Algorithmic Transparency

Jian Sun*

August 10, 2024

Abstract

We study optimal algorithmic disclosure in a lending market where a lender uses a predictive algorithm to screen a borrower and maximize profit. The algorithm, privately observed by the lender, uses borrower data as input, which can be manipulated by the borrower. Full disclosure is suboptimal due to excessive "gaming the system," while no disclosure is also suboptimal because the lender's ex-post efficient use of borrower data induces excessive ex-ante manipulation. Optimal algorithmic disclosure deters manipulation and improves data quality. Under the optimal policy, borrower data is used less intensively by the lender, reducing manipulation incentives. Despite receiving additional information about the predictive algorithm, the borrower's posterior belief remains significantly uncertain. Algorithmic disclosure can improve the lender's payoff even when she can commit to lending decisions or verify the borrower's type at a cost.

Keywords: FinTech Lending, Bayesian Persuasion, Algorithmic Transparency, Manipulation, Information Design

JEL Codes: D82, G38, G32, D83

*Singapore Management University. I am grateful to Hui Chen, Andrey Malenko and Haoxiang Zhu for their invaluable support and continuous guidance. I would also like to thank Ian Ball, Ben Bernanke, Taha Choukhmane, Marco Di Maggio, Daniel Greenwald, Peter Hansen, Nicolas Inostroza, Olivia Kim, Yaron Leitner, Marcus Opp, Emiliano Pagnotta, Jonathan Parker, Larry Schmidt, Antoinette Schoar, Yupeng Wang, Jiaheng Yu as well as seminar participants at MIT Sloan, Chicago Booth, Toronto Rotman, Tsinghua SEM, Tsinghua PBCSF, The University of Hong Kong, The Chinese University of Hong Kong, National University of Singapore, Singapore Management University, SWUFE and The 2022 Annual Conference in Digital Economics.

1 Introduction

Predictive algorithms have been widely used to mitigate adverse selection in various decision-making processes, including hiring, college admissions, and lending.[1] In these settings, decision makers use predictive models to establish relations between observable personal data and unobserved variables that are relevant to their decision-making problems. For example, employers score resumes to predict job performance, schools use standardized test results to assess academic potential, and FinTech lenders analyze alternative data to evaluate creditworthiness. In these cases, the specific relationship between the input (personal data) and the output (quality) is not transparent to the public, leaving economic agents (such as job candidates, students, and borrowers) with limited information about it. With the development of big data and data processing technology, predictive algorithms have become more complex and nonintuitive, involving variables that have no obvious relationships with each other, and thus have become even more opaque. Although there is a growing call for algorithmic transparency,[2] a key argument justifying the opaque nature of predictive algorithms is the risk of manipulation,[3] or "gaming the system." When economic agents know more about the predictive model, they are more likely to change their behavior strategically, which reduces the informativeness of the input data. Despite the importance of this question, the effects of algorithmic transparency and opacity on market outcomes are still underexplored in academic research. Although recent regulations have started to consider this issue,[4] the motivation typically stems from behavioral concerns, such as privacy or fairness, and largely ignores the effects on market efficiency. Furthermore, the limited understanding of the consequences of algorithmic transparency creates uncertainty about future regulations,[5] potentially

[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] Wang et al. (2020) discuss several examples about manipulation under algorithmic transparency.

[4] For example, the Digital Services Act ("DSA"), which takes effect on February 17, 2024, introduces due diligence and transparency obligations regarding algorithmic decision-making by online platforms. In the UK, the CDDO (Central Digital and Data Office) has launched an algorithmic transparency standard for government departments and the public sector.

[5] 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

adding another layer of inefficiency.

To better understand this question, this paper studies the optimal disclosure of a predictive algorithm that maximizes the algorithm user's objective in a FinTech lending setting. There are two players in this model: a borrower (he) and a lender (she), both of whom are risk-neutral. The borrower has a binary private type and owns a project whose payoff increases with the type. The lender and borrower have misaligned incentives regarding the financing decision. The lender wants to finance the project only if its expected payoff exceeds a constant cost, while the borrower desires financing regardless of the project's payoff. Although the lender cannot directly observe the borrower's type, the borrower generates personal data with binary realizations (e.g., using iOS or Android, or social media presence), which is informative about his type. A better-type borrower generates "better" data, while a worse-type borrower generates "worse" data. This personal data is subject to an unobservable manipulation problem. Specifically, the borrower can privately manipulate his data, changing its realization to the other one by paying a manipulation cost c, which is known only to the borrower. The lender only sees the final realization of the data, which may have been manipulated.

The exact relationship between the borrower's type and the project payoff is unobservable to the borrower but observable to the lender. For simplicity, we assume that the expected project payoff is zero for a worse-type borrower and v for a better-type borrower. The value v is known only to the lender, and intuitively, it influences the worse-type borrower's incentive to manipulate his data. In our model, v also represents the correlation between the project payoff and the borrower type. We refer to this value v as the "variable importance" of the predictive model.[6]

The borrower cares about the probability of the project being financed, and the lender finances the project only if the expected payoff exceeds a constant cost. The lender's evaluation of the expected payoff depends on two factors: the value of the variable importance v and the probability that the borrower is of a better type. The first factor, the variable importance v, is exogenous and is the lender's private in-

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.

[6]In fact, we borrow the term "variable importance" from the machine learning literature. "In machine learning, feature (variable) importance indicates how much each feature contributes to the model prediction. Basically, it determines the degree of usefulness of a specific variable for a current model and prediction." (What Is Feature Importance in Machine Learning?, Baeldung).

formation. The second factor depends on the borrower's data manipulation; less manipulation makes the data more informative about the borrower's type. When the lender discloses some information about the variable importance v to the borrower, it influences the borrower's assessment of the financing probability and his data manipulation incentives, which in turn affects the lender's payoff.

After the borrower receives some information about the variable importance v, a lending market equilibrium emerges accordingly. This equilibrium includes the borrower's data manipulation strategy and the lender's lending decision. A better-type borrower will never manipulate his data, while a worse-type borrower will manipulate if his manipulation cost is below a certain cutoff. The lender will not lend to a borrower with worse data, but will lend to one with better data if the variable importance v exceeds an endogenous cutoff. Therefore, the lending market equilibrium is defined by two cutoffs: a manipulation cost cutoff for the worse-type borrower's decision and an variable importance cutoff for the lender's lending decision. We then characterize the optimal disclosure of the variable importance v that maximizes the lender's payoff.

First, the lender receives zero payoff (the worst possible outcome) if she chooses full disclosure, i.e., disclosing the exact value of variable importance v to the borrower. This happens because, knowing the exact value of v, a worse-type borrower manipulates his data in such a way that the lender always becomes indifferent between lending and not lending to a borrower with better data, resulting in a zero payoff in equilibrium. This implies that full disclosure induces too much "gaming the system", and thus reduces model effectiveness. Second, if the lender chooses no disclosure, i,.e., not disclosing any information about the variable importance v to the borrower, she can achieve a positive payoff, demonstrating that no disclosure is better than full disclosure. This aligns with conventional wisdom that keeping the predictive model opaque can deter data manipulation and improve efficiency. However, we show that no disclosure is still suboptimal. To understand this, consider the case when the variable importance is just slightly above the equilibrium cutoff, and in this case, the lender lends to the borrower if he has better data. Although lending in this case is ex-post efficient , it marginally increases the ex-ante financing probability from the borrower's perspective. This slight increase in financing probability incentivizes worse-type borrower to manipulate his data, reducing its informativeness. It turns out that in this case, the payoff loss from reduced data informativeness outweighs the

marginal gains from lending. This also implies that the borrower data is used too often in the no disclosure equilibrium, because the lending decision depends on the borrower data only when the true variable importance is higher than the equilibrium cutoff.

We characterize the structure of the optimal disclosure policy, which features partial disclosure. The optimal policy divides the support of the borrower's prior belief into disjoint groups and reveals the group to which the true variable importance v belongs. Each revealed group leads to a different equilibrium in the lending market. Regarding the structure of these groups, there exists a group-independent cutoff v^* such that, regardless of which group is revealed, in equilibrium, the lender will lend to the borrower if and only if the borrower has better data and the true variable importance is above v^*. The lender will reject the borrower if he has worse data or if the true variable importance is below v^*. Although different revealed groups lead to different lending market equilibria, the effective lending cutoff remains v^*. This occurs because, for each group (except for at most one), there is a gap in the borrower's posterior belief between the regions above and below the cutoff v^*.

The structure of the optimal policy reveals key intuitions. First, in the optimal disclosure, regardless of which group is revealed, the lender will make the loan if and only if the variable importance v is higher than the group-independent cutoff v^*. Then unconditionally, the loan decision is monotone in variable importance which is efficient. Second, minimizing manipulation by worse-type borrowers is crucial because increased manipulation makes the data noisy, reducing the profitability of lending to a borrower with better data. Under the optimal policy, regardless of which group is revealed, there are always two regions in the borrower's posterior belief about v: one below v^* and one above it, with a positive gap between these regions. This gap maintains significant uncertainty in the borrower's posterior belief even after the lender's disclosure, which helps mitigate manipulation behavior.

Borrower data is used less frequently under the optimal policy compared to both full disclosure and no disclosure scenarios. The intuition is as follows: Under full disclosure, the probability of using borrower data is highest because the borrower's manipulation incentive monotonically increases with variable importance. Although the lender is less willing to use the data when variable importance is low, the borrower manipulates less in this case, leading the lender to use the data even if the variable importance is not very high. In the no disclosure case, this issue is mitigated because

the borrower's manipulation is independent of variable importance, deterring the use of borrower data when variable importance is not high. The optimal policy reduces data use further by varying the borrower's manipulation among different groups. For some groups, low equilibrium manipulation is sufficient to deter use when variable importance is below the cutoff v^*. For other groups, high equilibrium manipulation deters use even when variable importance is relatively high. Unconditionally, the optimal policy deters the use of borrower data more effectively.

We also provide a closed-form characterization of the optimal policy by imposing an assumption on the distribution of the borrower's manipulation cost. The optimal policy includes a discrete part, which induces an equilibrium with the lowest manipulation level among all posterior equilibria, and a continuous part, where manipulation levels vary continuously. This simplifies the optimal disclosure problem to a one-dimensional optimization problem, which can be solved by a differential equation.

We then consider the implementation of the optimal policy in two contexts. The first involves sharing training data to achieve algorithmic transparency. If the true correlation between the borrower's personal data and creditworthiness is below a threshold, the lender shares a representative or full sample with the public (borrowers). If it is above the threshold, the lender adds noise to the sample, making the correlation appear lower, thus keeping the borrower uncertain about the true correlation. The second example involves disclosing variable importance in machine learning models. If the true variable importance is below a threshold, the lender discloses its true value. If it is above the threshold, the lender understates the variable importance in a specific way, maintaining uncertainty in the borrowers' belief about the true variable importance.

Finally, I consider two extensions that explore the interaction between optimal disclosure and other methods of deterring manipulation. The first extension examines the case when the lender can also commit to lending decisions. In this extension, the lender can disclose information to the borrower after committing to a specific lending decision. We show that, under certain conditions, algorithmic disclosure can still improve the lender's payoff even after the lender commits to a lending decision. This result highlights the importance and robustness of information disclosure in deterring manipulation in our setting.

The second extension considers the costly verification of borrower type. In prac-

tice, lenders can verify borrower types through manual reviews, interviews, and fraud detection techniques. In this extension, the lender can verify the borrower's true type at a cost. We explore how costly verification interacts with algorithmic disclosure under the optimal policy. It turns out that in the optimal joint design, these two methods act as substitutes. The lender first reveals whether the true variable importance is above a threshold. If it is, the lender will randomly verify the borrower's type and will not disclose additional information. If it is below the threshold, costly verification is not used, and additional information about the true variable importance is disclosed similarly to the baseline model. This result confirms the importance of algorithmic disclosure in deterring manipulation even when costly verification is available.

The rest of this paper is organized as follows. In this section, we continue to discuss related literature. Section 2 provides a simple model to highlight the intuition, and Section 3 introduces the main model. In Section 4, we discuss the properties of the optimal disclosure policy. Section 5 studies two extensions, and Section 6 concludes.

Related Literature

This paper contributes to three main strands of literature. First, there is an emerging but rapidly expanding body of work on the impact and regulation of algorithmic decision-making, with most existing research focusing on behavior concerns like fairness, bias, and discrimination (e.g. Bartlett et al. (2021), Milone (2019), Gillis and Spiess (2019),Raghavan et al. (2020),Coston et al. (2021)). Our paper extends this literature by exploring algorithmic disclosure from a perspective of market efficiency. And we argue that even without these behavior concerns, algorithmic disclosure is still optimal. A related study by Wang et al. (2020) also examines strategic data manipulation but only compares full transparency and no disclosure policies. They consider both the correlational and causal observables, and focus on the trade off between the investment on these two types of features. In contrast, our research focuses on correlational features as inputs in predictive algorithms, and more importantly, we consider flexible disclosure policies and gain deeper insights into the design of optimal algorithmic disclosure. Additionally, Blattner et al. (2021) address the trade-off between model complexity and transparency and the role of algorithmic audits, which is different from our focus. Björkegren et al. (2020) empirically demonstrate data

manipulation when algorithms are transparent. There is also an expanding literature on algorithmic explainability and explainable AI (e.g., Bhatt et al. (2020), Carvalho et al. (2019), Lundberg and Lee (2017), Murdoch et al. (2019)), which primarily addresses the "black box" nature of machine learning algorithms, while our paper simplifies this nature to address an information design question in economics.

Second, our research contributes to the literature on Bayesian persuasion (Kamenica (2019) and Bergemann and Morris (2019) provide excellent surveys). We model the information structure following Kamenica and Gentzkow (2011) and address a persuasion problem with a continuous state, akin to Dworczak and Martini (2019) and Perez-Richet and Skreta (2022). Generally, Bayesian persuasion problems with continuous states are intractable, except in some special cases (e.g., Gentzkow and Kamenica (2016), Dworczak and Martini (2019), Goldstein and Leitner (2018)). Our theoretical results are derived using a novel "guess and verify" method. Bayesian persuasion has many applications in economics and finance, including shareholder voting (Malenko et al. (2021)), security design (Azarmsa and Cong (2020); Szydlowski (2021); Inostroza and Tsoy (2022)), 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)). Our paper contributes by addressing a new question (algorithmic disclosure) and providing a novel optimal signal structure.

Lastly, this research is related to the literature on strategic data manipulation (Frankel and Kartik (2019a), Frankel and Kartik (2019b); Ball (2019); Perez-Richet and Skreta (2022)) and signaling models. Our approach to modeling private information on the borrower side is similar to Frankel and Kartik (2019b). Ball (2019) examines multi-dimensional features and shows that the optimal scoring rule underweights some features to deter manipulation. These studies primarily focus on improving efficiency through commitment to decision rules. In contrast, we address an information design question, focusing on how commitment to information disclosure can enhance efficiency. Perez-Richet and Skreta (2022) consider how test design changes players' manipulation incentives. The key difference between our papers is that the decision maker has no private information in their model, while the decision maker's private information is the key part of our model.

2 A Simple Model

To fix ideas, let's first consider a simple model. There are two players: a lender (she) and a borrower (he), both are risk neutral. The borrower has zero initial wealth, and is protected by limited liability. He has a borrower-specific project which requires an investment $I = 3$. The project generates a positive cash flow $V = 10$ if it succeeds, and zero if it fails. The probability of success is a random variable. The lender has all bargaining power and collects all payoffs from the project, while the borrower receives a private benefit $b = 1$ if his project is successfully financed, irrespective of its outcome.

The borrower can be good or bad, with probability $\mu = 0.3$ and $1 - \mu = 0.7$, respectively. The borrower type is his private information. A good (bad) borrower has high (low) phone usage naturally, but a bad borrower can privately change the phone usage from low to high by incurring a private cost c, which is uniformly distributed between 0 and 1.[7] A key assumption is that manipulating phone usage does not change a borrower's inherent type. In this market, the only data that the lender can collect and use is the the borrower's phone usage after potential manipulation.

A bad borrower always fails, so his probability of success is always zero. A good borrower's probability of success v is drawn from a uniform distribution $U[0, 1]$, and its realization is only observable to the lender. The lender can commit to disclosing some information about v to the borrower before observing its realization, we examine the equilibria of three types of disclosure policies in this section.

No Disclosure

Suppose the lender does not disclose any information about v, it can be shown that there is a unique equilibrium which is characterized by two cutoffs c_N and v_N. A bad borrower with manipulation cost lower than c_N chooses to manipulate his phone usage from low to high. The lender always rejects a borrower with low phone usage and lends to a borrower with high phone usage if and only if $v > v_N$.

The equilibrium condition for a bad borrower with manipulation cost c_N is

$$\text{Prob}\,(v > v_N) \cdot b = c_N,$$

[7] Here we only allow a bad borrower to manipulate his phone usage for simplicity of exposition. But the result does not change if we allow a good borrower to manipulate his phone usage from high to low, because a good borrower will never manipulate this in equilibrium.

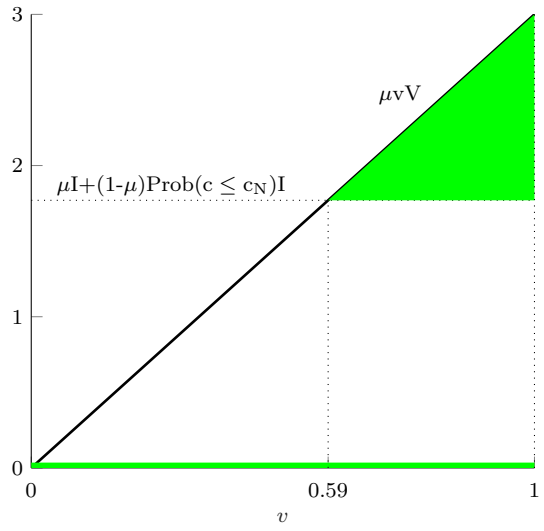

Figure 1: Lender's Payoff: No Disclosure

where $\text{Prob}(v > v_N) = 1 - v_N$ is the probability of a borrower with high phone usage receiving a loan. The lender's equilibrium condition is

$$\mu v_N V = (\mu + (1-\mu)\,\text{Prob}(c \leq c_N))\, I.$$

With our assumptions, the unique equilibrium values are $v_N = \frac{\mu I + (1-\mu) Ib}{\mu V + (1-\mu) Ib}$ and $c_N = b \cdot \frac{\mu(V-I)}{\mu V + (1-\mu) Ib}$. Then the lender's profit is

$$W_N = \mu V \int_{v_N}^{1} (v - v_N)\, dv.$$

We can show that in this equilibrium, $v_N = 0.59$, $c_N = 0.41$ and $W_N = 0.25$.

Figure 1 characterizes the above equilibrium. The green triangle in Figure 1 represents the profit W_N. In equilibrium, the loan is approved for a borrower with high phone usage when $v > 0.59$. The green line on the horizontal axis represents the support of posterior belief of v. In this no disclosure equilibrium, the posterior belief is the same as the prior belief.

Full Transparency

If the lender perfectly reveals the realization of v to the borrower, we can show that her profit equals zero, leading to the worst outcome. To see this, when $vV < I$, lending is always inefficient, resulting in no financing and hence zero profit. For any

$vV \geq I$, in equilibrium, the lender must be indifferent between lending to a borrower with high phone usage and not lending at all. Thus, the profit must also be zero for any $vV \geq I$.

A Binary Signal

Our question is, can the lender achieve a strictly higher outcome by disclosing some information about v to the borrower? The answer is yes. The definition of disclosure policy is formally introduced in Section 3.1, here we consider a simple case. Let's consider two sets $A = [0, 0.54) \cup (0.64, 0.91)$, and $A^{\complement} = [0.54, 0.64] \cup [0.91, 1]$.

Consider the following disclosure policy: *the lender reveals whether v is in the region A or $A^{\complement}$*. There are two possible equilibria depending on which region the realization v belongs to. If $v \in A$, the posterior belief about v is a uniform distribution on two disjoint intervals $[0, 0.54) \cup (0.64, 0.91)$, and we can show that the equilibrium outcomes are $v_1 = 0.54$, $c_1 = 0.34$, and $W_1 = 0.19$. Similarly, if $v \in A^{\complement}$, the posterior belief about v is a uniform distribution on $[0.54, 0.64] \cup [0.91, 1]$, and the equilibrium outcomes are $v_2 = 0.64$, $c_2 = 0.48$ and $W_2 = 0.09$.

Figure 2 characterizes these two equilibria. The red trapezoid in the left panel represents the profit in the equilibrium when observing $v \in A$, and the two red intervals on the horizontal line represent the support of the posterior belief in this equilibrium. In this case, the lender lends to the borrower with high phone usage only when $v \in (0.64, 0.91)$. Similarly, the right panel shows the profit on observing $v \in A^{\complement}$. The total profit with this disclosure policy is

$$W_s = W_1 + W_2 = 0.19 + 0.09 = 0.28 > 0.25 = W_N.$$

So the expected profit indeed improves.

Our analysis shows that the binary signal dominates both the no disclosure policy and full transparency policy. The result that full transparency policy is dominated is clear: when the exact information about the probability of success v is disclosed to the borrower, a bad type borrower manipulates his phone usage such that the profit from using the borrower data in lending decisions is always zero, and the lender is indifferent between using and not using the borrower data. The inefficiency in the no disclosure equilibrium arises due to the lender's lack of commitment problem, that is, the lender always makes the most efficient use of borrower data ex post in the lending

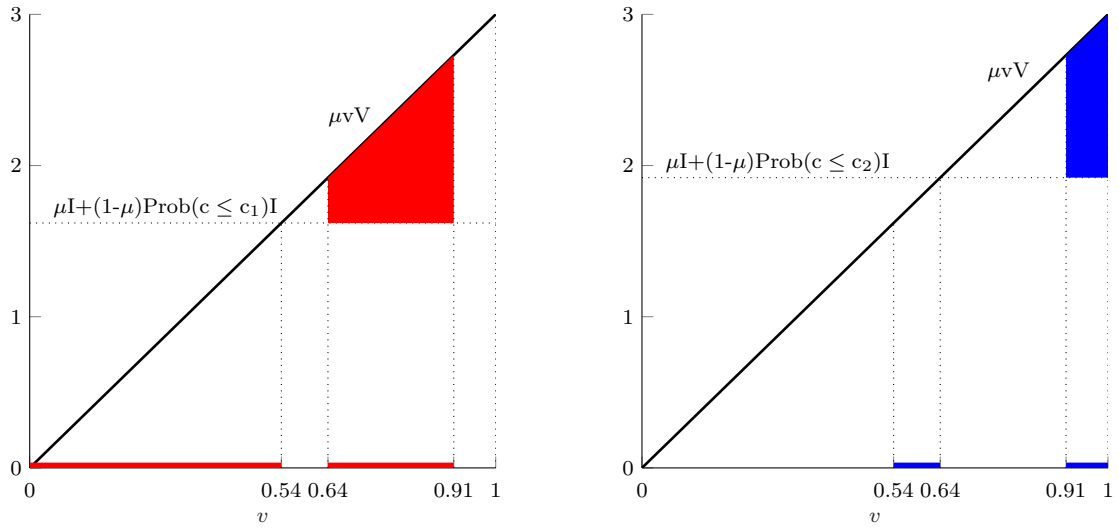

Figure 2: Lender's Payoff: binary signal

decision. To see this, suppose that the lender commits to lending to the borrower with high phone usage if $v > x$. We can calculate the equilibrium profit as a function of x, specifically, the lender's profit is[8]

$$W(x) = \int_x^1 \left[\mu v V - (\mu + (1-\mu)(1-x))\, I\right] dv.$$

Since v_N satisfies $\mu v_N V - (\mu + (1-\mu)(1-v_N))\, I = 0$, we can show that

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

The above result shows that the equilibrium cutoff v_N is inefficiently low from an ex ante perspective. Note that when $v < v_N$, the borrower data is effectively not used in the lender's lending decision, because the borrower will be rejected no matter what his data is. The borrower data will be used only when $v \geq v_N$, at which point a borrower with high and low phone usage will receive different loan approval decisions. So the probability that the borrower data is used in lending decisions, $(1 - v_N)$, is inefficiently high in this no disclosure equilibrium. When the lender uses the borrower data more often ex post in some states, a bad borrower is more likely to manipulate his data ex ante, and the total financing cost will increase for all other states from

[8]The bad type borrower chooses to manipulate when $c \leq c_x = b(1-x)$.

the ex ante perspective. This cross-state externality makes no disclosure equilibrium inefficient.

To mitigate the excess manipulation, the binary signal defers the lender's use of borrower data by differentiating the lending market equilibria. *Unconditionally*, with the binary signal, borrower's phone usage data is used when $v > 0.64$, while it is $v > 0.59$ in the no disclosure equilibrium. So the data is used less frequently under the binary signal. Intuitively, by differentiating the two equilibria, the "worse" equilibrium ($v \in A^{\complement}$) effectively guarantees that the borrower data will not be used in cases when it was indeed used in no disclosure equilibrium (when $v \in (0.59, 0.64)$), and the "better" equilibrium ($v \in A$) has lower level of data manipulation and generates more efficient outcome.

In the main model, I'll consider a general space of disclosure policies. But this binary signal has several notable properties that are still robust in the optimal disclosure policy in the general model. First, there exists a threshold ($v^* = 0.64$), such that *unconditionally*, the borrower data is used in lending decisions if and only if the true state is above the threshold. This cutoff property always holds for any posterior equilibria under any disclosure policy, and here we show it also holds unconditionally in the optimal policy. Second, for any signal realization of information disclosure ($v \in A$ or $v \in A^{\complement}$), the support of posterior belief can always be separated by this ex ante cutoff v^*. Third, the unconditional probability of using the borrower data in lending decisions is lower than that in the no disclosure equilibrium, implying that the borrower data is used less intensively under optimal disclosure. Lastly, the binary signal induces two posterior equilibria, one with higher ($v \in A^{\complement}$) and the other with lower ($v \in A$) data manipulation levels than the no disclosure equilibrium. All of these properties still hold in the optimal disclosure policy in the general model.

3 The Main Model

3.1 Model Setup

Now let's introduce the formal model. There are two players in this model: a deep pocket lender (she) and a representative borrower (he), both are risk neutral. The borrower has zero initial wealth but has access to an investment opportunity (a project) that costs $I > 0$. The project payoff V is a random variable. We assume the lender

has all the bargaining power and the borrower is protected by limited liability. If the lender finances the borrower, she obtains all the profit, $V - I$, from the investment; the borrower, on the other hand, gets private benefit $b > 0$ from undertaking the project. If the lender does not lend to the borrower, then both players obtain their outside options which are normalized to zero.

The borrower type $X \in \{0, 1\}$ is private information. The probability that the borrower type is $X = 1$ is denoted by μ, which is also the lender's prior belief. The project payoff for a borrower with type X is

$$V = X \cdot v + \epsilon, \tag{1}$$

where ϵ is pure noise, independent of X, and satisfies $\mathbb{E}(\epsilon) = 0$ and $\mathrm{Var}(\epsilon) < \infty$. The distributional information about ϵ is public information. Then it's clear that $\mathbb{E}(V|X) = X \cdot v$. The key feature of our model is that the value of v is drawn from a distribution, and its realization is privately observed by the lender. For simplicity, we assume the value v is drawn from a continuous distribution on $[0, \bar{v}]$ with a probability (cumulative) distribution function $g(\cdot)$ ($G(\cdot)$). Assume function g is continuous and strictly positive on $[0, \bar{v}]$. To focus on the interesting case, we assume that $\bar{v} > I$, otherwise, the lender will never lend to the borrower. The borrower only knows the distributional information g (and G) but not the realization. The equation (1) represents a predictive model, and the interpretation of the model parameter v may vary depending on the context. For example, v can be the variable importance or feature importance in machine learning models, or it can simply represent the correlation in a statistical model.[9] In practice, a FinTech lender might use long-term, historical, and unmanipulated data, along with their proprietary machine learning model, to extract the correlation between borrowers' choices of phone operating system and the default probability within the population. However, borrowers cannot observe this correlation information. For the rest of this paper, we refer to v as the *variable importance* to align with its interpretation in a machine learning setting.

Although the lender can't observe the borrower type X directly, she has access to some borrower data x, which is informative about the borrower type X. We assume that the borrower data x also takes values in $\{0, 1\}$. For instance, the data x could be

[9]Note that in our setting, the correlation between the outcome variable V and the type X is $\rho_{X,V} = \frac{1}{\sqrt{1+\frac{\mathrm{Var}(\epsilon)}{v^2\mu(1-\mu)}}}$, which is an increasing function of v.

the operating system used by the borrower, or more generally, his digital footprint, that is informative about the borrower's type (credit quality).[10] A borrower with type $X = 1$ ($X = 0$) naturally generates data $x = 1$ ($x = 0$), however, the data x is subject to the borrower's manipulation. Specifically, a borrower can change x from 1 to 0, or vice versa, by incurring a private cost c. This manipulation does not alter the borrower's true type X. Therefore, if the data is not manipulated, $x = X$, the data perfectly reveals the borrower's type. If the data is manipulated, then $x = 1 - X$. The cost c is independent of all other random variables in the model, and is drawn from a continuous distribution on $(0, \bar{c}]$ with a cumulative distribution function $K(\cdot)$. Only the borrower observes its realization. For simplicity, assume $K'(c)$ exists everywhere and is continuous and strictly positive at any $c \in [0, \bar{c}]$.[11]

Assumption 1. $\mu\bar{v} < [\mu + (1 - \mu) K(b)] I$.

Assumption 1 implies that in any equilibrium, the lender will never lend to the borrower with data $x = 1$ with 100% probability. If the lender indeed lends to the borrower with data $x = 1$ for sure, the probability that the borrower has data $x = 1$ becomes $[\mu + (1 - \mu) K(b)]$, and the expected cost of financing, $[\mu + (1 - \mu) K(b)] I$, is higher than the maximum payoff from the project, $\mu\bar{v}$, and thus lending is unprofitable for the lender. So this can not be an equilibrium with Assumption 1.

Disclosure About the Predictive Model

To maximize her expected profit, the lender can reveal some information about the predictive model (1) to the borrower, which is represented by the parameter v in our setting. Specifically, the lender commits to a policy that reveals information about v before observing its true value, and the borrower makes data manipulation decision

[10] Berg et al. (2020) mention that "...simple, easily accessible variables from the digital footprint proxy for income, character, and reputation and are highly valuable for default prediction. For example, the difference in default rates between customers using iOS (Apple) and Android (e.g., Samsung) is equivalent to the difference in default rates between a median credit score and the 80th percentile of the credit bureau score. " A related example is that in 2019, Goldman Sachs issued a new credit card, the "Apple Card," which can only be applied for on an iOS device. In this case, they effectively used the phone operating system as a screening variable in their credit card application.

[11] For example, in the case of the Apple Card, which can only be applied for on an iOS device, applicants without such devices might strategically alter their behavior by purchasing a new iOS device or borrowing one, based on their belief in the variable importance of operating system in predicting their true credit quality. Such behavioral changes, while costly, make the operating system data noisier and less effective in the predictive model.

after observing the revealed information. To make the disclosure policy as flexible as possible, we adopt the standard definition used in the information design literature.

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

$$\tilde{\sigma} : [0, \bar{v}] \to \Delta\mathcal{S}.$$

Let $s \in \mathcal{S}$, then (v, s) forms a joint distribution on $[0, \bar{v}] \times \mathcal{S}$. Let the marginal cumulative (probability) distribution function of s be F (f), and the conditional cumulative (probability) distribution function of $v|s$ be Π_s (π_s), then we must have

$$\int_{s \in \mathcal{S}} \pi_s dF = g.$$

For simplicity, we use $\{F, \pi_s\}$ to represents the distribution of posteriors induced by the signal, which is an element in $\Delta(\Delta\mathcal{S})$.

The lender discloses a signal s following the disclosure policy, and then the borrower updates his belief and chooses his manipulation strategy accordingly. As discussed in the literature, 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 to the borrower, and thus our model also sheds light on the boundary of the pure information channel mitigating strategic manipulation.

Later we will show that optimal disclosure policies can always be implemented by a specific class of policies known as deterministic policies. In these policies, the signal realization conditional on any state v is deterministic, and thus the disclosure policy can be represented by a message function, see below for the formal definition of a deterministic policy. For simplicity, let δ_x be the Dirac measure, which puts probability 1 on the state x, and zero otherwise.

Definition 3.2. A disclosure policy $(\mathcal{S}, \tilde{\sigma})$ is deterministic if for any $v \in [0, \bar{v}]$, the signal realization is deterministic, i.e., there exists a measurable function $\sigma : [0, \bar{v}] \to \mathcal{S}$, such that

$$\tilde{\sigma} = \delta_{\sigma(v)}.$$

We also call $\sigma(v)$ the message function.

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 a message function σ.[12]

Timeline

All events occur in the following order:

1. The lender chooses a signal structure $(\mathcal{S}, \tilde{\sigma})$; then v is randomly drawn from the distribution $g(v)$ and observed by the lender.
2. A signal realization s is generated based on $(\mathcal{S}, \tilde{\sigma})$, and disclosed to both players.
3. The borrower decides whether to manipulate his data x.
4. The lender decides whether to lend to the borrower or not based on the observed data x and the true value of v.
5. All random variables are realized, and both players receive their payoffs.

3.2 The Market Equilibrium

We first investigate the market equilibrium after a signal s is revealed under a disclosure policy $(S, \tilde{\sigma})$. Let's call this the subgame s. For a borrower with type X and manipulation cost c, his decision, $\gamma_s(X, c) \in [0, 1]$, is the probability of manipulating his data x upon observing signal s. For the lender, her decision, $\alpha_s(x, v) \in [0, 1]$, is the probability of lending to the borrower based on the borrower data x (after potential manipulation), the lender's privately observed variable importance v, and the public signal s.

The lender's expected profit of lending to a borrower with data $x = 1$ is

$$U_1 = \mu v \int_0^{\bar{c}} [1 - \gamma_s(1, c)] \, dK(c) - \left[\mu \int_0^{\bar{c}} [1 - \gamma_s(1, c)] \, dK(c) + (1 - \mu) \int_0^{\bar{c}} \gamma_s(0, c) \, dK(c) \right] I, \tag{2}$$

[12] For example, the full transparency policy can be represented by a signal space $\mathcal{S} = [0, \bar{v}]$ with a message function $\sigma(v) = v$; and the no disclosure policy can be represented by a signal space $\mathcal{S} = \{0\}$ with a message function $\sigma(v) \equiv 0$.

and zero otherwise. Similarly, her expected profit of lending to a borrower with data $x = 0$ is

$$U_0 = \mu v \int_0^{\bar{c}} \gamma_s(1,c)\, dK(c) - \left[\mu \int_0^{\bar{c}} \gamma_s(1,c)\, dK(c) + (1-\mu) \int_0^{\bar{c}} [1-\gamma_s(0,c)]\, dK(c)\right] I, \tag{3}$$

and zero otherwise. So the lender's choice $\alpha_s(x,v)$ solves the following problem:

$$\max_{\alpha\in[0,1]} \alpha U_x \tag{4}$$

for $x = 0, 1$. Then in the subgame s, the probability that the project owned by a borrower with data x can be financed is

$$d_s(x) = \int_0^{\bar{v}} \alpha_s \pi_s dv. \tag{5}$$

For a borrower with type X, when he manipulates the data, his data x becomes $1 - X$, then his profit is

$$d_s(1-X)\, b - c.$$

If he does not manipulate the data, his data x remains to be X, and his profit is

$$d_s(X)\, b.$$

Then his decision $\gamma_s(X,c)$ solves

$$\max_{\gamma\in[0,1]} (1-\gamma)\cdot d_s(X)\, b + \gamma\cdot(d_s(1-X)\, b - c). \tag{6}$$

We introduce a formal definition of the lending market equilibrium of subgame s below.

Definition 3.3. An equilibrium of subgame s consists of the borrower's manipulation strategy $\gamma_s(X,c)$ and the lender's lending strategy $\alpha_s(x,v)$, such that under the posterior belief π_s, the following two conditions are satisfied.

1. The lender's strategy $\alpha_s(x,v)$ solves (4).
2. The borrower's strategy $\gamma_s(X,c)$ solves (6), where $d_s(x)$ is defined in (5).

It turns out that the lending market equilibrium has a simple structure. First, in the appendix, we show that $\gamma_s(1,c) = 0$, this is because a borrower with data $x = 1$

is always perceived as being of better quality by the lender, so he has no incentive to manipulate his data. This implies that only a borrower with type $X = 0$ will possibly choose to manipulate his data, and thus the lender always rejects a borrower with $x = 0$, leading to $d_s(0) \equiv 0$. Therefore, we just need to focus on the borrower's strategy when his type is $X = 0$ and the lender's strategy when she faces to a borrower with data $x = 1$. In equilibrium, a borrower with type $X = 0$ decides to manipulate his data only if his manipulation cost is low enough, i.e.,

$$c \leq b \cdot d_s(1) = c_s.$$

The lender lends to a borrower with data $x = 1$ only if the privately observed v is high enough, i.e.,

$$\mu v - [\mu + (1-\mu) K(c_s)] \cdot I \geq 0 \Longleftrightarrow v \geq v_s = \frac{[\mu + (1-\mu) K(c_s)] \cdot I}{\mu}. \tag{7}$$

When $v = v_s$, the lender is indifferent to lending or not. The equilibrium can be summarized by these two cutoffs c_s and v_s. The following lemma characterizes the lending market equilibrium for the subgame s.

Lemma 3.1. *For any subgame s, if $\Pi_s(I) = 1$, there is no financing in any equilibrium; if $\Pi_s(I) < 1$, there exists a unique equilibrium with two cutoffs c_s and v_s, such that*

1. $\gamma_s(1, c) \equiv 0$, *and* $\gamma_s(0, c) = \begin{cases} 0 & \text{if } c > c_s \\ \in [0,1] & \text{if } c = c_s \\ 1 & \text{if } c < c_s \end{cases}$.

2. $\alpha_s(0, v) \equiv 0$, *and* $\alpha_s(1, v) = \begin{cases} 1 & \text{if } v > v_s \\ \in [0,1] & \text{if } v = v_s \\ 0 & \text{if } v < v_s \end{cases}$.

3. *c_s and v_s are solved by the following equilibrium conditions*

$$c_s = b \cdot d_s(1), \tag{8}$$

$$Prob(v > v_s | s) \leq d_s(1) \leq Prob(v \geq v_s | s), \tag{9}$$

and

$$v_s = \frac{[\mu + (1-\mu) K(c_s)] \cdot I}{\mu}, \tag{10}$$

where $Prob(\cdot|s)$ *is the probability function of the posterior belief about* v *under the subgame* s*.*

The lender's expected profit in the subgame s *is*

$$W_s = \mu \mathbb{E}_s \left[(v - v_s)^+\right] = \mu \int_{v_s}^{\bar{v}} (v - v_s) \pi_s(v) \, dv.$$

For the rest of the paper, we use (v_s, c_s) to represent the two equilibrium cutoffs mentioned in the above lemma under a signal realization s. Since there is no financing when $\Pi_s(I) = 1$, without loss of generality, we choose $c_s = 0$ and $v_s = 0$ for this case. By choosing a disclosure policy $(\mathcal{S}, \tilde{\sigma})$, the distribution of the signal s (represented by $F(s)$ or $f(s)$) is uniquely pinned down. Under each realization s, the belief about v is updated to π_s, and the lender obtains the expected profit W_s characterized in Lemma 3.1. Then the lender's expected profit is $\int_s W_s dF(s)$, and thus her optimization problem is

$$\max_{(\mathcal{S},\tilde{\sigma})} \int_s W_s dF(s).$$

For simplicity, it is standard in the literature to work with the distribution of posteriors, $\{F(s), \pi_s\}$,[13] rather than directly with disclosure policies. Thus, we can reformulate the lender's problem in the following proposition.

Proposition 3.1. *The lender solves the following equivalent problem:*

$$\max_{\mathcal{S},\{F,\pi_s\}} W = \mu \int_{s\in\mathcal{S}} \mathbb{E}\left[(v - v_s)^+ | s\right] dF(s) \tag{11}$$

$$s.t \int_{s\in\mathcal{S}} \pi_s dF(s) = g(v), \tag{12}$$

$$Prob(v > v_s | s) \leq \frac{K^{-1}\left(\frac{\mu v_s}{I(1-\mu)} - \frac{\mu}{1-\mu}\right)}{b} \leq Prob(v \geq v_s | s) \quad \forall s \in \mathcal{S}. \tag{13}$$

Here $Prob(v|s)$ *is the probability function associated with the posterior belief* π_s*.*

The last condition (13) solves the equilibrium cutoff v_s under the posterior belief π_s. This problem is known as a Bayesian persuasion problem with continuous states,

[13] See the Online Appendix of Kamenica and Gentzkow (2011) for the discussion with continuous states.

which is in general not tractable except in some special cases (Gentzkow and Kamenica (2016), Dworczak and Martini (2019)). Our model does not fit into any existing tractable framework and is solved by a "guess and verify" approach. Before analyzing the structures of the lender's solution, we have the following observation that helps simplify the problem.

Lemma 3.2. *For any disclosure policy, if there exist two distinct signal realizations s_1 and s_2 with equilibrium cutoffs $v_{s_1} = v_{s_2}$, then combing these two signals together will not change the market equilibrium.*

Lemma 3.2 implies that we can (without loss of generality) focus on policies $(\mathcal{S}, \tilde{\sigma})$, such that for any two distinct signal realizations $s_1, s_2 \in \mathcal{S}$, $v_{s_1} \neq v_{s_2}$. Then we can choose a signal space such that $s \equiv v_s$ for all $s \in \mathcal{S}$. We will use this simplification throughout the rest of this paper.

3.3 Suboptimality of No Disclosure

The key friction in our model is the adverse selection due to endogenous manipulation. A borrower with type $X = 0$ chooses his manipulation strategy based on the updated public belief on the distribution of v. For the optimal policy, a natural guess would be that the lender should not disclose any information, and thus make the model as opaque as possible. Denote the lending market equilibrium in this case as (v_N, c_N), where N represents "no disclosure." The lender's payoff is

$$W_N = \mu \int_{v_N}^{\bar{v}} (v - v_N)\, dG(v).$$

We can show that in this case, the use of borrower data x is too frequent in lending decisions, leading to excessive manipulation. This result arises from the lender's lack of commitment: she always makes the most efficient use of borrower data ex post, which is ex ante inefficient. To see this, note that given the borrower's manipulation, the lender always lends to a borrower with data $x = 1$ if the lending is ex post positive NPV, i.e., when $v > v_N$, where v_N satisfies

$$\mu v_N - [\mu + (1 - \mu) K(c_N)] \cdot I = 0. \tag{14}$$

Now, suppose the lender can commit to a slightly higher lending cutoff $v_N + \delta$ ($\delta \ll 1$), i.e., the lender lends to a borrower with data $x = 1$ only when the variable importance

$v \geq v_\delta = v_N + \delta$. Let her payoff in this case be W_δ. This leads us to the following result.

Lemma 3.3. $\frac{dW_\delta}{d\delta}\Big|_{\delta=0} > 0$.

Lemma 3.3 implies that, in the no disclosure equilibrium, a commitment to reduced lending improves the lender's payoff. Note that this commitment to reduced lending corresponds to a less frequent use of borrower data in lending decisions.[14] Similar results also show up in other economic settings where the information receivers commit to underweight some variables in decision rules to deter manipulation and improve efficiency (Ball (2019)). To see the intuition, rewriting $\frac{dW_\delta}{d\delta}\Big|_{\delta=0}$,

$$\frac{dW_\delta}{d\delta}\Big|_{\delta=0} = -\left[\mu v_N - \left(\mu + (1-\mu) K\left(c_N\right)\right) I\right] g\left(v\right) + \int_{v_\delta}^{\bar{v}} \left(-(1-\mu) I g\left(v\right) \frac{dK\left(c_\delta\right)}{d\delta}\Big|_{\delta=0} \right) dv. \tag{15}$$

The first term in (15) equals zero, reflecting the lender's optimality condition as specified in (14). However, the second term is positive because an increase in δ decreases the equilibrium cutoff of manipulation, denoted as c_δ. This highlights the suboptimality of the no disclosure policy. Although the lending decision is always positive NPV ex post, approving a borrower with data $x = 1$ incentivizes the borrower to engage in ex ante manipulation. This manipulation exacerbates adverse selection, thereby increasing the lending cost across all states, regardless of the realized variable importance v.

In our paper, committing to lending decisions is not allowed,[15] and only disclosure about the variable important v is considered. We can show that the lender can mitigate the manipulation and improve payoff by disclosing information about the variable importance v. The following proposition confirms that no disclosure is indeed suboptimal.

Proposition 3.2. *There exists a disclosure policy* $(\mathcal{S}, \tilde{\sigma})$ *with lender's payoff* W_1, *such that* $W_1 > W_N$.

[14] In our model, the lender rejects the borrower no matter what his data is if the variable importance v is below a certain cutoff, meaning the borrower data is essentially not used when rejecting. However, approval depends on the borrower data x; if v exceeds the cutoff, the lender lends only if the borrower has $x = 1$. Therefore, the lending decision depends on borrower data only when v is above the cutoff. As a result, a higher cutoff means a lower probability that the lending decision depends on borrower data, which effectively reduces the lender's usage of borrower data.

[15] We consider an extension about this assumption in Section 5.1.

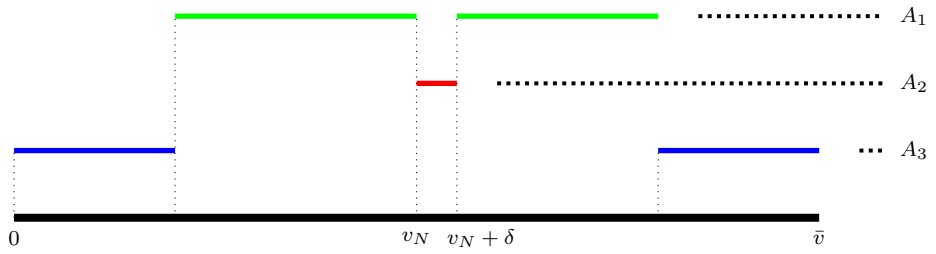

Figure 3: Suboptimality of No Disclosure Policy

Proposition 3.2 challenges the conventional wisdom that making predictive models more transparent will always hurt efficiency because of the "gaming the system" concern. The key to Proposition 3.2 is to find a disclosure policy under which the lender will use borrower data x less frequently in equilibrium, which will deter manipulation.

To gain intuitions on how it works, suppose the lender 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

$$\sigma(v) = s_1 \mathbb{1}_{A_1}(v) + s_2 \mathbb{1}_{A_2}(v) + s_3 \mathbb{1}_{A_3}(v),$$

where A_1, A_2, and A_3 are (unions of) intervals shown on Figure 3. The above disclosure policy effectively discloses which set of A_1, A_2, and A_3 that the true state belongs to. The boundaries of the intervals are chosen such that the following three conditions are satisfied. First, the equilibrium of subgame s_1 is the same as the no disclosure equilibrium, i.e., $(v_1, c_1) = (v_N, c_N)$; second, $A_2 = [v_N, v_N + \delta]$ is a small set, where $\delta \ll 1$; finally, the remaining regions is A_3, i.e., $A_3 = [0, \bar{v}] - A_1 \cup A_2$.

The equilibrium of subgame s_1 is the same as the no disclosure equilibrium. The signal s_2 reveals that the true state is in the interval $[v_N, v_N + \delta]$. In this subgame, the lender's payoff is less that $\mu\delta$ for any value of v, which is close to zero for any $v \in A_2$ because δ is very small. Then the change in lender's payoff is negligible in this region. In the equilibrium of subgame s_3, the probability of financing the borrower with data $x = 1$ is lower than that in the no disclosure equilibrium (note that the borrower with data $x = 1$ will be financed only if the true state v is in the right interval of A_3), which reduces the manipulation incentives and improves the outcome. Then the net effect on lender's payoff is positive. The example in Figure 3 shows that the lender can indeed improve the profit by disclosing some information about the variable importance v.

In the next section, we characterize general properties of optimal policies.

Another natural guess for the optimal disclosure policy is full transparency, i.e., disclosing the exact value of v to the borrower. We can show that full transparency leads to the worst outcome, and thus it must be suboptimal.

Lemma 3.4. *The lender's payoff when making the variable importance v fully transparent, W_F, is zero.*

Since the lender's payoff must be nonnegative, Lemma 3.4 implies that disclosing all the information about the true state of v leads to the worst payoff. The intuition behind the result is straightforward: when the borrower knows the exact value of the true state, in equilibrium, the data manipulation level is high enough such that there is zero surplus from financing a borrower with data $x = 1$, and the lender gets zero payoff when using borrower data x in lending. This result is consistent with the common view that disclosing too much information about the predictive model will hurt efficiency due to "gaming the system."

Remark 3.1. Proposition 3.2 and Lemma 3.4 jointly imply that the optimal disclosure policy must feature partial disclosure.

4 General Properties of Optimal Policies

Before analyzing the general properties of optimal policies, let's first consider the properties of the equilibrium in any subgame. First, Lemma 3.1 implies that for any two signal realizations s_1 and s_2, we must have

$$c_{s_1} \lesseqgtr c_{s_2} \iff v_{s_1} \lesseqgtr v_{s_2}.$$

When the manipulation cutoff c_s is lower, the borrower is more likely to manipulate his data, and the adverse selection is more severe when facing to a borrower with data $x = 1$. In this case, the variable importance has to be higher for loan approval. Consider an arbitrary signal s with posterior belief π_s. Let

$$A_s = \{v \in \text{supp}\,(\pi_s) \,|\alpha_s\,(1, v) > 0\}\,, \tag{16}$$

and

$$R_s = \{v \in \text{supp}\,(\pi_s) \,|\alpha_s\,(1, v) = 0\}\,. \tag{17}$$

A_s is the subset of the support of $v|s$, where the lender lends to a borrower with data $x = 1$ with positive probability, while R_s is the subset where the lender lends with zero probability. Let's call A_s and R_s the *acceptance region* and *rejection region,* respectively. Clearly, A_s and R_s are disjoint, and $A_s \cup R_s = \text{supp}(\pi_s)$. The following lemma shows that in any optimal policy, both A_s and R_s are non-empty sets.

Lemma 4.1. *Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, then for almost all $s \in \mathcal{S}$, both A_s and R_s are non-empty sets, and $A_s = \{v \in \text{supp}(\pi_s) \,|\alpha_s(1, v) = 1\}$.*

Since the lender's action $\alpha_s(x, v)$ is always increasing in v, for any subgame s, we must have

$$\sup R_s \leq v_s \leq \inf A_s,$$

showing that the acceptance region A_s and rejection region R_s are separated by a signal-contingent cutoff v_s for any signal s. This simple result stems from the lender's ex post optimality condition: it is always optimal for the lender to lend to a borrower with data $x = 1$ if the variable importance is higher. The following theorem demonstrates that this intuition also holds from an ex ante perspective under any optimal policy.

Theorem 4.1. *Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, then there must exist a **signal-independent** cutoff v^*, such that for almost all $s \in \mathcal{S}$, we have*

$$\sup R_s \leq v^* \leq \inf A_s,$$

where A_s and R_s are acceptance region and rejection region defined above.

Theorem 4.1 provides a necessary condition that all optimal policies must satisfy. From an ex ante perspective, the lender lends to a borrower with data $x = 1$ if and only if the variable importance v is sufficiently high. While this condition has been shown to hold under any subgame in previous analyses, Theorem 4.1 demonstrates that it must also hold ex ante under any optimal disclosure policy. It also implies that, in any optimal policy, for all signal realizations s, the following condition must hold:

$$\text{supp}(\pi_s) \cap (\min\{v_s, v^*\}, \max\{v_s, v^*\}) = \emptyset.$$

This implies that, the posterior distribution of all signal realizations must contain a "hole," which separates the rejection region and the acceptance region.

This special structure offers some insights into optimal policies. Optimal policies reveal information about the true state, but in the same time maintain significant uncertainty, regardless of signal realizations. Intuitively, creating a gap between the rejection and acceptance regions in posterior beliefs preserves uncertainty about the variable importance in the borrower's posterior belief, mitigating concerns about "gaming the system." The following proposition shows that the cutoff v^* is actually the maximum of (almost) all posterior lending cutoffs.

Proposition 4.1. *Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, then*

$$v^* \geq v_s$$

for almost all $s \in \mathcal{S}$.

In Section 3.3, we discussed how the no disclosure equilibrium leads to excessive lending and incentivizes borrower manipulation. Now we can see why this outcome can be improved from Proposition 4.1. To see the intuition, suppose the policy induces two signals, s_1 and s_2, with $v_{s_1} < v_{s_2} \leq v^*$. The higher approval probability in s_2 leads to more manipulation ($c_{s_2} > c_{s_1}$). Consequently, the lender requires a higher variable importance v in subgame s_2 to approve loans ($v_{s_2} > v_{s_1}$). This may shift the approval region near the cutoff v_N (e.g., $[v_N, v_N + \delta]$ for a small δ) to a rejection region under subgame s_2 if $v_{s_2} > v_N$, reducing the unconditional approval probability compared to no disclosure. This results in a higher overall lending cutoff v^* and less manipulation. The following proposition confirms this result.

Proposition 4.2. *Let $(\mathcal{S}, \tilde{\sigma})$ be an optimal policy in Theorem 4.1, and v_s be the lending cutoff under signal realization s under this policy, and v^* be the ex ante lending cutoff. Let v_N be the lending cutoff in the equilibrium with no disclosure. Then*

$$v^* > v_N > \inf_{s \in \mathcal{S}} v_s.$$

The result $v^* > v_N$ suggests that borrower data is used less frequently under optimal disclosure policies compared to the no disclosure equilibrium. Let c_N be the manipulation cutoff under the no disclosure equilibrium. Define $c_{\max} = \sup_{s \in \mathcal{S}} c_s$ and $c_{\min} = \inf_{s \in \mathcal{S}} c_s$. Proposition 4.2 also implies $c_{\max} > c_N > c_{\min}$. This proposition explains the differentiation across subgames. In the equilibrium with the highest borrower manipulation level ($c_s \to c_{\max}$), a high lending cutoff $v_s \to v^*$ is required

for approval of a borrower with $x = 1$, deterring the use of borrower data x in this subgame equilibrium. The cost is a high level of adverse selection, requiring the lender to finance inefficient borrower with high probability. However, the benefit is less adverse selection in other subgames (note that $c_{\min} < c_N$). Under the optimal policy, the positive effect dominates.

Finally, we show that for any optimal disclosure policy, there exists a deterministic policy that generates the same equilibria, as shown in the following lemma.

Lemma 4.2. *For any optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$, there must exist a deterministic policy $(\mathcal{S}, \sigma)$ with the same signal space $\mathcal{S}$, the same distribution of signals, and the same posterior belief for any signal $s \in \mathcal{S}$, such that the induced market equilibria are identical under any signal $s \in \mathcal{S}$ in both disclosure policies.*

Lemma 4.2 greatly simplifies our analysis and allows us to focus on the message function σ defined in Definition 3.2 rather than the distribution function $\tilde{\sigma}$ in Definition 3.1. The next theorem presents the characterization of optimal policies.

Theorem 4.2. *There exists a deterministic optimal policy $(\mathcal{S}, \sigma)$ such that:*

1. *$\sigma(v) = v_s$ for any v and s satisfying $v \in \operatorname{supp}(\pi_s)$, meaning the message space is chosen such that the message is the lender's equilibrium lending cutoff under any signal.*

2. *There exists a cutoff $v^* \in (0, \bar{v})$, such that:*

   (a) *$\sigma(v)$ is a weakly increasing function on $[0, v^*]$.*

   (b) *For any $s \in \mathcal{S}$, both R_s and A_s are nonempty.*

There are several implications from Theorem 4.2. First, as we discussed in Lemma 4.1, a borrower with data $x = 1$ is approved with positive probability in any subgame. This means that manipulation exists in any subgame. This implication rules out some disclosure policies. For example, if the lender chooses a policy that reveals whether v is below the investment I or not, financing the borrower is always inefficient if $v \leq I$ is revealed, regardless of the borrower's type. Therefore, if $v \leq I$ is revealed, there will be no loan approved and no manipulation. This disclosure policy is dominated by no disclosure because the lender does not benefit from states where $v \leq I$, but the borrower will choose to manipulate more in states where $v > I$. In fact, Theorem

4.2 shows that it is optimal to mix low states (where v is low) with high states (where v is high) to preserve uncertainty about the variable importance v in all subgames. Second, the message function σ is a weakly increasing function on $[0, v^*]$. This implies that lower values of v in rejection regions correspond to equilibria with lower lending cutoffs. This is intuitive because lower lending cutoffs indeed necessitate lower v values in rejection regions.

To make more precise predictions about the optimal policy, we impose a distributional assumption on the manipulation cost $K(c)$. With this assumption, we can derive optimal policies in closed form.

Assumption 2. *$cK(c)$ is a convex function of c on $[0, \bar{c}]$.*

With this assumption, the message function σ in Theorem 4.2 has a simple structure.

Theorem 4.3. *When Assumption 2 is satisfied, there exists a deterministic optimal policy $(\mathcal{S}, \sigma)$, that has the following structure on $[0, v^*]$: there exists a cutoff $v_a \in (0, v^*)$ such that*

$$\sigma(v) = \begin{cases} v_a & \text{if } v \in [0, v_a] \\ v & \text{if } v \in (v_a, v^*] \end{cases}.$$

Besides, $\sigma(v) = v_s$ for any $v \in \operatorname{supp}(\pi_s)$.

Theorem 4.3 characterizes the message function for the region where the variable importance v is lower than the lending cutoff v^*. In the region where v is higher than v^*, the characterization can be quite flexible. In any equilibrium, the borrower only cares about the probability that v exceeds v^*, not the specific value. Consequently, any message function that keeps the probability of v being higher than v^* consistent with the equilibrium can be optimal. The following proposition provides an example of optimal policy.

Proposition 4.3. *When Assumption 2 is satisfied, there exists a deterministic optimal policy with message function $\sigma(v)$. For $v \leq v^*$, function $\sigma(v)$ is defined in*

Theorem 4.3. For $v > v^*$, there exists $v_b \in (v^*, \bar{v})$, such that:

$$\sigma(v) = \begin{cases} v_a & \text{if } v \in (v^*, v_b] \\ \gamma(v) & \text{if } v \in (v_b, \bar{v}] \end{cases}.$$

Here $\gamma : [v_b, \bar{v}] \to [v_a, v^*]$ is a continuous, strictly increasing function satisfying the following differential equation:

$$\gamma(v) = \left[1 + \frac{1-\mu}{\mu} K\left(\frac{b \cdot g(v)}{\gamma'(v)\, g(\gamma(v)) + g(v)}\right)\right] I,$$

with boundary conditions $\gamma(v_b) = v_a$,$\gamma(\bar{v}) = v^*$, and $\frac{g(v_b)}{\gamma'(v_b) g(\gamma_a)} = \frac{G(v_b) - G(v^*)}{G(v_a)}$.

Based on the above results, for any given v^*, we can solve for the message function (including boundaries v_a and v_b) using the conditions in Theorem 4.3 and Proposition 4.3. Thus, finding the optimal message function reduces to finding the optimal cutoff v^*. The lender's problem in Proposition 3.1 becomes a one-dimensional maximization problem,

$$\max_{v^*} W,$$

where W is defined by (11) under a disclosure policy with signal space $\mathcal{S} = [v_a, v^*]$ and message function σ. The value v_a and message function σ are solved by Theorem 4.3 and Proposition 4.3. Figure 4 presents a numerical example of an optimal policy.

With the optimal policy, under the discrete signal v_a, values of the variable importance in $[0, v_a]$ are pooled with an interval above v^*, $(v^*, v_b]$. In this subgame, the lending cutoff is v_a, which is the upper bound of the posterior belief's support below v^*. Beyond this discrete signal, the optimal policy differentiates other states more precisely. Specifically, each state in $(v_a, v^*]$ is pooled with an infinitesimal point in $(v_b, \bar{v}]$, and all states in $(v_a, v^*]$ are separated from each other. Additionally, in the posterior equilibrium containing any state $v \in (v_a, v^*]$, the lending cutoff is exactly v, making the lender indifferent between lending or not in this state. For each subgame with signal $s = v_s$, the true variable importance is either v_s, leading the lender to not lend, or it is above v^*, prompting the lender to lend to a borrower with data $x = 1$. The equilibrium approval probability in this subgame is

$$\frac{1}{b} K^{-1}\left(\frac{\mu}{1-\mu}\left(\frac{v_s}{I} - 1\right)\right),$$

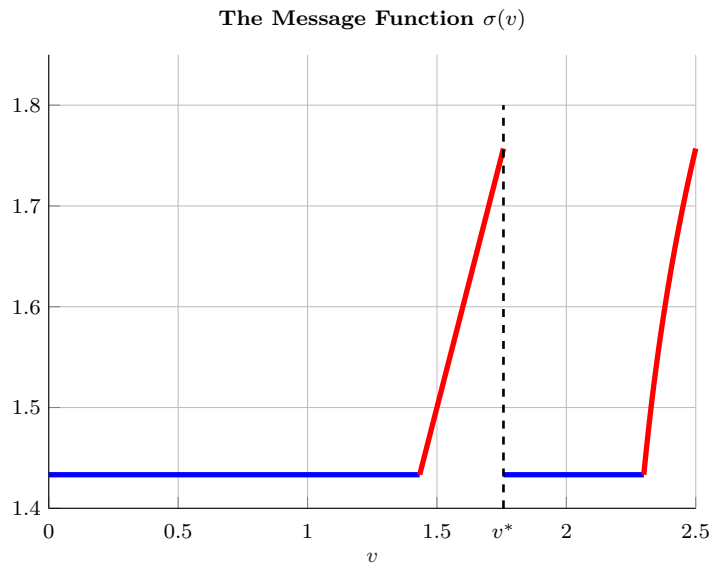

Figure 4: **Optimal Policy**. We numerically calculate an example of the optimal policy. In this exercise, we use the following parameters: $\mu = 0.35$, $\bar{c} = 2$, $b = 1.7$, $I = 1$, $\bar{v} = 2.5$. Both c and v follow uniform distribution. In this exercise, $v_a = 1.433$, $v^* = 1.757$, $v_b = 2.299$. As a benchmark, in this case, $v_N = 1.581$.

which is an increasing function of the signal $s = v_s$.

4.1 Implementation and Examples

We've characterized the optimal disclosure policy and will now discuss potential real-world implementations. In practice, algorithmic transparency can involve disclosures on human involvement, data, models, and more. Disclosures about data and models, in particular, have been emphasized in various settings. Although the discussion on algorithmic transparency is still in its early stages, our model provides insights into implementations that can improve the payoffs of algorithm users by mitigating data manipulation.

In the following examples, we illustrate how our optimal policy can be implemented in real-world scenarios. In the context of training data transparency, the optimal policy can be implemented by disclosing potentially biased training data. In the context of model disclosure, the optimal policy corresponds to understating the variable importance in machine learning models.

Disclosing Samples of Training Data with Added Noise

Suppose the parameter v represents the correlation between a certain variable (e.g., phone operating system) and an outcome (e.g., credit quality). While v is unknown to the public, it is learned by the lender after exploring the training data. Our optimal policy, as detailed in Proposition 4.3, can be implemented as follows: The lender commits to releasing the training data to the public, allowing them to estimate the correlation. If the correlation v in the original training data is below a cutoff v^*, which means the variable is not very informative, the lender releases the original data reflecting the true correlation v. If the correlation exceeds v^*, the lender releases the training data with added white noise, such that the correlation becomes $\gamma(v)$, which is lower than v. Intuitively, the lender releases the true training data if the correlation is low, but a noisier version with a lower correlation $\gamma(v) < v$ if the true correlation is high. This approach ensures that when the borrower observes the released training data, they remain uncertain whether it reflects the true correlation of the original data or is a noisy version. Our results highlight that committing to strategically releasing noisier training data may be optimal.

Understating Variable Importance in Machine Learning Models

Variable importance (or feature importance) is a crucial measure in machine learning models, reflecting a variable's effectiveness.[16] Given its informativeness, variable importance can also be used to enhance algorithmic transparency. Suppose the parameter v in our model represents variable importance in a machine learning model used by the lender. Our model suggests that if the variable importance is below a threshold v^*, the lender should disclose the true value. However, if it is above the threshold, the lender should disclose a modified, lower value $\gamma(v)$. Essentially, the lender discloses the true variable importance of the machine learning model when it is low and understates it when it is high. This creates uncertainty for the borrower about whether the reported variable importance is true or understated, preserving uncertainty in the posterior belief.

[16]For example, if for a given model with two input features "f_1" and "f_2", the variable importances are $\{f_1 = 5.8, f_2 = 2.5\}$, then the feature "f_1" is more "important" to the model than feature "f_2" (source: Google Developers, https://developers.google.com/machine-learning/decision-forests/variable-importances).

5 Extensions

5.1 Commitment: Information vs Decision

Our paper focuses on the role of algorithmic disclosure in mitigating gaming behavior, and Remark 3.1 shows that partial disclosure of the algorithm is optimal. The discussion in Section 3.3 demonstrates that committing to a higher lending cutoff can also improve the lender's payoff. In practice, regulating the decision-making process of algorithm users is also a significant part of the public debate.[17] In this section, we discuss the role of commitment to information (algorithm disclosure) versus commitment to action (lending decisions). We show that even when the lender commits to the optimal lending decision, she might still receive higher payoff by disclosing additional information about the predictive model, suggesting that the information disclosure can not be fully replaced by the lender's stronger commitment power on lending decisions.

To begin our analysis, consider a benchmark case of committing to lending actions. Suppose the lender commits to lending to a borrower with data $x = 1$ when $v \geq z$ and rejecting the borrower with $x = 0$ directly.[18] Then the lender's problem is:

$$\max_{z} \int_{z}^{\bar{v}} \left[\mu v - \mu I - (1-\mu) K(c) I\right] dG(v),$$

where c is determined by $b \cdot (1 - G(z)) = c$. Let the solution to this problem be $z^* = v_c$, and the borrower's manipulation cutoff in this case be

$$c_c = b \cdot (1 - G(v_c)).$$

We are interested in whether some additional disclosure about the state v can help the lender even if she has full commitment power over lending decisions. After committing to the optimal lending decision v_c, suppose the lender can choose a disclosure policy $(\mathcal{S}, \sigma)$ and disclose information about the true state accordingly, before the borrower

[17] For example, Fintech lenders are often required to exclude certain variables from their decision-making process and may set goals for default rate.

[18] Note that we can consider a more flexible lending rule. For instance, the lender could lend to a borrower with data $x \in \{0, 1\}$ when $v \geq v_x$, allowing the lender to lend to a borrower with $x = 0$. This flexibility might improve the lender's profit by deterring the borrower's incentive to manipulate. However, under certain distributional assumptions about $K(c)$, it is optimal for the lender to commit to always rejecting a borrower with $x = 0$. This case also allows for a comparable analysis with our baseline model.

chooses his manipulation strategy. The following proposition shows that the lender might receive higher payoff from additional disclosure about the true state v.

Proposition 5.1. *Suppose the lender commits to only lending to the borrower with data $x = 1$ when the true state $v \geq v_c$. If the function $xK(x)$ is strictly concave at $x = c_c$, then the lender can receive a higher payoff by disclosing additional information about v compared to disclosing nothing.*

Although the key friction in our model is the lender's lack of commitment problem, Proposition 5.1 suggests that the benefits of algorithmic disclosure cannot simply be replaced by committing to a lending decision. Instead, the lender can achieve a higher payoff by disclosing additional information about the predictive model even if she has full commitment power on lending decisions.

5.2 Costly Verification

In practice, some data manipulation behavior is classified as fraudulent activities (like misreporting personal information), and lenders can costly verify fraudulent activities using various methods, which is another way of mitigating data manipulation. In this extension, we consider how the disclosure policy interacts with costly verification in the lender's problem.

We introduce one more assumption to the main model. After a signal s is disclosed to the market, once the lender receives a loan application from a borrower, she has the option to verify and reveal the borrower's true type X by paying a cost $t > 0$. As we discussed in the main model, only a borrower with data $x = 1$ has non-zero probability of getting the loan, so we just need to consider the lender's choice of revealing the true type of a borrower with data $x = 1$. Consider an subgame with posterior belief π_s, a borrower with $X = 0$ chooses to manipulate if and only if his manipulation cost c is lower than a cutoff, denoted as c_s. Then in equilibrium, if the lender verifies the borrower's type, her profit is

$$W_V = \max\{\mu v - \mu I, 0\} - (\mu + (1-\mu) K(c_s)) t.$$

If the lender chooses to not verify the borrower's type, her profit is

$$W_{NV} = \max\{\mu v - \mu I - (1-\mu) K(c_s) I, 0\}.$$

When $t > \frac{(1-\mu)K(c_s)}{\mu+(1-\mu)K(c_s)}I$, $W_{NV} > W_V$ for all v. In this case, the lender will never verify the borrower's true type. When $t < \frac{(1-\mu)K(c_s)}{\mu+(1-\mu)K(c_s)}I$,

$$W_V > W_{NV} \Longleftrightarrow v > v_e = I + \frac{\mu + (1-\mu)\,K\,(c_s)}{\mu} t,$$

and the borrower's type must be verified before he successfully gets the loan. However, this implies that the borrower will never manipulate his data x, and thus the lender has no incentive to verify the borrower's type. Then $t < \frac{(1-\mu)K(c_s)}{\mu+(1-\mu)K(c_s)}I$ can't be true in the equilibrium of any subgame s. Then if in a subgame s, the lender verifies the borrower's type with positive probability, we must have $t = \frac{(1-\mu)K(c_s)}{\mu+(1-\mu)K(c_s)}I$, and thus there exists at most one subgame s, such that the verification happens with positive probability. The following theorem presents the complete characterization of an optimal policy with verification.

Theorem 5.1. *With costly verification, there exists t^v such that:*

1. *When $t \geq t^v$, the lender will never verify the borrower's type, and the optimal disclosure policy is the same as in the main model.*

2. *When $t < t^v$, there exists $v_v \in (0, \bar{v}]$, such that the optimal policy can be implemented by two steps:*

   (a) *The lender first discloses if $v > v_v$ or not.*

   (b) *If $v > v_v$, then the lender will verify the type of the borrower who has data $x = 1$ with probability $p^* = 1 - \frac{c_v}{b}$, and lends to the borrower if his type is not verified or verified to be $X = 1$.*

   (c) *If $v \leq v_v$, there is no verification, then information about v is further disclosed according to a policy $(\mathcal{S}^v, \sigma^v)$, where $(\mathcal{S}^v, \sigma^v)$ is an optimal disclosure policy characterized in Theorem 4.2 with a modified prior belief $G^*(v)$, where*

$$G^*(v) = \min\left\{\frac{G(v)}{G(v_v)}, 1\right\}.$$

Theorem 5.1 shows that the disclosure policy and verification technology interact in a simple way: when the true state v is sufficiently high ($v > v_v$), only verification is used to disincentivize manipulation behavior, and no further disclosure is

needed; while when the v is low ($v \leq v_v$), only disclosure is used to disincentivize the manipulation behavior and verification is never used.

6 Conclusion

We study the optimal algorithmic disclosure in a lending market where a Fintech lender uses privately observed predictive models to screen a borrower. The input to the predictive model is data collected from the borrower, which can be strategically manipulated. The optimal disclosure features partial disclosure, where information about the predictive model is partially disclosed to the borrower, differentiating the posterior lending market equilibrium by data manipulation levels. Under the optimal disclosure policy, the lender uses borrower data less intensively in her lending decisions, reducing the average data manipulation level and improving efficiency. Despite receiving additional information from the lender, the borrower's posterior belief remains quite uncertain. Algorithmic disclosure can improve the lender's payoff even when she has additional commitment power on lending decisions or can verify the borrower's type at a cost.

References

Azarmsa, E. and L. W. Cong (2020). Persuasion in relationship finance. *Journal of Financial Economics 138*(3), 818–837.

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*.

Berg, T., V. Burg, A. Gombović, and M. Puri (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies 33*(7), 2845–2897.

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.

Inostroza, N. and A. Tsoy (2022). Optimal information and security design. *Available at SSRN 4093333*.

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*.

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.

Perez-Richet, E. and V. Skreta (2022). Test design under falsification. *Econometrica 90*(3), 1109–1142.

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*.

Appendix

A Proofs

A.1 Proofs in Section 2

No Disclosure

Based on the distributional assumptions on the manipulation cost c and the probability of success v, the two equilibrium conditions are $b(1-v_N)=c_N$ and $\mu v_N V=(\mu+(1-\mu)c_N)I$. The unique solution is $\left(v_N=\frac{\mu I+(1-\mu)Ib}{\mu V+(1-\mu)Ib}, c_N=b\cdot\frac{\mu(V-I)}{\mu V+(1-\mu)Ib}\right)$.

A Binary Signal

When $v\in A$, let's verify that in equilibrium $v_1=0.54$, $c_1=0.34$, and $W_1=0.19$. The two equilibrium conditions are $\text{Prob}(v>v_1|A)\cdot b=c_1$, and $\mu v_1 V=(\mu+(1-\mu)\,\text{Prob}(c\leq c_1))I$. Since $A=[0,0.54)\cup(0.64,0.91)$, and $v_1=0.54$, $c_1=0.34$, the first condition becomes $\frac{0.91-0.64}{0.91-0.64+0.54}\times 1=0.34$, and the second condition becomes $0.3\times 0.54\times 10=(0.3+0.7\times 0.34)\times 3$. Both condition hold [19] under $A=[0,0.54)\cup(0.64,0.91)$, and $v_1=0.54$, $c_1=0.34$. Similarly, when $v\in A^{\complement}$, the two equilibrium conditions are $\frac{1-0.91}{1-0.91+0.64-0.54}\times 1=0.48$, and $0.3\times 0.64\times 10=(0.3+0.7\times 0.48)\times 3$. Both conditions hold under $A^{\complement}=[0.54,0.64]\cup[0.91,1]$ and $v_2=0.64$, $c_2=0.48$. Regarding the lender's profit, $W_1=0.3\times 10\times\int_{0.64}^{0.91}(v-0.54)\,dv=0.019$, and $W_2=0.3\times 10\times\int_{0.91}^{1}(v-0.64)\,dv=0.09$. So the total profit is $W_s=W_1+W_2=0.19+0.09=0.28$.

A.2 Proof of $\gamma_s(1,c)=0$

Suppose in a subgame s, a borrower with type X and manipulation cost $c>0$ chooses $\gamma_s(1,c)>0$ in equilibrium. For the borrower, his payoff of not manipulating data is $d_s(1)\cdot b$, while the payoff of manipulating data is $d_s(0)\cdot b-c$. Then he chooses to manipulate the data only if

$$d_s(0)\cdot b-c\geq d_s(1)\cdot b\Longleftrightarrow d_s(0)-d_s(1)>\frac{c}{b}.$$

[19] Since we just want to illustrate the key ideas, we only keep the first two decimal places of the results in this example, so the conditions only approximately hold.

Since $\gamma_s(1,c) > 0$, we must have

$$d_s(0) - d_s(1) \geq \frac{c}{b} > 0 \tag{18}$$

in equilibrium. For a borrower with type $X = 0$ and manipulation cost c_1, the payoff of not manipulating data is $d_s(0) \cdot b$, while the payoff of manipulating his data is $d_s(1) \cdot b - c_1$. Condition (18) implies that $d_s(1) \cdot b - c_1 < d_s(0) \cdot b$ for any c_1, which implies that in this equilibrium, a borrower with type $X = 0$ never chooses to manipulate data almost surely. Then for the lender, a borrower with data $x = 1$ will always has better quality than a borrower with data $x = 0$, which means $\alpha_s(1,v) \geq \alpha_s(0,v)$ for any v and s, and thus $d_s(1) \geq d_s(0)$. This is a contradiction to the condition (18). So in any equilibrium s, we must have $\gamma_s(1,c) = 0$ for any $c > 0$.

A.3 Proof of Lemma 3.1

We have already shown that $\gamma_s(1,c) \equiv 0$, then the lender's profit of lending to a borrower with data $x = 0$, represented by (3), must be negative or zero. When the profit is negative, we must have $\alpha_s(0,v) \equiv 0$ for any v and s. When the profit equals zero, it must be the case that observing a borrower with $x = 0$ is a zero-probability event on the equilibrium path. Then without loss of generality, we have $\alpha_s(0,v) \equiv 0$ for any v and s. This implies

$$d_s(0) = \int_0^{\bar{v}} \alpha_s(0,v)\,\pi_s dv = 0.$$

For a borrower with type $X = 0$ and manipulation cost c, the optimization problem (6) implies that he chooses to manipulate his data only if

$$d_s(1)\,b - c \geq d_s(0)\,b = 0 \Longleftrightarrow c \leq c_s = d_s(1)\,b.$$

When the lender faces to a borrower with data $x = 1$, if she makes the loan, her profit (2) becomes $\mu v - [\mu + (1-\mu)\,K(c_s)] \cdot I$, and she makes the loan ($\alpha_s(1,v) > 0$) only if $v \geq v_s = \frac{[\mu+(1-\mu)K(c_s)]\cdot I}{\mu}$. Since $d_s(1) = \int_0^{\bar{v}} \alpha_s(1,v)\,\pi_s dv$, we must have

$\text{Prob}\left(v > v_s | \, s\right) \leq d_s\left(1\right) \leq \text{Prob}\left(v \geq v_s | \, s\right)$. The lender's profit in the subgame s is

$$\begin{aligned} W_s &= \int_{v_s}^{\bar{v}} \left(\mu v - \left[\mu + \left(1-\mu\right) K\left(c_s\right)\right] \cdot I\right) \pi_s\left(v\right) dv \\ &= \int_{v_s}^{\bar{v}} \left(\mu v - \mu v_s\right) \pi_s\left(v\right) dv. \end{aligned}$$

A.4 Proof of Proposition 3.1

The lender's original problem is to find a disclosure policy $(\mathcal{S}, \tilde{\sigma})$, which induces a distribution of the signal $F\left(s\right)$, to maximize the expected profit

$$W = \mu \int_{s \in \mathcal{S}} \mathbb{E}\left[\left(v - v_s\right)^+ | s\right] dF\left(s\right). \tag{19}$$

It's standard in the literature to work with the distribution of posteriors $\{\pi_s\}_{s \in \mathcal{S}}$ instead of disclosure policies directly (Kamenica and Gentzkow (2011)), but an additional Bayesian plausibility condition is needed:

$$\int_{s \in \mathcal{S}} \pi_s dF\left(s\right) = g\left(v\right). \tag{20}$$

Under each posterior π_s, the equilibrium conditions are

$$\text{Prob}\left(v > v_s | \, s\right) \leq d_s\left(1\right) \leq \text{Prob}\left(v \geq v_s | \, s\right),$$

$$c \leq c_s = b \cdot d_s\left(1\right),$$

and

$$v_s = \frac{\left[\mu + \left(1-\mu\right) K\left(c_s\right)\right] \cdot I}{\mu}.$$

These conditions jointly imply

$$\text{Prob}\left(v > v_s | \, s\right) \leq \frac{K^{-1}\left(\frac{\mu v_s}{I(1-\mu)} - \frac{\mu}{1-\mu}\right)}{b} \leq \text{Prob}\left(v \geq v_s | \, s\right), \tag{21}$$

where $\text{Prob}\left(\cdot | \, s\right)$ is the probability function under the posterior π_s. So the lender's problem is equivalent to finding a distribution of posteriors, denoted by $\mathcal{S}$ and $\{F, \pi_s\}$, to maximize (19), subject to conditions (20) and (21).

A.5 Proof of Lemma 3.2

Consider a policy $(\mathcal{S}, \tilde{\sigma})$ with distribution of posterior beliefs $\{F, \pi_s\}$, and suppose there exist two distinct realizations $s_1, s_2 \in \mathcal{S}$, such that $(v_{s_1}, c_{s_1}) = (v_{s_2}, c_{s_2})$. Then consider a new policy $(\mathcal{S}', \tilde{\sigma}')$, where $\mathcal{S}' = \{s_0'\} \cup \mathcal{S} \backslash \{s_1, s_2\}$ and

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

for all $v \in [0, \bar{v}]$ and $s \in \mathcal{S}'$. Note that

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

for any $v \in [0, \bar{v}]$ 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 belief is the same under the two policies, i.e., $\pi_s = \pi_s'$. So the lending market equilibrium is the same for any $s \in \mathcal{S} \backslash \{s_1, s_2\} = \mathcal{S}' \backslash \{s_0'\}$ in these two policies. Besides, under the signal realization s_0' in the new disclosure policy, we can verify that the equilibrium manipulation cutoff c_s and lending cutoff v_s are all unchanged by verifying the equilibrium conditions for the subgames s_1 and s_2 under the policy $(\mathcal{S}, \tilde{\sigma})$.

A.6 Proof of Lemma 3.3

Given v_δ, the borrower's manipulation cutoff c_δ satisfies $b(1 - G(v_\delta)) = c_\delta$. The lender's payoff is

$$W_\delta = \int_{v_\delta}^{\bar{v}} [\mu v - (\mu + (1-\mu) K(c_\delta)) I] g(v) dv,$$

so

$$\left.\frac{dW_\delta}{d\delta}\right|_{\delta=0} = -[\mu v_N - (\mu + (1-\mu) K(c_N)) I] g(v) + \int_{v_\delta}^{\bar{v}} \left(-(1-\mu) I g(v) \left.\frac{dK(c_\delta)}{d\delta}\right|_{\delta=0} \right) dv. \tag{22}$$

The equilibrium condition of the no disclosure equilibrium is $\mu v_N - (\mu + (1-\mu) K(c_N)) I = 0$. And it's clear that $\left.\frac{dK(c_\delta)}{d\delta}\right|_{\delta=0} < 0$ because K' is always strictly positive. Then we must have $\left.\frac{dW_\delta}{d\delta}\right|_{\delta=0} > 0$.

A.7 Proof of Proposition 3.2

In the no disclosure equilibrium, the belief about v is the same as the prior, the lending market equilibrium is characterized by c_N and v_N, which satisfy conditions (8), (9) and (10). Let the lender's payoff be W_N in the no disclosure case. Now let's consider the following deterministic disclosure policy $(\mathcal{S}', \sigma')$, where $0 < \epsilon_1, \epsilon_2 \ll 1$, $\mathcal{S}' = \{s_1', s_2'\}$, and the message function is

$$\sigma'(v) = \begin{cases} s_1' & v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}] \\ s_2' & v \in (v_1' v_N) \cup (v_N + \epsilon_1, \bar{v} - \epsilon_2) \end{cases},$$

where $v_1' < v_N$ satisfies

$$\frac{\text{Prob}\left(v \in [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right)}{\text{Prob}\left(v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, 1]\right)} b = c_N.$$

Denote the equilibria under signals s_1' and s_2' as $\left(v_{s_1'}, c_{s_1'}\right)$ and $\left(v_{s_2'}, c_{s_2'}\right)$, respectively, it's easy to verify

$$\left(v_{s_1'}, c_{s_1'}\right) = \left(v_{s_2'}, c_{s_2'}\right) = (v_N, c_N).$$

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

$$\begin{aligned} W' =& \text{Prob}\left(v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_1'} \left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_1'}\right\}} \right] \\ &+ \text{Prob}\left(v \in (v_1', v_N) \cup (v_N + \epsilon_1, \bar{v} - \epsilon_2)\right) \cdot \mathbb{E}^{s_2'} \left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_2'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_2'}\right\}} \right]. \end{aligned} \tag{23}$$

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

$$\sigma''(v) = \begin{cases} s_1'' & v \in [v_N, v_N + \epsilon_1] \\ s_2'' & v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}] \\ s_3'' & v \in (v_1', v_N) \cup (v_N + \epsilon_1, \bar{v} - \epsilon_2) \end{cases}.$$

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 on $(v_1', v_N) \cup (v_N + \epsilon_1, \bar{v} - \epsilon_2)$. When signal s_1'' is disclosed, the borrower knows that the true state $v \in [v_N, v_N + \epsilon_1]$. In the no discloure case, the lender's payoff in state $[v_N, v_N + \epsilon_1]$ is close to zero (in order $o(\epsilon_1)$), as v_N is the equilibrium cutoff in lending decisions. So the difference between the lender's payoff in the subgame s_1'' under the disclosure policy $(\mathcal{S}'', \sigma'')$, and the lender's payoff in the no disclosure case when $v \in [v_N, v_N + \epsilon_1]$ is close to zero. However, the increase in lender'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, and thus the lender's payoff increases under the disclosure policy $(\mathcal{S}'', \sigma'')$. To see this, note that the lender's payoff under $(\mathcal{S}'', \sigma'')$ is

$$\begin{aligned} W'' =& \text{Prob}\left(v \in [v_N, v_N + \epsilon_1]\right) \cdot \mathbb{E}^{s_1''} \left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1''} \right) \right) I \right) \cdot \mathbb{1}_{\left\{ v \geq v_{s_1''} \right\}} \right] \\ &+ \text{Prob}\left(v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_2''} \left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_2''} \right) \right) I \right) \cdot \mathbb{1}_{\left\{ v \geq v_{s_2''} \right\}} \right] \\ &+ \text{Prob}\left(v \in (v_1', v_N) \cup (v_N + \epsilon_1, \bar{v} - \epsilon_2)\right) \cdot \mathbb{E}^{s_3''} \left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_3''} \right) \right) I \right) \cdot \mathbb{1}_{\left\{ v \geq v_{s_3''} \right\}} \right]. \end{aligned} \tag{24}$$

It's obvious that the last term in (24) equals the last term in (23), because equilibria

under signal realizations s_3'' and s_2' are the same. Then

$$
\begin{aligned}
& W'' - W' \\
=& \text{Prob}\left(v \in [v_N, v_N + \epsilon_1]\right) \cdot \mathbb{E}^{s_1''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_1''}\right\}}\right] \\
& + \text{Prob}\left(v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_2''}\right\}}\right] \\
& - \text{Prob}\left(v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_1'}\right\}}\right] \\
\geq& \text{Prob}\left(v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_2''}\right\}}\right] \\
& - \text{Prob}\left(v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_1'}\right\}}\right].
\end{aligned}
$$

Note that $v_{s_1'} = v_N$, we know

$$
\begin{aligned}
& \text{Prob}\left(v \in [0, v_1'] \cup [v_N, v_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq v_{s_1'}\right\}}\right] \\
=& \begin{array}{c} \text{Prob}\left(v \in [v_N, v_N + \epsilon_1]\right) \cdot \mathbb{E}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_N\right)\right) I\right) | v \in [v_N, v_N + \epsilon_1]\right] \\ + \text{Prob}\left(v \in [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_N\right)\right) I\right) | v \in [1 - \epsilon_2, 1]\right] \end{array}.
\end{aligned}
$$

Then

$$
\begin{aligned}
W'' - W' \geq & \text{Prob}\left(v \in [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \left[(1-\mu) I\left(K\left(c_N\right) - K\left(c_{s_2''}\right)\right)\right] - \\
& \text{Prob}\left(v \in [v_N, v_N + \epsilon_1]\right) \cdot \mathbb{E}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(c_N\right)\right) I\right) | v \in [v_N, v_N + \epsilon_1]\right].
\end{aligned}
$$

In the equilibrium of subgame s_2'',

$$
\begin{aligned}
c_{s_2''} &= b \frac{\text{Prob}\left(v \in [\bar{v} - \epsilon_2, \bar{v}]\right)}{\text{Prob}\left(v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}]\right)} \\
&= b \frac{1 - G\left(\bar{v} - \epsilon_2\right)}{1 - G\left(\bar{v} - \epsilon_2\right) + G\left(v_1'\right)}.
\end{aligned}
$$

Fix ϵ_2, and let $\epsilon_1 \to 0$, then

$$\begin{aligned}\frac{c_N - c_{s_2''}}{b} &= \frac{1 - G(\bar{v} - \epsilon_2) + G(v_N + \epsilon_1) - G(v_N)}{1 - G(\bar{v} - \epsilon_2) + G(v_1') + G(v_N + \epsilon_1) - G(v_N)} - \frac{1 - G(\bar{v} - \epsilon_2)}{1 - G(\bar{v} - \epsilon_2) + G(v_1')} \\ &= \frac{G(v_1')}{[1 - G(\bar{v} - \epsilon_2) + G(v_1')]^2} g(v_N)\epsilon_1 + o(\epsilon_1).\end{aligned}$$

Then

$$\begin{aligned}&\frac{W'' - W'}{\text{Prob}(v \in [\bar{v} - \epsilon_2, \bar{v}])} \\ \geq\ & \frac{\left[(1-\mu) I\left(K(c_N) - K\left(c_{s_2''}\right)\right)\right] -}{\frac{\text{Prob}(v \in [v_N, v_N + \epsilon_1])}{\text{Prob}(v \in [\bar{v} - \epsilon_2, \bar{v}])} \mu\left(\mathbb{E}(v | v \in [v_N, v_N + \epsilon_1]) - v_N\right)} \\ \geq\ & (1-\mu) I K'(c_N) b \frac{G(v_1')}{[1 - G(\bar{v} - \epsilon_2) + G(v_1')]^2} g(v_N)\epsilon_1 + o(\epsilon_1) - \frac{g(v_N)\epsilon_1 + o(\epsilon_1)}{1 - G(\bar{v} - \epsilon_2)} \mu \epsilon_1.\end{aligned}$$

It's clear that when $\epsilon_1 \to 0$, $W'' - W' > 0$, which means that the no disclosure policy is dominated by our new disclosure policy $(\mathcal{S}'', \sigma'')$.

A.8 Proof of Lemma 3.4

The full disclosure policy $(\mathcal{S}, \sigma)$ can be implemented by the space $\mathcal{S} = [0, \bar{v}]$ and a deterministic message function $\sigma(v) = v$. In this case, the true state v is perfectly revealed to the borrower. For any signal realization $s = v > I$, the lending market equilibrium of subgame s must satisfy $v_s = v$. Then the lender's payoff must be zero in this subgame. For any $v \leq I$, the lender will never make the loan, and thus her profit is zero. In summary, the lender's payoff is always zero for any subgame $s \in \mathcal{S}$, and thus her total payoff must be $W_F = 0$.

A.9 Proof of Lemma 4.1

We prove this result by contradiction. Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, with distribution of posteriors $\{F, \pi_s\}$.

First, Assumption 1 implies that in any subgame s, the approval probability $d_s(x)$ must be less than 1. Suppose for a signal s_0, R_{s_0} is an empty set, then we must have $\text{Prob}(v = v_{s_0} | s_0) > 0$, and the lender rejects a borrower with data $x = 1$ only when

$v = v_{s_0}$. R_{s_0} being an empty set means that $\alpha_{s_0}(1, v_{s_0}) \in (0,1)$. Then consider an alternative policy $(\mathcal{S}_1, \tilde{\sigma}_1)$, with $\mathcal{S}_1 = \mathcal{S} - \{s_0\}\{s_1, s_2\}$ and posterior belief:

$$\tilde{\pi}_s = \begin{cases} \delta_{v_{s_0}} & \text{if } s = s_1 \\ \frac{\left(1-\alpha_{s_0}\left(1,v_{s_0}\right)\right)\delta_{v_{s_0}} + \pi_{s_0}\mathbb{1}_{(v_{s_0},\bar{v}]}(v)}{1-\alpha_{s_0}\left(1,v_{s_0}\right)\cdot \text{Prob}\left(v=v_{s_0}|s_0\right)} & \text{if } s = s_2 \\ \pi_s & \text{if } s \in \mathcal{S} - \{s_0\} \end{cases}.$$

$(\mathcal{S}_1, \tilde{\sigma}_1)$ represents a simple policy, and it reveals information in the same way as $(\mathcal{S}, \tilde{\sigma})$, except that if $s = s_0$ is revealed, and if $v = v_{s_0}$, then with probability $\alpha_{s_0}(1, v_{s_0})$, the true state v_{s_0} is revealed. We can verify that the lender's payoff improves under signal s in this case, because the approval probability decreases in the subgame s, leading to a lower manipulation cutoff and less manipulation. However, the lender does not lose profit from financing, because she earns zero from financing the borrower when $v = v_s$. Since the improvement in the lender's payoff is positive in this case, the measure of all signals s satisfying R_s being an empty set must be zero.

Second, suppose for a signal s_0, A_{s_0} is an empty set, then in this case, we must have $\text{supp}(\pi_{s_0}) \leq I$, and thus the lender will always reject the borrower. It is obvious that combing this signal with any other signal will improve the lender's payoff, as the manipulation will be less severe on average, and the lender will not lose any profit from financing. As a result, the measure of the signal s that satisfies A_s being an empty set must be zero.

For the last point, we prove the result by contradiction. Note that $\alpha_s(1, v) \in (0,1)$ only if $v = v_s$. Suppose there exists a set $H \subset \mathcal{S}$, such that $\text{Prob}(H) > 0$ and for any $s \in H$, $v_s \in A_s$. Take any $s_0 \in H$, if $\text{Prob}(v_{s_0}|s_0) = k > 0$, and $\alpha_{s_0}(1, v_{s_0}) \in (0,1)$, then following the proof of the first part of this lemma, we can show that there is an alternative policy that generates higher lender's payoff for the subgame s_0 while keeping the lender's payoff unchanged for other states.

If $\text{Prob}(v_{s_0}|s_0) = k = 0$, since $v_{s_0} \in A_{s_0}$ then for any $\epsilon > 0$, we must have $\text{Prob}(v \in (v_{s_0}, v_{s_0} + \epsilon)|s_0) > 0$. Following the proof of Proposition 3.2, we know there must exist an alternative disclosure policy that reveals the states when $v \in (v_{s_0}, v_{s_0} + \epsilon)$ for a ϵ being small enough, such that the lender's payoff improves under signal s_0. We omit the details of the proof for this part as it essentially replicates the proof of Proposition 3.2.

Then in summary, for any set H satisfying $v_s \in A_s$ for all $s \in H$, we must have

$\text{Prob}(H) = 0$. Then for almost all $s \in \mathcal{S}$, we must have $A_s = \{v \in \text{supp}(\pi_s) | \alpha_s(1, v) = 1\}$.

A.10 Proof of Theorem 4.1

Suppose there exists an optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$ and it induces the distribution of posteriors $\{F, \pi_s\}$. If $\mathcal{S}$ is a singleton, then the policy is just the no disclosure policy. In this case, let $v^* = v_N$, then this result is obviously true. If $\mathcal{S}$ is not a singleton, let's prove the result by contradiction. For any s, let $\bar{R}_s = \sup R_s$ and $\underline{A}_s = \inf A_s$. If for all $s_1, s_2 \in \mathcal{S}$, $\bar{R}_{s_1} \leq \underline{A}_{s_2}$, then the statement must be true. If there exist s_1 and s_2, such that $\bar{R}_{s_1} > \underline{A}_{s_2}$, by the definition of $\bar{R}_s$ and $\underline{A}_s$, there must exist $v_1 \in R_{s_1}$ and $v_2 \in A_{s_2}$, such that $v_1 > v_2$. Since $v_1 \in \text{supp}(\pi_{s_1})$ and $v_2 \in \text{supp}(\pi_{s_2})$, then for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, and open balls $B(v_1, \epsilon_1)$, $B_2(v_2, \epsilon_2)$, we must have $\text{Prob}(B(v_1, \epsilon_1) | s_1) > 0$ and $\text{Prob}(B(v_2, \epsilon_2) | s_2) > 0$. Let us choose $\epsilon_1, \epsilon_2 < \frac{v_2 - v_1}{3}$, then $\inf B(v_1, \epsilon_1) > \sup B_2(v_2, \epsilon_2)$. Denote $\text{Prob}(B(v_1, \epsilon_1) | s_1) = K_1 > 0$ and $\text{Prob}(B(v_2, \epsilon_2) | s_2) = K_2 > 0$. If $f(s_1) K_1 \geq f(s_2) K_2$,[20] let's consider the following distribution of posteriors: $\left\{\hat{F}, \hat{\pi}_s\right\}$, where $\hat{\mathcal{S}} = \mathcal{S}$, $\hat{F} = F$, and

$$\hat{\pi}_s = \begin{cases} \pi_{s_1} + \frac{f(s_2)}{f(s_1)} \pi_{s_2} \mathbb{1}_{B(v_2, \epsilon_2)}(v) - \frac{f(s_2) K_2}{f(s_1) K_1} \pi_{s_1} \mathbb{1}_{B(v_1, \epsilon_1)}(v) & \text{if } s = s_1, \\ \pi_{s_2} - \pi_{s_2} \mathbb{1}_{B(v_2, \epsilon_2)}(v) + \frac{K_2}{K_1} \pi_{s_1} \mathbb{1}_{B(v_1, \epsilon_1)}(v) & \text{if } s = s_2, \\ \pi_s & o.w. \end{cases}$$

We can check that $\left\{\hat{F}, \hat{\pi}_s\right\}$ is Bayes-plausible, and thus there exists a disclosure policy $\left(\hat{S}, \hat{\sigma}\right)$ that induces this distribution of posteriors. But now in the new policy $\left(\hat{S}, \hat{\sigma}\right)$, $v_2 \notin \text{supp}(\pi_{s_2})$. And the lender's payoff is higher under the new policy $\left(\hat{\mathcal{S}}, \hat{\sigma}\right)$ because: 1. $\hat{F} = F$ for all $s \in \hat{\mathcal{S}} = \mathcal{S}$; 2. the lending market equilibrium (v_s, c_s) is the same under the two policies for any $s \in \mathcal{S} = \hat{\mathcal{S}}$; 3. the lender's payoff under any signal realization except for s_2 is unchanged; 4. the lender'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 s_2, the equilibrium variables (v_{s_2}, c_{s_2}) is unchanged compared to that with policy $(\mathcal{S}, \tilde{\sigma})$, so the total financing cost is unchanged. But the total payoff generated from the project increases by

$$f(s_2) \cdot K_2 \cdot [\mathbb{E}[\mu v | s_1, B(v_1, \epsilon_1)] - \mathbb{E}[\mu v | s_2, B(v_2, \epsilon_2)]].$$

[20] The other case when $f(s_1) K_1 < f(s_2) K_2$ can be proved using the same proof strategy.

Since $\inf B(v_1, \epsilon_1) > \sup B_2(v_2, \epsilon_2)$, the term $\mathbb{E}[\mu v | s_1, B(v_1, \epsilon_1)] - \mathbb{E}[\mu v | s_2, B(v_2, \epsilon_2)]$ is strictly positive. Since we assume that the policy $(\mathcal{S}, \tilde{\sigma})$ is optimal, then it must be the case that after removing a zero-measure set (denoted as Q) in $\mathcal{S}$, we can't find any s_1 and s_2 such that $\bar{R}_{s_1} > \underline{A}_{s_2}$ (otherwise the lender's payoff can be improved). This implies that, for all $s_1, s_2 \in \mathcal{S} - Q$, we must have $\bar{R}_{s_1} \leq \underline{A}_{s_2}$. Then there must exist a cutoff v^*, such that $\bar{R}_s \leq v^* \leq \underline{A}_s$ for all $s \in \mathcal{S} - Q$. Since Q is a zero-measure set, the statement of the theorem must be true.

A.11 Proof of Proposition 4.1

We establish the following lemma first.

Lemma A.1. *For any optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$, there must exist a deterministic optimal policy $(\mathcal{S}, \sigma)$ with the same signal space $\mathcal{S}$. Besides, let $\left\{\tilde{F}, \tilde{\pi}_s\right\}$ and $\{F, \pi_s\}$ be the distributions of posteriors for policies $(\mathcal{S}, \tilde{\sigma})$ and $(\mathcal{S}, \sigma)$, respectively, and let $(\tilde{v}_s, \tilde{c}_s)$ and (v_s, c_s) be equilibrium outcomes under policies $(\mathcal{S}, \tilde{\sigma})$ and $(\mathcal{S}, \sigma)$, respectively. Then the following properties hold:*

1. *$F = \tilde{F}$ and $(\tilde{v}_s, \tilde{c}_s) = (v_s, c_s)$ for almost all s, and the ex ante lending cutoffs defined in Theorem 4.1 are the same under these two policies, denoted as v^*.*

2. *For all $s \in \mathcal{S}$, both A_s and R_s are non-empty intervals, where A_s and R_s are defined in (16) and (17).*

3. *For all $s_1, s_2 \in S$ with $c_{s_1} < c_{s_2}$, we have*

$$\sup R_{s_1} \leq \inf R_{s_2} \tag{25}$$

and

$$\sup A_{s_1} \leq \inf A_{s_2}. \tag{26}$$

Proof. For the optimal disclosure policy $(\mathcal{S}, \tilde{\sigma})$, if $\mathcal{S}$ is a singleton, this lemma is obviously true. If $\mathcal{S}$ is not a singleton, there must exist two different signal realizations s_1 and s_2 in $(\mathcal{S}, \tilde{\sigma})$, with marginal probabilities $\tilde{f}(s_1)$ and $\tilde{f}(s_2)$, respectively. Based on Lemma 4.1, we just need to consider the case when all of A_{s_1}, A_{s_2}, R_{s_1} and R_{s_2} are non-empty. For simplicity, let's assume that both $\tilde{f}(s_1)$ and $\tilde{f}(s_2)$ represent probability densities and are positive, the proof for other cases (when $\tilde{f}(s_1)$ and $\tilde{f}(s_2)$

are probabilities but not densities) are basically the same. Denote the lending market outcomes as $(\tilde{v}_{s_1}, \tilde{c}_{s_1})$ and $(\tilde{v}_{s_2}, \tilde{c}_{s_2})$ under these two signal realizations. Without loss of generality, let's assume $\tilde{v}_{s_1} < \tilde{v}_{s_2}$. Denote the ex ante lending cutoff as v^* in this case. Suppose for s_1, s_2, the condition

$$\sup A_{s_1} \leq \inf A_{s_2}. \tag{27}$$

is not satisfied, let $B = [\inf A_{s_2}, \sup A_{s_1}]$. Then there must exist two non-negative functions w_1, w_2, such that

$$\tilde{f}(s_1) w_1(v) + \tilde{f}(s_2) w_2(v) = \tilde{f}(s_1) \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v) + \tilde{f}(s_2) \tilde{\pi}_{s_2} \cdot \mathbb{1}_B(v), \tag{28}$$

$$\sup \{\text{supp}(w_1(v)) \cap (v^*, \bar{v}]\} \leq \inf \{\text{supp}(w_2(v)) \cap (v^*, \bar{v}]\}$$

and

$$\int_0^{\bar{v}} w_1(v)\, dv = \int_0^{\bar{v}} \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v)\, dv. \tag{29}$$

Now let's consider the following distribution of posterior beliefs with signal space $\mathcal{S}$, denoted as $\{F, \pi_s\}$, where $F = \tilde{F}$ and

$$\pi_s = \begin{cases} \tilde{\pi}_{s_1} - \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v) + w_1(v) & \text{if } s = s_1 \\ \tilde{\pi}_{s_2} - \tilde{\pi}_{s_2} \mathbb{1}_B(v) + w_2(v) & \text{if } s = s_2 \\ \tilde{\pi}_s & o.w. \end{cases}.$$

We can check that the new distribution of posteriors $\{F, \pi_s\}$ is still Bayes-plausible, because

$$\int_0^{\bar{v}} w_1(v)\, dv = \int_0^{\bar{v}} \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v)\, dv$$

and

$$\int_0^{\bar{v}} w_2(v)\, dv = \int_0^{\bar{v}} \tilde{\pi}_{s_2} \cdot \mathbb{1}_B(v)\, dv.$$

The second condition is a direct result of (28) and (29). Then the distribution of posteriors $\{F, \pi_s\}$ can be induced by a disclosure policy $(\mathcal{S}, \sigma)$. Now the condition (27) is not violated anymore in the new policy. Then we just need to show that the lender'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 $\tilde{\pi}_{s_1}$, the approval probability

is

$$d_{s_1}(1) = \frac{\tilde{c}_{s_1}}{b}.$$

Note that since $\tilde{v}_{s_1} < \tilde{v}_{s_2}$, we know

$$\inf A_{s_2} \geq \tilde{v}_{s_2} > \tilde{v}_{s_1}.$$

Then for any $v \in B$, we must have $v > \tilde{v}_{s_1}$. Then under posterior belief π_{s_1}, the approval probability is still

$$d_{s_1}(1) - \int_0^{\bar{v}} \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v)\, dv + \int_0^{\bar{v}} w_1(v)\, dv = d_{s_1}(1) = \frac{\tilde{c}_{s_1}}{b}.$$

Based on this, we can check that all other equilibrium conditions are also satisfied, and this implies $(v_{s_1}, c_{s_1}) = (\tilde{v}_{s_1}, \tilde{c}_{s_1})$. Similarly, we can check $(v_{s_2}, c_{s_2}) = (\tilde{v}_{s_2}, \tilde{c}_{s_2})$. 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 the lender's payoff is the same under those two policies. We can continue to "modify" the disclosure policy in the above way such that the condition (27) is no longer violated.

The proof strategy still works if condition (29) is not satisfied in the optimal policy $(\mathcal{S}, \sigma)$. Besides, note that the third property in Lemma A.1 implies the second property in Lemma A.1, 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 v^* must be unchanged. □

Now we prove this result by contradiction. Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy, and $(\mathcal{S}, \sigma)$ is the deterministic policy satisfies the properties in Lemma A.1 which generates the same payoff for the lender. Let A_s and R_s be the acceptance region and rejection region for each s under the policy $(\mathcal{S}, \sigma)$. Lemma A.1 implies that both A_s and R_s are non-empty intervals.

Suppose there exists a set $\Omega \subset \mathcal{S}$, such that $\text{Prob}(\Omega) > 0$, and $v_s > v^*$ for all $s \in \Omega$. Our goal is to find another deterministic policy, $(\mathcal{S}_1, \sigma_1)$, such that the lender's payoff is strictly higher under $(\mathcal{S}_1, \sigma_1)$ compared to $(\mathcal{S}, \sigma)$. Consider any signal realization $s_m \in \Omega$. First, Lemma 4.1 implies that $\sup A_s = v_1 > v_{s_m}$. Since Lemma A.1 shows that A_s must be an interval, we consider the following two cases.

If $(v_{s_m}, v_1) \subset A_s$, then the lender can improve her payoff from the signal s_m by disclosing additional information. Specifically, following the proof strategy in

Proposition 3.2, after signal s_m is realized, the lender can strictly improve her payoff by disclosing whether the true state belongs to $(v_{s_m}, v_{s_m}+\epsilon)$ or not, for a sufficiently small but positive ϵ.

If $(v_{s_m}, v_1) \subset A_s$ is not satisfied, there must exist an interval $(v_{s_m}, v_{s_m}+\epsilon_m)$, such that $(v_{s_m}, v_{s_m}+\epsilon_m) \cap \text{supp}(\pi_s) = \emptyset$. We can choose a sufficiently small but positive ϵ_m, such that under the prior belief, $\text{Prob}(v \in (v_{s_m}, v_{s_m}+\epsilon_m)) < \text{Prob}(v \in [v_{s_m}+\epsilon_m, \bar{v}] \cap \text{supp}(\pi_s))$. Then we can find an interval $B \subset [v_{s_m}+\epsilon_m, \bar{v}] \cap \text{supp}(\pi_s)$, and a continuous, one-to-one mapping $z : (v_{s_m}, v_{s_m}+\epsilon_m) \to B$, with $z'(v) = \frac{g(v)}{g(z(v))}$. This implies that $\text{Prob}(v \in (v_{s_m}, v_{s_m}+\epsilon_m)) = \text{Prob}(v \in B)$. Now let's consider the following deterministic disclosure policy $(\mathcal{S}, \sigma')$ with

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

It's easy to check that all lending market equilibria are unchanged for all signal realizations under the new policy $(\mathcal{S}, \sigma')$, and thus the lender's payoff is the same under $(\mathcal{S}, \sigma)$ and $(\mathcal{S}, \sigma')$. However, under the new disclosure policy $(\mathcal{S}, \sigma')$, for the signal realization s_m, we have $(v_{s_m}, v_{s_m}+\epsilon_m) \subset \text{supp}(\pi_{s_m})$. As we discussed earlier, following the proof in Proposition 3.2, the lender can strictly improve her payoff by disclosing whether the true state belongs to $(v_{s_m}, v_{s_m}+\epsilon)$ for not, for a sufficiently small but positive ϵ.

We can apply the above operation to all signal realizations $s \in \Omega$. Since Ω has a positive measure, the improvement is lender's expected payoff is strictly positive, then the policy $(\mathcal{S}, \sigma)$ must be suboptimal. Since $(\mathcal{S}, \tilde{\sigma})$ generates the same payoff as $(\mathcal{S}, \sigma)$, it must also be suboptimal, a contradiction.

A.12 Proof of Proposition 4.2

Let's first introduce the following lemma to establish our results.

Lemma A.2. *For any two posterior beliefs π_{s_1} and π_{s_2}, with positive probabilities (densities) $f(s_1)$ and $f(s_2)$, and lending market equilibira (v_{s_1}, c_{s_1}) and (v_{s_2}, c_{s_2}) sat-*

isfying $v_{s_1} < v_{s_2}$. Let $\hat{s}$ be the "combined"signal with posterior belief

$$\pi\left(v|\hat{s}\right) = \frac{f\left(s_1\right)\pi_{s_1} + f\left(s_2\right)\pi_{s_2}}{f\left(s_1\right) + f\left(s_2\right)},$$

then the lending market equilibrium $\left(v_{\hat{s}}, c_{\hat{s}}\right)$ satisfies

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

and

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

Proof. If $c_{s_1} = 0$, the result is obviously true. When $c_{s_1} > 0$, first, it's impossible to have $v_{\hat{s}} \leq v_{s_1}$. Note that for the equilibria under s_1 and s_2, the equilibrium conditions are $\mu v_{s_1} = \left[\mu + (1-\mu) K\left(c_{s_1}\right)\right] I$ and $\mu v_{s_2} = \left[\mu + (1-\mu) K\left(c_{s_2}\right)\right] I$. For $\hat{s}$, we have $\mu v_{\hat{s}} = \left[\mu + (1-\mu) K\left(c_{\hat{s}}\right)\right] I$. If $v_{\hat{s}} \leq v_{s_1}$, then we must have $c_{\hat{s}} \leq c_{s_1}$. In equilibrium,

$$\text{Prob}\left(v > v_{\hat{s}}|\hat{s}\right) \leq \frac{c_{\hat{s}}}{b} \leq \text{Prob}\left(v \geq v_{\hat{s}}|\hat{s}\right),$$

where

$$\text{Prob}\left(v > v_{\hat{s}}|\hat{s}\right) = \frac{f\left(s_1\right)}{f\left(s_1\right) + f\left(s_2\right)}\text{Prob}\left(v > v_{\hat{s}}|s_1\right) + \frac{f\left(s_2\right)}{f\left(s_1\right) + f\left(s_2\right)}\text{Prob}\left(v > v_{\hat{s}}|s_2\right)$$

and

$$\text{Prob}\left(v \geq v_{\hat{s}}|\hat{s}\right) = \frac{f\left(s_1\right)}{f\left(s_1\right) + f\left(s_2\right)}\text{Prob}\left(v \geq v_{\hat{s}}|s_1\right) + \frac{f\left(s_2\right)}{f\left(s_1\right) + f\left(s_2\right)}\text{Prob}\left(v \geq v_{\hat{s}}|s_2\right).$$

If $v_{\hat{s}} \leq v_{s_1} < v_{s_2}$, we must have $\text{Prob}\left(v > v_{\hat{s}}|s_1\right) \geq \frac{c_{s_1}}{b}$, and $\text{Prob}\left(v > v_{\hat{s}}|s_2\right) \geq \frac{c_{s_2}}{b} > \frac{c_{s_1}}{b}$. Then we must have

$$\text{Prob}\left(v > v_{\hat{s}}|\hat{s}\right) > \frac{f\left(s_1\right)}{f\left(s_1\right) + f\left(s_2\right)}\frac{c_{s_1}}{b} + \frac{f\left(s_2\right)}{f\left(s_1\right) + f\left(s_2\right)}\frac{c_{s_1}}{b},$$

which implies $c_{\hat{s}} > c_{s_1}$, a contradiction! The same proof strategy works for the case $v_{\hat{s}} \geq v_{s_2}$. So the equilibrium must satisfy $v_{s_1} < v_{\hat{s}} < v_{s_2}$. □

Note that the signal in the no disclosure policy is a "combined" signal of all signals in the optimal policy $(\mathcal{S}, \sigma)$, Lemma A.2 implies that there exist sets $A, B \subset \mathcal{S}$ with

positive measures, such that $v_N < v_s$ for all $s \in A$ and $v_N > v_s$ for all $s \in B$. Then we must have $v_N > \inf_{s\in\mathcal{S}} v_s$. Together with Proposition 4.1, we can also conclude that $v^* > v_N$.

A.13 Proof of Lemma 4.2

This is an obvious result of Lemma A.1.

A.14 Proof of Theorem 4.2

The first point has been proved by Lemma 3.2. The second point has been proved by Theorem 4.1, Lemma A.1 and Lemma 4.1.

A.15 Proof of Theorem 4.3

Suppose $(\mathcal{S}, \sigma)$ is an optimal policy that has the structure characterized in Theorem 4.2. It induces the distribution of posteriors $\{F, \pi_s\}$, The probability (density) for each signal s is denoted as $f(s)$. We want to show that with Assumption 2, we can weakly improve it such that it has the structure characterized in Theorem 4.3. Let's choose any $v_3 \in (0, I)$, and $\mathcal{S}_1 = \{\sigma(v) | v \in [0, v_3]\}$. First, we want to show that the lender obtains weakly higher profit by pooling all signals in $\mathcal{S}_1$ together. Under the policy $(\mathcal{S}, \sigma)$, the lender's payoff from signals $s \in \mathcal{S}_1$ is:

$$\begin{aligned}
\tilde{W}_1 &= \int_{s\in\mathcal{S}_1} f(s) \cdot \mathbb{E}\left[(\mu v - (\mu + (1-\mu) K(c_s)) I)^+ | s\right] ds \\
&= \begin{array}{l} \int_{s\in\mathcal{S}_1} f(s) \cdot \mu \mathbb{E}(\alpha_s(1, v)(v - I) | s) ds - \\ \int_{s\in\mathcal{S}_1} f(s) \cdot d_s(1) \cdot (1-\mu) K(c_s) I ds \end{array} \\
&= \int_{s\in\mathcal{S}_1} f(s) \cdot \mu \mathbb{E}\left(\mathbb{1}_{[v^*, \bar{v}]}(v)(v - I) | s\right) ds - \int_{s\in\mathcal{S}_1} f(s) \cdot \frac{c_s}{b} \cdot (1-\mu) K(c_s) I ds.
\end{aligned}$$

Here we use the equilibrium condition $d_s(1) = \frac{c_s}{b}$ in the last equality. Note that $\int_{s\in\mathcal{S}_1} f(s) \cdot \frac{c_s}{b} \cdot ds = \int_{s\in\mathcal{S}_1} f(s) \cdot d_s(1) \cdot ds$. Theorem 4.1 implies that $\int_{s\in\mathcal{S}_1} f(s) \cdot d_s(1) \cdot ds = \text{Prob}(v \geq v^* \& s \in \mathcal{S}_1)$.

Let c_0 be the solution of

$$\left(\int_{s\in\mathcal{S}_1} f(s) \cdot ds\right) \frac{c_0}{b} = \text{Prob}(v \geq v^* \& s \in \mathcal{S}_1) = \int_{s\in\mathcal{S}_1} f(s) \cdot \frac{c_s}{B} \cdot ds,$$

it's clear that c_0 is just the manipulation cutoff in the equilibrium when we pool all

signals in $\mathcal{S}_1$ together. Lemma A.2 implies that $c_0 < \sup_{s\in\mathcal{S}_1} c_s$, and the equilibrium lending cutoff v_0 in this case also satisfies $v_0 < \max_{s\in\mathcal{S}_1} v_s < v^*$. Since function $xK(x)$ is convex on $x \in [0, \bar{c}]$, we must have

$$\left(\int_{s\in\mathcal{S}_1} f(s)\cdot ds\right)(1-\mu)\frac{c_0}{b}K(c_0) \leq \int_{s\in\mathcal{S}_1} f(s)\cdot\frac{c_s}{B}\cdot(1-\mu)K(c_s)\,ds. \tag{30}$$

The lender's payoff from all the pooling signal is:

$$\begin{aligned}\tilde{W}_2 &= \int_{s\in\mathcal{S}_1} f(s)\cdot\mathbb{E}\left[(\mu v-(\mu+(1-\mu)K(c_0))I)^+\right]ds\\ &= \begin{array}{l}\int_{s\in\mathcal{S}_1} f(s)\cdot\mu\mathbb{E}\left(\mathbb{1}_{[v^*,\bar{v}]}(v)(v-I)|s\right)ds-\\ \quad\int_{s\in\mathcal{S}_1} f(s)\cdot d_s(1)\cdot(1-\mu)K(c_0)Ids\end{array}\\ &= \int_{s\in\mathcal{S}_1} f(s)\cdot\mu\mathbb{E}\left(\mathbb{1}_{[v^*,\bar{v}]}(v)(v-I)|s\right)ds-\int_{s\in\mathcal{S}_1} f(s)\cdot\frac{c_0}{b}\cdot(1-\mu)K(c_0)Ids.\end{aligned}$$

Condition (30) implies that $\tilde{W}_2 \geq \tilde{W}_1$, then the policy $(\mathcal{S},\sigma)$ can be weakly improved if we pool all signals in $\mathcal{S}_1$ together. Then we can find an optimal policy that has the following structure: there exists a cutoff $v_4 \in (0, v^*)$, such that $\sigma(v) = \sigma(0)$ for all $v \leq v_4$, and $\sigma(v) > \sigma(v_4)$ for all $v > v_4$. $v_4 \neq v^*$ because we've already showed that the no disclosure policy is suboptimal. Let $s_0 = \sigma(0)$, A_{s_0} and R_{s_0} represent the acceptance region and rejection region under signal s_0.

Then we show that we can weakly improve the optimal policy further, which has the structure characterized in Theorem 4.3. Since we choose the equilibrium cutoff v_s as the message function, we can use $v_{\sigma(v)}$ to represent the equilibrium lending cutoff of the equilibrium whose support contains v. The next step is to show that we must have $\inf\operatorname{supp}\pi_{\sigma(v)} = v_{\sigma(v)}$ for all $v > v_4$. For the sake of contradiction, suppose there exists $v_1 > v_4$, and $\inf\operatorname{supp}\pi_{\sigma(v_1)} < v_{\sigma(v_1)}$. Let $s_1 = \sigma(v_1)$, and let the probability of the signal s_1 be $f(s_1)$. For the rest of the proof, we consider the case when $f(s_1)$ represents a positive probability. For the case when s_1 represents a positive density, the proof is basically the same. Then there must exist a set $E_{11} \subset \operatorname{supp}(\pi_{s_1})$, such that $\sup E_{11} < v_{s_1}$. Let $E_{12} = R_{s_1} - E_{11}$, where R_{s_1} is the rejection region under signal s_1. We can find sets F_{11} and F_{12} satisfying $F_{11}\cap F_{12} = \emptyset$, $F_{11}\cup F_{12} = A_{s_1}$, where A_{s_1} is the acceptance region under equilibrium s_1, and

$$\frac{\text{Prob}(v\in F_{11})}{\text{Prob}(v\in E_{11})+\text{Prob}(v\in F_{11})} = \frac{c_{s_1}}{b}.$$

Similarly, we can find sets E_{01}, E_{02}, F_{01} and F_{02} which are mutually exclusive, and satisfy $E_{01} \cup E_{02} = R_{s_0}$, $Y_{01} \cup Y_{02} = A_{s_0}$, $\sup E_{01} < I\,(< v^*)$, and

$$\frac{\text{Prob}\,(v \in F_{01})}{\text{Prob}\,(v \in E_{01}) + \text{Prob}\,(v \in F_{01})} = \frac{c_{s_0}}{b}.$$

Let $\epsilon = \min\{I - \sup E_{01}, v_{s_1} - \sup E_{11}\}$. Consider the following alternative disclosure policy with signal space $\mathcal{S}_1 = (\mathcal{S} \backslash \{s_0, s_1\}) \cup \{s_{01}, s_{02}\} \cup \{s_{11}, s_{12}\}$ and a message function:

$$\sigma_1(v) = \begin{cases} \sigma(v) & \text{if } v \notin \{\text{supp}\{\pi_{s_1}\} \cup \text{supp}(\pi_{s_0})\} \\ s_{01} & \text{if } v \in E_{01} \cup F_{01} \\ s_{02} & \text{if } v \in E_{02} \cup F_{02} \\ s_{11} & \text{if } v \in E_{11} \cup F_{11} \\ s_{12} & \text{if } v \in E_{12} \cup F_{12} \end{cases}.$$

It's clear that the lender's payoff under $(\mathcal{S}_1, \sigma_1)$ is the same as that under $(\mathcal{S}, \sigma)$. In particular, the lender's lending decision and borrower's manipulation decision are the same under signal s_{01}, s_{02} and s_0 (s_{11}, s_{12} and s_1).

Since $xK(x)$ is convex, for any δ, there must exist positive numbers c_0, c_1, p_0 and p_1, that satisfy $c_0 \in (c_{s_0}, c_{s_0} + \delta), c_1 \in (c_{s_1} - \delta, c_{s_1})$ and $p_1 + p_0 = f(s_{01}) + f(s_{11})$, such that

$$p_1 c_1 + p_0 c_0 = f(s_{01})\, c_{s_0} + f(s_{11})\, c_{s_1},$$

and

$$p_1 c_1 K(c_1) + p_0 c_0 K(c_0) \leq f(s_0)\, c_{s_0} K(c_{s_0}) + f(s_1)\, c_{s_1} K(c_{s_1}). \tag{31}$$

We can choose δ being small enough, such that $v_1 = \frac{\mu + (1-\mu)K(c_1)}{\mu} I > \sup E_{11}$ and $v_0 = \frac{\mu + (1-\mu)K(c_0)}{\mu} I < v^*$.

$xK(x)$ being convex and $v_{s_1} > v_{s_0}$ imply that $p_1 < f(s_{11})$ and $p_0 > f(s_{01})$. Then we must be able to find a set $D \subset Y_{11}$, such that

$$\text{Prob}\,(D) = f(s_{11}) - p_1.$$

Consider the following disclosure policy with signal space $\hat{\mathcal{S}}_1 = (\mathcal{S} \backslash \{s_0, s_1\}) \cup \{\hat{s}_{01}, s_{02}\} \cup$

$\{\hat{s}_{11}, s_{12}\}$ and a message function:

$$\hat{\sigma}_1(v) = \begin{cases} \sigma(v) & \text{if } v \notin \{\text{supp}\{\pi_{s_1}\} \cup \text{supp}(\pi_{s_0})\} \\ \hat{s}_{01} & \text{if } v \in E_{01} \cup F_{01} \cup D \\ s_{02} & \text{if } v \in E_{02} \cup F_{02} \\ \hat{s}_{11} & \text{if } v \in E_{11} \cup F_{11} - D \\ s_{12} & \text{if } v \in E_{12} \cup F_{12} \end{cases}.$$

It's clear the only differences between $(\mathcal{S}_1, \sigma_1)$ and $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$ are the signals s_{01} and s_{11} in $(\mathcal{S}_1, \sigma_1)$ and $\hat{s}_{01}$ and $\hat{s}_{11}$ in $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$. Since $(\mathcal{S}_1, \sigma_1)$ generates the same lender's payoff as $(\mathcal{S}, \sigma)$, to compare the lender's payoff under $(\mathcal{S}, \sigma)$ and $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$, we just need to compute the difference between lender's payoff under $\hat{s}_{01}$ and $\hat{s}_{11}$ in $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$ and lender's payoff under s_{01} and s_{11} in $(\mathcal{S}_1, \sigma_1)$. The lender's payoff under s_{01} in $(\mathcal{S}_1, \sigma_1)$ is

$$\begin{aligned} W_{01} &= \text{Prob}(v \in F_{01}) \mu \mathbb{E}(v - I | v \in F_{01}) - \text{Prob}(v \in F_{01})(1-\mu) K(c_{s_0}) I \\ &= \text{Prob}(v \in F_{01}) \mu \mathbb{E}(v - I | v \in F_{01}) - f(s_{01}) \frac{c_{s_0}}{b}(1-\mu) K(c_{s_0}) I. \end{aligned}$$

The lender's payoff under s_{11} in $(\mathcal{S}_1, \sigma_1)$ is

$$\begin{aligned} W_{11} &= \text{Prob}(v \in F_{11}) \mu \mathbb{E}(v - I | v \in F_{11}) - \text{Prob}(v \in F_{11})(1-\mu) K(c_{s_0}) I \\ &= \text{Prob}(v \in F_{11}) \mu \mathbb{E}(v - I | v \in F_{11}) - f(s_{11}) \frac{c_{s_1}}{b}(1-\mu) K(c_{s_1}) I. \end{aligned}$$

The lender's payoff under $\hat{s}_{01}$ in $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$ is

$$\begin{aligned} \hat{W}_{01} &= \text{Prob}(v \in F_{01} \cup D) \mu \mathbb{E}(v - I | v \in F_{01} \cup D) - \text{Prob}(v \in F_{01} \cup D)(1-\mu) K(c_0) I \\ &= \text{Prob}(v \in F_{01} \cup D) \mu \mathbb{E}(v - I | v \in F_{01} \cup D) - p_0 \frac{c_0}{b}(1-\mu) K(c_0) I. \end{aligned}$$

The lender's payoff under $\hat{s}_{11}$ in $\left(\hat{\mathcal{S}}_1, \hat{\sigma}_1\right)$ is

$$\begin{aligned}\hat{W}_{11} &= \text{Prob}\left(v \in F_{11} - D\right)\mu\mathbb{E}\left(v - I | v \in F_{11} - D\right) - \text{Prob}\left(v \in F_{11} - D\right)(1-\mu)K\left(c_1\right)I \\ &= \text{Prob}\left(v \in F_{11} - D\right)\mu\mathbb{E}\left(v - I | v \in F_{11} - D\right) - p_1\frac{c_1}{b}(1-\mu)K\left(c_1\right)I.\end{aligned}$$

Then the difference is

$$\begin{aligned}&\hat{W}_{01} + \hat{W}_{11} - W_{01} - W_{11} \\ =&(1-\mu)I\left[f\left(s_{01}\right)\frac{c_{s_0}}{b}K\left(c_{s_0}\right) + f\left(s_{11}\right)\frac{c_{s_1}}{b}K\left(c_{s_1}\right) - p_0\frac{c_0}{b}K\left(c_0\right) - p_1\frac{c_1}{b}K\left(c_1\right)\right] \\ \geq&0.\end{aligned}$$

The first equality is because the lender actually approves the project in the same states under these two policies, and this is ganrateeed by condition $v_1 > \sup E_{11}$ and $v_0 < v^*$. And the last inequality is because condition (31).

The above analysis shows that we can weakly improve the lender's payoff if there exists $v_1 > v_4$ such that $\inf \text{supp}\, \pi_{\sigma(v_1)} < v_{\sigma(v_1)}$. We can repeat the modification we've done above until this never happens in the disclosure policy. Then our optimal policy has the following structure: there exists $v_4 \in (0, v^*)$, such that $\sigma\left(v\right)$ is constant when $v \leq v_4$. For all $v > v_4$, $v_{\sigma(v)} = \sigma\left(v\right) = \inf \text{supp}\, \pi_{\sigma(v)}$ which is an increasing function. For the signal $s_0 = \sigma\left(0\right)$, since the lending cutoff satisfies $v_{s_0} \geq v_4$, and we know that $\lim_{v \to v_4^+} \sigma\left(v\right) = v_4$, the message function must be continuous at v_4.

A.16 Proof of Proposition 4.3

We consider a deterministic optimal policy $(\mathcal{S}, \sigma)$ such that function $\sigma\left(v\right)$ has the structure in Theorem 4.3 when $v \leq v^*$. When $v > v^*$, let's consider the following function form of σ on $v \in (v^*, \bar{v}]$:

$$\sigma\left(v\right) = \begin{cases} v_a & v \in (v^*, v_b] \\ \gamma\left(v\right) & v \in (v_b, \bar{v}] \end{cases},$$

where $\gamma\left(v\right)$ is to be solved, and satisfies $\gamma\left(v_b\right) = v_a$, $\gamma\left(\bar{v}\right) = v^*$. For any $v \in (v_b, \bar{v})$, the signal realization is $\sigma\left(v\right) = \gamma\left(v\right)$, and $\sigma\left(v\right) = \gamma_s$ is the equilibrium lending cutoff of the subgame that contains state v. The lender is indifferent between lending or

not when observing $\gamma(v) = \sigma(v)$. Then the lender's equilibrium condition is

$$\mu\gamma(v) = [\mu + (1-\mu) K(c_s)] I, \tag{32}$$

where c_s satisfies

$$\frac{c_s}{b} = \text{Prob}(v > v^* | s) = \frac{g(v)}{g(\gamma(v))\gamma'(v) + g(v)}.$$

Then the equilibrium condition (32) becomes

$$\mu\gamma(v) = \left[\mu + (1-\mu) K\left(\frac{bg(v)}{g(\gamma(v))\gamma'(v) + g(v)}\right)\right] I.$$

This is the differential equation in the proposition. When observing $s = v_a$, the equilibrium condition is $\mu v_a = [\mu + (1-\mu) K(c_{v_a})] I$, where

$$\frac{c_{v_a}}{b} = \frac{G(v_b) - G(v^*)}{G(v_b) - G(v^*) + G(v_a)}. \tag{33}$$

When the signal realization is $s = \lim_{v \to v_a^+} = v_a^+$, the equilibrium condition is $\mu v_a^+ = \left[\mu + (1-\mu) K\left(c_{v_a^+}\right)\right] I$, where

$$\frac{c_{v_a^+}}{b} = \frac{g(v_b)}{g(v_a)\gamma'(v_b) + g(v_b)}. \tag{34}$$

$c_{v_a^+} = c_{v_a}$ because of the continuity, then condition (33) and (34) imply that $\frac{G(v_b)-G(v^*)}{G(v_a)} = \frac{g(v_b)}{g(v_a)\gamma'(v_b)}$.

A.17 Proof of Proposition 5.1

Under the optimal commitment solution, the lender's payoff is

$$\begin{aligned} W_c &= \int_{v_c}^{\bar{v}} [\mu v - \mu I - (1-\mu) K(c_c) I] dG(v) \\ &= \int_{v_c}^{\bar{v}} (\mu v - \mu I) dG(v) - (1-\mu) I K(c_c) [1 - G(v_c)], \end{aligned}$$

where c_c is determined by $b\left(1-G\left(v_c\right)\right)=c_c$. Then

$$W_c=\int_{v_c}^{\bar{v}}\left(\mu v-\mu I\right)dG\left(v\right)-\frac{\left(1-\mu\right)I}{b}K\left(c_c\right)c_c.$$

For any constant $\epsilon>0$ that is small enough, we can find two cutoffs $v_1\left(\epsilon\right)>0$ and $v_2\left(\epsilon\right)>v_c$ that satisfy $\max\left\{v_1,v_2-v_c\right\}<\epsilon$,

$$\frac{1-G\left(v_2\right)}{G\left(v_c\right)-G\left(v_1\right)+1-G\left(v_2\right)}b=c_2>c_c,$$

$$\frac{G\left(v_2\right)-G\left(v_c\right)}{G\left(v_1\right)+G\left(v_2\right)-G\left(v_c\right)}b=c_1<c_c,$$

and $v_1<I<v_c<v_2<\bar{v}$. Let

$$p_1=G\left(v_1\right)+G\left(v_2\right)-G\left(v_c\right),$$

and

$$p_2=G\left(v_c\right)-G\left(v_1\right)+1-G\left(v_2\right),$$

then we can verify that $p_1+p_2=1$, $p_1,p_2>0$ and

$$p_1c_1+p_2c_2=b\left(1-G\left(v_c\right)\right)=c_c.$$

Consider the following policy with a binary signal space $\mathcal{S}=\left\{s_1,s_2\right\}$. s_1 is revealed if the true state is in the set $\left[0,v_1\right]\cup\left[v_c,v_2\right]$, and s_2 is revealed if the true state is in the set $\left(v_1,v_c\right)\cup\left(v_2,\bar{v}\right]$. And the lender commits to lending to the borrower if the true state is above v_c, irrespective of the signal realizations. Then under this policy, it's easy to verify that $\text{Prob}\left(s_1\right)=p_1$ and $\text{Prob}\left(s_2\right)=p_2$. Under signal s_1, in equilibrium, the lender approves the loan if and only if $v\in\left[v_c,v_2\right]$, and under signal s_2, the lender approves the loan if and only if $v\in\left(v_2,\bar{v}\right]$. The lender's expected payoff under this alternative disclosure policy is

$$W_1=\int_{v_c}^{\bar{v}}\left[\mu v-\mu I\right]dG\left(v\right)-\frac{\left(1-\mu\right)I}{b}\left[p_1K\left(c_1\right)c_1+p_2K\left(c_2\right)c_2\right].$$

Since $p_1c_1+p_2c_2=c_c$, and function $xK\left(x\right)$ is strictly concave at c_c, we can choose

$\epsilon > 0$ that is small enough, such that

$$p_1 K(c_1) c_1 + p_2 K(c_2) c_2 < c_c K(c_c),$$

and thus $W_0 < W_1$. So in this case, even the lender commits to an optimal lending cutoff v_c, she can still receive higher payoff by disclosure.

A.18 Proof of Theorem 5.1

Suppose the disclosure policy is $(\mathcal{S}, \tilde{\sigma})$. First, it's obvious that when the verification cost t is sufficiently high, the verification technology will never be used. In our analysis, we already show that in any equilibrium s such that the verification is used, the data manipulation cutoff c_v is uniquely pinned down by $t = \frac{(1-\mu)K(c_v)}{\mu+(1-\mu)K(c_v)} I$. And the lending market equilibrium is also uniquely pinned down. Then there is at most one signal s under which verification is used. Suppose under $s_v \in \mathcal{S}$, there is verification used in equilibrium, and $\text{Prob}(s_v) > 0$. Let v_v be the solution of $\mu v_v = [\mu + (1-\mu) K(c_v)] I$. Then we can weakly improve the lender's payoff if $\text{supp}(\pi_{s_v}) \cap [0, v_v] \neq \emptyset$. To see this, suppose $\text{supp}(\pi_{s_v}) \cap [0, v_v] = B$, and $\text{Prob}(B|s_v) > 0$. It's clear that the lender will never lend to the borrower if $v \in B$ in equilibrium s_v. Then let's consider a new disclosure policy which keeps everything unchanged except disclosing whether the true state $v \in B$ or not if the signal realization is s_v in the previous policy. It's clear that if the true state $v \in B$, the lender's payoff from these states is zero under the old policy, and is non-negative under the new policy, so it weakly improves. The lender's payoff from other states are unchanged, because the lender is always indifferent between verifying borrower type or not under this equilibrium s_v, and thus the lender's payoff will be unchanged from these states. Then the lender's payoff weakly increases under the new policy.

So the lender will reveals whether the state v is above v_v or not, and if $v > v_v$, the lender will verify the borrower type with positive probability. The borrower's equilibrium condition implies that the probability p^* satisfies

$$b(1 - p^*) = c_v,$$

which is $p^* = 1 - \frac{c_v}{b}$. When $v \leq v_v$ is revealed, the lender will not verify the borrower type, and in this case, our Theorem 4.2 shows that the lender can maximize her

profit by disclosing information optimally in the way characterized by the Theorem 4.2, while the only difference being the prior becoming $G^*(v) = \min\left\{\frac{G(v)}{G(v_v)}, 1\right\}$.

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 policy for any $t < t_x$. This result is straightforward. Suppose W_{NV} is the lender's payoff when there is no verification technology available, and $W_V(t)$ is lender's payoff when verification is available and the cost parameter is t. It's obvious 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 the optimal policy.