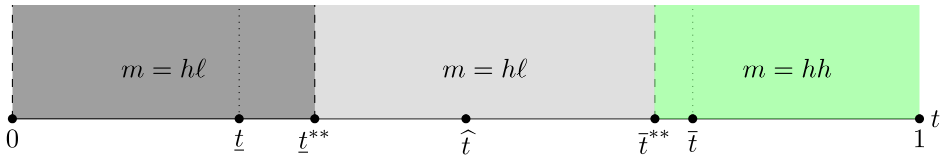


Figure 5: The structure of the supervisor's optimal communication strategy when she can commit to her follow-up supervisory ruling. The supervisor commits to approving only when $t \geq \underline{t}^{**}$.

ing. Then, the supervisor's optimal strategy involves $T^{h\ell} = [0, \overline{t}^{**})$ and $T^{hh} = [\overline{t}^{**}, 1]$ (i.e., reveal whether $t \geq \overline{t}^{**}$ or not), and $T^{h\ell}_{\text{Appr.}} = [\underline{t}^{**}, \overline{t}^{**})$ (i.e., commit to approval if $t \geq \underline{t}^{**}$) for some $\underline{t}^{**} \in (\underline{t}, \widehat{t})$ and $\overline{t}^{**} \in (\widehat{t}, \overline{t})$.

Figure 5 visualizes the structure of the supervisor's optimal strategy in Proposition 6. The most notable difference from the baseline case without commitment over supervisory ruling is the fact that $T^{\ell\ell}$ is empty. Recall that, as illustrated in Figure 3, $T^{\ell\ell} = [\underline{t}, \underline{t}^*) \neq \emptyset$ for some $\underline{t}^* \in (\underline{t}, \widehat{t})$ whenever (4) fails in the baseline case. Intuitively, the supervisor can now annex this region to $T^{h\ell}$ with the additional commitment power vested in her and set an approval cutoff $\underline{t}^{**} \in (\underline{t}, \underline{t}^*)$ (above $\underline{t}$). This is beneficial to the supervisor because it allows her to be more effective at discouraging the bank's high risk-taking and so induces the bank to take high risks also when $t \in [\underline{t}^{**}, \underline{t}^*)$ and $s = g$ (relative to our baseline outcome). Without commitment over supervisory ruling, this adjustment is not feasible because the supervisor would be tempted to approve the bank's high risk-taking when $t \in [\underline{t}, \underline{t}^{**})$ (assuming it is acting on its signal), which would disrupt the bank's incentives.

Figure 6 depicts how the two cutoffs in Proposition 6, $\underline{t}^{**}$ and $\overline{t}^{**}$, change according to γ and how they compare with the corresponding cutoffs in the baseline model. The key difference from the baseline case is that the set $T^{h\ell}_{\text{Appr.}} = [\underline{t}^{**}, \overline{t}^{**})$ does not shrink and vanish as γ approaches 1. Without commitment over supervisory ruling, the supervisor can use only the interval $[0, \underline{t})$ to reduce the likelihood of approval from the bank's perspective. However, as γ approaches 1, $[0, \underline{t})$ vanishes away, which in turn implies that $T^{h\ell}_{\text{Appr.}} = [\underline{t}^*, \overline{t}^*)$ in the baseline case also vanishes away. Commitment over supervisory ruling helps overcome this problem by enabling the supervisor to use a larger interval $[0, \underline{t}^{**})$ to provide incentives for the bank.

Notice that commitment over supervisory ruling is still not enough for the supervisor to obtain her first-best outcome when (4) fails: for example, the supervisor would ideally want to approve high risk-taking by the bank also when $t \in [\underline{t}, \underline{t}^{**})$ and $s = g$, but she commits to disapproving,

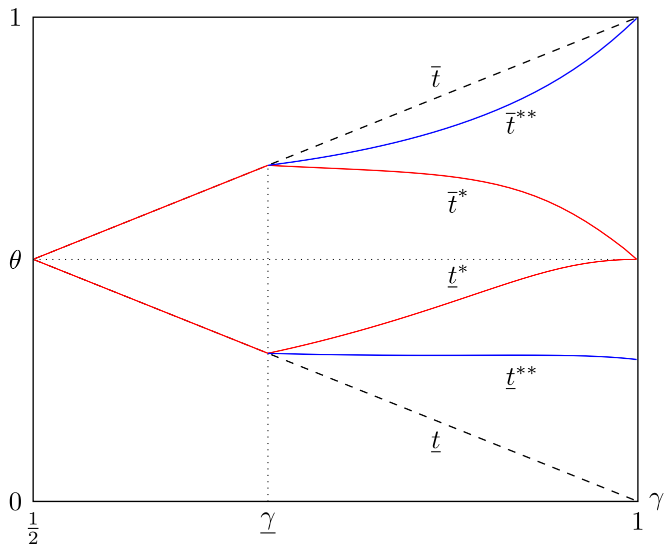


Figure 6: The two cutoffs, $\underline{t}^{**}$ and $\bar{t}^{**}$, as functions of γ, are depicted by blue solid curves, using the same parameterization as in Figure 4.

lest she disrupts the bank's incentives to take high risks only when its signal is g.

Nevertheless, commitment over supervisory ruling raises the supervisor's expected payoff sufficiently to ensure that welfare monotonically increases in γ; that is, it makes more information in the hands of the bank always good news to the supervisor. This is in stark contrast to the baseline case where the supervisor's indirect payoff is non-monotonic in γ and is minimized in the limit as γ approaches 1.

Proposition 7 *Suppose that the supervisor can commit to her follow-up supervisory action. Then, the supervisor's expected payoff is monotone-increasing in γ.*

Given that the supervisor still cannot obtain her first-best outcome, commitment to supervisory ruling is certainly not a perfect solution, but it would ensure that more information always results in higher welfare. Hence, it would be desirable to give the supervisor enough commitment power, particularly in the form of supervisory ruling. Commitment power over how much she discloses about her own information alone can be impotent in case the bank experiences a sudden boost in its private information.

9 Conclusion

Ensuring that banks operate in a safe and sound manner hinges critically on effective banking supervision, which in turn depends on supervisors' ability to make an informed assessment of risks taken by banks and to compel banks to take corrective actions. One important tool available to supervisors to influence banks' engagement with risky investments is supervisory guidance—communication prior to bank examinations. In this paper, we study a model of supervisory guidance that reflects the reality of incomplete information in banking supervision: supervisors and banks have correlated but different information of their own.

We characterize the structure of optimal communication and show that it features muddling: the supervisor tells the bank to act on its signal not only when she is moderate (when she really wants it to), but also when she is sufficiently pessimistic (when she is ready to disapprove of any high risk investments). This way, the supervisor induces the bank to use its signal to assess the likelihood that she will approve high risk investments, as well as to make its investment decisions.

An important lesson from our analysis is that a highly informative bank signal can compromise the effectiveness of banking supervision. In particular, the supervisor can rely ironically only on her own (imperfect) information when the bank has almost perfect information. This accentuates the fact that supervisors need adequate powers to keep banks safe and sound. To this end, our theoretical analysis identifies two dimensions along which supervisors should be granted more power in order to ensure more effective bank supervision: (i) raising the cost of supervisory disapproval for banks and (ii) giving the supervisor ex-ante commitment power over her supervisory ruling.

Our model can be extended to explore several interesting related issues. For example, whereas the bank is assumed to receive information according to an exogenously given technology in our model, banks' information is endogenous in reality; it would be interesting to allow the bank to engage in (flexible) information acquisition. While we consider the problem of a single bank, many banks are interconnected to one another in reality; one could consider the case where there are multiple banks (receivers) and the supervisor aggregates information across them in "horizontal" examinations.[28]

[28] See "What is a horizontal examination?" in Approaches to Bank Supervision.

References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 2016, *131* (4), 1593–1636.

Bergemann, Dirk and Stephen Morris, "Information Design: A Unified Perspective," *Journal of Economic Literature*, 2019, *57* (1), 44–95.

Berger, Allen N. and Gregory F. Udell, "Relationship Lending and Lines of Credit in Small Firm Finance," *Journal of Business*, 1995, *68* (3), 351–381.

Blackwell, David, "Comparison of experiments," in "Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability" University of California Press 1951, pp. 93–103.

Bouvard, Matthieu, Pierre Chaigneau, and Adolfo de Motta, "Transparency in the Financial System: Rollover Risk and Crises," *The Journal of Finance*, 2015, *70* (4), 1805–1837.

Bräuning, Falk and José L. Fillat, "The Impact of Regulatory Stress Tests on Bank Lending and Its Macroeconomic Consequences," 2020. Working Paper.

Colliard, Jean-Edouard, "Strategic Selection of Risk Models and Bank Capital Regulation," *Management Science*, 2019, *65* (6), 2591–2606.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Eckley, Peter, William Francis, and Aakriti Mathur, "In the dangerzone! Regulatory uncertainty and voluntary bank capital surpluses," 2021. Working Paper.

Egan, Mark, Stefan Lewellen, and Adi Sunderam, "The Cross-Section of Bank Value," *The Review of Financial Studies*, 2022, *35* (5), 2101–2143.

Faria-e-Castro, Miguel, Joseba Martinez, and Thomas Philippon, "Runs versus Lemons: Information Disclosure and Fiscal Capacity," *The Review of Economic Studies*, 2017, *84* (4), 1683–1707.

Gissler, Stefan, Jeremy Oldfather, and Doriana Ruffino, "Lending on hold: Regulatory uncertainty and bank lending standards," *Journal of Monetary Economics*, 2016, *81*, 89–101.

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

_ and _ , "Stress Test Disclosure: Theory, Practice, and New Perspectives," in J. Doyne Farmer, Alissa M. Kleinnijenhuis, Til Schuermann, and Thom Wetzer, eds., *Handbook of Financial Stress Testing*, Cambridge: Cambridge University Press, 2022, pp. 208–223.

Guo, Yingni and Eran Shmaya, "The Interval Structure of Optimal Disclosure," *Econometrica*, 2019, *87* (2), 653–675.

Huang, Jing, "Optimal Stress Tests in Financial Networks," 2021. Working Paper.

Inostroza, Nicolas, "Persuading Multiple Audiences: Strategic Complementarities and (Robust) Regulatory Disclosures," 2023. Working Paper.

_ and Alessandro Pavan, "Adversarial Coordination and Public Information Design," 2023. Working Paper.

Jehle, Geoffrey A. and Philip J. Reny, *Advanced Microeconomic Theory*, Pearson, 2010.

Julio, Brandon and Youngsuk Yook, "Political Uncertainty and Corporate Investment Cycles," *The Journal of Finance*, 2012, *67* (1), 45–83.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 2019, *11* (1), 249–272.

Kolotilin, Anton, "Optimal information disclosure: a linear programming approach," *Theoretical Economics*, 2018, *13* (2), 607–636.

_ , Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li, "Persuasion of a Privately Informed Receiver," *Theoretical Economics*, 2017, *85* (6), 1949–1964.

Leitner, Yaron and Basil Williams, "Model Secrecy and Stress Tests," *The Journal of Finance*, 2023, *78* (2), 1055–1095.

_ **and Bilge Yilmaz**, “Regulating a model,” *Journal of Financial Economics*, 2019, *131* (2), 251–268.

Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypacz, “The Design of Macroprudential Stress Tests,” *The Review of Financial Studies*, 2023, *36* (11), 4460–4501.

Parlatore, Cecilia and Thomas Philippon, “Designing Stress Scenarios,” 2023. Working Paper.

Petersen, Mitchell A. and Raghuram G. Rajan, “The Benefits of Lending Relationships: Evidence from Small Business Data,” *The Journal of Finance*, 1994, *49* (1), 3–37.

Quigley, Daniel and Ansgar Walther, “Inside and Outside Information,” 2023. *The Journal of Finance*, Forthcoming.

Rhee, Keeyoung and Keshav Dogra, “Stress tests and model monoculture,” *Journal of Financial Economics*, 2024, *152*, 103760.

Shapiro, Joel and Jing Zeng, “Stress Testing and Bank Lending,” 2023. *The Review of Financial Studies*, Forthcoming.

Sufi, Amir, “Information Asymmetry and Financing Arrangements: Evidence from Syndicated Loans,” *The Journal of Finance*, 2007, *62* (2), 629–668.

Williams, Basil, “Stress Tests and Bank Portfolio Choice,” 2017. Working Paper.

A Omitted Proofs

The following result will be useful for a number of proofs.

Lemma 5 *Fix $\gamma \in [1/2, 1]$. If*

$$\int_{T^{h\ell}\cap[\underline{t},1]} [(1-\gamma)t+\gamma(1-t)]\ \mathrm{d}F(t) = c\int_{T^{h\ell}\cap[0,\underline{t})} [(1-\gamma)t+\gamma(1-t)]\ \mathrm{d}F(t),$$

then the following inequality holds:

$$\int_{T^{h\ell}\cap[\underline{t},1]} [\gamma t+(1-\gamma)(1-t)]\ \mathrm{d}F(t) \geq c\int_{T^{h\ell}\cap[0,\underline{t})} [\gamma t+(1-\gamma)(1-t)]\ \mathrm{d}F(t).$$

Proof. Since $(1-\gamma)t+\gamma(1-t) = \gamma-(2\gamma-1)t$ is decreasing in t, the given equality implies that

$$\begin{aligned}&\int_{T^{h\ell}\cap[\underline{t},1]} [(1-\gamma)\underline{t}+\gamma(1-\underline{t})]\ \mathrm{d}F(t) \geq \int_{T^{h\ell}\cap[\underline{t},1]} [(1-\gamma)t+\gamma(1-t)]\ \mathrm{d}F(t)\\ &= c\int_{T^{h\ell}\cap[0,\underline{t})} [(1-\gamma)t+\gamma(1-t)]\ \mathrm{d}F(t) \geq c\int_{T^{h\ell}\cap[0,\underline{t})} [(1-\gamma)\underline{t}+\gamma(1-\underline{t})]\ \mathrm{d}F(t),\end{aligned}$$

which can be simplified to

$$\int_{T^{h\ell}\cap[\underline{t},1]} \mathrm{d}F(t) \geq c\int_{T^{h\ell}\cap[0,\underline{t})} \mathrm{d}F(t).$$

Combining this with the fact that $\gamma t+(1-\gamma)(1-t) = (2\gamma-1)t+1-\gamma$ is increasing in t,

$$\begin{aligned}&\int_{T^{h\ell}\cap[\underline{t},1]} [\gamma t+(1-\gamma)(1-t)]\ \mathrm{d}F(t) \geq \int_{T^{h\ell}\cap[\underline{t},1]} [\gamma\underline{t}+(1-\gamma)(1-\underline{t})]\ \mathrm{d}F(t)\\ &= [\gamma\underline{t}+(1-\gamma)(1-\underline{t})]\int_{T^{h\ell}\cap[\underline{t},1]} \mathrm{d}F(t) \geq [\gamma\underline{t}+(1-\gamma)(1-\underline{t})]\, c\int_{T^{h\ell}\cap[0,\underline{t})} \mathrm{d}F(t)\\ &= c\int_{T^{h\ell}\cap[0,\underline{t})} [\gamma\underline{t}+(1-\gamma)(1-\underline{t})]\ \mathrm{d}F(t) \geq c\int_{T^{h\ell}\cap[0,\underline{t})} [\gamma t+(1-\gamma)(1-t)]\ \mathrm{d}F(t).\end{aligned}$$

■

Proof of Lemma 1. Suppose there exists $T^{h\ell} \subset [0,\bar{t})$ such that $[\underline{t},\bar{t}) \subset T^{h\ell}$ and the following incentive constraints hold:

$$\int_{\underline{t}}^{\bar{t}} [\gamma t + (1-\gamma)(1-t)]\ \mathrm{d}F(t) \geq c \int_{T^{h\ell}/[\underline{t},\bar{t})} [\gamma t + (1-\gamma)(1-t)]\ \mathrm{d}F(t) \text{ and}$$
$$\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t) \leq c \int_{T^{h\ell}/[\underline{t},\bar{t})} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t).$$

Let $\tau \leq \underline{t}$ be the value such that

$$\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t) = c \int_{0}^{\tau} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t).$$

This value is well defined because the right-hand side is equal to 0 when $\tau = 0$, continuously rises, and is greater than the left-hand side when $\tau = \underline{t}$, that is,

$$\begin{aligned}\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t) &\leq c \int_{T^{h\ell}/[\underline{t},\bar{t})} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t) \\ &\leq c \int_{0}^{\underline{t}} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t).\end{aligned}$$

By Lemma 5, $\mathcal{T}^{h\ell} := [0,\tau) \cup [\underline{t},\bar{t})$ satisfies the other incentive constraint as well. ■

Proof of Lemma 2. We first show that it is impossible to induce the bank to take low risks when $s = g$ and high risks when $s = b$. Suppose there exists a message $m = \ell h$ that induces such action by the bank, and let $T^{\ell h}$ denote the set of supervisor types for which the supervisor sends $m = \ell h$. Then, if the bank takes high risks, the supervisor infers $s = b$, in which case it is optimal for her to approve if and only if $t \geq \bar{t}$. Given this supervisory ruling, for the bank to be willing to take high risks with $s = b$, the following inequality must hold:

$$\int_{T^{\ell h}\cap[\bar{t},1]} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t) \geq c \int_{T^{\ell h}\cap[0,\bar{t})} [(1-\gamma)t + \gamma(1-t)]\ \mathrm{d}F(t).$$

Since $(1-\gamma)t + \gamma(1-t)$ is strictly decreasing in t, the given inequality implies that

$$\begin{aligned}
&\int_{T^{\ell h}\cap[\bar{t},1]} \left[(1-\gamma)\bar{t} + \gamma(1-\bar{t})\right] \, \mathrm{d}F(t) \geq \int_{T^{\ell h}\cap[\bar{t},1]} \left[(1-\gamma)t + \gamma(1-t)\right] \mathrm{d}F(t) \\
\geq c &\int_{T^{\ell h}\cap[0,\bar{t})} \left[(1-\gamma)t + \gamma(1-t)\right] \mathrm{d}F(t) \geq c \int_{T^{\ell h}\cap[0,\bar{t})} \left[(1-\gamma)\bar{t} + \gamma(1-\bar{t})\right] \, \mathrm{d}F(t),
\end{aligned}$$

which can be simplified to

$$\int_{T^{\ell h}\cap[\bar{t},1]} \mathrm{d}F(t) \geq c \int_{T^{\ell h}\cap[0,\bar{t})} \mathrm{d}F(t). \tag{10}$$

Note that this inequality will hold strictly whenever $T^{\ell h}$ is non-negligible. Similarly, for the bank to be willing to take low risks with $s = g$,

$$\int_{T^{\ell h}\cap[\bar{t},1]} \left[\gamma t + (1-\gamma)(1-t)\right] \mathrm{d}F(t) \leq c \int_{T^{\ell h}\cap[0,\bar{t})} \left[\gamma t + (1-\gamma)(1-t)\right] \mathrm{d}F(t).$$

Using a similar argument as above and the fact that $\gamma t + (1-\gamma)(1-t)$ is strictly increasing in t, we have

$$\int_{T^{\ell h}\cap[\bar{t},1]} \mathrm{d}F(t) \leq c \int_{T^{\ell h}\cap[0,\bar{t})} \mathrm{d}F(t),$$

which would contradict (10), provided that $T^{\ell h}$ is non-negligible (in which case both inequalities would hold strictly).

For any communication strategy (M, π), let $a(m) \in \{hh, h\ell, \ell\ell\}$ denote the bank's action (conditional on its signal s) after observing m. Now, consider the following ternary signal: $M' = \{hh, h\ell, \ell\ell\}$ and

$$\pi'(t) = \begin{cases} hh & \text{if } \pi(t) \in M^{hh} := \{m \in M : a(m) = hh\} \\ h\ell & \text{if } \pi(t) \in M^{h\ell} := \{m \in M : a(m) = h\ell\} \\ \ell\ell & \text{if } \pi(t) \in M^{\ell\ell} := \{m \in M : a(m) = \ell\ell\}. \end{cases}$$

In other words, this alternative signal merges all messages that trigger the same action by the bank into one. For each $m \in M^{hh}$, the supervisor's optimal approval cutoff is $\widehat{t}$. Therefore, we have

$$\int_{T^m \cap [\widehat{t},1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \geq c \int_{T^m \cap [0,\widehat{t})} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \text{ and}$$
$$\int_{T^m \cap [\widehat{t},1]} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t) \geq c \int_{T^m \cap [0,\widehat{t})} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t).$$

Since both are linear in T^m, aggregating each inequality over M^{hh}, we arrive at

$$\int_{T^{hh} \cap [\widehat{t},1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \geq c \int_{T^{hh} \cap [0,\widehat{t})} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \text{ and}$$
$$\int_{T^{hh} \cap [\widehat{t},1]} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t) \geq c \int_{T^{hh} \cap [0,\widehat{t})} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t).$$

This means that merging messages that induce unconditional high risk-taking by the bank does not affect its incentives (as well as the supervisor's optimal approval decision). The same argument can be used to merge messages that induce $a(m) = h\ell$ or $a(m) = \ell\ell$. ■

Proof of Lemma 3. We show that (IC_b^{hh}) implies (IC_g^{hh}). Since $(1-\gamma)t + \gamma(1-t)$ is decreasing in t, (IC_b^{hh}) implies that

$$\int_{T^{hh} \cap [\widehat{t},1]} \left[(1-\gamma)\widehat{t} + \gamma(1-\widehat{t})\right] \mathrm{d}F(t) \geq \int_{T^{hh} \cap [\widehat{t},1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t)$$
$$\geq c \int_{T^{hh} \cap [0,\widehat{t})} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \geq c \int_{T^{hh} \cap [0,\widehat{t})} \left[(1-\gamma)\widehat{t} + \gamma(1-\widehat{t})\right] \mathrm{d}F(t),$$

which can be simplified to

$$\int_{T^{hh} \cap [\widehat{t},1]} \mathrm{d}F(t) \geq c \int_{T^{hh} \cap [0,\widehat{t})} \mathrm{d}F(t).$$

Combining this with the fact that $\gamma t + (1-\gamma)(1-t)$ is increasing in t,

$$\int_{T^{hh} \cap [\widehat{t},1]} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t) \geq \int_{T^{hh} \cap [\widehat{t},1]} \left[\gamma\widehat{t} + (1-\gamma)(1-\widehat{t})\right] \mathrm{d}F(t)$$
$$\geq c \int_{T^{hh} \cap [0,\widehat{t})} \left[\gamma\widehat{t} + (1-\gamma)(1-\widehat{t})\right] \mathrm{d}F(t) \geq c \int_{T^{hh} \cap [0,\widehat{t})} [\gamma t + (1-\gamma)(1-t)] \, \mathrm{d}F(t),$$

which yields (IC_g^{hh}). It is easy to modify this proof also to show that $(IC_g^{\ell\ell})$ implies $(IC_b^{\ell\ell})$.

Suppose that $V := T^{hh} \cap [0, \widehat{t}) \neq \emptyset$ at the optimal solution. Consider $\widetilde{T}^{hh} = T^{hh} \setminus V$ and $\widetilde{T}^{\ell\ell} = T^{\ell\ell} \cup V$. This change relaxes (IC_b^{hh}) and $(IC_g^{\ell\ell})$ by lowering the right-hand side of (IC_b^{hh}), while raising the right-hand side of $(IC_g^{\ell\ell})$. However, it leaves both $(IC_g^{h\ell})$ and $(IC_b^{h\ell})$ unaffected. For any $t \in V$, the supervisor is indifferent between sending $m = \ell\ell$ and $m = hh$ because, in the latter case, she would disapprove of the bank's high risk-taking. This implies that this alternative signal is also an optimal solution.

Now suppose that $V' := T^{\ell\ell} \cap [\widehat{t}, 1] \neq \emptyset$, and consider $\widetilde{T}^{\ell\ell} = T^{\ell\ell} \setminus V'$ and $\widetilde{T}^{hh} = T^{hh} \cup V'$. As above, this change does not affect the bank's incentives. However, it raises the supervisor's payoff because the bank now chooses hh, instead of $\ell\ell$, for $t \in V' \subset [\widehat{t}, 1]$, and the supervisor's payoff when the bank chooses hh $(t - (1-t)d)$ is larger than her payoff when the bank chooses $\ell\ell$ (0) whenever $t \geq \widehat{t}$. ■

Proof of Lemma 4. Let $\tau \in [0, \underline{t}]$ be the unique value such that

$$\int_{T^{h\ell} \cap [\underline{t}, 1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) = c \int_0^{\tau} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t).$$

Note that τ is well defined since the right-hand side is continuous and strictly increasing in τ, and

$$\begin{aligned} c \int_0^0 [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) &\leq \int_{T^{h\ell} \cap [\underline{t}, 1]} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \\ \leq c \int_{T^{h\ell} \cap [0, \underline{t})} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) &\leq c \int_0^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \end{aligned}$$

where the second inequality holds because $T^{h\ell}$ satisfies $(IC_b^{h\ell})$. Consider $\mathcal{T}^{h\ell} := [0, \tau) \cup (T^{h\ell} \setminus [0, \underline{t}))$. Then, by construction (the definition of τ), $\mathcal{T}^{h\ell}$ satisfies $(IC_b^{h\ell})$ with equality. By Lemma 5, it also satisfies $(IC_g^{h\ell})$. ■

Proof of Proposition 3. As the supervisor can always use her own information for supervision, her expected payoff is bounded below by

$$\underline{W} := \int_{\widehat{t}}^{1} [t - (1-t)d] \, \mathrm{d}F(t).$$

Based on the characterizations in Sections 5 and 6, it is straightforward to see that $W(\gamma) > \underline{W}$

whenever $\gamma \in (1/2, 1)$. Therefore, it suffices to show that $\lim_{\gamma \to \frac{1}{2}} W(\gamma) = \lim_{\gamma \to 1} W(\gamma) = \underline{W}$. For $\gamma \to 1/2$, the result follows from the fact that

$$\underline{W} \leq W(\gamma) \leq \overline{W}(\gamma) := \int_{\underline{t}}^{\bar{t}} [\gamma t - (1-\gamma)(1-t)d] \; \mathrm{d}F(t) + \int_{\bar{t}}^{1} [t - (1-t)d] \; \mathrm{d}F(t),$$

and that, as γ approaches $\frac{1}{2}$, both $\underline{t}$ and $\bar{t}$ converge to $\widehat{t}$, and so $\overline{W}(\gamma)$ converges to $\underline{W}$.

For $\gamma \to 1$, recall that $\underline{t}^*$ and $\bar{t}^*$ should satisfy

$$\int_{\underline{t}^*}^{\bar{t}^*} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) = c \int_{0}^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t).$$

As γ approaches 1, $\underline{t}$ converges to 0, which in turn implies that both $\underline{t}^*$ and $\bar{t}^*$ converge to $\widehat{t}$, and so $\lim_{\gamma \to 1} W(\gamma) = \underline{W}$. ■

Proof of Proposition 4. First, note that neither $\underline{t}$ nor $\bar{t}$ depends on c, so (4) holds, and the supervisor obtains her first-best outcome, if and only if $c \geq c^*$. The result for the case where $c \geq c^*$ then follows from the fact that the supervisor's first-best outcome also does not depend on c.

Next, suppose $c < c^*$. In this case, the supervisor's indirect payoff is given by

$$W(c) := \int_{\underline{t}^*}^{\bar{t}^*} [\gamma t - (1-\gamma)(1-t)d] \, \mathrm{d}F(t) + \int_{\bar{t}^*}^{1} [t - (1-t)d] \, \mathrm{d}F(t),$$

where $\underline{t}^*$ and $\bar{t}^*$ are the values that satisfy

$$\int_{\underline{t}^*}^{\bar{t}^*} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) = c \int_{0}^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t).$$

If c rises, $[\underline{t}^*, \bar{t}^*)$ has to expand to satisfy the last equality. Since $[\underline{t}^*, \bar{t}^*)$ is a strict subset of $[\underline{t}, \bar{t})$ whenever $c < c^*$, $\gamma t - (1-\gamma)(1-t)d > \max\{0, t - (1-t)d\}$ for all $t \in [\underline{t}^*, \bar{t}^*)$, so any expansion of $[\underline{t}^*, \bar{t}^*)$ is clearly beneficial to the supervisor. Hence, $W(c)$ is strictly increasing whenever $c < c^*$. ■

Proof of Proposition 6. It is without loss to assume that $T^{h\ell}_{\text{Disappr.}} = [0,1] \setminus (T^{h\ell}_{\text{Appr.}} \cup T^{hh})$ (i.e., $T^{\ell\ell} = \emptyset$), as it maximizes the right-hand side of (9) while leaving the objective function unaffected.

Then, the Lagrangian for this problem is given by

$$\begin{aligned}\mathcal{L} &= \int_{T^{h\ell}_{\text{Appr.}}} [\gamma t - (1-\gamma)(1-t)d]\,\mathrm{d}F(t) + \int_{T^{hh}} [t-(1-t)d]\,\mathrm{d}F(t)\\ &\quad + \lambda\left\{c\int_{T^{h\ell}_{\text{Disappr.}}} [(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t) - \int_{T^{h\ell}_{\text{Appr.}}} [(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t)\right\}\\ &= \int_{T^{h\ell}_{\text{Appr.}}} \{\gamma t - (1-\gamma)(1-t)d - \lambda(c+1)[(1-\gamma)t+\gamma(1-t)]\}\ \mathrm{d}F(t)\\ &\quad + \int_{T^{hh}} \{t-(1-t)d - \lambda c[(1-\gamma)t+\gamma(1-t)]\}\ \mathrm{d}F(t)\\ &\quad + \lambda c\int_0^1 [(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t).\end{aligned}$$

For each $\lambda \geq 0$, let $\underline{t}^{**}(\lambda)$ be the value of t such that

$$\gamma t - (1-\gamma)(1-t)d - \lambda(c+1)[(1-\gamma)t+\gamma(1-t)] = 0.$$

As before, if we let $\overline{\lambda}$ be the value of λ such that $\underline{t}^{**}(\lambda) = \widehat{t}$, then $\underline{t}^{**}(\lambda)$ continuously and strictly increases from $\underline{t}$ to $\widehat{t}$ as λ rises from 0 to $\overline{\lambda}$. Similarly, let $\overline{t}^{**}(\lambda)$ be the value of t such that

$$\gamma t - (1-\gamma)(1-t)d - \lambda[(1-\gamma)t+\gamma(1-t)] = t-(1-t)d.$$

Note that $\overline{t}^{**}(\lambda)$ is defined in exactly the same way as in Section 5.2.

By the definitions of $\underline{t}^{**}(\lambda)$ and $\overline{t}^{**}(\lambda)$, $t \in [\underline{t}^{**}(\lambda), \overline{t}^{**}(\lambda)]$ if and only if

$$\begin{aligned}&\gamma t - (1-\gamma)(1-t)d - \lambda(c+1)[(1-\gamma)t+\gamma(1-t)]\\ &\geq \max\{0, t-(1-t)d - \lambda c[(1-\gamma)t+\gamma(1-t)]\}.\end{aligned}$$

This implies that at the optimal solution, $T^{h\ell}_{\text{Appr.}} = [\underline{t}^{**}(\lambda), \overline{t}^{**}(\lambda))$ and $T^{hh} = [\overline{t}^{**}(\lambda), 1]$ for some $\lambda \geq 0$. The optimal value of λ can then be found from the following equation:

$$\int_{\underline{t}^{**}(\lambda)}^{\overline{t}^{**}(\lambda)} [(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t) = c\int_0^{\underline{t}^{**}(\lambda)} [(1-\gamma)t+\gamma(1-t)]\,\mathrm{d}F(t).$$

For the same reasons as in Section 5.2, there exists a unique value of λ that satisfies this equation (whenever (4) fails to hold). ■

Proof of Proposition 7. Fix $\gamma > \frac{1}{2}$, and consider the corresponding optimal solution, $\underline{t}^{**}$ and $\overline{t}^{**}$, when the supervisor can commit to her follow-up supervisory ruling. First, note that this optimal solution satisfies (9) with equality, that is,

$$\int_{\underline{t}^{**}}^{\overline{t}^{**}} [(1-\gamma)t + \gamma(1-t)]\,\mathrm{d}F(t) = c\int_0^{\underline{t}^{**}} [(1-\gamma)t + \gamma(1-t)]\,\mathrm{d}F(t).$$

This solution is also feasible for any $\gamma' > \gamma$ if and only if

$$\int_{\underline{t}^{**}}^{\overline{t}^{**}} [(1-\gamma')t + \gamma'(1-t)]\,\mathrm{d}F(t) \leq c\int_0^{\underline{t}^{**}} [(1-\gamma')t + \gamma'(1-t)]\,\mathrm{d}F(t).$$

Combining this inequality with the equation above, this inequality can be rewritten as

$$\int_{\underline{t}^{**}}^{\overline{t}^{**}} (1-2t)\,\mathrm{d}F(t) \leq c\int_0^{\underline{t}^{**}} (1-2t)\,\mathrm{d}F(t).$$

Dividing each side of this inequality by that of the equation above, this can also be written as

$$\frac{\int_{\underline{t}^{**}}^{\overline{t}^{**}} (1-2t)\,\mathrm{d}F(t)}{\int_{\underline{t}^{**}}^{\overline{t}^{**}} [t+\gamma(1-2t)]\,\mathrm{d}F(t)} \leq \frac{\int_0^{\underline{t}^{**}} (1-2t)\,\mathrm{d}F(t)}{\int_0^{\underline{t}^{**}} [t+\gamma(1-2t)]\,\mathrm{d}F(t)},$$

which can be simplified to

$$\int_0^{\underline{t}^{**}} tdF\left(t\left|t \in [0, \underline{t}^{**})\right.\right) \leq \int_{\underline{t}^{**}}^{\overline{t}^{**}} tdF\left(t\left|t \in [\underline{t}^{**}, \overline{t}^{**})\right.\right),$$

which is obviously true. Therefore, $(\underline{t}^{**}, \overline{t}^{**})$ is still feasible for any $\gamma' > \gamma$. It is easy to check that this strategy gives a strictly higher expected payoff to the supervisor when γ increases to $\gamma' > \gamma$. Therefore, the supervisor's optimal expected payoff must be strictly monotone-increasing in γ on the interval $(\frac{1}{2}, 1)$. ■

B Achieving the First-Best Outcome with a Ternary Bank Signal

The most restrictive assumption in our model is that the bank's signal is binary, that is, $s \in \{g, b\}$. In this appendix, we show how to extend some of our analyses and results for a more general case. To avoid an overly lengthy and technical discussion, we focus on the case where the bank's signal takes a simple ternary form and study only the condition under which the supervisor can achieve her first-best outcome.

Setup. The only difference from our main model is that now the bank's signal s can take on one of three values: g, b, or n ("neutral"). The probability of each signal realization, conditional on the state $\omega \in \{G, B\}$, is given in the following table:

state/realization	g	n	b
G	$\gamma - \varepsilon$	2ε	$1 - \gamma - \varepsilon$
B	$1 - \gamma - \varepsilon$	2ε	$\gamma - \varepsilon$

for some $\gamma \in (1/2, 1)$ and $\varepsilon \leq \min\{\gamma, 1 - \gamma\}$. This signal generalizes the binary signal adopted in the main model in the simplest way possible, by adding a truly uninformative signal realization n.

Supervisor's first-best outcome. As in Section 4.1, define $\widehat{t}$, $\underline{t}$, and $\overline{t}$ as follows:

$$\widehat{t} = \frac{d}{1+d}, \quad \frac{\underline{t}}{1-\underline{t}} = \frac{\widehat{t}}{1-\widehat{t}} \frac{1-\gamma-\varepsilon}{\gamma-\varepsilon}, \text{ and } \frac{\overline{t}}{1-\overline{t}} = \frac{\widehat{t}}{1-\widehat{t}} \frac{\gamma-\varepsilon}{1-\gamma-\varepsilon}.$$

Then, the supervisor's first-best outcome is to have the bank take high risks if and only if one of the following conditions holds: (i) $t \geq \overline{t}$ (regardless of $s \in \{g, n, b\}$), (ii) $t \in [\widehat{t}, \overline{t})$ and $s \in \{g, n\}$, and (iii) $t \in [\underline{t}, \widehat{t})$ and $s = g$. To achieve this first-best outcome, the supervisor's messages should distinguish among these three cases. Similarly to the main model, let hhh denote the message sent when $t \geq \overline{t}$, $hh\ell$ when $t \in [\widehat{t}, \overline{t})$, and $h\ell\ell$ when $t \in [\underline{t}, \widehat{t})$. Finally, we use $\ell\ell\ell$ to represent the supervisor's recommendation that the bank should avoid taking high risks regardless of its signal.

Incentive constraints. Let T^m denote the set of the supervisor's types on which the supervisor sends message $m \in \{hhh, hh\ell, h\ell\ell, \ell\ell\ell\}$, and IC_s^m denote the bank's relevant incentive constraint when the supervisor's message is $m \in \{hhh, hh\ell, h\ell\ell, \ell\ell\ell\}$ and the bank's signal is $s \in \{g, n, b\}$. For example, $IC_b^{hh\ell}$ is given by

$$\int_{\widehat{t}}^{\bar{t}} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t) \leq c \int_{T^{hh\ell}\cap[0,\widehat{t})} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t),$$

while $IC_n^{hh\ell}$ is given by

$$\int_{\widehat{t}}^{\bar{t}} [2\varepsilon t + 2\varepsilon(1-t)] \,\mathrm{d}F(t) \geq c \int_{T^{hh\ell}\cap[0,\widehat{t})} [2\varepsilon t + 2\varepsilon(1-t)] \,\mathrm{d}F(t).$$

Applying similar arguments to those in Sections 4 and 5, we find that the only relevant conditions for achieving the first-best outcome are $IC_b^{hh\ell}$ and $IC_n^{h\ell\ell}$. First, $T^{hhh} = [\bar{t}, 1]$ and $T^{lll} \subset [0, \underline{t})$ ensure that all (six) incentive constraints associated with hhh and $\ell\ell\ell$ hold. Next, it is easy to see that $IC_n^{hh\ell}$ implies $IC_g^{hh\ell}$, while $IC_n^{h\ell\ell}$ implies $IC_b^{h\ell\ell}$. Finally, we can use the fact that the bank intrinsically prefers to take more risks to conclude that the only potentially binding constraints are those associated with incentives to avoid high risks, namely $IC_b^{hh\ell}$ and $IC_n^{h\ell\ell}$.

The following lemma is important, as it allows us to focus on particularly simple structures of $T^{hh\ell}$ and $T^{h\ell\ell}$, depicted in Figure 7.

Lemma 6 *If the supervisor can attain her first-best outcome, there exist $\tau_1 \in [0, \underline{t}]$ and $\tau_2 \in [\tau_1, \underline{t}]$ such that she can achieve it with $T^{hh\ell} = [0, \tau_1) \cup [\widehat{t}, \bar{t})$ and $T^{h\ell\ell} = [\tau_1, \tau_2) \cup [\underline{t}, \widehat{t})$.*

Proof. The discussion above implies that the supervisor can achieve her first-best outcome if and only if there exist two disjoint subsets of $[0, \underline{t})$—$A = T^{hh\ell} \cap [0, \underline{t})$ and $B = T^{h\ell\ell} \cap [0, \underline{t})$—that satisfy the following two incentive constraints ($IC_b^{hh\ell}$ and $IC_n^{h\ell\ell}$):

$$\int_{\widehat{t}}^{\bar{t}} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t) \leq c \int_{A} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t), \quad (11)$$

$$\int_{\underline{t}}^{\widehat{t}} [2\varepsilon t + 2\varepsilon(1-t)] \,\mathrm{d}F(t) \leq c \int_{B} [2\varepsilon t + 2\varepsilon(1-t)] \,\mathrm{d}F(t) \Leftrightarrow \int_{\underline{t}}^{\widehat{t}} \mathrm{d}F(t) \leq c \int_{B} \mathrm{d}F(t). \quad (12)$$

Let $\tau_1 \in [0, \underline{t}]$ be the value such that

$$\int_{\widehat{t}}^{\bar{t}} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t) = c \int_{0}^{\tau_1} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \,\mathrm{d}F(t). \quad (13)$$

This value is well defined in $[0, \underline{t})$ since, as τ_1 increases from 0 to $\underline{t}$, the right-hand side increases from 0 to $c\int_0^{\underline{t}} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t) \geq c\int_A [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t)$.

Now, it suffices to show that

$$\int_B \mathrm{d}F(t) \leq \int_{\tau_1}^{\underline{t}} \mathrm{d}F(t), \tag{14}$$

which would imply that $\int_{\underline{t}}^{\widehat{t}} \mathrm{d}F(t) \leq c\int_B \mathrm{d}F(t) \leq c\int_{\tau_1}^{\underline{t}} \mathrm{d}F(t)$. Combining (11) and (13), we get

$$\int_0^{\tau_1} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t) \leq \int_A [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t),$$

which is equivalent to

$$\int_{[0,\tau_1)/A} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t) \leq \int_{A/[0,\tau_1)} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t).$$

Since $(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)$ is decreasing in t (for any $\gamma > 1/2$), we have

$$\begin{aligned}
&\int_{[0,\tau_1)/A} [(1-\gamma-\varepsilon)\tau_1 + (\gamma-\varepsilon)(1-\tau_1)]\ \mathrm{d}F(t) \\
\leq &\int_{[0,\tau_1)/A} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t) \\
\leq &\int_{A/[0,\tau_1)} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)]\ \mathrm{d}F(t) \\
\leq &\int_{A/[0,\tau_1)} [(1-\gamma-\varepsilon)\tau_1 + (\gamma-\varepsilon)(1-\tau_1)]\ \mathrm{d}F(t),
\end{aligned}$$

which yields

$$\int_{[0,\tau_1)/A} \mathrm{d}F(t) \leq \int_{A/[0,\tau_1)} \mathrm{d}F(t) \Leftrightarrow \int_0^{\tau_1} \mathrm{d}F(t) \leq \int_A \mathrm{d}F(t).$$

For the desired inequality (14), it suffices to observe that, since $A \cup B \subseteq [0, \underline{t})$ and $A \cap B = \emptyset$,

$$\int_B \mathrm{d}F(t) \leq \int_0^{\underline{t}} \mathrm{d}F(t) - \int_A \mathrm{d}F(t) \leq \int_0^{\underline{t}} \mathrm{d}F(t) - \int_0^{\tau_1} \mathrm{d}F(t) = \int_{\tau_1}^{\underline{t}} \mathrm{d}F(t).$$

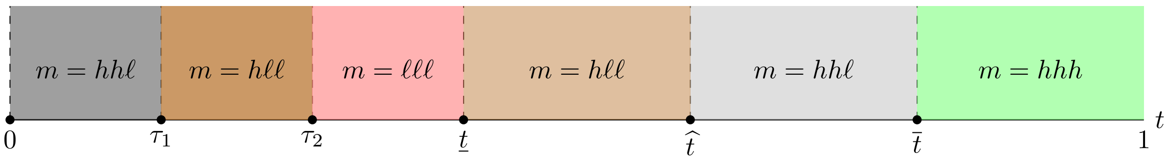


Figure 7: The structure of optimal communication for achieving the supervisor's first-best outcome when the bank's signal is ternary.

■

The above result immediately implies the following result, which serves as a ternary counterpart to Proposition 1.

Proposition 8 *When the bank's signal is ternary, the supervisor can obtain her first-best outcome if and only if there exists* $\tau \in [0, \underline{t}]$ *that satisfies* $\int_{\underline{t}}^{\widehat{t}} \mathrm{d}F(t) = c \int_{\tau}^{\underline{t}} \mathrm{d}F(t)$ *and*

$$\int_{\widehat{t}}^{\bar{t}} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \; \mathrm{d}F(t) \leq c \int_{0}^{\tau} [(1-\gamma-\varepsilon)t + (\gamma-\varepsilon)(1-t)] \; \mathrm{d}F(t).$$

As in our main model, the condition in Proposition 8 holds if γ is close to $\frac{1}{2}$ (i.e., the bank's signal is sufficiently uninformative); in that case, $\bar{t}$ is close to $\widehat{t}$, so the necessary inequality is easily satisfied. It is also straightforward to show that the condition fails when γ is sufficiently close to 1.

C Allowing for Mixed Strategy in the Bank's Risk-Taking

In the main text, we restricted attention to pure strategies by the bank; that is, the bank is assumed to choose whether to take high risks or low risks (with probability 1) after each (m, s). This appendix studies the extent to which this restriction limits the effectiveness of bank supervision. Specifically, we characterize the supervisor's optimal communication strategy assuming that the bank plays a particular mixed strategy and evaluate how the supervisor's (indirect) payoff depends on the bank's mixed strategy.

Preliminaries. Note that it suffices to consider the case where the supervisor's first-best outcome cannot be obtained. Therefore, we maintain the following assumption—the reverse of (4)—

throughout this appendix:

$$\int_{\underline{t}}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) > c \int_0^{\underline{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t). \tag{15}$$

We again consider the case where the supervisor uses a ternary signal $m \in \{hh, h\ell, \ell\ell\}$. However, we now assume that, following $m = h\ell$, the bank takes high risks with probability σ_s conditional on its own private signal $s \in \{g, b\}$; the case studied in the main text is a special case where $\sigma_g = 1$ and $\sigma_b = 0$. For the same reasons behind Lemma 3, the bank is always more willing to take high risks after $s = g$ than after $s = b$, so it suffices to focus on the case where either $\sigma_b = 0$ or $\sigma_g = 1$.

Suboptimality of the case where $\sigma_g \in (0, 1)$ and $\sigma_b = 0$. If the bank plays the mixed strategy $(\sigma_g, 0)$ after $m = h\ell$, the supervisor will approve the bank's high risk-taking if and only if $t \geq \underline{t}$ (as the bank takes high risks only when $s = g$). Importantly, this cutoff is independent of σ_g, which implies that the bank's incentives are also independent of σ_g. The supervisor's expected payoff as a function of σ_g is then

$$\int_{T^{h\ell} \cap [\underline{t}, 1]} \sigma_g \left[\gamma t - (1-\gamma)(1-t)d\right] \mathrm{d}F(t) + \int_{T^{hh} \cup [\widehat{t}, 1]} \left[t - (1-t)d\right] \mathrm{d}F(t).$$

Since $\gamma t - (1-\gamma)(1-t)d > 0$ for any $t > \underline{t}$, this payoff is increasing in σ_g; thus, it is never optimal for the supervisor to induce the bank to play a mixed strategy such that $\sigma_g < 1$ and $\sigma_b = 0$.

Optimal signal in the case where $\sigma_g = 1$ and $\sigma_b \in [0, 1]$. If the bank plays the mixed strategy $(1, \sigma_b)$ after $m = h\ell$, the supervisor will approve the bank's high risk-taking if and only if $t \geq \underline{t}(\sigma_b)$ where $\underline{t}(\sigma_b)$ is the value such that

$$\frac{\underline{t}(\sigma_b)}{1 - \underline{t}(\sigma_b)} = \frac{1 - \gamma + \gamma\sigma_b}{\gamma + (1-\gamma)\sigma_b} \frac{\widehat{t}}{1 - \widehat{t}}. \tag{16}$$

It is easy to see that $\underline{t}(\sigma_b)$ strictly increases from $\underline{t}$ to $\widehat{t}$ as σ_b rises from 0 to 1.

When the supervisor can obtain her conditionally first-best outcome. Conditional on $(\sigma_g, \sigma_b) = (1, \sigma_b)$, the supervisor's first-best outcome is that the bank takes high risks regardless of its signal s if $t \geq \bar{t}$ and with probabilities $(1, \sigma_b)$—depending on the bank's signal—if $t \in [\underline{t}(\sigma_b), \bar{t})$. Following the same logic as in Section 4.2, the supervisor can obtain this first-best outcome if and only if

$$\int_{\underline{t}(\sigma_b)}^{\bar{t}} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \leq c \int_0^{\underline{t}(\sigma_b)} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t). \tag{17}$$

Since $\underline{t}(\sigma_b)$ is increasing in σ_b, the left-hand side of (17) is decreasing, while the right-hand side is increasing, in σ_b. This implies that there exists $\overline{\sigma}_b \in (0,1]$ such that the supervisor obtains her first-best outcome conditional on $(1, \sigma_b)$ if and only if $\sigma_b \geq \overline{\sigma}_b$.

When the supervisor cannot obtain her conditionally first-best outcome. That is, suppose $\sigma_b < \overline{\sigma}_b$. Following the same steps as in Section 5, we can reduce the supervisor's problem to

$$\max_{\underline{t}^*, \bar{t}^*} \int_{\underline{t}^*}^{\bar{t}^*} [(\gamma + (1-\gamma)\sigma_b)t - (1-\gamma+\gamma\sigma_b)(1-t)d] \, \mathrm{d}F(t) + \int_{\bar{t}^*}^{1} [t - (1-t)d] \, \mathrm{d}F(t)$$

subject to

$$\int_{\underline{t}^*}^{\bar{t}^*} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \leq c \int_0^{\underline{t}(\sigma_b)} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t).$$

Since $\sigma_b < \overline{\sigma}_b$, this incentive constraint should be binding.

The supervisor's optimal strategy can be found as follows: for each $\lambda \geq 0$, let $\underline{t}^*(\lambda)$ be the value of t such that

$$(\gamma + (1-\gamma)\sigma_b)t - (1-\gamma+\gamma\sigma_b)(1-t)d - \lambda\left[t(1-\gamma) + (1-t)\gamma\right] = 0.$$

In addition, let $\bar{t}^*(\lambda)$ be the value of t such that

$$(\gamma + (1-\gamma)\sigma_b)t - (1-\gamma+\gamma\sigma_b)(1-t)d - \lambda\left[t(1-\gamma) + (1-t)\gamma\right] = t - (1-t)d.$$

Then, it suffices to find λ that satisfies

$$\int_{\underline{t}^*(\lambda)}^{\bar{t}^*(\lambda)} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) = c \int_0^{\underline{t}(\sigma_b)} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t). \tag{18}$$

We use the following lemma to analyze (18).

Lemma 7 *As* λ *increases from* 0 *to* $\overline{\lambda}$*,* $\underline{t}^*(\lambda)$ *rises from* $\underline{t}(\sigma_b)$ *to* $\widehat{t}$*, while* $\bar{t}^*(\lambda)$ *falls from* $\bar{t}$ *to* $\widehat{t}$*, where*

$$\overline{\lambda} := \frac{(2\gamma-1)(1-\sigma_b)\widehat{t}}{\gamma-(2\gamma-1)\widehat{t}}.$$

Proof. $\underline{t}^*(\lambda)$ and $\bar{t}^*(\lambda)$ can be explicitly solved as follows:

$$\begin{aligned}\underline{t}^*(\lambda) &= \frac{(1-\gamma+\gamma\sigma_b)d+\lambda\gamma}{\gamma+(1-\gamma)\sigma_b+(1-\gamma+\gamma\sigma_b)d+\lambda(2\gamma-1)}\\ \bar{t}^*(\lambda) &= \frac{\gamma(1-\sigma_b)d-\lambda\gamma}{(1-\gamma+\gamma d)(1-\sigma_b)-\lambda(2\gamma-1)}.\end{aligned}$$

All results follow from these explicit solutions, together with the definitions of $\underline{t}(\sigma_b)$, $\widehat{t}$, and $\bar{t}$. ■

By the same argument as in Section 5.2, there exists a unique positive value of λ that satisfies (18), and the corresponding cutoffs under the optimal communication strategy of the supervisor satisfy $\underline{t}(\sigma_b) < \underline{t}^* < \widehat{t} < \bar{t}^* < \bar{t}$ whenever $\sigma_b < \overline{\sigma}_b$. See the left panel of Figure 8 to see how these cutoffs are related to one another and depend on σ_b.

Welfare maximization over σ_b. We now analyze how the supervisor's expected payoff depends on σ_b, given that she adopts the optimal communication strategy for each σ_b. Let $W(\sigma_b)$ denote the supervisor's indirect expected payoff.

The above analysis has shown that there are two cases to consider, depending on whether $\sigma_b \geq \overline{\sigma}_b$ or not. The following result shows that $W(\sigma_b)$ reaches its maximum on the region where $\sigma_b \leq \overline{\sigma}$.

Lemma 8 *The supervisor's indirect payoff,* $W(\sigma_b)$*, is monotone-decreasing if* $\sigma_b \geq \overline{\sigma}_b$*.*

Proof. If $\sigma_b \geq \overline{\sigma}_b$, then (17) holds, so the supervisor obtains the following conditional ideal payoff:

$$W(\sigma_b) = \int_{\underline{t}(\sigma_b)}^{\bar{t}} \left[(\gamma+(1-\gamma)\sigma_b)t-(1-\gamma+\gamma\sigma_b)(1-t)d\right]\,\mathrm{d}F(t) + \int_{\bar{t}}^{1}\left[t-(1-t)d\right]\,\mathrm{d}F(t).$$

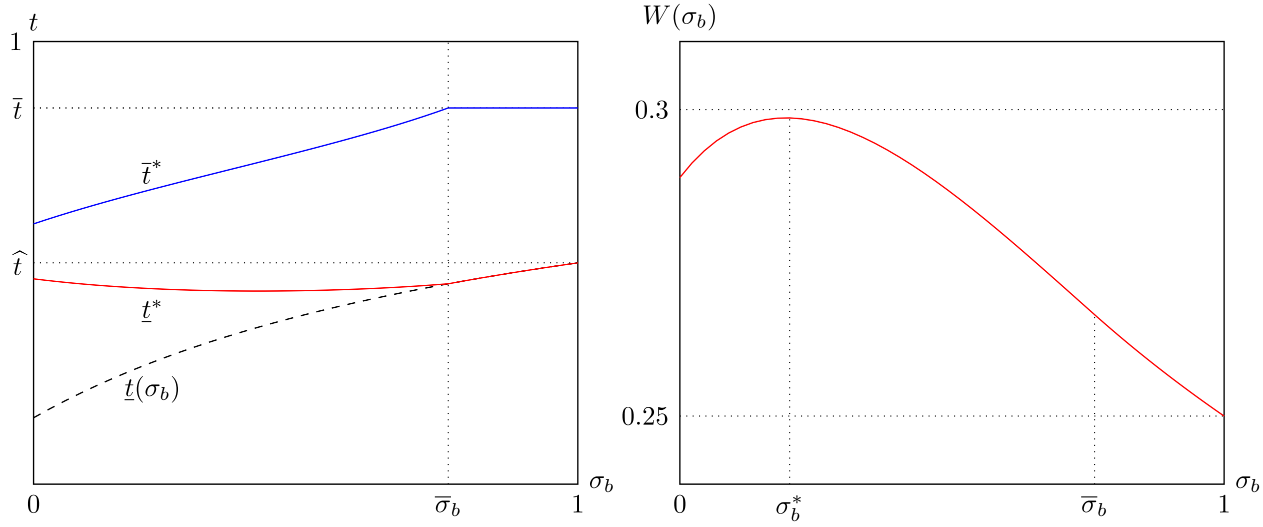


Figure 8: Optimal communication strategy of the supervisor (left) and the corresponding indirect expected payoff of the supervisor (right) as functions of σ_b when the supervisor's type t is uniform on $[0, 1]$ and $c = 1/2$, $d = 1$, $\gamma = 0.85$.

In this case,

$$\begin{aligned} W'(\sigma_b) = &\int_{\underline{t}(\sigma_b)}^{\overline{t}} \left[(1-\gamma)t - \gamma(1-t)d\right] \, \mathrm{d}F(t) \\ &- \left[(\gamma + (1-\gamma)\sigma_b)\underline{t}(\sigma_b) - (1-\gamma+\gamma\sigma_b)(1-\underline{t}(\sigma_b))d\right] \underline{t}'(\sigma_b) \\ = &\int_{\underline{t}(\sigma_b)}^{\overline{t}} \left[(1-\gamma)t - \gamma(1-t)d\right] \, \mathrm{d}F(t) < 0, \end{aligned}$$

where the second equality is due to (16) and the inequality is because $(1-\gamma)t - \gamma(1-t)d < 0$ whenever $t < \overline{t}$. ■

Intuitively, a higher σ_b means that the bank is more likely to take high risks even when its signal is bad ($s = b$). This reduces type I errors (i.e., low risk-taking when $\omega = G$) but increases type II errors (i.e., high risk-taking when $\omega = B$). In the relevant region $[\underline{t}(\sigma_b), \overline{t}] \subset [\underline{t}, \overline{t}]$, the latter (negative) effect always dominates the former (positive) one, so an increase of σ_b always lowers the supervisor's expected payoff.

As depicted in the right panel of Figure 8, $W(\sigma_b)$ is not monotone below $\overline{\sigma}_b$. To see why, let

λ^* denote the Lagrange multiplier corresponding to the optimal solution, which is strictly positive whenever $\sigma_b < \overline{\sigma_b}$. Then, by the envelope theorem,[29] we have

$$W'(\sigma_b) = \int_{\underline{t}^*}^{\bar{t}^*} [(1-\gamma)t - \gamma(1-t)d] \, \mathrm{d}F(t) + \lambda^* c \left[(1-\gamma)\underline{t}(\sigma_b) + \gamma(1-\underline{t}(\sigma_b))\right] \underline{t}'(\sigma_b).$$

The first term (integral) is negative, for the same reasons as in the proof of Lemma 8. However, the second (new) term is positive because $\lambda^*, t'(\sigma_b) > 0$. The presence of this second term makes the sign of $W'(\sigma_b)$ ambiguous in general.

Intuitively, as explained above, the direct effect of raising σ_b is negative as the relevant region $[\underline{t}^*, \bar{t}^*]$ is a subset of $[\underline{t}, \bar{t}]$ (where the supervisor wants the bank to take high risks only when $s = g$). But an increase of σ_b, by raising $\underline{t}(\sigma_b)$, also relaxes the constraint, thus enlarging the relevant region $[\underline{t}^*, \bar{t}^*]$ itself. The second term in the above expression for $W'(\sigma_b)$ captures this positive effect.

[29] See, e.g., Theorem A2.22 of Jehle and Reny (2010). Heuristically, because the constraint is binding, $W(\sigma_b)$ can be written in the following Lagrangian form:

$$\begin{aligned} W(\sigma_b) = &\int_{\underline{t}^*}^{\bar{t}^*} [(\gamma + (1-\gamma)\sigma_b)t - (1-\gamma+\gamma\sigma_b)(1-t)d] \, \mathrm{d}F(t) + \int_{\bar{t}^*}^{1} [t - (1-t)d] \, \mathrm{d}F(t) \\ &+ \lambda^* \left(c \int_0^{\underline{t}(\sigma_b)} [t(1-\gamma) + (1-t)\gamma] \, \mathrm{d}F(t) - \int_{\underline{t}^*}^{\bar{t}^*} [(1-\gamma)t + \gamma(1-t)] \, \mathrm{d}F(t) \right). \end{aligned}$$

Differentiating this (Lagrangian) function leads to the given expression for $W'(\sigma_b)$.