Model Transparency with Manipulable Input

Jian Sun*

June 10, 2025

Abstract

A lender uses a predictive model to assess borrower quality based on manipulable data. The model is unobservable to the borrower, but the lender can disclose information about it. Both full disclosure and no disclosure lead to excessive data manipulation. Under full disclosure, the borrower can perfectly "game" the model. Under no disclosure, the lender's equilibrium use of data exceeds the ex ante optimal level, which also exacerbates the borrower's incentive to manipulate. We fully characterize the optimal policy, which preserves significant uncertainty in the borrower's posterior by introducing gaps in beliefs. When restricted to monotone policies, the optimal disclosure policy always involves discrete signals, and we provide conditions under which no disclosure becomes optimal. Our results are robust to extensions with commitment to lending decisions and costly verification.

Keywords: Model Transparency, Data Manipulation, Information Design, Lending

JEL Codes: D82, G38, G32, D83

*Lee Kong Chian School of Business, Singapore Management University. Email: jiansun@smu.edu.sg. This paper was previously circulated under the title *"Algorithmic Transparency."* 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 models have been widely used in various decision-making processes, including hiring, college admissions, and lending.[1] In these settings, decision makers use predictive models to link observable personal data to unobserved variables that are relevant to their decision-making. For example, employers score resumes to predict job performance, schools use standardized test results to assess academic potential, and lenders analyze personal data to evaluate creditworthiness. In these cases, the mapping from the input (personal data) to the output (quality) is often not transparent. As a result, economic agents (such as job candidates, students, and borrowers) have limited information about how the data is used and how much it influences the final decision.

A common argument for keeping predictive models opaque is to prevent strategic manipulation of model input.[2] When individuals understand how their data affects outcomes, they may alter observable inputs to influence decisions–a behavior often referred to as "gaming the system" or described by Goodhart's Law: *"When a measure becomes a target, it ceases to be a good measure."* This issue is especially relevant in contexts where data can be easily manipulated at low cost, such as resume formatting, test preparation, or mobile app usage.

Since the effectiveness of a predictive model depends not only on its accuracy but also on the quality of the input data, to improve decision quality, it is essential to design mechanisms that discourage data manipulation. While keeping the model opaque is often used to achieve this, the optimal design of model transparency remains an open and underexplored question.

To better understand this question, this paper studies the optimal disclosure of an internal predictive model that maximizes the model user's payoff in a lending setting. There are two players: a borrower (he) and a lender (she), both of whom are risk-neutral. The borrower has a binary private type, and the lender's gross revenue from financing the borrower is increasing in the borrower's type. Their incentives are misaligned: the lender wants to finance the borrower only if the gross revenue exceeds a fixed cost, while the borrower always prefers to be financed. Although the lender cannot observe the borrower's type directly, the borrower generates personal data with binary realizations (e.g., using iOS or Android, or social media presence), which is informative about his type. Specifically, a better-type borrower generates "better" data, while a worse-type borrower generates "worse" data. However, this personal data is manipulable: the borrower can privately alter the data realization by incurring a manipulation cost, which is privately known to him. The lender only observes the final realization of the data, which may have been manipulated.

The lender possesses an internal predictive model that maps the borrower's type to the gross

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

[2] See discussion in the media article "Gaming the system: Loan applicants are reverse engineering the online lending algorithms" by Zack Miller, published on July 11, 2018.

revenue from financing the borrower. This internal model is unobservable to the borrower. For simplicity, we assume that financing a worse-type borrower always yields zero gross revenue, while financing a better-type borrower yields revenue equal to $v \geq 0$. The value v serves as the parameter of the internal model and is privately known to the lender. Intuitively, the parameter v influences both the worse-type borrower's incentive to manipulate data and the lender's financing decision in equilibrium.

The lender's evaluation of the expected gross revenue from financing a borrower depends on two factors: the value of the model parameter v and the probability that the borrower is of the better type. The first factor, v, is exogenous and privately observed by the lender. The second factor is endogenous and depends on the borrower's equilibrium data manipulation behavior: less manipulation makes the data more informative about the borrower's type. To maximize her expected payoff, the lender can disclose information about the model parameter v to the borrower. Once the borrower updates his belief about v, a subgame equilibrium arises. This equilibrium consists of the borrower's data manipulation strategy and the lender's lending decision. A better-type borrower never manipulates his data, while a worse-type borrower chooses to manipulate if his manipulation cost is below a certain threshold. The lender never finances a borrower with worse data but will lend to one with better data if the model parameter v exceeds an endogenous cutoff. Thus, the equilibrium can be characterized by two cutoffs: a manipulation cost cutoff for the worse-type borrower's decision and a model parameter cutoff for the lender's lending decision. Our central question is: what is the lender's optimal strategy for disclosing model information v to maximize her expected payoff? Before characterizing the optimal disclosure policy, we first build intuition by analyzing two benchmark cases: full transparency and no disclosure.

First, the lender receives zero payoff (the worst possible outcome) if she chooses full disclosure, i.e., disclosing the exact value of model parameter 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. As a result, the lender's equilibrium payoff is zero. This implies that full disclosure induces excessive "gaming the system," and thus reduces model effectiveness. Second, if the lender chooses no disclosure, i.e., not disclosing any information about the model parameter v to the borrower, she can achieve a strictly positive payoff. Hence, no disclosure is better than full disclosure. This finding 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 why, consider the case when the model parameter is just slightly above the equilibrium cutoff in no disclosure, 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 marginal increase in financing probability incentivizes the worse-type borrower to manipulate his data, thereby reducing its informativeness. The resulting payoff loss from decreased informativeness outweighs the marginal benefit of lending. This suggests that borrower data is overused in the no disclosure equilibrium. The reason no disclosure is suboptimal (and the key friction in this setting) is the lender's lack of commitment, and similar frictions arise in other economic contexts (see Ball (2025)). Specifically, the lender cannot commit ex ante to how she will use the borrower's data and therefore ends up using it excessively ex post. This overuse of data, in turn, encourages excessive manipulation by the borrower.

We then characterize the structure of the optimal disclosure policy, which features partial disclosure. The optimal policy divides the support of the prior belief about the model parameter v into disjoint groups and send a signal to the borrower to reveal the group to which the true value of v belongs. Each signal s induces a different posterior belief for the borrower. Under each signal realization s, in equilibrium, there exists a signal-dependent cutoff $\mathbf{v}_s$ such that the lender lends to the borrower if and only if the borrower has better data and the true model parameter v exceeds $\mathbf{v}_s$. The lender rejects the borrower if he has worse data or if $v < \mathbf{v}_s$.

Across all signal realizations, the cutoffs $\mathbf{v}_s$ are strictly less than, but converge to a signal-independent threshold $\mathbf{v}^*$. Moreover, under any signal realization s, the borrower's posterior belief must exclude the interval $(\mathbf{v}_s, \mathbf{v}^*)$. In other words, the borrower's posterior belief always contains two regions of v: one below the cutoff $\mathbf{v}_s$ and one above $\mathbf{v}^*$, with a positive gap between these two regions. As a result, $\mathbf{v}^*$ serves as the ex ante lending cutoff: regardless of the signal realization, the lender lends to a borrower with better data if and only if $v > \mathbf{v}^*$, and rejects the borrower otherwise. So although different signals induce different posterior beliefs and subgame equilibria, the effective lending cutoff remains fixed at $\mathbf{v}^*$.

Borrower data is used with lower probability under the optimal policy compared to both full disclosure and no disclosure. The intuition is as follows. Under full disclosure, the probability that the lending decision conditions on borrower data is the highest. This is because the borrower's incentive to manipulate data increases with the model parameter v. Although the lender is less likely to use the data when v is low, lower manipulation in that region keeps the data sufficiently informative, leading the lender to rely on it even when v is not very high. In the case of no disclosure, this issue is partially mitigated. Since the borrower's manipulation decision is independent of the true model parameter, the data is less informative when v is low, making the lender less likely to condition her approval decision on it in that region. This reduces the probability of data use overall. The optimal policy further reduces data usage by inducing variation in the borrower's manipulation incentives across different signals. Under signals where v tends to be low, manipulation is also low but calibrated just enough to discourage the lender from using the data. Under signals with higher v, the manipulation level is higher, again discouraging data use. As a result, the optimal policy lowers the unconditional probability of borrower data being used more effectively than either full or no disclosure.

We 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 consists of a discrete component, which induces an equilibrium with the lowest manipulation level among all posterior equilibria, and a continuous component, where manipulation levels vary continuously. This simplifies the optimal disclosure problem to a one-dimensional optimization problem.

A key feature of the optimal policy is its nonmonotonicity. Although the approval decision is monotonic in the model parameter v, meaning a borrower with better data is approved if and only if the true model parameter exceeds a cutoff, other equilibrium outcomes such as the intensity of data manipulation are not monotonic in the model parameter v. As an extension, we study the lender's optimal disclosure policy under a monotonicity constraint, where the lender is restricted to choosing monotone disclosure policies. Under such policies, the borrower's posterior belief about the model parameter is always a connected interval. This extension delivers two key results that align with real-world observations. First, unlike in the general case, the optimal monotone disclosure policy always consists of discrete signals, which are easy to implement in practice, for example through rating systems. Second, no disclosure can be optimal among monotone policies if the distribution of the model parameter satisfies a simple monotonicity condition. This helps explain why no disclosure remains common among model users in practice.

Finally, we consider two extensions that explore how optimal model disclosure interacts with other mechanisms for deterring manipulation. The first extension examines the case in which the lender can also commit to her lending decisions. Here, the lender first commits to an approval rule and then chooses what information about the model to disclose. We show that, under certain conditions, model disclosure can still improve the lender's payoff even after such commitment. This highlights the robustness of information disclosure as a tool for discouraging manipulation.

The second extension incorporates costly verification of borrower type. In practice, lenders may verify borrower types through manual reviews, interviews, or fraud detection tools. In this setting, the lender can pay a cost to verify the true type of the borrower. We study how costly verification interacts with model disclosure under the optimal joint design. We find that the two instruments act as substitutes. The lender first reveals whether the model parameter exceeds a threshold. If it does, she randomly verifies the borrower's type and does not disclose further information. If not, verification is not used and the lender discloses additional information about the model parameter, similar to the baseline policy. This result confirms that disclosure remains an important and effective tool for limiting 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 research is closely related to the literature on strategic data manipulation (e.g., Rajan and Xiong (2024), Frankel and Kartik (2019a), Frankel and Kartik (2019b), Ball (2025), Perez-Richet and Skreta (2022), Gamba and Hennessy (2024)). Our modeling approach to private information on the borrower side follows Fischer and Verrecchia (2000) and Frankel and Kartik (2019b). Ball (2025) studies a setting with multi-dimensional features and heterogeneous agents, and characterize the optimal scoring rule that improves inference. Perez-Richet and Skreta (2022) examine how test design affects agents' manipulation incentives. In these models, the decision maker faces a problem similar to ours: she uses data that can be strategically manipulated by agents. Rajan and Xiong (2024) also study a lending setting in which a lender uses alternative data to make lending decisions, while the borrower can strategically manipulate the data. Their analysis focuses on how much data the lender should commit to using in the lending decision. Greater data coverage improves screening accuracy but also increases the borrower's incentives to manipulate their data. A key insight from this literature is that it can be optimal for the decision maker to commit to underutilizing the data in order to reduce manipulation and improve the quality of data input. The key difference between our paper and this literature is that, in their settings, the decision maker can commit to a rule for *using* the data, while the information structure-such as the correlation between manipulable data and underlying quality-is assumed to be public. In contrast, our paper studies a setting where the decision maker *cannot* commit to how she uses the data, but instead *can* choose how much information about the model's structure to disclose. Our Section 5.1 further explores how these two mechanisms (commitment to action and control of information) interact.

Second, our research contributes to the literature on applications of Bayesian persuasion (Kamenica (2019) and Bergemann and Morris (2019) provide excellent surveys) in economics and finance, including shareholder voting (Malenko et al. (2021)), banking and financial network (Huang (2020)), security design (Azarmsa and Cong (2020); Szydlowski (2021); Inostroza and Tsoy (2022)), and bank stress test (Goldstein and Leitner (2018), Inostroza (2019), Inostroza and Pavan (2021), Leitner and Williams (2020)). Our paper considers a different setting. More importantly, the key friction that disclosure affects the decision maker's inference problem does not arise in these papers. On the theoretical side, we follow Kamenica and Gentzkow (2011) and consider 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.

Lastly, this paper relates to the broader discussion on algorithmic transparency. 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 main contribution is to consider a flexible model transparency problem and to provide a rational explanation for why model transparency can be optimal even in the absence of fairness or behavioral concerns. 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 models, and more importantly, we consider flexible disclosure policies and gain deeper insights into the design of optimal 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, and discuss the design to mitigate this issue. 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.

2 A Simple Model

To fix ideas, let's first consider a simple setting. 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 the bargaining power and collects all the revenue from the project if she finances the borrower, while the borrower receives a private benefit $b = 1$ if his project is successfully financed, regardless 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 private information. A good (bad) borrower chooses high (low) phone usage naturally,[3] but a bad borrower can choose high phone usage by incurring a private cost c, which is uniformly distributed over $[0, 1]$. We refer to the behavior of a borrower who does not choose his

[3]For example, a good-type borrower may have stronger communication and networking skills, higher information intake and market awareness, or greater self-motivation.

natural phone usage level as manipulation.[4] 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 observe is the the borrower's phone usage (after potential manipulation).

Regarding the probability of success, it is common knowledge that a bad borrower always fails, and a good borrower's probability of success v is drawn from a uniform distribution $U[0,1]$. The lender has an internal predictive model that perfectly reveals the realization of v, which is unknown to the borrower. The parameter v can be interpreted as either a model parameter or model-specific information in the lender's internal model. The lender can commit to a disclosure policy that determines how much information about v is revealed to the borrower before the borrower chooses his phone usage level and the lender chooses her lending decision. In this section, we examine the equilibria under three types of disclosure policies.

No Disclosure

Suppose the lender does not reveal any information about v, it can be shown that there is a unique equilibrium which is characterized by two cutoffs $\mathbf{c}_N$ and $\mathbf{v}_N$. A bad borrower chooses high phone usage level if his manipulation cost is lower than $\mathbf{c}_N$. The lender always rejects a borrower with low phone usage and lends to a borrower with high phone usage if and only if $v > \mathbf{v}_N$.

There are two equilibrium conditions. First, a bad type borrower with manipulation cost $\mathbf{c}_N$ is indifferent between manipulating or not, i.e.,

$$\text{Prob}\left(v > \mathbf{v}_N\right) \cdot b = \mathbf{c}_N,$$

where $\text{Prob}\left(v > \mathbf{v}_N\right) = 1 - \mathbf{v}_N$ is the probability of a borrower with high phone usage being financed. Second, the lender is indifferent between approving a borrower with high phone usage and rejecting him when the true value of v is $\mathbf{v}_N$, i.e.,

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

With our assumptions, the unique equilibrium values are $\mathbf{v}_N = \frac{\mu I + (1-\mu) I b}{\mu V + (1-\mu) I b}$ and $\mathbf{c}_N = b \cdot \frac{\mu (V - I)}{\mu V + (1-\mu) I b}$. The lender's payoff is equal to the gross revenue from financing the borrower minus the fixed financing cost I, and can be written as:

$$W_N = \mu V \int_{\mathbf{v}_N}^{1} \left(v - \mathbf{v}_N\right) dv.$$

We can show that in this equilibrium, $\mathbf{v}_N = 0.59$, $\mathbf{c}_N = 0.41$ and $W_N = 0.25$.

[4]For simplicity of exposition, we only allow a bad borrower to manipulate his phone usage. However, the results remain unchanged if we also allow a good borrower to manipulate his usage to low, since a good borrower would never do so in equilibrium.

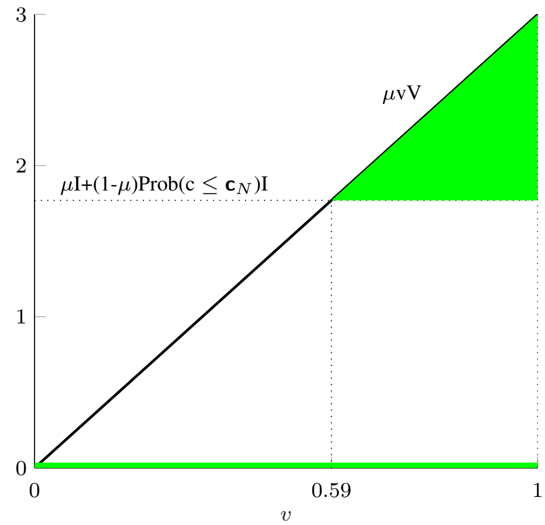

Figure 1: Lender's Payoff: No Disclosure

Figure 1 characterizes the above equilibrium. The green triangle in Figure 1 represents the payoff W_N. In equilibrium, a borrower with high phone usage is financed when the true value of v is above 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 payoff equals zero, leading to the worst outcome. To see this, when $v < \frac{I}{V}$, lending is always inefficient, leading to no financing and zero payoff. For any $v \geq \frac{I}{V}$, under our parameterization, the lender must be indifferent between lending to a borrower with high phone usage and not lending at all. As a result, the lender's payoff must also be zero for any $v \geq \frac{I}{V}$.

A Binary Signal

Our question is, can the lender achieve a strictly higher payoff by disclosing some model information 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 region A or $A^{\complement}$*. There are two possible equilibria depending on which region the realization v belongs to. If $v \in A$ is revealed, 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 lending cutoff and manipulation cutoff, defined in the same way as in the no disclosure case, are $\mathbf{v}_1 = 0.54$ and $\mathbf{c}_1 = 0.34$, respectively.

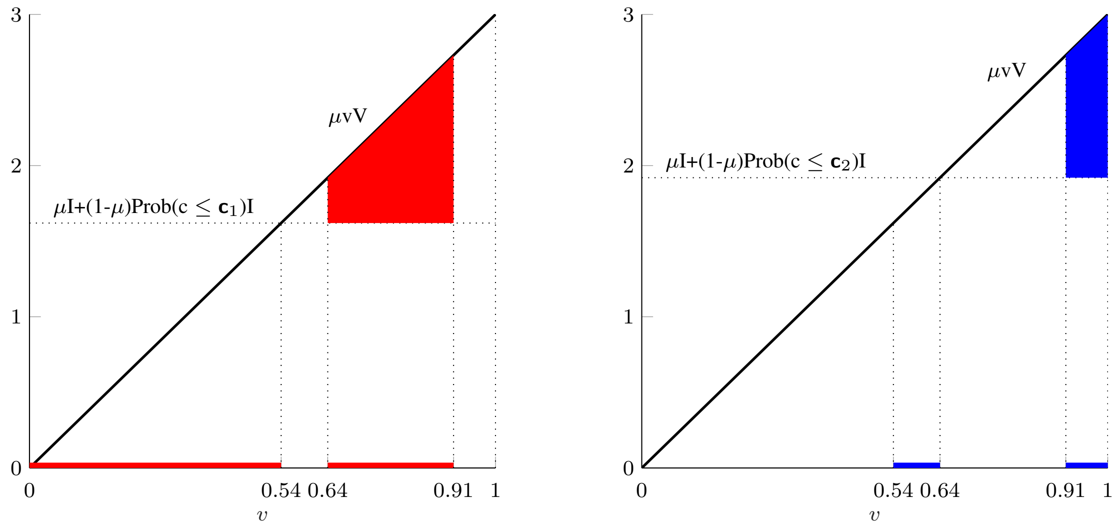

Figure 2: Lender's Payoff: binary signal

This means the lender approves a borrower with high phone usage when $v > \mathbf{v}_1 = 0.54$, and a bad borrower chooses high phone usage if his manipulation cost is lower than $\mathbf{c}_1 = 0.34$. We can show that the lender's payoff in this case is $W_1 = 0.19$. 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 $\mathbf{v}_2 = 0.64$, $\mathbf{c}_2 = 0.48$ and $W_2 = 0.09$.

Figure 2 characterizes these two equilibria. The red trapezoid in the left panel represents the lender's payoff in the equilibrium when revealing $v \in A$, 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 lender's payoff on observing $v \in A^{\complement}$. The total payoff with this disclosure policy is

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

So the lender's expected payoff indeed improves.

Our analysis shows that the binary signal dominates both the no disclosure policy and full transparency policy. The result that the full transparency policy is dominated is clear: when the exact model information v is disclosed to the borrower, the level of borrower's manipulation of phone usage makes the lender always indifferent between approving and rejecting the borrower for every value of v. Consequently, incorporating borrower's phone usage data into lending decisions provides zero value to the lender. The inefficiency in the no disclosure equilibrium arises from the lender's lack of commitment: ex post, the lender conditions the lending decision on phone usage data as efficiently as possible, which leads to excessive data manipulation from an ex ante perspective. To see this, suppose that the lender commits to lending to the borrower with high phone usage if $v > x$. We can calculate the lender's equilibrium profit as a function of x, denoted by $W(x)$:[5]

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

Since $\mathbf{v}_N$ satisfies $\mu \mathbf{v}_N V - (\mu + (1-\mu)(1-\mathbf{v}_N)) I = 0$, we can show that

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

The above result shows that the equilibrium cutoff $\mathbf{v}_N$ is inefficiently low from an ex ante perspective. Note that when $v < \mathbf{v}_N$, the phone usage data is effectively not used in the lender's lending decision, because the borrower will be rejected no matter what his data is. The phone usage data will be used only when $v \geq \mathbf{v}_N$, at which point a borrower with high and low phone usage will receive different loan approval decisions. So the probability that phone usage data is used in lending decisions, $(1 - \mathbf{v}_N)$, is inefficiently high in this no disclosure equilibrium.

To mitigate the excess manipulation, the binary signal makes the approval decision less dependent on phone usage data by differentiating the posterior distribution of model information v. *Unconditionally*, with the binary signal, borrower's phone usage data affects approval decision when $v > 0.64$, while it is $v > 0.59$ in the no disclosure equilibrium. So the lender's decision depends less on the phone usage data under the binary signal. Intuitively, by differentiating the posterior distribution of model information v, the equilibrium with more data manipulation ($v \in A^{\complement}$) ensures that borrower data does not affect the lender's decision in cases where it would have under the no disclosure equilibrium (when $v \in (0.59, 0.64)$), and the other equilibrium ($v \in A$) has lower level of data manipulation (compared to the no disclosure equilibrium) and generates more efficient outcome.

In the main model, I'll consider a general setting with more flexible 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 ($\mathbf{v}^* = 0.64$), such that *unconditionally*, lender's decision depends on phone usage data if and only if the true model parameter v 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, unconditionally, the probability that lender's decision depends on phone usage data is lower than that in the no disclosure equilibrium. Last, the binary signal induces two posterior equilibria, in both posterior equilibria, the lender's indifference cutoffs $\mathbf{v}_1$ and $\mathbf{v}_2$ are lower than the ex ante cutoff $\mathbf{v}^*$. Since the lender unconditionally rejects the borrower whenever $v < \mathbf{v}^*$, this implies that there must be gaps in the posterior distribution of the model parameter v. All of these properties still hold in the

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

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: a deep-pocketed lender (she) and a representative borrower (he), both of whom are risk neutral. The borrower's type $X \in \{0, 1\}$ is private information. The probability that the borrower is of type $X = 1$ is denoted by μ, which is common knowledge. The borrower has zero initial wealth and is protected by limited liability. He has a financing need and must borrow $I > 0$ from the lender. If the lender rejects the borrower, the borrower receives zero payoff; if the lender approves him, the borrower receives a constant private benefit of $b > 0$. The lender has an internal predictive model that evaluates the gross revenue from financing a borrower with type X, denoted by $V(X)$. For simplicity, we assume that

$$V = X \cdot v. \tag{1}$$

The lender's payoff is $V - I$ if she finances the borrower, and zero if she rejects the borrower.

The key feature of our model is that the model parameter 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.

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 $\{0, 1\}$. 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 privately 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 drawn from a continuous distribution on $(0, \bar{c}]$ with a cumulative distribution function $K(\cdot)$, and its realization is privately observed by the borrower. For simplicity, assume $K'(c)$ exists everywhere and is continuous and strictly positive at any $c \in (0, \bar{c}]$. We also assume that c is independent of all other random variables in the model.[6]

[6]It is natural to consider correlation between borrower type and manipulation cost, which can be modeled as type-specific cost distributions. This extension does not affect our results. With only two types and two data realizations in

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

Assumption 1 implies that, in any equilibrium, the lender will never approve a borrower with data $x = 1$ with probability one. Otherwise, the expected cost of financing, $[\mu + (1-\mu) K(b)] I$, is higher than the maximum gross revenue from lending, $\mu\bar{v}$, and thus lending is unprofitable for the lender. Therefore, this can not be an equilibrium under Assumption 1.

Disclosure About the Lender's Internal Model

To maximize her expected payoff, the lender can reveal some information about the model parameter v in (1) to the borrower. Specifically, the lender commits to a policy that reveals information about v before observing its true value, and the borrower makes data manipulation decision after observing the revealed information about v. To make the disclosure policy as flexible as possible, we adopt the standard definition used in the 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. 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 using pure information disclosure to mitigate 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

our model, the type $X = 1$ borrower never manipulates and thus his cost distribution is irrelevant.

function, see below for the formal definition. For simplicity, let δ_x be the Dirac measure, which puts probability one 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 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.[7]

Timeline

All events occur in the following order:

1. The lender chooses a signal structure $(\mathcal{S}, \tilde{\sigma})$; then v is 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 based on the observed data x and the true value of the internal model parameter v.

5. All random variables are realized, and both players receive their payoffs.

3.2 Manipulation and Lending Equilibrium

We first investigate the equilibrium after a signal s is revealed under a disclosure policy $(S, \tilde{\sigma})$. We refer to this as 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 internal model parameter v, and the signal s.

[7]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 $\sigma(v) \equiv 0$.

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

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

and zero if she rejects the borrower. Similarly, her expected payoff from lending to a borrower with data $x = 0$ is

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

and zero if she rejects the borrower. Therefore, 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 a borrower with data x can be financed is

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

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

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

If he does not manipulate the data, his data x remains to be X, and his payoff 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 now formally define the equilibrium of subgame s.

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).

The equilibrium has a simple structure. First, we show in the appendix that in equilibrium, $\gamma_s(1, c) = 0$ for all c and s. This is because a borrower with data $x = 1$ is always perceived as higher quality by the lender and thus 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 $d_s(0) \equiv 0$ in equilibrium. Therefore, we only need to focus on the borrower's strategy when his type is $X = 0$, and the lender's strategy when she observes 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.,

$$c \leq b \cdot d_s(1) = \mathbf{c}_s.$$

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

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

When $v = \mathbf{v}_s$, the lender is indifferent between lending and not lending. Then the equilibrium can be summarized by these two cutoffs $\mathbf{c}_s$ and $\mathbf{v}_s$. The following lemma characterizes the equilibrium in subgame s.

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

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

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

3. *$\mathbf{c}_s$ and $\mathbf{v}_s$ are solved by the following equilibrium conditions*

$$\mathbf{c}_s = b \cdot d_s(1), \tag{8}$$

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

and

$$\mathbf{v}_s = \frac{[\mu + (1-\mu) K(\mathbf{c}_s)] \cdot I}{\mu}, \tag{10}$$

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

The lender's expected payoff in the subgame s is

$$W_s = \mu \mathbb{E}_s \left[(v - \mathbf{v}_s)^+ \right] = \mu \int_{\mathbf{v}_s}^{\bar{v}} (v - \mathbf{v}_s) \, \pi_s (v) \, dv. \tag{11}$$

For the rest of the paper, we use bold, sans-serif letters $(\mathbf{v}_s, \mathbf{c}_s)$ 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 $\mathbf{c}_s = 0$ and $\mathbf{v}_s = \bar{v}$ 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 payoff W_s characterized in Lemma 3.1. Then the lender's expected payoff 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, $\{F(s), \pi_s\}$,[8] 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:*

$$\max_{\mathcal{S}, \{F, \pi_s\}} W = \mu \int_{s \in \mathcal{S}} \mathbb{E} \left[(v - \mathbf{v}_s)^+ | s \right] dF(s) \tag{12}$$

$$\textit{s.t} \int_{s \in \mathcal{S}} \pi_s dF(s) = g(v), \tag{13}$$

$$\textit{Prob} \left(v > \mathbf{v}_s | \, s \right) \leq \frac{K^{-1} \left(\frac{\mu \mathbf{v}_s}{I(1-\mu)} - \frac{\mu}{1-\mu} \right)}{b} \leq \textit{Prob} \left(v \geq \mathbf{v}_s | \, s \right) \quad \forall s \in \mathcal{S}. \tag{14}$$

Here $\textit{Prob}\,(v| \, s)$ is the probability function associated with the posterior belief π_s.

The last condition (14) solves the equilibrium cutoff $\mathbf{v}_s$ under the posterior belief π_s. This problem is known as a Bayesian persuasion problem 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 structure of the lender's solution, we present the following observation that simplifies the problem.

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

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

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 $s_1, s_2 \in \mathcal{S}$, $\mathbf{v}_{s_1} \neq \mathbf{v}_{s_2}$. Then we can choose a signal space such that $s \equiv \mathbf{v}_s$ for all $s \in \mathcal{S}$. We adopt this simplification for the rest of the paper.

3.3 Suboptimality of No Disclosure and Full Transparency

The borrower's data manipulation distorts the lender's inference about the borrower's type. A borrower with type $X = 0$ chooses his manipulation strategy based on the updated belief about v. For the optimal policy, a natural guess would be that the lender should not disclose any information, and thus make the internal model as opaque as possible. Denote the equilibrium in this case as $(\mathbf{v}_N, \mathbf{c}_N)$, where N represents "no disclosure." The lender's payoff is

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

We show that, in this case, the lender conditions her decision on borrower data x too frequently, that is, she conditions her decision on x at certain values of model parameter v where it would be (ex ante) optimal not to, and this leads to excessive data manipulation by the borrower. This friction 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 $x = 1$ whenever the lending yields a positive payoff, i.e., when $v > \mathbf{v}_N$, where $\mathbf{v}_N$ satisfies

$$\mu \mathbf{v}_N - [\mu + (1-\mu) K(\mathbf{c}_N)] \cdot I = 0. \tag{15}$$

Now, hypothetically, suppose the lender can commit to a slightly higher lending cutoff $\mathbf{v}_N + \delta$ $(0 < \delta \ll 1)$, i.e., the lender lends to a borrower with data $x = 1$ only when $v \geq \mathbf{v}_N + \delta$. Let her payoff in this case be W_δ. This leads us to the following result.

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

Lemma 3.3 implies that, in the no disclosure equilibrium, a commitment to reduced lending improves the lender's payoff.[9] To see the intuition, rewriting $\left.\frac{dW_\delta}{d\delta}\right|_{\delta=0}$:

$$\left.\frac{dW_\delta}{d\delta}\right|_{\delta=0} = -\underbrace{\left[\mu\mathbf{v}_N - \left(\mu + (1-\mu)\, K\left(\mathbf{c}_N\right)\right) I\right] g\left(v\right)}_{\text{= 0 in equilibrium}} + \int_{\mathbf{v}_N+\delta}^{\bar{v}} \left(-\left(1-\mu\right) I g\left(v\right) \left.\frac{dK\left(\mathbf{c}_\delta\right)}{d\delta}\right|_{\delta=0}\right) dv. \tag{16}$$

The first term in (16) is zero, which follows directly from the lender's equilibrium condition (15). The second term is positive because an increase in δ lowers the equilibrium cutoff for manipulation cost, denoted by $\mathbf{c}_\delta$, thereby increasing the lender's payoff. This result highlights the lender's commitment problem. In the no disclosure equilibrium, the lender always approves a borrower with $x = 1$ whenever the payoff is positive. However, such approval creates incentives for ex ante manipulation by the borrower, which ultimately undermines the lender's inference in equilibrium.

In our setting, the lender can not commit to lending decisions.[10] Instead, only disclosure of the model information v is considered. We show that, by strategically disclosing information about v, the lender can reduce lending and improve her payoff. 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 $W_1 > W_N$.*

Proposition 3.2 challenges the conventional wisdom that making internal models more transparent will always hurt efficiency because of the "gaming the system" concern. 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

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

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 model parameter v 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., $(\mathbf{v}_1, \mathbf{c}_1) = (\mathbf{v}_N, \mathbf{c}_N)$; second, $A_2 = [\mathbf{v}_N, \mathbf{v}_N + \delta]$ is a small set, where $\delta \ll 1$; finally, the remaining regions is A_3, i.e., $A_3 = [0, \bar{v}] - A_1 \cup A_2$.

[9]Note that this commitment to reduced lending corresponds to less frequent use of borrower data in lending decisions. In our model, the lender rejects the borrower regardless of his data when v falls below a certain cutoff, meaning the data is effectively not used in those cases. In contrast, when v exceeds the cutoff, the lender approves the loan only if the borrower has $x = 1$. Thus, the lending decision depends on borrower data only when v is above the cutoff. As a result, a higher cutoff implies a lower probability that the lending decision is conditioned on borrower data, which effectively reduces the lender's usage of borrower data. 000

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

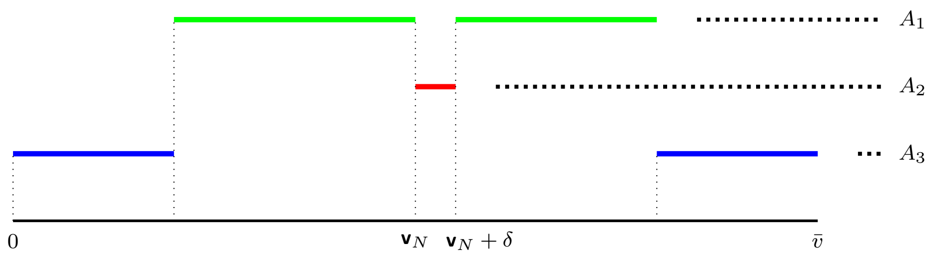

Figure 3: Suboptimality of No Disclosure Policy

We now compare the lender's payoff under the new disclosure policy and under no disclosure, for any true model parameter v. For $v \in A_1$, the lender's payoff remains unchanged, as the subgame equilibrium s_1 is identical to that of the no disclosure case. For $v \in A_2$, the probability measure of A_2 is of order $O(\delta)$. Moreover, in this case, the lender's payoff is strictly less than $\mu\delta$ under no disclosure and is nonnegative under the new policy. This implies that the difference of payoffs from this region under these two disclosure policies is of order $o(\delta)$. For $v \in A_3$, the new policy reduces the probability of financing a borrower with $x = 1$ relative to the no disclosure case, since financing occurs only when v lies in the right subinterval of A_3. This reduces data manipulation, improves the lender's inference, and increases her payoff, which is of order $O(\delta)$. When δ is sufficiently small, compared to the no disclosure policy, the magnitude of the payoff loss from A_2 is one order less than the payoff gain from A_3. Hence, the lender obtains a strictly higher payoff under the new disclosure policy.

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 model parameter v fully transparent, W_F, is zero.*

Since the lender's payoff must be nonnegative, Lemma 3.4 implies that making the model parameter v fully transparent leads to the worst payoff. The intuition is straightforward: when the borrower knows the exact value of v, in equilibrium, the lender must be indifferent between lending to a borrower with $x = 1$ and rejecting. This implies that the lender's payoff from financing a borrower with data $x = 1$ must be zero for all values of v. This result is consistent with the common view that disclosing too much information about an internal model will hurt the model user 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 any subgame equilibrium. Lemma 3.1 implies that for any two signal realizations s_1 and s_2, we must have

$$\mathsf{c}_{s_1} \lesseqgtr \mathsf{c}_{s_2} \Longleftrightarrow \mathsf{v}_{s_1} \lesseqgtr \mathsf{v}_{s_2}.$$

When the manipulation cost cutoff c_s is higher, the borrower is more likely to manipulate his data, making it harder for the lender to infer the borrower's true type when observing $x = 1$. Consider an arbitrary signal s with posterior belief π_s. Let

$$A_s = \{v \in \operatorname{supp}(\pi_s) \,|\alpha_s(1, v) > 0\}, \tag{17}$$

denote the subset of the support of $v|s$ in which the lender lends to a borrower with data $x = 1$ with positive probability, and let

$$R_s = \{v \in \operatorname{supp}(\pi_s) \,|\alpha_s(1, v) = 0\} \tag{18}$$

denote the subset in which the lender lends with zero probability. We refer to A_s and R_s as the *acceptance region* and *rejection region*, respectively. Clearly, A_s and R_s are disjoint, and together they cover the support of the posterior belief, i.e., $A_s \cup R_s = \operatorname{supp}(\pi_s)$. The following lemma shows that, under any optimal policy, both A_s and R_s must be non-empty, and the lender approves the borrower with data $x = 1$ with probability one for all $v \in A_s$.

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 $A_s = \{v \in \operatorname{supp}(\pi_s) \,|\alpha_s(1, v) = 1\}$.*

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

$$\sup R_s \le \mathsf{v}_s \le \inf A_s,$$

showing that the acceptance region A_s and rejection region R_s are separated by the signal-contingent equilibrium cutoff v_s for any signal s. This result follows directly from the lender's ex post optimality condition: it is always optimal to approve a borrower with data $x = 1$ when the true model parameter v 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 $\mathbf{v}^*$, such that for almost all $s \in \mathcal{S}$, we have

$$\sup R_s \leq \mathbf{v}^* \leq \inf A_s,$$

where A_s and R_s are the acceptance and rejection regions defined above.

Theorem 4.1 provides a necessary condition that all optimal policies must satisfy. Intuitively, the lender approves a borrower with data $x = 1$ if and only if v is sufficiently high. While this condition has been shown to hold under any subgame in previous analyses, Theorem 4.1 shows that it must also hold unconditionally across all signals under the optimal policy, i.e., for any signal realization s, the lender approves a borrower with $x = 1$ if and only if $v > \mathbf{v}^*$. For clarity, we refer to $\mathbf{v}^*$ as the *unconditional lending cutoff*. In contrast, we refer to $\mathbf{v}_s$ as the *ex post lending cutoff*, as it is the equilibrium lending cutoff conditional on signal s. Later, we will show that the unconditional lending cutoff $\mathbf{v}^*$ plays a central role in characterizing the optimal policy. To gain further insight, the following proposition establishes a connection between $\mathbf{v}^*$ and the ex post equilibrium cutoff $\mathbf{v}_s$ across all signal realizations.

Proposition 4.1. *Suppose $(\mathcal{S}, \tilde{\sigma})$ is an optimal policy. Then, up to a set of signals with measure zero, we have $\mathbf{v}_s < \mathbf{v}^*$, and*

$$\mathbf{v}^* = \sup_{s \in \mathcal{S}} \mathbf{v}_s.$$

Proposition 4.1 shows that, for almost all signal realizations, the ex post lending cutoff $\mathbf{v}_s$ satisfies $\mathbf{v}_s < \mathbf{v}^*$. In the subgame following any such signal s, the lender approves a borrower with $x = 1$ if and only if $v > \mathbf{v}_s$. However, Theorem 4.1 shows that under signal s in the optimal policy, a borrower with $x = 1$ is approved if and only if $v > \mathbf{v}^*$. These results together imply that the posterior belief must satisfy

$$\operatorname{supp}(\pi_s) \cap (\mathbf{v}_s, \mathbf{v}^*) = \emptyset,$$

so the interval $(\mathbf{v}_s, \mathbf{v}^*)$ has zero probability measure under π_s. That is, there must be a gap in the borrower's posterior for almost all signal realizations. As a result, even though the lender would approve a borrower with data $x = 1$ if $v > \mathbf{v}_s$, no borrower is actually approved for $v \in (\mathbf{v}_s, \mathbf{v}^*)$ because this region is excluded from the posterior. This feature holds for almost every signal realization.

By inducing different posterior beliefs, the optimal policy creates a sequence of signals and subgames with different ex post lending cutoffs $\mathbf{v}_s$, which converge to the unconditional lending cutoff $\mathbf{v}^*$. This structure reveals a key principle of optimal disclosure: while revealing partial information about v, the optimal disclosure preserves uncertainty by introducing gaps between the rejection and approval regions. This uncertainty reduces the borrower's ability to game the system.

To understand why the optimal information structure mitigates data manipulation and improves the lender's payoff, we first compare the ex post lending cutoffs under the optimal disclosure policy with the lending cutoff under no disclosure, as shown in the following proposition.

Proposition 4.2. *Let $(\mathcal{S}, \tilde{\sigma})$ denote the optimal policy in Theorem 4.1, let $\boldsymbol{v}^*$ denote the associated unconditional lending cutoff, and let $\boldsymbol{v}_s$ denote the ex post lending cutoff following signal realization s under this policy. Let $\boldsymbol{v}_N$ denote the lending cutoff in the equilibrium with no disclosure. Then*

$$\boldsymbol{v}^* > \boldsymbol{v}_N > \inf_{s \in \mathcal{S}} \boldsymbol{v}_s.$$

Now let us consider why the optimal disclosure policy can dominate the no disclosure policy. Consider two signals under the optimal policy, s_1 and s_2, with ex post lending cutoffs $\boldsymbol{v}_{s_1} < \boldsymbol{v}_N < \boldsymbol{v}_{s_2}$. Proposition 4.2 guarantees the existence of these two signals. In subgame s_2, compare to the no disclosure equilibrium, the lender requires a higher model parameter v to approve a borrower with data $x = 1$. As a result, the approval region $[\boldsymbol{v}_N, \boldsymbol{v}_N + \delta]$ ($\delta \ll 1$) in the no disclosure equilibrium becomes a rejection region under subgame s_2. In subgame s_1, although the lending cutoff $\boldsymbol{v}_{s_1}$ is lower than the cutoff under no disclosure, it is still sufficient to ensure that the lender rejects any borrower with data $x = 1$ for all $v < \boldsymbol{v}_{s_1}$. Compared to the no disclosure equilibrium, subgame s_2 results in a lower payoff for the lender due to more severe data manipulation, whereas subgame s_1 leads to a higher payoff due to reduced manipulation. Under the optimal policy, the positive effect from subgames like s_1 outweighs the negative effect from subgames like s_2, and thus the lender's payoff improves.

Finally, we show that for any optimal disclosure policy, there exists a deterministic policy that induces the same equilibrium outcomes, as stated in the following lemma.

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

The intuition behind Lemma 4.2 is as follows. For any signal realization s, the key equilibrium objects are the lending cutoff $\boldsymbol{v}_s$ and the probabilities $\text{Prof}(v > \boldsymbol{v}_s|s)$ and $\text{Prof}(v < \boldsymbol{v}_s|s)$. A type $X = 0$ borrower's data manipulation incentive depends only on the probability of being approved when switching data from $x = 0$ to $x = 1$. Therefore, as long as $\text{Prof}(v > \boldsymbol{v}_s|s)$ and $\text{Prof}(v < \boldsymbol{v}_s|s)$ remain unchanged, the equilibrium remains the same, regardless of the distribution of v conditional on $v > \boldsymbol{v}_s$ or $v < \boldsymbol{v}_s$. Lemma 4.2 follows directly from this observation.

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. $\sigma(v) = \mathbf{v}_s$ *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 ex post lending cutoff under that signal.*

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

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

   (b) *For any* $s \in \mathcal{S}$*, both the rejection region* R_s *and the acceptance region* A_s *are nonempty, and satisfy* $R_s \subset [0, \mathbf{v}^*]$ *and* $A_s \subset (\mathbf{v}^*, \bar{v}]$.

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 under any signal realization s. This means that data manipulation always exists under any signal realization. This implication rules out some disclosure policies. For example, it is suboptimal 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 $v \leq \mathbf{v}^*$) with high states (where $v > \mathbf{v}^*$) to preserve uncertainty about the model parameter v under any signal realizations. Second, the message function σ is a weakly increasing function on $[0, \mathbf{v}^*]$. This implies that lower values of v in rejection regions correspond to equilibria with lower ex post lending cutoffs. This is intuitive because lower ex post 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* $[0, \bar{c}]$.

A sufficient condition that guarantees Assumption 2 is that $K(\cdot)$ is convex. Many commonly used distribution functions satisfy this assumption, including the uniform distribution, power distribution, and truncated exponential distribution.

The role of this assumption is as follows. Under any signal s, condition (8) implies that $\mathbf{c}_s$ is linear in the probability that a borrower with data $x = 1$ is approved, while condition (10) implies that $K(\mathbf{c}_s)$ is linear in the ex post lending cutoff $\mathbf{v}_s$. The lender's objective (11) implies that the cost of financing a borrower with data $x = 1$ under signal s is proportional to $\mathbf{c}_s K(\mathbf{c}_s)$. Then Assumption 2 ensures that spreading the signal distribution more widely reduces the expected financing cost across all signals. In fact, the following theorem shows that the optimal policy consists of exactly one discrete signal and a continuum of continuous signals under this assumption.

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

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

Theorem 4.3 characterizes the message function in the region where the model parameter v is below the unconditional lending cutoff $\mathbf{v}^*$. The signal space consists of a discrete signal $s = v_a$ and continuous signals $s \in (v_a, \mathbf{v}^*]$. For any $s \in (v_a, \mathbf{v}^*]$, when the signal $s = v$ is realized, the borrower knows that the model parameter is either exactly v or strictly greater than $\mathbf{v}^*$. This confirms our earlier result that the borrower's posterior belief must have gaps. Moreover, since $s = v = \mathbf{v}_s$, the value v represents the highest model parameter that the lender would reject for a borrower with $x = 1$ under the associated signal. Thus, for any signal $s \in (v_a, \mathbf{v}^*]$, each value in the rejection region corresponds to the highest model parameter the lender is willing to reject under its signal. For the discrete signal $s = v_a$, it has the lowest ex post lending cutoff and the lowest manipulation cost cutoff. So in equilibrium, compared to all other signals, a borrower of type $X = 0$ is least likely to manipulate under this discrete signal, and a borrower with data $x = 1$ is least likely to be financed. The rejection region under the discrete signal consists of the bottom values $[0, v_a]$, where v_a is the highest model parameter under which the lender still chooses to reject a borrower with $x = 1$ under this signal.

In the region where $v > \mathbf{v}^*$, the message function is quite flexible. In equilibrium, the borrower cares only about the probability that v exceeds $\mathbf{v}^*$, not the specific distribution. As a result, any message function that preserves this probability can be part of an optimal policy. The following proposition provides an example of such an optimal policy.

Proposition 4.3. *When Assumption 2 is satisfied, there exists a deterministic optimal policy* $(\mathcal{S}, \sigma)$*. For* $v \leq \boldsymbol{v}^*$*, function* $\sigma(v)$ *is characterized in Theorem 4.3. For* $v > \boldsymbol{v}^*$*, there exists* $v_b \in (\boldsymbol{v}^*, \bar{v})$*, such that:*

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

Here $\gamma : [v_b, \bar{v}] \to [v_a, \boldsymbol{v}^*]$ 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 $\gamma(v_b) = v_a, \gamma(\bar{v}) = \boldsymbol{v}^*$, and $\frac{g(v_b)}{\gamma'(v_b) g(\gamma_a)} = \frac{G(v_b) - G(\boldsymbol{v}^*)}{G(v_a)}$.

Based on the above results, for any given $\mathbf{v}^*$, we can solve for the message function (including the boundaries v_a and v_b) using the conditions in Theorem 4.3 and Proposition 4.3. Thus, the only remaining part of the optimal policy to be determined is the cutoff $\mathbf{v}^*$. The optimal policy also has a simple structure. According to Theorem 4.3, the signal space consists of a discrete signal $s = v_a$ and continuous signals $s \in (v_a, \mathbf{v}^*]$. Under the discrete signal $s = v_a$, the borrower's posterior belief about v is supported on two disjoint intervals, $[0, v_a]$ and $(\mathbf{v}^*, v_b]$. For any continuous signal $s \in (v_a, \mathbf{v}^*]$, the posterior belief is a two-point distribution, with one point above $\mathbf{v}^*$ and one below. Thus, although the borrower receives additional information about v, his posterior belief remains highly uncertain.

Thus, finding the optimal message function reduces to finding the optimal cutoff $\mathbf{v}^*$. The lender's problem in Proposition 3.1 becomes a one-dimensional maximization problem,

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

where W is defined by (12) under a disclosure policy with signal space $\mathcal{S} = [v_a, \mathbf{v}^*]$ and message function σ. The value v_a and message function σ are solved by Theorem 4.3 and Proposition 4.3. Since the optimization problem does not admit a closed-form solution for $\mathbf{v}^*$, we present a numerical example of an optimal policy in Figure 4.

As discussed earlier, the optimal policy includes a discrete signal that induces a posterior belief supported on two disjoint intervals, shown as the two blue intervals in Figure 4. The policy also features a continuum of signals, represented by the red segments of the message function. For each realization of the continuous signal, the posterior belief is a two-point distribution: one point below $\mathbf{v}^*$ and one point above it (depicted as red points). The left red segment corresponds to a 45-degree straight line, while the slope of the red segment above $\mathbf{v}^*$, the function γ, determines the relative probability between the two points. Under this optimal policy, the equilibrium approval probability under each signal s is given by

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

which is an increasing function of the signal $s = \mathbf{v}_s$.

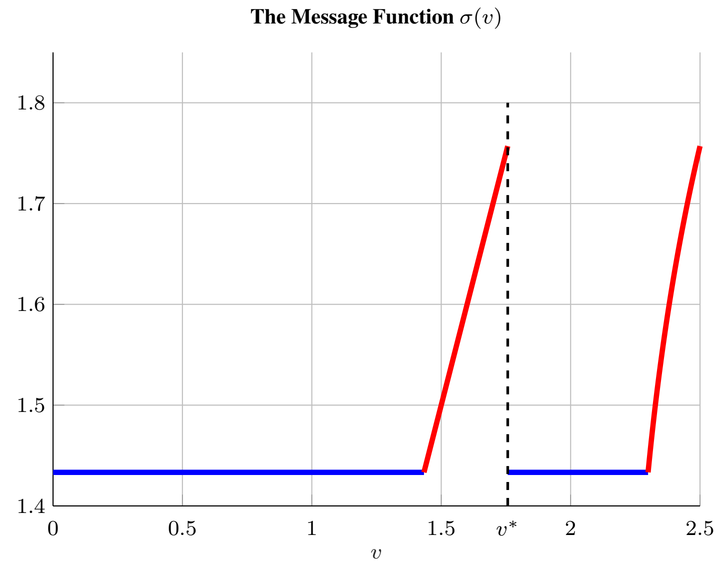

Figure 4: **Optimal Policy**. We present a numerical example of the optimal policy. In this exercise, we use the following parameters: $\mu = 0.35$, $\bar{c} = 2$, $b = 1.7$, $I = 1$, $\bar{v} = 2.5$. Both c and v follow uniform distribution. In this exercise, $v_a = 1.433$, $\mathbf{v}^* = 1.757$, $v_b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

4.1 Monotone Policies

The general form of optimal disclosure policies belongs to a special class of pairwise disclosure rules studied in Kolotilin and Wolitzky (2020). A key feature of the optimal disclosure is its non-monotonicity: equilibrium outcomes such as manipulation intensity can be non-monotonic in the underlying state.

In practice, however, model users (like the lender in our model) may face institutional or operational constraints that limit their ability to fully design arbitrary disclosure rules, and in fact, they often adopt relatively simple disclosure rules with only a limited number of signal realizations. In some cases, they choose not to disclose any information at all, keeping the model fully opaque,[11] a practice that our main analysis shows to be suboptimal.

In this section, we provide a rationale for these empirical observations by introducing an additional constraint. Specifically, we assume that the lender is limited to using monotone policies, where the message function must be monotonic in the true model parameter v. It is clear that under monotone disclosure policies, the borrower's posterior belief about v must be an interval. Monotone policies have been extensively studied in the literature and are simple to implement in practice, for instance through rating systems. The following is a formal definition of monotone

[11]This issue is especially pronounced when lenders adopt increasingly complex models to screen borrowers. See the media discussion in *The Secret (Controversial) Sauce*, published in the FinTech Collective newsletter on September 25, 2023.

policies.

Definition 4.1. A deterministic disclosure policy $(\mathcal{S}, \sigma)$ is called a monotone policy if and only if the message function $\sigma(v)$ is a monotonic function of v.

Unlike the results in our main model, where the optimal disclosure policy includes one discrete signal realization and a continuum of continuous signals, the following result shows that under monotone policies, the optimal disclosure consists only of discrete signals.

Proposition 4.4. *In any optimal monotone disclosure policy, the signal distribution is discrete.*

The intuition behind Proposition 4.4 is as follows. Under the restriction to monotone policies, any continuous signal realization must perfectly reveal a specific true model parameter v. However, this implies that the lender's payoff from such a signal is zero, which is the worst possible outcome. Since the lender strictly prefers to avoid fully revealing the true model parameter, the optimal monotone policy cannot include any continuous signals. As a result, it must consist only of discrete signal realizations.

Within the class of monotone policies, the following result provides a condition under which no disclosure is optimal.

Proposition 4.5. *If $g(v)$ is weakly decreasing, then the optimal monotone disclosure policy is no disclosure.*

Proposition 4.5 shows that, when disclosure is restricted to monotone policies, no disclosure is optimal if the probability density function of the model parameter v is decreasing. This condition implies that v is left-skewed. The intuition is as follows. When the distribution of v is left-skewed, the lending cutoff v_N in the no disclosure equilibrium is relatively low. As a result, the probability that the borrower chooses to manipulate his data is also low, and the lender's expected payoff is relatively high. Partitioning the domain of v into multiple intervals yields limited gains in the low-v region, where manipulation is already low in no disclosure equilibrium. However, in the high-v region, the equilibrium lending cutoff rises substantially, which reduces the lender's payoff. Therefore, under monotone disclosure and a left-skewed distribution of v, it is optimal not to disclose any model information v.

4.2 Other Applications

While our main analysis is framed in a lending setting, the underlying mechanism applies more broadly. In this section, we illustrate how the same logic can be applied to a variety of real-world contexts, including college admissions, bank stress testing, health insurance underwriting, and online marketplaces. These examples share the core features of our model.

School Admission. A graduate program uses GRE scores to screen applicants. The program decides whether to admit a student. Admitting a student incurs a fixed cost I. If the student eventually succeeds academically, the program receives a normalized reputation utility of 1 (e.g., from improved reputation); if the student fails, the reputation utility is 0. Academic success is unobservable at the time of admission, but the GRE score provides informative signals. Specifically, if students do not prepare for the GRE, getting a score below 300 always implies academic failure, while a score above 300 implies a probability of success equal to v, where the common prior is $v \sim G(\cdot)$. The school observes the realization of v from historical data, but students do not. Being admitted generates a utility of b for students. Based on their belief about v, students choose how much effort to invest in preparing for the GRE. Preparation does not affect the probability of their academic success, but can raise their GRE score. Students with a baseline score (without preparation) below 300 can boost it to above 300 by incurring a private effort cost $c \sim K(\cdot)$. The school decides how much information about v it discloses to students.

Bank Stress Test and Communication. A regulator must decide whether to pass a bank based on its underlying quality (e.g., its risk profile). Failing a bank yields a payoff of zero. Passing a high-quality bank generates a payoff of $v - I$, while passing a low-quality bank results in a loss of $-I$, where $v \sim G(\cdot)$. The regulator privately observes the realization of v, which may reflect broader objectives such as financial stability, risk tolerance, or expected market conditions. Passing the test always generates a benefit of b for the bank. The bank's true quality is unobservable but can be inferred from its self-reported stress test results. If the bank reports truthfully, passing the stress test always indicates high quality, while failing indicates low quality. However, a low-quality bank can manipulate its stress test results to appear as passing by incurring a private cost $c \sim K(\cdot)$. The regulator can commit to a disclosure policy over v to communicate with the bank, thereby influencing their incentives to manipulate test results and maximizing its expected payoff. This communication may involve conveying the regulator's regulatory goals or projections about future market conditions which are linked to the value of v.

Health Insurance Screening and Disclosure. A health insurance company decides whether to offer an insurance contract to a patient based on the patient's underlying health risk. Not offering coverage yields a payoff of zero. Offering insurance to a low-risk patient generates a payoff of $v - I$, while offering insurance to a high-risk patient results in a payoff of $-I$, where $v \sim G(\cdot)$. The insurer privately observes the realization of v, which may reflect the insurance company's risk tolerance, or internal forecasts such as expected claim costs, market dynamics, or regulatory considerations. The patient receives a benefit of b when covered by the insurance. The patient's true health risk is unobservable but can be inferred from self-reported medical history. If the patient reports truthfully, the medical record perfectly reveals their risk type. However, a high-risk patient can conceal part of their medical history to appear low risk by incurring a private cost $c \sim K(\cdot)$

(e.g., due to regulatory or reputational risks). The insurer can commit to a disclosure policy over v to communicate with patients. By strategically revealing information about v (like the company's risk tolerance or internal forecast model), the insurer can influence patients' incentives to conceal their medical history.

Online Seller Promotion and Rating Manipulation. In online marketplaces, customer ratings play a central role in determining which sellers gain visibility and platform support. A platform may choose to feature certain sellers, but featuring each seller incurs a cost of I for the platform. Being featured generates a benefit of b for each seller. The platform's payoff from featuring a seller depends on their unobserved quality: high-quality sellers generate a payoff of $v - I$ for the platform, while low-quality sellers result in a loss of $-I$, where $v \sim G(\cdot)$ is observed privately by the platform based on internal forecasts like user satisfaction metrics or long-term retention outcomes. Ratings provide information about seller quality. When ratings are unmanipulated, a high rating means high quality, and a low rating means low quality. However, low-quality sellers can manipulate their ratings by purchasing fake reviews at a private cost $c \sim K(\cdot)$. To mitigate such manipulation, the platform can design a disclosure policy over the realized value of v. By strategically revealing information about the benefit of featuring high-quality sellers (for example, historical sales data), the platform can deter sellers' rating manipulation incentives and improve the efficiency of its promotional decisions.

5 Extensions

5.1 Commitment: Information vs Decision Rules

Our paper focuses on the role of disclosing model information. Section 3.3 shows that committing to a higher model parameter threshold for approval decisions can also improve the lender's payoff.

In this section, we compare the role of commitment to information disclosure (disclosing model information) with commitment to action (committing to how the model is used). We show that even when the lender commits to the optimal approval decision by specifying which model parameters lead to borrower approval, she may still achieve a higher payoff by disclosing additional information about the model itself. Thus, the benefit of model transparency cannot be fully replaced by stronger commitment to the decision rule.

To begin our analysis, consider a benchmark case where the lender commits to approval decisions. Suppose the lender commits to approving a borrower with data $x = 1$ when $v \geq z$, and rejecting any borrower with $x = 0$ outright.[12] Then the lender's problem is:

[12]We could consider a more flexible lending rule where the lender approves a borrower with $x \in \{0, 1\}$ if $v \geq v_x$, allowing approval of borrowers with $x = 0$. This flexibility might improve the lender's profit by deterring manipulation

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

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

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

We are interested in whether additional disclosure about the model parameter v can further improve the lender's payoff, even when she has full commitment power over approval decisions. After committing to the optimal approval cutoff v_c, suppose the lender can also choose a disclosure policy $(\mathcal{S}, \sigma)$ and reveal information about v before the borrower chooses his manipulation strategy. The following proposition shows that the lender may achieve a higher payoff through additional disclosure about the model parameter v.

Proposition 5.1. *Suppose the lender commits to only approving the borrower with data $x = 1$ when the model parameter $v \geq v_c$. 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, Proposition 5.1 shows that the benefits of model transparency cannot be fully replaced by commitment to approval decisions. Instead, the lender can achieve a higher payoff by disclosing additional information about the model, even when she has full commitment power over approval 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 model transparency interacts with costly verification in the lender's problem.

We introduce one additional assumption to the main model. After a signal s is disclosed, the lender 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 nonzero probability of being approved, so we just need to focus on the lender's decision to verify the true type of a borrower with data $x = 1$. Consider a subgame with posterior belief π_s, a borrower with $X = 0$ chooses to

incentives. However, under certain distributional assumptions on $K(c)$, it is optimal for the lender to commit to always rejecting borrowers with $x = 0$. This case also allows for a comparable analysis with our baseline model.

manipulate if and only if his manipulation cost c is lower than $\mathbf{c}_s$. Then in equilibrium, if the lender verifies the borrower's type, her payoff is

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

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

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

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

$$W_V > W_{NV} \iff v > v_e = I + \frac{\mu + (1-\mu) K(\mathbf{c}_s)}{\mu} t,$$

and the borrower's type must be verified before he is approved. However, this implies that the borrower with type $X = 0$ will never manipulate his data x, and thus the lender ultimately has no incentive to verify the borrower's type. Thus, $t < \frac{(1-\mu)K(\mathbf{c}_s)}{\mu+(1-\mu)K(\mathbf{c}_s)} I$ cannot hold in equilibrium for any subgame s. If in a subgame s, the lender verifies the borrower's type with positive probability, it must be that

$$t = \frac{(1-\mu)K(\mathbf{c}_s)}{\mu + (1-\mu)K(\mathbf{c}_s)} I.$$

Therefore, there can be at most one subgame s in which verification occurs with positive probability. The following theorem provides a complete characterization of the optimal policy with costly type verification.

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

1. *When $t \geq t^v$, the lender never verifies the borrower's type, and the optimal disclosure policy coincides with that 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 in two steps:*

   (a) *The lender first discloses whether $v > v_v$.*

   (b) *If $v > v_v$, the lender verifies the type of a borrower with data $x = 1$ with a positive probability, and approves the borrower if his type is either not verified or verified to be $X = 1$.*

   (c) *If $v \leq v_v$, there is no verification. 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 disclosure and verification interact in a simple way: when the model parameter v is sufficiently high ($v > v_v$), only verification is used to deter manipulation, and no further disclosure is needed. When v is low ($v \leq v_v$), only disclosure is used to deter manipulation, and verification is not used.

6 Conclusion

We study the optimal model transparency in a lending setting where a lender uses a privately observed predictive model to screen a borrower. Under the optimal disclosure policy, information about the model is partially disclosed to the borrower. The lender uses borrower data less intensively in her lending decisions, reducing data manipulation and improving efficiency. Despite receiving additional information from the lender, the borrower's posterior belief about the model remains quite uncertain. No disclosure can be optimal when disclosure policies are restricted to be monotonic. Model transparency can improve the lender's payoff even when she has additional commitment power over lending decisions or the ability to verify the borrower's type at a cost. Overall, our results highlight that model transparency can serve as a powerful tool to mitigate strategic behavior in data-driven decision-making environments.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Fischer, P. E. and R. E. Verrecchia (2000). Reporting bias. *The Accounting Review 75*(2), 229–245.

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.

Gamba, A. and C. Hennessy (2024). Manipulable data, goodhartjs law, and credit risk prediction. *Goodhart's Law, and Credit Risk Prediction (June 03, 2024)*.

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.

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*.

Kolotilin, A. and A. Wolitzky (2020). Assortative information disclosure.

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.

Rajan, U. and Y. Xiong (2024). Data regulation in credit markets. *Available at SSRN 5064460*.

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 Proof of Lemma 3.3

Given $v_\delta = \mathbf{v}_N + \delta$, the borrower's manipulation cutoff $\mathbf{c}_\delta$ satisfies $b\left(1 - G\left(v_\delta\right)\right) = \mathbf{c}_\delta$. The lender's payoff is

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

so

$$\left.\frac{dW_\delta}{d\delta}\right|_{\delta=0} = -\left[\mu \mathbf{v}_N - \left(\mu + (1-\mu) K\left(\mathbf{c}_N\right)\right) I\right] g(v) + \int_{v_\delta}^{\bar{v}} \left(-(1-\mu) I g(v) \left.\frac{dK\left(\mathbf{c}_\delta\right)}{d\delta}\right|_{\delta=0} \right) dv. \tag{19}$$

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

A.2 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 $\mathbf{c}_N$ and $\mathbf{v}_N$, 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

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

where $v_1' < \mathbf{v}_N$ satisfies

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

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

$$\left(\mathbf{v}_{s_1'}, \mathbf{c}_{s_1'}\right) = \left(\mathbf{v}_{s_2'}, \mathbf{c}_{s_2'}\right) = \left(\mathbf{v}_N, \mathbf{c}_N\right).$$

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

$$\begin{aligned} W' =& \text{Prob}\left(v \in [0, v_1'] \cup [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_1'}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_{s_1'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq \mathbf{v}_{s_1'}\right\}}\right] \\ &+ \text{Prob}\left(v \in (v_1', \mathbf{v}_N) \cup (\mathbf{v}_N + \epsilon_1, \bar{v} - \epsilon_2)\right) \cdot \mathbb{E}^{s_2'}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_{s_2'}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq \mathbf{v}_{s_2'}\right\}}\right]. \end{aligned} \tag{20}$$

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

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

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 $(v_1', \mathbf{v}_N) \cup (\mathbf{v}_N + \epsilon_1, \bar{v} - \epsilon_2)$. When signal s_1'' is disclosed, the borrower knows that the true state $v \in [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]$. In the no discloure case, the lender's payoff in state $[\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]$ is close to zero (in order $o(\epsilon_1)$), as $\mathbf{v}_N$ 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 $v \in [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]$ 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

$$\begin{aligned} W'' =& \text{Prob}\left(v \in [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]\right) \cdot \mathbb{E}^{s_1''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_{s_1''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq \mathbf{v}_{s_1''}\right\}}\right] \\ &+ \text{Prob}\left(v \in [0, v_1'] \cup [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \mathbb{E}^{s_2''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_{s_2''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq \mathbf{v}_{s_2''}\right\}}\right] \\ &+ \text{Prob}\left(v \in (v_1', \mathbf{v}_N) \cup (\mathbf{v}_N + \epsilon_1, \bar{v} - \epsilon_2)\right) \cdot \mathbb{E}^{s_3''}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_{s_3''}\right)\right) I\right) \cdot \mathbb{1}_{\left\{v \geq \mathbf{v}_{s_3''}\right\}}\right]. \end{aligned} \tag{21}$$

It's obvious that the last term in (21) equals the last term in (20), because equilibria under signal realizations s_3'' and s_2' are the same. Then

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

Note that $\mathbf{v}_{s_1'} = \mathbf{v}_N$, we know

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

Then

$$
W'' - W' \geq \begin{array}{c} \text{Prob}\left(v \in [\bar{v} - \epsilon_2, \bar{v}]\right) \cdot \left[(1-\mu) I \left(K\left(\mathbf{c}_N\right) - K\left(\mathbf{c}_{s_2''}\right)\right)\right] - \\ \text{Prob}\left(v \in [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]\right) \cdot \mathbb{E}\left[\left(\mu v - \left(\mu + (1-\mu) K\left(\mathbf{c}_N\right)\right) I\right) | v \in [\mathbf{v}_N, \mathbf{v}_N + \epsilon_1]\right]. \end{array}
$$

In the equilibrium of subgame s_2'',

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

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

$$
\begin{aligned}
\frac{\mathbf{c}_N - \mathbf{c}_{s_2''}}{b} &= \frac{1 - G\left(\bar{v} - \epsilon_2\right) + G\left(\mathbf{v}_N + \epsilon_1\right) - G\left(\mathbf{v}_N\right)}{1 - G\left(\bar{v} - \epsilon_2\right) + G\left(v_1'\right) + G\left(\mathbf{v}_N + \epsilon_1\right) - G\left(\mathbf{v}_N\right)} - \frac{1 - G\left(\bar{v} - \epsilon_2\right)}{1 - G\left(\bar{v} - \epsilon_2\right) + G\left(v_1'\right)} \\
&= \frac{G\left(v_1'\right)}{\left[1 - G\left(\bar{v} - \epsilon_2\right) + G\left(v_1'\right)\right]^2} g\left(\mathbf{v}_N\right) \epsilon_1 + o\left(\epsilon_1\right).
\end{aligned}
$$

Then

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

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.3 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 $\mathbf{v}^* = \mathbf{v}_N$, 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 $v_1 \in \text{supp}\left(\pi_{s_1}\right)$ and $v_2 \in \text{supp}\left(\pi_{s_2}\right)$, then for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$, and open balls $B\left(v_1, \epsilon_1\right)$, $B_2\left(v_2, \epsilon_2\right)$, we must have $\text{Prob}\left(B\left(v_1, \epsilon_1\right) | s_1\right) > 0$ and $\text{Prob}\left(B\left(v_2, \epsilon_2\right) | s_2\right) > 0$. Let us choose $\epsilon_1, \epsilon_2 < \frac{v_2 - v_1}{3}$, then $\inf B\left(v_1, \epsilon_1\right) > \sup B_2\left(v_2, \epsilon_2\right)$. Denote $\text{Prob}\left(B\left(v_1, \epsilon_1\right) | s_1\right) = K_1 > 0$ and $\text{Prob}\left(B\left(v_2, \epsilon_2\right) | s_2\right) = K_2 > 0$. If $f\left(s_1\right) K_1 \geq f\left(s_2\right) K_2$,[13] 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 $\left(\hat{S}, \hat{\sigma}\right)$, $v_2 \notin \text{supp}\left(\pi_{s_2}\right)$. 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 $\left(\mathbf{v}_s, \mathbf{c}_s\right)$ 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 $\left(\hat{S}, \hat{\sigma}\right)$,

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

under s_2, the equilibrium variables $(\mathbf{v}_{s_2}, \mathbf{c}_{s_2})$ 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

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

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 $\mathbf{v}^*$, such that $\bar{R}_s \leq \mathbf{v}^* \leq \underline{A}_s$ for all $s \in \mathcal{S} - Q$. Since Q is a zero-measure set, the statement of the theorem must be true.

A.4 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 reveal 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.

Internet Appendix: Model Transparency with Manipulable Input

Jian Sun

I.A.1 Omitted Proofs

I.A.1.1 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 $b\left(1-\mathbf{v}_N\right)=\mathbf{c}_N$ and $\mu\mathbf{v}_N V=\left(\mu+(1-\mu)\,\mathbf{c}_N\right)I$. The unique solution is $\left(\mathbf{v}_N=\frac{\mu I+(1-\mu)Ib}{\mu V+(1-\mu)Ib}, \mathbf{c}_N=b\cdot\frac{\mu(V-I)}{\mu V+(1-\mu)Ib}\right)$.

A Binary Signal

When $v\in A$, let's verify that in equilibrium $\mathbf{v}_1=0.54$, $\mathbf{c}_1=0.34$, and $W_1=0.19$. The two equilibrium conditions are $\text{Prob}\left(v>\mathbf{v}_1|A\right)\cdot b=\mathbf{c}_1$, and $\mu\mathbf{v}_1 V=\left(\mu+(1-\mu)\,\text{Prob}\left(c\leq\mathbf{c}_1\right)\right)I$. Since $A=[0,0.54)\cup(0.64,0.91)$, and $\mathbf{v}_1=0.54$, $\mathbf{c}_1=0.34$, 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 conditions hold [14] under $A=[0,0.54)\cup(0.64,0.91)$, and $\mathbf{v}_1=0.54$, $\mathbf{c}_1=0.34$. 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 $\mathbf{v}_2=0.64$, $\mathbf{c}_2=0.48$. 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 $W_s=W_1+W_2=0.19+0.09=0.28$.

I.A.1.2 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

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

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

$$d_s(0)-d_s(1)\geq\frac{c}{b}>0 \tag{I.A.1.1}$$

[14] 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.

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 (I.A.1.1) 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 (I.A.1.1). So in any equilibrium s, we must have $\gamma_s(1, c) = 0$ for any $c > 0$.

I.A.1.3 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

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

When the lender faces to a borrower with data $x = 1$, if she makes the loan, her profit (2) becomes $\mu v - [\mu + (1 - \mu) K(\mathbf{c}_s)] \cdot I$, and she makes the loan ($\alpha_s(1, v) > 0$) only if $v \geq \mathbf{v}_s = \frac{[\mu + (1-\mu) K(\mathbf{c}_s)] \cdot I}{\mu}$. Since $d_s(1) = \int_0^{\bar{v}} \alpha_s(1, v) \pi_s dv$, we must have $\text{Prob}(v > \mathbf{v}_s | s) \leq d_s(1) \leq \text{Prob}(v \geq \mathbf{v}_s | s)$. The lender's payoff in the subgame s is

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

I.A.1.4 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 $F(s)$, to maximize the expected profit

$$W = \mu \int_{s \in \mathcal{S}} \mathbb{E}\left[(v - \mathbf{v}_s)^+ | s\right] dF(s). \tag{I.A.1.2}$$

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:

$$\int_{s\in\mathcal{S}} \pi_s dF(s) = g(v). \tag{I.A.1.3}$$

Under each posterior π_s, the equilibrium conditions are

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

$$c \leq \mathbf{c}_s = b \cdot d_s(1),$$

and

$$\mathbf{v}_s = \frac{\left[\mu + (1-\mu) K(\mathbf{c}_s)\right] \cdot I}{\mu}.$$

These conditions jointly imply

$$\text{Prob}\left(v > \mathbf{v}_s \middle| s\right) \leq \frac{K^{-1}\left(\frac{\mu \mathbf{v}_s}{I(1-\mu)} - \frac{\mu}{1-\mu}\right)}{b} \leq \text{Prob}\left(v \geq \mathbf{v}_s \middle| s\right), \tag{I.A.1.4}$$

where $\text{Prob}(\cdot| s)$ 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 (I.A.1.2), subject to conditions (I.A.1.3) and (I.A.1.4).

I.A.1.5 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 $(\mathbf{v}_{s_1}, \mathbf{c}_{s_1}) = (\mathbf{v}_{s_2}, \mathbf{c}_{s_2})$. Then consider a new policy $(\mathcal{S}', \tilde{\sigma}')$, where $\mathcal{S}' = \{s'_0\} \cup \mathcal{S} \backslash \{s_1, s_2\}$ and

$$\tilde{\sigma}'(s|v) = \tilde{\sigma}(s|v)\, \mathbb{1}_{\mathcal{S}\backslash\{s_1,s_2\}}(s) + \left(\tilde{\sigma}(s_1|v) + \tilde{\sigma}(s_2|v)\right) \mathbb{1}_{\{s'_0\}}(s)$$

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 $\mathbf{c}_s$ and lending cutoff $\mathbf{v}_s$ are all unchanged by verifying the equilibrium conditions for the subgames

s_1 and s_2 under the policy $(\mathcal{S}, \tilde{\sigma})$.

I.A.1.6 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 $\mathbf{v}_s = v$. 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 payoff 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 $W_F = 0$.

I.A.1.7 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 $\text{Prob}(v = \mathbf{v}_{s_0} | s_0) > 0$, and the lender rejects a borrower with data $x = 1$ only when $v = \mathbf{v}_{s_0}$. R_{s_0} being an empty set means that $\alpha_{s_0}(1, \mathbf{v}_{s_0}) \in (0, 1)$. 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:

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

$(\mathcal{S}_1, \tilde{\sigma}_1)$ 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 $v = \mathbf{v}_{s_0}$, then with probability $\alpha_{s_0}(1, \mathbf{v}_{s_0})$, the true state $\mathbf{v}_{s_0}$ 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 $v = \mathbf{v}_s$. 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 $v = \mathbf{v}_s$. Suppose there exists a set $H \subset \mathcal{S}$, such that Prob $(H) > 0$ and for any $s \in H$, $\mathbf{v}_s \in A_s$. Take any $s_0 \in H$, if Prob $(\mathbf{v}_{s_0}|s_0) = k > 0$, and $\alpha_{s_0}(1, \mathbf{v}_{s_0}) \in (0, 1)$, 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 Prob $(\mathbf{v}_{s_0}|s_0) = k = 0$, since $\mathbf{v}_{s_0} \in A_{s_0}$ then for any $\epsilon > 0$, we must have Prob $(v \in (\mathbf{v}_{s_0}, \mathbf{v}_{s_0} + \epsilon)|s_0) > 0$. Following the proof of Proposition 3.2, we know there must exist an alternative disclosure policy that reveals the states when $v \in (\mathbf{v}_{s_0}, \mathbf{v}_{s_0} + \epsilon)$ 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 $\mathbf{v}_s \in A_s$ for all $s \in H$, we must have Prob$(H) = 0$. Then for almost all $s \in \mathcal{S}$, we must have $A_s = \{v \in \text{supp}(\pi_s) | \alpha_s(1, v) = 1\}$.

I.A.1.8 Proof of Proposition 4.1

We establish the following lemma first.

Lemma I.A.1.1. *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 $(\tilde{\mathbf{v}}_s, \tilde{\mathbf{c}}_s)$ and $(\mathbf{v}_s, \mathbf{c}_s)$ be equilibrium outcomes under policies $(\mathcal{S}, \tilde{\sigma})$ and $(\mathcal{S}, \sigma)$, respectively. Then the following properties hold:*

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

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 (17) and (18).*

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

$$\sup R_{s_1} \leq \inf R_{s_2} \tag{I.A.1.5}$$

and

$$\sup A_{s_1} \leq \inf A_{s_2}. \tag{I.A.1.6}$$

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 $(\tilde{\mathbf{v}}_{s_1}, \tilde{\mathbf{c}}_{s_1})$ and $(\tilde{\mathbf{v}}_{s_2}, \tilde{\mathbf{c}}_{s_2})$ under these two signal realizations. Without loss of generality, let's assume $\tilde{\mathbf{v}}_{s_1} < \tilde{\mathbf{v}}_{s_2}$. Denote the ex ante lending cutoff as $\mathbf{v}^*$ in this case. Suppose for s_1, s_2, the condition

$$\sup A_{s_1} \leq \inf A_{s_2}. \tag{I.A.1.7}$$

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

$$\tilde{f}(s_1)\, w_1(v) + \tilde{f}(s_2)\, w_2(v) = \tilde{f}(s_1)\, \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v) + \tilde{f}(s_2)\, \tilde{\pi}_{s_2} \cdot \mathbb{1}_B(v), \tag{I.A.1.8}$$

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

and

$$\int_0^{\bar{v}} w_1(v)\, dv = \int_0^{\bar{v}} \tilde{\pi}_{s_1} \cdot \mathbb{1}_B(v)\, dv. \tag{I.A.1.9}$$

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 (I.A.1.8) and (I.A.1.9). Then the distribution of posteriors $\{F, \pi_s\}$ can be induced by a disclosure policy $(\mathcal{S}, \sigma)$. Now the condition (I.A.1.7) 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

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

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

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

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

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

Based on this, we can check that all other equilibrium conditions are also satisfied, and this implies $(\mathbf{v}_{s_1}, \mathbf{c}_{s_1}) = (\tilde{\mathbf{v}}_{s_1}, \tilde{\mathbf{c}}_{s_1})$. Similarly, we can check $(\mathbf{v}_{s_2}, \mathbf{c}_{s_2}) = (\tilde{\mathbf{v}}_{s_2}, \tilde{\mathbf{c}}_{s_2})$. 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 (I.A.1.7) is no longer violated.

The proof strategy still works if condition (I.A.1.9) is not satisfied in the optimal policy $(\mathcal{S}, \sigma)$. Besides, note that the third property in Lemma I.A.1.1 implies the second property in Lemma I.A.1.1, 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 $\mathbf{v}^*$ 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 I.A.1.1 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 I.A.1.1 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 $\mathbf{v}_s \geq \mathbf{v}^*$ 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 $\sup A_{s_m} = v_1 > \mathbf{v}_{s_m}$. Since Lemma I.A.1.1 shows that A_{s_m} must be an interval, we consider the following two cases.

If $(\mathbf{v}_{s_m}, v_1) \subset A_{s_m}$, 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 $(\mathbf{v}_{s_m}, \mathbf{v}_{s_m} + \epsilon)$ or not, for a sufficiently small but positive ϵ.

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

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

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 $(\mathbf{v}_{s_m}, \mathbf{v}_{s_m} + \epsilon_m) \subset \operatorname{supp}(\pi_{s_m})$. 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 $(\mathbf{v}_{s_m}, \mathbf{v}_{s_m} + \epsilon)$ 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.

Then we've shown that for almost all s, we must have $\mathbf{v}_s < \mathbf{v}^*$. Besides, since unconditionally the borrower with data $x = 1$ must be rejected for all $v \leq \mathbf{v}^*$, we conclude that $\mathbf{v}^* = \sup_{s \in \mathcal{S}} \mathbf{v}_s$.

I.A.1.9 Proof of Proposition 4.2

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

Lemma I.A.1.2. *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 $(\mathbf{v}_{s_1}, \mathbf{c}_{s_1})$ and $(\mathbf{v}_{s_2}, \mathbf{c}_{s_2})$ satisfying $\mathbf{v}_{s_1} < \mathbf{v}_{s_2}$. Let $\hat{s}$ be the "combined"signal with posterior belief*

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

then the lending market equilibrium $(\mathbf{v}_{\hat{s}}, \mathbf{c}_{\hat{s}})$ satisfies

$$\mathbf{v}_{s_1} < \mathbf{v}_{\hat{s}} < \mathbf{v}_{s_2}$$

and

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

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

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

where

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

and

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

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

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

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

Note that the signal in the no disclosure policy is a "combined" signal of all signals in the optimal policy $(\mathcal{S}, \sigma)$, Lemma I.A.1.2 implies that there exist sets $A, B \subset \mathcal{S}$ with positive measures, such that $\mathbf{v}_N < \mathbf{v}_s$ for all $s \in A$ and $\mathbf{v}_N > \mathbf{v}_s$ for all $s \in B$. Then we must have $\mathbf{v}_N > \inf_{s \in \mathcal{S}} \mathbf{v}_s$. Together with Proposition 4.1, we can also conclude that $\mathbf{v}^* > \mathbf{v}_N$.

I.A.1.10 Proof of Lemma 4.2

This is an obvious result of Lemma I.A.1.1.

I.A.1.11 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 I.A.1.1 and Lemma 4.1.

I.A.1.12 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:

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

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

Let c_0 be the solution of

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

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 I.A.1.2 implies that $c_0 < \sup_{s\in\mathcal{S}_1} \mathbf{c}_s$, and the equilibrium lending cutoff v_0 in this case also satisfies $v_0 < \max_{s\in\mathcal{S}_1} \mathbf{v}_s < \mathbf{v}^*$. Since function $xK(x)$ is convex on $x \in [0, \bar{c}]$, we must have

$$\left(\int_{s\in\mathcal{S}_1} f(s) \cdot ds\right) (1-\mu) \frac{c_0}{b} K(c_0) \leq \int_{s\in\mathcal{S}_1} f(s) \cdot \frac{\mathbf{c}_s}{B} \cdot (1-\mu) K(\mathbf{c}_s) ds. \tag{I.A.1.10}$$

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

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

Condition (I.A.1.10) 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 $v_4 \in (0, \mathbf{v}^*)$, such that $\sigma(v) = \sigma(0)$ for all $v \leq v_4$, and $\sigma(v) > \sigma(v_4)$ for all

$v > v_4$. $v_4 \neq \mathbf{v}^*$ 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 $\mathbf{v}_s$ as the message function, we can use $\mathbf{v}_{\sigma(v)}$ to represent the equilibrium lending cutoff of the equilibrium whose support contains v. The next step is to show that we must have $\inf \operatorname{supp} \pi_{\sigma(v)} = \mathbf{v}_{\sigma(v)}$ for all $v > v_4$. For the sake of contradiction, suppose there exists $v_1 > v_4$, and $\inf \operatorname{supp} \pi_{\sigma(v_1)} < \mathbf{v}_{\sigma(v_1)}$. 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

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

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

$$\frac{\text{Prob}(v \in F_{01})}{\text{Prob}(v \in E_{01}) + \text{Prob}(v \in F_{01})} = \frac{\mathbf{c}_{s_0}}{b}.$$

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:

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

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 $c_0 \in (\mathbf{c}_{s_0}, \mathbf{c}_{s_0} + \delta)$, $c_1 \in (\mathbf{c}_{s_1} - \delta, \mathbf{c}_{s_1})$ and $p_1 + p_0 = f(s_{01}) + f(s_{11})$, such that

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

and

$$p_1 c_1 K(c_1) + p_0 c_0 K(c_0) \leq f(s_0)\,\mathbf{c}_{s_0} K(\mathbf{c}_{s_0}) + f(s_1)\,\mathbf{c}_{s_1} K(\mathbf{c}_{s_1}). \tag{I.A.1.11}$$

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

$xK(x)$ being convex and $\mathbf{v}_{s_1} > \mathbf{v}_{s_0}$ 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 $(\hat{\mathcal{S}}_1, \hat{\sigma}_1)$ are the signals s_{01} and s_{11} in $(\mathcal{S}_1, \sigma_1)$ and $\hat{s}_{01}$ and $\hat{s}_{11}$ in $(\hat{\mathcal{S}}_1, \hat{\sigma}_1)$. 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 $(\hat{\mathcal{S}}_1, \hat{\sigma}_1)$, we just need to compute the difference between lender's payoff under $\hat{s}_{01}$ and $\hat{s}_{11}$ in $(\hat{\mathcal{S}}_1, \hat{\sigma}_1)$ 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

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

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

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

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

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

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

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

Then the difference is

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

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 $v_0 < \mathbf{v}^*$. And the last inequality is because condition (I.A.1.11).

The above analysis shows that we can weakly improve the lender's payoff if there exists $v_1 > v_4$ such that $\inf \operatorname{supp} \pi_{\sigma(v_1)} < \mathbf{v}_{\sigma(v_1)}$. 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 $v_4 \in (0, \mathbf{v}^*)$, such that $\sigma\left(v\right)$ is constant when $v \leq v_4$. For all $v > v_4$, $\mathbf{v}_{\sigma(v)} = \sigma\left(v\right) = \inf \operatorname{supp} \pi_{\sigma(v)}$ which is an increasing function. For the signal $s_0 = \sigma\left(0\right)$, since the lending cutoff satisfies $\mathbf{v}_{s_0} \geq v_4$, and we know that $\lim_{v \to v_4^+} \sigma\left(v\right) = v_4$, the message function must be continuous at v_4.

I.A.1.13 Proof of Proposition 4.3

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

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

where $\gamma(v)$ is to be solved, and satisfies $\gamma(v_b) = v_a$, $\gamma(\bar{v}) = \mathbf{v}^*$. For any $v \in (v_b, \bar{v})$, the signal realization is $\sigma(v) = \gamma(v)$, and $\sigma(v) = \gamma_s$ 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

$$\mu\gamma(v) = [\mu + (1-\mu) K(\mathbf{c}_s)] I, \tag{I.A.1.12}$$

where $\mathbf{c}_s$ satisfies

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

Then the equilibrium condition (I.A.1.12) 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

$$\frac{c_{v_a}}{b} = \frac{G(v_b) - G(\mathbf{v}^*)}{G(v_b) - G(\mathbf{v}^*) + G(v_a)}. \tag{I.A.1.13}$$

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

$$\frac{c_{v_a^+}}{b} = \frac{g(v_b)}{g(v_a)\gamma'(v_b) + g(v_b)}. \tag{I.A.1.14}$$

$c_{v_a^+} = c_{v_a}$ because of the continuity, then condition (I.A.1.13) and (I.A.1.14) imply that $\frac{G(v_b) - G(\mathbf{v}^*)}{G(v_a)} = \frac{g(v_b)}{g(v_a)\gamma'(v_b)}$.

I.A.1.14 Proof of Proposition 4.4

Under monotone disclosure policies, the support of the posterior belief must be either a single point or a non-degenerate interval. If all signal realizations induce non-degenerate intervals, then the signal distribution must be discrete. Suppose instead that there exists a set of signal realizations with positive measure for which the posterior beliefs are point masses. This cannot be optimal, since the lender obtains zero payoff under each of these signals, which is the lowest possible payoff. Therefore, such a policy cannot be optimal.

I.A.1.15 Proof of Proposition 4.5

Suppose we have two distinct signals, s_1 and s_2, corresponding to two adjacent intervals, $[v_L, v_M]$ and $(v_M, v_R]$.[15] We aim to show that pooling these two intervals into a single signal improves the lender's payoff. First, consider the case where $v_M \leq I$. In this case, pooling the two intervals must improve the lender's payoff, since the expected payoff from the first interval $[v_L, v_M]$ is zero. Therefore, we just need to focus on the case where $v_M > I$. Let the equilibrium cutoffs under the two separate signals be $(\mathbf{c}_1, \mathbf{v}_1)$ and $(\mathbf{c}_2, \mathbf{v}_2)$, and let $(\mathbf{c}_0, \mathbf{v}_0)$ denote the equilibrium cutoff when the two intervals are pooled.

Define x_0 to be the solution to

$$\int_{v_L}^{x_0} dG(v) = \int_{v_M}^{v_2} dG(v).$$

Now consider the following alternative disclosure policy over the interval $[v_L, v_R]$: disclose whether the true value of v lies in $[v_L, x_0] \cup [v_2, v_R]$ or in (x_0, v_2). The proof proceeds in two steps.

Step 1: We show that the disclosure policy that reveals whether the true value of v lies in the interval $[v_L, x_0] \cup [v_2, v_R]$ or in (x_0, v_2) yields a higher payoff for the lender than the original policy that discloses whether v is in $[v_L, v_M]$ or in $(v_M, v_R]$. Since

$$\int_{v_L}^{x_0} dG(v) = \int_{v_M}^{v_2} dG(v),$$

it follows that the posterior beliefs associated with the intervals $[v_L, x_0] \cup [v_2, v_R]$ and $(v_M, v_R]$ yield the same equilibrium behavior. Moreover, the probabilities assigned to these intervals are also identical. Therefore, the lender's payoffs under the two disclosure policies for these corresponding signals are the same. Next, we show that the payoff from the signal associated with the posterior belief (x_0, v_2) is strictly higher than that from the signal corresponding to $[v_L, v_M]$. The following lemma establishes this result.

Lemma I.A.1.3. *Let $M > 0$ be a constant. Let $[y, v_4(y)]$ be an interval such that $G(v_4) - G(y) = M$, and let $W(y)$ denote the lender's expected payoff when the value of v is drawn from this interval. Then $W(y)$ is increasing in y if the density function $g(\cdot)$ is weakly decreasing.*

Proof. Let the equilibrium cutoff under the posterior belief $[y, v_4]$ be $(\mathbf{c}_y, \mathbf{v}_y)$. Then $W(y) = \int_{v_y}^{v_4} (v - \mathbf{v}_y)\, dG(v)$. Besides, we have the constraint $G(v_4) - G(y) = M$. Equilibrium conditions imply

$$\frac{G(v_4) - G(\mathbf{v}_y)}{M} = \frac{\mathbf{c}_y}{b} \Longrightarrow \frac{g(v_4)\, dv_4 - g(\mathbf{v}_y)\, dv_y}{M} = \frac{d\mathbf{c}_y}{b}$$

[15] Whether these intervals are open or closed does not affect the result.

and

$$K(\mathbf{c}_y) = \frac{\mu}{1-\mu}\left(\frac{\mathbf{v}_y}{I} - 1\right) \Longrightarrow K'(\mathbf{c}_y)\, d\mathbf{c}_y = \frac{\mu}{1-\mu}\frac{1}{I} d\mathbf{v}_y.$$

These two conditions imply

$$\frac{g(v_4)\, dv_4 - g(\mathbf{v}_y)\, d\mathbf{v}_y}{M} = \frac{1}{b}\frac{\mu}{1-\mu}\frac{1}{I}\frac{d\mathbf{v}_y}{K'(\mathbf{c}_y)} \Longleftrightarrow g(v_4)\, dv_4 = \left[g(\mathbf{v}_y) + \frac{\mu M}{b(1-\mu) I}\frac{1}{K'(\mathbf{c}_y)}\right] d\mathbf{v}_y.$$

Then

$$\begin{aligned} dW &= (v_4 - \mathbf{v}_y)\, g(v_4)\, dv_4 - d\mathbf{v}_y \left[G(v_4) - G(\mathbf{v}_y)\right] \\ &= (v_4 - \mathbf{v}_y)\, g(v_4)\, dv_4 - \frac{g(v_4)\, dv_4}{g(\mathbf{v}_y) + \frac{\mu M}{b(1-\mu)I}\frac{1}{K'(\mathbf{c}_y)}}\left[G(v_4) - G(\mathbf{v}_y)\right]. \end{aligned}$$

Then

$$\begin{aligned} & \frac{dW}{dv_4} \geq 0 \\ \Longleftrightarrow\ & (v_4 - \mathbf{v}_y) \geq \frac{G(v_4) - G(\mathbf{v}_y)}{g(\mathbf{v}_y) + \frac{\mu M}{b(1-\mu)I}\frac{1}{K'(\mathbf{c}_y)}} \\ \Longleftrightarrow\ & (v_4 - \mathbf{v}_y)\left[g(\mathbf{v}_y) + \frac{\mu}{b(1-\mu)I}\frac{G(v_4) - G(y)}{K'(\mathbf{c}_y)}\right] \geq G(v_4) - G(\mathbf{v}_y) \\ \Longleftrightarrow\ & g(\mathbf{v}_y) + \frac{\mu}{b(1-\mu)I}\frac{G(v_4) - G(y)}{K'(\mathbf{c}_y)} \geq \frac{G(v_4) - G(\mathbf{v}_y)}{v_4 - \mathbf{v}_y}. \end{aligned}$$

Since g is a decreasing function, we know $g(v_y) \geq \frac{G(v_4) - G(\mathbf{v}_y)}{v_4 - \mathbf{v}_y}$, and thus the above condition must hold. □

Step 2: We now show that the lender's payoff is higher under the pooled disclosure policy that reveals whether $v \in [v_L, v_R]$ (i.e., no disclosure) than under the policy that separates $[v_L, x_0] \cup [v_2, v_R]$ from (x_0, v_2). To establish this result, we introduce the following lemma.

Lemma I.A.1.4. *Consider a disclosure policy that reveals whether the value of v lies in the interval $[v_L, x] \cup [v_2, v_R]$ or in (x, v_2), where $x_0 \leq x < v_2$. Suppose the equilibrium corresponding to $[v_L, x] \cup [v_2, v_R]$ is characterized by $(\mathbf{c}_{x2}, \mathbf{v}_{x2})$, and the equilibrium corresponding to (x, v_2) is characterized by $(\mathbf{c}_{x1}, \mathbf{v}_{x1})$, with $\mathbf{v}_{x1} < \mathbf{v}_{x2} \leq v_2$. Let W denote the lender's total payoff from the entire interval $[v_L, v_R]$. Then, if the function $cK(c)$ is weakly convex, we have $\frac{dW}{dx} > 0$.*

Proof. The lender's payoff is

$$W = \int_{\mathbf{v}_{x1}}^{v_2} (v - \mathbf{v}_{x1})\, dG(v) + \int_{v_2}^{v_R} (v - \mathbf{v}_{x2})\, dG(v).$$

Then
$$\frac{dW}{dx}=-\frac{d\mathbf{v}_{x1}}{dx}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right]-\frac{d\mathbf{v}_{x2}}{dx}\left[G\left(v_R\right)-G\left(v_2\right)\right],$$
and thus
$$\frac{dW}{dx}\geq 0\Longleftrightarrow-\frac{d\mathbf{v}_{x2}}{dx}\left[1-G\left(v_2\right)\right]\geq\frac{d\mathbf{v}_{x1}}{dx}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right].$$
The equilibrium conditions imply
$$\frac{1-G\left(v_2\right)}{1-G\left(v_2\right)+G\left(x\right)}=\frac{\mathbf{c}_{x2}}{b}\Longleftrightarrow\frac{G\left(x\right)}{1-G\left(v_2\right)}+1=\frac{b}{\mathbf{c}_{x2}}\Longrightarrow\frac{g\left(x\right)dx}{1-G\left(v_2\right)}=-\frac{b}{\mathbf{c}_{x2}^2}d\mathbf{c}_{x2}.$$

$$\begin{aligned}&\frac{G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)}{G\left(v_2\right)-G\left(x\right)}=\frac{\mathbf{c}_{x1}}{b}\Longleftrightarrow\frac{G\left(\mathbf{v}_{x1}\right)-G\left(x\right)}{G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)}+1=\frac{b}{\mathbf{c}_{x1}}\\ \Longrightarrow&\frac{-g\left(x\right)dx}{G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)}+\frac{\left[G\left(v_2\right)-G\left(x\right)\right]g\left(\mathbf{v}_{x1}\right)d\mathbf{v}_{x1}}{\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right]^2}=-\frac{b}{\mathbf{c}_{x1}^2}d\mathbf{c}_{x1}.\end{aligned}$$

$$K\left(\mathbf{c}_{x1}\right)=\frac{\mu}{1-\mu}\left[\frac{\mathbf{v}_{x1}}{I}-1\right]\Longrightarrow K'\left(\mathbf{c}_{x1}\right)d\mathbf{c}_{x1}=\frac{\mu}{1-\mu}\frac{\mathbf{v}_{x1}}{I}.$$
$$K\left(\mathbf{c}_{x2}\right)=\frac{\mu}{1-\mu}\left[\frac{\mathbf{v}_{x2}}{I}-1\right]\Longrightarrow K'\left(\mathbf{c}_{x2}\right)d\mathbf{c}_{x2}=\frac{\mu}{1-\mu}\frac{d\mathbf{v}_{x2}}{I}.$$
When $dx>0$, we must have $d\mathbf{v}_{x1}>0$, then
$$\frac{g\left(x\right)dx}{G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)}>\frac{b}{\mathbf{c}_{x1}^2}d\mathbf{c}_{x1}.$$

This implies
$$\begin{aligned}&\frac{-1}{\mathbf{c}_{x2}^2}d\mathbf{c}_{x2}\left[1-G\left(v_2\right)\right]>\frac{b}{\mathbf{c}_{x1}^2}d\mathbf{c}_{x1}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right]\\ &\frac{-1}{\mathbf{c}_{x2}^2}\frac{\mu}{1-\mu}\frac{d\mathbf{v}_{x2}}{IK'\left(\mathbf{c}_{x2}\right)}\left[1-G\left(v_2\right)\right]>\frac{1}{\mathbf{c}_{x1}^2}\frac{\mu}{1-\mu}\frac{d\mathbf{v}_{x1}}{IK'\left(\mathbf{c}_{x1}\right)}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right]\\ &\frac{1}{\mathbf{c}_{x2}^2}\frac{-d\mathbf{v}_{x2}}{K'\left(\mathbf{c}_{x2}\right)}\left[1-G\left(v_2\right)\right]>\frac{1}{\mathbf{c}_{x1}^2}\frac{d\mathbf{v}_{x1}}{K'\left(\mathbf{c}_{x1}\right)}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right].\end{aligned}$$

Then if $K'\left(\mathbf{c}_{x2}\right)\mathbf{c}_{x2}^2\geq K'\left(\mathbf{c}_{x1}\right)\mathbf{c}_{x1}^2$, we must have $-\frac{d\mathbf{v}_{x2}}{dx}\left[1-G\left(v_2\right)\right]\geq\frac{d\mathbf{v}_{x1}}{dx}\left[G\left(v_2\right)-G\left(\mathbf{v}_{x1}\right)\right]$ which implies $\frac{dW}{dx}>0$. When function $cK\left(c\right)$ is convex, we have
$$2K'\left(c\right)+cK''\left(c\right)\geq 0\Longrightarrow 2cK'\left(c\right)+c^2K''\left(c\right)\geq 0\Longrightarrow\left(K'\left(c\right)c^2\right)'\geq 0.$$

Since $\mathbf{c}_{x2}>\mathbf{c}_{x1}$, this implies $K'\left(\mathbf{c}_{x2}\right)\mathbf{c}_{x2}^2\geq K'\left(\mathbf{c}_{x1}\right)\mathbf{c}_{x1}^2$. □

Lemma I.A.1.4 implies that the lender can increase her payoff by raising the cutoff x, up to the point where $\mathbf{v}_{x1} = \mathbf{v}_{x2}$, which corresponds to the case of pooling the two intervals.

I.A.1.16 Proof of Proposition 5.1

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

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

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

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

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

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

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

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

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

and

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

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

$$p_1 c_1 + p_2 c_2 = b(1 - G(v_c)) = c_c.$$

Consider the following policy with a binary signal space $\mathcal{S} = \{s_1, s_2\}$. s_1 is revealed if the true state is in the set $[0, v_1] \cup [v_c, v_2]$, and s_2 is revealed if the true state is in the set $(v_1, v_c) \cup (v_2, \bar{v}]$. 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 Prob$(s_1) = p_1$ and Prob$(s_2) = p_2$. Under signal s_1, in equilibrium, the lender approves the loan if and only if $v \in [v_c, v_2]$, and under signal s_2, the lender approves the loan if and only if $v \in (v_2, \bar{v}]$. The lender's expected payoff under this alternative disclosure policy is

$$W_1 = \int_{v_c}^{\bar{v}} [\mu v - \mu I] \, dG(v) - \frac{(1-\mu) I}{b} [p_1 K(c_1) c_1 + p_2 K(c_2) c_2] .$$

Since $p_1 c_1 + p_2 c_2 = c_c$, and function $xK(x)$ 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.