Algorithmic Transparency

(Job Market Paper)

Jian Sun*

[Click here for most recent version]

November 16, 2021

Abstract

I study the optimal algorithmic disclosure in a lending market where lenders use a predictive algorithm to mitigate adverse selection. The predictive algorithm is unobservable to borrowers and uses a manipulable borrower feature as input. A regulator maximizes market efficiency by disclosing information about the statistical properties of variables embedded in the predictive algorithm to borrowers. Under the optimal disclosure policy, the posterior belief consists of two disjoint regions in which the borrower feature is more relevant and less relevant in predicting borrower quality, respectively. The optimal disclosure policy differentiates posterior lending market equilibria by the equilibrium data manipulation levels. Equilibria with more data manipulation hurt market efficiency, but also discourage lenders' use of the borrower feature. Equilibria with less data manipulation benefit from that and generate more efficient market outcomes. Unconditionally, the borrower feature is used less intensively under optimal disclosure. This information design problem can be reduced to a one-dimensional maximization problem by imposing a mild distributional assumption on manipulation cost. As an extension, I also discuss the joint design of algorithmic disclosure and costly verification.

Keywords: adverse selection, FinTech, Bayesian persuasion, algorithmic transparency

JEL Codes: D82, G38, G32, D83

*Sloan School of Management, Massachusetts Institute of Technology, jiansun@mit.edu. I am very grateful to my committee members, Ian Ball, Hui Chen (chair), Andrey Malenko and Haoxiang Zhu for their invaluable support and continuous guidance. I would also like to thank Ben Bernanke, Taha Choukhmane, Marco Di Maggio, Daniel Greenwald, Peter Hansen, Olivia Kim, Marcus Opp, Jonathan Parker, Larry Schmidt, Antoinette Schoar, Yupeng Wang, Jiaheng Yu as well as seminar participants at MIT Sloan for helpful discussion and comments.

1 Introduction

Predictive algorithms have been widely used to mitigate adverse selection in various decision making processes, including hiring, college admission, and lending[1]. In these settings, decision makers use predictive models that establish links between variables that are relevant in their decision making problems. For example, employers score resumes to predict capability, schools use results of standardized tests to predict academic potential, and FinTech lenders use alternative data to predict credit quality. In these examples, the exact relationship between input and output is opaque to the public, and economic agents (such as job candidates, students, and borrowers) have little information about that. With the development of big data and data processing technology, predictive algorithms have become more complex and nonintuitive, involving variables that have no obvious relationship with each other, and thus become even more opaque. A popular argument to justify this opaque nature of predictive algorithms is the extent to which they can be manipulated by "gaming the system", that is, when economic agents know more about the predictive model, they are more likely to change their behavior strategically, which hurts the informativeness of the input. Despite the importance of this question, the effects of algorithmic transparency/opacity on market outcomes is still underexplored in academic research. Although some of the recent regulations start to consider this issue[2], the motivation usually comes from concerns about fairness, and largely ignores the effects on market efficiency. Moreover, due to the limited understanding of the consequences of algorithmic transparency, there is still uncertainty about future regulation[3], which may add another layer of inefficiency.

To better understand this question, this paper studies the optimal disclosure of a predictive algorithm that maximizes market efficiency in a FinTech lending setup. There are three types of players in this model: borrowers, lenders, and a regulator. Borrowers have private types, which is either good or bad. There is a borrower feature, such as phone usage behavior or social media activities, that can be observed by lenders but can also be manipulated by borrowers privately and at cost. Borrower feature is perfectly correlated with borrower type if not manipulated. Each borrower owns a borrower-specific project that needs to be financed by lenders. The required initial investment is the same for all borrowers, and the project payoffs are independent random variables that depend on borrower type. A

[1]See Bogen and Rieke (2018) for algorithmic hiring, Kizilcec and Lee (2020) for algorithmic fairness in education, Bruckner (2018) and Di Maggio et al. (2021) for algorithmic lending.

[2]"...company using algorithmic decision-making must know what data is used in its model and how that data is used to arrive at a decision and explain that to the consumer."—Federal Trade Commission;

"Whenever personal data is subject to automated decision making, people have ....the right to an explanation"— General Data Protection Regulation

[3]For example, in June 2021, NCRC, Affirm, Lending Club, Oportun, PayPal Holdings Inc, Square and Varo Bank asked the Consumer Financial Protection Bureau (CFPB) to provide guidance on how it will apply disparate impact rules to any systems that use artificial intelligence (AI), machine learning (ML), algorithms, or alternative data to make lending decisions.

predictive algorithm reveals the statistical properties of fundamental random variables in an economic environment. In this paper, the predictive algorithm is a mapping from borrower type to payoff distribution[4]. Specifically, when projects are financed, bad type borrowers always receive zero payoff and thus always default, and good type borrowers will receive i.i.d. nonnegative random payoffs with the same cumulative distribution function indexed by a one-dimensional parameter: *relevance*, denoted by ρ. When the relevance is higher(lower), the expected value of the random payoff from good type borrowers is higher (lower)[5]. Borrowers do not observe the true value of the relevance but share a common belief about it. Lenders observe the exact value of it, which partially determines how to use borrower data in lending decisions. This two-sided private information, i.e., one side (lenders) privately observes the statistical properties of fundamental random variables in the economic environment, and the other side (borrowers) privately manipulates their data, is novel in the literature and is the key feature of this model.

The lending market equilibrium consists of the manipulation behavior of borrowers and the use of borrower data by lenders in lending decisions. The regulator can establish a disclosure rule and ask lenders to disclose any possible information about the true state of relevance to borrowers. It is clear that the lending market equilibrium is determined by the updated public belief in the relevance, so the market outcome depends on the choice of disclosure policy. In this paper, I consider the optimal design of a disclosure policy that maximizes market efficiency.

In this model, I focus on the informational role of algorithmic disclosure, but not the role as a commitment device. From the disclosure, borrowers receive new information about the true value of relevance, which updates their belief about the usefulness of their data. Lenders, on the other hand, cannot commit to how to use borrower data in their lending decisions. This lack of commitment problem turns out to be the source of the inefficiency in this model. Because lenders always make the most efficient use of borrower data, this ex post efficient use of borrower data gives borrowers excessive ex ante manipulation incentives, which in turn makes the feature noisier and also makes market outcome less efficient from the unconditional perspective. The optimal disclosure policy mitigates this problem and generates lower levels of data manipulation unconditionally.

I model this optimal disclosure problem as a Bayesian persuasion problem (Kamenica and Gentzkow (2011)), and characterize the optimal public disclosure of the relevance. First, I show that it is suboptimal to disclose nothing. In this no disclosure equilibrium, borrowers' manipulation behavior and lenders' lending decisions are jointly determined by the public

[4]In practice, a predictive algorithm usually refers to a mapping from the observed data to the output but not the unobserved type to the output. In Section 3.7, I provide a discussion on the equivalence of these two views when algorithmic disclosure only plays an informational role but not serves as a commitment device, which is the feature of this model.

[5]In this paper, I use *relevance* and *relevance of the feature* interchangeably, because ρ measures how useful the feature is in lending decisions if there is no manipulation.

prior belief about the relevance of the feature. Since lenders always make efficient lending decisions ex post using all information available, there must exist scenarios where the surplus from using the feature in lending decisions is small but positive, and the lenders choose to use it ex post because it is efficient to do so. However, this possibility gives borrowers extra incentives to manipulate their features ex ante, and hurts efficiency in other scenarios. This cross-state externality through data manipulation makes lenders use the feature too intensively in their lending decisions. Second, it is also suboptimal to disclose everything, and I show that this full transparency policy leads to the worst outcome. This result relies on our assumption that there are sufficiently many bad type borrowers, so the adverse section problem is severe when there is no borrower data available. The intuition is that when borrowers know exactly how relevant their feature is, they will choose their manipulation behavior such that in equilibrium lenders are indifferent between using the feature or not, and thus there is zero surplus from lending market, resulting in the worst market outcome.

The optimal disclosure policy features partial disclosure, and differentiates the posterior lending market equilibria by their equilibrium data manipulation levels. The regulator can implement the optimal disclosure policy by assigning a score to the feature based on its relevance. Notably, the score function is not monotone in relevance, that is, features with higher relevance may be assigned with lower scores. Under each score realization, the posterior belief about the relevance consists of two disjoint intervals, denoted as approval region and rejection region, respectively. The true relevance is either in the approval region in which manipulating feature can help to get financing, or in a rejection region in which lenders do not use the feature at all and therefore manipulation is useless. The relative fraction of these two regions in the posterior belief determines all the equilibrium outcomes, including the data manipulation level and loan approval rate. Specifically, for all the posterior equilibria, when the feature is used in lending decisions with higher probability, there will be more data manipulation, and lenders will use the feature only when the relevance is high enough.

Several economic implications follow accordingly. First, unconditionally, the use of the feature in lending decisions is monotone in relevance, i.e., there exists a cutoff such that lenders will use the feature in their lending decisions if and only if the true relevance is above the cutoff. This property is obviously true for any posterior equilibrium (no matter what the posterior belief is), but I show that it also holds unconditionally under the optimal disclosure policy. Second, under the optimal policy, the unconditional probability that the lenders use the feature in their lending decisions is strictly lower than that under the no disclosure equilibrium. This confirms our intuition that lenders use the feature too intensively without disclosure. Thirdly, compared to the no disclosure equilibrium, the "worst" posterior equilibrium under the optimal policy induces more manipulation, while the "best" posterior equilibrium induces less manipulation. This uncovers the intuition of why the optimal disclosure policy improves efficiency: although the worse equilibria induce more manipulation and hurt market efficiency, they also force lenders to have a higher standard for the use of the feature in their lending decisions, and in turn deter borrowers' manipulation incentives.

Better equilibria benefit from this and induce less data manipulation. Unconditionally, there is less data manipulation, and the negative cross-state externality is mitigated.

With the general structure of the optimal policy, I also provide a closed-form characterization by imposing a mild distributional assumption on the borrowers' manipulation cost. In this case, the optimal score function consists of a discrete part, which induces the equilibrium with lowest level of data manipulation, and a continuous part with data manipulation level continuously moving from the lowest level to the highest level. Furthermore, for any posterior belief, the highest relevance in the rejection region is exactly the level at which the lenders break even. With this result, I can simplify the optimal disclosure problem to a one-dimensional optimization problem, and the optimal score function is solved by an ordinary differential equation (ODE).

I also consider an extension with costly verification. In practice, lenders can verify the types of borrowers by manually reviewing their profiles, conducting interviews, and using various fraud detection techniques. In this extension, lenders can reveal the true type of any borrower with a linear cost function. I explore how costly verification interacts with algorithmic disclosure under the optimal policy. It turns out that in the optimal joint design, these two channels work as substitutes: verification is used when the relevance of the feature is higher than a threshold and disclosure becomes irrelevant in this case; otherwise, disclosure will be used and there is no verification in equilibrium. The optimal joint design can be implemented by two steps: The regulator first reveals if the relevance of the feature is above or below the threshold. If it is above the threshold, all lenders will verify all applicants with a positive probability; otherwise, the regulator reveals additional information under the updated belief and no verification will be used.

The rest of this paper is organized as follows. In this section, I continue to discuss related literature. Section 2 provides a simple model, and Section 3 introduces the general model. In Section 4, I discuss the source of friction and the intuition of improving market outcome from the no disclosure equilibrium. In Section 5, I discuss the results of the general model. Section 6 studies an extension where costly verification is available, and Section 7 concludes.

Related Literature

This paper mainly contributes to three strands of the literature. First, there is a nascent but growing literature on the impact and regulation of algorithmic decision-making. Most of the existing research mainly focuses on fairness, bias, or discrimination (e.g. Bartlett et al. (2021), Milone (2019), Gillis and Spiess (2019),Raghavan et al. (2020),Coston et al. (2021)). This paper contributes to the literature by considering the regulation on algorithmic disclosure from the perspective of market efficiency. A closely related paper that asks a similar question and also considers strategic manipulation of data is Wang et al. (2020). They consider both the correlational and causal observables in their model and only consider the full transparency and no disclosure policies. Compared to their work, my model only considers correlational features as input in the predictive algorithm, and thus the fundamental frictions are different. Besides, I consider flexible disclosure policies, the results on the optimal design of algorithmic disclosure in my paper do not have a counterpart in their paper. Another theoretical paper that also focuses on algorithmic transparency and governance is Blattner et al. (2021). They consider the trade-off between model complexity and transparency and the role of algorithmic audit in regulating algorithms, which is different from my focus. Björkegren et al. (2020) examine the interplay between strategic manipulation of data and algorithmic transparency. Although the question in their paper is quite different from mine, the results in their field experiment empirically verify the existence of data manipulation when people know more about the algorithms. There is also a growing literature in computer science about algorithmic explainability or explainable AI (Bhatt et al. (2020), Carvalho et al. (2019), Lundberg and Lee (2017), Murdoch et al. (2019)). But the computer science literature usually focuses on algorithmic audit and explainability which mainly consider the "black box" nature of machine learning algorithms, while my paper simplifies this "black box" nature and considers an information design question in a finance setting.

I also add to a growing literature on Bayesian persuasion (Kamenica (2019) and Bergemann and Morris (2019) provide excellent surveys) and its applications in finance. The way I model information structure follows Kamenica and Gentzkow (2011), and I consider a persuasion problem with continuous state as in Dworczak and Martini (2019). Methodologically, Bayesian persuasion problems with continuous states are in general not tractable, except for some special cases (for example, Gentzkow and Kamenica (2016), Dworczak and Martini (2019), Goldstein and Leitner (2018)). In my model, the information designer's objective function depends on the entire distribution of posterior beliefs, and this question does not fit into any existing tractable framework. I obtain my theoretical results using a novel "guess and verify" method. There are many applications of Bayesian persuasion in finance literature, including shareholder voting (Malenko et al. (2021)), security design (Szydlowski (2021)), bank stress test (Goldstein and Leitner (2018), Goldstein and Leitner (2020) Inostroza (2019), Inostroza and Pavan (2021), Leitner and Williams (2020)) and financial network (Huang (2020)). This paper contributes to this literature by considering a new question (algorithmic disclosure) and provides a novel optimal signal structure.

Lastly, this project is related to the literature on strategic manipulation of data (Frankel and Kartik (2019a), Frankel and Kartik (2019b), Ball (2019)), or more broadly, the signaling models. The way I model private information on the borrower side is similar to Frankel and Kartik (2019b). Ball (2019) considers a problem with multi-dimensional features, and shows that the optimal scoring rule underweights some features to deter data manipulation. All of these papers focus mainly on how committing to certain decision rules will improve efficiency. Relative to these work, I consider an information design question, and focus on how commitment on information structure (disclosure policy) will improve efficiency.

2 A Simple Model

To fix ideas, let's consider a simple model. There is a competitive lending market with many identical lenders and a unit mass of borrowers. Each borrower i is endowed with zero initial wealth and a borrower-specific project which requires an initial investment I at time 0. The project generates a positive cash flow V if it succeeds, and zero if it fails. The probability of success is a random variable, and its distribution is formally introduced later. All the borrowers can get private benefit b if their own borrower-specific project is successfully financed regardless of the outcome.

There is a manipulable feature for each borrower, which takes two possible levels: High or Low. Borrowers who are born with High (Low) feature make up μ $(1-\mu)$ of the entire population, and they are called good (bad) type borrowers. Manipulating feature is possible for bad borrowers[6], and they can privately change their feature to High by paying a cost c. The manipulation cost c follows a uniform distribution on $[0,1]$ among the bad borrowers. A key assumption here is that manipulation behavior does not change borrower type. In this lending market, the only data that lenders can collect and observe is the the borrower feature after potential manipulation.

Probability of success depends on borrower type (good or bad). Specifically, bad borrowers always fail. For good borrowers, the probability of success ρ is drawn from a uniform distribution on $[0,1]$. The true value of ρ is only observable to lenders, and all borrowers share the common prior about the distribution of ρ. In this example, we call ρ the *relevance* of the borrower feature, because it represents how relevant the borrower feature is in lending decisions when there is no manipulation. Besides, from the perspective of borrowers, the probability of success ρ is the only uncertain element in the mapping from borrower type to payoff distribution. In this example, let's impose the following assumption[7].

Assumption 1. *(Severe Adverse Selection)* $b \geq 1$ *and* $\mu V \leq I$.

Note that the manipulation cost follows $c \sim U[0,1]$, so $b \geq 1$ implies that if lenders lend to borrowers with feature High with probability 1, the private benefit always dominates the manipulation cost for all bad borrowers, and thus all of them will choose to manipulate their features. And the second condition $\mu V \leq I$ implies that lenders will never lend to any borrower if all bad type borrowers choose to manipulate their features. These two conditions jointly imply that in any equilibrium, not all bad borrowers choose to manipulate their features.

[6]Here we only allow bad borrowers to manipulate their features for simplicity of exposition. But this assumption is not necessary. Even if we assume that good type borrowers can costly manipulate their features, they will never do this in equilibrium.

[7]We'll have a similar assumption in the main model

No Disclosure On The Relevance ρ

First we consider the equilibrium when no additional information about ρ is disclosed to borrowers. Lenders can make their lending decisions based on the observed feature. The lending market is competitive, so lenders always make zero profit in equilibrium. In this case, it can be shown that there is a unique equilibrium which consists of two cutoffs $\bar{c}_N$ and ϱ_N, such that

- all bad type borrowers with manipulation cost lower than $\bar{c}_N$ choose to manipulate their features to High;
- lenders will lend to borrowers with feature High if $\rho > \varrho_N$.

For bad type borrowers with manipulation cost $\bar{c}_N$, the indifference condition is

$$\text{Prob}\left(\rho > \varrho_N\right) \cdot b = \bar{c}_N,$$

where $\text{Prob}\left(\rho > \varrho_N\right)$ is the probability that the loan application is approved for borrowers with feature High. For lenders, the total surplus from lending is zero when $\rho = \varrho_N$, implying

$$\mu \varrho_N V = \left(\mu + (1-\mu)\text{Prob}\left(c \leq \bar{c}_N\right)\right) I.$$

Based on our distributional assumptions, the unique solution of the equilibrium is

$$\left(\varrho_N = \frac{I}{\mu V + (1-\mu) I}, \bar{c}_N = b \cdot \frac{\mu(V-I)}{\mu V + (1-\mu) I}\right).$$

Let

$$k_N = \left(\mu + (1-\mu)\text{Prob}\left(c \leq \bar{c}_N\right)\right) I$$

be the effective financing cost, then the lending market surplus (measured by the net value of all projects financed) is

$$W_N = \int_{\varrho_N}^{1} \left(\mu \rho V - k_N\right) d\rho.$$

For simplicity, let's take the following parameters:

$$I = 3, V = 10, b = 1, \mu = 3/10,$$

then the equilibrium variables are

$$\left(\varrho_N = 0.59, \bar{c}_N = 0.41, W_N = 0.25\right). \tag{1}$$

Figure 1 summarizes the above equilibrium. The green triangle in Figure 1 is the surplus W_N. In equilibrium, the expected payoff outweighs the cost only when $\rho > 0.59$, and borrowers with High feature are financed only when ρ is in this region. The green line on the horizontal axis represents the support of posterior belief. In this no disclosure equilibrium, the posterior belief is the same as the prior belief, and thus the support of the posterior belief is the interval $[0, 1]$.

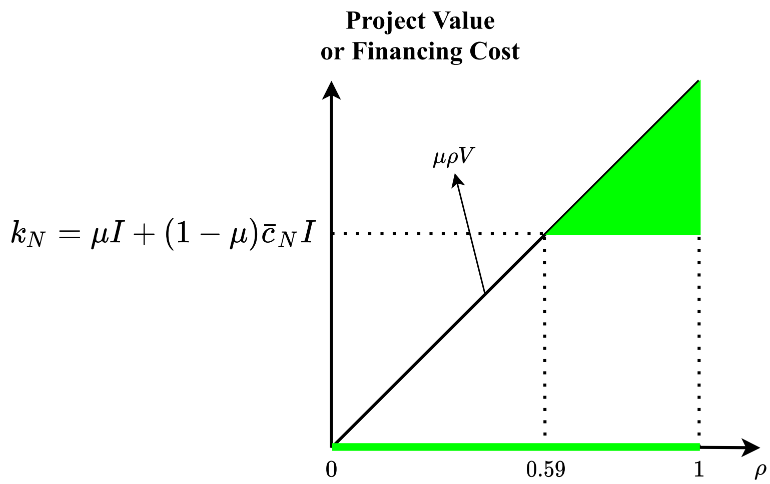

Figure 1: Equilibrium–No Disclosure

Full Transparency

Another natural disclosure policy is full transparency which reveals the true state of ρ perfectly to borrowers. It turns out that the surplus equals to zero in this case, which leads to the worst market outcome. To see this, suppose the true relevance ρ satisfies $\rho < \frac{I}{V}$, then even only lending to good type borrowers is inefficient, and thus there will be no financing and the market outcome must be zero. For any $\rho \geq \frac{I}{V}$, we know in equilibrium, the probability that lenders lend to borrowers with feature High must be less than 1, which means that they must be indifferent between lending and not lending. Then the surplus also must be zero for any $\rho \geq \frac{I}{V}$.

A Binary Color Signal

Our question is, can a regulator achieve a strictly higher outcome by designing a signal about ρ and disclosing it to the market? The answer is yes. The definition of disclosure policy is formally introduced in Section 3.4, here let's take the numbers from the no disclosure example and consider the following specific score function which consists of two levels $R(ed)$ and $B(lue)$:

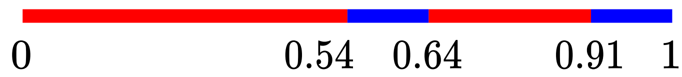

Figure 2: The Color Signal

$$s = \begin{cases} R(ed) & \text{if } \rho \in [0, 0.54) \cup (0.64, 0.91) \\ B(lue) & \text{if } \rho \in [0.54, 0.64] \cup [0.91, 1] \end{cases}. \tag{2}$$

This score function assigns colors to state of ρ, which is represented by Figure 2. It only reveals which region (Red or Blue) that the true state of ρ belongs to, and induces two possible posterior equilibria.

Specifically, if the signal realization is R, then the posterior belief about ρ is a uniform distribution on two disjoint intervals $[0, 0.54) \cup (0.64, 0.91)$, and it can be shown that the equilibrium outcomes are

$$\left(\bar{c}_R = 0.34, \underline{\rho}_R = 0.54, W_R = 0.19\right). \tag{3}$$

Similarly, if the signal realization is B, the posterior belief about ρ is a uniform distribution on $[0.54, 0.64] \cup [0.91, 1]$, and the equilibrium outcomes are

$$\left(\bar{c}_B = 0.48, \underline{\rho}_B = 0.64, W_B = 0.09\right). \tag{4}$$

Figure 3 summarizes the surpluses of these two equilibria. The red trapezoid in the left graph represents the surplus on observing R, and the two red intervals on the horizontal line represent the support of the posterior belief. In this equilibrium, lenders will lend to borrowers with feature High only when $\rho \in (0.64, 0.91)$. Similarly, the right graph in Figure 3 shows the surplus on observing B. The total surplus with this color signal (2) is

$$W_s = W_R + W_B = 0.19 + 0.09 = 0.28 > 0.25 = W_N.$$

So the surplus improves.

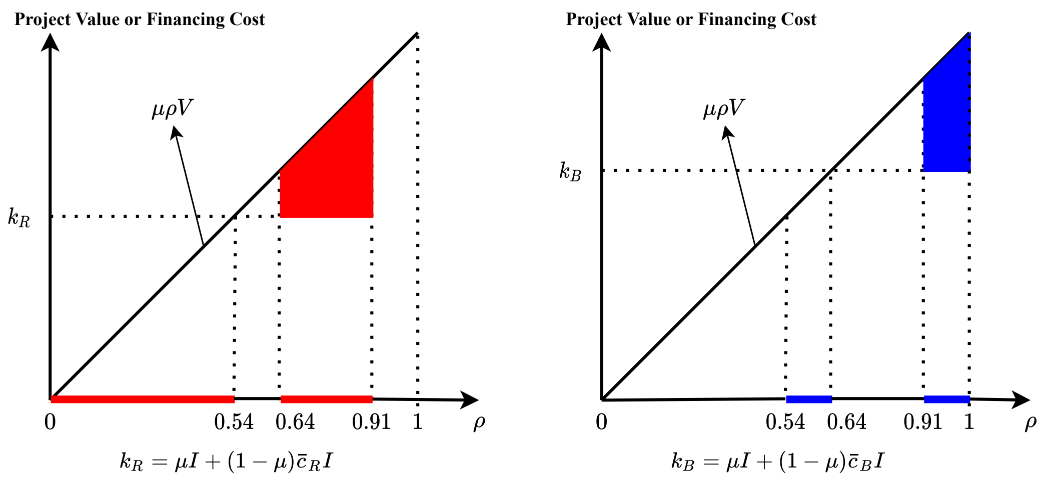

Figure 3: Equilibrium–Score R and B

Our analysis shows that the binary color signal dominates both the no disclosure policy and full transparency policy. But what is the intuition behind this result? The result that full transparency policy is dominated is clear: when the exact information about the relevance of the feature is disclosed to the market, bad type borrowers will manipulate their features such that in equilibrium the surplus from using the borrower data in lending decisions is always zero, and all lenders are indifferent between using and not using the borrower data in their lending decisions. The inefficiency embedded in the no disclosure equilibrium is the lenders' lack of commitment problem, that is, lenders always make the most efficient use of borrower data ex post in their lending decisions. To see this, suppose in the no disclosure equilibrium, the regulator is able to "force" the lenders to use the feature in their lending decisions only when ρ is greater than an exogenous cutoff $\varrho_x = \varrho_N + x$, where $x \ll 1$. Then the bad type borrowers with manipulation cost

$$c \leq \bar{c}_x = b \cdot \text{Prob}\left(\rho \geq \varrho_x\right)$$

will choose to manipulate their features, and the total surplus is a function of the exogenous cutoff ϱ_x:

$$W\left(x\right) = \int_{\varrho_x}^{1} \left[\mu\rho V - \left(\mu + (1-\mu)(1-\varrho_x)\right) I\right] d\rho. \tag{5}$$

Note $W\left(0\right) = W_N$, then

$$
\begin{aligned}
\left.\frac{dW(x)}{dx}\right|_{x=0} &= -\left[\mu \varrho_N V - \left(\mu + (1-\mu)(1-\varrho_N)\right) I\right] + \int_{\varrho_N}^{1} (1-\mu) I d\rho \\
&= \int_{\varrho_N}^{1} (1-\mu) I d\rho > 0.
\end{aligned} \tag{6}
$$

This is because the lender's equilibrium condition in the no disclosure equlibrium is

$$
\left[\mu \varrho_N V - \left(\mu + (1-\mu)(1-\varrho_N)\right) I\right] = 0. \tag{7}
$$

The result in (6) shows that the equilibrium cutoff ϱ_N is inefficiently low from the ex ante perspective. So the probability that the borrower data is used in lending decisions, $(1-\varrho_N)$, is inefficiently high. This result is based on condition (7),which is the ex post efficient use of borrower data in lending decisions. In equilibrium, when lenders use borrower data more often in some states ex post, more bad borrowers will choose to manipulate their features ex ante, and the effective financing cost will increase for all other states from the ex ante perspective. This cross-state externality makes no disclosure equilibrium inefficient. In the first order derivative (6), when lenders increase their lending cutoff by x, the approval probability decreases by

$$
\text{Prob}(\rho > \varrho_N) - \text{Prob}(\rho > \varrho_x) = x
$$

from the perspective of bad type borrowers, then the fraction of bad type borrowers who would like to manipulate decreases by

$$
\text{Prob}(c \leq \bar{c}_N) - \text{Prob}(c \leq \bar{c}_x) = x,
$$

implying that the effective financing cost decreases by

$$
(1-\mu) I \cdot x.
$$

Then the total cost saving from all states $\rho > \varrho_N$ is

$$
\int_{\varrho_N}^{1} (1-\mu) I d\rho \cdot x
$$

which corresponds to the last term in (6).

To mitigate the excess manipulation, the signal (2) defers lenders' use of borrower data by differentiating the two lending market equilibria by data manipulation levels. To see this, upon observing B, in equilibrium we have

$$
\bar{c}_B = 0.48 > \bar{c}_N = 0.41,
$$

which means there are more bad type borrowers manipulating their features compared to the no disclosure equilibrium. As a result, lenders have a more stringent lending standard, and lend to borrowers with feature High when $\rho > \underline{\rho}_B = 0.64$, which is greater than the cutoff in the no disclosure equilibrium ($\underline{\rho}_N = 0.59$). Then the lenders will not use borrower data in their lending decisions when $\rho \in [0.59, 0.64]$ under signal B. But note this is the region when lenders lend to borrowers with feature High in the no disclosure equilibrium. For $s = R$, there is less manipulation compared to the no disclosure equilibrium because $\bar{c}_R = 0.34 < 0.41 = \bar{c}_N$, and borrower data is used by lenders only when $\rho \in (0.64, 0.91)$.

Unconditionally, with the binary color signal, the feature is used when $\rho > 0.64$, while it is $\rho > \rho_N = 0.59$ in the no disclosure equilibrium. So the feature is used less frequently with the binary color signal. Intuitively, by differentiating the two equilibria by data manipulation levels, the "worse" equilibrium ($s = B$) effectively guarantees that the feature will not be used in cases when it was indeed used in no disclosure equilibrium, and the "better" equilibrium ($s = R$) has lower level of data manipulation and generates more efficient outcome.

Actually this binary color signal is optimal among all binary signals. In the main model, I'll consider a general space of disclosure policies. But this binary color signal has several notable properties that are still robust in the optimal disclosure policy in the main model. First, there exists a threshold ($\underline{\rho}^\star = 0.64$), such that unconditionally, the feature is used in lending decisions if and only if the true state is above the threshold. It is clear that this cutoff property always holds for any posterior equilibria, and here I show it also holds unconditionally. The intuition is clear: the relevance ρ represents how useful borrower feature is in lending decisions. When ρ is higher, borrowers with feature High are of better qualities and will have higher probability of success. Then it is efficient to lend to borrowers with feature High when the true relevance ρ is higher. Second, for any score realization (R or B), the support of posterior belief is always a union of two disjoint intervals. These two intervals correspond to lenders' equilibrium lending decisions. The interval below $\underline{\rho}^\star$ represents the rejection region, and the lenders will reject all borrowers when the true relevance is in this region; while the interval above $\underline{\rho}^\star$ represents the approval region and lenders will lend to all borrowers with feature High when the true relevance is in this region. Thirdly, the unconditional probability of using the feature in lending decisions is less than that in the no disclosure equilibrium, implying that the feature is used less intensively with optimal disclosure. Lastly, the binary color signal induces two posterior equilibria, with one equilibrium (B) having a higher data manipulation level than the no disclosure equilibrium, and the other one (R) having a lower data manipulation level. All of these properties still hold in the optimal disclosure policy in the main model.

3 The Main Model

The main model is a generalization of the simple model.

3.1 Players

There are three types of players in this model: a unit mass of borrowers, $N(>1)$ lenders, and a regulator. All players are risk neutral. We model borrowers in a similar way as the agents in Frankel and Kartik (2019b). Borrowers have two-dimensional private information: **quality** type $\theta \in \{G(ood), B(ad)\}$, and (manipulation) **cost** type c. For the joint distribution of (θ, c), I assume the unconditional probability of good type borrowers in the population is

$$\text{Prob}\,(\theta = G) = \mu > 0.$$

And the conditional probability $c|\theta$ is

$$c|\theta \begin{cases} \equiv \infty & \text{if } \theta = G \\ \sim F_c(\cdot) & \text{if } \theta = B \end{cases}, \tag{8}$$

where $F_c(\cdot)$ is the cumulative distribution function for a continuous random variable defined on $[0, \bar{c}]$[8]. Assume $F_c'(x) > 0$ and $F_c''(x)$ is bounded for all $x \in [0, \bar{c}]$

All lenders are identical and operate in a competitive lending market. At time 0, each borrower i receives a borrower-specific project (project i) and has zero initial wealth. Each project i requires an initial investment I; otherwise it fails, and is liquidated with zero liquidation value. If project i is financed at time 0, it will generate a nonnegative random payoff $\tilde{V}_i$[9] at $t = 1$, and the realization of the random payoff is publicly observable. Besides, borrower i also receives a constant nontransferable private benefit b if the project is successfully financed. Any project can be financed by at most one lender.

If the project i is financed by lender j with debt face value D_j^i, then when the payoff $\tilde{V}_i$ is realized, borrower i receives

$$\max\left\{\tilde{V}_i - D_j^i, 0\right\} + b,$$

lender j receives

$$\min\left\{\tilde{V}_i, D_j^i\right\} - I,$$

and the regulator's payoff (surplus) is the total outcome of lending market, which is[10]

$$\tilde{V}_i - I.$$

[8]Here I assume $c|\theta = G \equiv \infty$ for simplicity of exposition. Actually $c|\theta = G$ is irrelevant for all of my results. For example, we can assume $c|\theta \sim F_c(\cdot)$ for both $\theta \in \{G, B\}$, and all the results will be the same.

[9]For expositional convenience, we sometimes use $\tilde{V}$ to represent the random payoff for an arbitrary borrower.

[10]Note that the private benefit is not included in the regulator's utility, but this assumption is not crucial. Actually the key result, that partial disclosure is optimal, is still robust even if we include private benefit in the regulator's payoff function.

3.2 Predictive Algorithm

For each borrower i, his quality type θ_i is informative about his random payoff $\tilde{V}_i$. Specifically, when $\theta_i = B$, $\tilde{V}_i \equiv 0$, i.e., bad type borrowers always fail. When $\theta_i = G$, $\tilde{V}_i$ is a nonnegative, continuous random variable on $\left[0, \bar{V}\right]$, with cumulative distribution function $F(\cdot)$. The key feature of our model is that $F(\cdot)$ is drawn from a family of distribution functions $\{F_\rho(\cdot)\}_{\rho \in \mathcal{P}}$, where $\mathcal{P}$ is a subset of $\mathbb{R}$. Intuitively, since bad type borrowers always fail, ρ effectively measures how the quality type θ can be used to predict payoff distribution. Throughout the paper, I call ρ the *relevance*. In practice, machine learning algorithms adopted by FinTech lenders are hard to explain and interpret and can rarely be summarized by a one-dimensional parameter. In this paper, since I focus on disclosure instead of explainability (which is the primary focus of the computer science literature, see Lundberg and Lee (2017)), I abstract away the "black box" feature of the predictive algorithms and assume them to be summarized by a one-dimensional parameter ρ.

The relevance ρ is drawn from a continuous distribution with cumulative (probability) distribution function $\Pi_0(\rho)$ $(\pi_0(\rho))$. Without loss of generality, we assume ρ is drawn from a uniform distribution in $[0, 1]$[11], so $\mathcal{P} = [0, 1]$. The key assumption of our model is that ρ is only observable to all lenders but not borrowers, and we assume all borrowers share the common prior belief about the distribution of ρ.

Let

$$m(\rho) = \mathbb{E}\left(\tilde{V}|G, \rho\right) = \int_0^{\bar{V}} v \cdot dF_\rho(v)$$

be the expected payoff from any good type borrower if he is successfully financed. Then we impose the following assumptions on $m(\rho)$:

Assumption 2. *$m(\rho)$ satisfies the following conditions:*

1. *$m(\rho)$ is continuous and strictly increasing;*
2. *$m(0) = 0$ and $m(1) > I$;*
3. *$\mu m(1) \leq I$.*

The first assumption is mainly for expositional convenience; relaxing this assumption does not affect our main results. In the second assumption, $m(0) = 0$ is also mainly for expositional convenience so we can relax it without changing the main results. $m(1) > I$ is to make sure that when $\rho = 1$, it is efficient to lend to good type borrowers, otherwise it is always efficient to reject any borrower and the equilibrium becomes trivial. The last assumption means that the adverse selection in the market is severe and it is inefficient to lend to all borrowers, this assumption helps to establish a clear benchmark, but my main results do not rely on this specific assumption.

[11] Note that for any continuous random variable x with cumulative distribution function $T(\cdot)$, the new variable $y = T(x)$ always follows a uniform distribution on $[0, 1]$.

3.3 Feature and Manipulation

Although the quality type is informative about borrowers' riskness, it is the private information of borrowers, and thus can not be directly used by lenders in their lending decisions. There is a feature $\hat{\theta} \in \left\{\hat{G}, \hat{B}\right\}$ for each borrower and can be publicly observed by lenders. If borrowers do not manipulate their features, $\hat{\theta} = \hat{G}(\hat{B})$ if $\theta = G(B)$, i.e., borrower feature $\hat{\theta}$ can perfectly reveal borrower type θ. However, each borrower can change his feature to the other value by privately paying the non-pecuniary manipulation cost c. The cost structure is introduced in (8). Intuitively, good type borrowers are not able to manipulate their features, while bad type borrowers can manipulate their features by paying cost c, which follows a continuous distribution on $[0, \bar{c}]$ with cumulative distribution function $F_c(\cdot)$. The assumption that good type borrowers are not able to manipulate their features is actually redundant. We can show that good type borrowers will never manipulate in equilibrium even if they have finite manipulation cost (see Appendix A).

In equilibrium, lenders use the feature $\hat{\theta}$ to assess borrowers' riskness, but the informativeness of the feature $\hat{\theta}$ is determined by bad type borrowers' manipulation behavior. Lenders' lending decisions and bad type borrowers' data manipulation levels are jointly determined in equilibrium.

A key assumption in this model is that manipulation behavior does not change borrower type, i.e., the distribution of $\tilde{V}$ is not influenced by the manipulation behavior, so feature $\hat{\theta}$ only plays an informational role. This assumption is motivated by the "gaming the system" concern in the algorithmic transparency debate. For example, lenders find variables that can predict default risk using historical training data and machine learning algorithms, which focus more on correlation but not causation between input and output. If borrowers strategically change their behavior, their true riskness does not change but the predictive algorithm may become less effective.

I impose the following assumption which also shows up in Assumption 1 in the simple model.

Assumption 3. $b \geq \bar{c}$.

This assumption implies that if lenders lend to $\hat{G}$ borrowers for sure, then all of the bad type borrowers will choose to manipulate and the borrower feature becomes useless. This assumption, together with the condition $\mu m(1) \leq I$ in Assumption 2, jointly imply that in any equilibrium not all bad type borrowers choose to manipulate their features. This result that helps to characterize the optimal policy, but my main results can easily be extended to the case when this condition is violated.

3.4 Disclosure Policy

This project primarily focuses on the public disclosure of the relevance ρ. Although its realization is unobservable to borrowers, we consider the scenario in which the regulator can publicly reveal some information about the true state of ρ to all borrowers before they choose their manipulation behavior. Below is the formal definition of a disclosure policy.

Definition 3.1. A disclosure policy $(\mathcal{S}, \tilde{\sigma})$ consists of a signal space $\mathcal{S}$ and a mapping $\tilde{\sigma}$ from the realization $\rho \in \mathcal{P} = [0,1]$ to a distribution over signal space $\mathcal{S}$:

$$\tilde{\sigma}\left(s|\rho\right) : [0,1] \to \Delta\left(\mathcal{S}\right).$$

So $\tilde{\sigma}\left(s|\rho\right)$ is the (generalized) probability distribution function[12] of s conditional on state ρ. The regulator publicly announces the disclosure rule and then draws a realization of s based on it. After observing the realization s, all borrowers can update their beliefs on the distribution of ρ by Bayesian updating and then choose their manipulation strategies.

A special case of policies defined in Definition 3.1 is the deterministic policy. For these policies, the signal realization conditional on any state ρ is deterministic, so the conditional probability can be summarized by a deterministic function. Below is the definition of a deterministic policy. For notational simplicity, let's denote $\delta\left(x\right)$ as the Dirac function[13].

Definition 3.2. A disclosure policy $(\mathcal{S}, \tilde{\sigma})$ is deterministic if for any $\rho \in [0,1]$, the signal realization is deterministic, i.e., there exists a **message function**

$$\sigma : [0,1] \to \mathcal{S},$$

such that

$$\tilde{\sigma}\left(s|\rho\right) = \delta\left(s - \sigma\left(\rho\right)\right).$$

Throughout this paper, when there is no confusion, we use $(\mathcal{S}, \sigma)$ to represent a deterministic disclosure policy with signal space $\mathcal{S}$ and message function σ. To gain more intuitions on how disclosure policies work, note that the full transparency can be implemented by a deterministic policy with signal space $\mathcal{S} = [0,1]$, and the message function σ is

$$\sigma\left(\rho\right) = \rho.$$

[12] See Ziółkowski (2009) for discussion on generalized probability distribution function.

[13] A Dirac function $\delta\left(x\right)$ is defined as

$$\delta\left(x\right) = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{if } x \neq 0 \end{cases},$$

and $\int_{-\infty}^{+\infty} \delta\left(x\right) dx = 1$.

In this case, the regulator assigns a unique signal $s = \rho$ to each state ρ. When borrowers observe a realization s, the public belief will be updated and it is sure that the true state of relevance ρ is $\rho = s$. This disclosure policy effectively reveals all information about the true state of ρ. Another example is the no disclosure policy, i.e., the regulator does not reveal any information. It can also be implemented by a deterministic policy with only one element in the signal space. Then borrowers will always observe the same realization no matter what the true state of relevance ρ is, and thus they will learn nothing from the signal and no information is revealed by this disclosure policy.

A more complex but also commonly used signal structure is the cutoff disclosure, i.e, the regulator only reveals whether the true state of relevance ρ is above a threshold $\hat{\rho}$ or not. In this case, the disclosure can be implemented by a deterministic policy with two elements in the signal space $\mathcal{S} = \{s_L, s_H\}$, and the message function $\sigma(\rho)$ is

$$\sigma(\rho) = \begin{cases} s_L & \text{if } \rho \in [0, \hat{\rho}] \\ s_H & \text{if } \rho \in (\hat{\rho}, 1] \end{cases}.$$

So the regulator discloses s_H if $\rho > \hat{\rho}$ and s_L otherwise. Then borrowers can only learn if the true state of relevance ρ is above the threshold $\hat{\rho}$ or not.

The main advantage of modeling information disclosure in this way is the flexibility. Intuitively, the information structure defined in Definition 3.1 summarizes all possible ways of disclosing information, which also sheds light on the boundary of the pure information channel on mitigating manipulation in this problem.

Once we have a general signal structure $(\mathcal{S}, \tilde{\sigma})$, it will induce a distribution of posterior beliefs $\{f(s), \pi(\rho|s)\}_{s\in\mathcal{S}}$, where$f(s)$ is the (generalized) density function of the random variable s, and $\pi(\rho|s)$ is the public posterior belief (probability distribution function) of ρ conditional on observing the public signal realization s. To get sharp predictions on the optimal disclosure policy, we impose the following technical restriction on the posterior beliefs[14]:

Criterion 1. *We focus on disclosure policies such that for any s, and any $\rho \in \operatorname{supp}(\pi(\rho|s))$, there exists a closed set $B_{s.\rho} \subset \operatorname{supp}(\pi(\rho|s))$, such that $\rho \in B_{s,\rho}$ and $\mathbb{E}\left(\mathbb{1}_{B_{s,\rho}}(\rho)\,|s\right) > 0$, where $\mathbb{1}_A(x)$ is the indicator function:*

$$\mathbb{1}_A(x) = \begin{cases} 1 & \textit{if } x \in A, \\ 0 & \textit{if } x \notin A. \end{cases}$$

[14]This technical restriction is not important. In our model, any zero-measure change on the disclosure policy doesn't change the expected payoff. This restriction is to rule out some optimal policies that are almost the same as the optimal policies we characterize later.

The distribution of posteriors $\{\pi(\rho|s)\}_{s\in\mathcal{S}}$ must satisfy a necessary condition[15]:

$$\int_s \pi(\rho|s) f(s)\, ds = \mathbb{1}_{[0,1]}(\rho). \tag{9}$$

This is also known as the Bayes-plausible condition (Kamenica and Gentzkow (2011)). The interpretation is that the average of all posterior beliefs must be consistent with the prior belief. Then following the literature, instead of working with the signal structures directly, I work with distributions of posterior beliefs that satisfy condition (9) and [16]

$$\int_s f(s)\, ds = 1.$$

3.5 Lending Market

There are $N > 1$ identical lenders operating in a competitive lending market, then in equilibrium all lenders make zero profit[17]. The lending market equilibrium consists of the bad type borrowers' manipulation strategies and lenders' lending decisions. We only focus on symmetric equilibria in which all lenders choose the same strategy in equilibrium.

When the regulator commits to a disclosure policy $(\mathcal{S}, \tilde{\sigma})$, for each signal realization $s \in \mathcal{S}$, we call the lending market induced by this signal realization s the *subgame* s. Before exploring how $(\mathcal{S}, \tilde{\sigma})$ will change the market outcome, we solve the model backwards and first consider the lending market equilibrium under an arbitrary posterior belief.

Suppose updated public belief of ρ is $\pi(\rho|s)$. For any borrower i and lender j, let $\left(I_j^{s,\hat{\theta}}, D_j^{s,\hat{\theta}}\right)$ be the lender's strategy and γ_i^s be the borrower's manipulation decision where

1. $I_j^{s,\hat{\theta}} \in [0,1]$ represents the probability that lender j approves the loan application from $\hat{\theta}$ borrowers, and $D_j^{s,\hat{\theta}}$ represents the face value of the debt conditional on approval;

2. $\gamma_i^s \in \{0,1\}$ represents the probability that borrower i manipulates his feature $\hat{\theta}_i$.

It's clear that $\gamma_i^s = 0$ for all good type borrowers in any equilibrium because they have infinite manipulation cost, so the good type borrowers are passive in our model and do not play any strategic role. For all lenders, they'll never lend to $\hat{B}$ borrowers, as those must be

[15] The RHS of the condition represents the density of the prior belief of ρ, which is 1 under the uniform distribution on $[0,1]$.

[16] These two conditions are necessary conditions, and I'll verify the existence of the optimal policy later.

[17] This is not the key assumption of our model. Actually we can consider a model with a monopoly lender, and the results on the optimal disclosure polices are the same, as long as there is no screening by contracts. This is because in this model, the regulator wants to maximize the total surplus from financing activities, while market structure only changes the distribution of surplus but not the total surplus.

bad type borrowers who will default with probability 1. So we must have $I_j^{s,\hat{B}} = 0$, and the choice of $D_j^{s,\hat{B}}$ becomes irrelevant.

Let k_s be the total (effective) financing cost of lending to $\hat{G}$ borrowers, since all projects require the same initial investment, $\frac{k_s}{I}$ is the measure of $\hat{G}$ borrowers and $\left(\frac{k_s}{I} - \mu\right)$ is the measure of bad type borrowers who choose to manipulate their features. For a lender j, she lends to $\hat{G}$ borrowers only if

$$\mu m(\rho) - k_s \geq 0.$$

Then any lender j's lending decision can be summarized by[18]

$$I_j^{s,\hat{G}}(\rho) \begin{cases} = 1 & \text{if } \rho > \varrho_s \\ \in [0,1] & \text{if } \rho = \varrho_s \\ = 0 & \text{if } \rho < \varrho_s \end{cases},$$

where

$$\varrho_s = m^{-1}\left(\frac{k_s}{\mu}\right). \tag{10}$$

Since we only focus on symmetric equilibrium, in equilibrium we must have $\mathbb{E}\left(I_j^{s,\hat{G}}\right) = \mathbb{E}\left(I_k^{s,\hat{G}}\right)$ for any j, k. For simplicity, let I^s represent the equilibrium approval decision under signal s.

For a bad type borrower with private manipulation cost c_i, since he always fails, the only benefit he may get from the deal is the private benefit B, so he chooses to manipulate only if

$$\mathbb{E}(I^s) \cdot B \geq c_i.$$

Then it's clear that the bad type borrowers' manipulation strategy can be characterized by a cutoff $\bar{c}_s$, such that all bad type borrowers with manipulation cost $c \leq \bar{c}_s$ choose to manipulate their features, where

$$\bar{c}_s = \mathbb{E}(I^s) \cdot b. \tag{11}$$

Moreover, in equilibrium, we must have

$$k_s = [\mu + (1-\mu) F_c(\bar{c}_s)] I. \tag{12}$$

Then the equilibrium of the subgame s is characterized by (10),(11), and (12). Below is the formal definition of an equilibrium for any subgame s:

[18] In equilibrium, the debt face value $D_j^{s,\hat{G}}$ must satisfy the zero-profit condition: $\mu\mathbb{E}\left[\min\left\{\tilde{V}, D_j^{s,\hat{G}}\right\} | \theta = G\right] = k_s$. However, in our model, the debt face value only affects the distribution of surplus between lenders and borrowers and does not change the regulator's payoff (the surplus). So in this paper, we'll only focus on the approval decision.

Definition 3.3. An equilibrium of subgame s is a triple $(k_s, \underline{\rho}_s, \bar{c}_s)$, and a function I^s, where k_s is the total cost of financing $\hat{G}$ borrowers, $\underline{\rho}_s$ is the cutoff in lending approval decisions, $\bar{c}_s$ is the cutoff in bad type borrowers' manipulation decisions, and I^s is the probability that $\hat{G}$ borrowers are financed, such that the following conditions are satisfied:

1. Lender optimization: $I^s \begin{cases} = 1 & \text{if } \rho > \underline{\rho}_s \\ \in [0,1] & \text{if } \rho = \underline{\rho}_s \\ = 0 & \text{if } \rho < \underline{\rho}_s \end{cases}$, where $\underline{\rho}_s = m^{-1}\left(\frac{k_s}{\mu}\right)$;

2. Borrower optimization: $\bar{c}_s = \mathbb{E}\left(I^s\right) \cdot B$;

3. Consistency: $k_s = \left[\mu + (1-\mu) F_c\left(\bar{c}_s\right)\right] I$.

The regulator's utility in this subgame is all the surplus generated from the financing activities, which is

$$W_s = \mathbb{E}\left[\left(\mu m\left(\rho\right) - k_s\right)^+\right].$$

Then her unconditional expected utility is

$$W = \int_{s \in \mathcal{S}} W_s f\left(s\right) ds,$$

which is the expected surplus across all subgames. Then the regulator's information design problem is the following

$$\underset{\left\{\mathcal{S}, \left\{f(s), \pi(\rho|s)\right\}_{s\in\mathcal{S}}\right\}}{\text{maximize}} W = \int_{s\in\mathcal{S}} \mathbb{E}\left[\left(\mu m\left(\rho\right) - k_s\right)^+ |s\right] f\left(s\right) ds \tag{13}$$

$$\text{subject to} \quad \int_{s\in\mathcal{S}} f\left(s\right) ds = 1, \tag{14}$$

$$\int_{s\in\mathcal{S}} \pi\left(\rho|s\right) f\left(s\right) ds = \mathbb{1}_{\rho\in[0,1]}, \tag{15}$$

$$F_c\left(b \cdot \text{Prob}\left(\rho > m^{-1}\left(\frac{k_s}{\mu}\right)|s\right)\right) \le \frac{\mu}{1-\mu}\left(\frac{k_s}{\mu I} - 1\right)$$

$$\le F_c\left(b \cdot \text{Prob}\left(\rho \ge m^{-1}\left(\frac{k_s}{\mu}\right)|s\right)\right). \tag{16}$$

The solution to the regulator's problem is not unique, but in Section 5, we'll discuss and characterize the general properties of the optimal policies. The above regulator's problem is known as a Bayesian persuasion problem with continuous states. Bayesian persuasion models with continuous states are in general not tractable, except for some special cases (Gentzkow and Kamenica (2016), Dworczak and Martini (2019)). The regulator's problem in this paper does not fit into any existing tractable framework and I solve this model using a 'guess and verify' method.

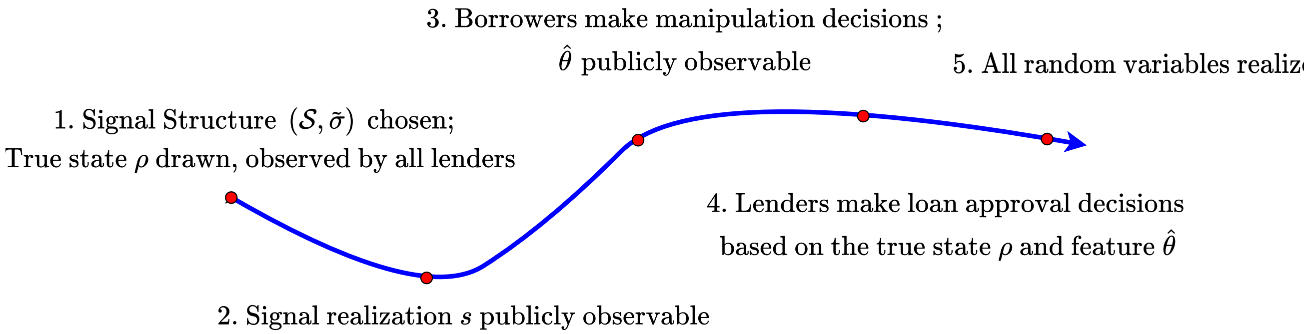

Figure 4: Timeline

3.6 Timeline

We summarize all the key ingredients of the model in Figure 4. All events occur in the following order:

1. the regulator chooses a signal structure $(\mathcal{S}, \tilde{\sigma})$; and Nature chooses the realization of ρ;
2. signal realization s is revealed, and is publicly observable to all the borrowers;
3. borrowers choose their manipulation strategies;
4. all lenders make their lending decisions simultaneously, and borrowers decide which contract to accept;
5. all random variables are realized, and all players receive their payoffs.

3.7 Discussion of the Assumptions

1. The notion of predictive algorithm. In practice, predictive algorithm usually refers to the mapping from observed input (which is the borrower feature after potential manipulation in this model) to the output (which is the future payoff distribution in this model). In this paper, I consider the disclosure policy from a pure informational perspective and it cannot serve as a commitment device. This means that, when lenders disclose their predictive algorithm, they are able to flexibly change their predictive algorithms privately as a response to the borrowers' manipulation behavior. Focusing

on the informational role of a predictive algorithm, disclosing information about the predictive algorithm is equivalent to disclosing the fundamental statistical properties of the random variables in the economic environment, which is the mapping from the borrower type to future payoff distribution in this model. By rational expectation, the manipulation behavior and lending decision rules are known by all players in equilibrium.

2. Disclosure vs regulating decision rules. In this paper, the disclosure is about the statistical properties of variables in the economy, and the regulator is not able to monitor or regulate lenders' lending decisions directly (for example, how they use certain variables in lending decisions). This feature is motivated by the challenge of regulating algorithmic lending in practice. First, although some regulations aim at regulating lending decisions directly (for example, prohibit the use of certain variables in lending decisions), the motivation usually comes from concerns on fairness and discrimination, and thus the regulation is independent of the statistical nature of the variables and easy to implement. With the focus on market surplus, in this paper, regulating lending decisions will depend on the statistical natural of variables, which is hard to monitor and implement. Second, regulating lending decisions by monitoring the use of certain variables may not be effective, because they can easily be deduced from other variables that correlate highly with them, known as the 'reconstruction problem' (Kleinberg et al. (2018)). Thirdly, the algorithms are dynamic and adjust over time depending on the availability of data and data processing technology, which makes it harder to monitor and regulate their decision rules directly.

3. Lending market structure. In this model, I assume all lenders are identical and the lending market is competitive. This assumption is mainly for expositional convenience. The regulator cares about the total surplus generated from all financing activities, but not the distribution of the surplus between borrowers and lenders. In this model, market structure only changes the distribution of surplus among borrowers and lenders but not the total surplus. In an extreme case with a monopoly lender, if the lender does not use differentiated contracts to screen borrowers[19], then all the results about optimal disclosure policies remain the same, and the only difference is the distribution of surplus between borrowers and lenders.

4. Bad type borrowers always fail. In the model, I assume the bad type borrowers always fail, and thus the only benefit they can receive from financing their projects is the private benefit. This is to simplify the lending market equilibrium, and make the analysis more concentrated on the disclosure side. Relaxing this assumption may

[19]In Appendix, I show that in our baseline model, lenders will not screen borrowers using differentiated contracts.

make the regulator's problem messy and intractable, but our key result, that partial disclosure policy is optimal, is still robust.

5. Only bad type borrowers are able to manipulate. In the model, I assume that only bad type borrowers are able to manipulate their features. But allowing good type borrowers to manipulate their features does not change the results. The key reason is that in equilibrium, $\hat{B}$ borrowers are always viewed as a worse group than $\hat{G}$ borrowers because bad type borrowers can always be $\hat{B}$ borrowers with no cost. Then good type borrowers have no incentive to manipulate and mimic the bad type. But this result relies on the assumption that the space of $\hat{\theta}$ is binary. In a model with general feature space, the good type borrowers may be able to signal their type by paying cost and differentiating themselves from the bad type borrowers further.

4 The Lack of Commitment Problem and the Inefficiency of No Disclosure

The only friction in our model is the adverse selection due to endogenous data manipulation behavior. Bad type borrowers change their manipulation behavior as a best response to the updated public belief on the relevance ρ. For the optimal policy, a natural guess would be that the regulator shouldn't disclose any information about the relevance ρ to the public and make it as opaque as possible. In this case, the lending market equilibrium is characterized by $(k_N, \varrho_N, \bar{c}_N)$, and the regulator's payoff is

$$W_N = \int_{\varrho_N}^{1} \left(\mu m\left(\rho\right) - k_N\right) d\rho.$$

However, in this scenario, the use of the feature $\hat{\theta}$ is too intensive from the regulator's perspective, and thus it creates too much manipulation unconditionally. This result comes from the lenders' lack of commitment problem: they always make the most efficient use of borrower data ex post. To see this, suppose the regulator can 'force' all lenders to choose a higher lending cutoff $\varrho_N + x$ ($x \ll 1$), so the lenders only use feature $\hat{\theta}$ in their lending decisions when the relevance $\rho > \varrho_N + x$. From the perspective of borrowers, the feature $\hat{\theta}$ will be used with lower probability, and thus discourage their manipulation incentives. The marginal change of regulator's payoff is

$$\left.\frac{dW}{dx}\right|_{x=0} = -\left(\mu m\left(\varrho_N\right) - k_N\right) + \int_{\varrho_N}^{1} \left(-\left.\frac{dk_N}{dx}\right|_{x=0}\right) d\rho. \tag{17}$$

In equilibrium we must have $-\left.\frac{dk_N}{dx}\right|_{x=0} < 0$, because the more stringent lending cutoff discourages borrowers' manipulation incentives, which in turn decreases the total financing cost. Besides, the ex post efficiency in the lending market equilibrium implies

$$\mu m\left(\rho_N\right)-k_N=0,$$

this is the lenders' break-even condition at $\rho=\rho_N$ in the no disclosure equilibrium. The above two observations jointly imply that

$$\left.\frac{dW}{dx}\right|_{x=0}>0.$$

This suggests that 'forcing' lenders to use the feature less frequently improves the outcome of the lending market. Similar results show up in other economics settings where the information receivers commit to underweight some variables in decision rules to deter manipulation and improve efficiency (for example, Ball (2019)).

Although committing to the lending decisions is impossible in our model, the regulator can mitigate (average) manipulation behavior by disclosing information about the true state of relevance ρ. This leads to our first key result: the suboptimality of no disclosure equilibrium.

Proposition 4.1. *There exists a disclosure policy* $(\mathcal{S},\tilde{\sigma})$ *with total surplus* W*, such that* $W>W_N$.

Proposition 4.1 challenges the conventional wisdom that making algorithms more transparent will always hurt efficiency because of the "gaming the system" concern. This is not true even if only correlational features are used in the predictive algorithm. The key to Proposition 4.1 is to find a disclosure policy under which the lenders will use feature $\hat{\theta}$ less frequently from the ex ante perspective, which will deter manipulation of the feature $\hat{\theta}$.

To gain intuitions on how it works, suppose the regulator designs a deterministic disclosure policy with three elements in the signal space $\mathcal{S}=\{s_1,s_2,s_3\}$, and the message function is

$$\sigma\left(\rho\right)=s_1\mathbb{1}_{A_1}\left(\rho\right)+s_2\mathbb{1}_{A_2}\left(\rho\right)+s_3\mathbb{1}_{A_3}\left(\rho\right)$$

where A_1, A_2, and A_3 are (unions of) intervals shown on Figure 5. The above disclosure policy effectively discloses which set of A_1, A_2, and A_3 that the true state belongs to. When signal s_i is disclosed to the borrowers, updated belief $\pi\left(\rho|s_i\right)$ is a uniform distribution conditional on set A_i. The boundaries of the intervals are chosen such that:

1. the equilibrium of subgame s_1 is the same as the no disclosure equilibrium, i.e.,

$$\left(k_1,\rho_1,\bar{c}_1\right)=\left(k_N,\rho_N,\bar{c}_N\right);$$

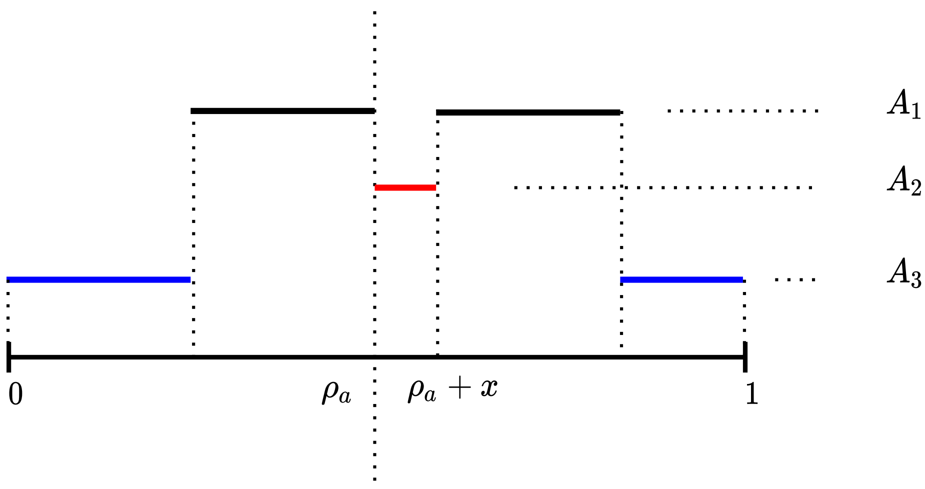

Figure 5: Suboptimality of No Disclosure Policy

2. $A_2 = [\rho_N, \rho_N + x]$, where $x \ll 1$;

3. $A_3 = [0,1] - A_1 \cup A_2$.

The equilibrium of subgame s_1 is the same as the no disclosure equilibrium, so it has no effect on the change of regulator's payoff. The signal s_2 reveals that the true state is in the interval $[\rho_N, \rho_N + x]$. Note that the equilibrium condition

$$\mu m\left(\rho_N\right) - k_N = 0$$

implies the surplus when $\rho \in [\rho_N, \rho_N + x]$ is close to zero in the no disclosure equilibrium. And in the equilibrium of subgame s_2, the surplus must be nonnegative, then the change of regulator's payoff is also negligible in this case. When s_3 is disclosed, in the equilibrium of subgame s_3, the probability of financing $\hat{G}$ borrowers is lower than that in the no disclosure equilibrium (note that $\hat{G}$ borrowers will be financed only if the true state is in the right interval of A_3), which mitigates the manipulation incentives of bad type borrowers and improves the market surplus. Then the net effect of marginally increasing lending cutoff is positive.

5 General Properties of Optimal Policies

In this section, I discuss the general properties of the optimal policies.

5.1 Structure of The Optimal Policies

We already show that no disclosure is suboptimal in Section 4. Another natural guess for the optimal disclosure policy is full transparency, i.e., disclosing all information about the relevance ρ to the public. We can show that full transparency leads to the worst outcome, and thus it must be suboptimal.

Lemma 5.1. *Suppose W_F is the regulator's payoff when she makes the true state of relevance ρ fully transparent, then $W_F = 0$.*

Note that the regulator's payoff must be nonnegative. Lemma 5.1 implies that disclosing all information about the true state of relevance ρ leads to the regulator's worst payoff, so it must be suboptimal. The intuition behind the result is straightforward: when bad type borrowers know perfectly about the true state of relevance ρ, then in equilibrium, the data manipulation level satisfies that there is zero surplus from financing $\hat{G}$ borrowers, and lenders are indifferent between using and not using borrower data in lending decisions. This result is consistent with the popular argument that disclosing too much information about the predictive model hurts efficiency.

Proposition 4.1 and Lemma 5.1 jointly imply that the optimal disclosure policy must feature partial disclosure. Before exploring the properties of the optimal policy, we show all the subgame equilibria are ranked by equilibrium variables.

Lemma 5.2. *For any disclosure policy $(\mathcal{S}, \tilde{\sigma})$, and any two signal realizations s_1 and s_2, we must have*

$$k_{s_1} \lesseqgtr k_{s_2} \Longleftrightarrow \bar{c}_{s_1} \lesseqgtr \bar{c}_{s_2} \Longleftrightarrow \varrho_{s_1} \lesseqgtr \varrho_{s_2},$$

where $(k_{s_1}, \varrho_{s_1}, \bar{c}_{s_1})$ and $(k_{s_2}, \varrho_{s_2}, \bar{c}_{s_2})$ are defined in Definition 3.3.

When more bad type borrowers manipulate their features, adverse selection is more severe in the pool of $\hat{G}$ borrowers, and the quality type θ needs to be a more relevant variable (higher ρ) in identifying borrowers with better quality in lending decisions. Another observation related to Lemma 5.2 is that if there exist two signal realizations s_1, s_2 such that

$$(k_{s_1}, \varrho_{s_1}, \bar{c}_{s_1}) = (k_{s_2}, \varrho_{s_2}, \bar{c}_{s_2}),$$

then "combining" these two signal realizations together does not change the equilibrium outcome. The following lemma is a formal statement of this result.

Lemma 5.3. *For an optimal signal structure $(\mathcal{S}, \tilde{\sigma})$ with distribution of posterior beliefs $\{f(s), \pi(\rho|s)\}_{s \in \mathcal{S}}$, if there exist two distinct realizations $s_1, s_2 \in \mathcal{S}$, such that*

$$(k_{s_1}, \varrho_{s_1}, \bar{c}_{s_1}) = (k_{s_2}, \varrho_{s_2}, \bar{c}_{s_2}),$$

then the signal structure $(\mathcal{S}', \tilde{\sigma}')$ *is also optimal, where* $\{s_0'\} \notin \mathcal{S}$ *and* $(\mathcal{S}', \tilde{\sigma}')$ *is defined by*

$$\mathcal{S}' = \{s_0'\} \cup \mathcal{S} \setminus \{s_1, s_2\}$$

and

$$\tilde{\sigma}'(s|\rho) = \tilde{\sigma}(s|\rho)\, \mathbb{1}_{\mathcal{S}\setminus\{s_1,s_2\}}(s) + (\tilde{\sigma}(s_1|\rho) + \tilde{\sigma}(s_2|\rho))\, \mathbb{1}_{\{s_0'\}}(s)$$

for all $\rho \in [0,1]$ *and* $s \in \mathcal{S}'$.

Lemma 5.3 is very intuitive. When there are two signal realizations s_1 and s_2 that lead to the equivalent equilibria, then instead of disclosing these two signal realizations separately, we can simply disclose that "the realization is either s_1 or s_2", and the equilibrium outcome will be unchanged. Based on this observation, without loss of generality, we impose the following restriction on optimal policies:

Criterion 2. *We focus on policies* $(\mathcal{S}, \tilde{\sigma})$ *such that for any* $s_1, s_2 \in \mathcal{S}$ *and* $s_1 \neq s_2$*, the lending market equilibria satisfy* $(k_{s_1}, \rho_{s_1}, \bar{c}_{s_1}) \neq (k_{s_2}, \rho_{s_2}, \bar{c}_{s_2})$.

Based on the above criterion and the suboptimality of no disclosure equilibrium, the optimal policy must differentiate the subgame equilibria by the data manipulation levels (and other equilibrium variables), which is measured by $\bar{c}_s$. The next lemma shows that data manipulation exists in all subgame equilibria, so there is no first best outcome for any subgame. And an implication of the lemma is that it is never optimal to confirm that a feature is not used in lending decisions for sure.

Lemma 5.4. *(Manipulation in all states) Suppose* $(\mathcal{S}, \tilde{\sigma})$ *is an optimal policy. Then for almost all* $s \in \mathcal{S}$*, we must have*

$$k_s > I, \rho_s > m^{-1}(I), \bar{c}_s > 0.$$

Lemma 5.4 rules out some disclosure policies. For example, suppose the regulator chooses a disclosure policy that reveals whether the relevance ρ is below $m^{-1}(I)$ or not. Note that for any $\rho < m^{-1}(I)$, it is inefficient to finance any borrowers, then there will be no loan approved and no manipulation. On the other hand, if $\rho > m^{-1}(I)$ is revealed, $\hat{G}$ borrowers will be financed and the unique lending market equilibrium is determined by conditions in Definition 3.3. This disclosure policy violates the result in Lemma 5.4, and thus it is inefficient. This is because compared to the no disclosure equilibrium, the regulator does not gain anything from states $\rho \leq m^{-1}(I)$ as she still only receives zero payoff, but more people will choose to manipulate in states $\rho > m^{-1}(I)$, as signal $\rho > m^{-1}(I)$ confirms the high relevance of the feature $\hat{\theta}$ and incentivizes more manipulation. This cutoff policy is dominated by the no disclosure policy, which effectively pools these two signals together. Actually, as we will discuss later, in the optimal policy, we want to mix low states (where

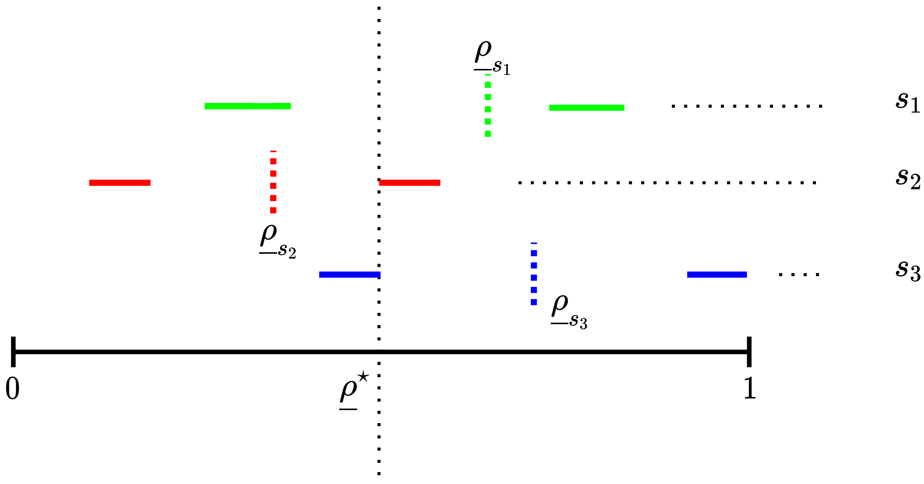

Figure 6: Efficiency

relevance ρ is low) with high states (where relevance ρ is high) and preserve uncertainty of the true state of relevance ρ in all posterior equilibria.

The second necessary condition of optimal policy features ex ante cutoff of lending decisions. Note that under any subgame s, the loan applications from $\hat{G}$ borrowers will be approved if the relevance ρ is high enough, i.e., when $\rho > \underline{\rho}_s$. This means that the lending decision is always a cutoff decision ex post, and this is a natural result in equilibrium: the feature $\hat{\theta}$ is more useful when ρ is higher. It turns out that this condition is also satisfied ex ante under the optimal disclosure policy. The following lemma states this result.

Lemma 5.5. *(Ex ante lending cutoff) Suppose that $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, with induced distribution of posteriors $\{f(s), \pi(\rho|s)\}_{s\in\mathcal{S}}$, then there must exist a constant $\underline{\rho}^{\star} \in (0,1)$, such that for almost all $s \in \mathcal{S}$, $\hat{G}$ borrowers are financed if and only if*

$$\rho \in (\underline{\rho}^{\star}, 1] \cap \operatorname{supp}(\pi(\rho|s)).$$

Figure 6 explains Lemma 5.5 by showing three specific signal realizations s_1, s_2, and s_3. The colored regions represent the posterior beliefs under these three signals, and cutoffs $\underline{\rho}_{s_1}$, $\underline{\rho}_{s_2}$ and $\underline{\rho}_{s_3}$ represent lenders' equilibrium lending cutoffs in these three equilibria. Consider signal s_1 with $\underline{\rho}_{s_1} > \underline{\rho}^{\star}$. Since in the subgame s_1, $\hat{G}$ borrowers will not be financed if $\rho \leq \underline{\rho}_{s_1}$, Lemma 5.5 implies that

$$\operatorname{supp}(\pi(\rho|s_1)) \cap (\underline{\rho}^{\star}, \underline{\rho}_{s_1}] = \emptyset.$$

Similar results can be obtained for all other signal realizations.

Then from the unconditional (ex ante) perspective, $\hat{G}$ borrowers will be financed if and only if

$$\rho > \varrho^{\star}.$$

This condition confirms the efficiency of optimal policies, in which lenders utilize the feature $\hat{\theta}$ if and only if the relevance ρ is high enough. Besides, note that for almost all subgame s, Lemma 5.5 implies that, in equilibrium we have

$$\sup \{\operatorname{supp}(\pi(\rho|s)) \cap [0, \varrho^{\star}]\} \leq \varrho_s \leq \inf \{\operatorname{supp}(\pi(\rho|s)) \cap (\varrho^{\star}, 1]\}.$$

Then the support of the posterior belief is divided into two parts: the rejection region

$$\operatorname{supp}(\pi(\rho|s)) \cap [0, \varrho^{\star}]$$

in which all loan applications are rejected, so the lending decision is independent of borrower data; and the approval region

$$\operatorname{supp}(\pi(\rho|s)) \cap (\varrho^{\star}, 1]$$

in which $\hat{G}$ borrowers are financed, so the lending decision making is dependent on borrower data. The following lemma shows that, without loss of generality, we can focus on deterministic disclosure policies in which both regions are intervals for all posterior equilibria, and all subgames are ranked by the equilibrium data manipulation levels.

Lemma 5.6. *For any optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$, there must exist a deterministic optimal policy $(\mathcal{S}, \sigma)$ with the same signal space $\mathcal{S}$. Let $\left\{\tilde{f}(s), \tilde{\pi}(\rho|s)\right\}_{s \in \mathcal{S}}$ and $\{f(s), \pi(\rho|s)\}_{s \in \mathcal{S}}$ be the distribution of posteriors for the policy $(\mathcal{S}, \tilde{\sigma})$ and $(\mathcal{S}, \sigma)$ respectively, and let $\left(\tilde{k}_s, \tilde{\varrho}_s, \tilde{\bar{c}}_s\right)$ and $(k_s, \varrho_s, \bar{c}_s)$ be equilibrium outcomes for the policy $(\mathcal{S}, \tilde{\sigma})$ and $(\mathcal{S}, \sigma)$ respectively. Then the following properties hold:*

1. *$f(s) = \tilde{f}(s)$ and $\left(\tilde{k}_s, \tilde{\varrho}_s, \tilde{\bar{c}}_s\right) = (k_s, \varrho_s, \bar{c}_s)$ for almost all s, and the ex ante lending cutoffs defined in Lemma 5.5 are the same under these two policies, denoted as $\varrho^{\star}$;*

2. *for almost all $s \in \mathcal{S}$, both*

$$\operatorname{supp}(\pi(\rho|s)) \cap [0, \varrho^{\star}]$$

and

$$\operatorname{supp}(\pi(\rho|s)) \cap (\varrho^{\star}, 1]$$

are non-empty intervals;

3. *for almost all $s_1, s_2 \in S$ with $\bar{c}_{s_1} < \bar{c}_{s_2}$,*

$$\sup \{\operatorname{supp}(\pi(\rho|s_1)) \cap [0, \varrho^{\star}]\} \leq \inf \{\operatorname{supp}(\pi(\rho|s_2)) \cap [0, \varrho^{\star}]\} \tag{18}$$

and

$$\sup \{\operatorname{supp}(\pi(\rho|s_1)) \cap (\varrho^{\star}, 1]\} \leq \inf \{\operatorname{supp}(\pi(\rho|s_2)) \cap (\varrho^{\star}, 1]\}. \tag{19}$$

Lemma 5.6 simplifies the space of optimal disclosure policies. It shows that, for any optimal policy $(\mathcal{S}, \tilde{\sigma})$, we can find a payoff-equivalent deterministic policy $(\mathcal{S}, \sigma)$ which induces the same posterior lending market equilibria. And for the deterministic optimal disclosure policy, the posterior belief always consists of two intervals representing the rejection region and the approval region. For almost all signal realizations, the posterior equilibria are ranked by the equilibrium data manipulation levels (measured by $\bar{c}_s$). Based on the above observations, we characterize the structure of an optimal disclosure policy in the following theorem.

Theorem 5.1. *There exists a deterministic optimal policy $(\mathcal{S}, \sigma)$ which consists of*

1. *a signal space $\mathcal{S} \subset [\bar{c}_{\min}, \bar{c}_{\max}]$;*
2. *a message function σ and cutoff $\varrho^\star \in (0, 1)$ such that both*

$$\sigma|_{[0,\varrho^\star]} : [0, \varrho^\star] \to \mathcal{S}$$

and

$$\sigma|_{(\varrho^\star,1]} : (\varrho^\star, 1] \to \mathcal{S}$$

are weakly increasing functions with the same range.

Under this optimal policy, for any subgame s, the equilibrium cutoff of data manipulation cost is $\bar{c}_s = s$, and $\hat{G}$ borrowers will be financed if and only if $\rho > \varrho^\star$ for all s.

Here we select a specific signal space such that the message sent to borrowers is actually the recommended data manipulation decision. Upon observing signal realization s, bad type borrowers are recommended to manipulate their features if and only if their manipulation cost satisfies $c \leq s$. Note that in Theorem 5.1 we only characterize the general structure of the optimal message function σ but not provide the exact functional form of it.

Figure 7 is a graphical illustration of Theorem 5.1. For each signal realization s (for example, the signal $s = \bar{c}_{\min}$ or $\bar{c}_{24}$), the posterior belief is a union of two disjoint intervals[20] which can always be separated by the cutoff $\varrho^\star$. These two disjoint intervals represent the rejection region and approval region in lending decisions. For example, the red intervals represent the posterior belief of signal $s = \bar{c}_{\min}$, and when the true state of ρ is in the red region, the recommended cutoff of data manipulation cost $\bar{c}_{\min}$ is sent to the borrowers. Upon observing the signal $s = \bar{c}_{\min}$, bad type borrowers update their belief and choose to manipulate their features if their manipulation cost satisfies $c \leq \bar{c}_{\min}$.

The optimal message function consists of two parts, denoted as

$$\sigma_L = \sigma|_{[0,\varrho^\star]}$$

[20]Note that a single point is also a closed interval.

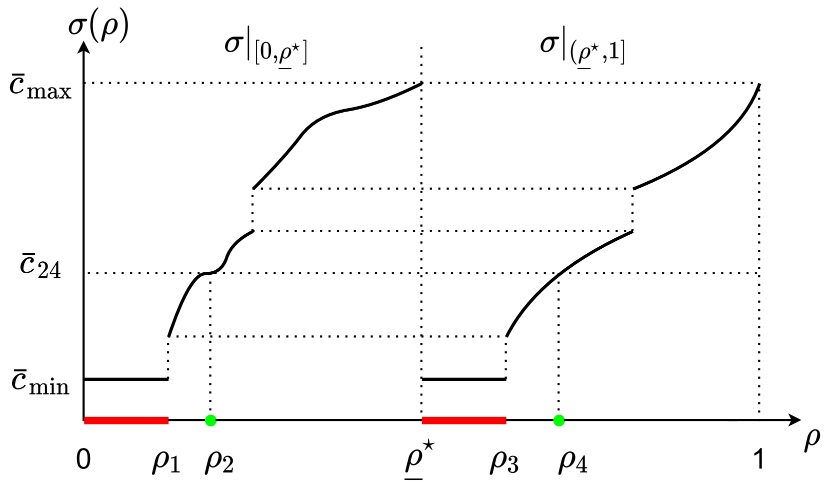

Figure 7: Graphical Illustration of Theorem 5.1

and

$$\sigma_R = \sigma|_{(\underline{\rho}^\star,1]}\,.$$

Then $\sigma_L(\rho)$ and $\sigma_R(\rho)$ can be viewed as the message functions for the rejection region and approval region respectively. For any $s \in \mathrm{Ran}(\sigma_L) = \mathrm{Ran}(\sigma_R)$[21], $\sigma_L^{-1}(s)$ (or $\sigma_R^{-1}(s)$) can either be an interval with positive length or a single point. In the first case, the signal is discrete and the equilibrium approval probability for $\hat{G}$ borrowers is

$$\frac{\mathrm{Prob}\left(\sigma_R^{-1}(s)\right)}{\mathrm{Prob}\left(\sigma_L^{-1}(s)\right) + \mathrm{Prob}\left(\sigma_R^{-1}(s)\right)}.$$

In the second case, the signal is continuous, and the equilibrium approval probability for $\hat{G}$ borrowers is[22]

$$\frac{\frac{1}{\sigma_R'\left(\sigma_R^{-1}(s)\right)}}{\frac{1}{\sigma_L'\left(\sigma_L^{-1}(s)\right)} + \frac{1}{\sigma_R'\left(\sigma_R^{-1}(s)\right)}},$$

[21] $\mathrm{Ran}(f)$ means the range of a function f.

[22] In this case, the probability of observing a specific signal is always zero. Then the distribution of signal is represented by a density function $f(s)$:

$$f(s) = \frac{1}{\tilde{\sigma}_L'\left(\tilde{\sigma}_L^{-1}(s)\right)} + \frac{1}{\tilde{\sigma}_R'\left(\tilde{\sigma}_R^{-1}(s)\right)},$$

where $\frac{1}{\tilde{\sigma}_L'\left(\tilde{\sigma}_L^{-1}(s)\right)}$ and $\frac{1}{\tilde{\sigma}_R'\left(\tilde{\sigma}_R^{-1}(s)\right)}$ represent the weights of the rejection region and approval region, respectively.

where $\frac{1}{\sigma_R'\left(\sigma_R^{-1}(s)\right)}$ $\left(\frac{1}{\sigma_L'\left(\sigma_L^{-1}(s)\right)}\right)$ is an analog of Prob $\left(\sigma_R^{-1}(s)\right)$ $\left(\text{Prob}\left(\sigma_L^{-1}(s)\right)\right)$ in the previous case.

5.2 Properties of Optimal Policies

In this subsection, we discuss some properties of the posterior equilibria. First, the prior belief of ρ is $\rho \sim U[0,1]$, then unconditionally, the probability that $\hat{G}$ borrowers are financed is

$$\text{Prob}\left(\rho > \underline{\rho}^{\star}\right) = 1 - \underline{\rho}^{\star}.$$

Similarly, in the no disclosure equilibrium, the probability that $\hat{G}$ borrowers are financed is

$$\text{Prob}\left(\rho > \underline{\rho}_N\right) = 1 - \underline{\rho}_N.$$

The following Proposition shows that $\hat{G}$ borrowers are financed less frequently under the optimal disclosure policy compared to the no disclosure case.

Proposition 5.1. *Suppose $\underline{\rho}^{\star}$ is the cutoff described in Lemma 5.5, then $\underline{\rho}^{\star} > \underline{\rho}_N$.*

Proposition 5.1 implies that borrower data are used less frequently under the optimal disclosure policy compared to the no disclosure case, which confirms our intuition why no disclosure equilibrium is suboptimal, and why it can be improved. In the no disclosure equilibrium, the feature $\hat{\theta}$ is used too intensively, resulting in too much manipulation. To mitigate this problem, the regulator prefers the feature $\hat{\theta}$ to be used less frequently, and this is achieved by the optimal policy.

The second property is about the data manipulation levels in posterior equilibria.

Proposition 5.2. $\bar{c}_{\max} > \bar{c}_N > \bar{c}_{\min}$.

$\bar{c}_{\max}$ and $\bar{c}_{\min}$ represents the highest and lowest equilibrium cutoffs of manipulation cost among all posterior equilibria. Proposition 5.2 explains the idea of differentiation of posterior equilibria. In the equilibrium with highest data manipulation level ($\bar{c}_s = \bar{c}_{\max}$), a higher ρ is required for the feature $\hat{\theta}$ to be used in lending decisions, and this deters the use of borrower data in this subgame equilibrium. The cost is the higher data manipulation level, and thus lenders have to finance more bad type borrowers, while the benefit is the less use of borrower data which discourages data manipulation unconditionally. As we discussed in Section 4, the positive effect dominates and surplus improves.

The last property is about the surplus in the posterior equilibria. Note in Section 4, we show that the inefficiency comes from states when ρ is close to ρ_N (see condition (17)). In these states the regulator's payoff is small, so the benefit of financing $\hat{G}$ borrowers cannot justify the negative externality it imposes on other states. The following Proposition shows that, to mitigate the negative externality, the surplus from lending activities must be large enough for any posterior equilibrium that occurs with positive probability.

Proposition 5.3. *(Positive surplus) Under any optimal policy characterized in Theorem 5.1, for any $\epsilon > 0$, there exists $\delta > 0$, such that for any posterior equilibrium with signal realization s satisfying*

$$Prob\,(s) > \epsilon,$$

the surplus from lending must be greater than δ for any $\rho > \varrho^\star$.

5.3 A Closed-Form Characterization

I characterize the general structure of optimal policies in Theorem 5.1, while leaving the functional form of message function $\sigma(\cdot)$ unsolved. In this subsection, I provide a closed-form characterization of the optimal policy by imposing a distributional assumption on the manipulation cost.

Assumption 4. *$xF_c(x)$ has at most one inflection point*[23] *on $[0, \bar{c}]$.*

Many commonly used distribution functions satisfy Assumption 4, including truncated normal distribution, uniform distribution, truncated exponential distribution, Beta distribution, Gamma distribution, Weibull distribution, etc. Since $xF_c(x)$ is locally convex around $x = 0$, Assumption 4 means that $xF_c(x)$ is either a weakly convex function on $[0, \bar{c}]$, or there exists $\tilde{c} \in (0, \bar{c})$ such that $xF_c(x)$ is weakly convex on $[0, \tilde{c}]$ and weakly concave on $[\tilde{c}, \bar{c}]$. With this assumption, the optimal policy has a simpler structure. In Theorem 5.1, the message is the recommended data manipulation decision, while in the following theorem, without loss of generality, I choose a different signal space to make the results simpler.

Theorem 5.2. *When Assumption 4 is satisfied, there exists a deterministic optimal policy $(\mathcal{S}, \sigma)$ characterized by*

1. *three cutoffs $(\varrho_a, \varrho^\star, \varrho_b)$ satisfying $0 < \varrho_a < \varrho^\star < \varrho_b < 1$;*
2. *a signal space $\mathcal{S} = [\varrho_a, \varrho^\star]$;*
3. *a continuous, strictly increasing function $\gamma : [\varrho_b, 1] \to [\varrho_a, \varrho^\star]$ satisfying $\gamma(\varrho_b) = \varrho_a$ and $\gamma(1) = \varrho^\star$, such that the message functions $\sigma(\rho)$ is*

$$\sigma|_{[0,\varrho^\star]} = \begin{cases} \varrho_a & \text{if } \rho \in [0, \varrho_a] \\ \rho & \text{if } \rho \in (\varrho_a, \varrho^\star] \end{cases}$$

and

$$\sigma|_{(\varrho^\star,1]} = \begin{cases} \varrho_a & \text{if } \rho \in (\varrho^\star, \varrho_b] \\ \gamma(\rho) & \text{if } \rho \in (\varrho_b, 1] \end{cases}.$$

[23] Inflection points are points where the function changes concavity.

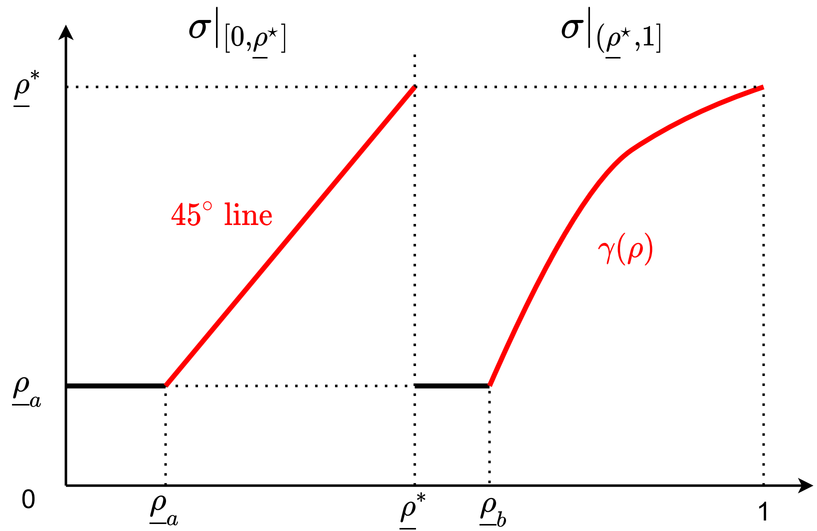

Figure 8: A closed-form characterization

For any $s \in \mathcal{S}$, the equilibrium data manipulation decision $\bar{c}_s$ satisfies

$$\mu m(s) = (\mu + (1-\mu) F_c(\bar{c}_s)) I.$$

The optimal policy is a simplified version of our general result in Theorem 5.1. Both $\sigma|_{[0,\underline{\rho}^{\star}]}$ and $\sigma|_{(\underline{\rho}^{\star},1]}$ are continuous and consist of a flat region and a strictly increasing region. In the signal space $\mathcal{S} = [\underline{\rho}_a, \underline{\rho}^{\star}]$, $s = \underline{\rho}_a$ is a discrete signal and the posterior belief is a uniform distribution conditional on $[0, \underline{\rho}_a] \cup (\underline{\rho}^{\star}, \underline{\rho}_b]$. For any $s \in (\underline{\rho}_a, \underline{\rho}^{\star}]$, the signal is continuous. Let $x = \gamma^{-1}(s)$, then the posterior distribution of relevance ρ is a lottery[24] with binary outcomes:

$$\left\langle (\gamma(x), x), \frac{(\gamma'(x), 1)}{\gamma'(x) + 1} \right\rangle. \tag{20}$$

For any subgame s, the equilibrium lending cutoff $\underline{\rho}_s$ satisfies the following condition:

Lemma 5.7. *Under the deterministic optimal disclosure policy $(\mathcal{S}, \sigma)$ characterized in Theorem 5.2, for any $s \in \mathcal{S}$, we must have $\underline{\rho}_s = \sup\{\operatorname{supp}(\pi(\rho|s)) \cap [0, \underline{\rho}^{\star}]\}$.*

This means that the lending cutoff $\underline{\rho}_s$ is chosen such that it equals the highest value in the rejection region. Lemma 5.7 and all equilibrium conditions jointly imply that for all

[24]A lottery $\langle (x_1, x_2, ...x_N), (p_1, p_2, ...p_N) \rangle$ is a discrete random variable with probability function $\text{Prob}(x = x_i) = p_i$.

$\rho \in (\varrho_b, 1]$, the function $\gamma(\rho)$ satisfies the following ODE:

$$m(\gamma(\rho)) = 1 + \frac{1-\mu}{\mu} F_c\left(b \cdot \frac{1}{1+\gamma'(\rho)}\right), \tag{21}$$

with boundary conditions

$$\gamma(\varrho_b) = \varrho_a \text{ and } \gamma(1) = \varrho^\star.$$

The equilibrium condition under the discrete signal $s = \varrho_a$ implies

$$\gamma'(\varrho_b) = \frac{\varrho_a}{\varrho_b - \varrho^\star}.$$

With the above characterization, all of $\gamma(\rho)$, ϱ_a and ϱ_b can be solved as a function of the ex ante lending cutoff $\varrho^\star$, so the equilibrium is uniquely determined by a single variable $\varrho^\star$. Then we can reduce the original infinite-dimensional optimization problem to a one-dimensional problem. The regulator's problem becomes

$$\begin{aligned}
\underset{\varrho^\star}{\text{maximize}} & \int_{\varrho^\star}^{1} m(x)\,dx - \int_{\varrho_b}^{1} m(\gamma(x))\,dx - (\varrho_b - \varrho^\star)\, m(\varrho_a) \\
\text{subject to } & m(\gamma(x)) = 1 + \frac{1-\mu}{\mu} F_c\left(b \cdot \frac{1}{1+\gamma'(x)}\right), \\
& \gamma(\varrho_b) = \varrho_a, \gamma(1) = \varrho^\star, \\
& \gamma'(\varrho_b) = \frac{\varrho_a}{\varrho_b - \varrho^\star}.
\end{aligned} \tag{22}$$

Lemma 5.7 also implies that approval probability in posterior equilibrium is an increasing function of s. Note that for any $s \in (\varrho_a, \varrho^\star)$, the posterior belief is a lottery represented by (20), then the approval probability is

$$\frac{1}{1+\gamma'(\gamma^{-1}(s))}.$$

These observations jointly implies that $\gamma'(\rho)$ is a strictly decreasing function of ρ. The following lemma states the formal result.

Lemma 5.8. *In the optimal disclosure policy characterized in Theorem 5.2, $\gamma(\rho)$ is a strictly concave function on $\rho \in (\varrho_b, 1)$.*

6 Extension: Costly Fraud Detection

In the main model, all bad type borrowers' manipulation decisions are unobservable to lenders. In practice, lenders can also costly identify fraudulent activities using various methods, which is another way of mitigating adverse selection. In this extension, I consider how the disclosure policy interacts with fraud detection in the regulator's problem.

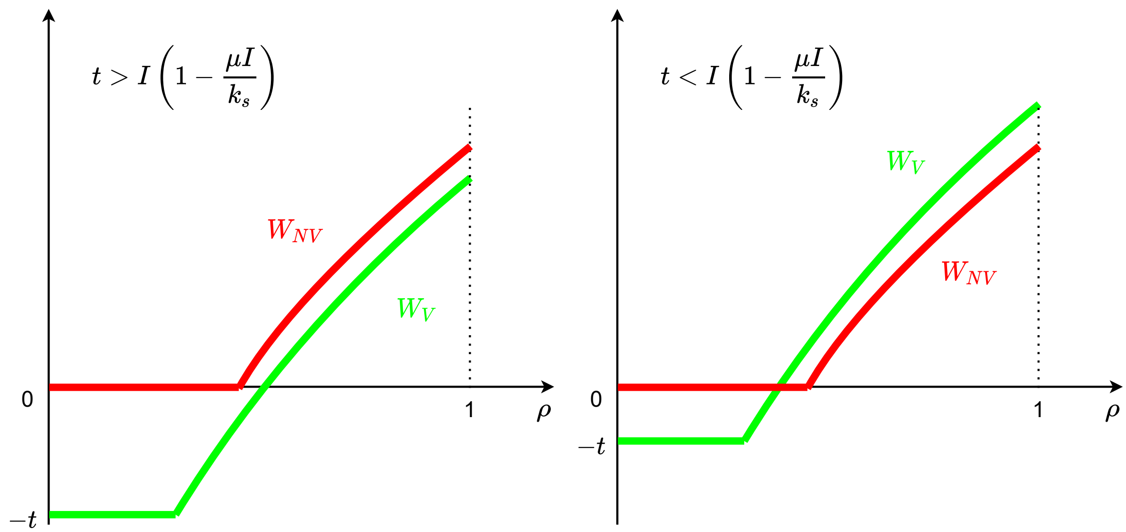

Figure 9: W_N vs W_{NV}

Assume all lenders have the identical linear cost function of fraud detection, i.e., each lender can verify and reveal any borrower's true type by paying cost $t > 0$. Once the type of a borrower is verified, it becomes public information. To consider the optimal disclosure policy with this fraud detection technology, note that for any equilibrium with posterior belief $\pi(\rho|s)$ and total financing cost k_s, the net value of verifying a $\hat{G}$ borrower's true type is

$$W_V = \max\left\{\frac{\mu I}{k_s}\left(m(\rho) - I\right), 0\right\} - t.$$

And the net value of not verifying the borrower's true type is

$$W_{NV} = \max\left\{\frac{\mu I}{k_s} m(\rho) - I, 0\right\}.$$

Figure 9 compares W_V and W_{NV}. When $t > I\left(1 - \frac{\mu I}{k_s}\right)$,

$$W_{NV} > W_V$$

for all ρ, then in this case, lenders will never verify any borrower's true type.

When $t < I\left(1 - \frac{\mu I}{k_s}\right)$,

$$W_{NV} < W_V \Longleftrightarrow \rho > \rho_e,$$

where ρ_e solves

$$\frac{\mu I}{k_s}\left(m(\rho_e) - I\right) - t = 0.$$

Then lenders will verify $\hat{G}$ borrowers with probability 1 when $\rho > \rho_e$, and lend to $\hat{G}$ borrowers only when they pass the verification. However, this cannot be an equilibrium because in this case no bad type borrower has the incentive to manipulate (since approval is possible only when they pass the verification). As a best response, lenders have no incentive to verify, which is a contradiction.

When $t = I\left(1 - \frac{\mu I}{k_s}\right)$, ρ_e solves

$$\mu m\left(\rho_e\right) - k_s = 0,$$

so $\rho_e = \rho_s$. Moreover,

$$W_{NV} = (>) W_V \iff \rho \geq (<) \rho_e.$$

In this case, lenders are indifferent between verifying or not when $\rho \geq \rho_s$, and will lend to $\hat{G}$ borrowers only when $\rho \geq \rho_s$.

In summary, if in subgame s, lenders verify any borrower's true type with positive probability, we must have

$$t = I\left(1 - \frac{\mu I}{k_s}\right) \iff k_s = k^v = \frac{\mu I^2}{I - t}.$$

And the data manipulation level $\bar{c}^v$ is

$$k^v = \left(\mu + (1-\mu) F_c\left(\bar{c}^v\right)\right) I \iff \bar{c}^v = F_c^{-1}\left(\frac{\mu t}{(1-\mu)(I-t)}\right).$$

In this subgame, when $\rho \geq m^{-1}\left(\frac{k^v}{\mu}\right)$, lenders verify $\hat{G}$ borrowers' true types with positive probability and lend to those who are not verified or verified to be good type borrowers. The verification probability p^v[25] satisfies the condition that bad type borrowers with cost $\bar{c}^v$ break even.

The following theorem presents the optimal disclosure policy with verification and confirms the robustness of our baseline result.

Theorem 6.1. *With costly verification, there exists t^v such that*

1. *when $t \geq t^v$, lenders will never use verification, and the optimal disclosure policy will not change;*

2. *when $t < t^v$, there exists $\rho^v \in (0, 1)$, such that the optimal disclosure is characterized as two steps:*

   (a) *The regulator first reveals if the true state ρ is above ρ^v or not.*

[25]Here I assume p^v to be constant for simplicity. The choice of verification probability p^v can depend on the true state ρ, and thus is not unique.

(b) *If the true state $\rho > \rho^v$, then the lenders will verify all $\hat{G}$ borrowers with probability $p^v = 1 - \frac{\bar{c}^v}{B}$, and lend to $\hat{G}$ borrowers who are not verified or verified to be good type borrowers.*

(c) *If the true state $\rho \in [0, \rho^v]$, then information about ρ is disclosed according to a policy $(\mathcal{S}^v, \sigma^v)$, where $(\mathcal{S}^v, \sigma^v)$ is an optimal disclosure policy characterized in Theorem 5.1 with prior belief $\rho \sim U[0, \rho^v]$.*

Theorem 6.1 shows that the disclosure policy and verification technology interact in a simple way: when the relevance ρ is sufficiently high ($\rho > \rho^v$), only verification is used to disincentivize manipulation behavior, and disclosure becomes irrelevant; while when the relevance ρ is not high enough ($\rho \leq \rho^v$), only disclosure policy is used to disincentivize the manipulation behavior and verification technology is never used.

7 Conclusion

I study the optimal algorithmic disclosure in a lending market. FinTech lenders use privately observed predictive algorithms to help make lending decisions. The input of the predictive algorithm is the data collected from borrowers, which is subject to a strategic manipulation problem. In the optimal public disclosure, the information about the predictive algorithm should be partially disclosed to the borrowers, which differentiates the posterior lending market equilibira by data manipulation levels. Under the optimal disclosure policy, lenders use borrower data less intensively in their lending decisions which decreases the average data manipulation level and improves efficiency.

There are some potential directions for future research. First, in my model, I abstract away the screening channel using contracts in the lending market. The joint design of information and contract will be a natural question for future research. Second, the feature in my model is a binary variable, and it will be interesting to consider a model with a general space for input. In a general model, all types of borrowers may signal their types by costly manipulation, and the interaction between the signaling and information design is also interesting. Thirdly, this paper mainly focuses on efficiency but not the distribution of surplus. Since fairness is also a crucial part of the regulator's objective, it will be interesting to consider the optimal algorithmic disclosure that achieves a particular surplus distribution. Finally, all lenders use the same predictive algorithm in my model, but it is natural to consider the setting where lenders use different but correlated algorithms, and in this case, algorithmic disclosure may change the lending market structure.

References

Ball, I. (2019). Scoring strategic agents. *arXiv preprint arXiv:1909.01888*.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2021). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*.

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

Bhatt, U., A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley (2020). Explainable machine learning in deployment. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 648–657.

Björkegren, D., J. E. Blumenstock, and S. Knight (2020). Manipulation-proof machine learning. *arXiv preprint arXiv:2004.03865*.

Blattner, L., S. Nelson, and J. Spiess (2021). Unpacking the black box: Regulating algorithmic decisions. *arXiv preprint arXiv:2110.03443*.

Bogen, M. and A. Rieke (2018). Help wanted: An examination of hiring algorithms, equity, and bias.

Bruckner, M. A. (2018). The promise and perils of algorithmic lenders' use of big data. *Chi.-Kent L. Rev. 93*, 3.

Carvalho, D. V., E. M. Pereira, and J. S. Cardoso (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics 8*(8), 832.

Coston, A., A. Rambachan, and A. Chouldechova (2021). Characterizing fairness over the set of good models under selective labels. *arXiv preprint arXiv:2101.00352*.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael (2021). Invisible primes: Fintech lending with alternative data. *Available at SSRN 3937438*.

Dworczak, P. and G. Martini (2019). The simple economics of optimal persuasion. *Journal of Political Economy 127*(5), 1993–2048.

Frankel, A. and N. Kartik (2019a). Improving information from manipulable data. *Journal of the European Economic Association*.

Frankel, A. and N. Kartik (2019b). Muddled information. *Journal of Political Economy 127*(4), 1739–1776.

Gentzkow, M. and E. Kamenica (2016). A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review 106*(5), 597–601.

Gillis, T. B. and J. L. Spiess (2019). Big data and discrimination. *The University of Chicago Law Review 86*(2), 459–488.

Goldstein, I. and Y. Leitner (2018). Stress tests and information disclosure. *Journal of Economic Theory 177*, 34–69.

Goldstein, I. and Y. Leitner (2020). Stress tests disclosure: Theory, practice, and new perspectives.

Huang, J. (2020). *Optimal stress tests in financial networks*. Ph. D. thesis, Duke University.

Inostroza, N. (2019). Persuading multiple audiences: An information design approach to banking regulation. *Available at SSRN 3450981*.

Inostroza, N. and A. Pavan (2021). Persuasion in global games with application to stress testing.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics 11*, 249–272.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kizilcec, R. F. and H. Lee (2020). Algorithmic fairness in education. *arXiv preprint arXiv:2007.05443*.

Kleinberg, J., J. Ludwig, S. Mullainathan, and A. Rambachan (2018). Algorithmic fairness. In *Aea papers and proceedings*, Volume 108, pp. 22–27.

Leitner, Y. and B. Williams (2020). Model secrecy and stress tests. *Available at SSRN 3606654*.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777.

Malenko, A., N. Malenko, and C. S. Spatt (2021). Creating controversy in proxy voting advice. *Available at SSRN 3843674*.

Milone, M. (2019). Smart lending. *Unpublished working paper*.

Murdoch, W. J., C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences 116*(44), 22071–22080.

Raghavan, M., S. Barocas, J. Kleinberg, and K. Levy (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 469–481.

Szydlowski, M. (2021). Optimal financing and disclosure. *Management Science 67*(1), 436–454.

Wang, Q., Y. Huang, S. Jasin, and P. V. Singh (2020). Algorithmic transparency with strategic users. *Available at SSRN 3652656*.

Ziółkowski, M. (2009). Generalization of probability density of random variables. *Scientific Issues of Jan Długosz University in Częstochowa. Mathematics 14*, 163–172.

Appendix

A When All Borrowers Can Manipulate

In this section, we consider the case that all borrowers can manipulate their features $\hat{\theta}$, while keeping the space of feature to be binary. In this section, we show that in any lending market equilibrium, the good type borrowers never manipulate their features.

Suppose any borrower can manipulate his characteristic by paying cost c_i, i.e., a good (bad) type borrower can change his characteristic to $\hat{\theta} = \hat{B}$ ($\hat{\theta} = \hat{G}$) by paying cost c_i, which follows a continuous distribution $D^{\theta}(c)$, for $\theta \in \{G, B\}$. So the distribution of manipulation cost is type-dependent. Similar to our baseline model, c_i is observable to borrowers but not to lenders. Under posterior belief $\pi(\rho|s)$, for borrowers with type $\theta \in \{G, B\}$, denote lender j's lending decision as $\left\{I_j^{s,\hat{\theta}}(\rho), D_j^{s,\hat{\theta}}(\rho)\right\}$, where $I_j^{s,\hat{\theta}}(\rho) \in [0,1]$ represents the probability that lender j approves the loan applications from borrowers with feature $\hat{\theta}$ conditional on the true state being ρ and the signal disclosed to borrowers being s; and $D_j^{s,\hat{\theta}}(\rho)$ represents the face value of debt that lender j offers to borrowers with feature $\hat{\theta}$, conditional on the true state being ρ and the signal disclosed to borrowers being s. We only focus on symmetric equilibria.

Similar to the baseline model, in this extension, the borrowers' manipulation strategy is summarized by a cutoff $\bar{c}^{\theta}$, such that borrowers with type θ will choose to manipulate if and only if their manipulation cost type c_i is no greater than $\bar{c}^{\theta}$. The following lemma shows that the good type borrowers will never manipulate in any subgame equilibrium, so our assumption in the baseline model that good type borrowers are not able to manipulate is without loss of generality.

Lemma A.1. *Given $\pi(\rho|s)$, in any equilibrium, $I_j^{s,\hat{B}}(\rho) = 0$ for all j and $\rho \in supp(\pi(\rho|s))$, and no good type borrower chooses to manipulate, i.e., $\bar{c}^G = 0$.*

Proof. Suppose the posterior belief is $\pi(\rho|s)$, and in equilibrium, all lenders choose $\left\{I_j^{s,\hat{\theta}}(\rho), D_j^{s,\hat{\theta}}(\rho)\right\}$. Then the fraction of borrowers with different types and features are summarized in the following table:

	$\hat{\theta} = \hat{G}$	$\hat{\theta} = \hat{B}$
$\theta = G$	$\mu\left(1 - F_c\left(\bar{c}^G\right)\right)$	$\mu F_c\left(\bar{c}^G\right)$
$\theta = B$	$(1-\mu) F_c\left(\bar{c}^B\right)$	$(1-\mu)\left(1 - F_c\left(\bar{c}^B\right)\right)$

.

Table 1: Fraction of borrowers.

In equilibrium, lenders will lend to borrowers with feature $\hat{\theta} = \hat{G}$ if and only if $\rho \geq \varrho^G$, where ϱ^G is solved by

$$\mu\left(1 - F_c\left(\bar{c}^G\right)\right) m\left(\varrho^G\right) - \left[\mu\left(1 - F_c\left(\bar{c}^G\right)\right) + (1-\mu) F_c\left(\bar{c}^B\right)\right] I = 0. \tag{23}$$

And for all $\rho \geq \varrho^G$, the equilibrium debt contract $D^{s,\hat{G}}(\rho)$ is solved by

$$\mathbb{E}\left(\min\left\{\tilde{V}, D^{s,\hat{G}}(\rho)\right\} | s, \rho, \theta = G\right) - \frac{\mu\left(1 - F_c\left(\bar{c}^G\right)\right) + (1-\mu) F_c\left(\bar{c}^B\right)}{\mu\left(1 - F_c\left(\bar{c}^G\right)\right)} I = 0. \tag{24}$$

Similarly, lenders will lend to borrowers with feature $\hat{\theta} = \hat{B}$ if and only if $\rho \geq \varrho^B$, where ϱ^B is solved by

$$m\left(\varrho^B\right) - \frac{\mu F_c\left(\bar{c}^G\right) + (1-\mu)\left(1 - F_c\left(\bar{c}^B\right)\right)}{\mu F_c\left(\bar{c}^G\right)} I = 0. \tag{25}$$

And for all $\rho \geq \varrho^B$, the equilibrium debt contract $D^{s,\hat{B}}(\rho)$ is solved by

$$\mathbb{E}\left(\min\left\{\tilde{V}, D^{s,\hat{B}}(\rho)\right\} | s, \rho, \theta = G\right) - \frac{F_c\left(\bar{c}^G\right) + (1-\mu)\left(1 - F_c\left(\bar{c}^B\right)\right)}{\mu F_c\left(\bar{c}^G\right)} I = 0. \tag{26}$$

If $\varrho^G \geq \varrho^B$, from (23) and (25), we have

$$\frac{\mu\left(1 - F_c\left(\bar{c}^G\right)\right) + (1-\mu) F_c\left(\bar{c}^B\right)}{\mu\left(1 - F_c\left(\bar{c}^G\right)\right)} \geq \frac{\mu F_c\left(\bar{c}^G\right) + (1-\mu)\left(1 - F_c\left(\bar{c}^G\right)\right)}{\mu F_c\left(\bar{c}^G\right)}.$$

However, in this case, for all the bad type borrowers, it's strictly profitable not to manipulate, which means in equilibrium we must have $\bar{c}^B = 0$, and thus

$$\frac{\mu\left(1 - F_c\left(\bar{c}^G\right)\right) + (1-\mu) F_c\left(\bar{c}^B\right)}{\mu\left(1 - F_c\left(\bar{c}^G\right)\right)} = 1 < \frac{\mu F_c\left(\bar{c}^B\right) + (1-\mu)}{\mu F_c\left(\bar{c}^B\right)},$$

a contradiction!

If $\varrho^G < \varrho^B$, by (23) and (25), we must have

$$\frac{\mu\left(1 - F_c\left(\bar{c}^G\right)\right) + (1-\mu) F_c\left(\bar{c}^B\right)}{\mu\left(1 - F_c\left(\bar{c}^G\right)\right)} < \frac{\mu F_c\left(\bar{c}^G\right) + (1-\mu)\left(1 - F_c\left(\bar{c}^G\right)\right)}{\mu F_c\left(\bar{c}^G\right)},$$

substitute this condition into (24) and (26), we can show that, for any $\rho \geq \varrho^B$ we must have

$$D^{s,\hat{B}}(\rho) > D^{s,\hat{G}}(\rho).$$

Then for all good type borrowers, manipulating is strictly dominated by not manipulating, and thus in equilibrium we must have $\bar{c}^G = 0$.

So in this equilibrium we must have $\bar{c}^G = 0$, which is the same as our baseline model. Then allowing all borrowers to manipulate their features doesn't change the equilibrium for any posterior belief $\pi(\rho, s)$, and thus it doesn't change our results. $\square$

B Proofs

B.1 Proofs in Section 2

No Disclosure On The Relevance ρ

Based on the distributional assumptions on the manipulation cost c and the relevance, the two equilibrium conditions are

$$B\left(1-\varrho_N\right)=\bar{c}_N$$

and

$$\mu\varrho_N V=\left(\mu+(1-\mu)\bar{c}_N\right)I.$$

The unique solution is

$$\left(\varrho_N=\frac{I}{\mu V+(1-\mu)I},\bar{c}_N=B\cdot\frac{\mu(V-I)}{\mu V+(1-\mu)I}\right).$$

Full Transparency

In this case, when $\rho\geq\frac{I}{V}$, the surplus from lending to $\hat{G}$ borrowers must be zero. To see this, if lenders lend to $\hat{G}$ borrowers for sure, since $B\geq 1$, all of the bad type borrowers must choose to manipulate their features. In this case, it's not profitable to lend to $\hat{G}$ borrowers for any $\rho<1$ because $\mu V\leq I$, a contradiction!

B.2 Proof of Proposition 4.1

The no disclosure policy is implemented by a signal $(\mathcal{S},\tilde{\sigma})$ with only one element in the signal space $\mathcal{S}=\{s_N\}$, and the mapping $\tilde{\sigma}\left(s|\rho\right)$ is trivial. The lending market equilibrium is characterized by $(k_N,\varrho_N,\bar{c}_N)$ which satisfy conditions in Definition 3.3 under the prior belief of ρ. Let the regulator's payoff be W_N in the no disclosure case. Now let's consider the following deterministic disclosure policy $(\mathcal{S}',\sigma')$, where $\mathcal{S}'=\{s_1',s_2'\}$, and

$$\sigma'\left(\rho\right)=\begin{cases}s_1' & \rho\in[0,\rho_1']\cup[\varrho_N,\varrho_N+\epsilon_1]\cup[1-\epsilon_2,1]\\ s_2' & \rho\in(\rho_1',\varrho_N)\cup(\varrho_N+\epsilon_1,1-\epsilon_2)\end{cases}$$

where $\rho_1'<\varrho_N$ satisfies

$$\frac{\text{Prob}\left(\rho\in[\varrho_N,\varrho_N+\epsilon_1]\cup[1-\epsilon_2,1]\right)}{\text{Prob}\left(\rho\in[0,\rho_1']\cup[\varrho_N,\varrho_N+\epsilon_1]\cup[1-\epsilon_2,1]\right)}=\bar{c}_N.$$

Denote the equilibria under signals s_1' and s_2' as $\left(k_{s_1'},\varrho_{s_1'},\bar{c}_{s_1'}\right)$ and $\left(k_{s_2'},\varrho_{s_2'},\bar{c}_{s_2'}\right)$, respectively, it's easy to verify

$$\left(k_{s_1'},\varrho_{s_1'},\bar{c}_{s_1'}\right)=\left(k_{s_2'},\varrho_{s_2'},\bar{c}_{s_2'}\right)=\left(k_N,\varrho_N,\bar{c}_N\right).$$

Then introducing the policy $(\mathcal{S}', \sigma')$ doesn't change the regulator's payoff, i.e., $W_N = W'$, where the regulator's payoff under disclosure policy $(\mathcal{S}', \sigma')$ can be written as

$$\begin{aligned} W' =& \text{Prob}\left(\rho \in [0, \rho_1'] \cup [\varrho_N, \varrho_N + \epsilon_1] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1'}\right\}}\right] \\ &+ \text{Prob}\left(\rho \in (\rho_1', \varrho_N) \cup (\varrho_N + \epsilon_1, 1-\epsilon_2)\right) \cdot \mathbb{E}^{s_2'}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_2'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_2'}\right\}}\right] \end{aligned} \tag{27}$$

Then, let's construct a new disclosure policy based on $(\mathcal{S}', \sigma')$, and show that the new disclosure policy increases regulator's payoff. Let's consider the deterministic disclosure policy $(\mathcal{S}'', \sigma'')$, with $S'' = \{s_1'', s_2'', s_3''\}$, and

$$\sigma'' = \begin{cases} s_1'' & \rho \in [\varrho_N, \varrho_N + \epsilon_1] \\ s_2'' & \rho \in [0, \rho_1'] \cup [1-\epsilon_2, 1] \\ s_3'' & \rho \in (\rho_1', \varrho_N) \cup (\varrho_N + \epsilon_1, 1-\epsilon_2) \end{cases}.$$

The signal realization s_3'' is "equivalent" to the signal realization s_2' in disclosure policy $(\mathcal{S}', \sigma')$, both induce the same posterior belief in $(\rho_1', \varrho_N) \cup (\varrho_N + \epsilon_1, 1-\epsilon_2)$. The difference is that policy $(\mathcal{S}'', \sigma'')$ further reveals if the true state is in $[\varrho_N, \varrho_N + \epsilon_1]$ or not. Note that the regulator's payoff in state $[\varrho_N, \varrho_N + \epsilon_1]$ is close to zero in the no disclosure case, as ϱ_N is the equilibrium cutoff in lending decisions. So revealing this information only changes the regulator's payoff marginally in states $\rho \in [\varrho_N, \varrho_N + \epsilon_1]$. However, the increase in regulator's payoff is non-trivial. Note that the approval probability is lower under s_2'' compared to the no disclosure case, so the equilibrium data manipulation level is lower under s_2''. As what we will show later, this is the dominating effect thus the regulator's payoff increases under the disclosure policy $(\mathcal{S}'', \sigma'')$. To see this, note that the regulator's payoff under $(\mathcal{S}'', \sigma'')$ is

$$\begin{aligned} W'' =& \text{Prob}\left(\rho \in [\varrho_N, \varrho_N + \epsilon_1]\right) \cdot \mathbb{E}^{s_1''}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1''}\right\}}\right] \\ &+ \text{Prob}\left(\rho \in [0, \rho_1'] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_2''}\right\}}\right] \\ &+ \text{Prob}\left(\rho \in (\rho_1', \varrho_N) \cup (\varrho_N + \epsilon_1, 1-\epsilon_2)\right) \cdot \mathbb{E}^{s_3''}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_3''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_3''}\right\}}\right] \end{aligned} \tag{28}$$

It's obvious that the last term in (28) is equal to the last term in (27), because equilibria under signal realizations s_3'' and s_2' are the same. Then

$$\begin{aligned}
& W'' - W' \\
=& \text{Prob}\left(\rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right) \cdot \mathbb{E}^{s_1''}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1''}\right\}}\right] \\
& + \text{Prob}\left(\rho \in [0, \rho_1'] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_2''}\right\}}\right] \\
& - \text{Prob}\left(\rho \in [0, \rho_1'] \cup [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1'}\right\}}\right] \\
\geq& \text{Prob}\left(\rho \in [0, \rho_1'] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_2''}\right\}}\right] \\
& - \text{Prob}\left(\rho \in [0, \rho_1'] \cup [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1'}\right\}}\right]
\end{aligned}$$

Note that $\rho_{s_1'} = \underline{\rho}_N$, we know

$$\begin{aligned}
& \text{Prob}\left(\rho \in [0, \rho_1'] \cup [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1] \cup [1-\epsilon_2, 1]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{\rho \geq \rho_{s_1'}\right\}}\right] \\
=& \begin{array}{c} \text{Prob}\left(\rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right) \cdot \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_N\right)\right) I\right) | \rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right] \\ +\text{Prob}\left(\rho \in [1-\epsilon_2, 1]\right) \cdot \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_N\right)\right) I\right) | \rho \in [1-\epsilon_2, 1]\right] \end{array}
\end{aligned}$$

Then

$$\begin{aligned}
W'' - W' =& \text{Prob}\left(\rho \in [1-\epsilon_2, 1]\right) \cdot \left[(1-\mu) I \left(F_c\left(\bar{c}_N\right) - F_c\left(\bar{c}_{s_2''}\right)\right)\right] \\
& - \text{Prob}\left(\rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right) \cdot \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c\left(\bar{c}_N\right)\right) I\right) | \rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right].
\end{aligned}$$

In the equilibrium of subgame s_1'',

$$\begin{aligned}
\bar{c}_{s_2''} =& \frac{\text{Prob}\left(\rho \in [1-\epsilon_2, 1]\right)}{\text{Prob}\left(\rho \in [0, \rho_1'] \cup [1-\epsilon_2, 1]\right)} \\
=& \frac{\bar{c}_N - x}{1 - x}
\end{aligned}$$

where

$$x = \frac{\text{Prob}\left(\rho \in [\underline{\rho}_N, \underline{\rho}_N + \epsilon_1]\right) \cdot (1 - \bar{c}_N)}{\text{Prob}\left(\rho \in [0, \rho_1']\right)}.$$

Consider the case when fixing ϵ_2, and let $\epsilon_1 \to 0$, then $x \to 0$ and $\bar{c}_{s_2''} = \bar{c}_N - x(1-\bar{c}_N) + o(x)$.

$$
\begin{aligned}
& W'' - W' \\
=& \text{Prob}\left(\rho \in [\varrho_N, \varrho_N + \epsilon_1]\right) \cdot \left[\begin{array}{c} \text{Prob}\left(\rho \in [1-\epsilon_2, 1]\right) \cdot \frac{(1-\mu) I \left(F_c(\bar{c}_N) - F_c\left(\bar{c}_{s_2''}\right)\right)}{\text{Prob}\left(\rho \in \left[\varrho_N, \varrho_N + \epsilon_1\right]\right)} - \\ \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c(\bar{c}_N)\right) I\right) | \rho \in [\varrho_N, \varrho_N + \epsilon_1]\right] \end{array} \right] \\
\approx& \text{Prob}\left(\rho \in [\varrho_N, \varrho_N + \epsilon_1]\right) \cdot \left[\begin{array}{c} \frac{\text{Prob}(\rho \in [1-\epsilon_2, 1])(1-\bar{c}_N)}{\text{Prob}\left(\rho \in \left[0, \rho_1'\right]\right)} \cdot \frac{(1-\mu) I (F_c(\bar{c}_N) - F_c(\bar{c}_N - x(1-\bar{c}_N)))}{x} - \\ \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c(\bar{c}_N)\right) I\right) | \rho \in [\varrho_N, \varrho_N + \epsilon_1]\right] \end{array} \right] \\
\approx& \text{Prob}\left(\rho \in [\varrho_N, \varrho_N + \epsilon_1]\right) \cdot \left[\begin{array}{c} \frac{\text{Prob}(\rho \in [1-\epsilon_2, 1])(1-\bar{c}_N)^2 (1-\mu) I}{\text{Prob}\left(\rho \in \left[0, \rho_1'\right]\right)} \cdot F_c'(\bar{c}_N) - \\ \mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c(\bar{c}_N)\right) I\right) | \rho \in [\varrho_N, \varrho_N + \epsilon_1]\right] \end{array} \right].
\end{aligned}
$$

Since $\epsilon_1 \to 0$, we must have

$$
\mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c(\bar{c}_N)\right) I\right) | \rho \in [\varrho_N, \varrho_N + \epsilon_1]\right] \to 0,
$$

because the equilibrium condition in the no disclosure case is

$$
\mathbb{E}\left[\left(\mu \tilde{V} - \left(\mu + (1-\mu) F_c(\bar{c}_N)\right) I\right) | \rho = \varrho_N\right] = 0.
$$

With $F_c'(\bar{c}_N) > 0$, then we must have

$$
W'' - W' > 0
$$

which means that the no disclosure policy is dominated by our new disclosure policy $(\mathcal{S}'', \sigma'')$.

B.3 Proof of Lemma 5.1

The full disclosure policy $(\mathcal{S}, \sigma)$ can be implemented by space $\mathcal{S} = [0, 1]$ and a deterministic message function $\sigma(\rho) = \rho$. In this case, the true state ρ is perfectly revealed to the public. Denote ϱ as the solution of

$$
m(\varrho) = I,
$$

For any $s = \rho > m^{-1}(I)$, the lending market equilibrium of subgame s, $(k_s, \varrho_s, \bar{c}_s)$, must satisfy $\varrho_s = \rho$, and thus

$$
\mu m(\rho) - k_s = 0. \tag{29}
$$

To see this, suppose $\mu(\rho) > k_s$, in equilibrium all $\hat{G}$ borrowers must be approved, and all bad type borrowers must choose to manipulate because of Assumption 3. Then the regulator's payoff of financing all $\hat{G}$ borrowers is

$$
\mu m(\rho) - I \leq \mu m(1) - I \leq 0,
$$

and the equality holds only when $\rho = 1$. As as result, lenders will not lend to $\hat{G}$ borrowers for all $\rho < 1$, a contradiction. So in equilibrium condition (29) must hold. And this condition implies that the regulator's payoff is zero.

Next, it's obvious that when $\rho \leq \underline{\rho}$, lender will never lend to any borrowers. In summary, regulator's payoff is zero for any $s \in \mathcal{S}$ thus the regulator's total payoff is $W_F = 0$ under full disclosure policy.

B.4 Proof of Lemma 5.2

The results are directly derived from the definition of lending market equilibrium. For any two equlibria $(k_{s_1}, \underline{\rho}_{s_1}, \bar{c}_{s_1})$ and $(k_{s_2}, \underline{\rho}_{s_2}, \bar{c}_{s_2})$, the first condition in Definition 3.3

$$\mu m \left(\underline{\rho}_s\right) = k_s$$

implies

$$k_{s_1} \geq k_{s_2} \Longleftrightarrow \underline{\rho}_{s_1} \geq \underline{\rho}_{s_2}, \tag{30}$$

because $m(\cdot)$ is an increasing function. The third condition

$$k_s = \left[\mu + (1-\mu) F_c \left(\bar{c}_s\right)\right] I$$

implies that

$$k_{s_1} \geq k_{s_2} \Longleftrightarrow \bar{c}_{s_1} \geq \bar{c}_{s_2}. \tag{31}$$

Then (30) and (31) complete the proof.

B.5 Proof of Lemma 5.3

We just need to verify that the regulator's payoff is unchanged under the new disclosure policy $(\mathcal{S}', \tilde{\sigma}')$. Notice that

$$\tilde{\sigma}' (s|\rho) = \tilde{\sigma} (s|\rho)$$

for any $\rho \in [0,1]$ and $s \in \mathcal{S} \backslash \{s_1, s_2\} = \mathcal{S}' \backslash \{s_0'\}$. Then for any $s \in \mathcal{S} \backslash \{s_1, s_2\} = \mathcal{S}' \backslash \{s_0'\}$, the posterior beliefs are the same under the two policies, i.e., for any $s \in \mathcal{S} \backslash \{s_1, s_2\} = \mathcal{S}' \backslash \{s_0'\}$, we have

$$\pi (\rho|s) = \pi' (\rho|s).$$

So the lending market equilibira are the same for any $s \in \mathcal{S} \backslash \{s_1, s_2\} = \mathcal{S}' \backslash \{s_0'\}$ in these two policies. Besides, the regulator's payoff from signal realization s_0' in the new disclosure policy is the sum of that under signal realizations s_1 and s_2 in policy $(\mathcal{S}, \sigma)$, this is because the lending market equilibrium under s_0, s_1 and s_2 are all the same, and the probability of observing s_0' in the new policy is the sum of the probabilities of observing s_1 and s_2 in policy $(\mathcal{S}, \tilde{\sigma})$. Since policy $(\mathcal{S}, \tilde{\sigma})$ is optimal, the new policy $(S', \tilde{\sigma}')$ must also be optimal.

B.6 Proof of Lemma 5.4

Given any policy $(\mathcal{S}, \tilde{\sigma})$ with distribution of posteriors $\{f(s), \pi(\rho|s)\}_{s\in\mathcal{S}}$, for any subgame s,

$$\bar{c}_s = 0 \Longleftrightarrow \rho_s = m^{-1}(I) \Longleftrightarrow k_s = I.$$

In this equilibrium, there is no manipulation, and lenders always reject all loan applications. The posterior belief must satisfy

$$\sup\{\text{supp}(\pi(\rho|s))\} \leq m^{-1}(I).$$

Since $m(1) > I$, there must exist at least one signal realization s_1, such that

$$\bar{c}_{s_1} > 0.$$

Suppose there also exists another signal realization s_2, such that

$$\bar{c}_{s_2} = 0.$$

Here assume both the probabilities of s_1 and s_2 are positive[26], then let's consider a new policy $(\mathcal{S}', \sigma')$ with distribution of posteriors $\{f'(s), \pi'(\rho|s)\}_{s\in\mathcal{S}'}$, signal space $\mathcal{S}' = \{s_0'\} \cup \mathcal{S}\setminus\{s_1, s_2\}$, and

$$\tilde{\sigma}'(s|\rho) = \tilde{\sigma}(s|\rho)\,\mathbb{1}_{\mathcal{S}\setminus\{s_1,s_2\}}(s) + (\tilde{\sigma}(s_1|\rho) + \tilde{\sigma}(s_2|\rho))\,\mathbb{1}_{\{s_0'\}}(s).$$

Obviously, any signal realization $s \in \mathcal{S}\setminus\{s_1, s_2\}$ must exist in the signal spaces of both disclosure polices, and induce the same lending market equilibrium. Besides, for signal realization s_0' in $(\mathcal{S}', \sigma')$ and $\{s_1.s_2\}$ in $(\mathcal{S}, \sigma)$, we have

$$f'(s_0) = f(s_1) + f(s_2),$$

and

$$\pi'(\rho|s_0') = \frac{1}{f(s_1) + f(s_2)}(f(s_1)\pi(\rho|s_1) + f(s_2)\pi(\rho|s_2)).$$

The equilibrium conditions in Definition 3.3 implies that $\bar{c}_s$ satisfies

$$\text{Prob}\left(\rho \geq m^{-1}\left(\frac{[\mu + (1-\mu)F_c(\bar{c}_s)]I}{\mu}\right)|s\right) \geq \frac{\bar{c}_s}{B} \geq \text{Prob}\left(\rho > m^{-1}\left(\frac{[\mu + (1-\mu)F_c(\bar{c}_s)]I}{\mu}\right)|s\right)$$

Note that

$$\Pi(\rho|s_0') > \Pi(\rho|s_1)$$

[26]The proof for the case when the signal is continuous is similar, in that case, we just need to deal with density functions.

for any $\rho > m^{-1}(I)$, and both $m^{-1}(\cdot)$ and $F_c(\cdot)$ are increasing functions, we conclude that

$$\bar{c}_{s'_0} < \bar{c}_{s_1}.$$

Since the regulator's payoff is always zero for any $\rho \leq m^{-1}(I)$, the difference of regulator's payoffs under $(\mathcal{S}', \sigma')$ and $(\mathcal{S}, \sigma)$ is

$$\begin{aligned} W' - W = & \begin{array}{c} f(s'_0)\,\mathbb{E}^{s'_0}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)F_c\left(\bar{c}_{s'_0}\right)\right)I\right)\mathbb{1}_{\left(\varrho_{s'_0},1\right]\cap\mathrm{supp}\left(\pi\left(\rho|s'_0\right)\right)}(s)\right] - \\ f(s_1)\,\mathbb{E}^{s_1}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)F_c\left(\bar{c}_{s_1}\right)\right)I\right)\mathbb{1}_{\left(\varrho_{s_1},1\right]\cap\mathrm{supp}\left(\pi\left(\rho|s_1\right)\right)}(s)\right] \end{array} \\ = & \begin{array}{c} f(s_1)\,\mathbb{E}^{s_1}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)F_c\left(\bar{c}_{s'_0}\right)\right)I\right)\mathbb{1}_{\left(\varrho_{s'_0},1\right]\cap\mathrm{supp}\left(\pi\left(\rho|s'_0\right)\right)}(s)\right] - \\ f(s_1)\,\mathbb{E}^{s_1}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)F_c\left(\bar{c}_{s_1}\right)\right)I\right)\mathbb{1}_{\left(\varrho_{s_1},1\right]\cap\mathrm{supp}\left(\pi\left(\rho|s_1\right)\right)}(s)\right] \end{array}. \end{aligned}$$

Since $\bar{c}_{s_0} < \bar{c}_{s_1}$ and $m^{-1}(I) < \varrho_{s'_0} < \varrho_{s_1}$, we have

$$\left(\varrho_{s'_0}, 1\right] \cap \mathrm{supp}\left(\pi\left(\rho|s'_0\right)\right) = \left(\varrho_{s_1}, 1\right] \cap \mathrm{supp}\left(\pi\left(\rho|s_1\right)\right),$$

and thus

$$W' - W > 0.$$

B.7 Proof of Lemma 5.5

Suppose there exists an optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$ and it induces the distribution of posteriors $\{f(s), \pi(\rho|s)\}_{s\in\mathcal{S}}$. If $\mathcal{S}$ is a singleton, then the policy is simply the no information policy. In this case, let $\varrho^\star = \varrho_N$ and this lemma is obviously true. Otherwise, if the statement is not true, there must exist $s_1, s_2 \in \mathcal{S}$, with lending market equilibria $(k_{s_1}, \varrho_{s_1}, \bar{c}_{s_1})$ and $(k_{s_2}, \varrho_{s_2}, \bar{c}_{s_2})$, such that

$$\rho_1 > \rho_2,$$

$$\rho_1 \in [0, \varrho_{s_1}] \cap \mathrm{supp}\left(\pi\left(\rho|s_1\right)\right)$$

and

$$\rho_2 \in (\varrho_{s_2}, 1] \cap \mathrm{supp}\left(\pi\left(\rho|s_2\right)\right).$$

Then there must exist intervals B_1, B_2[27], such that

$$\rho_1 \in B_1 \subset [0, \varrho_{s_1}] \cap \mathrm{supp}\left(\pi\left(\rho|s_1\right)\right),$$

[27]Note that a single point is also a closed interval.

$$\rho_2 \in B_2 \subset (\varrho_{s_2}, 1] \cap \operatorname{supp}(\pi(\rho|s_2)),$$

$$\inf B_1 > \sup B_2,$$

$$\operatorname{Prob}(B_1|s_1) = K_1 > 0,$$

and

$$\operatorname{Prob}(B_2|s_2) = K_2 > 0.$$

For this proof, let's assume both s_1 and s_2 occur with positive probability, then $f(s_1)$ and $f(s_s)$ represent the associated probabilities. The other case is when either s_1 or s_2 occur with zero probability and $f(s_1)$ or $f(s_2)$ represent the density functions. The proof strategy is basically the same.

In this case, if $f(s_1) K_1 \geq f(s_2) K_2$, then let's consider the following distribution of posteriors: $\left\{\hat{f}(s), \hat{\pi}(\rho|s)\right\}_{s\in\tilde{\mathcal{S}}}$, where $\hat{\mathcal{S}} = \mathcal{S}$, $\hat{f}(s) = f(s)$, and

$$\hat{\pi}(\rho|s) = \begin{cases} \pi\{\rho|s_1\} + \frac{f(s_2)}{f(s_1)}\pi\{\rho|s_2\}\mathbb{1}_{B_2}(\rho) - \frac{f(s_2)K_2}{f(s_1)K_1}\pi\{\rho|s_1\}\mathbb{1}_{B_1} & \text{if } s = s_1 \\ \pi\{\rho|s_2\} - \pi\{\rho|s_2\}\mathbb{1}_{B_2} + \frac{K_2}{K_1}\pi\{\rho|s_1\}\mathbb{1}_{B_1} & \text{if } s = s_2 \\ \pi(\rho|s) & o.w. \end{cases}$$

We can check that $\left\{\hat{f}(s), \hat{\pi}(\rho|s)\right\}_{s\in\tilde{\mathcal{S}}}$ is Bayes-plausible, and there exists a disclosure policy $\left(\hat{S}, \hat{\sigma}\right)$ that induces this distribution of posteriors. But now in the new policy $\left(\hat{S}, \hat{\sigma}\right)$, $\rho_2 \notin \operatorname{supp}(\pi(\rho|s_2))$. And the regulator's payoff is weakly increasing under the new policy $\left(\hat{\mathcal{S}}, \hat{\sigma}\right)$ because

1. $\hat{f}(s) = f(s)$for all $s \in \hat{\mathcal{S}} = \mathcal{S}$;

2. the lending market equilibria $(k_s, \varrho_s, \bar{c}_s)$ are the same under the two policies for any $s \in \mathcal{S} = \hat{\mathcal{S}}$;

3. the regulator's payoff under any signal realizations except for s_2 is unchanged;

4. the regulator's payoff under signal realization s_2 increases.

The last point holds because with the new disclosure policy $\left(\hat{S}, \hat{\sigma}\right)$, under the signal realization s_2, the equilibrium variables $(k_{s_2}, \varrho_{s_2}, \bar{c}_{s_2})$ is the same compared to that with policy $(\mathcal{S}, \sigma)$, so the total financing cost is unchanged, which is k_s. But the total payoff generated from projects increases by

$$f(s_2) \cdot K_2 \cdot [\mathbb{E}[\mu m(\rho)|s_1, B_1] - \mathbb{E}[\mu m(\rho)|s_2, B_2]]$$

which is positive because $\inf B_1 > \sup B_2$.

B.8 Proof of Lemma 5.6

For the optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$, if $\mathcal{S}$ is a singleton, this lemma is obviously true. Otherwise, there exist two different signals s_1 and s_2 with probabilities(densities) $\tilde{f}(s_1)$ and $\tilde{f}(s_2)$, respectively. For simplicity, let's assume that both $\tilde{f}(s_1)$ and $\tilde{f}(s_2)$ are positive, the proof for other cases are basically the same. Denote the lending market equilibrium variables as $\left(\tilde{k}_{s_1}, \tilde{\varrho}_{s_1}, \tilde{\bar{c}}_{s_1}\right)$ and $\left(\tilde{k}_{s_2}, \tilde{\varrho}_{s_2}, \tilde{\bar{c}}_{s_2}\right)$ under these two signals. Without loss of generality, let's assume $\tilde{\varrho}_{s_1} < \tilde{\varrho}_{s_2}$. Denote the ex ante lending cutoff as $\varrho^\star$ in this case. Suppose for s_1, s_2, the condition

$$\sup\left\{\operatorname{supp}\left(\tilde{\pi}\left(\rho|s_1\right)\right) \cap \left(\varrho^\star, 1\right]\right\} \leq \inf\left\{\operatorname{supp}\left(\tilde{\pi}\left(\rho|s_2\right)\right) \cap \left(\varrho^\star, 1\right]\right\} \tag{32}$$

is not satisfied, let

$$B = \left[\inf\left\{\operatorname{supp}\left(\tilde{\pi}\left(\rho|s_2\right)\right) \cap \left(\varrho^\star, 1\right]\right\}, \sup\left\{\operatorname{supp}\left(\tilde{\pi}\left(\rho|s_1\right)\right) \cap \left(\varrho^\star, 1\right]\right\}\right].$$

Then there must exist two non-negative functions v_1, v_2, such that

$$\tilde{f}(s_1)\, v_1(\rho) + \tilde{f}(s_2)\, v_2(\rho) = \tilde{f}(s_1)\, \tilde{\pi}\{\rho|s_1\} \cdot \mathbb{1}_B(\rho) + \tilde{f}(s_2)\, \tilde{\pi}\{\rho|s_2\} \cdot \mathbb{1}_B(\rho), \tag{33}$$

$$\sup\left\{\operatorname{supp}\left(v_1(\rho)\right) \cap \left(\varrho^\star, 1\right]\right\} \leq \inf\left\{\operatorname{supp}\left(v_2(\rho)\right) \cap \left(\varrho^\star, 1\right]\right\}$$

and

$$\int v_1(\rho)\, d\rho = \int \pi\{\rho|s_1\} \cdot \mathbb{1}_B(\rho)\, d\rho. \tag{34}$$

Now let's consider the following distribution of posterior beliefs with signal space $\mathcal{S}$: $\left\{\hat{f}(s), \hat{\pi}(\rho|s)\right\}_{s\in\mathcal{S}}$, where $\hat{f}(s) = f(s)$ and

$$\hat{\pi}(\rho|\tilde{s}) = \begin{cases} \tilde{\pi}\{\rho|s_1\} - \tilde{\pi}\{\rho|s_1\} \cdot \mathbb{1}_B(\rho) + v_1(\rho) & \text{if } \tilde{s} = s_1 \\ \tilde{\pi}\{\rho|s_2\} - \tilde{\pi}\{\rho|s_2\}\, \mathbb{1}_B(\rho) + v_2(\rho) & \text{if } \tilde{s} = s_2 \\ \tilde{\pi}(\rho|s) & o.w. \end{cases}.$$

We can check that the new distribution of posteriors $\left\{\hat{f}(s), \hat{\pi}(\rho|s)\right\}_{s\in\mathcal{S}}$ is still Bayes-plausible, because

$$\int v_1(\rho)\, d\rho = \int \tilde{\pi}\{\rho|s_1\} \cdot \mathbb{1}_B(\rho)\, d\rho$$

and

$$\int v_2(\rho)\, d\rho = \int \tilde{\pi}\{\rho|s_2\} \cdot \mathbb{1}_B(\rho)\, d\rho.$$

The second condition is a direct result of (33) and (34). And we can check that $\left\{\hat{f}(s), \hat{\pi}(\rho|s)\right\}_{s\in\mathcal{S}}$ can be induced by a disclosure policy $(\mathcal{S}, \hat{\sigma})$. Now the condition (32) is not violated anymore in the new policy. Then we just need to show that the regulator's payoff is unchanged under the new policy, and thus it is still optimal. To see this, with policy $(\mathcal{S}, \tilde{\sigma})$, we know under posterior belief $\tilde{\pi}(\rho|s_1)$

$$\mathrm{Prob}^{(\mathcal{S},\tilde{\sigma})}(\rho > \underline{\rho}_{s_1}|s_1) = \frac{\bar{c}_{s_1}}{B}.$$

Note that since $\underline{\rho}_{s_1} < \underline{\rho}_{s_2}$, we know

$$\inf\{\mathrm{supp}(\pi(\rho|s_2)) \cap (\underline{\rho}^\star, 1]\} \geq \underline{\rho}_{s_2} > \underline{\rho}_{s_1}.$$

Then for any $\rho \in B$, we must have $\rho > \underline{\rho}_{s_1}$. Then under posterior belief $\hat{\pi}(\rho|s_1)$, we know

$$\begin{aligned}\mathrm{Prob}^{(\mathcal{S},\hat{\sigma})}(\rho > \underline{\rho}_{s_1}|s_1) &= \mathrm{Prob}^{(\mathcal{S},\tilde{\sigma})}(\rho > \underline{\rho}_{s_1}|s_1) - \int \tilde{\pi}\{\rho|s_1\} \cdot \mathbb{1}_B(\rho)\, d\rho + \int v_1(\rho)\, d\rho \\ &= \mathrm{Prob}^{(\mathcal{S},\tilde{\sigma})}(\rho > \underline{\rho}_{s_1}|s_1) \\ &= \frac{\bar{c}_{s_1}}{B}.\end{aligned}$$

The second equality comes from condition (34). Based on this, we can check all other equilibrium conditions are also satisfied, and this implies $\left(\hat{\phi}_{s_1}, \hat{\rho}_{s_1}, \hat{\Delta}_{s_1}\right) = (\phi_{s_1}, \rho_{s_1}, \Delta_{s_1})$. Similarly, we can check $\left(\tilde{k}_s, \tilde{\underline{\rho}}_s, \tilde{\bar{c}}_s\right) = (k_s, \underline{\rho}_s, \bar{c}_s)$. For all other $s \in \mathcal{S}\backslash\{s_1, s_2\}$, it's obvious that the lending market equilibria are all the same under these two disclosure policies. Then we can easily show that the regulator's payoff is the same under those two policies.

The proof strategy still works if condition (34) is not satisfied in the optimal policy $(\mathcal{S}, \sigma)$. Besides, note that the third property in Lemma 5.6 implies the second property in Lemma 5.6, and these two jointly imply that the disclosure policy must be deterministic. Since all the posterior lending market equilibira are the same, the ex ante lending cutoff $\underline{\rho}^\star$ must be unchanged.

B.9 Proof of Theorem 5.1

Lemma 5.6 shows that for any optimal policy, there exists a deterministic optimal policy $(\mathcal{S}, \sigma)$ that induces almost equivalent lending market equilibria. Our Criterion 2 implies that for any two distinct signal realizations $s_1, s_2 \in \mathcal{S}$, we must have

$$\bar{c}_{s_1} \neq \bar{c}_{s_2}.$$

Then we consider a new signal space $\mathcal{S}' = [\bar{c}_{\min}, \bar{c}_{\max}]$, where

$$\bar{c}_{\min} = \inf_{s\in\mathcal{S}}\{\bar{c}_s\}$$

and

$$\bar{c}_{\max} = \sup_{s\in\mathcal{S}}\{\bar{c}_s\},$$

and a message function

$$\sigma'(\rho) = \bar{c}_{\sigma(\rho)}.$$

Then obviously $(\mathcal{S}', \sigma')$ is also a deterministic optimal policy, with the same lending market equilibira as $(\mathcal{S}, \sigma)$. And the cutoff $\rho^\star$ will be the same under these two optimal policies. For any two signals $s'_1, s'_2 \in \mathcal{S}'$, and $s'_1 < s'_2$ where both $\sigma'^{-1}(s'_1)$ and $\sigma'^{-1}(s'_2)$ are nonempty. Based on the construction of the new policy, we must have

$$\sup\left\{\sigma^{-1}(s'_1) \cap [0, \rho^\star]\right\} \leq \inf\left\{\sigma^{-1}(s'_2) \cap [0, \rho^\star]\right\}$$

and

$$\sup\left\{\tilde{\sigma}^{-1}(s'_1) \cap (\rho^\star, 1]\right\} \leq \inf\left\{\tilde{\sigma}^{-1}(s'_2) \cap (\rho^\star, 1]\right\}.$$

This means that for any $\rho \in [0, \rho^\star]$ or $\rho \in (\rho^\star, 1]$, $\sigma'(\rho)$ is a weakly increasing function, with $\inf \sigma'(\rho) = \bar{c}_{\min}$ and $\sup \sigma'(\rho) = \bar{c}_{\max}$.

B.10 Proof of Proposition 5.1

For any optimal policy, Lemma 5.6 shows that there exists another optimal policy that has the same ex ante lending cutoff $\rho^\star$, the same lending market equilibria, and satisfies conditions (18) and (19). So without loss of generality, we just need to focus on optimal policies that satisfy conditions (18) and (19). Let's introduce the following lemmas to establish our results.

Lemma B.1. *For any two posterior beliefs $\pi(\rho|s_1)$ and $\pi(\rho|s_2)$, with positive probabilities (densities) $f(s_1)$ and $f(s_2)$, and lending market equilibira $(k_{s_1}, \rho_{s_1}, \bar{c}_{s_1})$ and $(k_{s_2}, \rho_{s_2}, \bar{c}_{s_2})$ satisfying $\rho_{s_1} < \rho_{s_2}$. Let $\hat{s}$ be the "combined"signal with posterior belief*

$$\pi(\rho|\hat{s}) = \frac{f(s_1)\pi(\rho|s_1) + f(s_2)\pi(\rho|s_2)}{f(s_1) + f(s_2)},$$

Then the lending market equilibrium $(k_{\hat{s}}, \rho_{\hat{s}}, \bar{c}_{\hat{s}})$ satisfies

$$k_{s_1} < k_{\hat{s}} < k_{s_2},$$

$$\rho_{s_1} < \rho_{\hat{s}} < \rho_{s_2}$$

and

$$\bar{c}_{s_1} < \bar{c}_{\hat{s}} < \bar{c}_{s_2}.$$

Proof. First, it's impossible to have $\rho_{\hat{s}} \leq \rho_{s_1}$. Note that for the equilibria under s_1 and s_2, the equilibrium conditions are[28]

$$\mu m(\rho_{s_1}) = [\mu + (1-\mu) F_c(\bar{c}_{s_1})] I$$

[28]This is solved by the equilibrium conditions in Definition 3.3.

and

$$\mu m\left(\rho_{s_2}\right)=\left[\mu+(1-\mu)F_c\left(\bar{c}_{s_2}\right)\right]I.$$

For $\hat{s}$, we have

$$\mu m\left(\rho_{\hat{s}}\right)=\left[\mu+(1-\mu)F_c\left(\bar{c}_{\hat{s}}\right)\right]I.$$

If $\rho_{\hat{s}}\leq\rho_{s_1}$, Lemma 5.2 implies

$$\bar{c}_{\hat{s}}\leq\bar{c}_{s_1}<\bar{c}_{s_2}.$$

Following the equilibrium conditions in Definition 3.3, we have

$$\begin{aligned}\frac{\bar{c}_{\hat{s}}}{B}&=\mathbb{E}\left(I^{\hat{s}}|\hat{s}\right)\\&=\frac{f\left(s_1\right)}{f\left(s_1\right)+f\left(s_2\right)}\mathbb{E}\left(I^{\hat{s}}|s_1\right)+\frac{f\left(s_2\right)}{f\left(s_1\right)+f\left(s_2\right)}\mathbb{E}\left(I^{\hat{s}}|s_2\right).\end{aligned}\tag{35}$$

Since

$$\rho_{\hat{s}}\leq\rho_{s_1}<\rho_{s_2},$$

we must have

$$I^{\hat{s}}\left(\rho\right)\geq I^{s_1}\left(\rho\right)\geq I^{s_2}\left(\rho\right)$$

for all $\rho\in[0,1]$. Then

$$\mathbb{E}\left(I^{\hat{s}}|s_1\right)\geq\mathbb{E}\left(I^{s_1}|s_1\right)=\frac{\bar{c}_{s_1}}{B}$$

and

$$\mathbb{E}\left(I^{\hat{s}}|s_1\right)\geq\mathbb{E}\left(I^{s_2}|s_1\right)=\frac{\bar{c}_{s_2}}{B}.$$

Then condition (35) implies

$$\frac{\bar{c}_{\hat{s}}}{B}\geq\frac{f\left(s_1\right)}{f\left(s_1\right)+f\left(s_2\right)}\frac{\bar{c}_{s_1}}{B}+\frac{f\left(s_2\right)}{f\left(s_1\right)+f\left(s_2\right)}\frac{\bar{c}_{s_2}}{B}>\frac{\bar{c}_{s_1}}{B}\Rightarrow\bar{c}_{\hat{s}}>\bar{c}_{s_1},$$

contradiction!

The same proof strategy works for the case $\rho_{\hat{s}}\geq\rho_{s_2}$. So the equilibrium must satisfy

$$\rho_{s_1}<\rho_{\hat{s}}<\rho_{s_2}.$$

□

The following lemma provides an intermediate results about the structure of the deterministic optimal policy characterized in Lemma 5.6.

Lemma B.2. *Suppose $(\mathcal{S}, \sigma)$ is a deterministic optimal policy, then for almost any $s \in \mathcal{S}$, if there exists a constant $\epsilon_s > 0$, such that*

$$\operatorname{supp}\left(\pi\left(\rho|s\right)\right) \cap \left(\underline{\rho}_s - \epsilon_s, \underline{\rho}_s\right) = \emptyset \tag{36}$$

and

$$Prob\left(\rho \leq \underline{\rho}_s - \epsilon_s|s\right) > 0, \tag{37}$$

then there must exist a constant $\delta_s > 0$, such that

$$\operatorname{supp}\left(\pi\left(\rho|s\right)\right) \cap \left[\underline{\rho}_s, \underline{\rho}_s + \delta_s\right] = \emptyset.$$

Proof. Suppose for the sake of contradiction that there exists $s_0 \in \mathcal{S}$ satisfying conditions (36) and (37), and at least one of the following two scenarios is true:

1. There exists a constant δ, such that for any $0 < x < \delta$,

   $$\left(\underline{\rho}_{s_0}, \underline{\rho}_{s_0} + x\right) \subset \operatorname{supp}\left(\pi\left(\rho|s_0\right)\right) \cap \left(\underline{\rho}^{\star}, 1\right]$$

   and

   $$\operatorname{Prob}\left(\rho \in \left(\underline{\rho}_{s_0}, \underline{\rho}_{s_0} + x\right)|s_0\right) > 0.$$

2. The first condition doesn't hold and $\operatorname{Prob}\left(\rho = \underline{\rho}_{s_0}|s_0\right) > 0$.

If the first scenario is true, then let's consider another deterministic disclosure policy $(\mathcal{S}', \sigma')$ with signal space $\mathcal{S}' = \mathcal{S} \backslash \{s_0\} \cup \{s_0', s'\}$, and

$$\sigma'\left(\rho\right) = \begin{cases} \sigma\left(\rho\right) & \text{if } \rho \notin [0,1] \setminus \operatorname{supp}\left(\pi\left(\rho|s_0\right)\right) \\ s_0' & \text{if } \rho \in \operatorname{supp}\left(\pi\left(\rho|s_0\right)\right) \setminus \left(\underline{\rho}_{s_0}, \underline{\rho}_{s_0} + x\right) \\ s' & \text{if } \rho \in \left(\underline{\rho}_{s_0}, \underline{\rho}_{s_0} + x\right) \end{cases},$$

where $x < \delta$.

The increase in regulator's payoff under this new policy is

$$
\begin{aligned}
&\Delta W \\
=&\text{Prob}\left(\rho \in (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot \mathbb{E}^{s_1',G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_1'}\right)\right)\right) \cdot \mathbb{1}_{\left\{\rho \geq \varrho_{s_1'}\right\}}\right] \\
&+ \text{Prob}\left(\rho \in \text{supp}\left(\pi\left(\rho | s_0\right)\right) \setminus (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot \mathbb{E}^{s_0',G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0'}\right)\right)\right) \cdot \mathbb{1}_{\left\{\rho \geq \varrho_{s_0'}\right\}}\right] \\
&- \text{Prob}\left(\rho \in \text{supp}\left(\pi\left(\rho | s_0\right)\right)\right) \cdot \mathbb{E}^{s_0,G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0}\right)\right)\right) \cdot \mathbb{1}_{\left\{\rho \geq \varrho_{s_0}\right\}}\right] \\
\geq&\text{Prob}\left(\rho \in \text{supp}\left(\pi\left(\rho | s_0\right)\right) \setminus (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot \mathbb{E}^{s_0',G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0'}\right)\right)\right) \cdot \mathbb{1}_{\left\{\rho \geq \varrho_{s_0'}\right\}}\right] \\
&- \text{Prob}\left(\rho \in \text{supp}\left(\pi\left(\rho | s_0\right)\right)\right) \cdot \mathbb{E}^{s_0,G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0}\right)\right)\right) \cdot \mathbb{1}_{\left\{\rho \geq \varrho_{s_0}\right\}}\right] \\
=&\text{Prob}\left(\rho \in (\varrho^\star, 1] \cap \text{supp}\left(\pi\left(\rho | s_0\right)\right) \setminus (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot (1-\mu)\left[F_c\left(\bar{c}_{s_0}\right) - F_c\left(\bar{c}_{s_0'}\right)\right] \\
&- \text{Prob}\left(\rho \in (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot \mathbb{E}^{G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0}\right)\right)\right) | \rho \in (\varrho_{s_0}, \varrho_{s_0}+x)\right].
\end{aligned}
$$

Similar to the proof of Proposition 4.1, we know

$$\text{Prob}\left(\rho \in (\varrho^\star, 1] \cap \text{supp}\left(\pi\left(\rho | s_0\right)\right) \setminus (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot (1-\mu)\left[F_c\left(\bar{c}_{s_0}\right) - F_c\left(\bar{c}_{s_0'}\right)\right] = O(x)$$

and

$$\text{Prob}\left(\rho \in (\varrho_{s_0}, \varrho_{s_0}+x)\right) \cdot \mathbb{E}^{G}\left[\left(\mu\tilde{V} - \left(\mu + (1-\mu)\, F_c\left(\bar{c}_{s_0}\right)\right)\right) | \rho \in (\varrho_{s_0}, \varrho_{s_0}+x)\right] = O(x^2).$$

So we must have

$$\Delta W > 0$$

when x is sufficiently small, and this means the new disclosure policy generates higher regulator's payoff, a contradiction!

If the second scenario is true, the same proof strategy applies and we can also find a disclosure policy (not deterministic) that generates higher regulator's payoff, a contradiction. □

The next lemma in this proof presents a property about the "worse" subgame (the subgame with highest data manipulation level in equilibrium).

Lemma B.3. *For any deterministic optimal policy $(\mathcal{S}, \sigma)$ that satisfies properties in Lemma 5.6, we must have*

$$\sup_{s \in \mathcal{S}} \varrho_s = \varrho^\star. \tag{38}$$

Proof. First, we must have $\sup_{s\in\mathcal{S}} \varrho_s \geq \varrho^\star$, otherwise, there exists $\delta > 0$, such that for any s, $\varrho_s < \varrho^\star - \delta$. If this is true, consider any state $\rho \in (\varrho^\star - \delta, \varrho^\star)$. Since $\varrho_s < \varrho^\star - \delta$ for any s, lenders will finance $\hat{G}$ borrowers when the true state $\rho \in (\varrho^\star - \delta, \varrho^\star)$. However, by our definition of $\varrho^\star$, lenders will reject all loan applications if the state $\rho < \varrho^\star$, a contradiction. So we must have $\sup_{s\in\mathcal{S}} \varrho_s \geq \varrho^\star$.

Now we want to show it's impossible to have $\sup_{s\in\mathcal{S}} \varrho_s > \varrho^\star$. Suppose for the sake of contradiction that there exists $\delta > 0$, such that

$$\sup_{s\in\mathcal{S}} \varrho_s > \varrho^\star + \delta.$$

Then there must exist a signal realization, denoted as s_m, such that $\varrho_{s_m} > \varrho^\star + \delta$. From Lemma B.2, we know that there exists an interval $(\varrho_{s_m}, \varrho_{s_m} + \epsilon_m)$ such that

$$(\varrho_{s_m}, \varrho_{s_m} + \epsilon_m) \cap \operatorname{supp}(\pi(\rho|s_m)) = \emptyset,$$

and

$$\operatorname{Prob}(\rho \in (\varrho_{s_m}, \varrho_{s_m} + \epsilon_m)) < \operatorname{Prob}([\varrho_{s_m} + \epsilon_m, 1] \cap \operatorname{supp}(\pi(\rho|s_m))).$$

Then we can find an interval B, and a one to one mapping

$$z : (\varrho_{s_m}, \varrho_{s_m} + \epsilon_m) \to B,$$

with $z'(x) \equiv 1$, such that

$$B \subseteq [\varrho_{s_m} + \epsilon_m, 1] \cap \operatorname{supp}(\pi(\rho|s_m)).$$

Now let's consider the following deterministic disclosure policy $(\mathcal{S}', \sigma')$ with $\mathcal{S}' = \mathcal{S}$, and

$$\sigma'(\rho) = \begin{cases} \sigma(\rho) & \text{if } \rho \notin B \cup (\varrho_{s_m}, \varrho_{s_m} + \epsilon_m) \\ \sigma(z(\rho)) & \text{if } \rho \in (\varrho_{s_m}, \varrho_{s_m} + \epsilon_m) \\ \sigma(z^{-1}(\rho)) & \text{if } \rho \in B \end{cases}.$$

It's easy to check that all lending market equilibria are unchanged. Then the regulator's payoff is unchanged. However, under the new disclosure policy, for the signal realization s_m, we have

$$\rho_{s_m} = \inf\{\operatorname{supp}(\pi(\rho|s_m)) \cap (\varrho^\star, 1]\}.$$

But this violates Lemma B.2, a contradiction. So it's impossible to have

$$\sup_{s\in\mathcal{S}} \varrho_s > \varrho^\star,$$

and thus we must have

$$\sup_{s\in\mathcal{S}} \varrho_s = \varrho^\star.$$

□

Then Lemma 5.1 is a direct result of Lemma B.1 and B.3. Suppose for the sake of contradiction that $\varrho^\star \le \varrho_N$, then Lemma B.3 implies that

$$\varrho_s \le \varrho_N$$

for all $s \in \mathcal{S}$. Note that the signal in no information case is a "combined" signal of all signals in the optimal policy $(\mathcal{S}, \sigma)$, Lemma B.1 implies that

$$\varrho_N < \sup_{s \in \mathcal{S}} \varrho_s,$$

a contradiction! So we must have

$$\varrho^\star > \varrho_N.$$

B.11 Proof of Proposition 5.2

Suppose the deterministic optimal policy is $(\mathcal{S}, \sigma)$. Since there are at least two signals in the optimal policy, we must have

$$\bar{c}_{\max} > \bar{c}_{\min}.$$

Note that the no disclosure is the "combined" signal of optimal policy $(\mathcal{S}, \sigma)$. Then Lemma B.1 implies that

$$\bar{c}_{\max} > \bar{c}_N > \bar{c}_{\min}.$$

B.12 Proof of Proposition 5.3

Consider an optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$ with distribution of posteriors $\{f(s), \pi(\rho|s)\}_{s \in \mathcal{S}}$. Since the prior belief of ρ is a continuous distribution, for any s_0 satisfying

$$\text{Prob}(s_0) > \epsilon,$$

the equilibrium conditions imply that there must exist $\epsilon_{s_0} > 0$ and $\delta_{s_0} > 0$, such that

$$\text{Prob}(\rho \le \varrho^\star - \epsilon_{s_0} | s_0) = \delta_{s_0} > 0.$$

Let $M = \frac{\bar{c}_{s_0}}{B}$. Denote T as the solution of

$$\frac{T \cdot \text{Prob}(\rho \ge \varrho^\star | s_0)}{\delta_{s_0}} = \frac{M}{1 - M}.$$

Then let's consider a new signal space

$$\mathcal{S}_a = \mathcal{S} \backslash \{s_0\} \cup \{s_{a1}, s_{a2}\}$$

and a distribution of posteriors $\{f_a(s), \pi_a(\rho|s)\}_{s\in\mathcal{S}_a}$ where

$$f_a(s) = \begin{cases} f(s) & \text{if } s \in \mathcal{S}\setminus\{s_0\} \\ \frac{\delta_{s_0}}{1-M} f(s_0) & \text{if } s = s_{a1} \\ \left(1 - \frac{\delta_{s_0}}{1-M}\right) f(s_0) & \text{if } s = s_{a2} \end{cases},$$

and

$$\pi_a(\rho|s) = \begin{cases} \pi(\rho|s) & \text{if } s \in \mathcal{S}\setminus\{s_0\} \\ \frac{1-M}{\delta_{s_0}} \left[\pi(\rho|s_0)\,\mathbb{1}_{[0,\varrho^\star-\epsilon_{s_0}]}(\rho) + T\cdot\pi(\rho|s_0)\,\mathbb{1}_{[\varrho^\star,1]}(\rho)\right] & \text{if } s = s_{a1} \\ \frac{1}{1-\frac{\delta_{s_0}}{1-M}} \left[\pi(\rho|s_0)\,\mathbb{1}_{(\varrho^\star-\epsilon_{s_0},\varrho^\star)}(\rho) + (1-T)\cdot\pi(\rho|s_0)\,\mathbb{1}_{[\varrho^\star,1]}(\rho)\right] & \text{if } s = s_{a2} \end{cases}.$$

We can check the distribution of posteriors $\{f_a(s), \pi_a(\rho|s)\}_{s\in\mathcal{S}_a}$ is still Bayes-plausible, and there exists a disclosure policy that can induce this distribution of posteriors. Besides, we can check that the equilibrium variables $\{k_s, \varrho_s, \bar{c}_s\}$ are all the same for equilibria under signal s_0, s_{a1} and s_{a2}. Then by Lemma B.2[29], there must exists $t_s > 0$, such that

$$\text{supp}\left(\pi\left(\rho|s_{a2}\right)\right) \cap \left[\varrho_{s_0}, \varrho_{s_0} + t_s\right] = \emptyset,$$

which implies

$$\text{supp}\left(\pi\left(\rho|s_0\right)\right) \cap \left[\varrho_{s_0}, \varrho_{s_0} + t_s\right] = \emptyset$$

because of our construction of π_a. Then the surplus from lending must be greater than

$$\mu\left(m\left(\varrho_{s_0} + \frac{t_s}{2}\right) - m\left(\varrho_{s_0}\right)\right) > 0$$

for any $\rho > \varrho^\star$ in this posterior equilibrium s_0.

B.13 Proof of Theorem 5.2

The proof of Theorem 5.2 is established by three lemmas.

Lemma B.4. *(Pooling at the bottom) When Assumption 4 is satisfied, in any deterministic optimal policy $(\mathcal{S}, \sigma)$ characterized in Theorem 5.1, there must exist $\epsilon > 0$, such that for any $\rho_1, \rho_2 \in (0, \epsilon) \cup (\varrho^\star, \varrho^\star + \epsilon)$, we have $\sigma(\rho_1) = \sigma(\rho_2)$.*

[29] Although Lemma B.2 only considers deterministic optimal policies, it can be shown that it also holds for general optimal policies.

Proof. Suppose $(\mathcal{S},\sigma)$ is a deterministic optimal policy characterized in Theorem 5.1. Note Lemma B.3 implies $\varrho_s \leq \varrho^\star$ for all s, then let

$$\mathcal{S}_1 = \left\{ s \middle| \sup\left\{\operatorname{supp}\left(\pi\left(\rho|s\right)\right) \cap \left(0,\varrho^\star\right)\right\} < m^{-1}\left(I\right) \& \varrho_s < \frac{1}{2}\left(m^{-1}\left(I\right)+\varrho^\star\right) \right\},$$

$$B_1 = \cup_{s\in\mathcal{S}_1} \operatorname{supp}\left(\pi\left(\rho|s\right)\right),$$

and

$$\bar{c}_1 = \sup\left\{\bar{c}_s | s \in \mathcal{S}_1\right\}.$$

Suppose for the sake of contradiction that it doesn't satisfy the *Pooling at the bottom* property. Then there are infinite elements in $\mathcal{S}_1$. Regulator's ex ante surplus from all $s \in \mathcal{S}_1$ is

$$\begin{aligned}
\tilde{W}_1 &= \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot \mathbb{E}^s\left[\left(\mu m\left(\rho\right) - \left(\mu + \left(1-\mu\right)F_c\left(\bar{c}_s\right)\right)I\right)^+\right] ds \\
&= \begin{array}{l} \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot \operatorname{Prob}\left(\rho > \varrho^\star|s\right)\cdot \mu\mathbb{E}^s\left(m\left(\rho\right) - I|\rho>\varrho^\star\right) ds - \\ \quad \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot \operatorname{Prob}\left(\rho > \varrho^\star|s\right)\cdot\left(1-\mu\right)F_c\left(\bar{c}_s\right) I ds \end{array} \\
&= \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\frac{\bar{c}_s}{B}\cdot \mu\mathbb{E}^s\left(m\left(\rho\right)-I|\rho>\varrho^\star\right) ds - \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\frac{\bar{c}_s}{B}\cdot\left(1-\mu\right)F_c\left(\bar{c}_s\right) I ds, \\
&= \operatorname{Prob}\left(\rho\in\left(\varrho^\star,1\right]\cap B_1\right)\mu\mathbb{E}\left(m\left(\rho\right)-I|\rho\in\left(\varrho^\star,1\right]\cap B_1\right) - \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\frac{\bar{c}_s}{B}\cdot\left(1-\mu\right)F_c\left(\bar{c}_s\right) I ds
\end{aligned}$$

Here we use the equilibrium condition $\operatorname{Prob}\left(\rho > \varrho^\star|s\right) = \frac{\bar{c}_s}{B}$ in the last equality. Then we show that the regulator's payoff increases under another disclosure policy that satisfies the *Pooling at the bottom* property. To see this, in the above equilibrium,

$$\begin{aligned}
\int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\frac{\bar{c}_s}{B}\cdot ds &= \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\operatorname{Prob}\left(\rho>\varrho^\star|s\right)\cdot ds \\
&= \operatorname{Prob}\left(\rho\in\left(\varrho^\star,1\right]\cap B_1\right).
\end{aligned}$$

Let $\bar{c}_0$ be the solution of

$$\left(\int_{s\in\mathcal{S}_1} f\left(s\right)\cdot ds\right)\frac{\bar{c}_0}{B} = \operatorname{Prob}\left(\rho\in\left(\varrho^\star,1\right]\cap B_1\right) = \int_{s\in\mathcal{S}_1} f\left(s\right)\cdot\frac{\bar{c}_s}{B}\cdot ds,$$

obviously $\bar{c}_0 < \sup_{s\in\mathcal{S}_1}\bar{c}_s$. Based on Assumption 4, and using the concavification method (Kamenica and Gentzkow (2011)), we know there exist $\bar{c}_1 \leq \bar{c}_2 \leq \sup_{s\in\mathcal{S}_1}\bar{c}_s$, and two positive numbers p_1, p_2 satisfying $p_1+p_2=1$, such that

$$p_1 + p_2 = 1,$$

$$p_1\frac{\bar{c}_1}{B} + p_2\frac{\bar{c}_2}{B} = \frac{\bar{c}_0}{B},$$

and

$$\left(\int_{s\in\mathcal{S}_1} f(s)\cdot ds\right)(1-\mu)\left[p_1\frac{\bar{c}_1}{B}F_c(\bar{c}_1)+p_2\frac{\bar{c}_2}{B}F_c(\bar{c}_2)\right]<\int_{s\in\mathcal{S}_1} f(s)\cdot\frac{\bar{c}_s}{B}\cdot(1-\mu)F_c(\bar{c}_s)\,ds. \tag{39}$$

Here $\bar{c}_1$ and $\bar{c}_2$ represent the equilibrium data manipulation cutoffs for two signals $\hat{s}_1$ and $\hat{s}_2$. From the ex ante perspective. The regulator's payoff from financing good projects are unchanged for all states $\rho\in B_1$, while the ex ante surplus loss from financing bad projects decreases with the binary signals $\hat{s}_1$ and $\hat{s}_2$ because of condition (39). Then the new disclosure policy with signals $\hat{s}_1$ and $\hat{s}_2$ improves regulator's payoff, and this policy satisfies the *Pooling at the bottom* property. □

Then we want to show that there exists at most one discrete signal s satisfying $\text{Prob}(s)>0$. To get this result, let first provide an intermediate result:

Lemma B.5. *Suppose $(\mathcal{S},\sigma)$ is a deterministic optimal policy, then for any $s\in\mathcal{S}$ such that $\text{Prob}(s)>0$, function $xF_c(x)$ can not be strictly concave at $x=\bar{c}_s$.*

Proof. Suppose $(\mathcal{S},\sigma)$ is a deterministic optimal policy, and there exists $s_0\in\mathcal{S}$ such that $\text{Prob}(s_0)>0$. Suppose for the sake of contradiction that $xF_c(x)$ is strictly concave at $x=\bar{c}_{s_0}$. In this equilibrium, since $\text{Prob}(s_0)>0$, there must exists $\epsilon_0>0$ and $\delta_0>0$ such that

$$\text{Prob}\left(\rho\in(0,\varrho_{s_0}-\epsilon_0)\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right)\right)>\delta_0$$

and

$$\text{Prob}\left(\rho\in(\varrho_{s_0}+\epsilon_0,1)\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right)\right)>\delta_0.$$

Then there exists two sets L_1 and R_1, such that

$$L_1\subset(0,\varrho_{s_0}-\epsilon_0)\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right),$$
$$R_1\subset(\varrho_{s_0}+\epsilon_0,1)\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right),$$

$$\text{Prob}(L_1)>0,$$
$$\text{Prob}(R_1)>0,$$

and

$$\frac{\text{Prob}(L_1)}{\text{Prob}(R_1)}=\frac{\text{Prob}\left((0,\varrho_{s_0})\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right)\right)}{\text{Prob}\left(\left(\varrho^{\star}_{s_0},1\right)\cap\text{supp}\left(\pi\left(\rho|s_0\right)\right)\right)}.$$

Then consider the following deterministic policy with signal space $\mathcal{S}'=\mathcal{S}\backslash\{s_0\}\cup\{s_1',s_2'\}$ and message function

$$\sigma'(\rho)=\begin{cases}\sigma(\rho) & \text{if } \rho\notin\text{supp}\left(\pi\left(\rho|s_0\right)\right)\\ s_1' & \text{if } \rho\in L_1\cup R_1\\ s_2' & \text{if } \rho\in\text{supp}\left(\pi\left(\rho|s_0\right)\right)\backslash\left(L_1\cup R_1\right)\end{cases}.$$

The lending market equilibria under s'_1 and s'_2 are the same as the equilibrium under s_0 with equilibrium variables $\left(k_{s_0}, \varrho^\star_{s_0}, \bar{c}_{s_0}\right)$. However we can improve the regulator's payoff from states $\rho \in L_1 \cup R_1$ by disclosing additional information based on s'_1. Since $xF_c(x)$ is strictly concave at $x = \bar{c}_{s_0}$, there exists $\bar{\epsilon}_2 > 0$, such that for all $\epsilon < \bar{\epsilon}_2$, there exist two numbers $\bar{c}_1$ and $\bar{c}_2$ satisfying $\bar{c}_1, \bar{c}_2 \in (\bar{c}_{s_0} - \epsilon, \bar{c}_{s_0} + \epsilon)$ and two positive numbers p_1, p_2 satisfying $p_1 + p_2 = 1$ such that

$$p_1 \bar{c}_1 + p_2 \bar{c}_2 = \bar{c}_{s_0},$$

and

$$p_1 \bar{c}_1 F_c(\bar{c}_1) + p_2 \bar{c}_2 F_c(\bar{c}_2) < \bar{c}_{s_0} F_c(\bar{c}_{s_0}). \tag{40}$$

Let ϱ_1 and ϱ_2 be

$$\varrho_1 = m^{-1}\left(\frac{\mu + (1-\mu) F_c(\bar{c}_1)}{\mu} I\right)$$

and

$$\varrho_2 = m^{-1}\left(\frac{\mu + (1-\mu) F_c(\bar{c}_2)}{\mu} I\right),$$

then we can choose ϵ small enough, such that

$$\varrho_1, \varrho_2 \in (\varrho_{s_0} - \epsilon_0, \varrho_{s_0} + \epsilon_0).$$

Then there must exist L_{11}, L_{12}, R_{11}, R_{12} such that

$$L_{11} \cup L_{12} = L_1$$

$$R_{11} \cup R_{12} = R_1,$$

$$\frac{\text{Prob}(\rho \in R_{11})}{\text{Prob}(\rho \in R_{11}) + \text{Prob}(\rho \in L_{11})} = \frac{\bar{c}_1}{B}$$

and

$$\frac{\text{Prob}(\rho \in R_{12})}{\text{Prob}(\rho \in R_{12}) + \text{Prob}(\rho \in L_{12})} = \frac{\bar{c}_2}{B}.$$

Then let's consider the following deterministic policy with signal space $\mathcal{S}'_1 = \mathcal{S} \setminus \{s_0\} \cup \{s'_2\} \cup \{s'_{11}, s'_{12}\}$ and message function

$$\sigma'_1(\rho) = \begin{cases} \sigma(\rho) & \text{if } \rho \notin \text{supp}(\pi(\rho|s_0)) \\ s'_{11} & \text{if } \rho \in L_{11} \cup R_{11} \\ s'_{12} & \text{if } \rho \in L_{12} \cup R_{12} \\ s'_2 & \text{if } \rho \in \text{supp}(\pi(\rho|s_0)) \setminus (L_1 \cup R_1) \end{cases}.$$

It can be verified that, compared to the disclosure policy $(\mathcal{S},\sigma)$, the regulator's payoff is unchanged under the new policy $(\mathcal{S}_1',\sigma_1')$ for all states $\rho\in[0,1]\setminus(L_1\cup R_1)$. And for states $\rho\in L_1\cup R_1$, the regulator's payoff under $(\mathcal{S},\sigma)$ is

$$\begin{aligned}W_0 &= \text{Prob}\left(\rho\in R_1\right)\mu\mathbb{E}\left(m\left(\rho\right)-I|\rho\in R_1\right)-\text{Prob}\left(\rho\in R_1\right)\left(1-\mu\right)F_c\left(\bar{c}_{s_0}\right)I\\ &= \text{Prob}\left(\rho\in R_1\right)\mu\mathbb{E}\left(m\left(\rho\right)-I|\rho\in R_1\right)-\text{Prob}\left(\rho\in L_1\cup R_1\right)\frac{\bar{c}_{s_0}}{B}\left(1-\mu\right)F_c\left(\bar{c}_{s_0}\right)I\end{aligned}$$

while the regulator's payoff under $(\mathcal{S}_1',\sigma_1')$ is

$$W_{12}=\begin{array}{c}\text{Prob}\left(\rho\in R_1\right)\mu\mathbb{E}\left(m\left(\rho\right)-I|\rho\in R_1\right)-\text{Prob}\left(\rho\in L_{11}\cup R_{11}\right)\frac{\bar{c}_1}{B}\left(1-\mu\right)F_c\left(\bar{c}_1\right)I\\ -\text{Prob}\left(\rho\in L_{12}\cup R_{12}\right)\frac{\bar{c}_2}{B}\left(1-\mu\right)F_c\left(\bar{c}_2\right)I\end{array}.$$

Since

$$\text{Prob}\left(\rho\in L_{11}\cup R_{11}\right)+\text{Prob}\left(\rho\in L_{12}\cup R_{12}\right)=\text{Prob}\left(\rho\in L_1\cup R_1\right),$$

condition (40) implies that

$$W_{12}>W_0,$$

which implies that the regulator's payoff under $(\mathcal{S}_1',\sigma_1')$ is greater than her payoff under $(\mathcal{S},\sigma)$, a contradiction! □

Lemma B.6. *For any deterministic optimal policy characterized in Theorem 5.1, there exists a payoff-equivalent deterministic optimal policy $(\mathcal{S},\sigma)$, such that there exists only one $s\in\mathcal{S}$ that satisfies Prob$(s)>0$.*

Proof. Suppose $(\mathcal{S},\sigma)$is a deterministic optimal policy, and suppose for the sake of contradiction that there exists two signals s_1, $s_2\in\mathcal{S}$, such that

$$\text{Prob}\left(s_1\right)>0,$$

and

$$\text{Prob}\left(s_2\right)>0.$$

Denote the equilibrium variables under these two signals are $(k_{s_1},\rho_{s_1},\bar{c}_{s_1})$ and $(k_{s_2},\rho_{s_2},\bar{c}_{s_2})$, respectively. Without loss of generality assume $\bar{c}_{s_1}<\bar{c}_{s_2}$. Using the proof techniques in Lemma B.5, we can create two signals s_1' and s_2' based on s_1 and s_2, such that equilibrium under s_1' (s_2') is the same as the equilibrium under s_1 (s_2), and there exists two constant $\epsilon>0$, such that

$$\text{supp}\left(\pi\left(\rho|s_1'\right)\right)\cap\left(\rho_{s_1}-\epsilon,\rho_{s_1}+\epsilon\right)=\emptyset$$

and

$$\text{supp}\left(\pi\left(\rho|s_2'\right)\right)\cap\left(\rho_{s_2}-\epsilon,\rho_{s_2}+\epsilon\right)=\emptyset.$$

Lemma B.5 implies that function $xF_c(x)$ is weakly convex at both $\bar{c}_{s_1}$ and $\bar{c}_{s_2}$.

If the function $xF_c(x)$ is convex on $[\bar{c}_{s_1}, \bar{c}_{s_2}]$. Then for any $\delta > 0$ that is small enough, there exists positive numbers p_1, p_2, $\bar{c}_1 \in (\bar{c}_{s_1}, \bar{c}_{s_1} + \delta)$, $\bar{c}_2 \in (\bar{c}_{s_2} - \delta, \bar{c}_{s_2})$, such that

$$p_1 + p_2 = \text{Prob}(s_1') + \text{Prob}(s_2')$$

and

$$p_1\bar{c}_1 + p_2\bar{c}_2 = \text{Prob}(s_1')\,\bar{c}_{s_1} + \text{Prob}(s_2')\,\bar{c}_{s_2}.$$

Let ϱ_1 and ϱ_2 be

$$\varrho_1 = m^{-1}\left(\frac{\mu + (1-\mu)F_c(\bar{c}_1)}{\mu}I\right)$$

and

$$\varrho_2 = m^{-1}\left(\frac{\mu + (1-\mu)F_c(\bar{c}_2)}{\mu}I\right).$$

Then we can choose δ small enough, such that

$$\varrho_1 \in (\varrho_{s_1} - \epsilon, \varrho_{s_1} + \epsilon)$$

and

$$\varrho_2 \in (\varrho_{s_2} - \epsilon, \varrho_{s_2} + \epsilon).$$

Then following the proof strategy in Lemma B.5we can create another deterministic disclosure policy that generates higher regulator's payoff by creating two signals with equilibrium cutoffs ϱ_1 and ϱ_2.

If the function $xF_c(x)$ is not always convex on $[\bar{c}_{s_1}, \bar{c}_{s_2}]$, based on Assumption (4), we must have $\left.(xF_c(x))''\right|_{x=\bar{c}_{s_2}} = 0$. Let

$$L_1 = \text{supp}\left(\pi\left(\rho|s_2\right)\right) \cap (\varrho^\star, 1],$$

and

$$C_1 = \left[F_c^{-1}\left(\frac{\mu}{1-\mu}\left(\frac{m(\inf L_1)}{I} - 1\right)\right), F_c^{-1}\left(\frac{\mu}{1-\mu}\left(\frac{m(\sup L_1)}{I} - 1\right)\right)\right].$$

if $xF_c(x)$ is linear on C_1, then we can show that there exists a disclosure policy such that the message function is strictly increasing on L_1. If $xF_c(x)$ is not linear on C_1, then there must exist $c_2 \in C_1$, such that $xF_c(x)$ is strictly concave at c_2. The using the proof strategy in Lemma (B.5), we can show this disclosure policy must be suboptimal, a contradiction. $\square$

Besides,the based on the general characterization in Theorem 5.1, there must exists cutoff ϱ_a, ϱ_b and $\varrho^\star$ and a signal space $[\bar{c}_{\min}, \bar{c}_{\max}]$ such that the message function is weakly increasing on $[0, \varrho^\star]$ and $(\varrho^\star, 1]$. Since message function is $\sigma(\rho) = \bar{c}_{\sigma(\rho)}$, we can consider a different signal space $\mathcal{S}' = [\varrho_a, \varrho^\star]$, such that new message function is

$$\sigma'|_{[0,\varrho^\star]} = \begin{cases} \varrho_a & \text{if } \rho \in [0, \varrho_a] \\ \rho & \text{if } \rho \in (\varrho_a, \varrho^\star] \end{cases},$$

and

$$\sigma'|_{(\varrho^\star,1]} = \begin{cases} \varrho_a & \text{if } \rho \in (\varrho^\star, \varrho_b] \\ \gamma(\rho) & \text{if } \rho \in (\varrho_b, 1] \end{cases},$$

where $\gamma(x) = \operatorname{supp}(\pi(\rho|s = \sigma(x))) \cap [0, \varrho^\star]$. Then there must exist a deterministic policy that has the structure characterized in Theorem 5.2.

B.14 Proof of Lemma 5.7

Suppose $(\mathcal{S}, \sigma)$ is the deterministic optimal signal characterized in Theorem 5.2. For any $\rho \in (\varrho_a, \varrho^\star)$, the signal is $s = \rho$, and

$$\operatorname{supp}(\pi(\rho|s)) \cap [0, \varrho^\star] = s.$$

If

$$\varrho_s = \sup\{\operatorname{supp}(\pi(\rho|s)) \cap [0, \varrho^\star]\}$$

doesn't hold for $\rho_0 \in (\varrho_a, \varrho^\star)$, there must exist an interval $B_0 \in (\varrho_a, \varrho^\star)$ and a constant $\epsilon_0 > 0$, such that

$$\varrho_x > \sup\{\operatorname{supp}(\pi(\rho|x)) \cap [0, \varrho^\star]\} + 2\epsilon_0$$

for all $x \in B$. Besides, Lemma B.3 implies that there exists $B \in B_0$ and a constant $\epsilon < \epsilon_0$, such that

$$\varrho_x < \inf\{\operatorname{supp}(\pi(\rho|x)) \cap (\varrho^\star, 1]\} - 2\epsilon.$$

Then for all $x \in B$, we have

$$\varrho_x \in (\sup\{\operatorname{supp}(\pi(\rho|x)) \cap [0, \varrho^\star]\} + 2\epsilon, \inf\{\operatorname{supp}(\pi(\rho|x)) \cap (\varrho^\star, 1]\} - 2\epsilon).$$

Theorem 5.2 implies that there exists s_0 with

$$\operatorname{Prob}(s_0) > 0,$$

and $\varrho_{s_0} < \varrho_x$ for any $x \in B$. Without loss of generality, based on Assumption 4, we can focus on the cases when $xF_c(x)$ is concave on $x \in B$ or it's convex on $x \in B$. This is because there is only one inflection point for function $xF_c(x)$, so if this condition doesn't hold, we can always "truncate" it such that the concavity of function $xF_c(x)$ is unchanged on $x \in B$.

If $xF_c(x)$ is convex on $x \in B$, since $B \cap \text{supp}(\pi(\rho|s_0) \cap [0, \underline{\rho}^\star]) = \emptyset$, we can find $\bar{c}_1 > \bar{c}_{s_0}$, and two functions $f_n(\rho)$ and $\bar{c}_n(\rho)$ on $\rho \in B$, such that

$$\int_{\rho \in B} f_n(\rho)\, d\rho = \int_{\rho \in B} f(\rho)\, d\rho$$

$$\text{Prob}(s_0) \cdot \bar{c}_1 + \int_{x \in B} f_n(x)\, \bar{c}_n(x)\, d\rho = \text{Prob}(s_0) \cdot \bar{c}_{s_0} + \int_{x \in B} f(x)\, \bar{c}_x dx$$

and

$$\begin{aligned} &\text{Prob}(s_0) \cdot \bar{c}_1 F_c(\bar{c}_1) + \int_{x \in B} f_n(x)\, \bar{c}_n(x)\, F_c(\bar{c}_n(x))\, d\rho \\ <&\text{Prob}(s_0) \cdot \bar{c}_{s_0} F_c(\bar{c}_{s_0}) + \int_{x \in B} f(x)\, \bar{c}_x F_c(\bar{c}_x)\, dx, \end{aligned} \tag{41}$$

where the last condition is from the convexity of function $xF_c(x)$. We can always find $(\bar{c}_1, f_n(\rho), \bar{c}_n(\rho))$ such that

$$\bar{c}_1 < \inf_{x \in B} \bar{c}_n(x),$$

and

$$\inf_{x \in B} m^{-1}\left(\left(1 + \frac{1-\mu}{\mu} F_c(\bar{c}_n(x))\right) I\right) > \sup_{x \in B} \{\text{supp}(\pi(\rho|x)) \cap [0, \underline{\rho}^\star]\}.$$

This proof strategy replicates the idea in the proof of Lemma B.6, basically we want to design a new disclosure policy that generates higher regulator's payoff. And the conditions we impose here guarantee that under the new disclosure policy, the regulator's payoff from financing good projects is unchanged, while the cost from financing bad projects decreases because of the condition (41). The complete proof is omitted here because the rest is the same as the proof of Lemma B.6.

If $xF_c(x)$ is concave on $x \in B$, then we can follow the idea in proof of Lemma B.5 and show this is suboptimal. To see this, note we can find two functions $f_m(\rho)$ and $\bar{c}_m(\rho)$ on $\rho \in B$, such that

$$\int_{\rho \in B} f_m(\rho)\, d\rho = \int_{\rho \in B} f(\rho)\, d\rho$$

$$\int_{x \in B} f_m(x)\, \bar{c}_m(x)\, d\rho = \int_{x \in B} f(x)\, \bar{c}_x dx$$

and

$$\int_{x \in B} f_m(x)\, \bar{c}_m(x)\, F_c(\bar{c}_m(x))\, d\rho < \int_{x \in B} f(x)\, \bar{c}_x F_c(\bar{c}_x)\, dx \tag{42}$$

where the last condition is from the concavity of function $xF_c(x)$. We can always find $(f_m(\rho), \bar{c}_m(\rho))$ such that

$$\inf_{x \in B} m^{-1}\left(\left(1 + \frac{1-\mu}{\mu} F_c(\bar{c}_m(x))\right) I\right) > \sup_{x \in B} \{\text{supp}(\pi(\rho|x)) \cap [0, \underline{\rho}^\star]\}.$$

This proof strategy replicates the idea in the proof of Lemma B.5, basically we want to design a new disclosure policy that generates higher regulator's payoff. And the conditions we impose here guarantee that under the new disclosure policy, the regulator's payoff from financing good projects is unchanged, while the cost from financing bad projects decreases because of the condition (42). The complete proof is omitted here because the rest is the same as the proof of Lemma B.5.

B.15 Proof of Theorem 6.1

First, it's obvious that when verification cost t is sufficiently high, the verification technology will never be used. In our analysis, we already show that in any equilibrium s that the verification is used, we must have

$$k_s = k^v = \frac{\mu I^2}{I - t},$$

and the data manipulation level is

$$\bar{c}^v = F_c^{-1}\left(\frac{\mu t}{(1-\mu)(I-t)}\right).$$

Then the lending market equilibrium variables $(k_s, \varrho_s, \bar{c}_s)$ are uniquely determined whenever there is verification used in equilibrium. Suppose the disclosure policy is $(\mathcal{S}, \tilde{\sigma})$, then there is at most one signal s under which verification is used. Suppose under $s_0 \in \mathcal{S}$ there is verification used in equilibrium, and $\text{Prob}\,(s_0) > 0$, then we must have

$$\text{supp}\,(\pi\,(\rho|s_0)) \cap \left(0, m^{-1}\left(\frac{k^v}{\mu}\right)\right) = \emptyset.$$

To see this, suppose for the sake of contradiction that

$$\text{supp}\,(\pi\,(\rho|s_0)) \cap \left(0, m^{-1}\left(\frac{k^v}{\mu}\right)\right) = B,$$

and $\text{Prob}\,(B|s_0) > 0$. It's clear that lenders will never lend to any borrowers if $\rho \in B$ in equilibrium s_0. Then let's consider a new disclosure policy which keeps everything unchanged except disclosing whether the true state $\rho \in B$ or not if the signal realization is s_0 in the previous policy. It's clear that if the true state $\rho \in B$, the regulator's payoff from these states is zero under the old policy, and is non-negative under the new policy, so it weakly improves. The regulator's payoff from other states are unchanged, because lenders are always indifferent between verifying types or not under this equilibrium, and thus the regulator's payoff will be unchanged from these states. Then the regulator's payoff weakly increases under the new policy. Besides, we know that for all $s \in \mathcal{S} \backslash \{s_0\}$, we have $\underline{\rho}_s < m^{-1}\left(\frac{k^v}{\mu}\right)$. Then without loss of generality, we can consider the policy such that the signal s_0 reveals if the true state is above a threshold or not. Formally speaking,

Lemma B.7. *There exists an optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$ and a cutoff ρ^v such that*

$$\operatorname{supp}\left(\pi\left(\rho | s_0\right)\right) = \left(\rho^v, 1\right],$$

and

$$\operatorname{supp}\left(\pi\left(\rho | s\right)\right) \subset \left[0, \rho^v\right]$$

for any $s \in \mathcal{S} \backslash \{s_0\}$, where s_0 is the signal under which verification is used with positive probability.

Then all signals $s \in \mathcal{S} \backslash \{s_0\}$ can only reveal information about states below ρ^v. The following lemma shows that the disclosure policy conditional on $\mathcal{S} \backslash \{s_0\}$ is the optimal disclosure policy when the prior belief is $\rho \sim U\left[0, \rho^v\right]$.

Lemma B.8. *Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal disclosure policy characterized in Lemma B.7, then the disclosure policy $(\mathcal{S}_1, \tilde{\sigma}_1)$ where*

$$\mathcal{S}_1 = \mathcal{S} \backslash \{s_0\}$$

and

$$\tilde{\sigma}_1\left(s | \rho\right) = \tilde{\sigma}\left(s | \rho\right)\big|_{\rho \in [0, \rho^v]}$$

is an optimal disclosure policy when the prior $\rho \sim U\left[0, \rho^v\right]$.

The proof of Lemma B.8 is intuitive. Suppose $(\mathcal{S}_2, \tilde{\sigma}_2)$ is an optimal disclosure policy under prior belief $\rho \sim U\left[0, \rho^v\right]$. If

$$\sup_{s \in \mathcal{S}} \underline{\rho}_s \leq \rho^v,$$

then this optimal disclosure policy is consistent with the constraint of no verification: $\rho \leq \rho^v$, and thus this is optimal. If

$$\sup_{s \in \mathcal{S}} \underline{\rho}_s > \rho^v,$$

then including verification can actually increase the regulator's payoff from states $\rho \in \left[0, \rho^v\right]$, which means that $(\mathcal{S}, \tilde{\sigma})$ is not optimal, a contradiction!

The last part of the proof is to show that for any cost t_x, if when $t = t_x$, verification is used with positive probability under the optimal disclosure policy, then verification will always be used under optimal disclosure policy for any $t < t_x$. This result is straightforward. Suppose W_{NV} is the regulator's payoff when there is no verification technology available, and $W_V(t)$ is regulator's payoff when verification cost is available and the cost parameter is t. It's easy to show that $W_V(t)$ is decreasing in t, so if

$$W_V(t_x) > W_{NV},$$

we must have

$$W_V(t) > W_{NV}$$

for any $t < t_x$. This means that when t is below a threshold (denoted as t^v), verification will always be used under optimal disclosure. The above observation, together with Lemma B.7 and Lemma B.8, complete the proof.