Algorithmic Lending, Competition, and Strategic Information Disclosure

Qiaochu Wang Yan Huang Param Vir Singh*

Abstract

Machine learning (ML) algorithms used by financial lenders in their screening processes are hidden from the consumers who are affected by their decisions leading many consumers to make sub-optimal decisions when seeking credit. Despite increasing calls for greater transparency, only a few lenders provide personalized approval odds to consumers (e.g. via financial intermediaries like Credit Karma or pre-approval tools). We investigate how *competition* among algorithmic lenders affects their decisions to provide approval odds to consumers. We show that competitive pressures between lenders can undermine the disclosure incentives. Lenders use asymmetric disclosure of approval odds strategically to *soften* the competition when their algorithms are fairly accurate. The asymmetric disclosure of approval odds endogenously creates product differentiation and allows lenders to focus on different segments of consumers softening the competition on the interest rates. We find that consumer surplus is highest when both lenders provide approval odds and lowest when neither provides approval odds. However, our analysis also shows that any policy that mandates all lenders to provide personalized approval odds to consumers may not necessarily improve consumer surplus.

Keywords: Algorithmic transparency, competition, fintech, machine learning, financial intermediary, game theory.

1 Introduction

Financial lenders ("lenders" hereafter; e.g., banks) acquire proprietary information about borrowers and use machine learning (ML) algorithms in screening processes to predict their credit risk (Citron and Pasquale, 2014). These ML algorithms are usually hidden from the borrowers who are affected by the algorithms' decisions. Hence, potential borrowers face considerable uncertainty about their odds of approval when they seek a financial product (Experian, 2020a) [1]. Every lender has its own proprietary algorithm. As a result, a borrower may receive different decisions from different

*All authors are at Carnegie Mellon University.

[1]Consumers who are seeking credit are *potential* borrowers. Consumers who have received credit are borrowers. With some abuse of notation, we refer to potential borrowers as borrowers in the rest of the paper to be concise

lenders for comparable financial products.[2] Further, when a lender pulls a credit report to calculate approval odds for a borrower, it appears as a hard inquiry on the borrower's credit report, which can reduce the borrower's credit score (Experian, 2020b).

Financial intermediaries, such as Credit Karma, Quizzle, Credit Sesame, NerdWallet, and Wallethub, have emerged to help borrowers make better financial decisions when facing opaque algorithms by lenders (Andriotis, 2014).[3] These intermediaries give borrowers free access to their credit reports and use the credit information to serve them advertisements for credit cards, loans and other financial products (Popper and Merced, 2020). They earn a fee when a borrower purchases an advertised financial product. The financial intermediaries reverse-engineer lenders' screening algorithms using data they have on the borrowers who were approved in the past for the same product, and provide personalized odds of approval for a financial product to borrowers (Lockert, 2020). The estimated approval odds that intermediaries provide can reduce borrower uncertainty and help them more efficiently choose which lender to apply to.

Financial lenders can further reduce borrower uncertainty by offering pre-approval tools or by revealing their algorithm to intermediaries so that they can report accurate personalized approval odds to borrowers.[4] The resulting uncertainty reduction for borrowers can lead to *market expansion* for the revealing lenders. At the same time, the revealing lenders' algorithms are protected against gaming by the borrowers as only the outcome of the algorithms, not the algorithms themselves, is shared with borrowers. The provision of personalized approval odds to the borrowers by the lenders is akin to "partial algorithmic transparency".

Despite obvious advantages of the partial algorithmic transparency discussed above and increasing calls for algorithmic transparency in general from different sections of the society (Diakopoulos, 2016, Fu et al., 2020, Wang et al., 2020, Pasquale, 2015, Kroll et al., 2017), only two of the five largest lenders have chosen to share their secret screening algorithms with a financial intermedi-

[2] Differences in lenders' screening algorithms also lead to the "Winner's curse" that has been widely documented in the banking literature: If credit screening is imperfectly correlated across lenders and lenders are unaware of whether a borrower has been rejected by other lenders, borrowers rejected by one lender can apply to another lender. This systematically worsens the pool of applicants faced by all lenders (Broecker, 1990b).

[3] Credit Karma is the largest of these financial intermediaries. It provides free credit reports from TransUnion and Equifax to its more than 100 million registered users and allows them to shop for credit cards, loans and other financial products (Lunden, 2020).

[4] With pre-approval tools the lenders can directly provide approval odds to potential borrowers reducing their uncertainty. However, the pre-approval tools that the lenders can provide suffer from some limitations. In the pre-approval process, borrowers give the lender the authorization to conduct a soft inquiry on their credit report. Soft inquiries do not affect a borrower's credit score negatively. However, a soft inquiry does not return as detailed a credit information to the lender as a hard inquiry does. As a result, being pre-approved is an informative indicator but does not guarantee the approval of borrowers' formal application to the lender. Intermediaries like Credit Karma, on the other hand, have access to the full credit report of borrowers. In many cases, the intermediary and the lender may use credit reports from different agencies which may lead to some inaccuracies in the intermediary provided approval odds even when the lender reveals its screening algorithm to the intermediary. Throughout the paper, for simplicity, and without loss of generality, we assume that if the lender reveals its algorithm to the intermediary, the intermediary provided approval odds to the borrowers for this lender will be accurate.

ary that reports approval odds for financial products (Rudegeair and Andriotis, 2018). Relatedly, through a survey of several lenders' websites, we find a similar asymmetry in lenders' offering of pre-approval tools for their credit cards. We find that lenders do not offer pre-approval tools for their credit cards at all times, and at any time only a few offer those tools.[5] These observations are in contrast to the predictions of the "unraveling" theory (Milgrom, 1981, Grossman, 1981), which argues that a privately informed seller will voluntarily disclose all information of its product when such disclosure is costless and credible. In this study, we provide one potential reason in the algorithmic lending context that prevents the unraveling results from happening. We show that *competition* may prohibit some lenders from providing partial algorithmic transparency.

We investigate competition among lenders on intermediary platforms like Credit Karma that provide personalized approval odds for financial products to borrowers. We specifically focus on strategic algorithm revelation by lenders to the intermediary. We answer why asymmetric revealing of the algorithm to the intermediary by symmetric lenders is an equilibrium outcome. We further examine how the accuracy of the lenders' algorithms, the accuracy of the intermediary's reverse-engineered algorithm, and the riskiness of the market affect the lenders' decisions to reveal their algorithms. Finally, we investigate how a policy that mandates algorithm sharing with the intermediary affects borrower surplus?

To our knowledge, the strategic considerations in and the consequences of lenders' decisions to reveal screening algorithms to the intermediary or approval odds directly to borrowers has not been studied yet. The existing research has largely focused on either the consequences of sharing information related to product features with consumers in the strategic information sharing literature (e.g. Guo and Zhao (2009), Kuksov and Lin (2010)), or the consequence of acquiring or sharing borrowers' information with other lenders in the lender competition literature (e.g. Hauswald and Marquez (2003, 2006)). Our study examines a new type of information that can be shared to reduce borrower uncertainty, and discuss implications of such revealing behavior for competition, social welfare, and public policy. Hence, our paper provides new insights on strategic information sharing in the context of algorithmic lending under competition.

Competition introduces key dynamics that may affect a lender's decision to reveal its algorithm. Let us first illustrate how competition plays out in an algorithmic lending scenario with an example. First, consider a monopoly lender, and two types of borrowers, H (low default probability) and L (high default probability). The lender's algorithm predicts a borrower's type with 70% accuracy. In other words, the lender's algorithm will mistakenly predict a H type borrower as L type or predict a L type borrower as H type with a 30% probability. Under this algorithm, when the lender sees a

[5] For example, on October 19, 2021, Citi Bank, Chase and PNC were not offering the pre-approval tools for their credit cards whereas American Express, Capital One and Discover Card were.

borrower with a predicted type H, it knows the probability that this borrower's true type is H is 70% (i.e., $Pr(Type = H|Predict = H) = 0.7$). In this monopoly case, whether the lender reveals the algorithm to the intermediary (i.e. the approval odds to the borrowers) or not, it will not influence the lender's posterior belief (after observing the borrower's application behavior) about the borrower's type. However, this is no longer the case in a competitive environment.

Let us now consider a duopoly setting, where two competing lenders (i and j) are evaluating borrowers using their own algorithms. Assume that their algorithms are independent but have the same accuracy. If these algorithms are kept secret and borrowers randomly apply to one of the two lenders, the posterior probability $Pr(Type = H|Predict = H)$ as defined above is still 0.7 for both lenders. However, if a borrower knows her expected predicted type before applying to a lender (for example, through the approval odds provided by the intermediary or through pre-approval tools), she will be able to tell which lender is more likely to approve her application and would choose to apply to that lender. From lender i's perspective, the probability of a borrower that its own algorithm predicts as type H being an actual H type is no longer 0.7. The fact that this borrower applies to lender i but not lender j indicates that she may have gotten a worse predicted outcome from lender j and this fact contains useful information about the borrower's true type. The posterior probability of the borrower being of H type is in fact 0.65.[6] This posterior probability is lower than 0.7 because the lender is able to incorporate an *informative signal* from the competitor.

In competition, the borrowers can be divided into many segments based on the predictions of the two lenders. For some borrowers the predictions of type by the lenders will be same, whereas for others they may differ. Borrowers who are classified as H type by both the lenders appear less risky to the lenders and those who are classified as H type only by one lender appear riskier in comparison. However, to attract the borrowers who are classified as H type by both the lenders, the lenders have to intensely compete on the interest rate. In contrast, the riskier segment which only the focal lender classifies as H type can only be approved by the focal lender. As a result, the focal lender does not face competition for this segment of borrowers. However, the focal lender does not observe the predictions of the competing lender. Hence, it cannot differentiate whether an applicant that it classifies as H type, gets classified as H type or L type by the competing lender. As a result, borrowers in both the segments are offered the same interest rate. However, the lender can calculate the segment sizes as a well as their risks in expectation. The relative segment sizes and their riskiness determine the intensity of the competition. As the size (risk) of the segment that corresponds to borrowers who are classifies as H type by both the lenders increases (decreases), the

[6] The posterior probability of the borrower being of H type can be calculated as: $Pr(Type = H|Predict_i = H, Apply = i) = \frac{Pr(Apply=i|Type=H,Predict_i=H)Pr(Type=H)}{Pr(Apply=i)} = \frac{Pr(Predict_i \preceq Predict_j|Type=H,Predict_i=H)Pr(Type=H)}{Pr(Apply=i)} = \frac{(0.5\times 0.7+(1-0.7)\times 0.5)}{0.5} = 0.65$, where $Predict_i(Predict_j)$ denotes the prediction that the borrower receives from lender $i(j)$.

intensity of competition increases. In contrast, as the size (risk) of the segment that corresponds to borrowers who are classifies as H type by only one lender increases (decreases), it softens the competition. The revelation of the algorithms by lenders would affect the intensity of competition that lenders face via affecting the sizes and risk of the different borrower segments.

Formally, we analyze a multi-stage game within a duopoly of symmetric lenders who use ML algorithms to approve or reject borrowers who have apply for their financial products. The borrowers are modeled as H type and L type. The H type borrowers have a lower default rate than the L type borrowers. The borrowers decide which lender to apply to or they can choose not to apply to either. The two lenders are symmetric in their algorithms' accuracy and offer financial products that are identical in terms of non-price features (e.g. loan amount, credit line, and credit limit, etc.). The lenders decide (1) whether to reveal their algorithm to the intermediary or not and (2) the interest rate that they will charge for their financial product. We model the intermediary as non-strategic. The intermediary shows the financial products of both the lenders to the borrowers. The lenders pay the intermediary a fixed fee for any borrower who applies for the advertised product and is approved by the lender. The fee that the intermediary charges to the lenders is assumed to be exogenous. We solve for lenders' and borrowers' surplus in three scenarios – (i) both lenders reveal the algorithm to the intermediary, (ii) neither lender reveals the algorithm to the intermediary, and (iii) only one lender reveals the algorithm to the intermediary. We use the sub-game perfect Nash equilibrium (SPNE) as our solution concept.

Algorithmic lenders have to consider two key forces – *market expansion* and *competition* – while deciding whether to reveal their algorithm to the intermediary or not. The following points explain the intuition behind the market expansion and competition intensity effects in the scenarios (i)-(iii) defined above. Following the example above, we refer to the lenders as i and j where lenders only want to approve H type borrowers. For ease of understanding, we define "common" and "captive" segments. For lender i, the borrowers who both lender i's algorithm and lender j's algorithm (or the intermediary's reverse-engineered algorithm for lender j if lender j chooses not to reveal its algorithm) classify as H type constitute the common segment. In contrast, for lender i, the borrowers who lender i's algorithm classifies as H type but lender j's algorithm (or the intermediary's reverse-engineered algorithm for lender j if lender j chooses not to reveal its algorithm) classifies as L type constitute lender i's captive segment. Further, we use *profitability* of a segment to capture the size and risk of a borrower segment. A segment becomes more profitable if its size increases or its risk reduces or both. We find that in the three scenarios discussed next, an interest rate equilibrium in pure strategies does not exist; there exists an interest rate equilibrium in mixed strategies.

- If both lenders choose to reveal, the borrowers face no uncertainty. All the borrowers who would be approved by the lenders apply to either lender in this case. As a result, lenders benefit from the market expansion effect due to the reduced borrower uncertainty. However, the competition between the lenders in this case is the most intense. The signal that lender i receives about an borrower's type from lender j is the most precise in this case, and therefore, borrowers in the common segment are most profitable because their posterior probability of being of H type is the highest in this scenario. Further, borrowers in the captive segment for a lender are less profitable as their probability of being of H type is the lowest in this scenario. At the same time, the common segment is the largest and the captive segment is the smallest in this scenario. As a result, lenders compete aggressively on the interest rate, denoted as b, and set a low interest rate to attract borrowers in the common segment.

- If neither lender chooses to reveal the algorithm to the intermediary, borrowers face the most uncertainty. Borrowers make decisions based on the intermediary's noisy predictions rather than the output of the lenders' algorithms. Consequently, many borrowers make sub-optimal decisions (i.e., applying to a lender that would reject them, or not applying at all when they could have been approved by a lender). The market coverage by the lenders is the lowest in this case. However, the competition faced by the lenders is less intense than in the "both reveal" scenario. The signal that a lender receives from its rival is noisy and as a result is given less weight. Hence, compared to the "both reveal" case, the borrowers in the common segment are less profitable (both in terms of size and risk) in the "neither reveal" case. On the other hand, in the "neither reveal" case, the borrowers in the captive segment are more profitable (both in terms of size and risk) compared to in the "both reveal" case. As a result, lenders have less incentive to reduce interest rate to compete for the borrowers in the common segment and more incentive to set a high interest rate to exploit their captive segments.

- If only one lender (lender i) reveals its algorithm, it can benefit from the market expansion effect. By revealing its algorithm, it removes the noise added by the intermediary's prediction to the approval odds borrowers have access to. Hence, borrowers do not make any sub-optimal decisions regarding whether to apply to lender i. While lender j has to sacrifice the market expansion effect, it receives a more precise signal from lender i. This creates asymmetry in how the two lenders view their common and captive segments. Both lenders prefer their common to their "captive" segments. However, in comparison to lender i, lender j views its common segment to be relatively more profitable and its captive segment relatively less profitable. This consequently affects lenders' strategies in setting interest rate. While lender i would like to avoid intense competition for the common segment and generate more profit from its captive borrowers, lender j focuses more on the common segment and set competitive interest rate to capture these borrowers. As a result, lender i can take advantage of the market expansion effect and extract greater surplus from borrowers it approves, and lender j can capture a large portion of the common segment without having to reduce interest rate to a very low level.

We find that when the lenders' algorithms are accurate, the SPNE is *asymmetric* such that one lender chooses to reveal its algorithm to the intermediary while its competitor chooses not to reveal. Otherwise, both lenders should reveal their algorithm to the intermediary.

The asymmetric equilibrium in algorithm revelation is our most important and least intuitive result. We characterize the conditions needed to sustain an asymmetric SPNE in terms of the accuracy of the lenders' algorithms. One of the lenders always has an incentive to reveal its algorithm because of the market expansion effect. However, the competing lender would reveal its algorithm only when the accuracy of the lenders' algorithms is below a threshold. The relative profitability of the common segment to the captive segment determines the intensity of competition in interest rates. When the lenders' algorithms' accuracy is low, the intensity of competition is not very strong even when both lenders reveal their algorithms. This is because the lenders' algorithms would incorrectly classify many H type borrowers as L type, and vice versa. As a result, the common segment is only marginally profitable compared to the captive segment. At the same time, by revealing their algorithms, both lenders benefit from the market expansion effect. Hence, both lenders will choose to reveal their algorithms when their algorithms' accuracy is low. In contrast, an asymmetric equilibrium is observed when the accuracy of the lenders' algorithms is high. In this scenario, the non-revealing lender has no incentive to deviate from the non-revealing strategy, because if it does so, the resulting competition would be very intense. The negative effect of competition dominates the gains from the market expansion effect for the non-revealing lender. We further find that the threshold in lenders' algorithm accuracy over which asymmetric equilibrium is sustained is moderated by the accuracy of the intermediary's reverse-engineered algorithm and by the riskiness of the market.

Our results reveal a unique connection between lenders' algorithm revealing strategies and the degree of "product differentiation" that arises endogenously from lenders' revealing decisions. We find that lenders can use asymmetric revealing as a strategic tool to differentiate themselves and soften the competition. Asymmetric revealing of algorithm leads to asymmetry in the interest rates for products with identical non-price features which in turn diversify borrowers' preferences. In the asymmetric equilibrium, the revealing lender always gets a higher equilibrium payoff than the non-revealing lender, which implies that when conditions that support asymmetric equilibrium exist, a lender should take the opportunity to reveal its algorithm first.

From a social-welfare perspective, lenders' algorithm revealing behavior increases the efficiency of credit markets because it helps borrowers avoid non-optimal applying decisions. While borrower surplus is the highest when both lenders reveal their algorithms to the intermediary in our model, our analysis suggests regulating lenders to reveal their algorithms to the intermediary (or to provide pre-approval tools) may have some drawbacks. Particularly, lenders have less incentive to invest in their algorithm's accuracy under the "both reveal" case. The reason is that in the "both reveal" case, both lenders focus on the common segment of consumers. Increased algorithm accuracy increases the profitability of the common segment vis a vis captive segment which intensifies competition. In contrast, this is not the case in the asymmetric revealing equilibrium. When policy makers are considering regulations on algorithmic transparency, they should be aware that mandatory transparency may reduce lenders' incentive to invest in algorithmic screening technologies, which in the long term may not help allocate financial resources to more creditworthy borrowers.

The rest of this paper is organized as follows: §2 provides relevant literature and explain our contributions to it, §3 introduces the general setup of our model, §4 contains the bulk of the analysis, §5 discusses an important extension to the main model, and §6 concludes.

2 Contributions to Literature

In this section, we discuss the relevant literature and how our study contributes to it. Credit markets suffer from information asymmetry where lenders (firms) are uncertain about borrowers' (consumers') credit worthiness. A large stream of literature in economics has shown that in the presence of information asymmetry, firms can effectively employ screening techniques to learn more about consumers' type to make better selection of consumers (Stiglitz, 1975). In credit markets, lenders use secret screening algorithms to screen borrowers for their creditworthiness.

The first stream of literature relevant to our study investigates how strategic information acquisition affects the screening abilities of lenders and as a result credit market competition. Hauswald and Marquez (2003) show that investment in screening technologies by a lender would soften the competition in the short term but intensify it in the long term. On the one hand, the advanced screening technologies will give the informed lender an "information advantage" and thus discourage the uninformed lender to compete. On the other hand, the uninformed lender would also be able to observe some public signals from the informed lender without performing screening. Consequently, the informed lender's information advantage would deteriorate and competition would intensify. In a subsequent paper, Hauswald and Marquez (2006) study how lenders adapt their information acquisition strategies to soften the intensity of competition. They find that as competition increases (e.g., the number of lenders in the market increases), lenders respond by focusing more on the specialized screening but not broad-based screening. Doing so helps the lenders "specialize" and capture their captive segment of borrowers softening competition. Other related studies show that lenders can soften the competition and screen less by coordinating (Bouckaert and Degryse, 2004) or through product differentiation (Villas-Boas and Schmidt-Mohr, 1999).

Our paper contributes to the above stream of literature in that we identify asymmetric revealing of algorithms to the intermediary as a possible strategy to soften competition on interest rates. All of the information acquisition and exchange strategies discussed in the literature above directly affect the performance of the screening algorithm. Whereas lenders' strategy of revealing information on algorithmic decisions to borrowers, which is the focus of our study, does not directly affect the performance of the screening algorithm, but could change the intensity of competition through other channels. From the perspective of information flow, most of the above papers only consider information asymmetry in one direction, that is, lenders do not perfectly observe borrowers' credit worthiness. They overlook the information asymmetry that exists in the other direction, that is, borrowers do not know whether they will be approved ex ante. Our paper squarely focuses on the uncertainty faced by borrowers and how lenders' algorithm revealing decisions affect this uncertainty and the competitive structure of the credit market.

Beyond the lending context, our paper is connected to a broader literature on strategic information revealing. This stream of literature discusses firms' strategic decision on whether to reveal information to consumers to reduce consumer uncertainty. The information here usually refers to the features of the products. The first important result from this literature is the *unraveling theory* (Milgrom, 1981, Grossman, 1981), which shows that a privately informed seller will voluntarily disclose all information of its product when such disclosure is costless and credible. As a result, mandated disclosure is redundant. However, this prediction is not consistent with many empirical observations. As a result, several researchers propose explanations for this inconsistency, such as disclosure is usually not costless or consumers are not always aware of such disclosure (Fishman and Hagerty, 2003). Later, more involved theories have been put forward to study the influence of various market structures on firms' information revealing decisions, and many studies argue that competition could be the potential reason that prevents the unraveling results from happening (Board, 2009, Levin et al., 2009).

Our paper is closely related to the papers that study information revelation under competition. In these studies, typically, firms are modeled as vertically or horizontally differentiated in quality (of their products), and consumers are uncertain about the quality of the products. Each firm compares the competition intensity under uncertainty, i.e., consumers making decisions based on their perceived product quality, and under full information, i.e., consumers making decisions based on the actual product quality. Firms make a strategic decision on quality revelation based on whether such revelation would intensify the competition or soften it (Board, 2009, Levin et al., 2009, Kuksov and Lin, 2010, Gu and Xie, 2013). Our paper is similar to this stream of literature in that we also point out that competition may hinder the information revealing and show that it could lead to asymmetric revealing decisions even when the revealing cost is zero. However, our paper is differentiated in at least three important aspects: (1) while these papers study firms' decisions on revealing product quality, our paper studies the firms' decisions on revealing another dimension of information to reduce buyers' uncertainty, and thus expand the strategic information revealing to a boarder context which involves information asymmetry. (2) The asymmetric revealing equilibrium found in these papers is a result of differentiated firms. In contrast, our paper find asymmetric revealing equilibrium even when the two firms are symmetric and provide products with identical non-price features. Our paper identifies several new insights and a novel mechanism that could help explain the observed asymmetric revealing behavior by symmetric firms.

Our paper is also related to the literature on firms' strategies in differentiating products when facing heterogeneous consumers (See (Tirole and Jean, 1988) for a compete review). We extend this literature to the field which involves information asymmetry. Although in our model, we assume the products to be identical, and thus borrowers (consumers) should derive the same utility ex post from both products if the prices are the same, borrowers do have a preference between these identical products ex ante because they believe they have a greater chance of getting approved for one product than the other. We show that asymmetric revealing of the algorithm will play a strategic role in differentiating products and diversifying borrowers' preferences. By revealing the algorithm, a lender becomes more profitable to the borrowers for whom the algorithm predicts a positive approval outcome. As a result, the lender can extract greater surplus from them by charging a higher interest rate. At the same time, the competing lender has to keep a low interest rate to stay competitive in the market. Hence, endogenously, asymmetric algorithm revelation by ex ante symmetric lenders leads to differentiated products and softens competition.

Finally, our work is related to the emerging stream of literature on algorithmic transparency. This stream of literature has primarily considered the problem of designing optimal classification algorithms when facing strategic users who may manipulate the input to the system at a cost (e.g. (Hardt et al., 2016, Haghtalab et al., 2020, Kleinberg and Raghavan, 2019, Meir et al., 2012). However, Wang et al. (2020) argue that a firm can strategically leverage gaming by users to its benefit if it were to make its algorithm transparent. Our work is related to this stream of literature in that algorithm revelation by the lenders is akin to partial algorithmic transparency. However, we differ from the literature in that partial algorithmic transparency does not make the lender's algorithm susceptible to gaming by users. More importantly, none of the paper on algorithmic transparency have considered the effect of competition on a firm's decision to reveal its algorithm. Our paper squarely focuses on the competition effects in the context of algorithmic lending and transparency.

3 Model

We consider a duopoly credit market of two competing lenders that sell financial products to borrowers on a financial intermediary's platform. We begin by describing borrowers and lenders; then we explain the sequence of decisions.

3.1 Borrowers

We model two types of borrowers, high-quality (non-defaulter) and low-quality (defaulter), denoted as H and L, respectively. We use θ to denote the portion of H type borrowers in the market.[7] For simplicity, following a relevant stream of literature in Finance and Economics, we assume that a borrower's type is unknown to both herself and the lender before she enters into a credit relationship (Sharpe, 1990, Hauswald and Marquez, 2003, Ruckes, 2004).[8] This could be the case if say, a borrower and her lender gradually gather new information about her ability to manage her finances. However, the portion of H type borrowers, θ, is common knowledge. Borrowers are considering to apply for a financial product (e.g., credit card or loan). We assume that the amount of the loan is fixed, and we normalize it to 1. A borrower's utility from applying conditional on her type and the approval outcome is as follows:

$$U = \begin{cases} M_h - m - b & \text{if she is } H \text{ type and gets approved} \\ M_l - m - b & \text{if she is } L \text{ type and gets approved} \\ -m, & \text{if she gets rejected} \end{cases}$$

where $M_h(M_l)$[9] is the overall benefit that a $H(L)$ type borrower can get from the financial product (e.g., the monetary value and/or convenience brought by a credit card). b is the price of the financial product (e.g., the interest rate of the loan or the fee of the credit card), and $-m$ is the cost of applying (e.g., a hard inquiry on credit record and time/effort spent on the application process). The utility a borrower will receive if she does not apply is normalized to zero. We assume borrowers will apply to at most one of the lenders in the market.[10] Note that when a borrower makes the

[7] We model borrowers as either defaulters or non-defaulters, but our model can be easily generalized to the case where the H type borrowers default with a probability of ϵ_1 while the L type default with a probability of $1 - \epsilon_2$, where $\epsilon_1 < 0.5$ and $\epsilon_2 < 0.5$. The results of the paper still hold.

[8] This is not a critical assumption since our results hold even when we relax this assumption. We use this assumption to keep our model clean and easy to follow. There are other papers in Finance and Economics that assume that borrowers know their types. However, they assume away any self-selection devices and thus different types of borrowers' applying behavior is identical(Broecker, 1990a, Shaffer, 1998).

[9] In many other models, M_h is assumed to be larger than M_l, since L type borrowers have to incorporate the dis-utility associated with the reputation loss once they default. However, this assumption is not necessary in our setting.

[10] We can think of this as that the marginal benefit of getting a second financial product decreases sharply so that borrowers will not apply to multiple lenders simultaneously.

application decision, she does not know her type or whether she will be approved. Instead, she infers both her type and her probability of getting approved by lenders from the predictions she gets from the intermediary. Based on this information, she computes the expected utility she can receive from applying to each lender,[11] and applies to the lender that provides a higher expected utility (to be elaborated in Section 4.3). To facilitate discussion, hereafter, we will use personal loan as an example of the financial product.

3.2 Lenders

There are two symmetric lenders in the market, providing homogeneous loans to borrowers. That is, the loans have identical non-price features but may come with different interest rates. Each lender has its own screening algorithm. We assume that the algorithms of the two lenders are independent but have the same level of accuracy, and the algorithms are equally accurate in predicting the positive and negative cases (i.e., the true positive rate equals the true negative rate). Mathematically, the accuracy of the lenders' algorithms is characterized by $P_b \in (\frac{1}{2}, 1]$[12]: the algorithm predicts a H type applicant as H type with probability P_b, and a L type applicant as L type with probability P_b.

The value of a non-defaulter to a lender equals the interest rate set by the lender: b, and the value of a defaulter to a lender equals the negative of the loan amount, which has been normalized to -1. The utility that a lender receives if it approves n_H H type applicants and n_L L type applicants is given by

$$\Pi = n_H b - n_L \tag{1}$$

3.3 Intermediary

The intermediary helps borrowers evaluate their chances of getting approved. If a lender's algorithm is hidden from the intermediary, the intermediary provides the predicted odds based on its' own reverse-engineered algorithm. The accuracy of the intermediary's algorithm is characterized by

[11] Specifically, if a borrower knows that she will be approved by a lender with probability p_a and is a H type borrower with probability p_h, her expected utility from applying to the lender is $E(U) = p_a(p_h(M_h - m - b) + (1 - p_h)(M_l - m - b)) - (1 - p_a)m$. Here she infers both p_a and p_h from the predictions she gets from the intermediary and her knowledge about the lenders' revealing decisions (to be elaborated in Section 3.4). For example, if both lenders reveal their algorithms, and the intermediary tells her that she will be approved by both lenders (which is equivalent to both lenders' algorithms predicting that she is of H type and thus approving her application, since both lenders reveal their algorithms to the intermediary), she will infer that $p_a = 1$ and $p_h = \frac{P_b^2\theta}{P_b^2\theta + (1-P_b)^2(1-\theta)}$ (according to Bayes' rule), where P_b is the accuracy of the lenders' algorithms, which will be formally defined in Section 3.2.

[12] Hereafter, we use subscript b or B to denote all the lender side parameters since we use 'Bank' as an example for the lender.

parameter P_c[13]: The intermediary makes the same prediction as the lender with probability P_c.[14] Additionally, for simplicity, we assume that the intermediary has the same accuracy in predicting the outcomes of the algorithms of the two lenders. The intermediary's prediction is assumed to be independent of borrowers' true type given the lender's prediction: the intermediary is only interested in lenders' prediction but not borrowers' true type.[15] If the lender reveals its algorithm to the intermediary, the intermediary provides exactly the same prediction as the lender's algorithm. Whenever a borrower who applied to a lender through the platform is approved, the intermediary gets a fixed commission fee c from the lender. In this paper, we set $c = 0$ for the ease of discussion.[16]

We model the intermediary as non-strategic. It simply predicts (imperfectly) lenders' approving decisions and reports it to borrowers. Borrowers make applying decisions based on the intermediary's predictions but not the lender's predictions, since the latter are not observed by borrowers at the time of applying.

Our model can also be interpreted for a scenario where there is no intermediary. The intermediary described above provides approval odds that we consider here would represent the prior belief of the borrowers. In the absence of the intermediary, this belief may come from the information gathered from online forums and past experience or by communicating with other borrowers. The belief is positively correlated with the true outcome but is not perfectly accurate. Instead of revealing the algorithm to the intermediary, without the presence of the intermediary, lenders can provide information to borrowers through pre-approval tools.

3.4 Game Sequence

The sequence of stages in the game between the two lenders and the borrowers is as follows.

- **STAGE 1: Algorithm Revealing Stage** Each lender chooses whether or not to reveal its algorithm to the intermediary simultaneously. Each lender's decision is observed by the other lender but not observed by the borrowers.[17]

[13]Hereafter, we use subscript c or C to denote all the intermediary side parameters since we use 'Credit Karma' as an example for the intermediary.

[14]Note that P_c is not the accuracy of the intermediary's algorithm in predicting borrowers' type.

[15]Intermediaries like Credit Karma make money when borrowers who apply for a financial product through the intermediary website get approved by the lender. Lenders make approval decisions based on their own algorithms. Hence, it is incentive compatible for the intermediary to learn about the lenders' predictions and not about the true type of the borrowers.

[16]It is not difficult to see that setting $c > 0$ will not qualitatively change any of the results in this paper.

[17]We assume that borrowers could observe the number of lenders who reveals the algorithm, but not the specific revealing decisions of each lender. For example, Credit Karma is not willing to tell which two of the five lenders have revealed the algorithm. Additionally, by making this assumption, our model can be easily generalized to the case where borrowers have no knowledge on either how many lenders reveal their algorithms to the intermediary or which lenders do so.

- **STAGE 2: Pricing Stage** Each lender chooses an interest rate b ($b \in [0, \bar{b}]$)[18] for its financial product.

- **STAGE 3: Applying Stage** The intermediary shows the financial products of both lenders along with personalized approval odds to borrowers. Borrowers choose which lender to apply to or choose not to apply to either.

- **STAGE 4: Approving Stage** The lenders receive applications and use their algorithms to approve or reject borrowers. The payoffs to the lenders and the borrowers are then determined.

We use sub-game perfect Nash equilibrium (SPNE) as our solution concept. We solve for the SPNE using backward induction. We proceed with the analysis in the next section. A summary of notations can be found in Appendix A.

4 Analysis

4.1 Additional Assumptions

To avoid invalid and uninteresting equilibrium, we focus on the following parameter ranges:

Assumption 1

$$\bar{b}\theta - (1-\theta) \geq 0.$$

Assumption 1 says that granting loans to borrowers is *ex ante* efficient under interest rate $\bar{b}$. This assumption ensures that lenders can get non-negative equilibrium payoffs under mixed strategy settings in all sub-games.[19]

Assumption 2

$$m < \frac{P_c\left((M_h - \bar{b})P_b{}^2\theta + (\bar{b}\theta - \bar{b} + M_l - M_l\theta)(P_b - 1)^2\right)}{P_b{}^2(P_c\theta + P_c - 1) + P_b(1 - P_c) + (P_c - P_c\theta)(P_b - 1)^2}$$
$$m > \frac{M_h P_b{}^2\theta(P_c - 1) + M_l(P_c - P_c\theta + \theta - 1)(P_b - 1)^2}{P_b{}^2(P_c\theta + P_c - \theta) - P_bP_c + (P_c - P_c\theta + \theta - 1)(P_b - 1)^2}$$

The two conditions in Assumption 2 ensure that borrowers will always follow the intermediary's recommendation: The first condition ensures that the cost to apply (m) is small enough (or equivalently, the benefit from applying, M_h and M_l, are large enough) such that if a borrower is predicted

[18] $\bar{b}$ is the maximum interest rate that can be set by the lenders, for example, required by the law.

[19] Similar assumptions have been made in several papers, for example, in (Hauswald and Marquez, 2003) and (Hauswald and Marquez, 2006), to ensure that lenders will not have an incentive to engage in ruinous competition to obtain negative profits in equilibrium.

(possibly inaccurately) as H type by the intermediary for a lender, even when the lender charges the highest possible interest rate ($b = \bar{b}$), the borrower will still choose to apply. The second condition indicates that the cost to apply (m) is large enough (or equivalently, the benefit from applying, M_h and M_l, are small enough) such that if a borrower is predicted as L type by the intermediary for a lender, the borrower will not apply even if the interest rate is 0 ($b = 0$). The derivation of the two conditions in Assumption 2 can be found in Appendix B.1.

4.2 Analysis: Approving Stage

We begin our analysis from the last stage, the Approving Stage. The lenders' decisions in this stage are trivial: They make the approving decisions based on the algorithm's prediction. Specifically, they will approve (reject) borrowers who are predicted by their own algorithm as H (L) type.

4.3 Analysis: Applying Stage

After seeing the approval predictions from the intermediary, as well as the interest rates set by the two lenders, borrowers in this stage decide whether to apply and if so which lender to apply to. By Assumption 2, borrowers will only consider the lender(s) for which receive a prediction of H type from the intermediary. Specifically, (1) if a borrower learns from the intermediary that she will be predicted as L type by both lender, she will choose not to apply to either lender; (2) if a borrower learns from the intermediary that she will be predicted as H type by only one of the lenders, she will apply to that lender; (3) if a borrower learns from the intermediary that she will be predicted as H type by both lenders, she will apply to the lender who charges a lower interest rate. Note that in cases (1) and (2), borrowers make such decisions regardless of the interest rates the two lenders charge and the lenders' revealing decisions. In case (3), borrowers' decisions depend on the interest rate but not the revealing decisions of the two lenders, since even in the asymmetric revealing scenario, borrowers do not know which lender has revealed its algorithm, and thus the only asymmetry from the borrowers' perspective is the different interest rates the two lenders charge. Recall that Assumption 2 says it is always incentive compatible for borrowers to follow the intermediary's prediction. The constraints in Assumption 2 depend on P_c. Specifically, to ensure that borrowers will follow the noisy predictions by the intermediary, when P_c is high, conditions on m becomes milder, and when P_c is low, conditions on m becomes stricter .

4.4 Analysis: Pricing Stage

In this stage, the two lenders anticipate borrowers' applying strategies in Stage 3, and choose pricing strategies that depend on the revealing outcome of Stage 1. We begin the analysis by solving each sub-game that results from lenders' algorithm revelation choices in Stage 1. There are three possible outcomes (sub-games) in Stage 1:

R-R: Both lenders reveal their algorithms.

N-N: Neither lender reveals its algorithm.

N-R: One lender reveals its algorithm to intermediary while the other lender chooses not to.

We analyze the cases (sub-games) R-R, N-N, and N-R in succession.

4.4.1 Common Segment vs Captive Segment

Borrowers make their decisions to apply to a lender based on the type predictions that they expect to receive from the two lenders (or equivalently, the expected approval odds). There are four possible combinations (i.e., borrower segments) based on borrowers' belief in these predictions: HH – borrowers who believe they will be predicted as H type by both lenders, HL – borrowers who believe they will be predicted as H type for Lender 1 and as L type for Lender 2, LH – borrowers who believe they will be predicted as L type for Lender 1 and as H type for Lender 2, and LL – borrowers who believe they will be predicted as L type for both lenders.

The distribution of H and L type borrowers across the four segments will differ across the three sub-games: R-R, N-N and N-R. Figure 1 shows how the H and L type borrowers are distributed across the four segments in the R-R case. The top (bottom) expression in a cell computes the number of H (L) type borrowers that belong to the corresponding segment. For example, the top line in the HH segment reports number of H type borrowers who both Lender 1 and Lender 2 predict as H type. Each lender independently predicts the type of H type borrower correctly with probability P_b and θ is the fraction of borrowers who are H type in the population. Hence, $P_b^2\theta$ is the number of H type borrowers that fall in segment HH. The number of H and L type borrowers in other segments can be calculated similarly as reported in Figure 1.

Because we assume $P_b > 1/2$ (see Section 3.2), most of the H type borrowers are concentrated in the HH segment and most of the L type borrowers are concentrated in the LL segment. Similarly, HH segment has the lowest number of L type borrowers and the LL segment has the lowest number of H type borrowers. Therefore, the probability that a randomly chosen borrower from the segment is H type is highest for the HH segment and lowest for the LL segment. As a result, borrowers in HH are the most valuable (least risky) and LL are the least valuable (most risky) to lenders.

We can see that all borrowers who believe to receive a HH classification would be indifferent between the two symmetric lenders if they charge the same interest rate. As a result, the lenders will have to compete on this segment of borrowers in a Bertrand fashion. We refer to this segment of borrowers whose probability of applying to either of the lenders is non zero as the common

		Prediction from Lender 1	
		H	L
Prediction from Lender 2	H	$P_b^2\theta$ $(1-P_b)^2(1-\theta)$	$P_b(1-P_b)\theta$ $P_b(1-P_b)(1-\theta)$
	L	$P_b(1-P_b)\theta$ $P_b(1-P_b)(1-\theta)$	$(1-P_b)^2\theta$ $P_b^2(1-\theta)$

Figure 1: Common Segment vs Captive Segment for the R-R case

Note: From Lender 1's perspective, the upper left cell is the common segment while the lower left is its captive segment; From Lender 2's perspective, the upper left cell is the common segment while the upper right cell is its captive segment. The top line in each cell denotes the number of H type borrowers and the bottom line in each cell denotes the number of L type borrowers in the corresponding segment.

segment. In contrast, the borrowers in segment HL (LH) will only apply to Lender 1 (2) as they believe only Lender 1 (2) will approve them. We refer to them as the captive segment for Lender 1 (2). The lenders will compete to attract borrowers from the common segment but would face no competition for borrowers in the captive segment.

4.4.2 R-R: Both reveal

In the R-R case, borrowers observe accurate predictions from both lenders. A borrower knows exactly whether she will be approved by each lender or not.

To make the discussion easier, we use the following notations: We refer to the predictions of a borrower(k)'s type by Lender 1's and Lender 2's algorithms as $Y^1_{k,b}$ and $Y^2_{k,b}$, respectively. In the R-R case, lenders' predictions are directly passed on to borrowers by the intermediary. A is the set of all borrowers in the market. We define the following sets which correspond to borrowers in the segments defined in Section 4.4.1:

$$A^{rr}_{hh} = \{k \in A | Y^1_{k,b} = H, Y^2_{k,b} = H\}$$

$$A^{rr}_{hl} = \{k \in A | Y^1_{k,b} = H, Y^2_{k,b} = L\}$$

$$A^{rr}_{lh} = \{k \in A | Y^1_{k,b} = L, Y^2_{k,b} = H\}$$

$$A^{rr}_{ll} = \{k \in A | Y^1_{k,b} = L, Y^2_{k,b} = L\}$$

That is, A^{rr}_{hh} is the set of borrowers who are predicted as H type by both lenders, A^{rr}_{hl} is the set of borrowers who are predicted as H type by Lender 1 and predicted as L type by Lender 2, etc. Let

N_{xy}^{rr} denote the number of borrowers and V_{xy}^{rr} denote the fraction of H type borrowers in the set A_{xy}^{rr} where $x, y \in \{h, l\}$. The values of N_{xy}^{rr} and V_{xy}^{rr} can be calculated from Figure 1. Since neither lender will approve borrowers in A_{ll}^{rr}, we omit the calculations of N_{ll}^{rr} and V_{ll}^{rr} below.

$$N_{hl}^{rr} = P_b(1 - P_b)\theta + P_b(1 - P_b)(1 - \theta)$$
$$V_{hl}^{rr} = \frac{P_b(1 - P_b)\theta}{N_{hl}^{rr}} = \theta$$
$$N_{hh}^{rr} = P_b^2\theta + (1 - P_b)^2(1 - \theta)$$
$$V_{hh}^{rr} = \frac{P_b^2\theta}{N_{hh}^{rr}} > \theta$$

As discussed earlier, A_{hh}^{rr} is the common segment, and A_{hl}^{rr} and A_{lh}^{rr} are captive segments of Lender 1 and Lender 2 respectively. The segment A_{hh}^{rr} is the most valuable segment for the two lenders. The fact that the algorithms of both lenders predict these borrowers to be H type increases the posterior probability (after accounting for the competitor's prediction) that these borrowers are in fact H type. On the other hand, the posterior probability that the borrowers in the two lenders' respective captive segments, A_{hl}^{rr} and A_{lh}^{rr}, are in fact H type is lower and thus less profitable than those in segment A_{hh}^{rr} for lenders 1 and 2, respectively. As a result, lenders intensely compete for the segment A_{hh}^{rr}. The intensity of the competition is determined by the relative profitability of the segment A_{hh}^{rr} compared with the segment A_{hl}^{rr}, where the profitability of A_{hh}^{rr} intensifies the competition and profitability of A_{hl}^{rr} moderates the competition.

An important aspect of our context that is worth more clarification is that while a borrower can observe the accurate personalized odds of approval from both lenders in the R-R case, a lender cannot observe the approval odds that the competing lender gives to the borrower.[20] When a borrower, who the focal lender predicts as H type, applies to the lender, the lender cannot tell whether the competing lender would have classified her as H type or L type. As a result, a lender does not know whether the borrower is in segment A_{hh}^{rr} or A_{hl}^{rr}. In other words, lenders know the existence of the four segments defined above, but they do not know which segment each borrower belongs to. This implies that lenders cannot treat borrowers in the common and their own captive segment differently (e.g., charge different interest rates).

We next derive lenders' optimal strategies for setting their interest rates, b, in this case. The borrowers in segment A_{hh}^{rr} are indifferent between the two lenders if the interest rates offered by the two lenders are the same. They know that they will be approved by both lenders. The borrowers in the A_{hl}^{rr} segment will only apply to Lender 1 and those in the A_{lh}^{rr} segment will only apply to Lender 2. For borrowers in either lender's captive segment, the interest rate offered by the other

[20]Lenders do not have access to their competitors' algorithms or predictions.

lender does not matter as they will not be approved by that lender.

The two lenders set the interest rate (b) of their financial product simultaneously. A pure strategy equilibrium does not exist in this case. The reason is as follows: Both lenders will have an incentive to undercut the competitor's interest rate to attract borrowers in A^{rr}_{hh}. Since these borrowers are indifferent between the two lenders if the two lenders charge the same interest rate, choosing b that is just a bit smaller than the competitor's will attract all of the borrowers in the A^{rr}_{hh} segment. Once b drops below a certain level, one lender is better off raising b to only focus on its captive borrowers A_{hl} (or A_{lh}). Once a lender raises b, the other lender will have an incentive to also raise b to just a little lower than the first lender's, and then the war on cutting b continues. Therefore, we study the mixed strategy equilibrium for pricing in this case, and focus on the symmetric Nash equilibrium, i.e., the two lenders use the same mixed strategy. In a mixed strategy, lenders randomize their b.[21] Following the literature on mixed strategy pricing (Varian, 1980), we use a probability distribution with a cumulative density function (CDF) $F(b)$ to characterize lenders' strategy. The two lenders' equilibrium strategy is summarized in Lemma 1, the proof can be found in Appendix B.2.

Lemma 1 *The CDF of the distribution of lender's equilibrium pricing (interest rate setting) strategy is shown as follows:*

$$F^{rr}(b) = \begin{cases} 0 & \text{if } b < \underline{b} \\ \frac{bP_b\theta - P_b\theta - P_b\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right) + P_b + \theta - 1}{bP_b{}^2\theta + P_b{}^2\theta - P_b{}^2 - 2P_b\theta + 2P_b + \theta - 1} & \text{if } \underline{b} \le b < \bar{b} \\ 1 & \text{if } b \ge \bar{b} \end{cases}$$

Each lender's equilibrium payoff is:

$$\Pi_{rr} = P_b\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right)$$

An interesting observation from Lemma 1 is that lenders' equilibrium profit will decrease as the algorithms' accuracy increases. Taking derivative of lenders' equilibrium payoff with respect to P_b, we get

$$\frac{\partial \Pi_{rr}}{\partial P_b} = (\bar{b}\theta + \theta - 1)(1 - 2P_b) \le 0 \tag{2}$$

As P_b increases, the common segment becomes more profitable, which intensifies competition. As

[21] In the real world, such a strategy can be interpreted as lenders offering different interest rates for periods of a varying length, promotional interest rates or cash back rewards of a varying amount, which effectively changes b (Varian, 1980).

competition increases, the equilibrium interest rate gets lower, which decreases the profit from both the common and the captive segments.

4.4.3 N-N: Neither Reveal

In the N-N case, borrowers learn their chances of getting approved by both lenders from the intermediary's algorithm. In Figure 2, we summarize the distribution of H and L type borrowers in different segments based on the intermediary's and the lenders' predictions. As shown in Figure 2, there are 16 segments. Each segment is defined by the combination of the intermediary's and lenders' predictions in the following sequence – Lender 1's prediction, the intermediary's prediction for Lender 1, Lender 2's prediction and the intermediary's prediction for Lender 2. For example, the segment $HLHL$ includes all borrowers whom Lender 1 predicts as H type, the intermediary predicts as L type for Lender 1, Lender 2 predicts as H type, and the intermediary predicts as L type for Lender 2. In Figure 2, the number of H *(L)* type borrowers that belong to a segment are shown in the top (bottom) row in the corresponding cell. The calculations are straightforward. For example, to calculate the number of H type borrowers in $HLHL$, we multiply the following: (i) P_b, the probability of Lender 1 predicting the H type correctly, (ii) $1 - P_c$, the probability of the intermediary predicting the Lender 1's prediction incorrectly, (iii) P_b, the probability of Lender 2 predicting the H type correctly, (iv) P_c, the probability of the intermediary predicting the Lender 2's prediction correctly, and (v) θ, the fraction of borrowers who are H type in the population.

Without loss of generality, we discuss the results from Lender 1's perspective. Since the lender would not approve a borrower whom its own algorithm classifies as L type, Lender 1 will not approve any borrower who falls under columns 3 and 4 in Figure 2. Further, following our assumption in Section 4.1, borrowers who receive a prediction of L type from the intermediary for a lender would not apply to that lender. As a result, borrowers who fall under the second column in 2 would not apply to Lender 1. Hence, the only relevant borrowers for Lender 1 are the ones who fall under Column 1. We define the following sets to capture the borrowers in Column 1 of Figure 2 for ease of explanation. Note that $Y^1_{k,c}$ ($Y^2_{k,c}$) denotes the intermediary's prediction of an borrower(k)'s type for Lender 1 (2).

$$A^{nn}_{hh} = \{k \in A | Y^1_{k,c} = H \wedge Y^2_{k,c} = H \wedge Y^1_{k,b} = H\}$$

$$A^{nn}_{hl} = \{k \in A | Y^1_{k,c} = H \wedge Y^2_{k,c} = L \wedge Y^1_{k,b} = H\}$$

Specifically, A^{nn}_{hh} denotes the set of borrowers who are predicted as H type by the intermediary for both lenders and are predicted as H type by Lender 1. This is the common segment since borrowers in this segment do not have a preference between the two lenders. In Figure 2, the

				Prediction from Lender 1			
				H		L	
				Intermediary's prediction		Intermediary's prediction	
				H	L	H	L
Prediction from Lender 2	H	Intermediary's prediction	H	$P_c^2P_b^2\theta$ $P_c^2P_b^2(1-\theta)$	$P_c(1-P_c)P_b^2\theta$ $P_c(1-P_c)P_b^2(1-\theta)$	$P_c(1-P_c)P_b(1-P_b)\theta$ $P_c(1-P_c)P_b(1-P_b)(1-\theta)$	$P_c^2P_b(1-P_b)\theta$ $P_c^2P_b(1-P_b)(1-\theta)$
			L	$P_c(1-P_c)P_b^2\theta$ $P_c(1-P_c)P_b^2(1-\theta)$	$(1-P_c)^2P_b^2\theta$ $(1-P_c)^2P_b^2(1-\theta)$	$(1-P_c)^2P_b(1-P_b)\theta$ $(1-P_c)^2P_b(1-P_b)(1-\theta)$	$P_c(1-P_c)P_b(1-P_b)\theta$ $P_c(1-P_c)P_b(1-P_b)(1-\theta)$
	L	Intermediary's prediction	H	$P_c(1-P_c)P_b(1-P_b)\theta$ $P_c(1-P_c)P_b(1-P_b)(1-\theta)$	$(1-P_c)^2P_b(1-P_b)\theta$ $(1-P_c)^2P_b(1-P_b)(1-\theta)$	$(1-P_c)^2(1-P_b)^2\theta$ $(1-P_c)^2(1-P_b)^2(1-\theta)$	$P_c(1-P_c)(1-P_b)^2\theta$ $P_c(1-P_c)(1-P_b)^2(1-\theta)$
			L	$P_c^2P_b(1-P_b)\theta$ $P_c^2P_b(1-P_b)(1-\theta)$	$P_c(1-P_c)P_b(1-P_b)\theta$ $P_c(1-P_c)P_b(1-P_b)(1-\theta)$	$P_c(1-P_c)(1-P_b)^2\theta$ $P_c(1-P_c)(1-P_b)^2(1-\theta)$	$P_c^2(1-P_b)^2\theta$ $P_c^2(1-P_b)^2(1-\theta)$

Figure 2: Common Segment vs Captive Segment for the N-N case

Note: From Lender 1's perspective, the cells (Row 1, Column 1) and (Row 3, Column 1) combined are the common segment while the cell (Row 2, Column 1) and (Row 4, Column 1) combined are its captive segment. From Lender 2's perspective, cell (Row 1, Column 1) and (Row 1, Column 3) combined are the common segment while cells (Row 1, Column 2) and (Row 1, Column 4) combined are its captive segment. The top line in each cell denotes the number of H type borrowers while the bottom line in each cell denotes the number of L type borrowers.

cell (Row 1, Column 1) and (Row 3, Column 1) together constitute A^{nn}_{hh}. A^{nn}_{hl} denotes the set of borrowers who are predicted as H type by the intermediary for Lender 1, L type by the intermediary for Lender 2, and are predicted as H type by Lender 1. This is the captive segment for Lender 1 since the borrowers in this segment will always apply to Lender 1. In Figure 2, the cells (Row 2, Column 1) and (Row 4, Column 1) together constitute A^{nn}_{hl}.

We use N^{nn}_{xy} to denote the number of borrowers and V^{nn}_{xy} to denote the fraction of H type borrower in the set A^{nn}_{xy} where $x, y \in \{h, l\}$. From Figure 2, we can calculate:

$$
\begin{aligned}
N^{nn}_{hh} &= P_c^2 P_b^2\theta + P_c(1-P_c)P_b(1-P_b) + P_c^2(1-P_b)^2(1-\theta) \\
V^{nn}_{hh} &= \frac{P_c^2P_b^2\theta + P_c(1-P_c)P_b(1-P_b)\theta}{N^{nn}_{hh}} > \theta \\
N^{nn}_{hl} &= P_c(1-P_c)P_b^2\theta + P_c(1-P_c)(1-P_b)^2(1-\theta) + P_c^2P_b(1-P_b) \\
V^{nn}_{hl} &= \frac{P_b^2P_c(1-P_c)\theta + P_c^2P_b(1-P_b)\theta}{N^{nn}_{hl}} > \theta
\end{aligned} \tag{3}
$$

Compared to the R-R case, the competition between lenders in the N-N case is less intense. As before, the relative profitability of the common segment intensifies the competition whereas the relative profitability of the captive segment moderates the competition. Mathematically, we have

$N_{hh}^{nn} - N_{hl}^{nn} < N_{hh}^{rr} - N_{hl}^{rr}$ and $V_{hh}^{nn} - V_{hl}^{nn} < V_{hh}^{rr} - V_{hl}^{rr}$. This implies that compared to the R-R case, where the A_{hh}^{rr} segment is highly profitable relative to the A_{hl}^{rr} segment, in the N-N case, the difference between the profitability of A_{hh}^{nn} and A_{hl}^{nn} segment shrinks: While the intermediary predicts the borrowers in segment A_{hh}^{nn} to be H type for Lender 2 by definition, some of them may, in fact, be classified as L type by Lender 2's algorithm, because the intermediary's prediction is not perfect. Similarly, some of the borrowers in segment A_{hl}^{nn} may in fact be classified by Lender 2 as H type while the intermediary classifies them as L type for Lender 2. In other words, the intermediary's prediction of the competing lender's classification of a borrower's type provides a noisy signal for the competing lender's classification. Hence, compared to the R-R case, lenders give more weight to the prediction by their own algorithms when computing the posterior probability of a borrower's true type being H in the N-N case. As a result, the difference between the posterior probability that borrowers in segments A_{hh}^{nn} are H type and the posterior probability that borrowers in segment A_{hl}^{nn} are H type drops compared to the R-R case. At the same time, the difference in size between the two segments also drops compared to the R-R case.

We next solve for the optimal strategies in setting b for the two lenders in the N-N sub-game. As in the R-R case, a pure strategy equilibrium in interest rates does not exist in the N-N case. Therefore, we study the mixed strategy equilibrium, and focus on the symmetric Nash equilibrium, i.e., the two lenders use the same mixed strategy. We use a probability distribution with a CDF $F^{nn}(b)$ to characterize lenders' pricing strategy in the N-N case.

Lemma 2 *The CDF of the distribution of lender's equilibrium pricing strategy in b is as follows:*

$$F^{nn}(b) = \begin{cases} 0 & \text{if } b < \underline{b} \\ \frac{bN_{hh}^{nn}V_{hh}^{nn} + bN_{hl}^{nn}V_{hl}^{nn} + N_{hh}^{nn}V_{hh}^{nn} - N_{hh}^{nn} + N_{hl}^{nn}V_{hl}^{nn} - N_{hl}^{nn}\left(\bar{b}V_{hl}^{nn} + V_{hl}^{nn} - 1\right) - N_{hl}^{nn}}{N_{hh}^{nn}\left(bV_{hh}^{nn} + V_{hh}^{nn} - 1\right)} & \text{if } \underline{b} \leq b < \bar{b} \\ 1 & \text{if } b \geq \bar{b} \end{cases}$$

Each lender's equilibrium payoff is:

$$\Pi_{nn} = N_{hl}^{nn}\left(\bar{b}V_{hl}^{nn} + V_{hl}^{nn} - 1\right)$$

The proof of Lemma 2 can be found in Appendix B.3.

4.4.4 N-R: Asymmetric Revealing

In the N-R case, one lender reveals its algorithm so the predictions that borrowers receive from the intermediary reflect the true approval odds for the revealing lender. The other lender does not reveal the algorithm to intermediary and the borrowers have to rely on the predictions of the intermediary's reverse-engineered algorithm, which are not always accurate. Without loss of generality, in the following discussion, we consider Lender 1 as the lender that reveals its algorithm and Lender 2 as the lender that does not.

Similar to what we have done for the R-R and N-N case, in Figure 3, we show the distribution of H and L type borrowers across different segments based on the intermediary's and the lenders' predictions. As can be seen in Figure 3, there are 8 segments. Each segment is defined by the combination of the intermediary's and lenders' predictions in the following sequence – Lender 1's prediction, Lender 2's prediction and the intermediary's prediction for Lender 2. For example, the segment HLH includes all borrowers whom Lender 1 predicts as H type, Lender 2 predicts as L type, and the intermediary predicts as H type for Lender 2. In Figure 3, the number of *H(L)* type borrowers that belong to a segment is shown in the top (bottom) line in the corresponding cell.

				Prediction from Lender 1	
				H	L
Prediction from Lender 2	H	Intermediarys prediction	H	$P_cP_b^2\theta$ $P_cP_b^2(1-\theta)$	$P_cP_b(1-P_b)\theta$ $P_cP_b(1-P_b)(1-\theta)$
			L	$(1-P_c)P_b^2\theta$ $(1-P_c)P_b^2(1-\theta)$	$(1-P_c)P_b(1-P_b)\theta$ $(1-P_c)P_b(1-P_b)(1-\theta)$
	L	Intermediarys prediction	H	$(1-P_c)P_b(1-P_b)\theta$ $(1-P_c)P_b(1-P_b)(1-\theta)$	$(1-P_c)(1-P_b)^2\theta$ $(1-P_c)(1-P_b)^2(1-\theta)$
			L	$P_cP_b(1-P_b)\theta$ $P_cP_b(1-P_b)(1-\theta)$	$P_c(1-P_b)^2\theta$ $P_c(1-P_b)^2(1-\theta)$

Figure 3: Common Segment vs Captive Segment for the N-R case

Note: From Lender 1's perspective, the cells (Row 1, Column 1) and (Row 3, Column 1) together are the common segment, while the cells (Row 2, Column 1) and (Row 4, Column 1) together are its captive segment. From Lender 2's perspective, the cell (Row 1, Column 1) is the common segment, while the cell (Row 1, Column 2) is its captive segment. The top line in each cell denotes the number of H type borrowers while the bottom line in each cell denotes the number of L type borrowers.

Since the lender would not approve any borrower whom its own algorithm classifies as L type, Lender 1 will not approve any borrower who falls under Column 2 in Figure 3. Similarly, Lender 2 will not approve any borrower who falls in Row 3 or Row 4 in Figure 3. Further, following our assumption in Section 4.1, borrowers who receive a prediction of L type from the intermediary for a lender would not apply to that lender. Thus, borrowers who fall under the Column 2 in Figure

3 would not apply to Lender 1; those who fall in the Row 2 or Row 4 in Figure 3 would not apply to Lender 2. Hence, the only relevant borrowers for Lender 1 are the ones who fall under Column 1; the only relevant borrowers for Lender 2 are those who fall in Row 1. We define the following sets to capture the borrowers in Column 1 and/or Row 1 of Figure 3 for ease of explanation.

$$\begin{aligned}
A_{hh}^1 &= \{k \in A | Y_{k,b}^1 = H \wedge Y_{k,c}^2 = H\} \\
A_{hh}^2 &= \{k \in A | Y_{k,b}^1 = H \wedge Y_{k,c}^2 = H \wedge Y_{k,b}^2 = H\} \\
A_{hl}^1 &= \{k \in A | Y_{k,b}^1 = H \wedge Y_{k,c}^2 = L\} \\
A_{lh}^2 &= \{k \in A | Y_{k,b}^1 = L \wedge Y_{k,c}^2 = H \wedge Y_{k,b}^2 = H\}
\end{aligned}$$

Specifically, A_{hh}^1 denotes the set of borrowers who are predicted as H type by the intermediary for Lender 2 and are predicted as H type by Lender 1. A_{hh}^2 denotes the set of borrowers who are predicted as H type by both lenders and are predicted as H type by the intermediary for Lender 2. These are the common segment for Lender 1 and Lender 2 respectively since borrowers in these sets do not have a preference between the two lenders. In Figure 3, the cells (Row 1, Column 1) and (Row 3, Column 1) together constitute A_{hh}^1, the cell (Row 1, Column 1) corresponds to A_{hh}^2. A_{hl}^1 denotes the set of borrowers who are predicted as L type by the intermediary for Lender 2 and are predicted as H type by Lender 1. This is the captive segment for Lender 1 since the borrowers in this segment will always apply to Lender 1. A_{lh}^2 denotes the set of borrowers who are predicted as L type by Lender 1 and are predicted as H type by Lender 2 and by the intermediary for Lender 2. In Figure 3, the cells (Row 2, Column 1) and (Row 4, Column 1) together constitute A_{hl}^1, and the cell (Row 1, Column 2) corresponds to A_{lh}^2.

Again, N_{xy}^k denotes the number of borrowers, and V_{xy}^k denotes the fraction of H type borrowers in set A_{xy}^k where $x, y \in \{h, l\}$ and $k \in \{1, 2\}$. Lenders set interest rates, b_1 and b_2 simultaneously to compete. As before, a pure strategy equilibrium in interest rates does not exist. We study the mixed strategy equilibrium. We use the CDFs $F_1(b)$ and $F_2(b)$ to characterize Lender 1's (revealing lender) and Lender 2's (non-revealing lender) pricing strategies respectively. The lenders' equilibrium strategies are summarized in Lemma 3; the proof can be found in Appendix B.4.

Lemma 3 *The CDF for the revealing lender's equilibrium pricing strategy in b is*

$$F_1(b) = \begin{cases} 0 & \text{if } b < \underline{b} \\ \frac{bN_{hh}^2 V_{hh}^2 + bN_{lh}^2 V_{lh}^2 - k_2 + N_{hh}^2 V_{hh}^2 - N_{hh}^2 + N_{lh}^2 V_{lh}^2 - N_{lh}^2}{N_{hh}^2 \left(bV_{hh}^2 + V_{hh}^2 - 1\right)} & \text{if } \underline{b} \leq b < \overline{b} \\ 1 & \text{if } b \geq \overline{b} \end{cases}$$

The CDF for the non-revealing lender's equilibrium pricing strategy in b is

$$F_2(b) = \begin{cases} 0 & \text{if } b < \underline{b} \\ \frac{bN^1_{hh}V^1_{hh} + bN^1_{hl}V^1_{hl} - k_1 + N^1_{hh}V^1_{hh} - N^1_{hh} + N^1_{hl}V^1_{hl} - N^1_{hl}}{N^1_{hh}\left(bV^1_{hh} + V^1_{hh} - 1\right)} & \text{if } \underline{b} \leq b < \overline{b} \\ 1 & \text{if } b \geq \overline{b} \end{cases}$$

The equilibrium payoffs to the revealing lender and the non-revealing lender are

$$\Pi^1_{nr} = k_1$$

$$\Pi^2_{nr} = k_2$$

where

$$\begin{cases} k_1 = & N^1_{hl}\left(\overline{b}V^1_{hl} + V^1_{hl} - 1\right) \\ k_2 = & (\overline{b}(N^2_{hh}N^1_{hl}V^2_{hh}V^1_{hl} + N^1_{hl}N^2_{lh}V^1_{hl}V^2_{lh}) + N^1_{hh}N^2_{hh}(V^2_{hh} - V^1_{hh}) - N^1_{hh}N^2_{lh}(-V^1_{hh} + V^2_{lh}) \\ & +N^2_{hh}N^1_{hl}V^2_{hh}V^1_{hl} - N^2_{hh}N^1_{hl}V^1_{hl} + N^1_{hl}N^2_{lh}V^1_{hl}V^2_{lh} - N^1_{hl}N^2_{lh}V^1_{hl})/(N^1_{hh}V^1_{hh} + N^1_{hl}V^1_{hl}) \end{cases}$$

$$\begin{cases} N^1_{hh} = & {P_b}^2P_c\theta + P_b\theta\left(-P_b+1\right)\left(-P_c+1\right) + P_b\left(-P_b+1\right)\left(-P_c+1\right)\left(-\theta+1\right) \\ & +P_c(-P_b+1)^2\left(-\theta+1\right) \\ N^2_{hh} = & {P_b}^2P_c\theta + P_c(-P_b+1)^2\left(-\theta+1\right) \\ N^1_{hl} = & {P_b}^2\theta\left(-P_c+1\right) + P_bP_c\theta\left(-P_b+1\right) + P_bP_c\left(-P_b+1\right)\left(-\theta+1\right) \\ & +(-P_b+1)^2\left(-P_c+1\right)\left(-\theta+1\right) \\ N^2_{lh} = & P_bP_c\theta\left(-P_b+1\right) + P_bP_c\left(-P_b+1\right)\left(-\theta+1\right) \\ V^1_{hh} = & \frac{{P_b}^2P_c\theta + P_b\theta(-P_b+1)(-P_c+1)}{{P_b}^2P_c\theta + P_b\theta(-P_b+1)(-P_c+1) + P_b(-P_b+1)(-P_c+1)(-\theta+1) + P_c(-P_b+1)^2(-\theta+1)} \\ V^2_{hh} = & \frac{{P_b}^2P_c\theta}{{P_b}^2P_c\theta + P_c(-P_b+1)^2(-\theta+1)} \\ V^1_{hl} = & \frac{{P_b}^2\theta(-P_c+1) + P_bP_c\theta(-P_b+1)}{{P_b}^2\theta(-P_c+1) + P_bP_c\theta(-P_b+1) + P_bP_c(-P_b+1)(-\theta+1) + (-P_b+1)^2(-P_c+1)(-\theta+1)} \\ V^2_{lh} = & \frac{P_bP_c\theta(-P_b+1)}{P_bP_c\theta(-P_b+1) + P_bP_c(-P_b+1)(-\theta+1)} \end{cases}$$

Mathematically, we have $F_1(b) \leq F_2(b)$, $\forall b \in [0, \overline{b}]$. This implies that the revealing lender sets a higher interest rate than the non-revealing lender in general. Figure 4 illustrates the CDFs of the two lenders' strategies in the N-R case, together with the lenders' equilibrium strategies in the N-N and R-R cases. As is shown in the figure, the interest rate set in the R-R case is the lowest, indicating intense competition between the two lenders in that case.

It is easy to see that $V^1_{hh} < V^2_{hh}$, which confirms that the common segment for the revealing lender is riskier than the common segment for the non-revealing lender. In contrast, the revealing

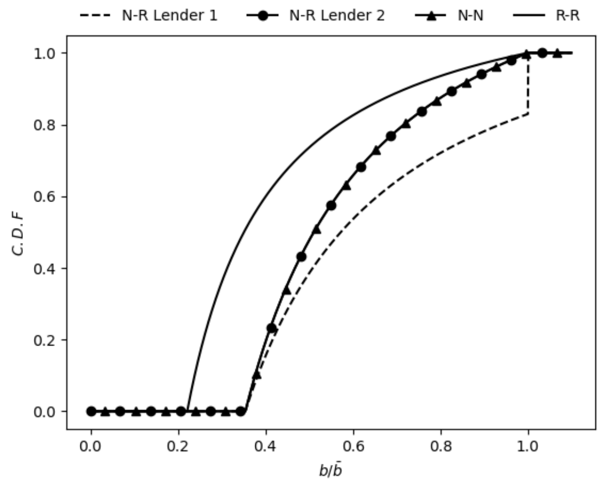

Figure 4: Lenders' equilibrium pricing strategies in each sub-game; $P_b = 0.8$, $P_c = 0.8$, $\theta = 0.7$

lender faces a less risky captive segment as the fraction of H type borrowers are higher in its captive segment. In other words, $V^1_{hl} > V^2_{lh}$. The reason is that the signal from the revealing lender is more precise compared to the signal from the non-revealing lender as the intermediary's noise is added to the latter.

While both lenders still prefer their common segment to the their captive segment, the revealing lender's preference for its common segment over its captive segment is not as strong as the non-revealing lender's. As a result, the revealing lender is more willing to sacrifice the common segment and chooses a higher interest rate to extract greater surplus from its captive segment. In contrast, the non-revealing lender sets a lower interest rate, which allows it to capture more borrowers in its common segment while sacrificing surplus from its captive segment.

Figure 5 shows the equilibrium payoff breakdown by different borrower segments for each lender in the N-R case. Since the lenders are playing mixed strategies, we plot the profit breakdown by lender (revealing vs non-revealing) and segment as a function of b when the competing lender is using the equilibrium mixed strategy. Both solid lines decrease in b, because setting a higher b will make the focal lender's product less competitive in the common segment (A^1_{hh} or A^2_{hh}). In contrast, the dashed lines increase in b, because borrowers in segments A^1_{hl} or A^2_{lh} are captive and the profit generated from them is proportional to b. Notice that compared with the non-revealing lender, a larger portion of the revealing lender's profit is generated from its captive segment.

It is not difficult to check that $\frac{\Pi^2_{nr}}{\Pi^1_{nr}} = P_c \leq 1$. This means that the revealing lender gets a higher profit in equilibrium than the non-revealing lender, and the difference in the two lenders' profits decreases in P_c. Note that $(1 - P_c)$ is the amount of noise that the intermediary introduces to the approval prediction borrowers receive. When a lender reveals its algorithm, it removes this noise

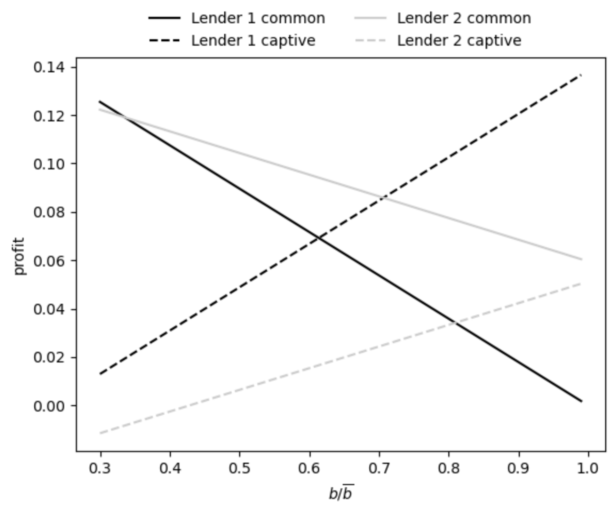

Figure 5: Lenders' equilibrium profit break down in the N-R case

Note: The solid lines represent the profit generated from common segment borrowers; the dashed lines represent the profit generated from captive segment borrowers.

from the process. The revealing lender benefits from the market expansion effect as it attracts more qualified borrowers to apply who would not have applied if the algorithm is not revealed due to the errors in intermediary's predictions. At the same time, the non-revealing lender benefits from an information advantage as it gets a more accurate signal from the revealing lender's algorithm. Both the market expansion and information advantage effects are a function of $(1 - P_c)$. However, compared to the marginal effect of market expansion, the marginal effect of information advantage increases at a slower rate with $(1 - P_c)$. Hence, by revealing the algorithm, the revealing lender is in a more advantageous position relative to the non-revealing lender.

4.5 Analysis: Revealing Stage

4.5.1 The Payoff Matrix and the Algorithm Revealing Equilibrium

So far, we have solved all possible sub-games after Stage 2. We now solve for the equilibrium in Stage 1. The "revealing game" that the two lenders are facing in Stage 1 is shown in the following table,

		Lender 2: Not Reveal	Lender 2: Reveal
Lender 1	Not Reveal	(Π_{nn}, Π_{nn})	(Π^1_{nr}, Π^2_{nr})
	Reveal	(Π^2_{nr}, Π^1_{nr})	(Π_{rr}, Π_{rr})

where Π_{rr}, Π_{nn}, Π^1_{nr} and Π^2_{nr} denote lenders' equilibrium payoffs under different sub-games, as defined in Lemmas 1, 2, and 3.

We prove that $\Pi^1_{nr} > \Pi_{nn}$ always holds (see Appendix B.5), which rules out N-N as an equilibrium outcome. One of the lenders in the N-N case always has the incentive to deviate and reveal its algorithm to benefit from the market expansion effect.

The relationship between Π^2_{nr} and Π_{rr} depends upon P_b: $\Pi^2_{nr} > \Pi_{rr}$ when P_b is relatively high. The non-revealing lender in the N-R case would choose to stay in the N-R or switch to R-R case based on the balance among three forces (a) market expansion (b) information advantage, and (c) competition. By revealing the algorithm, the non-revealing lender can benefit from market expansion but will lose the "information" advantage it holds over its competitor. As we discussed earlier, the market expansion effect from revealing the algorithm dominates the information advantage effect from not revealing. However, in the R-R case, the lenders become symmetric and the competition on the interest rate is intensified. The intensity of the competition on the interest rate is determined by the relative profitability of the common segment over the captive segment which is directly proportional to P_b. When P_b is low, even when both lenders reveal their algorithms, the competition is not very intense. Hence, the non-revealing lender in the N-R case would choose to switch to the R-R case and benefit from the market expansion effect when P_b is low. In contrast, when P_b is high, it would prefer not to reveal its algorithm and stick to the N-R case to avoid intense competition.

We summarize the algorithm revealing equilibrium in Stage 1 in the following proposition.

Proposition 1 *The algorithm revealing equilibrium is determined by lenders' algorithm's accuracy, P_b, as follows:*
1. If lenders' algorithm's accuracy P_b is sufficiently low, i.e., $P_b < P_b^0$, then $\Pi^2_{nr} < \Pi_{rr}$, and the revealing equilibrium is symmetric such that both lenders choose to reveal the algorithm to the intermediary.
2. If lenders' algorithm's accuracy P_b is sufficiently high, i.e., $P_b \geq P_b^0$, then $\Pi^2_{nr} \geq \Pi_{rr}$, and the revealing equilibrium is asymmetric such that only one lender chooses to reveal the algorithm to the intermediary.
The threshold P_b^0 above is a function of P_c and θ:

$$P_b^0 = \frac{\bar{b}P_c\theta + \bar{b}\theta + 3P_c\theta - 3P_c + \theta - 1 + \sqrt{\Delta}}{4\bar{b}P_c\theta + 2\bar{b}\theta + 4P_c\theta - 4P_c + 2\theta - 2}$$

$$\begin{aligned}\Delta =& \bar{b}^2 {P_c}^2\theta^2 + 2\bar{b}^2 P_c\theta^2 + \bar{b}^2\theta^2 - 2\bar{b}{P_c}^2\theta^2 + 2\bar{b}{P_c}^2\theta + 4\bar{b}P_c\theta^2 - 4\bar{b}P_c\theta + 2\bar{b}\theta^2 \\ &- 2\bar{b}\theta + {P_c}^2\theta^2 - 2{P_c}^2\theta + {P_c}^2 + 2P_c\theta^2 - 4P_c\theta + 2P_c + \theta^2 - 2\theta + 1\end{aligned}$$

The proof can be found in Appendix B.5. The dependence of revealing equilibrium on P_b and P_c

is shown in figure 6, which graphically illustrates Proposition 1. It can be seen that an asymmetric equilibrium will occur unless the accuracy of lenders' algorithm is very low.

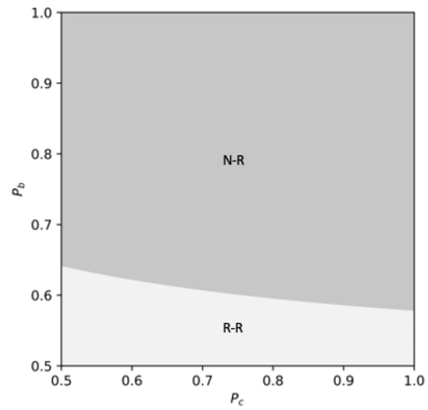

Figure 6: Algorithm revealing equilibrium depending on P_b and P_c when $\theta = 0.8$, $\bar{b} = 0.5$

Figure 7 further illustrates the equilibrium profits under the pricing equilibrium in the three possible sub-games resulting from the two lenders' revealing decisions (i.e., R-R, N-N, and N-R). SPNE profits are denoted in black, and off-equilibrium profits are denoted in gray. By deviating from the symmetric N-N case to the N-R case, the revealing lender is able to get a much higher profit due to the market expansion effect. On the other hand, the non-revealing lender does not become worse off because of the softened competition on b.

4.5.2 Effects of P_c and θ on revealing SPNE

The effects of P_c and θ on SPNE are summarized in Proposition 2.

Proposition 2 *P_b^0 decreases with P_c and increases with θ.*

The proof can be found in Appendix B.6. Proposition 2 states that the threshold that P_b needs to exceed for asymmetric revealing equilibrium to be sustained decreases with P_c. When P_c is high, the noise introduced by the intermediary is very small. As a result, the market expansion effect that a lender can benefit from and the information advantage it loses from revealing the algorithm are both small. At the same time, there is little difference in the intensity of competition between the N-R case and the R-R case. However, if the non-revealing lender in the N-R case reveals its algorithm and switches to the R-R case, as P_c increases, the positive market expansion effect from revealing would decrease at a faster rate than the rate at which the negative effects of the intensified competition and the loss of information advantage would decrease. Therefore, when P_c is large, the non-revealing lender in the N-R case would prefer not to switch to the R-R case by revealing its algorithm even at some relatively low values of P_b.

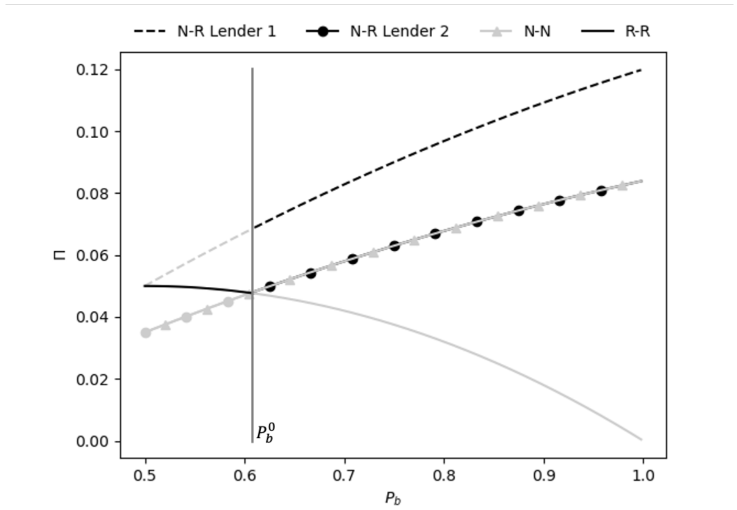

Figure 7: Equilibrium profits in each algorithm revealing sub-games when $P_c = 0.7$, $\theta = 0.8$, $\bar{b} = 0.5$

Note. SPNE profits are denoted in black, and off-equilibrium profits are denoted in gray. When $P_b < P_b^0$ (the vertical line in the figure), the SPNE is both lenders revealing their algorithms in Stage 1 and setting interest rates according to Lemma 1 in Stage 2, and the corresponding equilibrium profits for the two lenders are represented by the black portion of the "R-R case" line. When $P_b \geq P_b^0$, the SPNE is one lender revealing its algorithm and the other not revealing its algorithm in Stage 1, and the revealing and non-revealing lenders setting interest rates according to Lemma 3. The corresponding equilibrium profit for the revealing lender is represented by the black portion of the "N-R case Lender 1" line and that for the non-revealing lender is represented by the black portion of the "N-R case Lender 2" line.

Moreover, when θ, the portion of H type borrowers in the population, increases, the threshold on P_b that is required to sustain asymmetric equilibrium increases. In comparison to the intensity of competition, the market expansion effect is affected at a greater rate by θ. As a result, a higher value of θ encourages the non-revealing lender in the N-R case to reveal its algorithm and switch to the R-R case to take advantage of the market expansion effect.

4.6 Borrower Surplus and Social Welfare

We next examine borrowers' surplus under each sub-game resulting from lenders' decisions on algorithm revelation in Stage 1. We proceed by first calculating total surplus of lenders and borrowers, and then subtracting the lenders' equilibrium payoff. Since the interest rates set by the lender do not directly affect total surplus,[22] our approach to calculate borrower surplus avoids integration over b. Previously, we have divided borrowers into several segments according to the lenders' and intermediary's predictions. Some segments of borrowers never apply to either of the lenders, and

[22] The reason is as follows. Borrowers who do not apply to either lender generate no surplus to the society; those who apply and are rejected each generate a negative surplus of $-m$; those who apply and are approved also generates a fixed amount of surplus, which depends on their type but not on b, because the interest payments are surplus transferred from borrowers to lenders. Under Assumption 2, the number of borrowers in each of the segments is not affected by b either, because as long as a borrower receives at least one positive ("H") prediction, she will apply to one lender regardless of the interest rate.

those borrowers generate zero total surplus to the system. Some segments of borrowers will apply but will not be approved, and those borrowers generate a negative surplus because there is a cost associated with being rejected. Other segments of borrowers will apply and will be approved, and whether they generate a positive total surplus to the system or not depends on their types.

The comparison of total surplus and borrower surplus in the three algorithm revealing cases is summarized in Proposition 3.

Proposition 3 *Both total surplus and borrower surplus are the highest in the R-R case, followed by the asymmetric case, and the lowest in the N-N case. That is,* $TS^{rr} > TS^{nr} > TS^{nn}$*, and* $CS^{rr} > CS^{nr} > CS^{nn}$*.*

The full derivation and expressions of total surplus and borrower surplus can be found in Appendix B.7. Here we focus on discussing the comparison between the N-R case and the R-R case because these are the two cases that can be sustained in the equilibrium. We first look at why $TS^{rr} > TS^{nr} > TS^{nn}$. In the R-R case, there is no uncertainty faced by borrowers regarding their approval odds. In contrast, borrowers face the most uncertainty about their approval odds in the N-N case. The uncertainty regarding approval odds leads borrowers to make sub-optimal application decisions, e.g. borrowers who will be approved choose not to apply whereas borrowers who will be rejected choose to apply. The number of such non-optimal decisions is the lowest in the R-R case and highest in the N-N case. These non-optimal decisions create negative surplus to the society. A borrower who applies but gets rejected will create $-m$ social surplus, and a borrower who would be approved but does not apply will result in an opportunity cost to social surplus with a size of $\theta(M_h - m) + (1-\theta)(M_l - m - 1)$.[23] Hence, social surplus is the highest in the R-R case, followed by the N-R case and then the N-N case.

As for borrower surplus, when comparing CS^{nr} with CS^{rr}, notice that (1) the number of borrowers who are approved in the R-R case is larger than in the N-R case; (2) the number of borrowers who are rejected in the R-R case is smaller than in the N-R case; (3) the interest rate set in the R-R case is on average lower than in the N-R case. (This is the case because $F^{rr}(b) > F_1(b), F^{rr}(b) > F_2(b), \forall b \in [0, \bar{b}]$.) In other words, in the R-R case, more borrowers are approved, fewer borrowers are rejected, and the interest rates set by the lenders are lower compared with the N-R case. Consequently, borrowers' welfare is higher in the R-R case than in the N-R

[23] This value is always positive. Rewrite the expression as $[\theta\bar{b} + (1-\theta)(\bar{b}-1)] + [\theta(M_h - m - \bar{b}) + (1-\theta)(M_l - m - \bar{b})]$. The first part is greater than 0 under Assumption 1, and the second part is greater than 0 following the first inequality in Assumption 2. To see the logic of the latter, recall that the first inequality of Assumption 2 ensures that borrowers will always follow the intermediary's predictions to apply to lenders regardless of the interest rate and revealing decisions the lenders make. Imagine that in the R-R case, a borrower receives a positive type prediction from Lender 1 but a negative prediction from Lender 2. Her expected utility from applying to Lender 1 is exactly the same as the second part of the expression above, and this utility has to be greater than 0 to make it consistent with Assumption 2

case. The borrower surplus in the N-N case, CS^{nn}, is the smallest among the three. The reason is that there are a large number of borrowers who make sub-optimal applying decisions because of the large market friction. Even when one lender sets a higher interest rate in the N-R case than it would do in the N-N case, moving from N-N to N-R will still benefit borrowers.

5 Mandating Algorithm Revelation

As discussed previously, both borrower surplus and social surplus are the highest in the R-R case. However, when P_c is high, only one lender would want to reveal the algorithm. Does it imply that mandatory disclosure could be socially desirable from the policy makers' perspective? In this extension, we tackle this question by showing the other side of the story – the effect of mandatory revealing on lenders' incentive to improve their algorithms. While in the previous analysis we assume that the lenders' algorithm accuracy P_b is exogenous, in this extension we consider it a decision that lenders have to make prior to Stage 1 in the current model.

In this extension, we study two scenarios: a mandatory scenario where lenders are required to reveal the algorithms, and a voluntary scenario where such requirements are absent. We will focus on the parameter range where the two lenders make asymmetry revealing decisions voluntarily, because outside of this range, mandatory revealing is redundant. In other words, we assume that $P_b > P_b^0$, where P_b^0 is defined in Proposition 1. Under this assumption, both lenders will reveal their algorithms when revealing is mandatory, while only one lender will reveal its algorithm without such a policy intervention. We assume there are advances in technology that could help lenders increase their algorithms' accuracy from P_b to P_b^* without any cost, and study lenders' incentives to take on this opportunity in both scenarios. Intuitively, it may appear that a lender should always improve its algorithm's accuracy when such improvement is cost-less. However, we will demonstrate that is not the case. Below, we start with the mandatory scenario. The lenders' equilibrium strategies in upgrading their algorithms in this scenario are summarized in Proposition 4.

Proposition 4 *In the mandatory revealing scenario, there is a unique symmetric mix-strategy equilibrium in upgrading screening algorithms, where each lender chooses to upgrade its algorithm with probability* $P^M = \frac{1-P_b}{P_b^*}$.

The proof of Proposition 4 can be found in Appendix B.8. An interesting observation is that P^M decreases in P_b and P_b^*, which means that when lenders' algorithms' accuracy is relatively high, lenders will have little incentive to further increase it. The reason is that when lenders' algorithms' accuracy is high, competition is especially intense in the "both reveal" case. Any further increase in algorithms' accuracy would make the common segment even more profitable compared to the captive segment thus further intensifying the competition on the interest rate, which would drive down lenders' equilibrium profits. Since the mandatory revealing policy is only relevant in the case where P_b is relatively large (i.e., $P_b > P_b^0$), lenders will have less incentive to improve their screening algorithms if revealing is made mandatory.

We next consider the voluntary scenario. The lenders' equilibrium strategies in upgrading their algorithm in the voluntary scenario are shown in Proposition 5

Proposition 5 *In the voluntary revealing scenario, the revealing lender always chooses to upgrade its algorithm, while the non-revealing lender chooses to upgrade the algorithm if $\theta \leq \theta^0$ or $P_c \leq P_c^0$, where*

$$\begin{cases} \theta^0 = & \frac{2P_bP_b^*P_c - P_bP_b^* - 2P_bP_c + P_b + 2P_b^{*2}P_c - P_b^{*2} - 3P_b^*P_c + 2P_b^* + P_c}{-2P_bP_c + P_b + P_b^{*2}(2P_c - 1)(\bar{b}+1) + P_b^*((\bar{b}+1)(2P_bP_c - P_b - P_c) + 2 - 2P_c) + P_c} \\ P_c^0 = & \frac{P_b + P_b^*}{2P_b + 2P_b^* - 1} \end{cases}$$

The proof of Proposition 5 can be found in Appendix B.9. Proposition 5 says that in the voluntary scenario, both lenders will adopt the higher accuracy algorithm unless both θ and P_c are high, in which case only the revealing lender chooses to upgrade the algorithm. We show in the appendix that upgrading the algorithm is a dominant strategy for the revealing lender. The intuition is as follows: In Section 4, we have shown that in the N-R case, the non-revealing lender could incorporate a more accurate signal from its rival than the revealing lender, thus it chooses to focus on capturing the common segment while the revealing lender chooses to focus on its captive segment, softening the competition. If the revealing lender increases its algorithm's accuracy, the signal that the non-revealing lender could incorporate will be even more accurate, and on the other hand the revealing lender can also better screen borrowers in its captive segment, and therefore, competition will be further softened. By contrast, the non-revealing lender has to consider the relative effect sizes of improved screening ability and increased competition due to an increase in its algorithm's accuracy. When $\theta \leq \theta^0$, the market is risky in the sense that there is large fraction of L type borrowers, and there are significant benefits from a better screening ability. Similarly, when $P_c \leq P_c^0$, the direct impact of an increase in the accuracy of the non-revealing lender's algorithm on intensifying the competition is fairly weak. As a result, the non-revealing lender would choose to increase its algorithm's accuracy only when $\theta \leq \theta^0$ or $P_c \leq P_c^0$.

Our analysis above suggests that the implications of policy makers' decision on regulating algorithm revelation is not as straightforward as it appears intuitively. Since increasing the accuracy of screening algorithms can help allocate fund to more creditworthy borrowers, failing to do so may potentially hurt borrower surplus. The following proposition specifies a sufficient condition under which mandatory algorithm revealing will hurt consumer surplus.

Proposition 6 *Mandatory algorithm revealing will hurt borrower surplus if* $\theta \leq \theta^0$ *and* $P_c > \underline{P_c}$*, Where* θ^0 *is defined in Proposition 5, and*

$$\begin{cases} \underline{P_c} = & \frac{-B+\sqrt{B^2-4AC}}{2A} \\ A = & 2\bar{b}P_b^{*2}\theta - \bar{b}P_b^{*}\theta + 2P_b^{*2}\theta - 2P_b^{*2} - 3P_b^{*}\theta + 3P_b^{*} + \theta - 1 \\ B = & m\theta + (P_b^{*2} - P_b)\left(\bar{b}\theta + 3m - 2M_h\theta + 2M_l\theta - 2M_l - \theta + 1\right) + P_b^{*}(m - 2m\theta) \\ C = & 1 - M_h\theta + 2P_b{}^2\left(-\bar{b}\theta - \theta + 1\right) + 2P_b\left(\bar{b}\theta + \theta - 1\right) - \theta \\ & +P_b^{*2}\left(-\bar{b}\theta - 2m + M_h\theta - M_l\theta + M_l\right) + 2P_b^{*}(m + M_l\theta - M_l) \\ & +(-m + M_h\theta - M_l\theta + M_l + \theta - 1)(P_b{}^4 - 2P_b{}^3 + P_b{}^2)/P_b^{*2} \end{cases}$$

Proposition 6 suggests that mandating revealing of screening algorithms could hurt consumer surplus when P_c is high but θ is low. First, the condition on θ is to ensure that the lenders have an incentive to increase accuracy in the N-R case, according to Proposition 5. The intuition behind the P_c related condition will become clear once we layout the upside and downside of such a mandatory revealing policy. On the upside, mandatory revealing will reduce market friction by eliminating borrowers' uncertainties. On the downside, it will impede lenders from adopting more accurate algorithms. The benefit from the upside is decreasing in P_c since the benefit of revealing the algorithm is low if the intermediary can already provide accurate odds of approval. Moreover, the loss due to the downside is increasing in P_c: when predictions the intermediary provides are more accurate (i.e., P_c is higher), borrowers sub-optimal applying behavior is reduced, and the benefit of lenders adopting more advanced algorithms becomes larger since any improvement in screening accuracy will directly improve the chances that credit is allocated to the more deserving H type borrowers.

6 Conclusions

6.1 Summary of Results

The use of ML algorithms and data driven decision making by financial lenders seems win-win for both the lenders and the borrowers – lenders are able to better screen borrowers for their credit worthiness and deserving borrowers are more likely to get approved for credit (Fu et al., 2021). However, as lenders get better at screening the borrowers, the later continue to face considerable uncertainty in their chances of getting approved for credit by a lender (Citron and Pasquale, 2014, Fu et al., 2020).

In this paper, we investigate forces that affect lenders' decisions to reveal their algorithms to the intermediary or provide approval odds to the borrowers. We provide explanation for the empirical observation that only a few lenders have revealed their algorithms to intermediaries or provide approval odds via pre-approval tools to borrowers. We show that lenders use asymmetric revealing of approval odds strategically to soften the competition. Asymmetric revealing allows lenders to focus on different segments of borrowers. Our results show that asymmetric revealing of the algorithms occurs when the lenders' algorithms' accuracy is high. Further, asymmetric revealing equilibrium can also be sustained with a relatively lower algorithmic accuracy when the accuracy of the intermediary's algorithm is high or the market is very risky due to the presence of a large portion of borrowers with low credit worthiness.

The asymmetric revealing equilibrium leads to endogenous product differentiation. The revealing lender focuses more on its captive segment and on average charges a higher interest rate. In contrast, the non-revealing lender focuses more on the common segment and charges a low interest rate. The revealing lender receives a greater payoff compared to the non-revealing lender.

The borrower surplus and total surplus are the highest in the case where both lenders reveal their algorithms and the lowest in the case where neither lender reveals its algorithm. We further find that a mandatory policy that requires lenders to provide accurate approval odds generated by their algorithms may not necessarily improve borrower surplus. This is because lenders have less incentive to improve their algorithms' accuracy when algorithm revelation is mandated compared to when it is voluntary.

6.2 Managerial Implications

Our analysis provides a number of insights that may guide practice and future research on algorithm revelation (or provision of approval odds more specifically) in the financial lending scenario. First, if algorithm revelation by a lender can intensify competition in one case, but soften competition in another, it implies that provision of personalized approval odds is a complex decision. Symmetric revelation always intensifies the competition whereas asymmetric revelation softens it. However, besides competition lenders should also consider the potential market expansion effect. While making these revelation decisions it is important for us to ascertain the mediators for these two type of effects. Our model suggests that these mediators are most likely related to the accuracy of the lenders' algorithms, the accuracy of the intermediary's algorithm and the composition of the market (i.e. fraction of H type borrowers). Second, our results show that at least one lender should always provide approval odds to the borrowers under all conditions. And both lenders should choose to reveal their algorithm to the intermediary when specific conditions are met. Given the increasing calls for algorithmic transparency, the fact that market forces can lead to partial transparency by at least one lender and in some cases by both the lenders is quite promising.

Third, the competitive implications of algorithm revelation to intermediary by both lenders are more severe in the markets where lenders' algorithms are quite accurate. For example, in these markets if both the lenders provide approval odds, the borrowers would face no uncertainty. Further, due to high accuracy of the lenders' algorithm, the common segment would be even more profitable to the lenders as the chances of type II errors in this segment are minimal when both lenders provide approval odds. Whereas the captive segment would become less profitable as the chances of type II errors in this segment increase when both lenders provide approval odds. Hence, if both lenders provide approval odds, they could get into self debilitating competition on interest rates. As a result, a lender should avoid providing personalized approval odds if its competitor is already providing them when their algorithms accuracy is high.

Fourth, the prior belief of the borrowers regarding their approval odds can moderate the competitive and market expansion effects of provision of approval odds by lenders. In our model, the prior belief of the borrowers is captured by the accuracy of the intermediary's algorithm. When the intermediary's algorithm is inaccurate, many borrowers would have inaccurate beliefs and make sub-optimal decisions. The market expansion effect from provision of approval odds is stronger when the intermediary's algorithm is less accurate. Hence, both the lenders can be better off by revealing their algorithms when intermediary's accuracy is low even when they have relatively higher algorithm accuracy. Fifth, when the market is composed of many H type borrowers, any sub-optimal decision by borrowers due to uncertainty regarding approval odds is costlier to the lenders. Hence, both the lenders should reveal their algorithm to the intermediary when the fraction of H type borrowers in the market is high even for relatively high accuracy of lenders' algorithms.

Sixth, under conditions where only one lender should reveal its algorithm, our results show that the revealing lender is better off than the non-revealing lender. Surprisingly, the revealing lender is not better off at the cost of the non-revealing lender. The asymmetric revealing generates asymmetry in both the captive and common segments faced by the two lenders. As a result, they target different segment of borrowers softening the competition. Hence, when conditions for asymmetric revealing equilibrium are met, a lender should take the opportunity to be the first mover and reveal its algorithm if its competitor has not done so yet. At the same time, the non-revealing lender should not worry that it will be hurt because of the asymmetric action of the competing lender. This result highlights the role of asymmetric revelation as a new mechanism for endogeneously creating product differentiation to soften competition among financial lenders (Tirole and Jean, 1988).

Finally, financial products with similar non price features would be equally profitable to the borrowers and, hence, one would think lenders would not be able to charge different prices for their products. However, lenders' algorithms' accuracy allows lender differentiation. When the lenders' algorithms are less accurate they will incorrectly classify many borrowers. On first look this may appear all bad for the lenders. However, there is a positive aspect to low accuracy. When lender algorithms are less accurate, they would differ in their predictions. Different approval odds then lead to borrowers having relatively different preferences for the lenders. The difference in preferences for the two lenders by different customers leads to softening of competition. While there are advantages to an accurate algorithm (less loss due to misclassification), lenders should carefully consider its impact on competition as well.

6.3 Implications for Public Policy

Our analysis suggests that from a social-welfare perspective, lenders' algorithm revealing behavior increases the efficiency of credit markets, because it helps borrowers avoid non-optimal applying decisions which are socially wasteful. However, this does not mean that policy makers should always compel lenders to reveal pre-approval odds or the screening algorithms to intermediaries. Policy makers are suggested to understand the strategic reasons as to why lenders do not want to reveal such information, and be aware of the potential impact of mandatory algorithmic revealing on lender competition. Our analysis shows that when lenders are mandated to reveal their algorithm, they will have less incentive to invest in screening technologies. When policy makers are considering regulations on compulsory information revealing, they are suggested to consider how such regulations may affect subsequent competition and to keep in mind that such regulations may reduce lenders' incentive to invest in algorithmic screening technologies, which in the long term may not help allocate credit to more creditworthy borrowers. Our analysis suggests that in general, when the market is relatively risky and the intermediary's algorithm is relatively accurate, policy makers should be especially concerned with mandatory algorithmic revealing since in those cases, the benefit of reduced market friction will be overshadowed by the loss due to lack of adoption of better screening technologies by lenders.

6.4 Limitations

Our paper attempts to model the most important forces in the algorithmic lending and revelation context. This allows us to provide a deep and robust analysis on the algorithm revelation strategies. However, there are many forces and strategic actions which we do not capture which open up interesting avenues for future researchers as the algorithmic lending market further evolves. First, we model the intermediary as non-strategic. It is reasonable because even the largest intermediaries at present have limited market power to influence a large lender's decision. However, as these intermediaries gain market power they could strategically influence lenders' algorithm revelation decisions. Second, we do not model algorithmic gaming by borrowers. Borrowers only get to know whether they will be approved or not by a lender, as a result gaming of the algorithm by borrowers via changing their attributes is going to be very limited in scope if any. Future research can consider how algorithm revelation strategies would change if algorithm revelation could lead to gaming by borrowers which may hurt the predictive performance of the algorithm. Third, we ignore other avenues through which borrowers could apply to a lender. Intermediary platforms are only one such avenue. Borrowers can also directly apply to lenders through their websites. This is important because, lenders do not have to pay the intermediary any fee if a borrower applies directly to the lender but have to pay a fee if the borrower applies via an intermediary. It would be interesting to investigate how algorithm revelation strategies of lenders would change if one considers these multiple avenues through which borrowers could apply.

Finally, the financial lending market is quite complex with a large number of competitors. For simplicity we only consider a duopoly competition. However, as the emergence of common and captive segments is a phenomenon that could be observed even in the presence of more than two lenders in the market, we are inclined to believe that the findings from our model are easily generalizable to scenarios where oligopolistic competition is more prevalent as compared to duopolistic competition.

Despite the limitations discussed earlier, we believe our paper provides an insightful and robust analysis on algorithmic lending and specifically on the algorithm revelation strategies of lenders. In conclusion, our paper is among the first to examine the effect of competition between lenders on their decision to reveal approval odds to borrowers. We provide novel explanations for the observed asymmetric information revealing behavior in credit markets and recognize a new dimension of strategy to leverage competition, which has not been captured by previous models of strategic information revealing and lender competition.

References

A. Andriotis. Looking past a credit score. *WSJ*, Mar 2014.

O. Board. Competition and disclosure. *The Journal of Industrial Economics*, 57(1):197–213, 2009.

J. Bouckaert and H. Degryse. Softening competition by inducing switching in credit markets. *The Journal of Industrial Economics*, 52(1):27–52, 2004.

T. Broecker. Credit-worthiness tests and interbank competition. *Econometrica*, pages 429–452, 1990a.

T. Broecker. Credit-worthiness tests and inter-bank competition. *Econometrica*, 58(2):429–452, 1990b.

D. K. Citron and F. Pasquale. The scored society: Due process for automated predictions. *Wash. L. Rev.*, 89:1, 2014.

N. Diakopoulos. Accountability in algorithmic decision making. *CACM*, 59(2):56–62, 2016.

Experian. Survey findings: How do consumers feel about credit cards? *Experian*, 2020a.

Experian. How long do hard inquiries stay on your credit report? *Experian*, Sep 2020b.

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

R. Fu, Y. Huang, and P. V. Singh. Artificial intelligence and algorithmic bias: Source, detection, mitigation, and implications. In *Pushing the Boundaries: Frontiers in Impactful OR/OM Research*, pages 39–63. INFORMS, 2020.

R. Fu, Y. Huang, and P. V. Singh. Crowd, lending, machine, and bias. *Information Systems Research*, 31(1):72–92, 2021.

S. J. Grossman. The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, 24:461–483, 1981.

Z. J. Gu and Y. Xie. Facilitating fit revelation in the competitive market. *Management Science*, 59(5):1196–1212, 2013.

L. Guo and Y. Zhao. Voluntary quality disclosure and market interaction. *Marketing Science*, 28 (3):488–501, 2009.

N. Haghtalab, N. Immorlica, B. Lucier, and J. Wang. Maximizing welfare with incentive-aware evaluation mechanisms. In *29th International Joint Conference on Artificial Intelligence*, 2020.

M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 111–122, 2016.

R. Hauswald and R. Marquez. Information technology and financial services competition. *Review of Financial Studies*, 16:921–948, 02 2003.

R. Hauswald and R. Marquez. Competition and Strategic Information Acquisition in Credit Markets. *Review of Financial Studies*, 19(3):967–1000, 2006.

J. Kleinberg and M. Raghavan. How do classifiers induce agents to invest effort strategically? In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 825–844, 2019.

J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu. Accountable algorithms. *University of Pennsylvania Law Review*, 165, 2017.

D. Kuksov and Y. Lin. Information provision in a vertically differentiated competitive marketplace. *Marketing Science*, 29(1):122–138, 2010.

D. Levin, J. Peck, and L. Ye. Quality disclosure and competition. *Journal of Industrial Economics*, 57(1):167–196, 2009.

M. Lockert. What your credit karma approval odds really mean? *Credit Karma*, Sep 2020.

I. Lunden. Intuit confirms that it is buying credit karma for $7.1b in cash and stock. *Techcrunch*, Feb 2020.

R. Meir, A. Procaccia, and J. Rosenschein. Algorithms for strategy-proof classification. *Journal of Artificial Intelligence*, 186:123–156, 07 2012.

P. R. Milgrom. Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics*, 12(2):380–391, 1981.

F. A. Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press, 2015.

N. Popper and M. Merced. Intuit to buy credit karma to create financial data giant. *The New York Times*, Feb 2020.

M. Ruckes. Bank competition and credit standards. *The Review of Financial Studies*, 17(4): 1073–1102, 2004.

P. Rudegeair and A. Andriotis. Lenders share their underwriting secrets with credit karma. *WSJ*, Oct 2018.

S. Shaffer. The winner's curse in banking. *Journal of Financial Intermediation*, 7(4):359–392, 1998. ISSN 1042-9573.

S. A. Sharpe. Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships. *The Journal of Finance*, 45(4):1069–1087, 1990.

J. E. Stiglitz. The theory of "screening," education, and the distribution of income. *The American Economic Review*, 65(3):283–300, 1975.

J. Tirole and T. Jean. *The Theory of Industrial Organization.* MIT press, 1988.

H. R. Varian. A Model of Sales. *American Economic Review*, 70(4):651–659, 1980.

J. M. Villas-Boas and U. Schmidt-Mohr. Oligopoly with asymmetric information: Differentiation in credit markets. *The RAND Journal of Economics*, 30(3):375–396, 1999.

Q. Wang, Y. Huang, S. Jasin, and P. V. Singh. Algorithmic transparency with strategic users, 2020.

A Notation Summary

Notation	Meaning
P_b	Accuracy of lenders' algorithms
P_c	Accuracy of intermediary's algorithm
θ	Percentage of H type borrowers in the population
m	Borrowers' cost of applying
M_h	Utility of being approved for a H type borrower
M_l	Utility of being approved for a L type borrower
b	Interest rate set by the lenders
$\overline{b}$	Maximum interest rate can be set by the lenders
$Y_{k,b}^{i}$	Lender i's prediction of borrower k's type
$Y_{k,c}^{i}$	Intermediary's guess of lender i's prediction of borrower k's type
$N_{hh}^{rr}(N_{hh}^{nn})$	The number of borrowers in a lender's common segment in the R-R (N-N) case
$N_{hl}^{rr}(N_{hl}^{nn})$	The number of borrowers in a lender's captive segment in the R-R (N-N) case
$V_{hh}^{rr}(V_{hh}^{nn})$	The percentage of H type borrowers in a lender's common segment in the R-R (N-N) case
$V_{hl}^{rr}(V_{hl}^{nn})$	The percentage of H type borrowers in a lender's captive segment in the R-R (N-N) case
$N_{hh}^{1}(N_{hh}^{2})$	The number of borrowers in Lender 1(2)'s common segment in the N-R case
$N_{hl}^{1}(N_{lh}^{2})$	The number of borrowers in Lender 1(2)'s captive segment in the N-R case
$V_{hh}^{1}(V_{hh}^{1})$	The percentage of H type borrowers in Lender 1(2)'s common segment in the N-R case
$V_{hl}^{1}(V_{lh}^{2})$	The percentage of H type borrowers in Lender 1(2)'s captive segment in the N-R case

Table 1: Notation summary

B Mathematical Appendix

B.1 Derivation of Conditions in Assumption 2

This assumption is to ensure borrowers will always follow the intermediary platform's suggestion. Specifically, if a borrower is predicted as H type by the intermediary for either lender, she will apply to the lender even under the least favorable conditions (i.e., the interest rate is $\overline{b}$ and the other lender predicts her as L type[24]). We want the above argument to be true for all possible

[24]The two lenders' algorithms are independent given her type but not independent without conditioning on her type. That is, without knowing her true type, the prediction from the other lender (or intermediary) will influence her posterior belief on whether she will be approved or not by the focal lender.

combinations of predictions that a borrower may receive in all possible sub-games. In the R-R sub-game, this is satisfied trivially, since being predicted as H type by a lender will ensure a borrower being approved the lender. In the N-R sub-game, the extra information from the other lender's (or the intermediary's) predictions is stronger than that in the N-N sub-game, since in the N-N sub-game all signals are from the intermediary's noisy predictions. This means if the above argument holds for the N-R case, it must also hold for the N-N case. Consequently, we only need to make sure that in the N-R case, if a borrower receives H from the non-revealing lender, she will apply to this lender even when the revealing lender predicts her as L type.[25] According to numbers of borrowers in each segments shown in Figure 3, given that a borrower is predicted as H type by the intermediary for the non-revealing lender and as L type by the revealing lender, her beliefs are specified as follows: with probability $\frac{P_b(1-P_b)P_c}{D}$ she is a H type and will be approved thus gets utility $M_h - \bar{b} - m$, with probability $\frac{P_c P_b(1-P_b)}{D}$ she is a L type and will be approved thus gets utility $M_l - \bar{b} - m$, with probability $\frac{(1-P_b)^2(1-P_c)+{P_b}^2(1-P_c)}{D}$ she will be rejected and gets utility $-m$, where $D = P_b(1-P_b)P_c + P_c P_b(1-P_b) + (1-P_b)^2(1-P_c) + {P_b}^2(1-P_c)$ is a common denominator that ensures all probabilities add up to one. By calculating the expected utility from the probabilities above and making it greater than 0, we can get the first condition in Assumption 2.

As for the second condition, the purpose is to ensure that if a borrower is predicted as L type by the intermediary for a lender, she will not apply to this lender even under the most favorable conditions (i.e., the interest rate is as low as 0 and the other lender predicts her as H type. Following a similar procedure as that for the first condition, we can calculate the expected utility and set it to be smaller than 0, which gives us the second condition. The details are omitted here.

B.2 Proof of Lemma 1

Let $F^{rr}(b)$ denote the CDF of both lender's equilibrium mixed strategy for the pricing decision. Facing Lender 2's mixed strategy characterized by $F^{rr}(b)$, Lender 1's expected profit is:

$$\begin{aligned}\mathbb{E}[\Pi_1^{rr}](b) =& (1 - F^{rr}(b))(N_{hh}^{rr}(V_{hh}^{rr}b - 1 + V_{hh}^{rr}) + N_{hl}^{rr}(V_{hl}^{rr}b - 1 + V_{hl}^{rr})) \\ &+ F^{rr}(N_{hl}^{rr}(V_{hl}^{rr}b - 1 + V_{hl}^{rr}))\end{aligned} \tag{B.1}$$

The first line in Equation (B.1) corresponds to the case where Lender 1 sets a lower interest rate b than Lender 2 and thus Lender 1 gets segments A_{hh}^{rr} and A_{hl}^{rr}. The second line corresponds to the case where Lender 1 sets a higher b than Lender 2 and thus Lender 1 gets only segment A_{hl}^{rr}.

To make sure that Lender 1 is using the same mixed strategy characterized by $F^{rr}(b)$, Lender

[25] In fact borrowers do not know which lender has revealed its algorithm, and therefore, the above argument will form a sufficient but not necessary condition under which borrowers will always follow the intermediary's suggestions.

1 has to be indifferent in setting any b on the support of $F^{rr}(b)$, mathematically,

$$\mathbb{E}[\Pi_1^{rr}](b) = k \tag{B.2}$$

where k is a constant. The maximum b that can be set by the lenders is $\bar{b}$, i.e.,

$$F^{rr}(\bar{b}) = 1 \tag{B.3}$$

Using Equations (B.1), (B.2), and (B.3), we can solve for $F^{rr}(b)$. The mixed strategy equilibrium in this sub-game is as shown in Lemma 1.

B.3 Proof of Lemma 2

Facing Lender 2's pricing strategy characterized by the CDF $F^{nn}(b)$, Lender 1's expected profit is:

$$\begin{aligned}\mathbb{E}[\Pi_1](b) =& (1 - F^{nn}(b))(N_{hh}^{nn}(V_{hh}^{nn}b - (1 - V_{hh}^{nn}))b) + N_{hl}^{nn}(V_{hl}^{nn}b - (1 - V_{hl}^{nn}))) \\ &+ F^{nn}(b)N_{hl}^{nn}(V_{hl}^{nn}b - (1 - V_{hl}^{nn}))\end{aligned} \tag{B.4}$$

The first line corresponds to the case where Lender 1 sets a lower b than Lender 2, and thus Lender 1 gets segments A_{hh}^{nn} and A_{hl}^{nn}. The second line corresponds to the case where Lender 1 sets a higher b than Lender 2, and thus Lender 1 only gets the A_{hl}^{nn} segment.

In a symmetric equilibrium, Lender 1 is using the same mixed strategy characterized by $F^{nn}(b)$, therefore, Lender 1 has to be indifferent in setting any b on the support of $F^{nn}(b)$. Mathematically,

$$\mathbb{E}[\Pi_1](b) = k \tag{B.5}$$

where k is a constant. The maximum b that can be set by the lenders is $\bar{b}$, i.e.,

$$F^{nn}(\bar{b}) = 1 \tag{B.6}$$

Combining Equations (B.4), (B.5), and (B.6), we can solve for $F^{nn}(b)$. The mixed strategy equilibrium in this sub-game is summarized in Lemma 2.

B.4 Proof of Lemma 3

If $b_1 > b_2$, Lender 1 will only get the A_{hl}^1 segment, and Lender 2 will get the A_{hh}^2 and A_{lh}^2 segments. If $b_1 < b_2$, Lender 1 will get the A_{hh}^1 and A_{hl}^1 segments, and Lender 2 will get only the A_{lh}^2 segment of borrowers. We do not need to consider the case where $b_1 = b_2$ since this will happen with probability 0 under the mixed strategy setting.[26]

Lender 1's expected payoff can be written as:

$$
\begin{aligned}
\mathbb{E}[\Pi_1^{nr}] = &(1 - F_2(b)) \left(N_{hh}^1 \left(V_{hh}^1 b - 1 + V_{hh}^1\right) + N_{hl}^1 \left(V_{hl}^1 b - 1 + V_{hl}^1\right)\right) \\
&+ F_2(b) N_{hl}^1 \left(V_{hl}^1 b - 1 + V_{hl}^1\right)
\end{aligned}
\tag{B.7}
$$

Lender 2' expected payoff can be written as:

$$
\begin{aligned}
\mathbb{E}[\Pi_2^{nr}] = &(1 - F_1(b)) \left(N_{hh}^2 \left(V_{hh}^2 b - 1 + V_{hh}^2\right) + N_{lh}^2 \left(V_{lh}^2 b - 1 + V_{lh}^2\right)\right)) \\
&+ F_1(b) N_{lh}^2 \left(V_{lh}^2 b - 1 + V_{lh}^2\right)
\end{aligned}
\tag{B.8}
$$

Facing the competitor's strategy, Lender 1 (Lender 2) should be indifferent in any b in the support of $F_1(b)$ and $F_2(b)$. Mathematically:

$$\mathbb{E}[\Pi_1^{nr}](b) = k_1 \tag{B.9}$$

$$\mathbb{E}[\Pi_2^{nr}](b) = k_2 \tag{B.10}$$

Additionally, consider the maximum b that can be set by both firm is $\bar{b}$, we have

$$F_1(\bar{b}) = 1 \tag{B.11}$$

$$F_2(\bar{b}) = 1 \tag{B.12}$$

Another condition that need to be satisfied is that the CDF $F_1(b)$ should be greater than or equal to 0 in any region where a positive probability is assigned to $F_2(b)$. Similarly, the CDF $F_2(b)$ should be greater than or equal to 0 in any region where a positive probability is assigned to $F_1(b)$. The logic is that setting $b_1 = min(b_2)$ ensures that Lender 1 gets the common segment with probability 1, so Lender 1 has no incentive to further reduce b_1. Similarly, setting $b_2 = min(b_1)$ ensures that Lender 2 gets the common segment with probability 1 so Lender 2 has no incentive to further reduce b_2. This requires:

$$F_1(\underline{b}) = 0 \tag{B.13}$$

$$F_2(\underline{b}) = 0 \tag{B.14}$$

Solving Equations (B.7) - (B.14), we get each lender's equilibrium strategy in the N-R sub-game. The mixed strategy equilibrium in this sub-game is summarized in Lemma 3.

[26] It can be shown that there is no mass point on b_2's distribution.

B.5 Proof of Proposition 1

We first prove that $\Pi_{nr}^1 > \Pi_{nn}$. Expand and rewrite Π_{nn} and Π_{nr}^1 as

$$\Pi_{nn} = -P_c\left(\bar{b}P_b\theta\left(P_b\left(P_c-1\right)+P_c\left(P_b-1\right)\right)+P_bP_c\left(P_b-1\right)\left(\theta-1\right)+\left(P_b-1\right)^2\left(P_c-1\right)\left(\theta-1\right)\right)$$

$$\Pi_{nr}^1 = -\bar{b}P_b\theta\left(P_b\left(P_c-1\right)+P_c\left(P_b-1\right)\right)-P_bP_c\left(P_b-1\right)\left(\theta-1\right)-\left(P_b-1\right)^2\left(P_c-1\right)\left(\theta-1\right)$$

It is straightforward to see that $\Pi_{nr}^1 > P_c\Pi_{nr}^1 = \Pi_{nn}$.

We next prove that $\Pi_{nr}^2 > \Pi_{rr}$ if and only if $P_b > P_b^0$: Expand and rewrite the two terms as:

$$\Pi_{nr}^2 = -P_c\left(\bar{b}P_b\theta\left(P_b\left(P_c-1\right)+P_c\left(P_b-1\right)\right)+P_bP_c\left(P_b-1\right)\left(\theta-1\right)+\left(P_b-1\right)^2\left(P_c-1\right)\left(\theta-1\right)\right)$$

$$\Pi_{rr} = P_b\left(-\bar{b}P_b\theta+\bar{b}\theta-P_b\theta+P_b+\theta-1\right)$$

We first prove that the two functions, Π_{nr}^2 and Π_{rr} cross at most once. Take the first derivative w.r.t. P_b. The goal is to show $\frac{\partial\Pi_{nr}^2}{\partial P_b} \geq \frac{\partial\Pi_{rr}}{\partial P_b}$ for any value of P_b.

$$\frac{\partial\Pi_{nr}^2}{\partial P_b} = P_c\left(-4\bar{b}P_bP_c\theta+2\bar{b}P_b\theta+\bar{b}P_c\theta-4P_bP_c\theta+4P_bP_c+2P_b\theta-2P_b+3P_c\theta-3P_c-2\theta+2\right)$$

$$\frac{\partial\Pi_{rr}}{\partial P_b} = -2\bar{b}P_b\theta+\bar{b}\theta-2P_b\theta+2P_b+\theta-1$$

Since it is difficult to compare these two values directly, we go one step further and take the second derivative.

$$\frac{\partial^2\Pi_{nr}^2}{\partial P_b^2} = 2P_c\left(-2\bar{b}P_c\theta+\bar{b}\theta-2P_c\theta+2P_c+\theta-1\right)$$

$$\frac{\partial^2\Pi_{rr}}{\partial P_b^2} = -2\bar{b}\theta-2\theta+2$$

We next prove that $\frac{\partial^2\Pi_{nr}^2}{\partial P_b^2} \leq \frac{\partial^2\Pi_{rr}}{\partial P_b^2}$: rewrite $\frac{\partial^2\Pi_{nr}^2}{\partial P_b^2} = 2P_c(2P_c-1)(1-\theta-\bar{b})$, so we have $\frac{\partial^2\Pi_{nr}^2}{\partial P_b^2}/\frac{\partial^2\Pi_{rr}}{\partial P_b^2} = P_c(2P_c-1) \leq 1$. Now since $\frac{\partial^2\Pi_{nr}^2}{\partial P_b^2} \leq \frac{\partial^2\Pi_{rr}}{\partial P_b^2}$, we only need to show $\frac{\partial\Pi_{nr}^2}{\partial P_b} \geq \frac{\partial\Pi_{rr}}{\partial P_b}$ holds at the right most point, that is, $\frac{\partial\Pi_{nr}^2}{\partial P_b}|_{P_b=1} \leq \frac{\partial\Pi_{rr}}{\partial P_b}|_{P_b=1}$. Expand and rewrite these two terms:

$$\frac{\partial\Pi_{nr}^2}{\partial P_b}|_{P_b=1} = P_c(-3\bar{b}P_c\theta+2\bar{b}\theta-P_c\theta+P_c)$$

$$\frac{\partial\Pi_{rr}}{\partial P_b}|_{P_b=1} = -\bar{b}\theta-\theta+1$$

The latter is invariant in P_c while the former is a quadratic function of P_c. Specifically, the former first increases and then decreases in P_c, and therefore, the minimum value is reached at either $P_c = 0.5$ or at $P_c = 1$ Since $\frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1,P_c=0.5} = (\bar{b}\theta - \theta + 1)/4 > 0 > \frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1}$ and $\frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1,P_c=1} = -2\bar{b}\theta - 2\theta + 2 = \frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1}$, $\frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1} \geq \frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1}$ for any value of P_c.

Combine the facts that $\frac{\partial^2 \Pi_{nr}^2}{\partial P_b^2} \leq \frac{\partial^2 \Pi_{rr}}{\partial P_b^2}$ and $\frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1} \geq \frac{\partial \Pi_{nr}^2}{\partial P_b}|_{P_b=1}$, we get $\frac{\partial \Pi_{nr}^2}{\partial P_b} \geq \frac{\partial \Pi_{nr}^2}{\partial P_b}, \forall P_b$. This is sufficient to ensure that the two curves intersect at most once. We then plug the expression of P_b^0 into Π_{nr}^2 and Π_{rr} and get

$$\Pi_{nr}^2|_{P_b=P_b^0} = \Pi_{rr}|_{P_b=P_b^0}$$

Thus we have $\Pi_{nr}^2 > \Pi_{rr}$ if and only if $P_b > P_b^0$

B.6 Proof of Proposition 2

We first prove that $\frac{\partial P_b^0}{\partial P_c} < 0$. Rewrite the expression of P_b^0 by collecting the P_c terms as:

$$P_b^0 = \frac{P_c\left(\bar{b}\theta + 3\theta - 3\right) + \bar{b}\theta + \theta - 1 + \sqrt{{P_c}^2\left(\bar{b}\theta - \theta + 1\right)^2 + 2P_c\left(\bar{b}\theta + \theta - 1\right)^2 + \left(\bar{b}\theta + \theta - 1\right)^2}}{4P_c\left(\bar{b}\theta + \theta - 1\right) + 2\left(\bar{b}\theta + \theta - 1\right)} \tag{B.15}$$

Let $x = \bar{b}\theta - \theta + 1$ and $y = \bar{b}\theta + \theta - 1$. Rewrite Equation B.15 as:

$$P_b^0 = \frac{\frac{P_c(2y-x)}{y} + 1 + \sqrt{\frac{P_c^2 x^2}{y^2} + 2P_c + 1}}{2(2P_c + 1)} \tag{B.16}$$

Now take the derivative w.r.t. P_c. Note that P_c is not contained in either x or y.

$$\frac{\partial P_b^0}{\partial P_c} = \frac{x^2 P_c - x\sqrt{x^2{P_c}^2 + 2y^2 P_c + y^2} - 2y^2 P_c - y^2}{2y\sqrt{x^2{P_c}^2 + 2y^2 P_c + y^2}\,(2P_c + 1)^2} \tag{B.17}$$

Also notice that according to Assumption 1, $y > 0$. Furthermore, $x = y + 2(1-\theta) > y > 0$. It is easy to see that the denominator of the right-hand site expression in Equation B.17 is greater than 0. Rewrite the numerator as $x(\sqrt{x^2 P_c^2} - \sqrt{x^2{P_c}^2 + 2y^2 P_c + y^2}) + y^2(-2P_c - 1)$, and we can see both terms in the numerator are negative. Thus, we have $\frac{\partial P_b^0}{\partial P_c} < 0$.

We next prove that $\frac{\partial P_b^0}{\partial \theta} > 0$. Rewrite the expression of P_b^0 while collecting the terms involving θ as

$$P_b^0 = \frac{\theta(\bar{b}P_c + \bar{b} + 3P_c + 1) - 3P_c - 1 + \sqrt{\Delta'}}{\theta(2\bar{b}P_c + \bar{b} + 2P_c + 1) - 2P_c - 1} \tag{B.18}$$

where $\Delta' = \theta^2(\bar{b}^2(P_c^2 + 2P_c + 1) + 2\bar{b}(-P_c^2 + 2P_c + 1) + (P_c^2 + 2P_c + 1)) + \theta(-2\bar{b}(-P_c^2 + 2P_c + 1) - 2(P_c^2 + 2P_c + 1)) + P_c^2 + 2P_c + 1$. Rewrite Δ' as

$$\Delta' = (\theta\bar{b}P_c - \theta P_c + P_c)^2 + (2P_c + 1)(\bar{b}\theta + \theta - 1)^2 \tag{B.19}$$

Note that for the second term $\frac{\partial((2P_c+1)(\bar{b}\theta+\theta-1)^2)}{\partial\theta} = 2(2P_c+1)(\bar{b}\theta+\theta-1)(\bar{b}+1) > 0$ according to Assumption 1. Combine the fact that the partial derivative of the denominator in Equation B.18 w.r.t. θ equals $2\bar{b}P_c + \bar{b} + 2P_c + 1$, which is greater than 0, we have

$$\begin{aligned}
\frac{\partial P_b^0}{\partial\theta} &= \frac{\partial \frac{\theta(\bar{b}P_c+\bar{b}+3P_c+1)-3P_c-1+\sqrt{\Delta'}}{\theta(2\bar{b}P_c+\bar{b}+2P_c+1)-2P_c-1}}{\partial\theta} \\
&> \frac{\partial \frac{\theta(\bar{b}P_c+\bar{b}+3P_c+1)-3P_c-1+\sqrt{(\theta\bar{b}P_c-\theta P_c+P_c)^2}}{\theta(2\bar{b}P_c+\bar{b}+2P_c+1)-2P_c-1}}{\partial\theta} \\
&= \frac{\partial \frac{\theta(\bar{b}P_c+\bar{b}+3P_c+1)-3P_c-1+(\theta\bar{b}P_c-\theta P_c+P_c)}{\theta(2\bar{b}P_c+\bar{b}+2P_c+1)-2P_c-1}}{\partial\theta} \\
&= \frac{\partial \frac{\theta(2\bar{b}P_c+\bar{b}+2P_c+1)-2P_c-1}{\theta(2\bar{b}P_c+\bar{b}+2P_c+1)-2P_c-1}}{\partial\theta} \\
&= 0
\end{aligned}$$

Thus $\frac{\partial P_b^0}{\partial\theta} > 0$.

B.7 Proof of Proposition 3

In this proof, we first compare total surplus in each of the sub-games resulting from the two lenders' algorithm revealing decisions, and then subtract the lenders' equilibrium payoff from the total surplus and compare the borrower (consumer) surplus. We calculate the total surplus by adding together the social surplus generated by borrowers from different segments. When a H type borrower gets the loan, a ($TS_h = M_h - m$) amount of social welfare is generated. When a L type borrower gets the loan, a ($TS_l = M_l - m - 1$) amount of social welfare is generated (i.e., the lender loses 1, the loan amount, while the borrower loses m, the negative impact on the credit score). When a borrower is rejected, a $-m$ amount of social surplus is generated. To calculate the social surplus generated in each sub-game, we need to find out the number of H type borrowers who are approved ($N_{H,a}$), the number of L type borrowers who are approved ($N_{L,a}$), and the number of borrowers who are rejected (N_r). We do not care about the borrowers who do not apply since they will generate 0 surplus to the social welfare. For the R-R case, there are 4 segments of borrowers: HH HL LH and LL. Borrowers in the HH HL and LH segments will apply and will be approved. Thus we have $N_{H,a}^{rr} = N_{hh}^{rr}V_{hh}^{rr} + 2N_{hl}^{rr}V_{hl}^{rr}$, $N_{L,a}^{rr} = N_{hh}^{rr}(1-V_{hh}^{rr}) + 2N_{hl}^{rr}(1-V_{hl}^{rr})$, and $N_r^{rr} = 0$. (We use the superscript rr, nn, and nr to denote the R-R, N-N, and N-R case respectively.) The total surplus generated in the R-R case is

$$TS^{rr} = N_{H,a}^{rr}TS_h + N_{L,a}^{rr}TS_l \tag{B.20}$$

Similarly, in the N-N case, we have

$$TS^{nn} = N^{nn}_{H,a} TS_h + N^{nn}_{L,a} TS_l + N^{nn}_r (-m) \tag{B.21}$$

where

$$\begin{aligned}
N^{nn}_{H,a} &= P_b P_c \theta \left(-2P_b P_c + P_b + P_c + 1\right) \\
N^{nn}_{L,a} &= P_c \left(P_b - 1\right)\left(\theta - 1\right)\left(2P_b P_c - P_b \left(P_c - 1\right) - P_c \left(P_b - 1\right) + 2\left(P_b - 1\right)\left(P_c - 1\right)\right) \\
N^{rr}_r &= {P_b}^2(-3{P_c}^2 + 5P_c - 2) + 4P_b P_c \theta (P_c - 1) + P_b P_c (P_c - 3) + 2P_b - 2P_c \theta (P_c - 1)
\end{aligned}$$

In the N-R case, we have

$$TS^{nn} = N^{nr}_{H,a} TS_h + N^{nr}_{L,a} TS_l + N^{nr}_r (-m) \tag{B.22}$$

where

$$\begin{aligned}
N^{nr}_{H,a} &= P_b \theta \left(-P_b P_c + P_c + 1\right) \\
N^{nr}_{L,a} &= \left(P_b - 1\right)\left(\theta - 1\right)\left(2P_b P_c - P_b \left(P_c - 1\right) - P_c \left(P_b - 1\right) + \left(P_b - 1\right)\left(P_c - 1\right)\right) \\
N^{nr}_r &= \left(P_c - 1\right)\left({P_b}^2 \left(\theta - 1\right) - \theta (P_b - 1)^2\right)
\end{aligned}$$

Next, we show that $TS^{rr} > TS^{nr} > TS^{nn}$. The comparison is straightforward but cumbersome and thus the details are omitted here. The intuition is as follows. Note that $N^{nn}_r > N^{nr}_r > N^{rr}_r$ and $N^{nn}_{H,a} + N^{nn}_{L,a} < N^{nr}_{H,a} + N^{nr}_{L,a} < N^{rr}_{H,a} + N^{rr}_{L,a}$. In other words, in the N-N case, the number of borrowers who are approved is the lowest and the number of borrowers who are rejected is the highest. In the R-R case, the number of borrowers who are approved is the highest and the number of borrowers who are rejected is the lowest. The N-R case is between the two extremes. Since rejected borrowers will generate a negative social surplus, and accepted borrowers on average will generate a positive social surplus, we have $TS^{rr} > TS^{nr} > TS^{nn}$.

We next subtract the lenders' equilibrium profits from the social surplus to compute borrower surplus:

$$CS^{rr} = TS^{rr} - 2\Pi_{rr} \tag{B.23}$$

$$CS^{nn} = TS^{nn} - 2\Pi_{nn} \tag{B.24}$$

$$CS^{nr} = TS^{nr} - \Pi^1_{nr} - \Pi^1_{nr} \tag{B.25}$$

where TS^{rr}, TS^{nn}, and TS^{nr} are defined in Equations B.20, B.21, and B.22 respectively. Π_{rr}, Π_{nn}, and Π_{nr} are defined in Lemma 1, 2, and 3 respectively. We can plug the expression of total surplus and lenders' profit into these expressions and compare them, and we will get $CS^{rr} > CS^{nr} > CS^{nn}$. Again, the comparison is straightforward but cumbersome, so we omitted the details here. The intuition of $CS^{rr} > CS^{nr}$ is simple: (1) The number of borrowers who are approved in the R-R case is larger than in the N-R case. (2) $F^{rr}(b) > F_1(b)$ and $F^{rr}(b) > F_2(b)$, which means on average borrowers face lower interest rates in the R-R case than in the N-R case. Points (1) and (2) above ensure that borrowers' welfare in the R-R case is higher than in the N-R case. We next discuss why $CS^{nr} > CS^{nn}$. Comparing the difference in total surplus and the difference in lenders' profit in the N-N and N-R cases, we have $TS^{nr} - TS^{nn} > \Pi^1_{nr} + \Pi^2_{nr} - 2\Pi_{nn} = \Pi^1_{nr} - \Pi_{nn}$ (Note that $\Pi^2_{nr} = \Pi_{nn}$), which means the lenders only take a portion of the increased surplus, and the remaining is left to the borrowers.

B.8 Proof of Proposition 4

We consider sub-games resulting from the lenders' decisions on algorithm upgrade (*NU-NU*,*NU-U*,*U-NU*,*U-U*), where *NU* stands for "not upgrade" and the algorithm's accuracy stays at P_b, and *U* stands for "upgrade" and the algorithm's accuracy rises to P_b^*. The equilibrium payoff in the two symmetric cases, *NU-NU* and *U-U*, have already been calculated and shown in Lemma 1. Since the two lenders' decisions on revealing is symmetric under mandatory revealing, the lenders' equilibrium payoff in *NU-U* case mirrors the payoff in the *U-NU* case. Compared with the *NU-NU* case, in the *NU-U* case, the only things that will change are the number of borrowers in each segment. Similar to the analysis in Appendix B.2, we can find the lenders' equilibrium payoff when their algorithms' accuracy are P_b and P_b^* respectively. The payoff matrix of the "algorithm upgrading" game is shown in the following table:

		Lender 2: Not Upgrade	Lender 2: Upgrade
Lender 1	Not Upgrade	$(\Pi^{rr}_{NU\text{-}NU}, \Pi^{rr}_{NU\text{-}NU})$	$(\Pi^{rr}_{NU\text{-}U}, \Pi^{rr}_{U\text{-}NU})$
	Upgrade	$(\Pi^{rr}_{U\text{-}NU}, \Pi^{rr}_{NU\text{-}U})$	$(\Pi^{rr}_{U\text{-}U}, \Pi^{rr}_{U\text{-}U})$

Where

$$\Pi^{rr}_{NU\text{-}NU} = P_b\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right)$$
$$\Pi^{rr}_{U\text{-}U} = P_b^*\left(-\bar{b}P_b^*\theta + \bar{b}\theta - P_b^*\theta + P_b^* + \theta - 1\right)$$
$$\Pi^{rr}_{NU\text{-}U} = P_b\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right)$$
$$\Pi^{rr}_{U\text{-}NU} = P_b^*\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right)$$

We focus on the symmetric mixed strategy equilibrium and assume both lenders choose to upgrade with probability P^M. The following equation must hold to ensure both lenders are indifferent between upgrading or not:

$$(1-P^M)\Pi_{NU\text{-}NU} + P^M\Pi^1_{NU\text{-}U} = (1-P^M)\Pi^2_{NU\text{-}U} + P^M\Pi_{U\text{-}U} \tag{B.26}$$

Solve Equation B.26 and we get $P^M = \frac{1-P_b}{P_b^*}$.

B.9 Proof of Proposition 5

Assume in this case Lender 1 reveals its algorithm but Lender 2 does not. We solve for lenders equilibrium payoff in sub-games followed by every combination of the two lenders' upgrading decisions, the payoff matrix of the game is shown in the following table:

		Lender 2: Not Upgrade	Lender 2: Upgrade
Lender 1	Not Upgrade	$(\Pi^{nr1}_{NU\text{-}NU}, \Pi^{nr2}_{NU-NU})$	$(\Pi^{nr1}_{NU\text{-}U}, \Pi^{nr2}_{U-NU})$
	Upgrade	$(\Pi^{nr1}_{U\text{-}NU}, \Pi^{nr2}_{NU-U})$	$(\Pi^{nr1}_{U\text{-}U}, \Pi^{nr2}_{U-U})$

where

$$\begin{cases}
\Pi^{nr1}_{NU\text{-}NU} = & -\bar{b}P_b\theta\left(P_b\left(P_c-1\right)+P_c\left(P_b-1\right)\right)-\left(P_b+P_c-1\right)\left(\theta-1\right)\left(P_b-1\right) \\
\Pi^{nr2}_{NU\text{-}NU} = & P_c\left(P_b{}^2\left(-2\bar{b}P_c\theta+\bar{b}\theta-2P_c\theta+2P_c+\theta-1\right)+P_b\left(\bar{b}P_c\theta+(\theta-1)(3P_c-2)\right)\right. \\
 & -P_c\theta+P_c+\theta-1 \\
\Pi^{nr1}_{U\text{-}NU} = & -\bar{b}P_b^*\theta\left(P_b\left(P_c-1\right)+P_c\left(P_b-1\right)\right)-P_bP_c\left(P_b^*-1\right)\left(\theta-1\right) \\
 & -\left(P_b-1\right)\left(P_b^*-1\right)\left(P_c-1\right)\left(\theta-1\right) \\
\Pi^{nr2}_{NU\text{-}U} = & P_c(P_b{}^2\left(P_b^*\left(-2\bar{b}P_c\theta+\bar{b}\theta-2P_c\theta+2P_c+\theta-1\right)+2P_c\theta-2P_c-\theta+1\right) \\
 & +P_b\left(P_b^*\left(\bar{b}P_c\theta+P_c\theta-P_c-\theta+1\right)-P_c\theta+P_c\right)+P_b^*\left(\theta-1\right))/P_b^* \\
\Pi^{nr1}_{NU\text{-}U} = & -\bar{b}P_b\theta\left(P_b^*\left(P_c-1\right)+P_c\left(P_b^*-1\right)\right)-P_b^*P_c\left(P_b-1\right)\left(\theta-1\right) \\
 & -\left(P_b-1\right)\left(P_b^*-1\right)\left(P_c-1\right)\left(\theta-1\right) \\
\Pi^{nr2}_{U\text{-}NU} = & P_c(P_b\left(P_b^{*2}\left(-2\bar{b}P_c\theta+\bar{b}\theta+(\theta-1)(1-2P_c)\right)+P_b^*\left(\bar{b}P_c\theta+(\theta-1)(P_c-1)\right)\right) \\
 & +\theta-1+P_b^{*2}\left(2P_c\theta-2P_c-\theta+1\right)+P_b^*\left(-P_c\theta+P_c\right))/P_b \\
\Pi^{nr1}_{U\text{-}U} = & -\bar{b}P_b^*\theta\left(P_b^*\left(P_c-1\right)+P_c\left(P_b^*-1\right)\right)-P_b^*P_c\left(P_b^*-1\right)\left(\theta-1\right) \\
 & -\left(P_b^*-1\right)^2\left(P_c-1\right)\left(\theta-1\right) \\
\Pi^{nr2}_{U\text{-}U} = & P_c\left(-2\bar{b}P_b^{*2}P_c\theta+\bar{b}P_b^{*2}\theta+\bar{b}P_b^*P_c\theta-2P_b^{*2}P_c\theta+2P_b^{*2}P_c+P_b^{*2}\theta-P_b^{*2}\right. \\
 & +3P_b^*P_c\theta-3P_b^*P_c-2P_b^*\theta+2P_b^*-P_c\theta+P_c+\theta-1
\end{cases}$$

We first prove that Lender 1 has a dominate strategy "U" by showing $\Pi^{nr1}_{U\text{-}NU} > \Pi^{nr1}_{NU\text{-}NU}$ and

$\Pi^{nr1}_{U\text{-}U} - \Pi^{nr1}_{NU\text{-}U}$.

$$\Pi^{nr1}_{U\text{-}NU} - \Pi^{nr1}_{NU\text{-}NU} = (P_b - P_b^*)\left(\bar{b}\theta\left(P_b\left(P_c - 1\right) + P_c\left(P_b - 1\right)\right) + (\theta - 1)\left(P_bP_c + \left(P_b - 1\right)\left(P_c - 1\right)\right)\right)$$

Since $P_b < P_b^*$, $P_b < 1$, $P_c < 1$ and $\theta < 1$, $\Pi^{nr1}_{U\text{-}NU} - \Pi^{nr1}_{NU\text{-}NU} > 0$. Further,

$$\Pi^{nr1}_{U\text{-}U} - \Pi^{nr1}_{NU\text{-}U} = (P_b - P_b^*)\left(\bar{b}\theta\left(P_b^*\left(P_c - 1\right) + P_c\left(P_b^* - 1\right)\right) + (\theta - 1)\left(P_b^*P_c + \left(P_b^* - 1\right)\left(P_c - 1\right)\right)\right)$$

Again, since $P_b < P_b^*$, $P_b^* < 1$, $P_c < 1$ and $\theta < 1$, $\Pi^{nr1}_{U\text{-}U} - \Pi^{nr1}_{NU\text{-}U} > 0$.

As Lender 1 always chooses to upgrade the algorithm, Lender 2 makes the upgrading decision by comparing $\Pi^{nr2}_{NU\text{-}U}$ with $\Pi^{nr2}_{U\text{-}U}$. Let $\delta = \Pi^{nr2}_{U\text{-}U} - \Pi^{nr2}_{NU\text{-}U}$. It is easy to check that δ is linearly decreasing in θ. Since $\delta|_{\theta=\theta^0}$, $\delta > 0$ when $\theta < \theta^0$. Notice that θ^0 is decreasing in P_c, and $\theta^0|_{P_c = P_c^0} = 1$, which means $\theta^0 > 1$ when $P_c < P_c^0$. Thus $\theta < \theta^0$ will always hold.

B.10 Proof of Proposition 6

In the mandatory revealing scenario, both lenders choose to upgrade the algorithm with probability P_M, that is, with probability P_M^2 both lenders' algorithms' accuracy will be P_b^*, with probability $(1 - P_M)^2$ both lenders' algorithms' accuracy will be P_b, and with probability $2P_M(1 - P_M)$ one lender's accuracy will be P_b and the other's will be P_b^*. Borrower surplus is thus the expected total surplus minus the lenders' expected profit, as shown in Equation B.27.

$$\begin{aligned} CS^{rr*} =& (1 - P_M)^2\left(TS^{rr}_{P_b,P_b} - 2\Pi^{rr}_{P_b,P_b}\right) + {P_M}^2\left(TS^{rr}_{P_b^*,P_b^*} - 2\Pi^{rr}_{P_b^*,P_b^*}\right) \\ &+ 2P_M\left(1 - P_M\right)\left(TS^{rr}_{P_b,P_b^*} - \Pi^{rr}_{P_b,P_b} - \Pi^{rr}_{P_b,P_b^*}\right) \end{aligned} \tag{B.27}$$

where

$$\begin{cases} TS^{rr}_{P_b,P_b} = & P_b\left(2 - P_b\right)\left(-\theta\left(m - M_h\right) + (\theta - 1)\left(m - M_l + 1\right)\right) \\ TS^{rr}_{P_b,P_b^*} = & \left(\theta\left(m - M_h\right) - (\theta - 1)\left(m - M_l + 1\right)\right)\left(-P_bP_b^* + P_b\left(P_b^* - 1\right) + P_b^*\left(P_b - 1\right)\right) \\ TS^{rr}_{P_b^*,P_b^*} = & P_b^*\left(2 - P_b^*\right)\left(-\theta\left(m - M_h\right) + (\theta - 1)\left(m - M_l + 1\right)\right) \\ \Pi^{rr}_{P_b,P_b} = & P_b\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right) \\ \Pi^{rr}_{P_b,P_b^*} = & P_b^*\left(-\bar{b}P_b\theta + \bar{b}\theta - P_b\theta + P_b + \theta - 1\right) \\ \Pi^{rr}_{P_b^*,P_b^*} = & P_b^*\left(-\bar{b}P_b^*\theta + \bar{b}\theta - P_b^*\theta + P_b^* + \theta - 1\right) \end{cases}$$

In the voluntary scenario, when $\theta < \theta^0$, both lenders upgrade their algorithms to an accuracy of P_b^*. Thus borrower surplus can be calculated as:

$$CS^{nr*} = TS^{nr}_{P_b^*,P_b^*} - \Pi^{nr1}_{P_b^*,P_b^*} - \Pi^{nr2}_{P_b^*,P_b^*} \tag{B.28}$$

where

$$\begin{cases} TS^{nr}_{P_b^*,P_b^*} = & -m\left(P_c-1\right)\left(P_b^{*2}\left(\theta-1\right)-\theta(P_b^*-1)^2\right) \\ & +P_b^*\theta\left(m-M_h\right)\left(-P_b^*P_c+P_b^*\left(P_c-1\right)+2P_c\left(P_b^*-1\right)\right) \\ & -\left(P_b^*-1\right)\left(\theta-1\right)\left(m-M_l+1\right)\left(2P_b^*P_c-P_c\left(P_b^*-1\right)+\left(P_b^*-1\right)\left(P_c-1\right)\right) \\ \Pi^{nr1}_{P_b^*,P_b^*} = & -\bar{b}P_b^*\theta\left(P_b^*\left(P_c-1\right)+P_c\left(P_b^*-1\right)\right)-P_b^*P_c\left(P_b^*-1\right)\left(\theta-1\right) \\ & -(P_b^*-1)^2\left(P_c-1\right)\left(\theta-1\right) \\ \Pi^{nr2}_{P_b^*,P_b^*} = & P_c\left(-2\bar{b}P_b^{*2}P_c\theta+\bar{b}P_b^{*2}\theta+\bar{b}P_b^*P_c\theta-2P_b^{*2}P_c\theta+2P_b^{*2}P_c+P_b^{*2}\theta-P_b^{*2}\right. \\ & +3P_b^*P_c\theta-3P_b^*P_c-2P_b^*\theta+2P_b^*-P_c\theta+P_c+\theta-1 \end{cases}$$

Define $\delta = CS^{rr*} - CS^{nr*}$. Combining and rearranging terms, we can see that δ is a quadratic function in P_c: $\delta = AP_c^2 + BP_c + C$, where A, B, and C are defined in Proposition 6. We can then check that $\delta|_{P_c=0} = C < 0$, and $\delta|_{P_c=1} = A + B + C > 0$ in the parameter ranges defined in Assumptions 1 and 2. Thus one of the roots of the quadratic function, $P_c^0 = \frac{-B+\sqrt{B^2-4AC}}{2A}$, must be in range $(0,1)$. Thus $\delta > 0$ if $P_c > P_c^0$.