The BigTech-Bank Lending Ecosystem*

Yiyao Wang ⓡ Ben Charoenwong†

Abstract

We develop a lending model featuring banks with low funding costs and bigtech lenders with superior data for screening borrowers and show that bigtech-bank collaboration at the data level dominates other collaboration models in the lending ecosystem. Under data-level collaborations, bigtech lenders serve as an on-ramp to the financial system by lending to risky borrowers and then selling borrowers' credit history data to banks who take over subsequent lending. Allowing bigtech-bank collaborations improves financial inclusion but reduces bigtech investment in screening technology, and its welfare effect depends crucially on lending market competition and banks' funding-cost advantage.

JEL Classifications: G21, G23, G33, O32, O34, L84
Keywords: data governance, asymmetric information, relationship lending, bigtech

*The authors thank Zhiguo He, Yunzhi Hu, Jing Huang (discussant), Kai Li, Jun Pan, Ajay Patel (discussant), Wenlan Qian, Anand Srinivasan, participants at the 2021 CICF, Risk Management Institute 2021 Annual Meeting, 2023 Asian Finance Association Meetings, 2023 Vietnam Symposium in Banking and Finance, 2023 Australasian Finance and Banking Conference, 2023 Global AI in Finance Conference, 2024 ANSCSE, 2024 GFFC at Cornell University, Southern University of Science and Technology seminar, and brownbag participants at National University of Singapore, Shanghai Advanced Institute of Finance. Yiyao thanks Xingjian Zheng and Yitian Zhang for excellent research assistance. Any remaining errors are our own. Yiyao Wang, Email: yiyaowang@saif.sjtu.edu.cn, Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University, 211 West Huaihai Road, Office 708, Shanghai, China 200120. Ben Charoenwong, Email: ben.charoenwong@insead.edu, INSEAD, INSEAD Asia Campus, 1 Ayer Rajah Ave #621, Singapore 138676.

†Certified randomized at the AEA link. The typographical error in the record is due to Ben misspelling his own last name.

1 Introduction

Bigtech lending – lending done or facilitated by large technology-based companies whose primary business is not financial services – stands at over US$500 billion worldwide, more than double the amount of fintech lending at US$220 billion as of 2019 (Cornelli et al., 2020).[1] Compared to traditional banks, one advantage of bigtech lenders is their access to proprietary data of their users that they can use to screen borrowers. For example, Ant Financial's loans to small consumer and businesses in China were over US$238 billion, made using their proprietary credit scoring system Sesame Credit, which relies on data from its commonly-owned online retail platform Taobao (Hau et al., 2018). However, how the development of these new kinds of lenders affects the traditional banking system is not clear (Vives, 2017). While they may compete in some dimensions for some segments of borrowers, they can also gain from cooperating, especially when bigtech lenders have advantages in screening borrowers with alternative data and banks enjoy lower funding costs due to the regulated ability to take insured deposits and access central-bank liquidity.[2] For example, Amazon in the United States has lent over US$1 billion to small merchants both directly and with third-party financial institutions such as Goldman Sachs.[3] Liu et al. (2022) document that around 60% of global bigtech lending is done in collaboration with banks. But exactly how traditional banks and bigtech lenders interact can affect financial inclusion, borrowing costs, bigtech investments in data processing technology, and social welfare. As BigTech lending grows rapidly, it's crucial to understand how these new lenders interact with traditional banks. This interaction could significantly impact financial inclusion, systemic risk, and the overall structure of credit markets. Unlike previous studies focusing on competition between BigTech and banks, we study different

[1] In contrast to bigtech, fintech lenders may have an advantage vis-à-vis traditional banks in terms of information or monitoring (e.g., Berg et al. 2020; Council 2021; Huang 2021), enforcement (e.g., Huang 2021), hardening soft information (He et al. 2024), alternative collateral (e.g., Dai et al. 2020), or convenience (e.g., Buchak et al. 2018b; Berg et al. 2022). We focus on bigtech lenders and an information advantage, given the growth of bigtech lending done by companies whose primary business lines were not in lending but in other things like search, advertising, and social media, which generates credit-risk-relevant alternative data (Charoenwong and Kwan 2021).

[2] Traditional banks may be unable to emulate the financial analytics capability due to more stringent regulation, legacy information technology, organization frictions, and the lack of alternative data access (Stulz 2019).

[3] See https://www.businessinsider.com/amazon-seller-lending-program-growing-recession-fears-2023-1?op=1 and https://www.nerdwallet.com/article/small-business/amazon-lending-small-business-loans.

forms of collaboration between bigtech lenders and banks in the lending market.

To answer these questions, we develop a model of a lending market featuring banks with low funding costs and bigtech lenders with superior borrower screening data and technology.[4] Our model rests on three key assumptions: BigTech lenders have superior data for screening but higher funding costs than banks, BigTech lenders are protected by limited liability, and the screening intensity is non-contractible. Bigtech lenders can increase financial inclusion by lending to risky borrowers otherwise under-served by traditional banks. In doing so, bigtech lenders create and accumulate borrowers' credit history data, which is valuable for banks' lending decisions and thus gives rise to potential bigtech-bank collaborations. We define the "BigTech-bank lending ecosystem" as the set of competitive and collaborative interactions between BigTech firms engaged in lending activities and traditional banks. This ecosystem encompasses various modes of competition and collaboration. We show that the privately optimal collaboration is at the data level; that is, the bigtech lender sells the borrowers' credit history data to the bank, which can be implemented in practice through an alternative credit scoring service (Dash et al., 2021; Bradford, 2023). With this type of collaboration, bigtech lenders effectively serve as an on-ramp to the traditional financial system by selling borrowers' credit history data to banks who take over subsequent lending. The bigtech-bank collaboration through data trading lowers bigtech lenders' incentives to screen, which has an ambiguous effect on welfare. We find that social welfare is more likely to fall if banks have a significant funding-cost disadvantage over bigtech lenders. This justifies the need for regulatory restrictions on bigtech-bank collaborations and data transactions.

In the model, there is a continuum of borrowers who need to finance their projects. Specifically, each borrower has an investment project that can be either good or bad. A good-type project yields a positive net present value (NPV), while a bad-type project has negative NPV because it may fail with some probability. Borrowers can apply for financing from two types of lenders: banks and bigtech lenders. We assume that banks have lower funding costs while bigtech lenders possess superior screening technologies that can help detect bad-type projects. We show that under

[4]BigTech firms often have access to alternative data sources, such as social media activity or e-commerce transactions, which can improve credit scoring (Frost et al. 2019; Berg et al. 2020).

certain conditions, bigtech lenders can improve financial inclusion by lending to risky borrowers underserved by banks.

By lending to a risky borrower, a bigtech lender privately owns the credit history data of the borrower. In the model, conditional on no defaults, the borrower's perceived credit quality improves over time. Therefore, a borrower's credit history data becomes more valuable over time as the lender can capture the information-monopoly rent in servicing the borrower's future financing needs. Consequently, banks—who have lower funding costs than bigtech lenders but find it too risky to initiate lending to high-risk borrowers —may find it profitable to collaborate with bigtech lenders so that they can utilize their funding-cost advantages in lending to these borrowers when they have good credit histories.

To our knowledge, ours is the first to analyze different modes of collaboration between BigTech lenders and banks, showing that data-level collaboration dominates other forms of partnership. We explore three different types of collaboration models between bigtech lenders and banks: (1) funding-level collaboration, in which the bank provides partial funding for the bigtech lending, (2) data-level collaboration, in which the bigtech lender sells the borrower's credit history data to the bank, and (3) asset-level collaboration, in which the bigtech lender sells the borrower loan to the bank. We show that the privately optimal bigtech-bank collaboration is in the data level. The data-level collaboration contract dominates because it best incentivizes bigtech to invest in screening technology by having the bigtech lender bear the entire loan risk for the initial financing period when the borrower's credit quality is relatively low.

Under data-level collaborations, bigtech lenders provide loans to borrowers underserved by banks and later serve as an on-ramp to the financial system by selling borrowers' credit history data to banks who take over subsequent lending. In practice, this collaboration can be implemented as bigtech lenders providing alternative credit scoring services to banks who screen their new loan applications. This form of data contracting allows the bigtech lender to monetize their access to alternative data and screening technology by converting them into more standardized data, such as alternative credit scoring and credit reports, that other financial institutions can use.

Bigtech-bank collaborations tend to reduce bigtech screening investment, yet whether such a decrease in screening investment is welfare increasing or decreasing is ambiguous. We show that the decrease in screening investment brought by bigtech-bank collaboration can improve welfare when the bigtech lending market is competitive and when the funding-cost difference between banks and bigtech lenders is small. In the model, because of asymmetric information in the lending market, bigtech lenders tend to over-invest in screening technologies compared to the first-best level when facing fierce competition. In this case, allowing bigtech-bank collaboration via data trading reduces over-investment in bigtech screening, which can improve welfare when the funding-cost difference between the bank and the bigtech lender is relatively small. However, when the bigtech lender enjoys monopoly power in the lending market or when the bank enjoys a significant funding-cost advantage, allowing bigtech-bank collaborations via data trading leads to underinvestment in screening technologies. In those cases, a bigtech lender tends to sell the borrower's credit history data too early to enjoy the funding-cost savings, rendering their skin in the game too low in screening.

Empirically, our model predicts that markets with more prevalent BigTech-bank collaborations should experience increased lending to previously underserved borrowers. Additionally, the model predicts that the nature of these collaborations (e.g., data-level vs. funding-level) should impact both the extent of financial inclusion and the investment in screening technologies. These predictions could be tested by analyzing the lending practices and outcomes in markets with different types of BigTech-bank partnerships.

Our results have implications for policymakers and regulators. Encouraging data-sharing arrangements between BigTech lenders and banks could enhance financial inclusion and market efficiency. However, our results also show that regulators should also be mindful of potential risks to data privacy that such collaborations might pose. To this end, our paper also highlights the role of financial data governance regulations (e.g., McNulty et al. 2023; Arner et al. 2023), particularly pertaining to the rules on credit information sharing. Importantly, our results show financial data governance is related to but distinct from other privacy-related considerations (Ramadorai

et al. 2020). Policies restricting the use of certain information in credit score evaluation affect the comparative advantage of the bigtech screening but need not necessarily hamper the sale or sharing of credit history data. Bigtech lending is a mechanism that allows bigtech to convert their proprietary data and technology into a standardized form that can preserve their proprietary edge, avoid potential database reconstruction (2014; 2019), and resolve some privacy concerns. In addition, because the models' predictions depend on the existing industrial organization of the banking sector, implications may vary across different regulatory environments in practice.

After discussing the related literature, the remainder of the paper is as follows: Section 2 develops a model of bigtech lending, Section 3 analyzes different forms of bigtech-bank collaborations, Section 4 extends the model to allow for competition among bigtechs as well as competition between bigtechs and banks and discusses potential implementations concerns like data exclusivity and reverse-engineering concerns as a result of data selling, and Section 5 concludes.

Related Literature

Our research contributes to the nascent literature on financial technology and banking. A close paper to ours is Liu et al. (2022), who document widespread collaboration of bigtech lenders with bank. While they focus on potential reasons for why specialization of bigtech appear to specialize in short-term liquidity financing, we extend the analysis to consider various forms of data and loan sharing arrangements. In this sense, our paper is closer to Parlour et al. (2022), who study the impact of bigtech competition in payment services when banks and fintech companies can use payment data to learn about consumers' credit quality. While they focus on both payment processing costs and the lending market, we focus exclusively on the lending market but instead adopt an optimal contracting framework to study collaborative models. Consequently, our findings are more relevant to lending market regulations, such as types of data lenders may use, data sharing policies, and the governance of bigtech-bank partnerships. Schweitzer and Barkley (2017) compares traditional bank borrowers and online platform borrowers and show that the latter resemble applicants who were denied credit by traditional banks (Cole and Sokolyk, 2016). Jagtiani and Lemieux (2018) and Chioda et al. (2024) show that fintech lending penetrates areas underserved

by traditional banks.

Our study of information sharing builds on existing works like Pagano and Jappeli (1993) who study the determinants of and implications when banks choose to share information of borrowers with each other through the establishment of a credit bureau. They find that the incentives to share data among banks increase with more technological availability for screening or information, the size of the credit market, and decreases with regulatory safeguards for privacy. They find that increased competition among banks decreases the incentives to share the data, similar to the results in our model. Relatedly, Bennardo et al. (2015) shows information sharing between banks can mitigate overborrowing incentives and emphasize that private and social incentives for information sharing are not aligned. In contrast, we study data agreements in an ecosystem with different types of lenders and consider the unique aspects of bigtech-bank collaborations and their impact on financial inclusion and screening efficiency.

On the value of data, Farboodi et al. (2019) and Jones and Tonetti (2020) show that the value of data has increasing returns in helping borrowers do price differentiation. The latter discusses situations in which data sharing would increase economic efficiency and concludes that giving data rights to consumers who trade off between privacy and economic gains from selling their data generates an allocation close to socially optimal. He et al. (2023) consider a setting with an open bank whereby borrowers own their borrowing data and find that open banking can leave borrowers worse off even if they could share their data. Our model consider data-level collaboration between different lenders. In our setting, giving the borrowers their credit history data is not credible because it would reduce bigtech lenders' incentives to create screening and credit history data in the first place, causing increased borrowing costs and credit rationing. Our consideration of the market value of credit history also relates to Chatterjee et al. (2020) who develop a structural model to estimate the value of credit scores, although the latter does not consider interactions across financial institutions.

In addition, our research relates to the industrial organization of the financial system, including the consideration of banking competition and fintech on incumbents. For example, Mitchell and

Pearce (2011) finds that more competitive lending markets provide minority small business owners with more access, while Berger et al. (2015) find that more small bank presence yields more lending with slightly lower failure rates of small borrowers during normal times, but the differences in failure rates disappear during the financial crisis. In terms of the interaction with technology-based lenders, Fuster et al. (2019) studies how technology affects mortgage origination in terms of processing time and find that fintech lenders process applications faster and adjust supply more elastically than other lenders in response to exogenous mortgage demand. Our research is also relevant for the more recent European regulations aimed at providing a framework for fintech-bank data sharing (Borgogno and Colangelo, 2020).

2 A Model of BigTech Lending

2.1 Setup

Time is continuous and runs to infinity. The economy is populated with three types of agents, borrowers, bigtech lenders, and banks. All agents are risk-neutral. Bigtech lenders and banks discount cashflows at the rate of ρ, and borrowers discount cashflows at the rate of $\rho_b > \rho$.

Borrowers

There is a continuum of borrowers[5], which we normalize to be of measure one. At date 0, each borrower is endowed with an investment project. The project requires 1 dollar of initial investment and generates ydt cashflows per dt unit of time. Whether invested or not, the project matures with poisson rate λ, and if invested, the project generates a terminal cashflow of $X = 1$ at maturity. After the project matures, the borrower exits the economy. The project can be either a good or a bad type. A good-type project never fails, whereas a bad-type project fails with poisson rate γ. For simplicity, we assume that the residual value of a failed project is 0. We assume the type of a project is its borrower's private information. Among all borrowers, π of them have good projects, while

[5]We use the term "borrower" rather than "entrepreneur" or "firm", because the model also applies to both firms as well as individual borrowers who may be seeking consumer loans or trying to reconcile existing debt, to the extent that some borrowers have positive net present value "projects" and others do not.

the remaining $1-\pi$ have bad projects. π, which is assumed to be public information, measures the observable quality of borrowers. We assume $y \in (\rho, \rho+\eta)$ so that the net present value (NPV) of a good project is positive and that a bad project is negative.

Lenders

The borrower has no net worth at date 0, so it has to get external funding to finance its investment project. Specifically, the borrower can get external financing from either a traditional bank or a bigtech lender.[6] These two types of financiers are long-lived and deep-pocketed. To keep the problem tractable, we restrict our attention to debt contracts that offer constant flows of interest payment and can be fully repaid at the borrower's discretion. In this regard, the debt contract in our model can be interpreted as a line of credit to the borrower.

There are two main differences between traditional banks and bigtech lenders in the model. First, we assume traditional banks have lower funding costs due to their deposit-taking ability compared to bigtech lenders whose financing typically comes from private markets. Specifically, we assume a bank enjoys bdt units of subsidy per unit of lending during a dt period. The funding-cost difference captures banks' advantage in their access to insured deposits and central bank liquidity. Second, we assume bigtech lenders have superior screening technology compared to banks. A bigtech lender has a screening technology that can generate informative signals about the borrower's creditworthiness (e.g., Council 2021) not available to banks. Specifically, the bigtech lender can pay to receive and receive a binary signal $s \in \{g, b\}$ (for good and bad, respectively) about project quality before making loans. The signal is always good ($s = g$) when the borrower is a good type while the signal is bad ($s = b$) with probability $\theta \in [0,1]$ when the borrower is a bad type. In other words, if the bigtech lender defines a borrower being good as the null hypothesis in the screening process, the signal makes a Type II error (false negative) with probability $1-\theta$ and makes no Type I errors (false positive).

In the baseline model, we assume that only one bigtech lender can access the screening tech-

[6] We can assume the measure of bigtech lenders is larger than 2, so the number of lenders is larger than the largest possible number of investment projects that are waiting to be financed at every date.

nology. This assumption captures a situation, for example, where the bigtech lender's screening advantage comes from its unique possession of borrowers' payment data (Parlour et al. 2022) or other types of alternative data (Charoenwong and Kwan 2021). We extend the model to allow for bigtech competition and discuss its welfare implication in Section 4. In screening, the lender chooses $\theta \in [0,1]$ at a convex cost $c(\theta)$ where $c'(\theta) > 0$, $c'(0) \to \infty$, and $c''(\theta) > 0$. Therefore, θ can be interpreted as the bigtech lender's investment in screening technology, or equivalently, its screening intensity. An alternative way to model the screening cost is to assume that the borrower generates a disutility that is increasing and convex in the bigtech lender's screening intensity. This disutility term captures the borrower's concern about data privacy. The main results of the paper are robust when we introduce such concern about data privacy. Therefore, we interpret the screening cost in the model more broadly as the combination of the bigtech lender's investment in screening technology as well as the borrower's disutility over data privacy.

In contrast, we assume banks have no access to such a screening technology, so they can only make loan decisions based on π, the publicly observable quality of the borrower. Therefore, the screening technology in the model can be interpreted as the bigtech lender's ability to develop analytics and alternative data pipelines in evaluating borrower quality.[7] For instance, the Ant Financial Group, a bigtech lender in China, can utilize consumers' online shopping records when screening their consumer loan applications, and can also extract information from online retailers' sales history when screening their business loan applications (Netzer et al. 2019). In this example, commercial banks and other bigtech lenders are unable to adopt the advanced screening technology, simply because they have no access to those alternative data (Djeundje et al., 2021). In addition, banks are lagging in screening with alternative data because of their slow adoption of financial technology and their conservativeness regarding data privacy violations (Stulz 2019).

To study the value of bigtech lenders' credit history data and the effect on potential bigtech-bank collaborations, we focuses on the group of borrowers that have no access to financing from the traditional banking sector. Specifically, banks find it too risky to lend to a borrower if and only

[7]For example, Duarte et al. (2012) shows using psychometrics that people who appear more trustworthy have better credit scores and Berg et al. (2020) shows a credit application's digital footprints are informative of credit quality.

if the borrower's observable credit quality π is below a threshold value π_b, where

$$\pi_b \frac{y+\lambda+b}{\rho+\lambda} + (1-\pi_b)\frac{y+\lambda+b}{\rho+\lambda+\eta} = 1. \tag{1}$$

In practice, there are a few reasons why some borrowers are underserved by traditional banks. First, small and medium enterprises may lack sufficient collateral assets, which are sometimes required for business loans (Cole and Sokolyk, 2016). Second, borrowers with low credit scores and limited credit history may be denied credit by banks due to a lack of data to evaluate their creditworthiness (Stiglitz and Weiss, 1981). For these borrowers, bigtech lenders can reduce lending risks for these high-risk borrowers by using alternative data for screening, improving their access to financing.[8] Specifically, a borrower with $\pi < \pi_b$ cannot obtain financing from banks and turns to a matched bigtech lender who has the borrower's alternative data for credit screening.

The bigtech lender chooses to screen the borrower with intensity level θ. If receiving a bad signal, the bigtech lender learns that the borrower has a bad project and will reject the loan application. If receiving a good signal, the lender's posterior belief about the borrower's credit quality improves, and the lender lends to the borrower at the quoted interest rate. Denote the bigtech lender's lending rate by R. In the baseline model, the borrower cannot get a loan from a bank and the bigtech lender is the only one who has the screening technology, so the bigtech lender faces no competition when servicing this low-quality borrower. Consequently, the lender can charge the highest possible interest rate $R = y$ to extract all the cashflows from the borrower's project. We assume that the borrower is willing to take such a loan because it generates some non-pledgeable cashflows or control rents for the borrower. After investing, the borrower pays the interest rate Rdt to the bigtech lender per dt period, until the project matures or fails.

[8]For example, Square Capital - the lending arm of the Square payment processor - has a page titled "Expanding Access" with statistics on how it has made loans to small businesses typically underserved by banks. See https://squareup.com/us/en/capital/access, accessed November 2023.

2.2 Bigtech Lending

We first analyze conditions under which bigtech lenders can improve financial inclusion, that is, financing borrowers that are underserved by banks. The lender's expected payoff from screening a borrower with $\pi < \pi_b$ can be written as

$$V_0(\pi) = \max_{\theta \in [0,1]} -c(\theta) + \pi[\frac{y+\lambda}{\rho+\lambda} - 1] + (1-\pi)(1-\theta)[\frac{y+\lambda}{\rho+\lambda+\eta} - 1]$$

The bigtech lender screens the borrower at the cost $c(\theta)$. With probability $(1-\pi)\theta$, the screening signal is bad, in which case the bigtech lender rejects the loan application. With probability $\pi + (1-\pi)(1-\theta)$, the bigtech lender receives a good signal and updates its posterior belief about the borrower being of good type from π to $\frac{\pi}{\pi+(1-\pi)(1-\theta)}$. In this case, the bigtech lender lends to the borrower at the interest rate $R = y$ and her expected payoff is the sum of the expected profit from lending to a good-type borrower ($\pi[\frac{y+\lambda}{\rho+\lambda} - 1] > 0$) and that from lending to a bad-type borrower ($(1-\pi)(1-\theta)[\frac{y+\lambda}{\rho+\lambda+\eta} - 1] < 0$). The screening technology allows the lender to reduce the expected loss from lending to a potential bad-type borrower.

The bigtech lender's optimal screening intensity balances the marginal screening cost and the marginal screening benefit of reducing expected loan losses of lending to bad-type borrowers. Denote the optimal screening intensity by θ_0, and the bigtech lender's expected payoff from lending by V_0. The optimal screening $\theta = \theta_0^*$ solves

$$c'(\theta) = (1-\pi)[1 - \frac{y+\lambda}{\rho+\lambda+\eta}]. \tag{2}$$

The bigtech lender finds it profitable to lend to the borrower if and only if $V_0(\pi) < 0$. Proposition 1 formally states the conditions under which bigtech lenders improve financial inclusion.

Proposition 1. *Assume that $\pi < \pi_b$. There exists a threshold value π_f^0 such that the bigtech lender screens the borrower with $\pi \in [\pi_f^0, \pi_b)$ and lends after receiving a good signal. $[\pi_f^0, \pi_b)$ is nonempty if and only if $b < \bar{b}$ and $y \in (y_1, y_2)$.*

The proof and the functional form of π_f^0, $\bar{b}$, and (y_1, y_2) are provided in the Appendix. Proposition 1 states that the bigtech lender can improve financial inclusion under two parameter conditions. Firstly, the funding-cost advantage of banks b must be relatively small. When b is too large, the bigtech lender's screening advantage is overwhelmingly dominated by its disadvantage in funding costs, making it impossible for the bigtech lender to improve financial inclusion.

Secondly, the cashflow from the project, which reflects the borrower's ability to service its debt, must stay in an intermediate range (y_1, y_2). When $y < y_1$, borrowers do not have many cashflows to service their debt, so only borrowers with high observable quality will be financed. For these borrowers, bigtech lenders' screening advantages over banks are relatively low, so they are unable to finance borrowers underserved by banks. When $y > y_2$, borrowers have plenty of cashflows to service their debt, so even borrowers with low observable quality will be financed by banks. In this case, the credit risk of a borrower that is underserved by banks is too high to justify the bigtech lender's initial screening cost. As a result, the bigtech lender is unable to improve financial inclusion neither.[9] However, when $y \in (y_1, y_2)$, banks find it too risky to lend to borrowers with intermediate levels of credit quality, whereas bigtech lenders may extend credits to these borrowers for whom the bigtech screening advantage is the largest. This result suggests that bigtech lenders improve financial inclusion for borrowers with intermediate levels of cashflows that can be used for debt payment.

2.3 Credit History Data

As the bigtech lender screens and lends to a borrower, she obtains information about the borrower, including the initial screening signal and information about whether the borrower has defaulted on the loan. These pieces of information will be referred to as the borrower's credit history data. A borrower's credit history data can be summarized by three variables $\Phi_t = (t, i_s, i_{d,t})$, where t denotes the time interval between the date when the borrower applied for a loan (date 0) and the current date, $i_s \in \{0, 1\}$ denotes whether the borrower has passed the initial bigtech screening, and

[9] While bigtech lenders do not improve financial inclusion in this case, they may compete with banks and take some borrowers away from banks, especially those borrowers with intermediate levels of credit quality. Such an analysis of lending market competition is beyond the scope of this paper, but we do discuss its implication in the Appendix.

$i_{d,t} \in \{0,1\}$ denotes whether the borrower has defaulted on the loan in the past. To simplify the analysis, we assume that whether the borrower's project is matured or not, that is, whether the borrower needs financing at the current moment, is public information. We call that a borrower has good credit history at date t if the borrower passes the initial bigtech screening ($i_s = 1$) and has not defaulted yet ($i_{d,t} = 0$).

The value of credit history data changes dynamically. Conditioning on the fact that the borrower has not defaulted on the loan, the bigtech lender's posterior belief about the borrower's credit quality improves over time. This result is driven by the assumption that a bad-type borrower defaults at a higher poisson rate than a good-type borrower does. We use π_τ to denote the lender's belief about the probability that the borrower is of good type conditioning on that the borrower has not defaulted at date τ. Then π_τ evolves as $d\pi_\tau = \eta\pi_\tau(1-\pi_\tau)d\tau$, with $\pi_0 = \frac{\pi}{\pi+(1-\pi)(1-\theta)}$. We have $\pi_\tau = \frac{\pi}{\pi+(1-\pi)(1-\theta)e^{-\eta\tau}}$, which is an increasing function on τ.

In Section 2.2, we assume that the borrower's credit history data is privately owned by the bigtech lender. If the borrower's credit history data is publicly available, and assuming the borrower does not default on the loan, its perceived creditworthiness improves over time. This gives the borrower an incentive to search for another lender in order to obtain a lower borrowing rate. If the cost of searching and switching lenders is relatively low, the initial lender cannot profit from the lending relationship. As a result, the initial lender may have little motivation to screen the borrower, leading to credit rationing.

Therefore, we primarily focus on the scenario where the bigtech lender owns the borrower's credit history data. Under this assumption, although the lender's belief about the borrower's credit quality improves over time, the borrower's credit history data remains confidential and not publicly known. In that case, the lender has no incentive to adjust the interest rate[10] and can capture information rents due to her private data ownership. However, because bigtech lenders possess valuable credit history data, there is gain from trade for bigtech lenders to collaborate with tradi-

[10]The borrower's outside option of not borrowing from the current lender is to find another bigtech lender who is uninformed of her credit history. In that case, the new lender would offer the same interest rate as the borrower gets from the current lender. As a result, neither a good-type nor a bad-type borrower has an incentive to borrow from a new lender were there a tiny cost of searching and switching.

tional banks or other low-funding-cost financiers. Such a bigtech-bank collaboration combines a bigtech lender's screening advantage with a bank's low funding cost. In the next section, we study the impact of introducing bigtech-bank collaborations under various contractual and regulatory environments.

3 BigTech-Bank Collaborations

By lending to ex-ante risky borrowers rationed by banks, bigtech lenders create valuable credit history data, which can be valuable to banks in making lending decisions. In this section, we study how bigtech-bank collaborations under different contractual environments affect bigtech screening and financial inclusion in the economy. A bigtech-bank collaboration requires both the bigtech lender's screening skills and the bank's low-cost funding. If there are no contractual frictions, the bank would directly fund the loan to the borrower while the bigtech lender screens at the optimal level that balances marginal benefit and cost. We introduce two additional assumptions to make the interaction between bigtechs and banks more realistic.

First, we assume that bigtech lenders and banks cannot contract on the screening intensity. In reality, the screening costs incurred may include the development of machine learning models and the hiring of data scientists that may not be observable to external parties. This assumption also effectively prevents the bank from acquiring the bigtech lender, as such an acquisition would make the screening technology controllable by the bank. With this assumption in mind, the bigtech lender in our model is better understood as bigtech companies like Google, Apple, or Ant Financial Group, which may either be too large and complex to be acquired by traditional banks due to their other lines of business.[11] Second, we assume that the bigtech lender is protected by limited liability. Under this assumption, the bank cannot force the bigtech lender to cover its losses in the event of

[11] While our framework focuses on the lending function within these bigtech companies, in practice due to the non-rivalrousness of data (Jones and Tonetti 2020; Charoenwong et al. 2024), the data generated by the bigtech is likely not only used for the borrower screening function and may be used for other things like targeted ads. These other revenue streams by the data generator also make acquisition of the bigtech lender untenable, so we believe this assumption is realistic.

a default on a bigtech-related loan.[12]

Under these two assumptions, the only way to incentivize bigtech screening is for the bigtech lender to hold a fraction of the loan for a certain period, which is costly due to the funding-cost difference between the bank and the bigtech lender. To study the bigtech-bank collaboration, we set the timing of the model as the following:

1. At date 0, the bigtech lender enters into a collaboration contract with a bank.

2. Then, the bigtech lender is matched with a borrower with $\pi < \pi_b$, and the lender can choose the screening intensity θ and lend if receiving a good signal at the interest rate of $R = y$.

3. The bigtech lender and the bank execute the collaboration contract at a pre-determined date specified in the contract.

In terms of the collaboration contract, we study three different types of collaboration contracts the bank and the bigtech lender collaborate at different levels: (1) funding-level collaboration (Section 3.1), (2) data-level collaboration (Section 3.2), and (3) asset-level collaboration (Section 3.3). Section 3.4 compares these contracts and discusses their welfare implications. Note that after the bigtech lender initiates the collaboration contract, the borrower's credit history is now privately owned by the collaborating bank, who can capture information rents by keeping the interest rate unchanged even when the borrower's credit quality improves over time (conditional on no defaults). To simplify the analysis, we assume that the bigtech lender has full bargaining power when collaborating with the bank.

3.1 Funding-level Collaboration

First, we analyze the funding-level collaboration contract in which the bank contributes funding to the lending. In reality, such a contract can be carried out in several ways. For instance, the bigtech lender and the bank make a syndicated loan, also known as joint lending, to the borrower after it

[12] If the bigtech lender is allowed to compensate the bank when the bigtech loan defaults, the optimal contract would trivially have the bigtech lender bear all the lending risks and the bank make a risk-free loan to the bigtech lender. Another way to rationalize this assumption is that if the bank lends to the bigtech lender, the lending is subject to a stringent capital requirement that eliminates the bank's funding-cost advantage.

passes the bigtech screening. Alternatively, the bank can just finance the bigtech lender for it to make the lending.

In a funding-level collaboration, we assume the bigtech lender executes the collaboration contract right after being matched with a borrower at date 0. The contract specifies two variables $\{\alpha, P\}$. α captures the bigtech lender's share of the loan. We assume the bigtech lender and the bank receive interest and principal payments that are proportional to their initial loan contributions. P captures the one-time service fee paid from the bank to the bigtech lender at date 0.

Given the funding-level collaboration contract $\{\alpha, P\}$, the bigtech lender chooses the screening intensity to maximize her expected payoff from a borrower with $\pi < \pi_b$, that is,

$$V_1(\alpha, P) = \max_{\theta \in [0,1]} P + \alpha\pi(\frac{y+\lambda}{\rho+\lambda} - 1) + \alpha(1-\pi)(1-\theta)(\frac{y+\lambda}{\rho+\lambda+\eta} - 1) - c(\theta).$$

The optimal screening $\theta = \theta^*(\alpha)$ solves $c'(\theta^*(\alpha)) = \alpha(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$, equalizing the marginal cost and benefit screening if the solution is in the interior region $\theta^*(\alpha) \in (0,1)$. Similar to Section 2, the interest rate of the loan is $R = y$ because the borrower is underserved by banks and is connected with only one bigtech lender that can effectively screen with the borrower's alternative data.

At the beginning of date 0, the bigtech lender chooses the optimal contract to maximize her expected payoff, subject to the bank's participation constraint,

$$\max_{\alpha \in [0,1]} V_1(\alpha, P) \quad \text{subject to}$$
$$P \geq (1-\alpha)[\pi(\frac{y+\lambda+b}{\rho+\lambda} - 1) + (1-\pi)(1-\theta^*(\alpha))(\frac{y+\lambda+b}{\rho+\lambda+\eta} - 1)].$$

In the optimal contract, the bigtech loan share and the screening intensity (α, θ) must solve the following two equations:

$$c'(\theta) = \alpha L_1 \tag{3}$$

$$(1-\alpha)L_2\frac{L_1}{c''(\theta)} = b(\frac{\pi}{\rho+\lambda}+\frac{(1-\pi)(1-\theta)}{\rho+\lambda+\eta}), \tag{4}$$

where $L_1 = (1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$ and $L_2 = (1-\pi)(1-\frac{y+\lambda+b}{\rho+\lambda+\eta})$. Equation 3 is the optimality condition for choosing the screening intensity, which equalizes the marginal cost and benefit of screening, taking the funding-level collaboration contract as given. Equation 4 is the optimality condition of the bigtech loan share in the optimal contract. The left-hand side of the equation is the marginal benefit of increasing the bigtech lender's loan share. Intuitively, increasing the bigtech lender's loan share at the margin gives the lender more incentives to screen and thus raises the screening intensity by $\frac{L_1}{c''(\theta)}$, and one unit of increase in the screening intensity at the margin would increase the bigtech lender's expected payoff by $(1-\alpha)L_2$. The right-hand side of the equation is the marginal cost of increasing the bigtech loan share, which captures the discounted sum of all forgone bank funding-cost savings. To establish the uniqueness of the optimal contract, we make the following assumptions about model parameters:

Assumption 1. *$c'''(\theta) \geq 0$, $y < \rho+\eta-2b$, and $b < \eta(1-\pi)$.*

Under Assumption 1, the optimal funding-level collaboration contract is unique. Denote the bigtech loan share, the screening intensity, and the bigtech lender's expected payoff by α_1^*, θ_1^*, and V_1^*, respectively.

Because our model is a one-loan lender, the optimal contract can be interpreted either as the bank and the bigtech lender making a syndicated loan or as the bank financing the bigtech lender for its lending business. The bigtech lender has skin-in-the-game to screen because she holds a fraction of the loan and thus bears loan losses proportional to her initial loan contribution. Because we assume that the recovery value of a failed project is zero, the seniority structure of the bank's and the bigtech lender's loan does not affect the outcome of the model. If we assume the recovery value is positive, the collaboration contract can specify how the bigtech lender and the bank will

divide the recovery value. In this case, we conjecture that the optimal contract would give the lowest possible recovery value to the bigtech lender, effectively making the bigtech lender's claim junior to the bank's. Reducing the bigtech lender's recovery value from a failed project creates a higher incentive for the bigtech lender to invest in better screening technology, allowing the bigtech lender to earn a higher profit and improve financial inclusion from bigtech-bank collaboration. Therefore, when the recovery value is positive, the contract can also be interpreted either as the bank funding the bigtech lender via senior claims such as debt or preferred shares or as the bigtech lender selling the senior tranche of the loan to the bank.

3.2 Data-level Collaboration

Now we analyze the data-level collaboration contract in which the bigtech lender sells the borrower's screening and credit history data to the bank. Under this contract, the the bigtech lender screens and lends to the borrower first. Then, the bigtech lender sells the borrower's screening and credit history data to the bank for subsequent lending. In reality, such data selling can be implemented by bigtech firms providing alternative credit scoring service to banks. In practice, these services are typically offered via Application Programming Interfaces. For example, providers in the United States, United Kingdom, and Brazil include Plaid, Finicity (part of MasterCard), Genify, Quovo, True Layer, and Instantor. Banks would only make loans to new customers for which there are sufficient time series data.

In this case, the collaboration contract specifies two variables $\{\tau, P\}$. τ captures the predetermined date at which the bigtech lender executes the contract and sells the borrower's credit history data. Here we implicitly assume that the borrower's past credit history is verifiable and contractable, so the bank has no concern about the authenticity of the borrower's credit history data. We will discuss more about this assumption in detail in Section 4. P captures the one-time service fee paid from the bank to the bigtech lender at date τ, which can also be interpreted as the selling price of the credit history data.

Given a data-level collaboration contract $\{\tau, P\}$, the bigtech lender chooses the screening intensity to maximize her expected payoff from a borrower with $\pi < \pi_b$, that is,

$$V_2(\tau,P) = \begin{array}{l} \max_{\theta\in[0,1]} e^{-(\rho+\lambda)\tau}P + (1-e^{-(\rho+\lambda)\tau})\pi(\frac{y+\lambda}{\rho+\lambda}-1) \\ +(1-e^{-(\rho+\lambda+\eta)\tau})(1-\pi)(1-\theta)(\frac{y+\lambda}{\rho+\lambda+\eta}-1)-c(\theta). \end{array}$$

In the above expression, note that the service fee or the price of the credit history data is discounted at the rate of ρ, the discount rate, plus λ, the maturity rate of the project. Therefore, the collaboration contract is executed if and only if the borrower's project has not matured at date τ. Unlike in the funding-level collaboration where the bigtech lender holds a constant α share of the loan regardless of borrower type, in the data-level collaboration, the bigtech lender effectively enjoys $(1-e^{-(\rho+\lambda)\tau})$ fraction of loan profits from a good-type borrower while suffers $(1-e^{-(\rho+\lambda+\eta)\tau})$ fraction of loan losses from a bad-type borrower. The optimal screening $\theta=\theta^*(\tau)$ solves $c'(\theta^*(\tau)) = (1-e^{-(\rho+\lambda+\eta)\tau})(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$, equalizing the marginal cost and benefit from screening, assuming an interior solution. Again, the interest rate of the loan takes away all the pledgeable cashflows from the borrower who has no outside options.

At the beginning of date 0, the bigtech lender chooses the optimal contract to maximize her expected payoff, subject to the bank's participation constraint,

$$\begin{array}{l} \max_{\tau\geq 0} V_2(\tau,P) \quad \text{subject to} \\ P \geq \pi(\dfrac{y+\lambda+b}{\rho+\lambda}-1) + e^{-\eta\tau}(1-\pi)(1-\theta^*(\tau))(\dfrac{y+\lambda+b}{\rho+\lambda+\eta}-1). \end{array}$$

At date τ, the bank purchases the borrower's credit history data if the borrower's project hasn't matured. For the bank, the probability that the borrower has a good credit history, that is, the probability that the borrower has passed the initial screening and hasn't defaulted on the loan, is $\pi + e^{-\eta\tau}(1-\pi)(1\text{-}\theta^*(\tau))$. After purchasing the credit history data, the bank will lend to the borrower (for him to refinance the bigtech loan) if he has a good credit history. Note that the value of credit history data is increasing in τ, because the posterior credit quality of the borrower (conditional on no defaults) $\pi_\tau = \frac{\pi}{\pi+(1-\pi)(1-\theta)e^{-\eta\tau}}$ is higher when the length of his good credit

history becomes longer.

In the optimal contract, the length of the borrower's credit history and the screening intensity (τ, θ) solve the following two equations:

$$c'(\theta) = (1 - e^{-(\rho+\lambda+\eta)\tau})L_1 \tag{5}$$

$$L_2 \frac{(\rho+\lambda+\eta)e^{-(\rho+\lambda+\eta)\tau}L_1}{c''(\theta)} = b(e^{\eta\tau}\pi + (1-\pi)(1-\theta)), \tag{6}$$

where $L_1 = (1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$ and $L_2 = (1-\pi)(1-\frac{y+\lambda+b}{\rho+\lambda+\eta})$. Equation 5 is the optimality condition for choosing the screening intensity, which equalizes the marginal cost and benefit of screening, taking the data-level collaboration contract as given. Equation 6 is the optimality condition of the bigtech lender's lending period length in the optimal contract. Intuitively, increasing the credit history length raises the total surplus because it requires the bigtech lender to bear the loan risk for a longer period of time and thus provides her with a higher screening incentive. Nevertheless, it is costly because selling the credit history data at a later date reduces the benefit of the bank's funding-cost advantage in a bigtech-bank collaboration.

Similar to Section 3.1, we can show that the optimal data-level collaboration contract is unique as long as Assumption 1 holds. Denote the credit history length, the screening intensity, and the bigtech lender's expected payoff under the optimal data-level collaboration contract by τ_2^*, θ_2^*, and V_2^*, respectively. In the optimal contract, bigtech lenders serve as an on-ramp to the financial system by lending to risky borrowers for the first τ^* period and then selling borrowers' credit history data to banks, who take over subsequent lending. In this case, a bigtech lender has skin in the game to screen because she holds the entire loan and bears the loan risk for the initial financing period.

3.3 Asset-level Collaboration

Last, we study the asset-level collaboration contract in which the bigtech lender sells the loan to the bank. Under this contract, the bigtech lender screens and lends to the borrower first. Then,

the bigtech lender sells the entire loan to the bank. In practice, such a loan-level collaboration can be implemented via securization. However, different from the funding-level collaboration contract, the asset-level collaboration contract forces the bigtech lender to sell the entire loan to the bank. Compared with data-level collaboration contracts, asset-level collaboration contracts are executed only for non-defaulted outstanding loans, while data-selling contracts are executed for all borrowers regardless of their default status.

The asset-level collaboration contract specifies two variables $\{\tau, P\}$. τ captures the pre-determined date at which the bigtech lender executes the contract and sells the loan to the bank. P captures the one-time service fee paid from the bank to the bigtech lender at date τ, which can also be interpreted as the price of the asset/loan. Note that the bigtech lender has an outstanding loan to sell at date τ if and only if the borrower has passed the initial screening and hasn't defaulted before that date. Therefore, under the contract, the bigtech lender is effectively selling the credit history data of those borrowers with good credit records.

Given the asset-level collaboration contract $\{\tau, P\}$, the bigtech lender chooses the screening intensity to maximize her expected payoff from a borrower with $\pi < \pi_b$, that is,

$$V_3(\tau, P) = \max_{\theta \in [0,1]} e^{-(\rho+\lambda)\tau}(\pi + (1-\pi)(1-\theta)e^{-\eta\tau})P + (1-e^{-(\rho+\lambda)\tau})\pi(\tfrac{y+\lambda}{\rho+\lambda} - 1) + (1-e^{-(\rho+\lambda+\eta)\tau})(1-\pi)(1-\theta)(\tfrac{y+\lambda}{\rho+\lambda+\eta} - 1) - c(\theta).$$

The bigtech lender's payoff under the asset-level collaboration contract is very similar to that under the data-level collaboration contract, except for one big difference. Under the loan-selling contract, the service fee or the price of the loan is paid from the bank to the bigtech lender only when the borrower has a good credit history at date τ. The probability of that event is given by $(\pi + (1 - \pi)(1-\theta)e^{-\eta\tau})$.

The optimal screening $\theta = \theta^*(\tau)$ solves $c'(\theta^*(\tau)) + e^{-(\rho+\lambda+\eta)\tau}(1-\pi)P = (1-e^{-(\rho+\lambda+\eta)\tau})(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$, equalizing the marginal cost and benefit of screening. Under the asset-level collaboration contract, increasing screening intensity not only raises the direct screening cost, but

also reduces the expected proceeds from future loan selling. Therefore, compared to the data-level collaboration, the asset-level collaboration gives the lender a disincentive to screen, all else being equal. In other words, it is *less efficient* to incentivize bigtech screening under a *asset-level collaboration* than under a *data-level collaboration*, all else being equal.

At the beginning of date 0, the bigtech lender chooses the optimal contract to maximize her expected payoff, subject to the bank's participation constraint,

$$\max_{\tau \geq 0} V_3(\tau, P) \quad \text{subject to}$$
$$P \geq \frac{\pi}{(\pi+(1-\pi)(1-\theta^*(\tau))e^{-\eta\tau})}(\frac{y+\lambda+b}{\rho+\lambda}-1)+\frac{(1-\pi)(1-\theta^*(\tau))e^{-\eta\tau}}{(\pi+(1-\pi)(1-\theta^*(\tau))e^{-\eta\tau})}(\frac{y+\lambda+b}{\rho+\lambda+\eta}-1)].$$

At date τ, the bank purchases the loan if and only if the borrower has a good credit history at date τ. For such a borrower, the posterior belief about the probability that the borrower is a good type is $\frac{\pi}{(\pi+(1-\pi)(1-\theta^*(\tau))e^{-\eta\tau})}$. In the optimal contract, the length of the borrower's credit history and the screening intensity (τ, θ) solve the above maximization problem. Denote the credit history length, the screening intensity, and the bigtech lender's expected payoff under the optimal asset-level collaboration contract by τ_3^*, θ_3^*, and V_3^*, respectively.

3.4 Data-as-a-Service as the Optimal Contract

Now we compare the bigtech lender's expected profits and screening intensity under four scenarios: (1) bigtech lending with no collaborations (θ_0^*, V_0), (2) the optimal funding level collaboration contract (θ_1^*, V_1^*), (3) the optimal data-level collaboration contract (θ_2^*, V_2^*), and (4) the optimal asset-level collaboration contract (θ_3^*, V_3^*). Proposition 2 summarizes the comparison.

Proposition 2. *Assume that $\pi \in [\pi_f^0, \pi_b)$, so that the bigtech lender finds it profitable to lend to the borrower who is credit rationed by banks. We have*

1. $V_2^* > max\{V_1^*, V_3^*\} > V_0$.

2. $\theta_i^* < \theta_0$ for $i = 1, 2, 3$. There exists a b_T such that $\theta_2^* < \theta_1^*$ if $b < b_T$ and $\theta_2^* \geq \theta_1^*$ otherwise.

The proof is provided in the Appendix. Figure 1 plots the bigtech lender's expected payoffs under different collaboration contracts. As shown in the figure, allowing bigtech-bank collaboration can increase the bigtech lender's expected payoff. This is so because collaborating with a bank in lending enables the bigtech lender to reduce the effective funding cost.

Among the three types of collaborations, the optimal data-level collaboration contract generates the highest profit. First, we compare the funding-level collaboration and the data-level collaboration. The former incentivizes bigtech screening by having the bigtech lender hold a fraction of the loan, while the latter incentivizes screening by having the bigtech lender hold the loan for the initial period of time. Conditional on no defaults, the borrower's posterior credit quality improves over time. Therefore, it is more efficient to incentivize screening by letting the bigtech lender bear the loan losses for the initial period when the default risk is high. Consequently, bigtech lenders prefer selling data to banks over getting funding from banks.

Second, we compare the data-level collaboration and the asset-level collaboration. The key difference between the two is that under an asset-level collaboration, the bigtech lender has an additional disincentive to screen so as to sell the borrower's loan to the bank at a higher probability. Therefore, compared with the asset-level collaboration, the data-level collaboration is again more efficient in inducing screening effort and generates more profits for the bigtech lender.

Note that because we assume no competition among bigtech lenders, the bigtech lender always charges the highest possible interest rate $R = y$ to these borrowers who face credit rationing from banks in all four cases analyzed above. In Section 4, extend the model by introducing bigtech competition and study its implications on borrowing costs.

Figure 1: Bigtech Profits under Different Collaboration Contracts

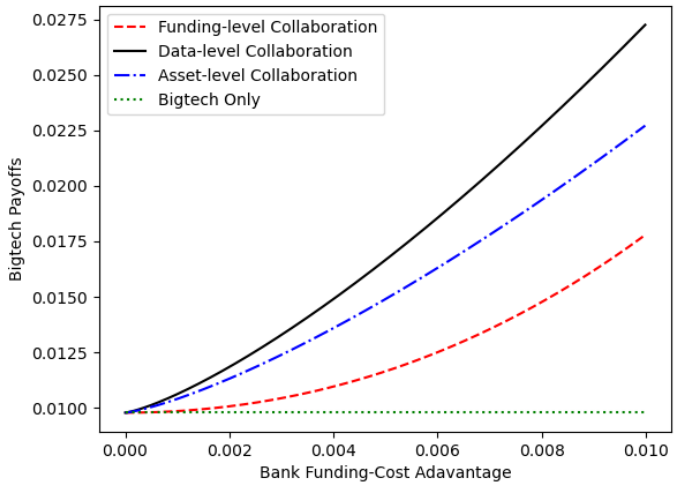
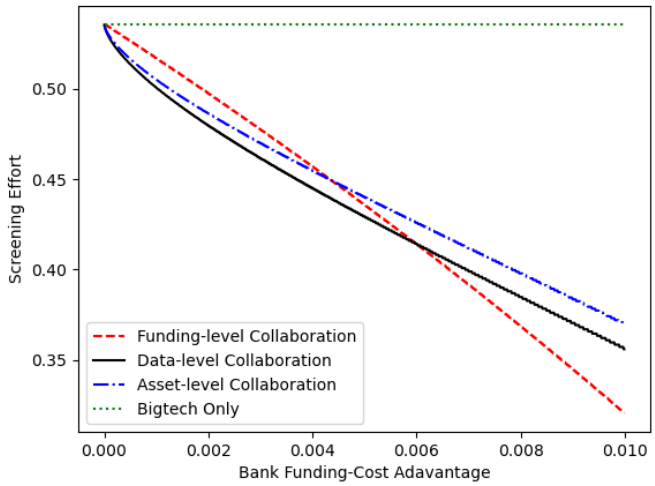


Figure 2 plots the bigtech lender's screening intensity under different collaboration contracts. We see that allowing bigtech-bank collaboration reduces the bigtech lender's screening effort under all three types of collaboration contracts. Introducing bigtech-bank collaborations reduces the bigtech lender's screening effort. When the bigtech lender and the bank collaborate, the bigtech lender is incentivized to invest more in screening technology either by lending a fraction of the loan or by making the entire lending for a certain period. In both cases, due to the funding-cost difference between bigtech lenders and banks, bigtech screening is associated with additional opportunity costs of not taking up the cheap bank funding. As a result, the bigtech screening intensity under bigtech-bank collaboration is lower than that when the bigtech lender finances the borrower alone.

Now we compare bigtech screening intensity among the three optimal collaboration contracts. Intuitively, whether the screening intensity under the optimal data-selling contract is larger or smaller than that under the other two optimal contracts is theoretically ambiguous. On the one hand, because the data-level collaboration contract is more effective in exerting screening intensity, keeping the bigtech lender's effective share of the loan fixed, the bigtech lender would screen more under the data-level collaboration. On the other hand, under the data-level collaboration

contract, the bigtech lender has an incentive to reduce her effective share of the loan to take more advantage of the bank's low funding cost, which leads to a lower screening intensity. Proposition 2 states that when comparing the optimal data-level collaboration contract to the optimal funding-level collaboration contract, the former effect is dominated by the latter if and only if the bank's funding cost advantage is relatively small.[13]

Figure 2: Bigtech Screening under Different Collaboration Contracts



In sum, we find that bigtech lenders prefer using data-level collaboration contracts to collaborate with banks. Under a data-level collaboration contract, the bigtech lender serves as an on-ramp to the traditional banking system by lending to borrowers underserved by banks, accumulating credit history data over time, and selling the data to banks who make subsequent lending to these borrowers with good credit records. Data-level collaboration has the bigtech lender bear the entire loan risk for the initial financing period, thereby inducing screening most effectively. Nevertheless, we show that compared to other forms of collaboration contracts, data-level collaboration is likely to induce a lower level of screening technology, especially when the funding-cost difference between bigtech lenders and banks is relatively small.

[13] In Figure 2, the screening intensity under the optimal data-level collaboration contract is always smaller than that under the optimal asset-level collaboration contract, however, this result is not robust and is sensitive to parameter choices.

3.5 Welfare Implications

We analyze the welfare implications of bigtech-bank collaborations focusing on the level of financial inclusion and the efficiency in choosing the screening intensity. For financial inclusion, we define π_f^i to be the lowest level of observable credit quality that the bigtech lender is willing to lend under funding-level collaboration ($i = 1$), data-level collaboration ($i = 2$), and asset-level collaboration ($i = 3$). The lower the threshold value, the higher the level of financial inclusion in the economy. Note that we can back out π_f^i by setting $V_i(\pi_f^i) = 0$ under these three types of collaboration contracts.

For screening intensity, we first determine the first-best level of screening in this economy. Specifically, we assume there is a social planner who can choose the bigtech screening intensity to maximize the total surplus from the investment project, that is, $\pi[\frac{y+\lambda}{\rho+\lambda} - 1] + (1-\pi)(1-\theta)[\frac{y+\lambda}{\rho+\lambda+\eta} - 1] - c(\theta)$. Here we implicitly assume that the bank's funding-cost advantage does not add real values to the investment surplus. This advantage comes from banks' access to deposit insurance and central bank liquidity provision, which we view as unmodeled transfers from taxpayers to banks in our setting. The first best screening intensity θ_{fb} solves

$$c'(\theta_{fb}) = (1-\pi)(1 - \frac{y+\lambda}{\rho+\lambda+\eta}).$$

We focus on borrowers with $\pi < \pi_f^0$, those who face credit rationing in the traditional banking system. For these borrowers, $\theta_{fb} = \theta_0$, where θ_0 is the optimal screening intensity with no bigtech-bank collaborations. Corollary 1 compares financial inclusion and screening intensity under different forms of bigtech-bank collaborations.

Corollary 1. *Assume $\pi < \pi_b$, we have $\pi_f^2 < min\{\pi_f^1, \pi_f^3\} < \pi_f^0$ and $\theta_{fb} = \theta_0$.*

Compared with other forms of bigtech-bank collaborations, the data-level collaboaration contract offers the highest level of financial inclusion. Intuitively, selling credit history data is the most efficient way to incentivize screening, which generates the highest possible bigtech profit and allows the lender to extend credits to risky borrowers that are otherwise underserved under other

collaboration arrangements.

Nevertheless, allowing for bigtech-bank collaborations lowers the bigtech lender's screening intensity, because the funding-cost savings from bank funding creates an opportunity cost for bigtech screening. In the baseline model where we focus on borrowers that are underserved by banks, the screening intensity with pure bigtech lending coincides with the first best level. Therefore, introducing bigtech-bank collaborations leads to under-investment in screening. Specifically, according to Proposition 2 and Corollary 1, the data-level collaboration contract, the most preferred collaboration contract in the private economy, may yield the lowest level of screening among all types of collaboration contracts.

The above result implies that for these underserved borrowers, allowing bigtech-bank collaborations typically leads to a tradeoff between financial inclusion and screening intensity. On the one hand, getting access to cheap bank funding improves the bigtech lending profit and the level of financial inclusion. On the other hand, it creates an opportunity cost for investing in screening technology and thus results in under-investment. Such a tradeoff is specifically sharp for the privately optimal data-selling contract, as it generates the highest level of financial inclusion but may cause the highest level of under-investment in bigtech screening. Nevertheless, we argue in Section 4.1 that this welfare implication can be very different when we allow for bigtech competition in the lending market.

4 Extensions and Discussions

4.1 Bigtech Lending Competition

In this section, we extend our model in two dimensions. First, we allow for competition among bigtechs in the lending market to study the model's implication of borrowing rates in the economy. We assume that there are many bigtech lenders who can screen borrowers with the same screening technology introduced in Section 2.2. Specifically, at date 0, after entering into a potential bigtech-bank collaboration contract, each bigtech lender can quote an interest rate R of her loan to attract one borrower. After being matched with a borrower, the bigtech lender screens the borrower with

an intensity level θ. Then the lender lends to the borrower at the quoted interest rate level if she receives a good screening signal.

Second, we lift the assumption $\pi < \pi_b$. Now, unlike the baseline model that focuses on risky borrowers that are underserved by banks, the extended model also studies those high-quality borrowers who can otherwise obtain loans from banks. This allows us to study how bigtech lenders compete with banks in financing low-risk borrowers. For a low-risk borrower, his borrowing rate R is likely to be smaller than the cashflow rate y. In that case, we impose the additional assumption that $y < \rho_b$, so that good-type borrowers find it optimal to immediately consume the net income (or pay out as dividends). Then bad-type borrowers must mimic the behavior of good-type borrowers, otherwise they will be recognized as bad-type borrowers and their projects will no longer be funded. As a result, there is no incentive for the borrower to keep private savings or repay the loan prematurely.

To simplify the analysis, we will discuss only two cases, the pure bigtech lending case and the data-level collaboration case. In the extended model, we can still prove that the data-level collaboration contract is the most preferred one among the three types of collaboration contracts introduced above. Therefore, focusing on the data-level collaboration contract is without loss of generality.

For the pure bigtech lending case, we first argue that given Assumption 1, the economy does not have an equilibrium in which both bigtech lenders and banks are active in lending to borrowers with a certain observable credit quality π. Intuitively, good-type borrowers are always going to pass the bigtech screening, so they choose to apply for loans from financiers who offer the lowest possible interest rate. Then bad-type borrowers have to mimic the behavior of good-type ones, otherwise they will be recognized as bad types in equilibrium and are unable to obtain financing. Therefore, borrowers with the same observable quality π either all borrow from banks or all borrow from bigtech lenders, except when bigtech lenders and banks offer the same borrowing rate. In that same-borrowing-rate case, bad-type borrowers strictly prefer borrowing from banks so that they will not be screened out. However, given $b < \eta(1-\pi)$ in Assumption 1, such a same-borrowing-

rate case cannot be an equilibrium of the model. Then for borrowers with a fixed observable credit quality, we can just compute the interest rate of bigtech lending and that of bank lending, and compare the two to determine which type of lender is active in equilibrium. [14]

For a bigtech lender, she chooses the interest rate and the screening intensity (R, θ) to minimize the interest rate offer to the market, subject to her break-even condition, that is,

$$\min_{\theta \in [0,1], R>0} R \quad \text{subject to}$$
$$c(\theta) \leq \pi[\frac{R+\lambda}{\rho+\lambda} - 1] + (1-\pi)(1-\theta)[\frac{R+\lambda}{\rho+\lambda+\eta} - 1],$$

which has a unique solution denoted by $(R_0^{**}, \theta_0^{**})$. For bank lending, the bank set the interest rate R_B to satisfy the break-even condition $\pi\frac{R_B+\lambda+b}{\rho+\lambda} + (1-\pi)\frac{R_B+\lambda+b}{\rho+\lambda+\eta} = 1$. In the equilibrium, borrowers are financed by bigtech lenders if $R_0^{**} < R_0^B$ and by banks if $R_0^{**} > R_B$.

For the data-level collaboration case, we still have bigtech lenders and banks competing by quoting interest rates in the lending market. Here, each bigtech lender first enters into a data-level collaboration contract with a bank specifying (τ, P), the credit history length of the borrower when the data is traded and the price of the data. Then, taking the collaboration contract as given, the bigtech lender chooses the interest rate and the screening intensity (R, θ) to minimize the interest rate offer to the market, subject to her break-even condition, that is,

$$R(\tau, P) = \min_{\theta \in [0,1], R>0} R \quad \text{subject to}$$
$$c(\theta) + e^{-(\rho+\lambda)\tau}P \leq (1-e^{-(\rho+\lambda)\tau})\pi[\frac{R+\lambda}{\rho+\lambda} - 1] + (1-e^{-(\rho+\lambda+\eta)\tau})(1-\pi)(1-\theta)[\frac{R+\lambda}{\rho+\lambda+\eta} - 1].$$

Denote the solution by $\theta^*(\tau, P)$ and $R^*(\tau, P)$. Given this expectation about the screening intensity and interest rate, the optimal data-level collaboration contract is chosen to minimize the

[14] For off-equilibrium beliefs, we assume that an inactive lender believes that a matched borrower is a good type with probability π.

interest rate quoted to the market, due to perfect competition, subject to the bank's participation constraint in the collaboration, that is

$$\min_{P\geq 0,\tau\geq 0} R^*(\tau,P) \quad \text{subject to}$$
$$P \leq \pi[\frac{R^*+\lambda}{\rho+\lambda}-1]+e^{-\eta\tau}(1-\pi)(1-\theta^*)[\frac{R^*+\lambda}{\rho+\lambda+\eta}-1].$$

Denote the optimal collaboration contract by (τ_2^{**},P_2^{**}), and let $\theta_2^{**}=\theta^*(\tau_2^{**},P_2^{**})$ and $R_2^{**}=R^*(\tau_2^{**},P_2^{**})$. In the equilibrium, borrowers are financed by bigtech lenders (with bigtech-bank collaboration arrangements) if $R_2^{**}<R_0^B$ and by banks if $R_2^{**}>R_0^B$.

Proposition 3 summarizes the effect of introducing bigtech lending competition and the effect of allowing bigtech-bank collaboration (at the data level) when bigtech lenders and banks perfectly compete in the lending market.

Proposition 3. *Assume that $b<\bar{b}$, $y\in(y_1,y_2)$, and that assumption 1 holds. We have*

1. In the pure bigtech lending case, there exists a π_b^0 such that 1. if $\pi\in[\pi_f^0,\pi_b^0)$, bigtech lenders screen borrowers at intensity level θ_0^{**} and lend to borrowers with good screening signals at the interest rate R_0^{**}, and 2. if $\pi>\pi_b^0$, banks finance borrowers at the interest rate R_B.

2. In the data-level collaboration case, there exists a π_b^2 such that 1. if $\pi\in[\pi_f^2,\pi_b^2)$, bigtech lenders screen borrowers at intensity level θ_2^{**} and lend to borrowers with good screening signals at the interest rate R_2^{**}, and 2. if $\pi>\pi_b^2$, banks finance borrowers at the interest rate R_B.

3. $\pi_f^2<\pi_f^0$ and $\pi_b^2>\pi_b^0>\pi_b$. When $\pi\in[\pi_f^0,\pi_b^0)$, R_0^{**}>R_2^{**}, $\theta_0^{**}>\theta_2^{**}$ and $\theta_0^{**}>\theta_{fb}$.

Proposition 3 states that for both cases, bigtech lenders and banks tend to serve borrowers with different risk profiles. Banks can provide lower interest rates for low-risk borrowers because of their low funding costs, while bigtech lenders out-compete banks in financing high-risk borrowers

due to their advanced screening technology. Specifically, we have $\pi_b^2 > \pi_b^0 > \pi_b$. Therefore, in both cases, bigtech lenders not only lend to borrowers that are underserved by banks, but also outcompete banks in financing some borrowers that can otherwise get bank funding. In other words, besides improving financial inclusion for high-risk borrowers, bigtech lenders also benefit some low-risk borrowers by reducing their borrowing costs.

Comparing these two cases in the extended model, we find that bigtech lenders can expand their lending businesses if they are allowed to collaborate with banks under data-level collaboration contracts. The ability to utilize cheap bank funding through bigtech-bank collaboration allows the bigtech lender to provide financing to borrowers that otherwise face credit rationing ($\pi_f^2 < \pi_f^0$), and it also allows the bigtech lender to steal banks' customers in the lending market ($\pi_b^2 > \pi_b^0$). Therefore, banks' willingness to collaborate with bigtech lenders may lead to a decrease in the total share of bank financing in the lending market.

However, different from the baseline model, the effect of bigtech-bank collaboration on the screening intensity in the extended model is ambiguous. In the case of pure bigtech lending, Proposition 3 says that the private economy over-invests in screening technology. Intuitively, due to asymmetric information about borrower types and the negative NPV assumption about bad-type borrowers, the equilibrium is inevitably pooling as bad-type borrowers mimic the behavior of their good-type counterparts. In such a pooling equilibrium, the bigtech lender's objective is to minimize the interest rate quoted to the market, which would maximize the expected payoff of good-type borrowers. In contrast, the first-best screening intensity would maximize the expected payoff from the project, which takes potential bad-type borrowers' payoffs into account. Consequently, the private economy with pure bigtech lending tends to invest excessively in screening technology.

Therefore, our extended model suggests limiting bigtech lenders' investment in screening technology can improve overall welfare. However, directly regulating these investments may prove challenging as they are observable or measurable. According to the extended model, encouraging collaborations between bigtech lenders and banks can be an alternative way to alleviate the over-investment in bigtech screening technology. Allowing bigtech lenders to collaborate with banks or

other lower-funding-cost lenders increases the opportunity cost of screening, as the bigtech lender now needs to contribute their high-funding-cost capital to demonstrate a commitment to proper screening. Consequently, allowing bigtech-bank collaborations can effectively mitigate the issue of overinvestment in bigtech screening technologies. However, whether the screening intensity under the optimal data-level collaboration contract is higher than the first-best level is ambiguous. Underinvestment is likely to occur when the bank's relative funding-cost advantage is large. In that case, bigtech-bank collaborations rely too little on bigtech funding, causing a lack of skin-in-the-game for screening.

To sum up, in the extended model with bigtech lending competition, bigtech-bank collaboration can be viewed as a way to reduce bigtech lenders' over-investment in screening technology. In the real world, over-investment in bigtech screening can be socially costly not only due to the excessive deployment of computing powers but also due to concerns about data privacy. Encouraging bigtech-bank collaboration can reduce such excessive investment, but under certain conditions, it may lead to under-investment.

In this regard, government regulation may be needed to nudge bigtech lenders and banks to collaborate in a way that strikes a balance among financial inclusion, borrowing costs, and screening intensity. For instance, in regions where traditional banks enjoy a large funding-cost advantage, although allowing bigtech-bank collaboration can greatly reduce borrowing costs and improve financial inclusion, certain regulatory measures need to be taken to increase bigtech lenders' skin-in-the-game when investing in their screening technologies. If left unregulated, bigtech lenders may under-invest in screening, resulting in too many bad projects being financed. As these loans to bad projects may eventually enter into banks' balance sheets under bigtech-bank collaboration contracts, unregulated collaborations between bigtech lenders and banks may post a significant systemic risk to the traditional banking system.

4.2 Data Exclusivity and Reverse-Engineering Concerns

Our analysis above implicitly assumes that the bigtech lender can exclusively let one bank use the borrower's credit history data. However, in practice, the data cannot be used exclusively to one

entity if there is no mechanism to restrict the duplication and resale of the data for other uses. This property of data is known as non-rivalrousness (Jones and Tonetti 2020). In our setting, a bank may not be willing to purchase or subscribe the borrower's credit history data knowing that other banks may do so because competition in the lending market would reduce the information edge of the credit history data. Thus, the viability of the data-selling contract depends crucially on whether the economy has a mechanism to facilitate exclusive use of data or, more generally speaking, limited data resalability.

Exclusive use of data can be implemented in at least two ways. One way is to have the bigtech lender facilitate and manage the bank loan associated with the sale of credit history data. For instance, the optimal data-level collaboration contract can require the bigtech lender to receive the service fee after she successfully helps facilitate the bank loans to those borrowers with good credit histories. This is an indirect incentive-based method to ensure data exclusivity; the participation of the bigtech lender in the lending can assure the bank that the use of credit history data is exclusive.

Another way to ensure exclusive use of credit history data, even in settings with many sellers and buyers, is possible due to recent developments in financial technology due to tokenization, a method that combines cryptography with record-keeping on a ledger. Data tokenization is a process in which sensitive data is replaced with a unique identifier called a token which is mapped back to the original data through the tokenization system. The tokenization allows security, anonymity, and most importantly, access control and auditability. For example, a token may be constructed with an auditable and automatically executing smart contract which ensures that not only is the initial data encrypted, but no copies of the data can be made, and access and ownership of that data "token" can also be tied to a private key which can be transferred (Chaleenutthawut et al.). The tokenization system can also be externally audited by a trusted party to ensure the bigtech lender has not made copies of the data before encryption (Goharshady 2021). Once transferred, the originating party can no longer view the data.[15] By tokenizing data, firms are able to exclusively

[15]Data tokens in the cryptocurrency ecosystem frequently contain such a feature. One example of such a system is Ocean Protocol, which tokenizes data to transfer ownership and access in a privacy-protected manner. In this case, a credit scoring system at a bank need only take in "blinded" tokens for analytics without any human directly viewing the data.

transfer it to another entity while maintaining its security and privacy. This allows for the secure exchange of data without exposing the actual sensitive information to unauthorized parties. This is a direct contract-based method to ensure data exclusivity.

Another related concern about data-level collaboration is that the collaborating bank may use the credit history data of these relevant borrowers to reverse-engineer the bigtech lender's screening model. The extent to which a bigtech lender concerned about such reverse engineering depends crucially on the source of the bigtech lender's screening advantage. If the bigtech lender's screening advantage comes largely from its collection or possession of alternative data, it is very hard for the bank to distil useful information from analyzing borrowers' credit history data together with standard credit-risk-relevant data. In this case, data-level collaboration would not raise much concerns about reverse engineering, and thus should be preferred due to its effectiveness in inducing screening. However, in practice, if the bigtech lender's screening advantage comes largely from applying advanced machine learning techniques on standard financial data, bigtech lenders should be more concerned about reverse engineering, so they may instead choose funding-level or asset-level collaborations, even when data-level collaborations are more efficient in reducing lending risks.

5 Conclusion

In this paper, we present a model of the bigtech-bank lending ecosystem that allows bigtech lenders and banks to compete and collaborate in various ways. In the model, encouraging bigtech-bank collaborations can improve financial inclusion and reduce excessive investment in screening technology. Comparing different collaboration models, we find data-level collaboration provides a more effective way to incentivize bigtech investments in screening technology. Under such collaboration contracts, bigtech lenders serve as an on-ramp to the financial system by selling borrowers' credit history data to banks, who take over subsequent lending. Introducing bigtech-bank collaborations improves financial inclusion but reduces bigtech investment in screening. However, the decrease in screening investment can be welfare improving or decreasing, depending on bigtech lenders' funding-cost disadvantage. Therefore, we do not expect to see a one-size-fits-all finan-

cial regulation on the lending ecosystem but expect to find different types of economies reacting differently.

References

Arner, D. W., Castellano, G. G., and Selga, E. (2023). Financial data governance: The datafication of finance, the rise of open banking and the end of the data centralization paradigm. *University of Hong Kong Faculty of Law Research Paper*.

Bennardo, A., Pagano, M., and Piccolo, S. (2015). Multiple bank lending, creditor rights, and information sharing. *Review of Finance*, 19(2):519–570.

Berg, T., Burg, V., Gombović, A., and Puri, M. (2020). On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies*, 33(7):2845–2897.

Berg, T., Fuster, A., and Puri, M. (2022). Fintech lending. *Annual Review of Financial Economics*, 14(1):187–207.

Berger, A. N., Cerqueiro, G., and Penas, M. F. (2015). Market Size Structure and Small Business Lending: Are Crisis Times Different from Normal Times? *Review of Finance*, 19(5):1965–1995.

Borgogno, O. and Colangelo, G. (2020). The data sharing paradox: BigTechs in finance. *European Competition Journal*, 16(2-3):492–511.

Bradford, T. (2023). âgive me some credit!â: Using alternative data to expand credit access. *Payments System Research Briefing*, pages 1–6.

Buchak, G., Matvos, G., Piskorski, T., and Seru, A. (2018b). Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of financial economics*, 130(3):453–483.

Chaleenutthawut, Y., Davydov, V., Kuzmin, A., and Yanovich, Y. (2021). Practical blockchain-based financial assets tokenization. In *Proceedings of the 2021 4th International Conference on Blockchain Technology and Applications*, pages 51–57.

Charoenwong, B., Kowaleski, Z. T., Kwan, A., and Sutherland, A. G. (2024). Regtech: Technology-driven compliance and its effects on profitability, operations, and market structure. *Journal of Financial Economics*, 154:103792.

Charoenwong, B. and Kwan, A. (2021). Alternative data, big data, and applications to finance. In *Fintech with Artificial Intelligence, Big Data, and Blockchain*, pages 35–105. Springer.

Chatterjee, S., Corbae, D., Dempsey, K., and Rios-Rull, J.-V. (2020). A Quantitative Theory of the Credit Score. *SSRN Electronic Journal*.

Chioda, L., Gertler, P., Higgins, S., and Medina, P. (2024). Fintech lending to borrowers with no credit history. *Working Paper*.

Cole, R. and Sokolyk, T. (2016). Who needs credit and who gets credit? Evidence from the surveys of small business finances. *Journal of Financial Stability*, 24(March 2014):40–60.

Cornelli, G., Frost, J., Gambacorta, L., Rau, R., Wardrop, R., and Ziegler, T. (2020). Fintech and big tech credit-a new database. *BIS Working Papers*, (887).

Council, F. B. (2021). How fintech lending trends benefit small businesses. *Forbes*. Accessed on 2023-10-29.

Dai, L., Han, J., Shi, J., and Zhang, B. (2020). Digital Footprints as Collateral for Debt Collection. *Working Paper*, pages 1–61.

Dash, R., Kremer, A., and Petrov, A. (2021). Designing next-generation credit-decisioning models. *McKinsey & Company, White paper, Dec*.

Djeundje, V. B., Crook, J., Calabrese, R., and Hamid, M. (2021). Enhancing credit scoring with alternative data. *Expert Systems with Applications*, 163:113766.

Duarte, J., Siegel, S., and Young, L. (2012). Trust and credit: The role of appearance in peer-to-peer lending. *Review of Financial Studies*, 25(8):2455–2483.

Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

Farboodi, M., Mihet, R., Philippon, T., and Veldkamp, L. (2019). Big Data and Firm Dynamics. *AEA Papers and Proceedings*, 109:38–42.

Frost, J., Gambacorta, L., Huang, Y., Shin, H. S., and Zbinden, P. (2019). Bigtech and the changing structure of financial intermediation. *Economic Policy*, 34(100):761–799.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The Role of Technology in Mortgage Lending. *Review of Financial Studies*, 32(5):1854–1899.

Goharshady, A. K. (2021). Irrationality, extortion, or trusted third-parties: Why it is impossible to buy and sell physical goods securely on the blockchain. In *2021 IEEE International Conference on Blockchain (Blockchain)*, pages 73–81. IEEE.

Hau, H., Huang, Y., Shan, H., and Sheng, Z. (2018). Fintech credit, financial inclusion and entrepreneurial growth. *Working Paper*.

He, Z., Huang, J., and Zhou, J. (2023). Open banking: Credit market competition when borrowers own the data. *Journal of financial economics*, 147(2):449–474.

Huang, J. (2021). Fintech expansion. *SSRN Working Paper*, 3957688.

Jagtiani, J. and Lemieux, C. (2018). Do fintech lenders penetrate areas that are underserved by traditional banks? *Journal of Economics and Business*, 100(March):43–54.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American economic review*, 110(9):2819–2858.

Liu, L., Lu, G., and Xiong, W. (2022). *The big tech lending model*. Number 30160. NBER Working Paper.

Markatou, E. A. and Tamassia, R. (2019). Full database reconstruction with access and search pattern leakage. In *Information Security: 22nd International Conference, ISC 2019, New York City, NY, USA, September 16–18, 2019, Proceedings 22*, pages 25–43. Springer.

McNulty, D., Miglionico, A., and Milne, A. (2023). Data access technologies and the new governance techniques of financial regulation. *Journal of Financial Regulation*, 9(2):225–248.

Mitchell, K. and Pearce, D. K. (2011). Lending technologies, lending specialization, and minority access to small-business loans. *Small Business Economics*, 37(3):277–304.

Netzer, O., Lemaire, A., and Herzenstein, M. (2019). When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications. *Journal of Marketing Research*, 56(6):960–980.

Pagano, M. and Jappeli, T. (1993). Information Sharing in Credit Markets. *The Journal of Finance*, 48(5):1693–1718.

Parlour, C. A., Rajan, U., and Zhu, H. (2022). When fintech competes for payment flows. *The Review of Financial Studies*, 35(11):4985–5024.

Ramadorai, T., Uettwiller, A., and Walther, A. (2020). The market for data privacy.

Schweitzer, M. and Barkley, B. (2017). Is fintech good for small business borrowers? impacts on firm growth and customer satisfaction. *Working Paper*.

Stiglitz, B. J. E. and Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *American Economic Review*, 71(3):393–410.

Stulz, R. M. (2019). FinTech, BigTech, and the Future of Banks. *Journal of Applied Corporate Finance*, 31(4):86–97.

Vives, X. (2017). The impact of fintech on banking. *European Economy*, (2):97–105.

Appendix

Proof of Proposition 1

For each π, denote the bank's break-even interest rate by R_b, which is the solution to $\pi(\frac{R_b+b+\lambda}{\rho+\lambda}-1)+(1-\pi)(\frac{R_b+b+\lambda}{\rho+\lambda+\eta}-1)=0$. Denote the bigtech lender's break-even interest rate and the associated optimal screening intensity by R and θ respectively. $(R(\pi),\theta(\pi))$ solves

$$\begin{aligned} c'(\theta) &= (1-\pi)(1-\frac{R+\lambda}{\rho+\lambda+\eta}) \\ c(\theta)+c'(\theta)(1-\theta) &= \pi(\frac{R+\lambda}{\rho+\lambda}-1). \end{aligned}$$

We can rewrite the interest rate difference between the two by $R_b-R=:g(\pi)=-b+\frac{c'(\theta)\theta-c(\theta)}{\frac{\pi}{\rho+\lambda}+\frac{(1-\pi)}{\rho+\lambda+\eta}}$. Then the bigtech lender can provide financing to borrowers with $\pi<\pi_b$ iff 1.$\Phi=\{\pi\in[0,1]|g(\pi)>0\}$ is non-empty and 2. $\pi_b\in\Phi$. For the first condition, note that $g(0)=g(1)=-b$. Further we can prove that there exists a unique π_M such that $g'(\pi)>0$ if $\pi<\pi_M$ and $g'(\pi)<0$ if $\pi>\pi_M$. So$\{\pi\in[0,1]|g(\pi)>0\}$ is non-empty if and only if $g(\pi_M)>0$, that is, $b<\bar{b}=:\frac{c'(\theta(\pi_M))\theta(\pi_M)-c(\theta(\pi_M))}{\frac{\pi_M}{\rho+\lambda}+\frac{(1-\pi_M)}{\rho+\lambda+\eta}}$. For the second condition, note that if $g(\pi_M)>0$, due to the hump shape of the g function, there exists two threshold values π_1 and $\pi_2<\pi_1$ such that $g(\pi_2)=g(\pi_1)=0$. Then we can define y_i as the solution to $\pi_i(\frac{y_i+b+\lambda}{\rho+\lambda}-1)+(1-\pi_i)(\frac{y_i+b+\lambda}{\rho+\lambda+\eta}-1)=0$ for $i=1,2$. As a result, $\pi_b\in\Phi$ if and only if $y\in(y_1,y_2)$.

Therefore, if $b<\bar{b}$ and $y\in(y_1,y_2)$, the bigtech lender finds it profitable to finance some borrowers that are underserved by banks. Specifically, borrowers with $\pi\in[\pi_f^0,\pi_b)$ will be financed, where π_f^0 is the solution to $R(\pi_f^0)=y$.

Proof of Proposition 2

Firstly, we show that$V_2^* > max\{V_1^*, V_3^*\} > V_0$. Note that the bigtech lender can always choose not to collaborate with a bank. So, the bigtech lender's expected profit when it is allowed to collaborate with a bank (V_i^* for $i = 1,2,3$) should be higher than that in the case of pure bigtech lending (V_0). Further, the inequality is strict as long as $b > 0$. Then we prove that $V_2^* > V_1^*$. Imagine a special joint-lending contract with (α, P) so that the bigtech lender execute the contract at time $\tau > 0$. The bigtech's expected payoff is

$$W_1(\tau,\alpha) = \begin{array}{c} e^{-(\rho+\lambda)\tau}P + (1-e^{-(\rho+\lambda)\tau}(1-\alpha))\pi(\frac{y+\lambda}{\rho+\lambda}-1) \\ +(1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\pi)(1-\theta^*)(\frac{y+\lambda}{\rho+\lambda+\eta}-1) - c(\theta^*), \end{array}$$

where $P = (1-\alpha)\pi(\frac{y+\lambda+b}{\rho+\lambda}-1) + (1-\alpha)e^{-\eta\tau}(1-\pi)(1-\theta^*(\tau))(\frac{y+\lambda+b}{\rho+\lambda+\eta}-1)$, and $\theta^*(\tau)$ solves $c'(\theta^*) = (1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$.

Now consider an increase of τ to τ' and a decrease of α to α' such that $e^{-(\rho+\lambda+\eta)\tau}(1-\alpha)$=$e^{-(\rho+\lambda+\eta)\tau'}(1-\alpha')$ stays constant. After such a change of collaboration contract, the optimal screening intensity θ^* remains the same. But we can show that $W(\tau',\alpha') > W(\tau,\alpha)$.

As a result, for every funding-level collaboration contract (with $\alpha > 0$ and $\tau = 0$), we can increase the bigtech lender's expected profit by increasing the time length of bigtech lending and decreasing the bigtech lender's share of joint lending. The maximized profit is achieved when we push the joint-lending loan share to zero, in which case the contract becomes a data-level collaboration contract (with $\alpha = 0$ and $\tau > 0$). Therefore, the optimal data-level collaboration contract must yield a higher expected profit when compared to the optimal funding-level collaboration contract, that is, $V_2^* > V_1^*$.

To show that $V_2^* > V_3^*$, we first look at the optimal asset-level collaboration contract. Under this contract, the credit history length, the screening intensity, and the bigtech lender's expected payoff under the optimal asset-level collaboration contract are τ_3^*, θ_3^*, and V_3^*, respectively. Then we look at a data-level collaboration contract with the same credit history length τ_3^*. Denote the screening

intensity and the expected payoff under this data-selling contract by θ_3 and V_3, respectively. It is easy to see that $\theta_3^* < \theta_3 < \theta_0$. Intuitively, compared with the data-level collaboration, the asset-level collaboration gives the lender a disincentive to screen because it wants to sell more loans to the bank. Then, we have $V_3^* < V_3 < V_2^*$, the former is due to the data-level collaboration contract inducing screening more effectively, and the latter is due to the re-optimization of τ under the optimal data-level collaboration contract.

Secondly, we show that $\theta_i^* < \theta_0$ for $i = 1,2,3$ and that there exists a b_T such that $\theta_2^* < \theta_1^*$ if $b < b_T$ and $\theta_2^* \geq \theta_1^*$ otherwise. The former result is obvious when looking at the first-order conditions under these three types of collaborations contracts: as long as $\alpha_1^* > 0$, $\tau_2^* > 0$ and $\tau_3^* > 0$, the bigtech lender has a strictly lower incentive to screen when she is allowed to collaborate with a bank that has a lower funding cost. The latter result is to compare the screening intensity under the optimal data-level collaboration contract and the optimal funding-level collaboration contract. Define $\alpha_2 = (1 - e^{-(\rho+\lambda+\eta)\tau_2^*})$, and we can rewrite the two optimality condition for the optimal data-level collaboration contract as

$$\begin{aligned} c'(\theta_2^*) &= \alpha_2 L_1 \\ \frac{L_1 L_2}{c''(\theta_2^*)}(1-\alpha_2) &= b(Z(\alpha_2)\frac{\pi}{\rho+\lambda} + \frac{(1-\pi)(1-\theta_2^*)}{(\rho+\lambda+\eta)}), \end{aligned}$$

where $Z(\alpha_2) = \frac{(\rho+\lambda)}{\rho+\lambda+\eta}(1-\alpha_2)^{\frac{-\eta}{\rho+\lambda+\eta}}$. Note that these two equations are very similar to equation 3 and 4, re-written below, except for the $Z(\alpha_2)$ term.

$$\begin{aligned} c'(\theta_1^*) &= \alpha_1^* L_1 \\ (1-\alpha_1^*)\frac{L_1 L_2}{c''(\theta_1^*)} &= b(\frac{\pi}{\rho+\lambda} + \frac{(1-\pi)(1-\theta_1^*)}{\rho+\lambda+\eta}). \end{aligned}$$

Under Assumption 1, we can show that when $\theta_2^* < \theta_1^*$ if and only if $Z(\alpha_2) > 1$. Because α_2 is

decreasing in b, there exists a $b_T \geq 0$ such that $\theta_2^* < \theta_1^*$ if $b < b_T$ and $\theta_2^* \geq \theta_1^*$ otherwise.

Proof of Proposition 3.

As argued in Section 4, the economy does not have an equilibrium in which both bigtech lenders and banks are active in lending to borrowers with a certain observable credit quality π. Therefore, we can just compare the interest rate of bigtech and bank lending to determine who is active for borrowers of a given credit quality π. Under the assumption that $b < \bar{b}$ and $y \in (y_1, y_2)$, we can show that in the pure bigtech lending case,$R_b - R =: g(\pi) = -b + \frac{c'(\theta)\theta - c(\theta)}{\frac{\pi}{\rho+\lambda} + \frac{(1-\pi)}{\rho+\lambda+\eta}}$ is positive if and only if $\pi \in [\pi_f^0, \pi_1)$, both thresholds defined in the proof of Proposition 1. Borrowers with $\pi \in [\pi_f^0, \pi_b^0)$ apply for loans from bigtech lenders, while borrowers with $\pi > \pi_b^0$ apply for loans from banks, where $\pi_b^0 = \pi_1$.

In the data-level collaboration case, note that when bigtech lenders are active, we have$R_2^{**} < R_0^{**}$ and $\theta_2^{**} < \theta_0^{**}$. Introducing data-level collaboration can reduce the effective funding-cost of the lending, which leads to a lower borrowing rate, but it reduces the bigtech lender's incentive to screen. The proofs for these two results are similar to those for Proposition 2, so we leave them out. Because $R_2^{**} < R_0^{**}$, introducing bigtech-bank collaboration helps the bigtech lender expand its lending share in the market both by improving financial inclusion ($\pi_f^2 < \pi_f^0$) and by stealing customers from banks ($\pi_b^2 > \pi_b^0$).

Lastly, we show that $\theta_0^{**} > \theta_{fb}$. The privately optimal screening intensity under bigtech competition is given by $c'(\theta_0^{**}) = (1-\pi)[1 - \frac{R_0^{**}+\lambda}{\rho+\lambda+\eta}]$, while the socially optimal screening intensity satisfies $c'(\theta_{fb}) = (1-\pi)(1 - \frac{y+\lambda}{\rho+\lambda+\eta})$. As long as $R_0^{**} < y$, we have $\theta_0^{**} > \theta_{fb}$. Intuitively, bigtech lenders compete in the lending market by offering lower interest rates to attract good-type borrowers in the pooling equilibrium. As a result, the screening choice does not internalize its effect on bad-type borrowers. So, bigtech lenders tend to invest too much in screening technologies when compared to the social planner's optimal choice.