INSEAD The Business School for the World®

Working Paper

2025/32/FIN
(Revised version of 2025/11/FIN)

BigTech-Bank Collaborations in the Lending Market

Yiyao Wang
Shanghai Advanced Institute of Finance, yiyaowang@saif.sjtu.edu.cn

Ben Charoenwong
INSEAD, ben.charoenwong@insead.edu

We model lending markets where bigtechs possess superior screening technology while banks enjoy funding-cost advantages. The optimal external collaboration occurs at the data level: bigtechs initially lend to risky borrowers before selling their credit histories to banks for subsequent financing. This arrangement concentrates bigtechs' screening incentives in the highrisk early lending phase while leveraging banks' funding advantage thereafter. By serving as on-ramps to the financial system, bigtechs expand credit access to previously excluded borrowers. Under certain conditions, restricting bigtech-bank integrations and encouraging external collaborations at the data level can improve welfare by preventing excessive screening investments.

JEL Classifications: G21; G23; G33; O32; O34; L84
Keywords: Data Governance; Asymmetric Information; Relationship Lending; Bigtech

Electronic copy available at: https://ssrn.com/abstract=4785114

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 the 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.
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 – credit provision by large technology companies whose primary business is not financial services – has grown to over US$500 billion worldwide as of 2019 (Cornelli et al., 2020). This expansion reflects a fundamental complementarity in financial intermediation: bigtech firms possess proprietary user data that enhances borrower screening,[1] while traditional banks enjoy significantly lower funding costs through deposit-taking privileges and access to central bank liquidity. This complementarity has fueled collaborations that now account for 60% of global bigtech lending (Liu et al., 2022).[2] Given the increasing prevalence of these collaborations, understanding their optimal structure is crucial for the industrial organization of financial intermediationâdetermining whether technological advances will reshape market structure through collaborative specialization or integration of traditionally separate financial functions. These questions are relevant to the broader research on how technological change affects financial intermediation efficiency and market structure (Philippon, 2015, 2016).

In this paper, we develop a model based on an optimal contracting framework to analyze how bigtech and bank collaborations affect credit accessibility, screening intensity, and social welfare with three key frictions: (1) a bigtech lender's exclusive access to borrowers' alternative data for screening, (2) banks' access to low-cost funding, and (3) contractual limitations in bigtech-bank collaborations. Using its superior screening technology, the bigtech lender can extend credit to risky borrowers otherwise underserved by traditional banks. Meanwhile, the bigtech lender has an incentive to collaborate with banks in lending due to their access to low-cost funding. The central tension emerges from the inability to directly contract on screening intensity – requiring

[1] In practice, traditional banks may be unable to emulate the financial analytics capability due to more stringent regulations, legacy information technology, organization frictions, and the lack of alternative data access (Stulz, 2019). Bigtech firms with core businesses in areas like e-commerce, search, and social media bring uniquely valuable alternative data to lending markets (Berg et al., 2020; Charoenwong and Kwan, 2021).

[2] For example, Ant Financial's loans to small consumers and businesses in China exceed US$238 billion, powered by their proprietary Sesame Credit scoring system that leverages data from the affiliated Taobao e-commerce platform (Hau et al., 2018). Similarly, Amazon partners with Goldman Sachs to extend over US$1 billion in loans to small merchants using seller performance data unavailable to traditional lenders. 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.

collaboration structures that provide sufficient incentives for information production without sacrificing the benefits of low-cost bank funding. This setup allows us to analyze how the interplay between information production and funding advantages shapes the lending market organization in the presence of incomplete contracts.

Our theoretical analysis proceeds in three stages. First, we characterize the equilibrium outcomes under different collaboration structures: (1) funding-level collaboration, where banks and bigtechs jointly finance loans from inception, (2) data-level collaboration, where bigtechs share borrowers' credit history with banks after an initial lending period, and (3) asset-level collaboration, where bigtechs sell entire loans to banks. Second, we solve for the endogenous screening intensity in each arrangement and identify which structure emerges as optimal from the private perspectives of the collaborating entities. Finally, we analyze welfare implications by comparing the privately optimal arrangement against the social optimum, focusing on how the divergence between private and social incentives depends on the magnitude of banks' funding advantage and borrowers' non-pledgeable benefits. This approach allows us to derive conditions under which specific policy interventions regarding bigtech-bank relationships would enhance welfare.

We show that data-level collaboration – where bigtech lenders initially screen and fund risky borrowers then sell borrowers' credit history data to banks who then take over subsequent lending – dominates other forms of external collaborations in the private economy. With this type of collaboration, which can be implemented in practice through an alternative credit scoring service (Dash et al., 2021; Bradford, 2023), bigtech lenders effectively serve as an *on-ramp* to the traditional financial system by allowing 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, in an incentive-compatible way. Under certain conditions, external collaboration through data selling can generate a higher level of welfare than internal collaboration through bigtech-bank integration. Specifically, our model provides a rationale for the government to restrict bigtech firms' acquiring banks and instead build the necessary data-market infrastructure to facilitate data-level collaborations between lenders with different comparative advantages.

In our model, borrowers seek to finance their investment projects. Each project generates stable pledgeable cashflows and non-pledgeable benefits before project maturity. The investment project can be either a good type or a bad type. A good-type project never fails prematurely, yet a bad-type project may fail at a given Poisson arrival rate. We assume symmetric information: no one knows a borrower's project type, and the ex-ante probability of the project being a good type, referred to as the project quality or the borrower's credit quality, is common knowledge.

To finance the investment project, a borrower can apply for loans from banks or from a bigtech lender. Banks have lower funding costs but can only make loan approval decisions based on the borrower's credit quality. Meanwhile, there is a bigtech lender that faces higher funding costs but can screen these borrowers by processing and analyzing their alternative data. Specifically, the bigtech lender can choose the screening intensity, which affects the information quality of the screening signal, at a quadratic cost. The bigtech lender's unique access to alternative data allows it to reduce the lending risk and make credit accessible to borrowers facing credit rationing from the traditional banking sector.

By lending to risky borrowers, the bigtech lender generates two valuable information assets (credit history data): initial screening signals and loan performance data over time. As borrowers service their debt, their perceived credit quality improves, and the value of their credit history data increases. Due to banks' access to cheap funding, the bigtech lender has an incentive to collaborate with banks in financing borrowers. Without any contractual limitations, the bigtech lender and the bank would contract on the screening intensity and have the bank contribute all the funding. We refer to this case as internal collaboration because it can be easily implemented by integrating bigtech lenders and banks through mergers and acquisitions. In contrast, we focus mainly on cases with contractual limitations. Specifically, we assume that the bigtech lender and the bank cannot directly contract on the screening intensity. We refer to these cases as external collaboration. In an external collaboration, the bigtech lender must contribute a fraction of the funding to provide skin-in-the-game for screening.

We consider the following contract space for bigtech-bank collaborations. At date 0, the bigtech enters into a collaboration contract specifying: (1) the length of the initial period during which the bigtech lender makes the entire loan to the borrower, (2) the service fee paid to the bigtech lender, (3) whether the service fee payment is contingent on the status of the loan, and (4) the share of the loan the bank acquires after the initial period. This contract space is wide enough to span three distinct types of external collaboration structures observed in the reality: (1) funding-level collaboration, where banks and bigtechs jointly fund loans from inception; (2) data-level collaboration, where bigtechs give banks access to borrowers' credit history data after an initial bigtech-lending period; and (3) asset-level collaboration, where bigtechs sell entire loans to banks after an initial bigtech-lending period.

Data-level collaboration dominates other forms of external collaboration because it balances screening incentives with funding efficiency. By requiring the bigtech lender to bear the entire loan risk during the initial high-risk period, this kind of collaboration creates strong screening incentives precisely when they matter most. As borrower quality becomes more transparent over time as they service their debt without default, the collaboration transitions to bank funding, optimally leveraging banks' funding advantage without undermining initial screening incentives. This collaboration structure creates value by having each institution specialize in its comparative advantage: bigtechs in information production and banks in low-cost intermediation.

Our analysis challenges three prevailing views in the financial technology literature. First, contrary to the common assumption that bigtech firms will primarily compete with banks, we show that collaboration creates greater value than competition in market segments where ex ante quality is hard to gauge. Second, unlike existing models that treat information sharing as ancillary to lending relationships, we demonstrate that information is the primary asset exchanged in optimal arrangements. Third, we contradict the notion that joint lending arrangements are optimal when lenders have complementary strengths, showing instead that clear separation of functions — with bigtechs screening initially and banks funding subsequently — maximizes value.

The welfare comparison between external and internal collaboration hinges on the divergence between private and social screening incentives. From a private perspective, internal collaboration is always preferred–allowing the bigtech lender to fully leverage bank funding while maintaining optimal screening intensity. However, from a social welfare perspective, this arrangement can lead to excessive screening investment because the bigtech lender fails to internalize borrowers' non-pledgeable control rents or, more generally, their gain from the investment. External collaboration through data selling introduces a counterbalancing force. Requiring bigtechs to commit their own capital during the initial lending period increases the marginal cost of screening and reduces equilibrium screening intensity. This contractual friction, while suboptimal from a private perspective, can paradoxically enhance welfare when borrowers' non-pledgeable benefits are substantial relative to banks' funding advantage.

Our model delivers three empirical predictions that can guide future empirical work on bigtech-bank collaborations. First, markets with more prevalent bigtech-bank external collaborations should exhibit higher lending volumes to previously underserved borrowers compared to markets where such collaborations are restricted. Second, bigtech lenders in markets with strong bank funding advantages should invest less in screening technology than those in markets with weaker bank funding advantages. Third, under the optimal external collaboration, both the bigtech lending period and screening intensity decrease in the marginal screening cost and the bank's funding-cost advantage.

Our results also suggest specific policy approaches depending on market structure. Broadly, encouraging data-sharing arrangements between bigtech lenders and banks could enhance financial inclusion and market efficiency. Our paper highlights the role of financial data governance regulations (e.g., McNulty et al. 2023; Arner et al. 2023), particularly regarding the rules on credit information sharing. Importantly, we 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 scoring 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.

After discussing the related literature, the remainder of the paper proceeds as follows. Section 2 develops a model of lending with two types of lendersâbanks with funding advantages and bigtech firms with screening advantages. Section 3 analyzes and compares three specific forms of collaboration between these lenders: funding-level collaboration, where banks provide capital; data-level collaboration, where bigtechs sell credit histories; and asset-level collaboration, where loans are transferred. Section 4 explores how competition among bigtechs affects our results and addresses implementation concerns. Section 5 concludes with implications for market structure and financial regulation.

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 banks. While they focus on potential reasons why bigtech specializes 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, showing that the latter resemble applicants 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 about borrowers with each other through establishing 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 decrease 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 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 considers data-level collaboration between different lenders. In our setting, giving 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. Regarding 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 to the more recent European regulations aimed at providing a framework for fintech-bank data sharing (Borgogno and Colangelo, 2020).

2. The Model

Our model captures two main features in lending markets: complementary skills between different types of lenders and contractual limitations that prevent first-best arrangements.

The economy is populated with three types of agents: borrowers, bank lenders, and a bigtech lender. All agents are risk-neutral and discount cashflows at the rate of ρ. Time is continuous and runs to infinity.

2.1. Borrowers

There is a continuum of borrowers, which we normalize to be of measure one.[3] At date 0, each borrower is endowed with an investment project. The project requires 1 dollar of initial investment. During dt period of time, the project generates ydt cashflows, which can be pledged to lenders, and a non-pledgeable cashflow or control rent of cdt for the borrower, which cannot be pledged to lenders due to moral hazard and/or other types of contract incompleteness. The project matures with Poisson rate λ, and upon maturity, the project generates a terminal cash flow of $X = 1$.

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 η. For simplicity, we assume that the residual value of a failed project is zero. Information is symmetric in that no agent in the economy knows the type of a borrower's project. Among all borrowers, π of them have good projects, while the remaining $1-\pi$ have bad projects. π, which is assumed to be public information, measures the observable credit quality of borrowers.

[3] We use the term "borrower" rather than "entrepreneur" or "firm" because the model also applies to both firms and 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.

Assumption 1. $y \in [\rho, \rho + \eta - c]$.

Under the assumption, the net present value of the pledgeable cashflows from a good project is positive, while the net present value, including both the pledgeable cashflows and non-pledgeable control rent, of a bad project is negative.

2.2. Lenders

A borrower has no net worth at date 0, so she has to get external funding to finance her investment project. Specifically, the borrower can get external financing from either a traditional bank or a bigtech lender. 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 payments 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 the bigtech lender 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 that the bigtech lender has a superior screening technology to banks. Such a superior screening technology is interpreted as the bigtech lender's ability to collect and analyze the borrower's alternative data. These alternative data records economic activities happening between the borrower and the bigtech lender, and are thus not accessible to traditional bank lenders. In the model, we assume only the bigtech lender can generate informative signals about the borrower's creditworthiness (e.g., Council 2021). Specifically, the bigtech lender can 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 positives).

Our main model assumes that only one bigtech lender can access the screening technology. 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). In screening, the lender chooses $\theta \in [0,1]$ at a quadratic cost $C(\theta) = \frac{1}{2}k\theta^2$. 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 incurs 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 a concern about data privacy. Therefore, we interpret the screening cost in the model more broadly as combining the bigtech lender's investment in screening technology and 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.[4] 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 cannot adopt the advanced screening technology simply because they have no access to 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 highlight the bigtech lender's unique screening ability and the value of potential collaborations between bigtech lenders and banks, we focus on the group of borrowers that have no access to financing from the traditional banking sector by imposing the following assumption

Assumption 2. $\pi < \bar{\pi}$, *where* $\bar{\pi}\frac{y+\lambda+b}{\rho+\lambda} + (1-\bar{\pi})\frac{y+\lambda+b}{\rho+\lambda+\eta} = 1$.

[4]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.

Under $\pi < \bar{\pi}$, bank lenders find it too risky to lend to these borrowers, and financing cannot happen without the bigtech's screening technology. In practice, there are a few reasons why traditional banks underserve some borrowers. 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.[5]

2.3. *Bigtech Lending and Bigtech-Bank Collaborations*

Because a borrower with $\pi < \bar{\pi}$ cannot obtain bank financing, she turns to the bigtech lender, who can utilize her alternative data for credit screening. Upon receiving a loan application, the bigtech may screen the borrower with an intensity level θ. If receiving a bad signal, the bigtech lender learns that the borrower has a bad project and will reject her loan application. If receiving a good signal, the lender's posterior belief about the borrower's credit quality improves, and the lender may find it profitable to grant credit to the borrower. Denote the bigtech lender's lending rate by R. After investing, the borrower pays an interest rate of Rdt to the bigtech lender per dt period, until the project matures or fails.

Because the bigtech lender is the only one with the screening technology, it faces no competition when servicing these high-risk borrowers. Consequently, the lender can charge the highest possible interest rate, $R = y$, to extract all the pledgeable cashflows from the borrower's project, while leaving the control rent c to the borrower.[6] In Section 4, we argue that the model's main results still hold when we introduce lending market competition between multiple bigtech lenders.

Due to banks' access to low-cost funding, the bigtech lender has an incentive to collaborate with a bank in financing these high-risk borrowers. To simplify the analysis, we assume banks

[5]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.

[6]Alternatively, we can interpret $y + c$ as the total cashflow from the project, and $R = y$ as the exogenously given loan interest rate.

face perfect competition, so they are willing to collaborate with the bigtech lender as long as they break even. By combining the bigtech lender's screening skills and the bank's low-cost funding, the bigtech-bank collaboration can increase the bigtech lender's expected payoff from lending, allowing it to extend more credit to borrowers who are credit rationed by bank lenders.

3. Bigtech-Bank Collaborations

In the model, the effect of the bigtech-bank collaboration on the lending market depends crucially on the contractual environment. We first study cases where the bigtech lender collaborates with a bank through external contracting in Section 3.1. We discuss different forms of external collaboration in Section 3.2 and derive the optimal contract in 3.3. Then we analyze the case where the bigtech lender's screening intensity is contractable between the bigtech lender and the bank in Section 3.4. Lastly, we compare these collaboration models and discuss the model's welfare implications in Section 3.5.

3.1. External Collaboration

Before analyzing specific collaboration models, we establish a general framework for *external collaboration* between bigtech lenders and banks. This framework allows us to compare different arrangements and identify the optimal structure. We introduce two additional assumptions to make the external contracting between the bigtech lender and the bank 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 hiring 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, which will be analyzed in Section 3.4, would effectively 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 be too large and complex to be acquired by traditional banks due to their other lines of business.[7] Second, we assume that the

[7] 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

bigtech lender is protected by limited liability. Under this assumption, the bank cannot force the bigtech lender to cover loan losses in the event of a default on a bigtech-originated loan.[8]

Under these two assumptions, external collaboration features a tradeoff between utilizing low-cost funding and incentivizing bigtech screening. Intuitively, because the screening intensity is not contractable, the only way to incentivize bigtech screening is to have the bigtech lender bear a fraction of the loan losses. Further, because of limited liability, the only way to have the bigtech lender bear some loan losses is to have the bigtech lender hold a fraction of the loan, and/or for a certain period of time, which is costly due to the funding difference between the bank and the bigtech lender.

We assume an external collaboration contract specifies four variables (τ, P, i, α). Under this contract, the bigtech lender finances the borrower by itself before date τ. $\tau \in [0, +\infty)$ is the length of the bigtech lending period, which is also the length of the borrower's credit history at the bigtech lender.

At date τ, the bank first pays one-time upfront service fee $P \in [0, +\infty)$ to the bigtech lender under a certain condition which is captured by the term $i \in \{0, 1\}$ in the collaboration contract. $i = 0$ means that the bank needs to pay the service fee to the bigtech lender unconditionally. $i = 1$ means that the bank only pays if the borrower's loan is outstanding at date τ, which happens when the borrower has neither defaulted nor repaid her loan prior to date τ.

After paying the service fee P under the condition i, if the borrower's loan is still outstanding, the bank can pay $1-\alpha$ to obtain $1-\alpha$ fraction of the loan's future interest incomes. Here, $\alpha \in [0, 1]$ is the bigtech lender's share of lending after date τ.

Under an external collaboration contract, the timing of the model is the follows,

1. At date 0, the bigtech lender enters into a collaboration contract (τ, P, i, α) with a bank.

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.

[8] 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 makes 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.

2. The bigtech lender matches with a (representative) borrower with $\pi < \bar{\pi}$. The lender can choose the screening intensity θ and lend to the borrower at an interest rate $R = y$ if receiving a good screening signal.

3. At date τ, the bank pays the bigtech lender the service fee P according to the term i. If the borrower's loan is still outstanding, the bank pays the bigtech lender $1-\alpha$ in exchange for $1-\alpha$ fraction of the loan's future interest incomes.

We solve the model in backward induction. First, taking an external collaboration contract (τ, P, i, α) as given, the bigtech lender chooses the optimal screening intensity to maximize its expected payoff, that is,

$$\max_{\theta}[-C(\theta)+e^{-\rho\tau}q(i,\theta,\tau)P]+(1-e^{-(\rho+\lambda)\tau}(1-\alpha))\pi(\tfrac{y+\lambda}{\rho+\lambda}-1)$$
$$+(1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\pi)(1-\theta)(\tfrac{y+\lambda}{\rho+\lambda+\eta}-1).$$

The bigtech lender's expected payoff consists of three components: (1) screening costs paid at date and service fee income, (2) expected gains from financing good projects, and (3) expected losses from financing bad projects that were not screened out. The contract variables (τ, P, i, α) affect each component differently, creating tradeoffs between screening incentives and funding efficiency.

The first term includes the bank's screening cost at date 0 and the service fee income at date τ. Note that the expected value of the service fee depends on $q(i, \theta, \tau)$, the probability of the bank paying the service fee at date τ, which equals

$$q(i,\theta,\tau)=\begin{cases}1 & i=0\\ e^{-\lambda\tau}(\pi+(1-\pi)(1-\theta)e^{-\eta\tau}) & i=1.\end{cases}$$

For $i=0$, the bank always pays the service fee, so we have $q(0,\theta,\tau)=1$. For $i=1$, the bank pays the service fee only if the borrower has a loan outstanding at the bigtech lender at date τ, which occurs with probability $q(1,\theta,\tau)=e^{-\lambda\tau}(\pi+(1-\pi)(1-\theta)e^{-\eta\tau})$.

The second term is the bigtech lender's expected gain from financing a good project, which is the product of the bigtech lender's effective share on the loan to a good project $(1-e^{-(\rho+\lambda)\tau}(1-\alpha))$, the probability of financing a good project π, and the expected gain of financing a good project $(\frac{y+\lambda}{\rho+\lambda}-1)$. Note that under the collaboration contract, the bank will finance $(1-\alpha)$ of the loan at date τ, conditional on the loan being non-defaulted and non-matured at that date. So, the bank's effective share on a good project from the collaboration is $e^{-(\rho+\lambda)\tau}(1-\alpha)$, which discounts the bank's loan share $(1-\alpha)$ according to time preference and project maturity. As such, the bigtech lender's effective share on the project can be computed as 1 minus the bank's effective share.

Similarly, the third term in the payoff function is the bigtech lender's expected gain from financing a bad project, which is the product of the bigtech lender's effective share on the loan to a bad project $(1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))$, the probability of financing a bad project $(1-\pi)(1-\theta)$, and the expected gain of financing a bad project $(\frac{y+\lambda}{\rho+\lambda+\eta}-1)$. All three variables here are slightly different from those in the second term. First, the bigtech lender's effective share on a bad project is higher than on a good project. This is caused by the Poisson arrival of loan default for a bad project before date τ. Second, the probability of financing a bad project $(1-\pi)(1-\theta)$ now depends on the screening intensity. Intuitively, screening at a higher intensity can reduce the amount of financing going to bad projects, which helps reduce expected loan losses. Third, the expected gain of financing a bad project is $(\frac{y+\lambda}{\rho+\lambda+\eta}-1)$, which is negative under assumption 1.

The bigtech lender chooses θ, the screening intensity, to maximize its payoff function. We can write the optimality condition in the case of an interior solution[9] as

$$C'(\theta)-e^{-\rho\tau}\frac{\partial q}{\partial\theta}P=(1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta}). \quad (1)$$

The optimality condition above equalizes the marginal cost and the marginal benefit of screening. The marginal benefit of screening is proportional to $(1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))$, the bigtech lender's effective share on the loan to a bad project, because screening helps the bigtech lender

[9]Hereinafter, we focus on parameter values that give rise to an interior solution to the optimality condition. A sufficient condition to ensure an interior solution is to assume $C'(1)>(1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$.

reduce its expected losses from financing bad projects. The marginal cost contains both the direct screening cost and the indirect effect of screening on the bigtech lender's service fee. Given the definition of q, we have $\frac{\partial q}{\partial \theta} = 0$ if $i = 0$ and $\frac{\partial q}{\partial \theta} < 0$ if $i = 1$. Intuitively, a contract with $i = 1$ makes the service fee contingent on whether the borrower has a loan outstanding at date τ, which creates a disincentive for the bigtech lender to screen. Such a disincentive is absent for a contract with $i = 0$, in which case the bigtech lender receives the service fee unconditionally.

Denote the optimal screening intensity that solves equation (1) as $\theta^* = \theta^*(\tau, P, i, \alpha)$. Now we solve for the optimal external collaboration contract at date 0. The bigtech lender designs the collaboration contract to maximize its expected payoff, which is, $W_{ec} = \max_{\tau,P,i,\alpha} W(\tau, P, i, \alpha)$, where

$$W(\tau,P,\alpha,i) = \begin{array}{l}(-C(\theta^*) + e^{-\rho\tau} q(i,\theta^*,\tau)P) + (1 - e^{-(\rho+\lambda)\tau}(1-\alpha))\pi(\frac{y+\lambda}{\rho+\lambda} - 1) \\ \quad + (1 - e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\pi)(1-\theta^*)(\frac{y+\lambda}{\rho+\lambda+\eta} - 1),\end{array}$$

subject to the bank's break-even condition for participating in the collaboration:

$$P = \begin{cases} e^{-\lambda\tau}[(1-\alpha)\pi(\frac{y+\lambda+b}{\rho+\lambda} - 1) + e^{-\eta\tau}(1-\alpha)(1-\pi)(1-\theta^*)(\frac{y+\lambda+b}{\rho+\lambda+\eta} - 1)] & i = 0 \\ \frac{[(1-\alpha)\pi(\frac{y+\lambda+b}{\rho+\lambda} - 1) + e^{-\eta\tau}(1-\alpha)(1-\pi)(1-\theta^*)(\frac{y+\lambda+b}{\rho+\lambda+\eta} - 1)]}{\pi + (1-\pi)(1-\theta^*)e^{-\eta\tau}} & i = 1 \end{cases}$$

Let $P(\tau, \alpha, i)$ denote the service fee that satisfies the break-even condition above. Plugging this expression for the service fee in the maximization problem, we can rewrite the problem as the bigtech lender choosing (τ, α, i) to maximize its expected payoff.

3.2. Discussions of External Collaboration Models

Before solving the optimal collaboration contract, we discuss mapping different collaboration contracts (τ, P, i, α) in our theoretical framework to different real-world contracts. Specifically, we discuss three different collaboration models and relate them to three types of collaboration contracts within the contract space.

Joint Lending (Funding-Level Collaboration). One common collaboration model between multiple lenders uses syndicated loans to fund one borrower. Unlike the traditional syndicated loan market, where mainly high-quality large firms obtain debt financing from financial institutions, our model focuses on the collaboration between bigtech firms and bank lenders in financing consumers with low credit scores and/or small firms with limited credit access. Therefore, we use the word joint lending to describe this type of bigtech-bank collaboration model.

In our model, a joint lending contract refers to one where $\tau = 0$ and $\alpha > 0$. Under this contract, once the borrower passes the bigtech lender's screening at date 0, the bigtech lender contributes α fraction of the loan funding and receives α fraction of the loan's interest income, while the rest of the loan funding is paid and the rest of the interest income is earned by the bank. The bigtech lender has skin in the game to screen because it holds a fraction of the loan and thus bears loan losses proportional to its initial loan contribution.

Because there is no recovery value from a failed project, the seniority ranking of the two debt claims becomes irrelevant. Suppose we assume the recovery value is positive. In that case, the contract space for external collaboration can be extended to introduce a new variable that governs the division of the recovery value of a defaulted loan. In that case, we conjecture (without offering a formal proof) that the optimal contract would give all the recovery value to the bank, effectively making the bigtech lender's claim junior to the bank's to give the bigtech lender more skin in the game to screen the borrower. Therefore, when the recovery value is positive, the optimal joint-lending contract features the bank holding the senior tranche and the bigtech lender holding the junior tranche of the loan.

In practice, examples of joint lending arrangements include Ant Financial, which partners with numerous regional banks. The banks provide funding for loans, while Ant contributes its advanced risk assessment technology. Under this arrangement, Ant's MYbank typically retains 2-30% of each loan, providing sufficient skin in the game to ensure proper screening incentives while allowing bank partners to deploy their lower-cost capital for the majority of the funding. Similarly, Amazon's partnership with Goldman Sachs for merchant lending represents a joint lending arrangement where Amazon leverages its seller data for screening while Goldman contributes funding expertise and balance sheet capacity. Our model explains why these arrangements typically involve bigtech firms maintaining a meaningful but limited stake in the resulting loans.

Loan Selling (Asset-Level Collaboration). Another collaboration model is for the bigtech lender to sell the loan to the bank at a given price. Such a loan-selling collaboration can be implemented via securitization. In our framework, a loan selling contract is one where $\tau > 0$, $\alpha = 0$ and $i = 1$. Under this contract, when a borrower passes the initial bigtech screening, the bigtech lender lends to the borrower for a certain period of time τ. If the loan is still outstanding at date τ, the bigtech lender sells the entire loan to the bank at the price $1 + P$. Here, we implicitly assume the bank makes a lump-sum payment to purchase the loan. Therefore, the loan-selling collaboration maps into a collaboration contract with $i = 1$, in which the service fee is paid only when the loan is outstanding at date τ. In this case, a bigtech lender has skin in the game to screen because it holds the entire loan and bears the loan risk for the initial financing period. Nevertheless, making the service fee contingent on loan status creates a disincentive for screening, because screening at a higher intensity would reduce the bigtech lender's probability of earning the service fee at date τ.

In practice, loan selling arrangements have become increasingly common in the fintech ecosystem. For example, Upstart, a company using AI for credit decisions, facilitates loan origination through partner banks like Cross River Bank and Customers Bank, who originate and sell the loans to institutional investors. Similarly, LendingClub's marketplace model involves originating loans and selling them to banking partners, allowing banks to acquire loan assets while benefiting from LendingClub's technology-driven underwriting. Goldman Sachs has also utilized this model with Apple Card, where Apple provides the technology interface and customer acquisition while Goldman's Marcus business line holds the loan assets. These arrangements allow technology companies to focus on customer acquisition and screening while transferring balance sheet risk to regulated banks with lower funding costs.

Data Selling (Data-Level Collaboration). The bigtech lender can also sell the borrower's data, including her screening signals and credit history, to the bank. By screening and lending to ex-ante risky borrowers, bigtech lenders create valuable data, which can be valuable to banks in making lending decisions. Bigtechs can implement this kind of data selling by providing alternative credit scoring services to banks.

In our model, such a data selling contract is one where $\tau > 0$, $\alpha = 0$ and $i = 0$. Under this contract, when a borrower passes the initial bigtech screening, the bigtech lender lends to the borrower for a certain period of time τ. Then at date τ, the bigtech lender sells the access to the borrower's screening result and credit history data to one bank for a service fee P.[10] If the borrower has a good credit history, which means that she passes the bigtech screening and has not defaulted on her lending, the bank can offer a new loan to the borrower. We assume that for unmodelled reasons such as payment convenience and relationship building, the borrower weakly prefers borrowing from a bank over borrowing from a bigtech lender at the same interest rate. Then the bank can set the interest rate at $R = y$, and the borrower would refinance her entire bigtech loan with the new bank loan, corresponding to $\alpha = 0$ in the contract. In this case, a bigtech lender has skin in the game to screen because it holds the entire loan and bears the loan risk for the initial financing period.

In practice, data-level collaboration - also called alternative credit scoring - has become increasingly prevalent. These services are typically offered via Application Programming Interfaces. For example, providers in the United States, the United Kingdom, and Brazil include Plaid, Finicity (part of MasterCard), Genify, Quovo, True Layer, and Instantor. After purchasing borrowers' credit history data and/or alternative credit scores from bigtech lenders, banks can make loans to new customers with sufficient time-series data and/or high alternative credit scores. Plaid, which Visa nearly acquired for \$5.3 billion in 2020 before regulatory concerns ended the deal, provides API-based access to customer financial data that banks can use for lending decisions. Similarly, Experian Boost allows consumers to add alternative data like utility payments to their credit files, effectively selling this data to traditional lenders. In China, Tencent's WeBank and Ant Group have developed sophisticated data-sharing arrangements with traditional banks, providing credit

[10] We can alternatively assume that the bank needs to pay an upfront service fee at date 0 for getting access to these data.

scoring services based on their vast alternative datasets. The European Union's PSD2 regulatory framework has institutionalized this approach by requiring banks to share customer data (with permission) through standardized APIs, creating a regulatory infrastructure for the exact type of data-level collaboration our model identifies as optimal.

3.3. Data Selling as the Optimal Contract

Having established the three possible collaboration structures, we now demonstrate why data-level collaboration emerges as the optimal arrangement through two key economic mechanisms. First, we show that unconditional service fees dominate conditional fees (Lemma 1). Second, we prove that longer bigtech lending periods with smaller retained stakes dominate shorter periods with larger stakes (Lemma 2). Together, these properties lead to data selling as the uniquely optimal contract. A formal proof of the two lemmas is in the Appendix.

The first key insight concerns how service fee payments should be structured to maximize screening incentives:

Lemma 1. Let $P(\tau,\alpha,i)$ be the service fee that satisfies the bank's break-even condition. We have $W(\tau,P(\tau,\alpha,0),0,\alpha) > W(\tau,P(\tau,\alpha,1),1,\alpha)$.

Lemma 1 compares two collaboration contracts with the same (τ,α) and that satisfy the bank's break-even condition, but have different conditions for the service fee payment. Compared to the contract that pays the service fee conditional on the borrower having an outstanding loan, the contract that pays the service fee unconditionally yields a higher payoff for the bigtech lender. Collaborating with the bank in the lending market reduces the bigtech lender's effective loan share, lowering the screening intensity. As a result, when two contracts offer the same amount of benefits from the bank's low-cost funding, the contract that induces a higher screening intensity generates a higher expected payoff for the bigtech lender.

The intuition is that when the service fee is conditional on the loan being outstanding $(i = 1)$, the bigtech lender faces conflicting incentives. Higher screening intensity reduces expected loan losses by rejecting more bad borrowers, but simultaneously reduces the probability of receiving the service fee at date τ. This screening disincentive is eliminated when fees are paid unconditionally ($i = 0$), leading to higher screening intensity and greater expected profits. This can be seen from the optimality condition for screening intensity in equation (1).

Following Lemma 1, we see that the bigtech lender prefers data-selling contracts over loan-selling contracts when collaborating with banks. The comparison of these two types of contracts is just a special case of Lemma 1, setting $\alpha = 0$. The second key insight concerns the optimal balance between the lending period (τ) and the bigtech's loan share (α):

Lemma 2. For any (α, τ) and (α', τ') such that $\tau' > \tau$ and $(1-\alpha)e^{-(\rho+\lambda+\eta)\tau} = (1-\alpha')e^{-(\rho+\lambda+\eta)\tau'}$, we have $W(\tau', P(\tau', \alpha', 0), 0, \alpha') > W(\tau, P(\tau, \alpha, 0), 0, \alpha)$.

Lemma 2 compares two contracts with the same $i = 0$, satisfying the bank's break-even condition but having different values of (τ, α). Specifically, we set $(1-\alpha)e^{-(\rho+\lambda+\eta)\tau} = (1-\alpha')e^{-(\rho+\lambda+\eta)\tau'}$, so that the optimal screening intensity under the two contracts is the same. Lemma 2 states that the bigtech lender prefers the contract with a larger τ.

This result reflects the time-varying nature of default risk. Because borrower credit quality improves over time (conditional on no defaults), screening incentives are most efficiently provided by having the bigtech lender bear full risk during the initial high-risk period rather than partial risk over the entire loan duration. This timing advantage explains why data selling ultimately dominates joint lending arrangements.

In an external collaboration, the bigtech lender has skin in the game to screen either because the bigtech lender bears the entire loan losses for the first τ period of time or because it holds α fraction of the loan after the collaboration is carried out. Conditional on no defaults, the borrower's posterior credit quality improves over time. Therefore, keeping the amount of bank funding fixed, it is more efficient to incentivize screening by letting the bigtech lender bear the loan losses for the initial time period during which the default risk is high. In other words, for two contracts that induce the same screening intensity, the contract with a higher τ and a lower α is preferred, because it allows the bigtech lender to capture more benefits from the bank's low-cost funding.

According to Lemma 2, keeping the value of $(1-\alpha)e^{-(\rho+\lambda+\eta)\tau}$ fixed, the higher the τ, the larger the bigtech lender's expected payoff. Therefore, we can improve the outcome by pushing τ higher and higher, until α is squeezed to the left corner ($\alpha = 0$). The final contract features $i = 0$ and $\alpha = 0$, which is interpreted as a data-selling contract, and is the optimal external collaboration contract.

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. These two properties lead directly to our central result. Proposition 1 summarizes the main property of the optimal data-selling contract.

Proposition 1. *The optimal external collaboration contract is a data-selling contract, which sets both α and i to zero. Assuming that $y < \rho + \eta - 2b$ and $k < (1-\pi)(1-\frac{y+\lambda}{\rho+\lambda+\eta})$, we have*

1. The optimal data-selling date τ^* is unique and strictly positive.

2. Under the optimal contract, the bigtech lender can finance a borrower only if the borrower's initial credit quality is higher than a threshold value π^*_{ec}. Further, $\pi^*_{ec} < \bar{\pi}$ holds if the marginal screening cost k is smaller than a threshold value $\bar{k}$.

3. Let θ^*_{ec} be the screening intensity under the optimal contract. Both θ^*_{ec} and τ^* are decreasing in the marginal screening cost k and the bank's funding cost advantage b.

Proof. The proof of Proposition 1 and the expressions for $(\tau^*, \pi^*_{ec}, \bar{k}, \theta^*_{ec})$ are given in the Appendix. □

The optimal external collaboration is a data-selling contract featuring three key characteristics. First, bigtech lenders bear the entire loan risk during the initial period ($\alpha = 0$), creating maximum screening incentives when they matter most. Second, payments from banks to bigtechs are unconditional ($i = 0$), eliminating counterproductive screening disincentives. Third, the optimal length of bigtech lending (τ^*) balances two forces: longer periods improve screening incentives but delay the utilization of banks' funding advantage. The data-selling date τ^* measures the length of the borrower's credit history at the bigtech lender, which can be interpreted more broadly as the quality of the bigtech lender's alternative credit scoring data sold to the bank.

Proposition 1 states that the optimal τ^* is unique and strictly positive. This result suggests that bigtech lenders always prefer collaborating with banks through data sales over lending to borrowers without any collaboration. Under the no collaboration case, the bigtech lender can only make loans from its own funding source, equivalent to a collaboration contract with $\alpha = \tau = 0$. While the cheap bank funding is completely untapped, the bigtech lender's effective share on the loan to a bad project is at its highest possible value, 1, which would lead to the highest possible level of screening intensity among all possible external collaboration contracts. We show that such a no-collaboration case is strictly dominated if the bank's funding advantage is strictly positive, $b > 0$.

Next, Proposition 1 defines π_{ec}^* as the bigtech lender's lending standard: the lowest-possible initial credit quality at which the bigtech lender is willing to extend credits, under the optimal data selling contract. Among all external collaboration contracts, the optimal data selling contract leads to the lowest bigtech lending standard or, equivalently, the highest level of credit accessibility for high-risk borrowers. This is true because the bigtech lender's expected payoff is the highest under the optimal data-selling contract for a borrower at any given credit quality π.

Further, we show that as long as the marginal screening cost is not too high, the bigtech lending standard under the optimal data selling contract will be lower than the bank's lending standard, that is, $\pi_{ec}^* < \bar{\pi}$. As such, introducing bigtech-bank collaboration can make credit more accessible to borrowers who face credit rationing in the traditional banking system. The credit rationing problem, which is caused by banks' lack of credit-relevant data for screening, is alleviated by bigtech lenders' abilities to generate screening signals, to accumulate credit history data through lending, and to sell these data to banks for subsequent bank screening and lending.

Lastly, we discuss how the date-selling date τ^* and the screening intensity θ_{ec}^* depend on several key model parameters. The date-selling date τ^* in our model can be measured by the length of the borrower's bigtech credit history at the time when the bank offers a loan to the borrower. Proposition 1 states that θ_{ec}^* and τ^* are decreasing in the marginal screening cost k and the bank's funding cost advantage b.

Intuitively, requiring the bigtech lender to bear the loan risk for a longer period induces a higher screening intensity and thus increases the bigtech payoff. Nevertheless, it is costly because selling the credit history data at a later date reduces the bank's time-weighted share of the loan, lowering the expected benefit from the bank's low-cost funding. An increase in k makes screening more costly, which reduces the benefit of inducing a higher screening intensity. In response, the bigtech lender would choose a smaller τ^*, which induces a smaller θ_{ec}^* to economize on the screening cost. Meanwhile, an increase in b increases the benefit of relying on the bank's low-cost funding, so the bigtech lender would choose a smaller τ^*, which induces a smaller θ_{ec}^* but leads to a higher expected gain from utilizing bank funding.

In sum, bigtech lenders prefer using data-selling contracts to collaborate with banks. Under such a 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 borrowers with good credit records. Among all external collaboration contracts, the data-selling contract is the most efficient one in inducing the bigtech screening intensity without unduly forgoing the benefit of the bank's low-cost funding. Under certain conditions, bigtech firms' data-as-a-service to banks allows banks to onboard borrowers who are unable to obtain bank loans directly.

3.4. Internal Collaboration

Now we analyze the case where the bigtech lender's screening intensity is contractable between the bigtech lender and the bank. Without the contractual friction, the bank will directly fund the entire loan to the borrower from date 0 while the bigtech lender screens at a level to maximize the total gain from lending. This case can be interpreted as the *internal collaboration* model where the bigtech lender merges with a bank, combining the screening skill and low-cost funding under the same corporate umbrella.

Internal collaborations have emerged through three distinct mechanisms in practice. First, direct acquisitions integrate screening capabilities with banking infrastructure–Goldman Sachs' $2.2 billion acquisition of GreenSky in 2021 and JPMorgan's $300 million purchase of WePay in 2017 exemplify this approach. Second, banks develop in-house technology platforms, as seen with Goldman's Marcus platform, which built consumer lending algorithms internally and deployed $4 billion in consumer loans within its first year. Third, hybrid structures using corporate venture capital create partial integration–Santander InnoVentures' $135 million investment in Kabbage aligned incentives while preserving operational independence until Kabbage was ultimately acquired by American Express in 2020.

Unlike external collaboration, internal collaboration eliminates the need to incentivize screening through contract design. The integrated entity can directly implement the screening intensity that maximizes joint profits, fully leveraging both the screening technology and funding advantages. The decision maker only needs to choose the screening intensity to maximize the expected payoff from screening the representative borrower:

$$W_{ic} = \max_{\theta} -C(\theta) + \pi(\frac{y+\lambda+b}{\rho+\lambda} - 1) + (1-\pi)(1-\theta)(\frac{y+\lambda+b}{\rho+\lambda+\eta} - 1).$$

The payoff function is the sum of the screening cost and the expected gain from lending to a good project, subtracted by the expected loss from lending a bad project. The bank's funding cost advantage term b is present as long as the borrower's loan is outstanding. The first-order condition for the optimal screening intensity is

$$C'(\theta) = (1-\pi)(1 - \frac{y+\lambda+b}{\rho+\lambda+\eta}).$$

Let θ^*_{ic} and π^*_{ic} be the optimal screening intensity, which satisfies the above optimality condition, and the lending standard under the optimal screening intensity, which is the value of π that makes $W_{ic} = 0$, respectively. For comparison, we also define θ^*_{nc} and π^*_{nc} as the optimal screening intensity and the lending standard under the case where the bigtech lender finances the borrower without collaborating with a bank. The no collaboration case is just a special case of the internal collaboration case setting $b = 0$, and therefore we have $\theta^*_{nc} = \theta^*_{ic}(b = 0)$ and $\pi^*_{nc} = \pi^*_{ic}(b = 0)$.

Comparing this condition with equation (1) for external collaboration reveals the key difference: the internal collaboration eliminates the term $e^{-\rho\tau}\frac{\partial q}{\partial \theta}P$, which represents the screening disincentive in external arrangements. This elimination of contractual friction suggests internal collaboration might dominate, yet our welfare analysis below will show this intuition is incomplete.

3.5. *Welfare Implications*

To evaluate the social welfare implications of different collaborations, we first define the socially optimal screening level and then compare private and social incentives across collaboration models. A key insight of our analysis is that private screening incentives systematically diverge from social optima, with the direction and magnitude of this divergence depending critically on two parameters: the bank's funding advantage and the borrower's private benefit from investment.

First, we define the socially optimal screening level in the economy. Specifically, we assume there is a social planner who can dictate the bigtech lender's choice of the screening intensity to maximize the total surplus from the investment project:

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

In the surplus function, the bank's funding cost advantage b is not included, reflecting the implicit assumption that the bank's funding-cost advantage does not add real value. In reality, the funding-cost advantage comes mainly from banks' access to deposit insurance and central bank liquidity provision, which we leave unmodeled.

However, the borrower's control rent c, which is not considered in the bigtech lender's screening decision in the private economy studied in Section 3.1 and 3.4, enters into the surplus calculation. As will be discussed more in Section 4, c captures the borrower's gain from the investment project. When the bigtech lender screens the borrower, it only considers the effect of screening on its expected loan losses, without considering the effect on the borrower's gain from the project.

The optimal screening intensity in the social planner problem θ_{sp} solves

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

Then, we analyze the welfare implications of the model by comparing the lending standard π_j^*, the screening intensity θ_j^*, and the welfare level $S(\theta_j)$ under three different collaboration models: 1. no collaboration ($j = nc$), 2. the optimal external collaboration (data-selling) contract ($j = ec$), and 3. the optimal internal collaboration contract ($j = ic$). The next proposition summarizes the comparison.

Proposition 2. *1. For the lending standard, we have $\pi_{ic}^* < \pi_{ec}^* < \pi_{nc}^*$. 2. For the screening intensity, we have $\theta_{ec}^* < \theta_{ic}^* < \theta_{nc}^*$. Moreover, $\theta_{ic}^* > \theta_{sp}$ if and only if $c > b$. 3. There exist two threshold values $(\underline{c}, \bar{c})$, where $\underline{c} < b < \bar{c}$, such that the welfare level $S(\theta_j)$ is the highest at $j = nc$ if $c \in [0, \underline{c})$, at $j = ic$ if $c \in [\underline{c}, \bar{c})$, and at $j = ec$ if $c \geq \bar{c}$. 4. There exist two threshold values $(\underline{b}, \bar{b})$, where $\underline{b} < c < \bar{b}$, such that the welfare level $S(\theta_j)$ is the highest at $j = ec$ if $b \in [0, \underline{b})$, at $j = ic$ if $b \in [\underline{b}, \bar{b})$, and at $j = nc$ if $b \geq \bar{b}$.*

Proof. The proof of Proposition 2 and the expressions for $(\underline{b}, \bar{b}, \underline{c}, \bar{c})$ are given in the Appendix. □

Proposition 2 states that the economy with internal collaboration between bigtech lenders and banks has the lowest lending standard and thus the highest level of credit accessibility. Intuitively, the ability to access bank funding determines the bigtech lender's expected payoff and thus its ability to provide credit to low-quality borrowers. Among the three cases, the internal collaboration between a bigtech firm and a bank fully utilizes the bank's access to low-cost funding. The optimal external collaboration contract partially utilizes bank funding because the contract requires the bigtech lender to hold the loan for a certain period to provide a proper screening incentive. In the case of no collaboration, bigtech lenders cannot access cheap bank funding.

Consequently, allowing internal collaboration between banks and bigtech lenders can generate the highest lending profit, resulting in the lowest lending standard and the highest level of credit accessibility for low-quality borrowers. Because internal collaboration generates the highest lending profit, if left unregulated, bigtech lenders have incentives to merge with banks to better combine their advanced screening skills with banks' cheap funding. However, as stated in Proposition 2, the internal collaboration model, which the bigtech lender prefers, may not lead to the highest welfare level among the three collaboration models.

Before discussing the welfare level, we compare the screening intensity under the three collaboration models. Proposition 2 states that the economy with external collaboration has the lowest screening intensity, and the economy without collaboration has the highest. Compared with the internal collaboration case, the economy under the optimal external collaboration contract induces a lower screening intensity. This is because, in an external collaboration, inducing a higher screening intensity requires the bigtech lender to hold the loan for a longer period of time, which is costly due to the bank's access to cheaper funding. Compared with the internal collaboration case, the no collaboration case yields a higher screening intensity. Intuitively, allowing the bigtech lender to tap cheap bank funding via internal collaboration reduces the loss from a defaulted loan, lowering the bigtech lender's incentive to screen.

Now, we compare the welfare level, measured by the total surplus from the investment project, using the three different collaboration models. We show that the welfare comparison depends on the bank's funding-cost advantage (b) and the borrower's gain from the project (the control rent c).

Figure 1 compares screening intensity and welfare level under different values for the borrower's gain from the investment, which is the control rent c in the model. The green dashed line, the blue dashed line, and the red dashed line are the screening intensities in the no collaboration, internal collaboration, and external collaboration cases, respectively. Figure 1 shows the case with the highest welfare level by its corresponding line changing from dashed to solid.

When $c < \bar{c} = 0.005$, the welfare level is the highest under no collaboration. Intuitively, because the bigtech lender does not internalize the borrower's gain from the investment when making

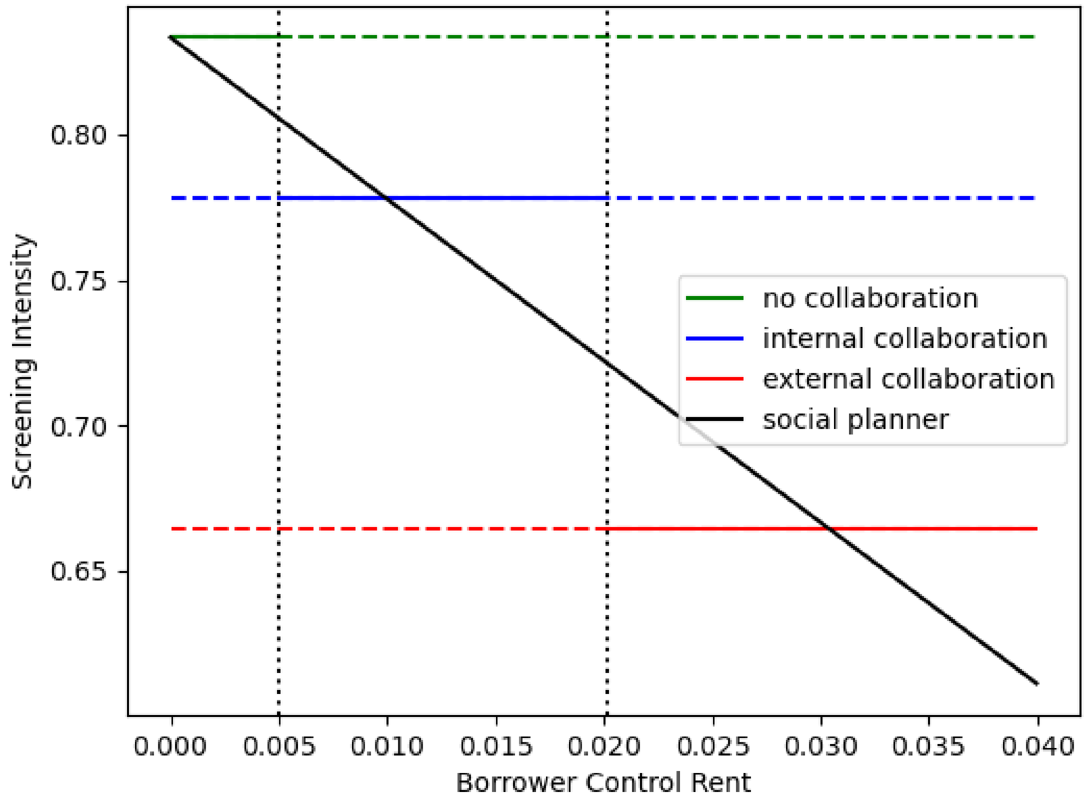

Figure 1 compares the screening intensity in the economy with no collaboration (the green flat line on the top), the economy with internal collaboration (the blue flat line in the middle), the economy with external collaboration (the red flat line on the bottom), and the social planner problem (the black downward-sloping line) for different values of c. Among the three collaboration models, the welfare level is the highest under no collaboration if $c < \underline{c} = 0.005$, under external collaboration if $c > \bar{c} = 0.02$, and under internal collaboration if $c \in (\underline{c}, \bar{c})$. The parameter values are $\rho = 0.05, y = 0.1, k = 0.2, \pi = 0.5, \lambda = 0.2, \eta = 0.2, b = 0.01$.

the screening decision, the no collaboration case features over-investment in the screening technology, that is, $\theta_{nc} > \theta_{sp}$, When the borrower's gain from investment is relatively small, the over-investment problem is relatively small, and the screening intensity in the no collaboration case is close to the socially optimal level. In contrast, both the internal and external collaboration cases feature significant under-investment in the screening technology because using cheap bank funding reduces the bigtech lender's expected loss from a defaulted loan and thus its ex-ante screening incentive.

Because $\underline{c} < b$, we can show that the no collaboration case is strictly dominated in terms of welfare under the assumption that $c > b$. In other words, when the borrower's flow gain from investment exceeds the bank's flow funding-cost advantage, an assumption we think is reasonable to make in reality, introducing bigtech-bank collaboration can increase the total welfare in the economy. Whether internal or external collaborations can yield a higher welfare level further depends on the value of c and b in the economy.

The internal collaboration case yields the highest welfare level when c is in $[0.005, 0.020]$. Intuitively, as the borrower's gain from investment increases, the over-investment of screening in the no collaboration case becomes more severe. Allowing internal collaboration can alleviate the over-investment problem, because using cheap bank funding reduces the ex-post loan losses and thus the bigtech lender's ex-ante screening incentive.

Meanwhile, in this region, the external collaboration case causes significant under-investment in the screening technology due to the agency problem in screening, which is also dominated by the internal collaboration case in terms of welfare. Specifically, the optimal external collaboration contract needs to incentivize screening by having the bigtech lender use its funding to finance the loan for a certain period. Therefore, increasing the screening intensity now comes with an additional cost of foregoing cheap bank funding, which makes the screening intensity in the external collaboration case the lowest among the three cases.

However, when c is larger than 0.020, the external collaboration case yields the highest welfare level. When the borrower's gain from investment is very high, the bigtech lender's over-investment in screening technology is very large. In this case, introducing external collaboration can generate a higher welfare level by alleviating the over-investment problem. Collaborating with a bank through external contracting allows the bigtech lender to tap cheap bank funding, which reduces its expected loan loss and thus its screening incentive. Meanwhile, it also puts in place the contractual friction that elevates the marginal cost of screening, further reducing the intensity of the screening.

Figure 2 compares screening intensity and welfare level under different values for the bank's funding-cost advantage b. In Figure 2, we only present parameter values that satisfy $c > b$, so the

$b > \bar{b}$ case does not arise in the figure.

Under the assumption $c > b$, we have shown that introducing bigtech-bank collaborations can improve welfare in the economy. When $b < \underline{b} = 0.016$, the screening intensity in the external collaboration case is closer to the socially optimal level. Intuitively, when the bank's funding cost advantage is relatively small, the internal collaboration case leads to a relatively high level of over-investment in screening. The contractual friction in the external collaboration case can alleviate the over-investment and improve welfare.

However, when $b \in [\underline{b}, c]$, the contractual friction in the external collaboration case pushes down the screening intensity so much that the economy features severe under-investment in the screening technology. In that case, the internal collaboration yields a higher welfare level, although the screening intensity is still higher than the socially optimal level.

In sum, we find that under the reasonable assumption that the borrower's gain from investment is higher than the bank's funding advantage, introducing bigtech-bank collaborations can increase the bigtech lender's expected payoff, the credit accessibility of low-quality borrowers, and the welfare level in the economy.

We further compare the optimal internal collaboration in which bigtech lenders and banks merge into one entity, and the optimal external collaboration in which bigtech lenders sell borrowers' credit history data to banks. We find that the internal collaboration case always generates the highest credit accessibility for low-quality borrowers because cheap bank funding is the entire funding source. However, the external collaboration case may generate a higher level of welfare. In short, the contractual friction in external collaboration reduces the bigtech lender's excessive screening incentive, possibly leading to a screening intensity closer to the socially optimal level.

Our result suggests that encouraging external collaborations between bigtech lenders and banks through data sales, e.g., imposing stricter regulations on bigtech-bank conglomerates and/or building proper infrastructure for data transactions, may be desirable for welfare considerations. Specifically, compared to internal collaboration, external collaboration generates a higher welfare level when the borrower's gain from investment is high and/or when the bank's funding advantage is

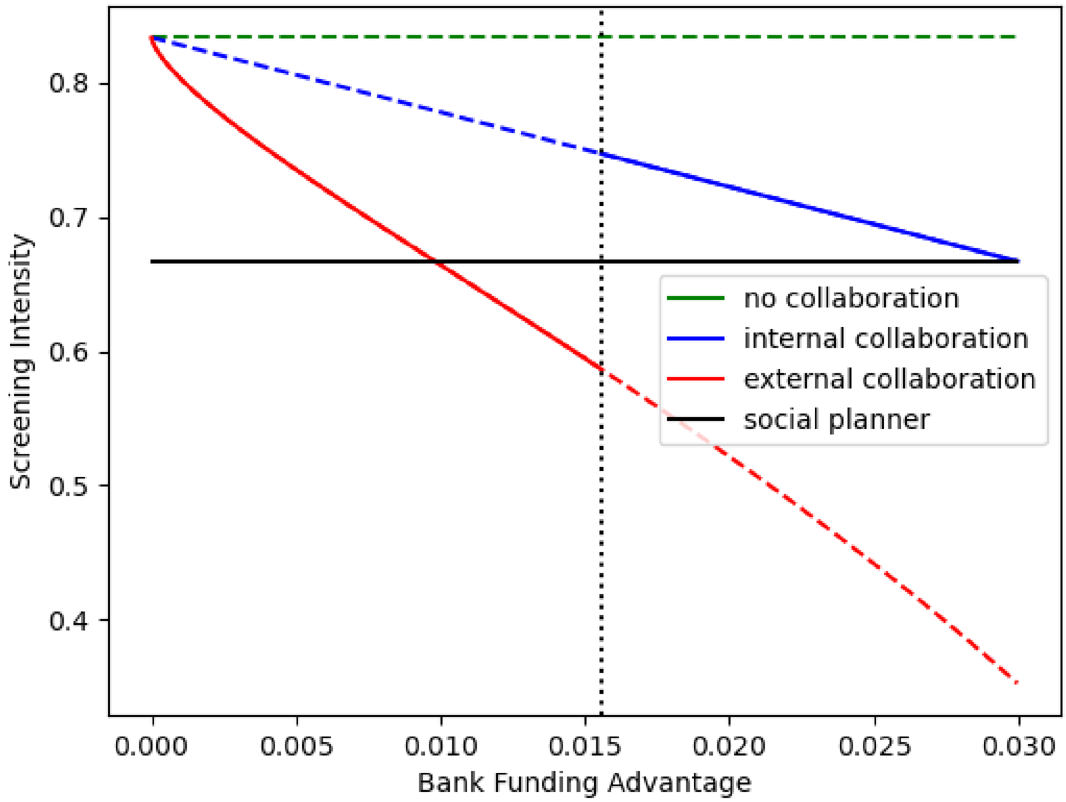

Figure 2 compares the screening intensity in the economy with no collaboration (the green dashed flat line on the top), the economy with internal collaboration (the blue solid-dashed downward-sloping line), the economy with external collaboration (the red dashed-solid downward-sloping line), and the social planner problem (the black solid line) for different values of b. Among the three collaboration models, the welfare level is the highest under external collaboration if $b < \bar{b} = 0.016$ and under internal collaboration if $b > \bar{b}$. The parameter values are $\rho = 0.05, y = 0.1, k = 0.2, \pi = 0.5, \lambda = 0.2, \eta = 0.2, c = 0.03$.

weak.

4. Discussions

4.1. Lending Market Competition

In the main model, we assume that there is only one bigtech lender equipped with the advantageous screening technology. We interpret this assumption as the bigtech lender's unique access to the borrower's alternative data. Under this interpretation, the bigtech lender sets the highest-possible interest rate $R = y$ to capture all the pledgeable cashflows from the borrower's investment project, leaving the borrower the non-pledgeable part c.

If the borrower can share its alternative data with multiple bigtech lenders, these bigtech lenders would compete in the lending market. While the specific equilibrium outcome depends on how we model the competition, which we do not explicitly specify, introducing lending market competition would generally reduce the equilibrium lending rate R and increase the screening intensity θ. The decrease in the lending rate is driven by the competition between bigtech lenders, whereas the increase in the screening effort reflects the decrease in the marginal screening benefit now that the bigtech lender captures less cashflows from the investment project. In

However, our main result regarding data selling as the optimal external collaboration contract is robust after introducing lending market competition. Intuitively, data selling is preferred because it incentivizes screening the most efficiently: compared with other external collaboration contracts that provide the same screening incentive, the data selling contract allows the bigtech lender to tap the highest amount of cheap bank funding. When bigtech lenders compete in the lending market, they would also prefer using data-selling contracts to collaborate with banks because it helps them quote lower interest rates and stay competitive in the lending market.

Moreover, the welfare trade-off between internal and external collaborations is also robust after introducing lending market competition. Intuitively, the internal collaboration model, the privately optimal collaboration arrangement, may lead to over-investment in the screening technology because the bigtech lender fails to internalize a borrower's gain from the project. The contractual friction in external collaboration models reduces the over-investment by increasing the private cost of screening, but it may lead to under-investment in screening, especially when the bank's funding advantage is large. These intuitions would hold regardless of how the lending market determines the interest rate. Furthermore, after introducing competition in the lending market, the borrower's effective share from the investment project $c+y-R$ would increase because R is pushed below y. This suggests that the over-investment problem brought by internal collaborations is more severe, making external collaborations more likely to be preferred from a societal perspective.

4.2. *Data Exclusivity*

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 by 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 to 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 the 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 data-selling collaboration contract can require the bigtech lender to receive the service fee after it successfully helps facilitate the bank loans to those borrowers with good credit histories. However, this method ensures that the exclusivity of data sale is captured as the loan selling contract in our model. As shown above, the loan-selling contract is not as good as the data-selling contract, giving the bigtech lender a disincentive for screening and making the skin in the game for screening more expensive.

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, such as tokenization. This method 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. 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.[11] By tokenizing data, firms can exclusively transfer it to another entity while maintaining its security and privacy. This allows for a secure data exchange without exposing sensitive information to unauthorized parties. This is a direct contract-based method to ensure data exclusivity.

4.3. Reverse-Engineering (Database Reconstruction)

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 is concerned about such reverse engineering depends on the source of the bigtech lender's screening advantage. Suppose the bigtech lender's screening advantage comes largely from collecting or possessing alternative data. In that case, it is very hard for the bank to distill useful information from analyzing borrowers' credit history data and standard credit-risk-relevant data. In this case, allowing banks to access the borrowers' credit history data would not raise many concerns about reverse engineering, and thus should be preferred due to its effectiveness in inducing screening. However, in practice, if a bigtech lender's screening advantage comes largely from applying advanced machine learning techniques on standard financial data, it should be more concerned about reverse engineering so they may instead choose to collaborate with banks in the form of joint lending or loan selling, even if data sales are more efficient in incentivizing screening and reducing lending risks.

5. Conclusion

In this paper, we present a model where bigtech lenders and banks collaborate in the lending market in different forms. We show that encouraging bigtech-bank collaborations can improve borrowers' access to credit and social welfare. Among all external collaborations, we find that

[11] 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 needs to only take in "blinded" tokens for analytics without any human directly viewing the data.

data selling is 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 financial regulation on the lending ecosystem; instead, we expect to find different types of economies reacting differently.

This optimal structure yields several important implications for market participants and regulators. For bigtech lenders, data-level collaboration provides a pathway to monetize their screening technology and alternative data advantages without requiring permanent balance sheet allocation. For banks, these arrangements enable access to new customer segments that would otherwise remain outside their risk tolerance. For borrowers, particularly those with limited credit histories, these collaborations expand financial inclusion by creating a bridge between alternative data-based assessment and traditional banking relationships.

Our findings suggest specific regulatory approaches depending on market characteristics. When markets feature significant competition among bigtech lenders and modest bank funding advantages, regulators should facilitate data sharing by standardizing data formats and establishing clear ownership rights. In markets where banks enjoy substantial funding advantages or bigtech firms have significant market power, minimum holding periods before data sales may be necessary to prevent underinvestment in screening. These targeted interventions can enhance financial inclusion while mitigating risks to credit quality. Future research could explore how these collaboration models evolve under different privacy regimes, regulatory frameworks, and technological changes in credit assessment. As financial data governance continues to develop globally, understanding the optimal structure of bigtech-bank partnerships will remain critical for promoting both innovation and stability in the lending ecosystem.

References

Arner, D. W., Castellano, G. G., 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., Piccolo, S., 2015. Multiple bank lending, creditor rights, and information sharing. Review of Finance 19, 519–570.

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

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

Borgogno, O., Colangelo, G., 2020. The data sharing paradox: BigTechs in finance. European Competition Journal 16, 492–511.

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

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

Charoenwong, B., Kowaleski, Z. T., Kwan, A., 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., Kwan, A., 2021. Alternative data, big data, and applications to finance. In: *Fintech with Artificial Intelligence, Big Data, and Blockchain*, Springer, pp. 35–105.

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

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

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

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

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

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

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

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

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

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

Fuster, A., Plosser, M., Schnabl, P., Vickery, J., 2019. The Role of Technology in Mortgage Lending. Review of Financial Studies 32, 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)*, IEEE, pp. 73–81.

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

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

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

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

Liu, L., Lu, G., Xiong, W., 2022. The big tech lending model. No. 30160, NBER Working Paper.

Markatou, E. A., 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*, Springer, pp. 25–43.

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

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

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

Pagano, M., Jappeli, T., 1993. Information Sharing in Credit Markets. The Journal of Finance 48, 1693–1718.

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

Philippon, T., 2015. Has the US finance industry become less efficient? On the theory and measurement of financial intermediation. American Economic Review 105, 1408–1438.

Philippon, T., 2016. The Fintech Opportunity. NBER Working Paper pp. 1–24.

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

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

Stiglitz, B. J. E., Weiss, A., 1981. Credit Rationing in Markets with Imperfect Information. American Economic Review 71, 393–410.

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

Appendix

Proof of Proposition 1

Define $H_f = \pi(\frac{y+\lambda}{\rho+\lambda} - 1) > 0$ and $H_b = \pi(\frac{y+\lambda+b}{\rho+\lambda} - 1) > 0$ as the (unconditional) expected profit from lending to a good project with the fintech lender's funding and with the bank's funding, respectively. Define $L_f = (1-\pi)(\frac{y+\lambda}{\rho+\lambda+\eta} - 1) < 0$ and $L_b = (1-\pi)(\frac{y+\lambda+b}{\rho+\lambda+\eta} - 1) < 0$ as the (unconditional) expected loss from lending to a bad project with the fintech lender's funding and with the bank's funding, respectively.

First, we prove Lemma 1 and Lemma 2. The bigtech lender chooses the optimal collaboration contract to maximize its expected payoff, subject to the bank's participation constraint. We can plug the constraint into the objective function and rewrite it as

$$W(\tau,P,i,\alpha) = \begin{array}{c} -C(\theta^*) + e^{-(\rho+\lambda)\tau}(1-\alpha)H_b + e^{-(\rho+\lambda+\eta)\tau}(1-\alpha)(1-\theta^*)L_b \\ +(1-e^{-(\rho+\lambda)\tau}(1-\alpha))H_f + (1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(1-\theta^*)L_f, \end{array}$$

so P does not enter the objective function. Lemma 1 compares contracts with the same τ,α but different values for i. Note that i has an impact on the above payoff function through its effect on θ^*. So we can use $\theta^* = \theta^*(\tau,i,\alpha)$ to replace i as the variable by letting $w(\tau,\alpha,\theta=\theta^*) = W(\tau,P,i,\alpha)$, where $\theta^* = \theta^*(\tau,i,\alpha)$ is given by the optimality condition

$$C'(\theta^*) - e^{-\rho\tau}\frac{\partial q}{\partial\theta}P = (1-e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(-L_f),$$

where $\frac{\partial q}{\partial\theta} = 0$ if $i=0$ and $\frac{\partial q}{\partial\theta} < 0$ if $i=1$. Therefore, we have $\theta^*(i=1) < \theta^*(i=0)$, so the contract with $i=0$ induces a higher screening intensity.

To compare the bigtech payoff under the two contracts, we take the derivative:

$$\frac{\partial w}{\partial \theta} = -C'(\theta) + (1 - e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(-L_f) + e^{-(\rho+\lambda+\eta)\tau}(1-\alpha)(-L_b),$$

Then we can show that $\frac{dW}{d\theta} > 0$ at $\theta = \theta^*(i=0)$. Moreover, if the screening cost function is increasing and convex, which holds because we assume a quadratic cost function for simplicity, we can show that $\frac{dw}{d\theta} > 0$ holds for all $\theta < \theta^*(i=0)$. Consequently, we have $w(\theta^*(i=0)) > w(\theta^*(i=1))$ for any given value of (α, τ). So for any feasible contract with $i = 1$, we can find a feasible contract with $i = 0$ that yields a higher bigtech payoff. Therefore, we have$W(\tau, P(\tau, \alpha, 0), \alpha, 0) > W(\tau, P(\tau, \alpha, 1), \alpha, 1)$.

For Lemma 2, we can further rewrite the objective function W as

$$W(\tau, P, i, \alpha) = -C(\theta^*) + H_f + (1-\theta^*)L_f + e^{-(\rho+\lambda)\tau}(1-\alpha)\frac{\pi b}{\rho+\lambda} + e^{-(\rho+\lambda+\eta)\tau}(1-\alpha)\frac{(1-\pi)(1-\theta^*)b}{\rho+\lambda+\eta}$$

For a feasible contract with $i = 0$ and (τ, α) such that $\alpha > 0$, we are able to find another feasible contract with $i = 0$, $\alpha' = \alpha - \delta$, where $\delta > 0$, and $\tau' > \tau$ satisfying $e^{-(\rho+\lambda+\eta)\tau}(1-\alpha) = e^{-(\rho+\lambda+\eta)\tau'}(1-\alpha')$. Note that the bigtech screening intensity θ^*, which is given by $C'(\theta^*) = (1 - e^{-(\rho+\lambda+\eta)\tau}(1-\alpha))(-L_f)$, is the same under the contract $(\tau, P(\tau, \alpha, 0), 0, \alpha)$ and the contract $(\tau', P(\tau', \alpha', 0), 0, \alpha')$. Yet, we have

$$W(\tau', P(\tau', \alpha', 0), 0, \alpha') - W(\tau, P(\tau, \alpha, 0), 0, \alpha) = e^{-(\rho+\lambda)\tau}(1-\alpha)\frac{b}{\rho+\lambda}(e^{\eta\tau'} - e^{\eta\tau}) > 0$$

. Therefore, among all contracts that induce the same screening intensity, the bigtech lender prefers contracts with the longest initial lending period τ, because it allows the bigtech lender to utilize the highest amount of cheap bank funding.

Given the two lemmas, the optimal external collaboration contract is a data-selling contract with $\alpha = 0$ and $i = 0$. Now we prove the rest of Proposition 1.

First, note that under the assumption $k < (1-\pi)(1 - \frac{y+\lambda}{\rho+\lambda+\eta})$, the optimal screening effort is always an interior solution, that is $\theta^*_{ec} \in (0,1)$. The screening effort and the optimal data-selling date τ^* solve

$$k\theta = (1 - e^{-(\rho+\lambda+\eta)\tau})(-L_f)$$
$$(1-\pi)(1-\theta) = e^{-(\rho+\lambda+\eta)\tau}\frac{\rho+\lambda+\eta}{bk}L_f L_b - e^{\eta\tau}\pi.$$

The first equation is the optimality condition for choosing the screening effort, and the second equation is the optimality condition for choosing the data-selling date. Defining $m = e^{-(\rho+\lambda+\eta)\tau}$, we can write $\theta_1(m)$ and $\theta_2(m)$ as the solution of the above two equations, respectively. We have $\theta_1'(m) < 0$ and $\theta_2'(m) < 0$. Under the assumption $y < \rho + \eta - 2b$, we have $|\theta_1'(m)| < |\theta_2'(m)|$ for all $m \in [0,1]$. Thus, the equation $\theta_1(m) - \theta_2(m) = 0$ has a unique solution, which means that there is a unique pair of (τ^*, θ^*_{ec}) that solves the two optimality conditions stated above.

Second, we can write the bigtech lender's payoff under the optimal contract as a function of the borrower's observable credit quality π and the marginal screening cost k. Using the envelope theorem, we can show that the bigtech payoff $W(\pi,k)$ is increasing in π and decreasing in k. Define π^*_{ec} as the unique solution to $W(\pi^*_{ec},k) = 0$. Then, π^*_{ec} is increasing in k. Let $\bar{k}$ be the solution to $W(\bar{\pi},\bar{k}) = 0$. We have $\pi^*_{ec} < \bar{\pi}$ if and only if $k < \bar{k}$. Last, we show that both (τ^*, θ^*_{ec}) are decreasing in k and b. The two optimality conditions that uniquely determines (τ^*, θ^*_{ec}) can be written as $\theta = \theta_1(\overset{(-)}{m}, \overset{(-)}{k})$ and $\theta = \theta_2(\overset{(-)}{m}, \overset{(+)}{b}, \overset{(+)}{k})$, where the $+/-$ sign above a variable indicates the sign of the partial derivative with respect to that variable. Given $|\theta_1'(m)| < |\theta_2'(m)|$, we can show that m is increasing in b and k. Then, we can show that both (τ^*, θ^*_{ec}) are decreasing in k and in b.

Proof of Proposition 2

First, we prove that $\pi^*_{ic} < \pi^*_{ec} < \pi^*_{nc}$. π^*_{nc} is the value of π that makes $W_{nc}(\pi) = 0$, where $W_{nc}(\pi) = W_{ic}(\pi, b=0) = W(\pi, b=0)$. Using the envelope theorem, we can show that the bigtech payoff under the optimal external collaboration contract W is increasing in b and π. Then we have $W(\pi,b) > W_{nc}(\pi)$ as long as $b > 0$, which leads to $\pi^*_{ec} < \pi^*_{nc}$. To show that $\pi^*_{ic} < \pi^*_{ec}$, note that we always have $W(\pi) < W_{nc}(\pi)$ for any value of π. Under internal collaboration, the bigtech lender can freely choose the amount of banking funding and the screening effort, while under external collaboration, there is an additional skin-in-the-game constraint (the optimality condition for choosing the screening intensity) that ties the amount of banking funding with the bigtech screening intensity. Therefore, the bigtech lender would achieve a higher payoff under internal collaboration. Then, the lending standard would be lower under internal collaboration, that is, $\pi_{ic}^* < \pi_{ec}^*$.

Then, we prove that $\theta_{ec}^* < \theta_{ic}^* < \theta_{nc}^*$, and that $\theta_{ic}^* > \theta_{sp}$ if and only if $c > b$. First, note that θ_{nc}^* is equal to θ_{ic}^* if $b = 0$, and that θ_{ic}^* is decreasing in b. These two combined get us $\theta_{ic}^* < \theta_{nc}^*$. Compared with internal collaboration, external collaboration introduces the additional skin-in-the-game constraint, which elevates the marginal cost of screening. Therefore, we have $\theta_{ec}^* < \theta_{ic}^*$. Further, given the equations for θ_{ic}^* and θ_{sp}, it is obvious that $\theta_{ic}^* > \theta_{sp}$ if and only if $c > b$.

Lastly, we show that there exist two threshold values $(\underline{c}, \bar{c})$, where $\underline{c} < b < \bar{c}$, such that the welfare level $S(\theta_j)$ is the highest at $j = nc$ if $c \in [0, \underline{c})$, at $j = ic$ if $c \in [\underline{c}, \bar{c})$, and at $j = ec$ if $c \geq \bar{c}$. Note that an increase in c only affects θ_{sp} and have no impact on $\theta_{ec}^*, \theta_{ic}^*, \theta_{nc}^*$.

Then we can show that 1. if $c = 0$, $W(\theta_{nc}^*; c) > W(\theta_{ic}^*; c) > W(\theta_{ec}^*; c)$, 2. if $c \in (0, b)$, $W(\theta_{nc}^*; c)$ is decreasing in c and both $W(\theta_{ic}^*; c)$ and $W(\theta_{ec}^*; c)$ are increasing in c, 3. if $c = b$ we have $W(\theta_{ic}^*; c) > \max\{W(\theta_{nc}^*; c), W(\theta_{ec}^*; c)\}$, 4. if $c \in (b, c_0)$, where c_0 solves $W(\theta_{sp}, c_0) = W(\theta_{ec}^*; c_0)$, $W(\theta_{ec}^*; c)$ is increasing in c and both $W(\theta_{ic}^*; c)$ and $W(\theta_{nc}^*; c)$ are decreasing in c.

As a result, there exists a unique $\underline{c} \in (0, b)$ that solves $W(\theta_{nc}^*; c) = W(\theta_{ic}^*; c)$ and a unique $\bar{c} \in (b, c_0)$ that solves $W(\theta_{ic}^*; c) = W(\theta_{ec}^*; c)$. Moreover, the welfare comparison under different values of c stated in Proposition 2 holds under this unique pair of threshold values $(\underline{c}, \bar{c})$. The welfare comparison under different values of b can be proved similarly.