Regulating Information and Competition: Evidence from Fintech SME Loans*

Yingju Ma
MIT, Luohan Academy

André Sztutman
Carnegie Mellon University

Robert M. Townsend
MIT

June 26, 2025

[Click here for the latest version]

Abstract

How do market power and information availability affect the efficiency of emerging informational-rich lending markets? Leveraging data from a major fintech lending provider for SMEs in China and a natural experiment from a temporary and geographically discontinuous interest rate discount policy, we analyze the presence of information asymmetries, selection, and market power. Our findings reveal a decrease in the average probability of loan repayment following interest rate reductions, indicating that selection is advantageous. Alternative explanations, such as moral hazard, observable heterogeneity, and dynamic portfolio optimization, cannot fully account for that pattern. Building on our estimated demand elasticities and selection effects, we assess the welfare implications of information-sharing and pro-competition policies, finding that among a restricted set of policies, ensuring competition and mandating information to be shared across financial providers results in the largest surplus gains.

*Ma: yingju@mit.edu; Sztutman: asztutma@andrew.cmu.edu; Townsend: rtownsen@mit.edu. We benefited from the comments of Martin Beraja, Liyan Shi, and Shengxing Zhang. We thank the data support of Ant Group, which has helped create de-identified data that are selectively available to be used for academic research. We are grateful to the data scientists team at Luohan Academy for their assistance in constructing the datasets. Owen Engbretson, Zhiren Hu, and Jiaheng Zhao have provided outstanding research assistance. The opinions expressed are those of the authors alone and do not represent Ant Group's views.

1 Introduction

Recent years have witnessed an ever-increasing availability of data and methods to analyze it. That richness has affected, in particular, lending markets. However, the increasing availability of data and methods does not necessarily translate to the disappearance of information asymmetries or the elimination of barriers to the efficient functioning of lending markets. The use of data could destroy implicit cross-subsidies between borrowers, possibly resulting in fewer transactions, and smaller surplus (Levin, 2001). This increasing availability of data, moreover, has not been shared among all potential lenders. Large technological conglomerates have an inherent advantage in gathering more data and may be able to better position themselves as data suppliers and repositories in lending markets. This scenario raises concerns about the interplay between market power and information availability. Should the use of information be regulated in lending markets? Do the effects of monopoly power amplify or mitigate changes in information asymmetries?

To address these questions, this paper empirically investigates the prevalence of informational frictions in lending markets for small and medium-sized enterprises (SMEs) in China, and the strength of market power of a major lending provider. We find evidence that information asymmetries remain, despite the rich availability of data on potential borrowers, and that selection is advantageous. Moreover, we find that the residual demand for loans is not perfectly elastic. This finding raises the possibility that market power, by restricting the supply of loans, may attenuate the effects of advantageous selection.

Theoretically, we address the interplay of information and market power by i) constructing empirically informed counterfactual scenarios where information use is restricted or unrestricted and ii) by deriving sufficient statistics conditions for the local desirability of changing information policies under different competitive regimes, extending results from Immorlica et al. (2022) and Kolotilin (2018). Among the counterfactual scenarios we consider, we find that increasing competition and mandating information to be shared would achieve the largest surplus gains. Alternatively, allowing preserving the exclusivity of information and market power of the loan provider, but mandating it to increase the supply of loans, and its risk exposure, would achieve similarly sized surplus gains.

Our empirical strategy builds on an approach that has been used to study both adverse selection in insurance markets (Einav, Finkelstein, and Cullen, 2010), and market power in labor markets (Bassier, Dube, and Naidu, 2022; Azar and Marinescu, 2024). Both strands of the literature share the idea that exogenous variation in prices allows us to trace, in the first case, average cost curves or, in the second case, residual demand curves. The slope of average cost curves informs us about the degree of selection and

the distortions it introduces, while the slope of residual demand curves reveals the degree of market power and the distortions it causes.

To apply that strategy, we leverage comprehensive data from a major lending provider that utilizes various sources to assess borrowers' creditworthiness. Crucially, we exploit a specific policy change enacted by this lender during the COVID-19 pandemic: the introduction of interest rate discounts for new loans over a period of approximately three months to SMEs located in Hubei province, home to the city of Wuhan. While Wuhan faced stringent COVID-19 measures, other cities within Hubei experienced policies similar to those in neighbouring provinces. Therefore, we use surrounding cities outside Hubei as a control group and the cities within Hubei that received the interest rate discounts as the treatment group. To ensure comparable exposure to the COVID-19 shock, we focus on neighbouring provinces—Jiangxi, Hunan, and Henan—that had quarantine timings closest to Hubei (excluding Wuhan). We also select cities approximately equidistant from Wuhan. Our control group comprises SMEs in six cities: Anqing, Changde, Lu'an, Jiujiang, Xinyang, and Yueyang, while the treatment group includes SMEs in five cities with in Hubei province: Jingzhou, Jingmen, Suizhou, Xiangyang, and Yichang (see Figure 1).

When comparing control and treatment, before and after the interest rate discount policy is implemented, we find that the average probability of repayment falls as a response to the drop in the offered interest rates. The fall in repayment probabilities happened not only on average across different borrowers' profiles but also within each profile, as identified by an internal credit score assigned by the credit provider. This finding aligns with the idea that there is *advantageous selection* in lending markets for SMEs in China.

We consider a variety of alternative explanations, including moral hazard, dynamic optimization of repayment decisions of borrowers, observable heterogeneity in elasticities and repayment probabilities. None of those explanations can fully account for the patterns we observe in the data. First, moral hazard would predict that repayment probabilities would go up, not down, as interest rates decrease, which is at odds with our empirical evidence. Second, when looking only at new borrowers, we also observe that a negative effect of the interest rate discount on average repayment probabilities. Finally, those patterns are present at even finer-grained levels than the interest rate group the provider uses to classify borrowers. The findings also hold after further subdividing borrowers by their credit limits, and are also robust to the inclusion of additional variables the provider could have used to classify borrowers.

To extract further lessons from our estimates, we consider a variety of counterfactual scenarios and we extend theoretical results on information design from Immorlica et al. (2022), Kolotilin (2018) and Kolotilin et al. (2025)[1]. We compute equilib-

[1]See also Farboodi et al. (2025).

rium prices and quantities under a variety of competitive regimes: perfect competition, monopoly, and a monopoly with regulated risk exposure – which more closely describes the empirical prices and quantities we see in the data. Moreover, we also consider a variety of informational regimes: not allowing any information to be used when pricing loans, and allowing and mandating information to be shared across lender providers (i.e., in a similar way to what has been described as *open banking*). We find that an *open banking* policy coupled with ensuring competition delivers the highest surplus gains. Similarly sized surplus gains however could be alternatively achieved by allowing the financial provider to use the exclusive information it has, but mandating it to increase its risk exposure level.

Additionally, we show that, the result in Immorlica et al. (2022) that under perfect competition, under an borrower surplus maximizer information policy, there should be an increasing relationship between prices and the ratio of the price derivatives of profits over quantities,[2] can be modified to accommodate the case of a monopolist. In this case, we show that there should be an increasing relationship between prices and the ratio of *second derivatives* of the profit function over quantities. Also can be modified the result that under perfect competition, at each bundle, across signals, there should be an affine relationship between inverse markups and elasticities of profits. In the monopoly case, we show that there should be an affine relationship between *modified* markups – the ratio of revenues over the derivative of profits – and the elasticity of the *derivative* of profits. We verify in the data that these conditions are violated for all simple policies we considered, illustrating that there are further gains from more elaborate information disclosure policies.

Related Literature This paper builds and contributes to four strands of literature on lending markets and information imperfections.

First, on the literature on *empirical estimation of adverse selection and moral hazard*. There is a long line of research on the adversely selected markets, starting with Akerlof (1970), and recently summarized in Einav, Finkelstein, and Mahoney (2021). More recently, a vast growing literature has appeared on empirical tests for adverse selection and estimation of supply and demand in adversely selected markets (Finkelstein and Poterba, 2004; Cohen and Einav, 2007; Einav et al., 2010; Hendren, 2013, 2017; Handel and Kolstad, 2015; Kurlat and Stroebel, 2015; Finkelstein and Notowidigdo, 2019; Cabral et al., 2022). In particular, we are closely related to research on the credit markets such as Dobbie and Skiba (2013), DeFusco, Tang, and Yannelis (2022), and Collier and Ellis (2024). We document that there is likely substantial *advantageous selection* in credit markets for SMEs, where reducing the interest rate results in a decrease in repay-

[2]Or, equivalently, there should be an increasing relationship between prices and the elasticities of the average value to the lender with respect to prices.

ment probabilities, and estimate the full picture of a credit market, from the probability of application and loan demand to the repayment rate, enabling evaluations of the impacts of various policies.

This paper also relates to recent studies that reveal the role of information in the credit market and its welfare implications. Blattner, Hartwig, and Nelson (2022) show that the adverse selection in the US credit card market generates substantial welfare loss, and they find that requiring histories to be shorter can significantly improve social welfare. Agarwal et al. (2024) study the interaction of search and screening. They show that screening is informative and improved screening technology can increase interest rates even for the most creditworthy groups, driven by the spillover in lender's rate posting. Nelson (2023) studies the consequences of the CARD Act price restrictions and finds that after limiting credit card lenders' ability to raise interest rates based on new information, average prices decrease, and consumer surplus rises at all credit scores. Jansen, Nagel, Yannelis, and Zhang (2022) shows that the data advantage can create the transfer of surplus from borrowers to lenders, removing prior bankruptcy information can substantially increase the social surplus of affected borrowers while mildly reducing social welfare.

Second, we contribute to a rising literature on the *competition between fintech lenders and banks*. As demonstrated by seminal works in Broecker (1990) and Shaffer (1998), increasing competition in lending markets with information asymmetries can reduce efficiency. The growing adoption of digital technologies has introduced new factors into the competitive dynamics and market structure of the credit market. On one hand, FinTech lenders, often companies that have access to digital payment data (Ghosh, Vallee, and Zeng, 2023; Liu, Lu, and Xiong, 2022; Gambacorta, Huang, Li, Qiu, and Chen, 2023) or possess advanced capabilities in processing alternative data (Berg et al., 2020; Agarwal et al., 2019; Jagtiani and Lemieux, 2019; Nam, 2023), have rapidly risen by leveraging their superior risk assessment abilities. The fast-growing empirical literature on Fintech lenders shows that they are catching up in credit markets and intensifying competition (Berg, Fuster, and Puri, 2022; Vives, 2019). On the other hand, traditional banks are swiftly adopting digital technologies to enhance their ability to process information (Koont, 2023), making the competitive outcomes between banks and fintech lenders unclear. Several theoretical works have studied the driving forces and possible outcomes in the credit market when fintech lenders compete with banks. The advantage of fintech lenders may arise from flexibility and convenience (Liu et al., 2022), better screening ability (He, Huang, and Zhou, 2023), superior flexibility in pricing (Vives and Ye, 2024), the ability to harden the soft information (He, Huang, and Parlatore, 2024), and platform advantage in repayment enforcement (Li and Pegoraro, 2022). Our contribution on this literature is to empirically quantify the welfare improvements driven by sharing information and by increased competition in credit

market, respectively. It provides a benchmark for designing policies regarding data-sharing and competition among Fintech lenders and banks.

Third, on the literature of *information design in markets with asymmetric information*. The theory works including Levin (2001) and Mahoney and Weyl (2017), both pointing to a negative relationship between reducing information asymmetry and social welfare. Mahoney and Weyl (2017) studies the interaction of market power and selections, showing that the methods used to offset adverse selection can reduce the social surplus. Different from earlier works that take a concavication approach (Kamenica and Gentzkow, 2011), some recent works apply a linear programming approach, including Kolotilin (2018), Immorlica, Sztutman, and Townsend (2022), and Kolotilin et al. (2025). In particular, Immorlica, Sztutman, and Townsend (2022) analyze the optimal design of credit scores when there are information asymmetries in the credit markets. They show properties that optimal disclosure policies need to satisfy in a competitive environment, and, in particular, across different markets, there should be an increasing relationship between prices and the price elasticities of the expected value of the loans for investors. We extend their results to non-competitive markets, showing that, across markets, there should be an increasing relationship between prices and the curvature of the profit function.

Fourth, our work is related the theory literature in *price discrimination and market segmentation*. Bergemann, Brooks, and Morris (2015) first shows that any inefficient market can be segmented in a way that is Pareto improving in a single product market. Haghpanah and Siegel (2023) shows that, in the multi-products case, generically market segmentation can be Pareto-improving from any inefficient market. Other studies consider the optimal segmentation (Braghieri, 2019; Hidir and Vellodi, 2021), maximum consumer surplus can be achieved (Ichihashi, 2020), and when the surplus triangle can be achieved (Bergemann, Brooks, and Morris, 2015; Haghpanah and Siegel, 2022). The rise of big data technologies enhances the sellers' ability to segment the market with consumer data. It brings a theoretical question that whether/when data collection can improve the social surplus. Farboodi, Haghpanah, and Shourideh (2025) characterize the conditions of market demand curves under which the use of data is always good for social surplus in a monopolist setting. Deb, Pai, and Roesler (2024) analyze a similar question in the bilateral trade. Different from these works that mainly focus on the product market, where the cost of products is fixed (or depends on private information), we study the credit market, where the cost of each loan depends on private information and private action. We characterize and estimate the conditions under which full disclosure is seller (lender)-optimal and carry out counterfactual analysis of the welfare implications of different information disclosure policies. We also provide evidence on how noisy information can sometimes increase social welfare, as suggested in theoretical works.

The remainder of the paper is structured as follows. Section 2 introduces the institutional background of the credit market and our data. Section 3 presents the empirical strategy and the sample selection. Section 4 presents the empirical results with various robustness analyses. Section 5 presents our counterfactual analysis. Section 6 concludes.

2 Institutional Background and Data

Alipay is a payment platform that provides consumers and merchants with online and mobile payment services. As of June 2020, Alipay is used by 1.3 billion consumers and 80 million merchants. It was created in 2004 by Alibaba to provide a trustworthy third-party payment method for e-commerce buyers and sellers. With the adoption of QR code payment in 2014, it has quickly expanded from online e-commerce to offline use cases, becoming a universally accepted payment tool in China.[3] MYbank is a privately-owned bank initiated by Ant Group, which holds a 30% stake and leads its operation. Ant Group shares data with MYbank, and the bank then provides and manages credit loans to borrowers through Alipay and its own mobile application.

The loans are unsecured and require no collateral. Upon approval, the bank gives borrowers a menu detailing their credit limits and interest rates, allowing them to borrow within their limits. The loan terms are six or twelve months. Borrowers can repay early at any time, paying interest solely for the actual days and amounts borrowed. The annualized interest rates range from 5.4% to 21.6%. Approximately 96% of borrowers' interest rates fall into nine rate groups, ranging from 7.2% ($= 1.8\% \times 4$) to 21.6% ($= 1.8\% \times 12$), with each group increasing by 1.8 percentage points.[4] The average credit limits is 9,300 CNY ($\approx$ 1,350 USD) in 2019. Figure A3 shows the distribution of the loan amount.

MYbank sets and adjusts the credit limits and interest rates for borrowers based on their payment activities, loan usage, repayment history, and other individual activities and information shared with Alipay. If borrowers want to increase their credit limit, they can provide additional documentation and request a manual review. Apart from this, all decisions are determined by an algorithm-driven system.

The common feature of these loan borrowers is that they accept digital payments from customers. Given the widespread adoption of digital payments and Alipay in China, this group covers almost all types of merchants. The top five types of busi-

[3] Alipay is the largest payment platform in the mobile payment market in China, with about 45% market share, followed by WeChat Pay and UnionPay.

[4] The annual interest rate of 1.8% corresponds to a daily interest rate of 0.005%, or 0.5 base points. When applying for a loan, borrowers will see the interest rate displayed as both a daily and an annual rate. As the loan usage period is relatively short, the daily rate allows borrowers to more conveniently estimate the necessary interest.

nesses are restaurants, grocery stores, clothing shops, home decor stores, and beauty services. The average annual sales through Alipay is 92,100 CNY ($\approx$ 13,300 USD) in 2019. There are other Fintech banks in China that were lending to borrowers using big data during the period of the study, but their customer bases were much smaller than those of the bank we are examining. Traditional bank loans typically require collateral and are usually issued in much larger amounts[5], therefore does not constitute direct competition for MYbank.

2.1 Covid-19 and Interest Rate Discount Policy

Timeline of Lockdowns Beginning on January 23, 2020, with the outbreak of the COVID pandemic in Wuhan, the capital city of Hubei province, nearly all cities in China were placed under complete lockdown, permitting only essential personnel and vehicle mobility. By late February, some cities started to gradually ease restrictions, and by April 30, with Wuhan also reopening, the lockdown was largely lifted. During this time, Chinese authorities designated specific risk levels for each district, which are administrative units smaller than cities and counties. The risk level —low, medium, and high— is based on the pandemic situation and corresponding management measures are applied.[6]

Interest Discount Policy In early February 2020, MYbank announced a policy to offer a 20% discount on loan interest rates for all small and micro businesses in Hubei province, as support for these merchants to live through the pandemic. The policy went into effect on February 10 and lasted until April 30, ending without any prior announcement. The policy applies for 1.5 million merchants in the province.

[5] Liu, Lu, and Xiong (2022) study the borrowers that borrow from both the big tech lender and banks. They compare the loan and usage characteristics and find that big tech loans are smaller in amount, and borrowers tend to repay far before maturity and borrow more frequently, mainly serving for the short-term liquidity rather than long-term financing needs.

[6] On February 17, 2020, the State Council's *Joint Prevention and Control Mechanism* released the "Guidance on Scientific Prevention and Control and Targeted Policies for Effective COVID-19 Prevention and Control by Zones and Levels." The *Guidance* specifies the criteria for determining risk levels: areas (counties, cities, districts) without any confirmed cases or with no new confirmed cases for 14 consecutive days are designated as *low*-risk; areas with new confirmed cases within the past 14 days, totaling no more than 50 cases overall, or areas with more than 50 cases but no new clusters within 14 days, are classified as *medium*-risk; and areas with over 50 cases in total and new clusters within 14 days are classified as *high*-risk.

Low-risk areas are required to implement the strategy of "preventing external input" and fully restore normal production and living order, and lift road traffic restrictions. Medium-risk areas should adopt the strategy of "preventing external input and internal spread" and restore normal production and living order in an orderly manner. High-risk areas should implement the strategy of "preventing internal spread, preventing external output, and strict control."

Table 1: Timeline of Lockdown Policy

City	Lockdown starts	Low risk	Lift all lockdown measures	Duration (days)	Distance to Wuhan (km)
Wuhan	Jan. 23	Apr. 18	Apr. 30	98	-
Treatment					
Jingmen	Jan. 24	Mar. 16	Mar. 25	61	242
Jingzhou	Jan. 24	Mar. 13	Mar. 17	53	220
Suizhou	Jan. 24	Mar. 13	Mar. 25	61	170
Xiangyang	Jan. 24	Mar. 13	Mar. 25	61	314
Yichang	Jan. 24	Mar. 13	Mar. 25	61	325
Control					
Anqing	Jan. 24	Mar. 2	Mar. 18	54	323
Changde	Jan. 23	Mar. 22	Mar. 22	59	386
Jiujiang	Jan. 24	Mar. 6	Mar. 21	57	240
Lu'an	Jan. 24	Mar. 3	Mar. 18	54	304
Xinyang	Jan. 25	Mar. 7	Mar. 16	51	216
Yueyang	Jan. 23	Mar. 22	Mar. 22	59	228

Note: "Low risk" indicates that all areas in the city are turning into low-risk areas. We use "lifting all lockdown measures" to indicate that all residential communities will be released from lockdown, and any restrictions such as access to public transportation and entry to buildings are solely based on the status of the individual's personal digital health code. The code is a mobile application that contains personal health information and recent travel history, and serves as an individual's electronic pass. The road distance is measured by the road distance suggested by Google map and Amap.

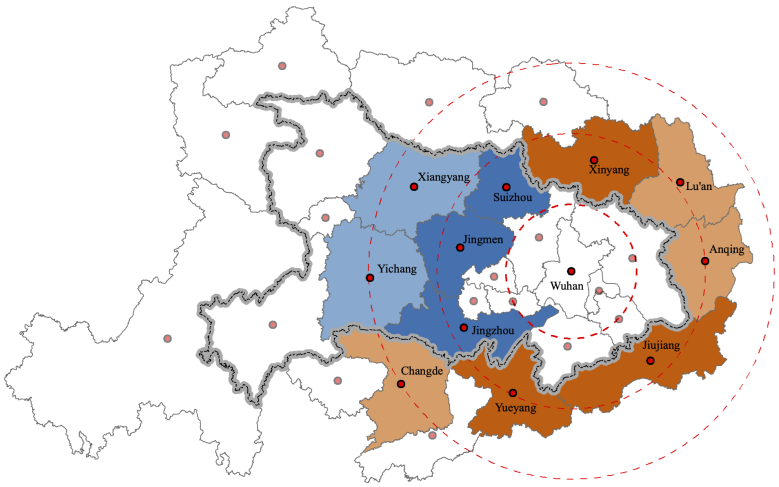


Figure 1: Treatment and Control Cities Map

Note: This figure shows cities in the sample in relation to Wuhan. The grey outline denotes the borders of Hubei province. Blue prefecture-level cities are the treatment cities; orange are control cities. For each set, the darker colour indicates the inner circle and the light colour represents the outer circle cities. The three rings centered on Wuhan represent straight-line distances of 100 km, 200 km, and 300 km, respectively.

Treatment and Control Group During the lockdown period, each province developed its specific control measures for pandemic prevention, with China's administrative divisions structured by provinces. The provinces neighboring Hubei imposed relatively strict pandemic control policies on the cities adjacent to Hubei to reduce the risk of importing cases from Hubei into these areas. Therefore, in terms of pandemic risk levels, the duration of lockdown measures, and the extent of economic impact, we found that cities adjacent to Hubei and the cities within Hubei Province, excluding Wuhan, were quite similar. Hence, the merchants in the cities adjacent to Hubei experienced comparable pandemic impacts, but were not granted the interest rate discount. We utilize these features of lockdown policies to select the treatment and control group.

We select cities based on the principle of ensuring similar levels of pandemic impact. The primary economic impact stems from the strictness of lockdown measures, which, in turn, depends on the number of cases. Case numbers are correlated with the city's distance from Wuhan. On the other hand, Wuhan's stringent lockdown measures resulted in spillover economic effects, with the severity of these effects also linked to the city's proximity to Wuhan. Accordingly, we selected two groups of cities with similar distances from Wuhan: one within Hubei Province, and the other consists of cities in neighbouring provinces that are directly adjacent to Hubei's border.

Figure 1 shows the cities we selcted. We choose 5 cities in Hubei as our treatment group, including Jingmen, Jingzhou, Suizhou, Xiangyang, and Yichang, and 6 in surrounding provinces as the control group, including Anqing, Changde, Jiujiang, Lu'an, Xinyang, and Yueyang. Within each group of cities, we further divided them into inner-circle cities, located within a road distance of 200 kilometres from Wuhan, and outer-circle cities, situated at a road distance of approximately 300 kilometres from Wuhan.[7]

Table 1 presents the timing and duration of lockdown policies for the two groups of cities, demonstrating that the policy implementations are highly similar across the groups. In addition, we analyze the similarity of two groups with a series of economic activity indexes, as shown in Figure A5. These indexes capture the total consumption and sales of each city at the weekly level. It is aggregate data from the payment platform Alipay. Figure A6 present the difference of these indexes between the two groups, showing that the differences in consumption and sales between the two groups are statistically insignificant.

2.2 Data

The data is provided by Alipay and covers the credit and business histories of small and medium enterprises (SMEs) enrolled in the credit provider's lending platform.

[7]See Table 1 for the detailed road distance information. As road distances differ from straight-line distances, the inner-circle cities are approximately 150 kilometres from Wuhan in the straight-line distance, while the outer-circle cities are about 200 to 250 kilometres away, as in Figure 1.

Our full sample consists of borrowers across the 11 cities in the treatment and control group, amounting to just under 470,000 unique SMEs. In our baseline specification, we further limit our focus to borrowers in industries highly exposed to the local economy,[8] this corresponds to 40% of the sample, at just under 200,000 unique borrowers. For borrowers in the covered cities, we observe the terms of originated loans (such as loan amount, interest rates, applied discounts, and disbursement, due and settlement dates), and information on the daily payment behavior over the life of each loan. To analyze changes in repayment behavior as a result of the policy, we identify treated loans as loans for borrowers in Hubei with disbursement dates in the policy period (February 1st, 2020 - May 10th, 2021), recording the final repayment proportion for each loan. Guo et al. (2022) analyze the recovery of business for 80 million small merchants, including borrowers in our sample. They show revenue had rebounded to around 80% in early April, where they should be seven weeks after the COVID-19 outbreak, and had remained at this level until the end of April, when their data ends.

To to look at changes in borrower loan demand, we construct a balanced panel for each borrower covering September 2019 through the end of 2020, which summarizes weekly new loan and repayment amounts. Borrowers are included in the panel if they are recorded as having loan activity (originating or making a payment towards a debt) in the dataset— a time period spanning from September 2019 to February 2022. All borrowers included in the panel have either originated loans, or made payments to pre-existing loans in this 2 year period. We define week 0 of the panel as the week of September 1st 2019, and summarise all loan activity for each week for the following 70 week period. With this dating scheme, the policy period corresponds to weeks 24 through 36. In addition to credit data, the dataset includes information on SME business activity and demographic characteristics, such as: weekly amounts of sales and purchases, proprietor's age, gender and city of residence. Summary statistics for variables of interest in the panel and loan variables, in both control and treatment groups, are described in Table A1

Furthermore, as the platform's interest rate policy coincides with the timing of lockdown policies, we also include controls for city-level exposure to shock from COVID shock that could confound our estimates on loan performance and demand. In particular, we use a measure of trade exposure to Wuhan based on e-commerce flows between cities. This index is defined as follows: for a given month t, if $out_{c,j,t}$ are the outflows from city c to city j, and $in_{c,j,t}$ are the flows in to city c from city j, we define our trade intensity index for each city c vis-a-vis Wuhan as $I_{t,c} = (in_{c,Hubei,t} + out_{c,Hubei,t})/(\sum_j in_{c,j,t} + \sum_j out_{c,j,t})$

[8]Our selected industry categories are the grocery, catering, and apparel industries. The pre-policy industry composition of both control and treatment cities is reported in Table A2.

3 Empirical Strategy

In this section, we describe our strategy to estimate the elasticities of demand for loans with respect to interest rates as well as the elasticities of repayment rates with respect to interest rates. Those elasticities are informative both about the degree of market power and about the presence of informational frictions, such as adverse selection and moral hazard.

Elasticities of residual demand for loans can be interpreted through the lenses of market power, as in the literature on monopsony in labor markets (Sokolova and Sorensen, 2021). If there is no market power, then as the financial provider raises interest rates above the market prevailing interest rate, the demand for their loans would collapse quickly. Conversely, if they lower interest rates below the prevailing rate, they would attract all the borrowers. Non-extreme elasticities of demand, therefore, can be interpreted as evidence of some degree of market power.

Elasticities of repayment rates can be interpreted through the lenses of informational frictions, as in Karlan and Zinman (2009) and Einav et al. (2010). If there is adverse selection, when interest rates exogenously decrease, a different pool of borrowers would take up a loan, and in particular, those who had just been attracted by the lower interest rate would have higher expected repayment rates, and the average quality of the pool of borrowers from the point of view of investors would improve. Conversely, if the selection is advantageous, when interest rates decrease, the borrowers who had just been attracted by the lower interest rate would have lower expected repayment rates, and the average quality of the pool of borrowers from the point of view of investors would deteriorate.[9] If there is moral hazard, holding fixed the contract, when interest rates exogenously decrease, safer projects become relatively more attractive.[10] In response to an interest rate decrease, therefore repayment rates may increase. Therefore, exogenous variation can be informative for both moral hazard and adverse selection. Interestingly, moral hazard has a one-sided prediction, whereas selection can work both ways: it can be adverse or advantageous, that is, repayment rates can either go up or down.

One alternative explanation is that, under lower interest rates, penalties for being overdue are smaller, which could lead to lower repayment rates. However, for this particular loan, the overdue penalty ranges between 1 to 1.5 times the interest the borrower needs to pay at the time of overdue. For these loans, the average amount is 6,854 *yuan*, outstanding balance 1,220 *yuan* with average total interest of 440 *yuan*. The penalty fee of these loans has an average of 110.3 *yuan*, median of 1.14 *yuan*, and third quantile of 15.95 *yuan*, suggesting that the difference in penalties due to changes in interest rates

[9] See Appendix E.1 for the formal identification argument.
[10] See Appendix E.2 for the formal claim.

is minimal and unlikely to impact repayment behaviour significantly.

To identify the supply of bonds and their expected values as a function of prices, we explore the policy change Alipay enacted during Covid-19 as a source of exogenous variation in prices. MYbank offered a 20% discount on interest rates only for SMEs in Hubei province, from February 10 to April 30, and did not offer the discount for the neighbouring provinces.

The key challenge in leveraging this source of variation is distinguishing the effects of the pandemic and containment measures from the interest rate discount policy. To do so, we rely on the different geographical distributions of the two. The discount policy was restricted to Hubei. The pandemic and containment measures were not. In fact, the Covid pandemic started in Wuhan (the capital and main city of Hubei), where a stringent lockdown policy was implemented. Cities in the region were directly and indirectly impacted by the restriction implemented in Wuhan and their own restrictions, as well as by the pandemic itself, and cities within and outside Hubei were similarly affected.

Thus we select a treatment group within Hubei consisting of cities within a 200 and 300 km road distance from Wuhan. Similarly, we select as a control group, cities within the same road distance (Figure 1). For each specification, we assess the robustness of our estimates separately for the inner-circle group and the outer-circle group. We also perform robustness checks for borrowers in different industries. In our baseline specification, we focus on three key categories that are highly dependent on the local economy: restaurants, grocery stores, and clothing shops, which collectively account for 40% of borrowers in the sample.

4 Empirical Results

In this section we present the empirical results on the effect of interest discount policy on repayment, borrower composition, and loan demand.

4.1 Policy Effects on Repayment

To exploit the natural experiment outlined above, we employ a standard differences-in-differences design: for a loan originated in week t and in city c we estimate:

$$Repay_{t,c} = \beta_1 Treat_t + \beta_2 H_c + \beta_3 (H_c \times Treat_t) + \Gamma I_{t,c} + \mu_c + \epsilon_{t,c} \tag{1}$$

Where $Treat$ is a time dummy equal to 1 when the loan is originated during the policy period and H_c is a group dummy equal to 1 if the city c belongs to Hubei province.

Furthermore, to control for the exposure to Wuhan and the economic effects of the COVID shock, which may confound repayment behaviour, we include an exposure index I interacted with the time dummy $Treat$; we also include city-level fixed effects μ_c to control for any unobserved factors determining repayment. The results of estimating the model described above are reported in Table 2.

Table 2: Repayment Rates

	Repayment Rate At Maturity		
	(1)	(2)	(3)
$H_c \times Treat_t$	−0.41	−1.14	−1.19
	(0.07)	(0.07)	(0.07)
$Treat_t$	0.15	0.72	0.76
	(0.04)	(0.04)	(0.04)
H_c	−0.57	−1.31	−1.29
	(0.03)	(0.04)	(0.09)
Covid Controls	No	Yes	Yes
City FE	No	No	Yes

Note: Table reports difference-in-differences results on loan-level final repayment (repayment ranging from 0-100%). H is a province dummy indicating a loan is originated in Hubei province; $Treat_k$ is a time dummy indicating origination occurs during the policy. COVID control is the trade intensity index described in preceding sections. Robust standard errors are reported in parentheses.

The effect of the policy is summarized in the first row of Table 2. Column (1) reports a baseline specification with no controls for COVID exposure nor fixed effects; Column 2 adds the control (with the index being the logarithm of the inflow-index described in Section 3; Column 3 includes city fixed effects. In the full specification, as a result of the rate-decreasing policy, repayment rates in Hubei dropped around −1.2 percentage points (p.p.).

Despite controlling for exposure to the COVID shock, and conditioning on borrowers within industries that would be uniformly affected by lockdown policies regardless of province, pre-trends in the differences between control and treatment are still a concern. To evaluate assumptions around parallel pre-trends that are required for the unbiasedness of our estimates in Table 2, we also estimate a dynamic differences-in-differences (event study) model. For time dummies T_t, we estimate Equation 2 for loans in city c, where the estimates of each dynamic treatment effect β_t are normalized by the pre-policy difference (week 23).

$$Repay_{t,c} = \sum_{t=0;t\neq 23}^{70} \beta_t (H_c \times T_t) + I_{c,t} + \mu_c + \epsilon_{c,t} \tag{2}$$

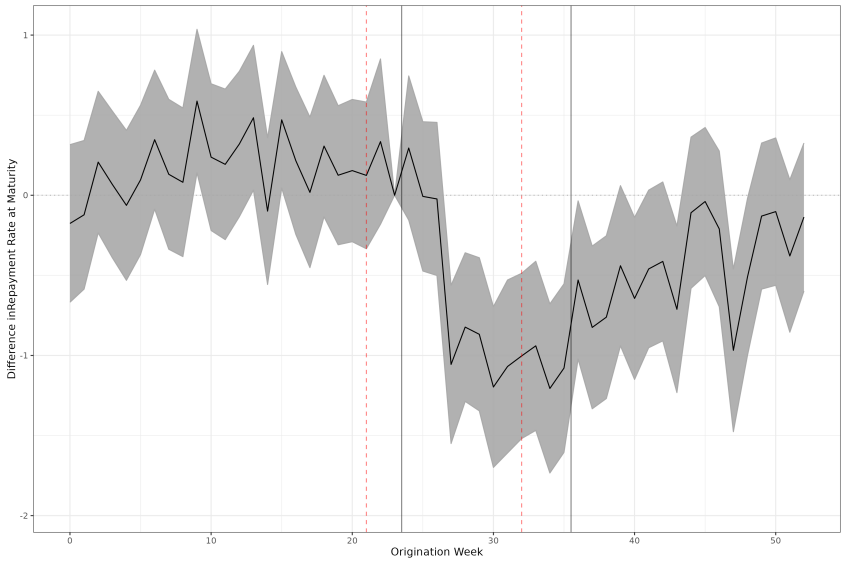


Figure 2: Dynamic DiD: Repayment Rate at Maturity

Note: This figures reports dynamic difference-in-differences estimates for repayment rates. Estimates are normalized to the week 23 level; 95% confidence intervals are reported by the shaded bands. . Week 0 corresponds to the week of September 1st, 2019. Policy starts week 24 and persists until week 36 (shown by black the vertical bars). Important Covid events are shown by the dotted red lines: line 1 shows the start of the Wuhan lockdown (January 23rd 2020), line 2 shows the end of the lockdown (April 8th, 2020)

The estimates from the dynamic model are reported in Figure 2.[11] As can be seen, for loans originated in the pre-policy period, differences in repayment are largely not statistically significant from zero. Within a few weeks of the onset of the interest rate policy, repayment rates fall in Hubei by around -1.1 p.p., relative to control cities.

4.2 Borrower Composition

Borrowers in the sample are roughly split into 9 credit categories, as indicated by the interest rate they are offered on new loans. The vast majority of loans[12] have initial rates given by annual rates of 1.8% increments of 7.2% to 21.6%. Given the frequency of these assigned rates, they provide a sufficient proxy for the lending platform's assessment of the borrower's risk profile, as determined by the vast amounts of data

[11] For brevity, only the first 52 weeks of estimates are reported

[12] These 9 pre-discount interest rates account for 94.62% of all interest rates, with 14.4% being the most common (22.67% of loans), and 7.2 being the least common at roughly 2.13% of loans

the lending platform has access to. To construct approximate counterfactuals of when lending platforms do not have this rich information, we also estimate an elasticity of repayment with respect to the price of the loan, where the price varies with the interest rate paid□

$$\epsilon^r = \frac{\Delta Repay}{\Delta P}\frac{P_0}{R_0} \tag{3}$$

Where $\Delta Repay$ is the slope provided by the DiD estimates and R_0 is the baseline repayment rate. For a given interest rate i, P_0 is the price of a unit loan, given by $\frac{100}{100+i}$ ΔP is the change in price implied by the 20% policy discount, given by $\Delta P = \frac{100}{100+.8i} - P_0$. To explore how the policy may have heterogeneously affected repayment across these borrowers and to estimate calculate of repayment, we estimate Equation 1 for each interest rate separately and report the results in Table 3

Table 3: Repayment Rates: Across Credit Categories

Estimate	9	10.8	12.6	14.4	16.2	18	19.8	21.6
				A. Repay Rate				
Baseline	98.196	98.464	98.086	97.69	97.678	97.491	97.502	98.167
	(11.645)	(9.886)	(11.198)	(11.968)	(11.877)	(12.509)	(11.957)	(10.765)
DiD Slope	-1.004	-1.309	-1.055	-1.187	-1.281	-1.713	-1.224	-0.45
	(0.35)	(0.227)	(0.201)	(0.159)	(0.169)	(0.205)	(0.214)	(0.181)
Elasticity	-0.609	-0.669	-0.47	-0.47	-0.457	-0.558	-0.367	-0.124
	(0.22)	(0.1)	(0.09)	(0.06)	(0.06)	(0.07)	(0.06)	(0.05)
Obs	25222	81217	135086	184099	167638	130355	110057	111627

Note: Table reports regression results across the 8 borrower groups as defined by the origination interest rate; results are from the full specification, including controls for COVID exposure and city fixed effects. The first row reports the pre-policy average repayment rate across all cities, with the sample standard error given in parentheses. The second row reports the DiD coefficient, with robust standard errors reported in parentheses. The third row reports the repayment elasticity with respect to the interest rate discount, standard errors are calculated with the sub-sample (using a fixed proportion $p = .25$) bootstrap (200 replications).

The first row of the table reports baseline repayment rates corresponding to the pre-policy average repayment rate across all cities in the sample. As can be seen, repayment rates are high, nearly 98% across all borrower groups. The second row shows the change in repayment due to the policy, as given by the difference-in-differences coefficient. These coefficients are generally negative across credit categories, indicating that lower interest rates are associated with a decline in repayment rates, indicating that repayment rates are lower when interest rates are decreasing. The third row shows the elasticity, as calculated from the preceding rows according to Equation 3. These values capture the percentage change in repayment rates for a one-percentage-point change in interest rates. The elasticities are negative for all the credit categories, confirming

the advantageous selection.

4.3 Borrower Portfolio Optimization

In addition to selection, repayment rates could also decline due to changes in borrower behavior. For borrowers with loans originated before and after the discount, it may be optimal to pay off the higher interest rate loans first. This portfolio optimization on the part of the borrower, holding their risk category fixed, could account for the decline in repayment rates.

To evaluate the changes in repayment more likely to come from selection, we consider loans from borrowers who have only discounted loans. We consider a loan "active" if it is originated or has an outstanding balance during the policy period. Loans are then considered "policy" loans if originated during the policy period, or "legacy" if originated before. Borrowers who have no "legacy" loans, but at least 1 policy loan would not be faced with a portfolio optimization problem that could induce a decline in repayment for policy loans. To evaluate loan repayment for these borrowers, we consider the difference of means estimator given by Equation 4:

$$Repay_{t,c} = \beta_0 + \gamma_1 H_c + \gamma_2 I_c + \mu_c + \epsilon_{t,c} \tag{4}$$

For borrowers with no legacy loans, the only loans we observe in the data are originated during the policy period. Hence, we estimate an equation with the dummy H_c indicating if the borrower originates the loan in a city c in Hubei, controlling for the levels of Wuhan exposure given by the index I_c, and including city fixed effects μ_c. The results are shown in Table 4.

Table 4: Repayment Rates: No Legacy Loans

	Repayment Rate At Maturity		
	(1)	(2)	(3)
Hubei	−0.59	−1.52	−2.04
	(0.06)	(0.07)	(0.15)
Covid Controls	No	Yes	Yes
City Fes	No	No	Yes

Note: This table reports difference in means estimates on loan-level final repayment for borrowers who only have loans originated during the policy (i.e. no pre-policy outstanding loans that have carried over). H is a province dummy indicating a loan is originated in Hubei province. COVID control is an inflow index indicating the relative importance of Wuhan as a source for individuals flowing into a given city; index is the pre-lockdown (January 2020) monthly average. Robust standard errors reported in parentheses.

The table shows that, after ruling out the selection effect of paying back old loans, loans with a discount policy still show lower repayment rates, consistent with the findings in the previous section.

4.4 Policy Effects on Loan Demand

In addition to repayment behavior, we also evaluate the change in demand for loans as a result of the policy. Utilizing the borrower-week, balanced panel mentioned prior, we again utilize the policy to estimate a standard differences-in-differences model. For a borrower j, in city c during a week t, we estimate:

$$Y_{j,t,c} = \mu_c + \alpha_0 Treat_t + \alpha_1 H_c + \beta(H_c \times Treat_t) + \Gamma(I_{t,c}) + \epsilon_{j,t,c} \tag{5}$$

Where Y is an outcome (new loan amount in *yuan*, probability of a new loan, and average loan amount[13]). $Treat_t$ is an indicator for the week falling in the policy period (weeks 24 through 36); H_c indicates if the borrower is in a city c in Hubei province; I_c is an exposure index, and μ_c are city fixed effects. This equation is again estimated for all 8 credit groups: in the panel, borrowers are assigned the borrower group corresponding to the interest rate on their most previous originated loan. Furthermore, for use in constructing counterfactuals of the lending environment, we estimate elasticities

[13] Average loan amount is the average amount across all loans originated in a week, conditional on a borrower taking out a new loan.

of loan demand, provided the baseline levels of new loan amounts and the slope from the DiD estimate.

Table 5: Loan Demand: Across Credit Categories

Estimate	9	10.8	12.6	14.4	16.2	18	19.8	21.6
A. New Loan Amount								
Baseline	2015.425	1747.175	1535.977	1265.334	929.33	673.15	352.996	355.811
	(13212.598)	(11447.776)	(9268.47)	(8125.367)	(6116.432)	(4827.717)	(3054.076)	(3131.326)
DiD Slope	968.683	842.769	571.95	439.102	425.305	259.949	152.307	101.017
	(138.324)	(64.829)	(44.422)	(32.451)	(23.235)	(20.001)	(11.916)	(12.965)
Elasticity	28.624	24.261	16.266	13.438	15.955	12.272	12.622	7.708
	(4.04)	(1.8)	(1.25)	(1.05)	(0.84)	(1.01)	(0.96)	(1.07)
Obs	243522	744354	1212007	1727123	1804560	1472317	1705571	1726102
B. P(New Loan)								
Baseline	0.089	0.094	0.1	0.097	0.089	0.082	0.065	0.066
	(0.285)	(0.292)	(0.301)	(0.296)	(0.285)	(0.275)	(0.247)	(0.247)
DiD Slope	0.021	0.025	0.02	0.029	0.031	0.032	0.033	0.015
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Elasticity	14.065	13.332	8.698	11.551	12.148	12.378	14.778	6.213
	(2.16)	(1.02)	(0.64)	(0.49)	(0.45)	(0.49)	(0.44)	(0.44)
Obs	243522	744354	1212007	1727123	1804560	1472317	1705571	1726102
C. Average New Loan Amount								
Baseline	18039.991	14155.732	11917.274	9974.072	8162.975	6341.46	4250.699	4251.78
	(32800.939)	(26129.608)	(20161.544)	(17677.28)	(13963.449)	(11541.263)	(8377.826)	(8711.532)
DiD Slope	3794.744	2772.832	1261.405	230.593	462.974	-357.594	-584.148	20.407
	(1141.868)	(491.845)	(304.21)	(220.688)	(177.056)	(171.177)	(134.685)	(127.902)
Elasticity	12.528	9.852	4.624	0.895	1.977	-1.792	-4.02	0.13
	(3.97)	(1.56)	(1.19)	(0.85)	(0.72)	(0.8)	(0.87)	(0.88)
Obs	21366	68992	120938	166640	159224	119391	108545	113860

Note: Table reports results across 3 variables: new loan amount (Panel A), probability of a new loan in a week (Panel B), and average new loan amount, conditional on having a new loan (Panel C). Reported in each panel is the baseline quantity (given by the pre-policy mean), the slope of the DiD coefficient from the specification in Equation 5, and the elasticity calculated from the preceding two quantities. In parantheses, sample standard errors are reported under the baseline quantity, robust standard errors from the regressions underneath the slope quantites, standard errors for elasticities are calculated with the sub-sample (using a fixed proportion $p = .25$) bootstrap (200 replications). The baseline and DiD slopes in Panels 1 and 3 are in nominal amounts of *yuan*.

The results of estimating Equation 5 across the 8 borrower categories are reported in Table 5. Panel A shows the response to the applied new loan amount. It shows a significant increase in new loan amounts under the discount policy. The loan demand is more sensitive to interest rate changes in the lower-interest groups. Panel B presents the extensive margin of the probability of applying for a new loan. Elasticities of the probability of applying for new loans are higher in the low-interest rate groups. Panel C is for the average amount of new loans, conditional on a borrower applying for a loan. The responses are significant and relatively larger for the low-interest groups. The response is mostly driven by extensive margin: additional borrowers instead of larger loans from existing borrowers.

4.5 Robustness Analysis with Machine Learning Controls

An important concern in our empirical strategy is that changes in the composition of borrowers following the interest rate policy could drive our results through observable characteristics rather than through selection on unobservables. To address this concern, we employ machine learning techniques to flexibly control for a wide array of borrower characteristics that the lending platform observes at the time of loan origination.

Following the approach of Athey et al. (2019), we implement regression forests to generate predicted repayment rates $\hat{R}^f_{i,t} = f(X_{i,t})$ for each loan i at time t, where $X_{i,t}$ represents the vector of observable borrower characteristics. Specifically, we include:

1. Personal characteristics: occupation, education, gender, age, native flag, city
2. Credit: Sesame credit score, consumer credit amount, business credit amount
3. Payments/sales: Sales and personal spending (amount and number of payments)
4. Current loan variables: loan origination amount, total balance, current balance
5. Past loan variables: lagged loans paid, lagged payment amount, rolling average repayment across loans

We then augment our baseline difference-in-differences specification with this predicted repayment rate:

$$R_{i,t,c} = \hat{R}^f_{i,t} + \beta_1 \text{Treat}_t + \beta_2 H_c + \beta_3 (H_c \times \text{Treat}_t) + I_{c,t} + \mu_c + \epsilon_{i,c,t} \tag{6}$$

We train random forests (see details in Appendix C.1) and estimate separately for each interest rate group. We repeat this exercise for two cases of training samples for the random forest: (1) a pre-policy (23-week) in-sample, and (2) the full 70-week sample. While we limit the set of information for a given loan to that accessible to the platform at the loan's origination, training the random forest on the full sample equates to us allowing the platform to account for the changing relationship between observables and repayment rates during COVID and the interest rate policy.

The difference-in-differences estimates with machine learning controls for the pre-policy in-sample and full in-sample are reported in A5 and A6, respectively. Compared to our baseline results, the estimated elasticities are attenuated, though we still see significant results indicative of the presence of advantageous selection.

Futhermore, we show in Table A8 and A9 a measure of variable importance captured in the random forest estimation for repayment rates. It shows that the combination of interest rates and credit utilization effectively summarizes the fintech lender's

information in predicting the repayment. Adding more covariates does not significantly change the results. The persistence of negative repayment elasticities, even after controlling for rich observables using flexible machine learning methods, provides strong evidence of advantageous selection in this lending market.

5 Counterfactuals

With the estimates of baselines and elasticities of repayment and loan demand described in the prior section, we turn towards a simple model describing the lending environment of a platform such as Alipay, and consider the welfare implications of reforming policies around bank competition and information access. We consider a model in which the banks purchase debt from the SMEs; for our counterfactuals: the model varies according to the competition amongst the purchasers (Monopolist vs. Perfect Competition) and the level to which banks can price discriminate due to information about types of borrowers (Informed vs. No Information). We categorize the lending environment that persisted through the COVID-era policy as the case where Alipay held a monopolist in the SME lending market with information. We compare the welfare of the status quo with three other cases: "open banking" (competition and information), "banning use of exclusive information" (monopolist and no information), and "closed banking" (competition and no information).

5.1 Model Setup

For the purpose of our counterfactual calculations, we assume that, at each market segment i, the supply of bonds Q_i (which we can also call *borrowers' demand function for loans*) and repayment rates R_i are isoelastic with respect to prices (p). That is, for each market segment, supply of bonds can be described by Equation 7, below, where $\epsilon_{Q,i}$ is the elasticity of the supply of bonds, $P_{0,i}$ and $Q_{0,i}$ are baseline prices and quantities. To fix the scale across scenarios where these markets are considered separately or when aggregated into one singular market, we assign a weight ω_i to each market reflecting the typical importance of each interest rate group in a week of the panel,[14] such that weights across all markets sum to one.

$$Q_i(p) = \omega_i \max\left\{0, Q_{0,i} \cdot \left(\frac{p}{P_{0,i}}\right)^{\epsilon_{Q,i}}\right\} \tag{7}$$

[14]Details on weight construction: For a given week t of the panel i we take the share of observations of a given rate in the selected weights as $\omega_i^t = \frac{N_{i,t}}{\sum N_{i,t}}$. For all 9 selected interest rates, the final weight is just the average proportion across all $T = 104$ weeks of the panel $\omega_i = \frac{1}{T}\sum_t \omega_i^t$

Similarly, for each market segment, the average repayment rates can be similarly described by an equation like Equation 8, below, where $\epsilon_{R,i}$ is the elasticity of the repayment rates with respect to prices, and $P_{0,i}$ and $R_{0,i}$ are baseline prices and repayment rates.

$$R_i(p) = \min\left\{100, R_{0,i} \cdot \left(\frac{p}{P_{0,i}}\right)^{\epsilon_{R,i}}\right\} \tag{8}$$

Lastly for each market segment, we define the inverse supply as:

$$P_i(q) = P_{0,i} \cdot \left(\frac{q}{\omega_i Q_{0,i}}\right)^{\frac{1}{\epsilon_{Q,i}}} \tag{9}$$

These market segments correspond to the 8 credit categories[15] that roughly partition borrowers on the lending platform. From these groups, we calculate baseline prices as $P_{0,i} = \frac{100}{100+i}$. We discipline the model with estimates obtained in the prior section. Our estimates of $\epsilon_{Q,i}$, $Q_{0,i}$ come from the first and third row of Panel 1 of Table 5, and estimates of $\epsilon_{R,i}$ and $R_{0,i}$ are from Table 3. We further assume that average values are equal to repayment probabilities times a discount factor $\delta = \frac{1}{1.02}$. That is

$$AvgV_i(q) = \delta R_i(P_i(q)) \tag{10}$$

Equation 7 and Equation 10 form the basis under which we can compute a variety of counterfactuals. For example, given estimates of Equation 7 and Equation 10 for different market segments, we can simulate what would happen if the information were concealed and the different market segments were merged into a single market.

5.2 Market Structure and Data Accessibility

To extract insights about the desirability of different competitive and information policies, we compute the equilibrium and welfare under several alternative assumptions on the market structure and data accessibility.

Case 1: Informed monopolist lender If there is a single provider, who can differentiate between different profiles of borrowers (i.e. market segments), we define the equilibrium to be such that at each market segment i, the provider picks a point in the borrowers' demand curve (q_i^m, p_i^m) to maximize profits.

Denote $P_i(q)$, the borrowers' inverse demand function. Denote the average value function as a function of quantities $AvgV_i(q) = AvgV_i(p_i(q))$. Then we can write the monopolist problem as:

$$\max_{q_i} q_i \cdot (AvgV_i(q_i) - P_i(q_i)) \tag{11}$$

[15]These 8 annualized rates are 9%, 10.8%, 12.6%, 14.4%, 16.2%, 18%, 19.8%, and 21.6%

Given (q_i^m, p_i^m), we compute surplus at each market by calculating consumer surplus and the financial provider's profits, which are then added across markets to compute the total surplus.

Case 2: Competitive equilibrium with informed lenders In our second case, we assume that financial providers have access to the information that allows them to differentiate each market segment i. Then, at each market segment i, the competitive equilibrium is described by a standard Akerlof condition, as in Equation 12:

$$p_i^* = AvgV_i(p_i^*) \tag{12}$$

Total surplus, then, is equal to consumer surplus, which can be computed by integrating the area under the demand curve at each market, and then summing up the surplus across markets.

Case 3: Competition with restricted information In our third case, we assume that no financial provider can use the information that allowed them before to differentiate across markets i. That is, we assume that all market segments are merged into a single one.

We thus aggregate the supply functions (7) and the average value functions (10) to compute equilibrium as the solution to Equation 13 and to calculate surplus by integrating the aggregate supply function.

$$p_A^* = AvgV_A(p_A^*) \tag{13}$$

Case 4: Monopolist lender with restricted information In this case, we consider the situation where there is a single financial provider, and that financial provider does not have information to differentiate borrowers (or is not allowed to use it). We define equilibrium to be such that the provider picks a point in the demand curve (q_A^m, p_A^m) to maximize profits.

Denote $P_A(q)$, the borrower's aggregate inverse demand function. Denote the average value function as a function of quantities $AvgV_A(q) = AvgV_A(p_A(q))$. Equilibrium is determined by the profit-maximizing condition:

$$\max_{q_A} q_A \cdot (AvgV_A(q_A) - P_A(q_A)) \tag{14}$$

We then calculate total surplus as the sum of the lender profits and borrowers' surplus, which can be computed by integrating the area under the borrowers' aggregate demand curve.

Case 5: Regulated and informed monopolist lender As challengers in the credit market, fintech lenders operate under significant regulatory uncertainty. As a precaution, they proactively manage risk within a certain threshold to ensure that their activities do not trigger financial instability or systemic risk. In this case, we consider the possibility that the financial provider limits its exposure to risk. That is the product of the discounted average repayment rate and the volume of loans the lender issues should be below a target $\bar{A}$. Thus equilibrium is determined by the constrained monopolist problem below:

$$\max_{\{q_i\}} \sum_i q_i \cdot (AvgV_i(q_i) - P_i(q_i)) \tag{15}$$

$$\text{s.t.} \sum_i q_i \cdot AvgV_i(q_i) \geq \bar{A} \tag{16}$$

We then compute surplus by summing across markets consumer surplus and the financial provider's profits.

Welfare Comparisons The resulting equilibrium prices and quantities are shown in full in Table 8; for each case described above, quantities and prices are reported in subsequent rows. For cases 1 and 2 with informed financial providers, the outcomes of each separate market is reported in columns 3-12 corresponding to the interest rate group; the average market size-weighted average price and the total quantity are reported in the final two columns. In cases 3 and 4 with uninformed financial providers, only price and total quantity from the single aggregated market are reported in the final two columns. Finally, the market size weights used in the calculations of Equation 7, ω_i are reported in the last row.

The welfare of these markets is described in Table A12. Given the estimates used to discipline the model, the counterfactuals we calculate represent the weekly lending equillibrium for a prospective borrower. This welfare for a single borrower, in a single week, is reported in Yuan in column one of Table A12. To contextualize the welfare impact of different policy regimes: we scale this per capita quantity by the number of unique borrowers in our panel and then multiply by 52 weeks to get an annual welfare quantity.[16] This is then divided by the regional GDP, to give a percentage of GRP each policy represents, as reported in column 2.

[16]Our panel data includes 480 thousands of borrowers in Hubei. GRP of Hubei is 5,001.29 billion *yuan* in 2021, see data.stats.gov.cn/english.

Table 6: Average Parameter Values Comparison

	Sample Averages
$\square_Q$	14.071
Q_0	914.187
$\square_r$	-0.424
R_0	0.978
Avg Rate	16.657
Avg Implied Price	0.857

Note: This table presents the parameter used in the counterfactual analysis.

Table 7: Welfare Comparison: Compared to Optimal

	Consumer Surplus		Producer Surplus		Per Bororwer Outcomes		Total Surplus	
Case	Per Borrower	% Hubei GDP	Per Borrower	% Hubei GDP	Total Quantities	Average Price	% Opt	% Hubei GDP
Monopsony, Information	45.700	0.007	72.900	0.011	825.800	0.860	84.800	0.018
Competition, Information	130.800	0.020	0	0	2,204.530	0.920	93.524	0.020
Competition, No Information	129.300	0.020	0	0	2,139.890	0.920	92.451	0.020
Monopsony, No Information	47.800	0.007	68.100	0.010	793.620	0.860	82.870	0.018
Regulated Monopsony, Information	47.400	0.007	72.900	0.011	856.970	0.860	86.016	0.018

Note: This table compares welfare across the counterfactual scenarios. Columns 1 and 3 shows weekly consumer and producer surplus per borrower; for information cases, this is the sum of surpluses across the 9 interest rate markets, for no-information cases it is the surplus in the one combined market. Columns 2,4, and 9 scales the per-borrower surplus by the number of borrowers in the sample used for parameter estimation (152,819), and by 52 weeks; this quantity is then divided by Hubei's regional GDP. Columns 6 and 7 show the total weekly credit quantity and borrower-weighted average prices. Column 8 shows the total surplus as a proportion of the welfare produced in the optimal pricing scenario.

5.2.1 Monopoly vs Competition

A simple way to gain intuition on the welfare comparisons that we are going to compute is to think of each of the cases (monopoly, competition, with information, without information), through the lenses of a simple markup equation, that states that prices, in each of these situations, are a markup over marginal value.

$$p_i = \theta_i MV_i \tag{17}$$

Each model implies a different θ. When there is a single, monopolist lender, the markup $\theta_m = \frac{1}{1+\varepsilon_D^{-1}}$, where ε_D is the elasticity of demand for loans. When there is competition, equilibrium prices are described by Akerlof lemons condition, $p_i = AVG_i(p_i)$. This can be rewritten, similarly, as a markup over marginal cost where $\theta_c = \frac{1}{1+\varepsilon_{AVG}}$, where $\varepsilon_{AVG} = \frac{d \log AVG}{d \log Q}$ is the elasticity of the average value for the investors.

Thus, choosing between monopoly and competition, in this case, amounts, broadly speaking, to choosing between two wedges: if there is competition, the markup (or

markdown) is driven by selection. If there is a monopolist, the markup is driven by the elasticities of demand and how expanding or restricting the supply of credit affects the price that the monopolist can ask the borrowers to accept. Interestingly, the monopolist, when deciding which price to ask, internalizes the effects of selection, and its markup over marginal cost does not directly depend on selection effects and the elasticities of average value.

Which regime generates higher prices (the inverse of the interest rate) tells us which would be better for the borrowers. Which regime generates θ's that are closest to 1 tells us, broadly speaking, which of them gets prices that are closer to the prices that support first-best efficient allocations, where there would be no markups or markdowns.

5.2.2 Data-sharing versus less data-sharing

To analyze whether sharing data delivers a better outcome than sharing less of the data, we can consider the consequences of merging two signals that generate similar prices, as in Immorlica et al. (2022).

In a competitive regime, when two signals are combined, the resulting price for the combined pool of borrowers should be such that investors break even; that is, profits are zero. Letting $\pi_i(p)$ be the profits investors would enjoy at price p from signals i. If we combine one unit of a signal i with α units of a signal j, investors would break even only if the following equation holds:

$$\pi_i(p) + \alpha\pi_j(p) = 0 \tag{18}$$

Approximating each profit function $\pi_i(p)$ around its competitive price, we have that:

$$\pi_i(p) \approx \pi_i(p_i) + \pi'(p_i)(p - p_i) = \pi'(p_i)(p - p_i) \tag{19}$$

Where we used that profits are zero at its competitive equilibrium price ($\pi_i(p_i) = 0$). Fixing $p = \frac{p_i+p_j}{2}$, we conclude that, as a first-order approximation, for investors to break even at the new price, $\alpha = \frac{\pi_i'(p_i)}{\pi_j'(p_j)}$, i.e., signals should be combined in a ratio that is inversely proportional to the slope of its profit function.

Moreover, to a first-order approximation, for each signal, the change in consumer surplus is equal to the product of their demand at the original price and the change in prices. Assuming without loss that $p_i < p_j$, there is an overall increase in consumer surplus only if $Q_i(p_i) > \alpha Q_j(p_j)$. We conclude that the following conditions should hold (as in Proposition 3 from Immorlica et al. (2022)):

Proposition 1 (Immorlica et al. (2022))**.** *Suppose full disclosure equilibrium prices p_i are dense over an interval $I = (\underline{p}, \overline{p})$. Then if full disclosure is optimal, it must be the case that if*

$p_i < p_j$, $\frac{\pi_i'(p_i)}{Q_i(p_i)} < \frac{\pi_i'(p_i)}{Q_i(p_i)}$.

In contrast to the competitive case, in a monopolistic regime, when combining two signals, the resulting price should be such that the lender maximizes profits. Thus, instead of the levels of profits, the derivative of profits should be zero at the new prices (provided that profits are concave).

If we combine one unit of a signal i with α units of a signal j, investors would be maximizing profits only if the following equation holds:

$$\pi_i'(p) + \alpha \pi_j'(p) = 0 \tag{20}$$

In this case, we conclude that, as a first-order approximation, for the monopolist to be willing to accept the new price, $\alpha = \frac{\pi_i''(p_i)}{\pi_j''(p_j)}$, i.e., signals should be combined in a ratio that is inversely proportional to the second derivative of its profit function. Thus, by the same logic above, there should be a tight relationship between monopolistic prices and the second derivative of the profit function across different market segments.

Proposition 2. *Suppose full disclosure monopolistic prices p_i are dense over an interval $I = (\underline{p}, \overline{p})$. Then if full disclosure is optimal, it must be the case that if $p_i < p_j$, $\frac{\pi_i''(p_i)}{Q_i(p_i)} < \frac{\pi_i''(p_i)}{Q_i(p_i)}$.*

Proposition 1 and Proposition 2, as well as Figure 3 and Figure 4 can help us interpret the results from Table A12. Figure 3 shows that there is not a monotonic relationship between prices p_i^c and elasticities of the average value $\epsilon_{i,AVG}(p_i^c)$. In light of the results from Proposition 1, this means that when comparing pair by pair, we would find that, for some, but not all, pairs of signals, there would be welfare gains from disclosing the information of whether a borrower belongs to one interest rate group or another. This highlights that, under competition, simple information policies such as full disclosure or no disclosure would not be optimal. However, full disclosure still achieves a considerable fraction of the possible welfare gains, as seen in Table A12.

Similarly, Figure 4 shows that there is not a monotonic relationship between prices p_i^c and normalized second derivative of profits $\frac{\pi_i''(p_i^m)}{Q_i(p_i^m)}$. In light of the results from Proposition 2, this means that when comparing pair by pair, we would find that, for some, but not all, pairs of signals, there would be welfare gains from allowing the monopolist to discriminate between the borrowers from one interest rate group or another. This again highlights that, under monopoly, simple information policies such as full disclosure or no disclosure would not be optimal. Moreover, neither information policy alone is sufficient to undo the distortions caused by monopoly power, and it would be desirable to find an arrangement that increases the volume of loans, as shown in Table A12.

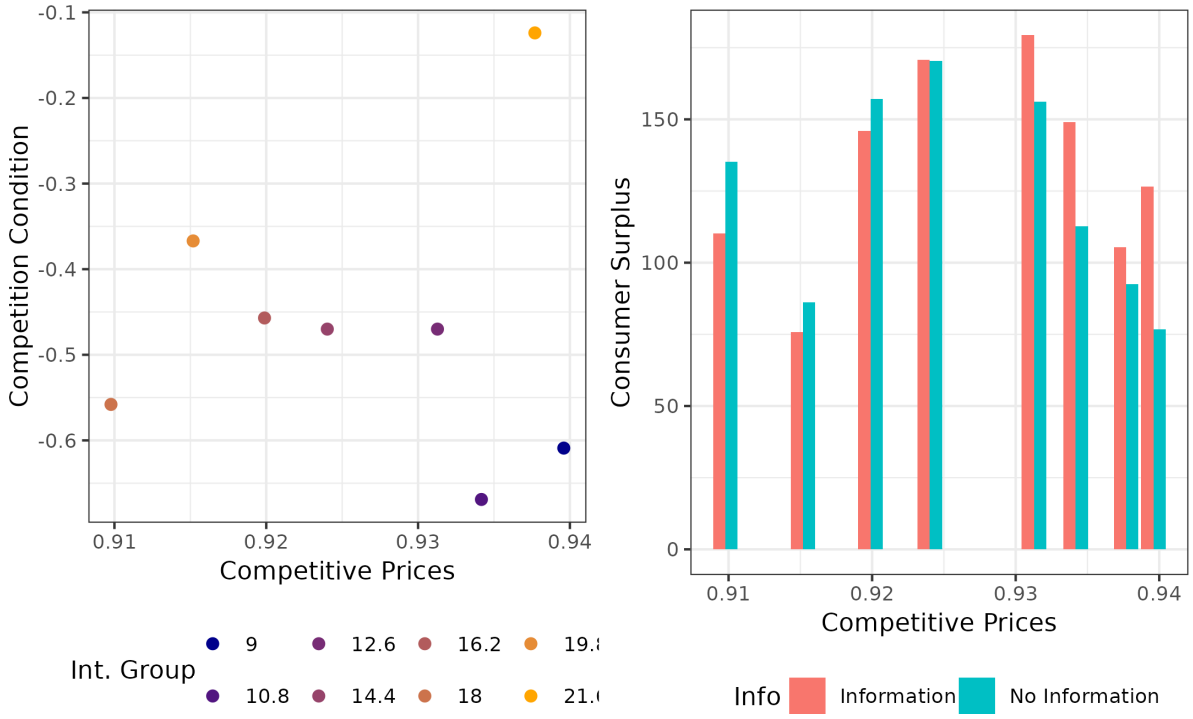


Figure 3: Competition and Information

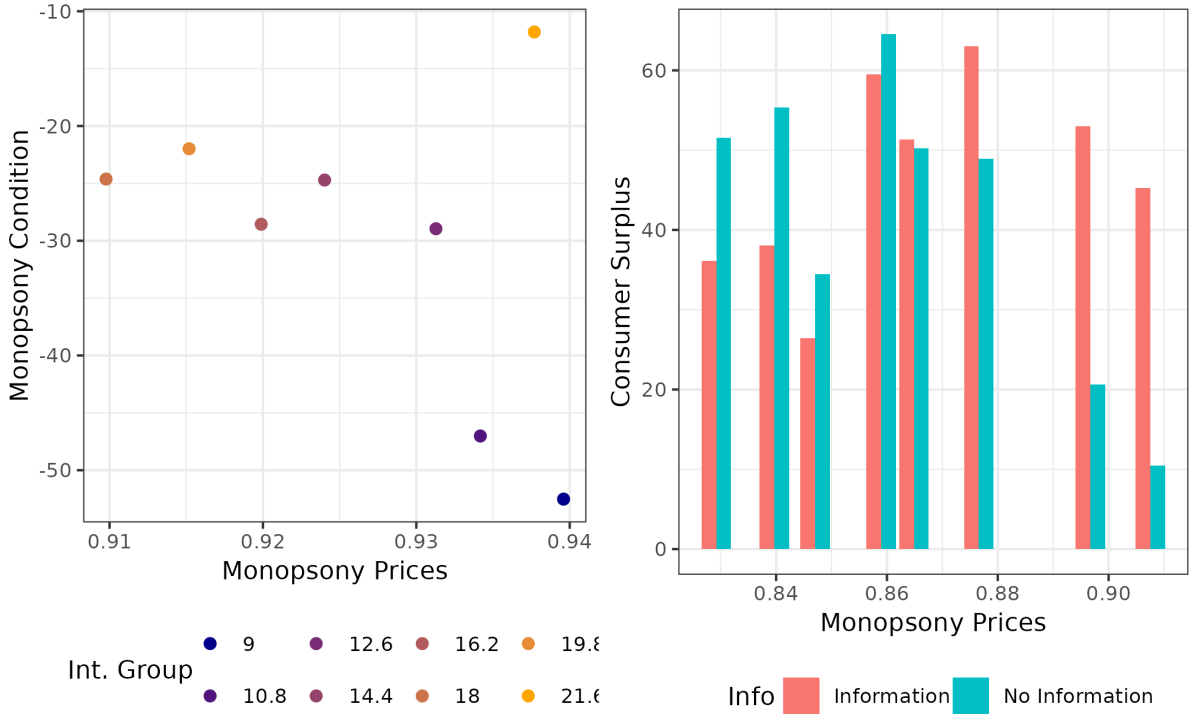


Figure 4: Monopolistand Information

Table 8: Weekly Borrowing Equillibria Comparison

Case	Outcome	9	10.8	12.6	14.4	16.2	18	19.8	21.6
Monop, Info	Q	33.8	104.5	141.5	162.3	170.7	83.2	68.1	61.5
Monop, Info	P	0.908	0.897	0.877	0.859	0.865	0.84	0.847	0.829
Monop, Info	R	98.8	98.9	98.7	98.5	97.5	98	97	98.1
Monop, Info	CS	1	3.7	7.2	9.7	8.7	5.3	4.2	5.9
Monop, Info	PS	1.8	6.6	11.5	15.8	13.9	9.3	6.4	7.6
Comp, Info	Q	91.4	282.2	378.9	433.3	456.6	222.6	180.9	158.8
Comp, Info	P	0.94	0.934	0.931	0.924	0.92	0.91	0.915	0.938
Comp, Info	R	96.8	96.2	95.9	95.2	94.7	93.7	94.3	96.6
Comp, Info	CS	2.9	10.4	20.4	27.7	24.8	15.3	12.2	17.1
Comp, Info	PS	0	0	0	0	0	0	0	0
Comp, No Info	Q	56.4	215.5	332.6	432.2	489	268.8	203.7	141.6
Comp, No Info	P	0.924	0.924	0.924	0.924	0.924	0.924	0.924	0.924
Comp, No Info	R	97.8	96.9	96.3	95.2	94.6	92.9	93.9	96.8
Comp, No Info	CS	1.8	7.9	17.8	27.7	26.6	18.7	13.8	15
Comp, No Info	PS	1.4	3.7	3.6	0.1	-2.8	-5.9	-2.4	2.2
Monop, No Info	Q	8.2	42.2	111.4	175.1	167.3	117.8	87.2	84.3
Monop, No Info	P	0.864	0.864	0.864	0.864	0.864	0.864	0.864	0.864
Monop, No Info	R	100	100	99.4	98.2	97.5	96.5	96.3	97.6
Monop, No Info	CS	0.2	1.4	5.6	10.5	8.5	7.7	5.5	8.4
Monop, No Info	PS	0.9	4.5	11.3	15.8	13.9	8.6	6.2	7
Reg Monop, Info	Q	36	109.9	147.1	167.6	177.3	85.6	70.4	63
Reg Monop, Info	P	0.91	0.899	0.879	0.861	0.867	0.842	0.849	0.832
Reg Monop, Info	R	98.7	98.8	98.6	98.4	97.3	97.9	96.9	98
Reg Monop, Info	CS	1.1	3.9	7.5	10	9.1	5.4	4.4	6
Reg Monop, Info	PS	1.8	6.6	11.5	15.8	13.9	9.3	6.4	7.6
Optimal	Q	62.3	187.9	273.9	312.9	332.1	154.3	137.7	142.1
Optimal	P	0.927	0.919	0.913	0.902	0.902	0.883	0.896	0.924
Optimal	R	97.6	97.3	96.8	96.3	95.6	95.3	95	96.8
Optimal	CS	2	6.8	14.5	19.5	17.7	10.3	9.1	15.1
Optimal	PS	1.3	4.9	7.4	10.2	8.8	6.5	3.7	2.1
Empirical Quantities		46.2	122.3	175	205.5	157.7	93.2	56.6	57.7
Empirical Prices		0.917	0.903	0.888	0.874	0.861	0.847	0.835	0.822
Empirical Repay		98.196	98.464	98.086	97.69	97.678	97.491	97.502	98.167
Empirical CS		1.4	4.4	9	12.4	8	6	3.5	5.5
Empirical PS		1.7	6.5	11.2	15.3	13.8	9.2	6.3	7.5
Group Shares		0.02	0.07	0.11	0.16	0.17	0.14	0.16	0.16

Note: This table shows the equilibrium outcomes produced across all 9 interest rate groups for each counterfactual situation. For each situation, as well as outcomes under optimal pricing and empirical outcomes, quantities (Q), prices (P), repayment rates (R), consumer surplus (CS), and producer surplus (PS) are reported.

5.2.3 No data-sharing versus more data-sharing

To analyze whether not sharing data delivers a better outcome than sharing more of the data, we can consider the consequences of picking pairs of signals that are merged together in a single message and creating new bundles out of those, as in Immorlica et al. (2022).

In a competitive environment, the resulting price for each bundle we create should be such that investors break even. One way to achieve that is to assign a pair of signals i and j to a price p_m, another pair of i' and j' to a price $p_{m'}$, respectively in proportions $\alpha_i, \alpha_j, \alpha_{i'}, \alpha_{j'}$ where:

$$\alpha_i = \frac{\frac{1}{\pi_i(p_m)}}{\frac{\pi_j(p_0)}{\pi_j(p_m)} - \frac{\pi_i(p_0)}{\pi_i(p_m)}} \quad \alpha_j = \frac{-\frac{1}{\pi_j(p_m)}}{\frac{\pi_j(p_0)}{\pi_j(p_m)} - \frac{\pi_i(p_0)}{\pi_i(p_m)}} \tag{21}$$

$$\alpha_{i'} = \frac{\frac{1}{\pi_{i'}(p_{m'})}}{\frac{\pi_{i'}(p_0)}{\pi_{i'}(p_{m'})} - \frac{\pi_{j'}(p_0)}{\pi_{j'}(p_{m'})}} \quad \alpha_{j'} = \frac{-\frac{1}{\pi_{j'}(p_{m'})}}{\frac{\pi_{i'}(p_0)}{\pi_{i'}(p_{m'})} - \frac{\pi_{j'}(p_0)}{\pi_{j'}(p_{m'})}} \tag{22}$$

And we restrict our choice of pairs of signals to those for which $\alpha_i, \alpha_j, \alpha_{i'}$, and $\alpha_{j'}$, as defined above, are positive. That set of α's guarantees that investors would break even at each of the bundles in the market, at the prices p_m, $p_{m'}$, and p_0.

Define $\sigma_i(p)$ as the borrower's surplus per unit of signal i at price p. This deviation generates a change in surplus equal to:

$$\frac{\left(\frac{\sigma_i(p_m)-\sigma_i(p_0)}{\pi_i(p_m)}\right) - \left(\frac{\sigma_j(p_m)-\sigma_j(p_0)}{\pi_j(p_m)}\right)}{\frac{\pi_j(p_0)}{\pi_j(p_m)} - \frac{\pi_i(p_0)}{\pi_i(p_m)}} - \frac{\left(\frac{\sigma_{i'}(p_{m'})-\sigma_{i'}(p_0)}{\pi_{i'}(p_{m'})}\right) - \left(\frac{\sigma_{j'}(p_{m'})-\sigma_{j'}(p_0)}{\pi_{j'}(p_{m'})}\right)}{\frac{\pi_{j'}(p_0)}{\pi_{j'}(p_{m'})} - \frac{\pi_{i'}(p_0)}{\pi_{i'}(p_{m'})}} \tag{23}$$

In the limit where $p_m \downarrow p_0$, and $p_{m'} \uparrow p_0$, the change in surplus is proportional to

$$\frac{\frac{\sigma_i'(p_0)}{\pi_i(p_0)} - \frac{\sigma_j'(p_0)}{\pi_j(p_0)}}{\frac{\pi_i'(p_0)}{\pi_i(p_0)} - \frac{\pi_j'(p_0)}{\pi_j(p_0)}} - \frac{\frac{\sigma_{i'}'(p_0)}{\pi_{i'}(p_0)} - \frac{\sigma_{j'}'(p_0)}{\pi_{j'}(p_0)}}{\frac{\pi_{i'}'(p_0)}{\pi_{i'}(p_0)} - \frac{\pi_{j'}'(p_0)}{\pi_{j'}(p_0)}} \tag{24}$$

Whereas, if we assign i' and j' to p_m and i and j to $p_{m'}$, in the limit where $p_m \downarrow p_0$, and $p_{m'} \uparrow p_0$, we will conclude that the change in surplus is proportional to the negative of the expression above. Thus, if the original bundled was designed optimally it needs to be the case that:

$$\frac{\frac{\sigma_i'(p_0)}{\pi_i'(p_0)} - \frac{\sigma_j'(p_0)}{\pi_j(p_0)}}{\frac{\pi_i'(p_0)}{\pi_i(p_0)} - \frac{\pi_j'(p_0)}{\pi_j(p_0)}} = \frac{\frac{\sigma_{i'}'(p_0)}{\pi_{i'}(p_0)} - \frac{\sigma_{j'}'(p_0)}{\pi_{j'}(p_0)}}{\frac{\pi_{i'}'(p_0)}{\pi_{i'}(p_0)} - \frac{\pi_{j'}'(p_0)}{\pi_{j'}(p_0)}} = \gamma(p_0) \tag{25}$$

for some constant $\gamma(p_0)$. Noticing that $\sigma_i'(p) = Q_i(p)$, Equation 25 can be easily rearranged to show that there should be an affine relationship between inverse markups $m_i(p_0) = p_0/(AVG_i(p_0) - p_0)$ and elasticities of profits $\epsilon_{\pi_i,p}(p_0) = \frac{d\log \pi_i(p_0)}{d\log p}$ as in Immorlica et al. (2022).

Proposition 3 (Immorlica et al. (2022)). *Under competition, for any surplus maximizing*

information disclosure policy, all signals that are pooled together at p_0 must satisfy and affine relationship:

$$m_i(p_0) = \kappa(p_0) + \gamma(p_0)\epsilon_{\pi_i,p}(p_0) \tag{26}$$

for some constants $\kappa(p_0)$ and $\gamma(p_0)$, and where $m_i(p_0) = p_0/(AVG_i(p_0) - p_0)$ and $\epsilon_{\pi_i,p}(p_0) = \frac{d\log \pi_i(p_0)}{d\log p}$.

The same argument above can be modified to accommodate a monopolist lender. The key fact to note is that, whereas in the competitive case investors must break-even and thus expected profits should be zero, in the monopolist case the lender must be setting prices optimally so the expected derivative of profits should be zero at each bundle, as in Farboodi et al. (2025). To achieve that we can assign a pair of signals i and j to a price p_m, another pair of i' and j' to a price $p_{m'}$, respectively in proportions $\tilde{\alpha}_i$, $\tilde{\alpha}_j$, $\tilde{\alpha}_{i'}$, $\tilde{\alpha}_{j'}$ where:

$$\tilde{\alpha}_i = \frac{\frac{1}{\pi_i'(p_m)}}{\frac{\pi_j'(p_0)}{\pi_j'(p_m)} - \frac{\pi_i'(p_0)}{\pi_i'(p_m)}} \quad \tilde{\alpha}_j = \frac{-\frac{1}{\pi_j'(p_m)}}{\frac{\pi_j'(p_0)}{\pi_j'(p_m)} - \frac{\pi_i'(p_0)}{\pi_i'(p_m)}} \tag{27}$$

$$\tilde{\alpha}_{i'} = \frac{\frac{1}{\pi_{i'}'(p_{m'})}}{\frac{\pi_{i'}'(p_0)}{\pi_{i'}'(p_{m'})} - \frac{\pi_{j'}'(p_0)}{\pi_{j'}'(p_{m'})}} \quad \tilde{\alpha}_{j'} = \frac{-\frac{1}{\pi_{j'}'(p_{m'})}}{\frac{\pi_{i'}'(p_0)}{\pi_{i'}'(p_{m'})} - \frac{\pi_{j'}'(p_0)}{\pi_{j'}'(p_{m'})}} \tag{28}$$

Thus, following analogous steps that go into showing that there should be an affine relationship between inverse markups and elasticities of profits, we can show that in this case there should be an affine relationship between a modified inverse markup $\tilde{m}_i(p_0) = \frac{Q_i(p_0)p_0}{\pi_i'(p_0)}$, and the elasticity of the derivative of profits $\tilde{\epsilon}_{\pi_i',p}(p_0) = \frac{d\log \pi_i'(p_0)}{d\log p}$.

Proposition 4. *With a monopolist lender, for any surplus maximizing information disclosure policy, all signals that are pooled together at p_0 must satisfy and affine relationship:*

$$\tilde{m}_i(p_0) = \tilde{\kappa}(p_0) + \tilde{\gamma}(p_0)\tilde{\epsilon}_{\pi_i',p}(p_0) \tag{29}$$

for some constants $\tilde{\kappa}(p_0)$ and $\tilde{\gamma}(p_0)$, and where $\tilde{m}_i(p_0) = Q_i(p_0)p_0/\pi_i'(p_0)$ and $\tilde{\epsilon}_{\pi_i',p}(p_0) = \frac{d\log \pi_i'(p_0)}{d\log p}$.

Violations of the affine relationships above imply that there would be first-order gains from creating new sub-bundles from an existing bundle of signals. Those conditions are tested in the data in Appendix Figure A10 and Appendix Figure A11. In both cases, there is close to an affine relationship between the inverse markups and elasticities, suggesting some surplus gains from moving away from non-disclosure, especially

in the monopolist case. However, as shown in Table A12, these gains are much smaller than what would be achieved by moving to a competitive arrangement or by regulating quantities.

6 Conclusion

In this paper, we study an innovative fintech lending market by using this cash flow information to provide unsecured small business loans to 40 million merchants. In the history of microcredit, this could be one of the largest and most commercially sustainable financial innovations compared to many microfinance institutions (MFIs). We utilize a policy that provides an 20% discount on interest rates for borrowers in some areas in a certain period. With rich data, the fintech lender can identify borrower's risk with a precise interest rate ranging from 5.4% to 21.6%. The findings reveal a decrease in the average probability of loan repayment following interest rate reductions, indicating advantageous selection.

Building on our estimated demand elasticities and selection effects, we assess the welfare implications of information-sharing and pro-competition policies, finding that among a restricted simple set of policies, ensuring competition and mandating information to be shared across financial providers results in the largest welfare gains. This work sheds light on the policy debates on open banking and offers evidence for designing competitive policies among Fintech lenders and traditional financial institutions.

References

Sumit Agarwal, Shashwat Alok, Pulak Ghosh, and Sudip Gupta. Fintech and Credit Scoring for the Millennials: Evidence using Mobile and Social Footprints. *SSRN Electronic Journal*, 2019. ISSN 1556-5068. doi: 10.2139/ssrn.3507827. URL https://www.ssrn.com/abstract=3507827.

Sumit Agarwal, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. Searching for Approval. *Econometrica*, 92(4):1195–1231, 2024. ISSN 0012-9682. doi: 10.3982/ECTA18554. URL https://www.econometricsociety.org/doi/10.3982/ECTA18554.

George A. Akerlof. The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488, August 1970. ISSN 00335533. doi: 10.2307/1879431. URL https://academic.oup.com/qje/article-lookup/doi/10.2307/1879431.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2), April 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1709. URL https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-2/Generalized-random-forests/10.1214/18-AOS1709.full.

José Azar and Ioana Marinescu. Monopsony Power in the Labor Market: From Theory to Policy. *Annual Review of Economics*, 16(1):491–518, 2024. Publisher: Annual Reviews.

Ihsaan Bassier, Arindrajit Dube, and Suresh Naidu. Monopsony in movers: The elasticity of labor supply to firm wage policies. *Journal of Human Resources*, 57(S):S50–s86, 2022. Publisher: University of Wisconsin Press.

Tobias Berg, Valentin Burg, Ana Vanjak, and Manju Puri. On the rise of FinTechs: Credit Scoring using Digital Footprints. *Review of Financial Studies*, 33(7):2845–2897, 2020.

Tobias Berg, Andreas Fuster, and Manju Puri. FinTech Lending. *Annual Review of Financial Economics*, 14(1):187–207, November 2022. ISSN 1941-1367, 1941-1375. doi: 10.1146/annurev-financial-101521-112042. URL https://www.annualreviews.org/doi/10.1146/annurev-financial-101521-112042.

Dirk Bergemann, Benjamin Brooks, and Stephen Morris. The Limits of Price Discrimination. *American Economic Review*, 105(3):921–57, March 2015. doi: 10.1257/aer.20130848. URL https://www.aeaweb.org/articles?id=10.1257/aer.20130848.

Laura Blattner, Jacob Hartwig, and Scott Nelson. Information design in consumer credit markets. 2022.

Luca Braghieri. Targeted Advertising and Price Discrimination in Intermediated Online Markets. *SSRN Electronic Journal*, 2019. ISSN 1556-5068. doi: 10.2139/ssrn.3072692. URL https://www.ssrn.com/abstract=3072692.

Thorsten Broecker. Credit-Worthiness Tests and Interbank Competition. *Econometrica*, 58(2):429, March 1990. ISSN 00129682. doi: 10.2307/2938210. URL https://www.jstor.org/stable/2938210?origin=crossref.

Marika Cabral, Can Cui, and Michael Dworsky. The Demand for Insurance and Rationale for a Mandate: Evidence from Workers' Compensation Insurance. *American Economic Review*, 112(5):1621–1668, May 2022. ISSN 0002-8282. doi: 10.1257/aer.20190261. URL https://pubs.aeaweb.org/doi/10.1257/aer.20190261.

Alma Cohen and Liran Einav. Estimating Risk Preferences from Deductible Choice. *American Economic Review*, 97(3):745–788, May 2007. ISSN 0002-8282. doi: 10.1257/aer.97.3.745. URL https://pubs.aeaweb.org/doi/10.1257/aer.97.3.745.

Benjamin Collier and Cameron Ellis. A Demand Curve for Disaster Recovery Loans. *Econometrica*, 92(3):713–748, 2024. ISSN 0012-9682. doi: 10.3982/ECTA20417. URL https://www.econometricsociety.org/doi/10.3982/ECTA20417.

Rahul Deb, Mallesh Pai, and Anne-Katrin Roesler. How Informed Do You Want Your Principal to Be? Technical report, 2024.

Anthony A DeFusco, Huan Tang, and Constantine Yannelis. Measuring the welfare cost of asymmetric information in consumer credit markets. *Journal of Financial Economics*, 146(3):821–840, 2022. Publisher: Elsevier.

Will Dobbie and Paige Marta Skiba. Information Asymmetries in Consumer Credit Markets: Evidence from Payday Lending. *American Economic Journal: Applied Economics*, 5(4):256–282, October 2013. ISSN 1945-7782, 1945-7790. doi: 10.1257/app.5.4.256. URL https://pubs.aeaweb.org/doi/10.1257/app.5.4.256.

Liran Einav, Amy Finkelstein, and Mark R Cullen. Estimating welfare in insurance markets using variation in prices. *The quarterly journal of economics*, 125(3):877–921, 2010. Publisher: MIT Press.

Liran Einav, Amy Finkelstein, and Neale Mahoney. The IO of selection markets. In *Handbook of Industrial Organization*, volume 5, pages 389–426. Elsevier, 2021. ISBN 978-0-323-98887-2. doi: 10.1016/bs.hesind.2021.11.014. URL https://linkinghub.elsevier.com/retrieve/pii/S1573448X21000145.

Maryam Farboodi, Nima Haghpanah, and Ali Shourideh. Good Data and Bad Data: The Welfare Effects of Price Discrimination, February 2025. URL http://arxiv.org/abs/2502.03641. arXiv:2502.03641 [econ].

Amy Finkelstein and Matthew J Notowidigdo. Take-Up and Targeting: Experimental Evidence from SNAP. *The Quarterly Journal of Economics*, 134(3):1505–1556, August 2019. ISSN 0033-5533, 1531-4650. doi: 10.1093/qje/qjz013. URL https://academic.oup.com/qje/article/134/3/1505/5484907.

Amy Finkelstein and James Poterba. Adverse Selection in Insurance Markets: Policyholder Evidence from the U.K. Annuity Market. *Journal of Political Economy*, 112 (1):183–208, February 2004. ISSN 0022-3808, 1537-534X. doi: 10.1086/379936. URL https://www.journals.uchicago.edu/doi/10.1086/379936.

Leonardo Gambacorta, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen. Data versus Collateral. *Review of Finance*, 27(2):369–398, March 2023. ISSN 1572-3097, 1573-692X. doi: 10.1093/rof/rfac022. URL https://academic.oup.com/rof/article/27/2/369/6569096.

Pulak Ghosh, Boris Vallee, and Yao Zeng. FinTech lending and cashless payments. *Journal of Finance*, forthcoming, 2023.

Feng Guo, Yiping Huang, Jingyi Wang, and Xue Wang. The informal economy at times of COVID-19 pandemic. *China Economic Review*, 71:101722, 2022. Publisher: Elsevier.

Nima Haghpanah and Ron Siegel. The Limits of Multiproduct Price Discrimination. *American Economic Review: Insights*, 4(4):443–458, December 2022. ISSN 2640-205X, 2640-2068. doi: 10.1257/aeri.20210426. URL https://pubs.aeaweb.org/doi/10.1257/aeri.20210426.

Nima Haghpanah and Ron Siegel. Pareto-Improving Segmentation of Multiproduct Markets. *Journal of Political Economy*, 131(6):1546–1575, June 2023. ISSN 0022-3808, 1537-534X. doi: 10.1086/722932. URL https://www.journals.uchicago.edu/doi/10.1086/722932.

Benjamin R. Handel and Jonathan T. Kolstad. Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare. *American Economic Review*, 105(8):2449–2500, August 2015. ISSN 0002-8282. doi: 10.1257/aer.20131126. URL https://pubs.aeaweb.org/doi/10.1257/aer.20131126.

Zhiguo He, Jing Huang, and Jidong Zhou. Open banking: Credit market competition when borrowers own the data. *Journal of Financial Economics*, 147(2):449–474, 2023. ISSN 0304-405X. doi: https://doi.org/10.1016/j.jfineco.2022.12.003. URL https://www.sciencedirect.com/science/article/pii/S0304405X22002471.

Zhiguo He, Jing Huang, and Cecilia Parlatore. Information Span in Credit Market Competition. 2024.

Nathaniel Hendren. Private Information and Insurance Rejections. *Econometrica*, 81 (5):1713–1762, 2013. ISSN 0012-9682. doi: 10.3982/ECTA10931. URL http://doi.wiley.com/10.3982/ECTA10931.

Nathaniel Hendren. Knowledge of Future Job Loss and Implications for Unemployment Insurance. *American Economic Review*, 107(7):1778–1823, July 2017. ISSN 0002-8282. doi: 10.1257/aer.20151655. URL https://pubs.aeaweb.org/doi/10.1257/aer.20151655.

Sinem Hidir and Nikhil Vellodi. Privacy, Personalization, and Price Discrimination. *Journal of the European Economic Association*, 19(2):1342–1363, April 2021. ISSN 1542-4766, 1542-4774. doi: 10.1093/jeea/jvaa027. URL https://academic.oup.com/jeea/article/19/2/1342/5851746.

Shota Ichihashi. Online Privacy and Information Disclosure by Consumers. *American Economic Review*, 110(2):569–595, February 2020. ISSN 0002-8282. doi: 10.1257/aer.20181052. URL https://pubs.aeaweb.org/doi/10.1257/aer.20181052.

Nicole Immorlica, Andre M Sztutman, and Robert M Townsend. Optimal Credit Scores Under Adverse Selection. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 737–738, 2022.

Julapa Jagtiani and Catharine Lemieux. The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform. *Financial Management*, 48(4):1009–1029, 2019. Publisher: Wiley Online Library.

Mark Jansen, Fabian Nagel, Constantine Yannelis, and Anthony Lee Zhang. Data and Welfare in Credit Markets. Technical Report w30235, National Bureau of Economic Research, Cambridge, MA, July 2022. URL http://www.nber.org/papers/w30235.pdf.

Emir Kamenica and Matthew Gentzkow. Bayesian Persuasion. *American Economic Review*, 101(6):2590–2615, October 2011. ISSN 0002-8282. doi: 10.1257/aer.101.6.2590. URL https://pubs.aeaweb.org/doi/10.1257/aer.101.6.2590.

Dean Karlan and Jonathan Zinman. Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment. *Econometrica*, 77(6): 1993–2008, 2009. ISSN 0012-9682. doi: 10.3982/ECTA5781. URL http://doi.wiley.com/10.3982/ECTA5781.

Anton Kolotilin. Optimal information disclosure: A linear programming approach: Optimal information disclosure. *Theoretical Economics*, 13(2):607–635, May 2018. ISSN 19336837. doi: 10.3982/TE1805. URL http://doi.wiley.com/10.3982/TE1805.

Anton Kolotilin, Roberto Corrao, and Alexander Wolitzky. Persuasion and Matching: Optimal Productive Transport. *Journal of Political Economy*, 133(4):1334–1381, April 2025. ISSN 0022-3808, 1537-534X. doi: 10.1086/734095. URL https://www.journals.uchicago.edu/doi/10.1086/734095.

Naz Koont. The Digital Banking Revolution: Effects on Competition and Stability. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4624751. URL https://www.ssrn.com/abstract=4624751.

Pablo Kurlat and Johannes Stroebel. Testing for Information Asymmetries in Real Estate Markets. *Review of Financial Studies*, 28(8):2429–2461, August 2015. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhv028. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhv028.

Jonathan Levin. Information and the Market for Lemons. *RAND Journal of Economics*, pages 657–666, 2001. Publisher: JSTOR.

Jian Li and Stefano Pegoraro. Borrowing from a Bigtech Platform. *SSRN Electronic Journal*, 2022. ISSN 1556-5068. doi: 10.2139/ssrn.4206016. URL https://www.ssrn.com/abstract=4206016.

Lei Liu, Guangli Lu, and Wei Xiong. The Big Tech Lending Model. Technical Report w30160, National Bureau of Economic Research, Cambridge, MA, June 2022. URL http://www.nber.org/papers/w30160.pdf.

Neale Mahoney and E Glen Weyl. Imperfect competition in selection markets. *Review of Economics and Statistics*, 99(4):637–651, 2017. Publisher: MIT Press.

Rachel J Nam. Open banking and customer data sharing: Implications for fintech borrowers. *SAFE Working Paper*, 2023.

Scott Nelson. Private information and price regulation in the US credit card market. *Unpublished working paper*, 2023.

P. M. Robinson. Root-N-Consistent Semiparametric Regression. *Econometrica*, 56(4): 931, July 1988. ISSN 00129682. doi: 10.2307/1912705. URL https://www.jstor.org/stable/1912705?origin=crossref.

Sherrill Shaffer. The Winner's Curse in Banking. *Journal of Financial Intermediation*, 7 (4):359–392, October 1998. ISSN 10429573. doi: 10.1006/jfin.1998.0251. URL https://linkinghub.elsevier.com/retrieve/pii/S1042957398902510.

Anna Sokolova and Todd Sorensen. Monopsony in Labor Markets: A Meta-Analysis. *ILR Review*, 74(1):27–55, January 2021. ISSN 0019-7939, 2162-271X. doi: 10.1177/0019793920965562. URL https://journals.sagepub.com/doi/10.1177/0019793920965562.

Xavier Vives. Digital Disruption in Banking. *Annual Review of Financial Economics*, 11(1):243–272, December 2019. ISSN 1941-1367, 1941-1375. doi: 10.1146/annurev-financial-100719-120854. URL https://www.annualreviews.org/doi/10.1146/annurev-financial-100719-120854.

Xavier Vives and Zhiqiang Ye. Fintech Entry, Lending Market Competition, and Welfare. *Working Paper*, 2024.

A Additional Tables

A.1 Summary Statistics

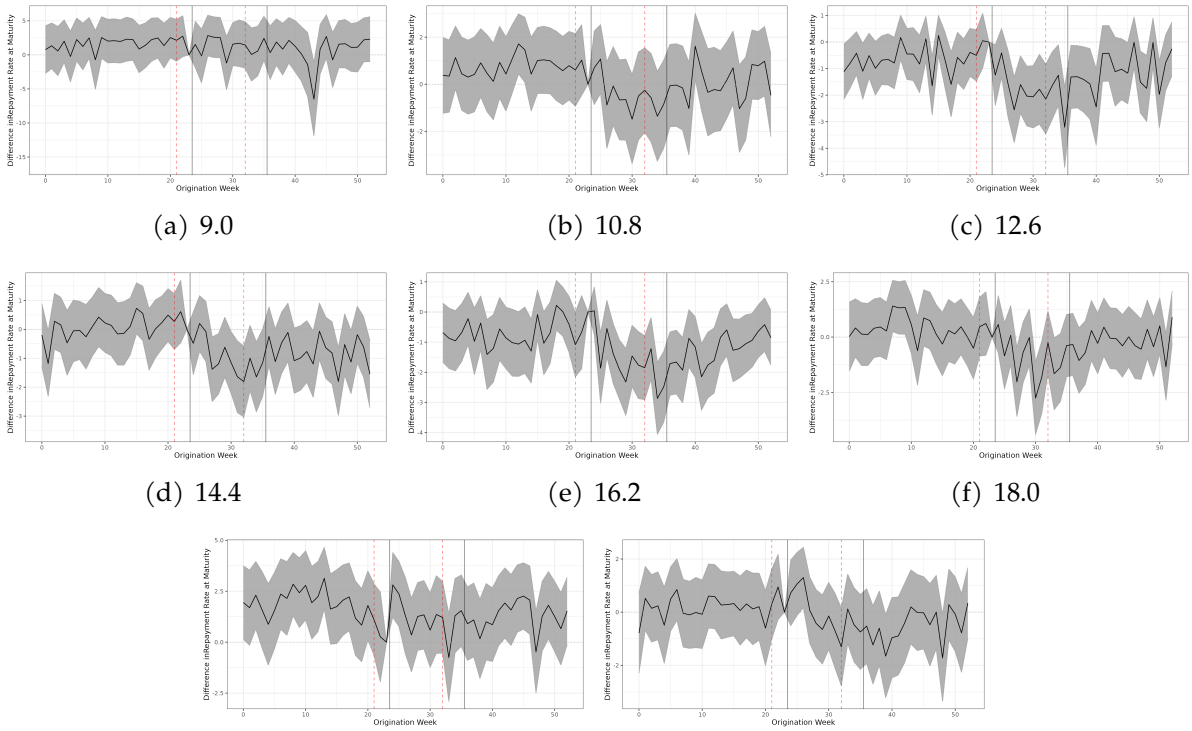
Table A1: Summary Statistics

I. Loan-Level Summaries						
		Pre-Policy			**F**ull Sample	
Variable	Control	Treatment	Difference	Control	Treatment	Difference
Repayment Rate At Maturity	98.55	98.43	0.11	97.93	97.45	0.47
	(0.0212)	(0.0313)	(0.0373)	(0.0131)	(0.0204)	(0.0234)
Share of 12-mo Terms	64.26	63.69	0.56	66.78	66.64	0.14
	(0.1011)	(0.1438)	(0.1756)	(0.0544)	(0.0744)	(0.0921)
Share of Loans Repaid in Full	97.49	97.32	0.17	93.59	93.11	0.48
	(0.033)	(0.0483)	(0.0579)	(0.0283)	(0.04)	(0.0485)
Share of Loans Repaid on Origination Day	7.19	6.76	0.43	6.42	6.15	0.27
	(0.0545)	(0.0751)	(0.0936)	(0.0283)	(0.0379)	(0.0476)
Share of Loans In 1st Ring	46.01	51.84	-5.84	45.88	51.14	-5.26
	(0.1051)	(0.1494)	(0.1828)	(0.0575)	(0.0789)	(0.0976)
Observations	224876	111809		750646	401872	
II. Panel-Level Summaries						
		Pre-Policy			**F**ull Sample	
Variable	Control	Treatment	Difference	Control	Treatment	Difference
New Loan Amount	868.65	984.75	-116.1	1108.88	1161.48	-52.6
	(3.1357)	(3.2937)	(4.5478)	(3.791)	(3.5716)	(5.2084)
Number of New Loans	0.1	0.1	0	0.12	0.13	-0.01
	(2e-04)	(2e-04)	(3e-04)	(2e-04)	(3e-04)	(4e-04)
Has New Loan	7	7.18	-0.18	8.59	8.73	-0.14
	(0.0114)	(0.0115)	(0.0097)	(0.0125)	(0.0126)	(0.0178)
Average New Loan Amount	9858.97	10737.68	-878.71	10019.2	10176.02	-156.82
	(8.6921)	(8.4368)	(12.113)	(9.238)	(8.3069)	(12.4236)
Outstanding Balance	172.59	182.91	-10.32	163.15	127.08	36.07
	(3.2115)	(3.2478)	(4.5675)	(3.8328)	(3.4765)	(5.1746)
Repayment Amount	696.06	801.84	-105.78	945.73	1034.4	-88.67
	(2.5668)	(2.7327)	(3.7494)	(3.107)	(3.1366)	(4.4149)
Line of Credit Amount	38768.48	40771.68	-2003.19	41104.96	42437.95	-1332.98
	(23.2255)	(21.2596)	(31.4845)	(24.9497)	(21.4882)	(32.9277)
Observations	1687393	1689406		4994069	5001250	

This table shows summary statistics for loan and panel variables for control and treatment groups. The top Panel I reports loan variable summaries, and the bottom Panel II reports panel variable summaries. Both panels report sample averages and standard errors (in parentheses) for control and treatment observations; both panels also report statistics for pre-policy (week 0 to week 23), and the full-sample period (week 0 to week 70). Sample differences, with pooled standard errors, are also reported.

A.2 Dynamic Difference-in-Differences: Interest Rate Groups

Figure A1: Dynamic DiD Estimates by Interest Rate Group: Repayment Rates



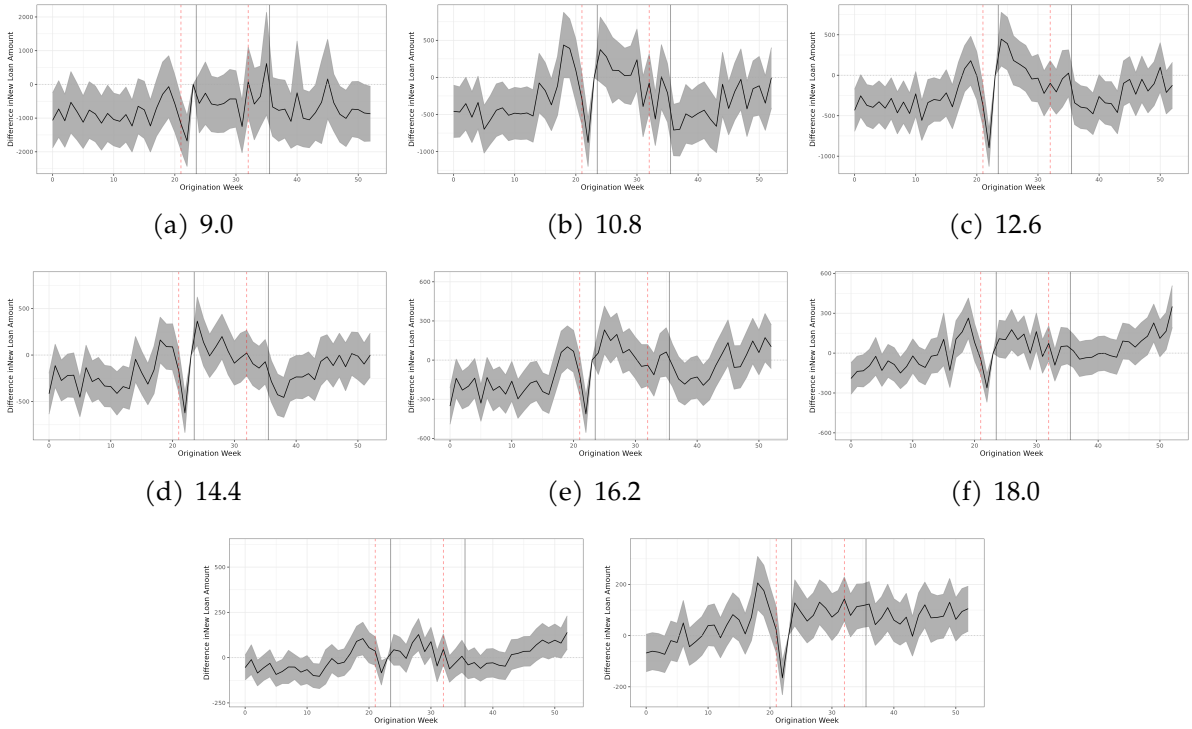
(a) 9.0 (b) 10.8 (c) 12.6

(d) 14.4 (e) 16.2 (f) 18.0

(g) 19.8 (h) 21.6

Note: Figure shows dynamic difference-in-differences estimates for repayment rates. Week 0 corresponds to the week of September 1st, 2019. Policy starts week 24 and persists until week 36 (shown by the vertical bars). Important Covid events are shown by the dotted red lines: line 1 shows the start of the Wuhan lockdown (January 23rd 2020), line 2 shows the end of the lockdown (April 8th, 2020)

Figure A2: Dynamic DiD Estimates by Interest Rate Group: New Loan Amount



(a) 9.0 (b) 10.8 (c) 12.6

(d) 14.4 (e) 16.2 (f) 18.0

(g) 19.8 (h) 21.6

Note: Figure shows dynamic difference-in-differences estimates for new loan amounts. Week 0 corresponds to the week of September 1st, 2019. Policy starts week 24 and persists until week 36 (shown by the vertical bars). Important Covid events are shown by the dotted red lines: line 1 shows the start of the Wuhan lockdown (January 23rd 2020), line 2 shows the end of the lockdown (April 8th, 2020)

A.3 Industry Composition

Table A2: Top 20 Industry Composition: Pre-Policy

	Proportion of Loans			**P**roportion of Borrowers		
Industry	Control	Treatment	Difference	Control	Treatment	Difference
Non-Classified	17.66	18.7	-1.04 (0.09)	27.31	28.04	-0.72 (0.03)
Catering	13.64	16.39	-2.75 (0.08)	13.13	12.97	0.16 (0.02)
Grocery	15.41	14.4	1.01 (0.08)	13.43	13.31	0.12 (0.02)
Apparel	11.27	9.98	1.29 (0.07)	9.49	9.39	0.1 (0.02)
Home Decor	7.05	6.5	0.56 (0.06)	6.12	6.08	0.04 (0.02)
Beauty Services	3	3.32	-0.32 (0.04)	2.81	2.79	0.03 (0.01)
Auto Parts	2.66	2.74	-0.08 (0.04)	2.38	2.35	0.03 (0.01)
Internet Services	2.59	2.33	0.26 (0.04)	2.11	2.11	0 (0.01)
Agriculture	1	1.57	-0.57 (0.02)	1.12	1.11	0.01 (0.01)
Accommodation	1.25	1.44	-0.19 (0.03)	1.19	1.18	0.02 (0.01)
Courier Services	1.55	1.41	0.13 (0.03)	1.26	1.24	0.02 (0.01)
Advertising	1.35	1.4	-0.05 (0.03)	1.13	1.12	0.01 (0.01)
Medical Services	1.36	1.38	-0.03 (0.03)	1.16	1.15	0.01 (0.01)
Mobile Sales	1.42	1.22	0.2 (0.03)	1.07	1.06	0.01 (0.01)
Logistics	1.05	1.12	-0.07 (0.02)	0.93	0.91	0.01 (0.01)
Electrochemical Products	1.29	0.95	0.34 (0.02)	1	1	0.01 (0.01)
Home Appliances	0.89	0.85	0.04 (0.02)	0.73	0.73	0 (0.01)
Childcare	1.04	0.83	0.21 (0.02)	0.85	0.83	0.01 (0.01)
Renovation	0.9	0.76	0.14 (0.02)	0.77	0.76	0.01 (0.01)
Fruit Store	0.61	0.68	-0.07 (0.02)	0.6	0.59	0 (0)

This table shows the proportion of observations in each industry classification. The top 20 industries by share are shown for the pre-policy period (week 0 to week 23). For both the proportion of loans, and proportion of panel observations (borrowers x weeks), the industry share for control, treatment, and their differences (control - treatment) are reported; pooled standard errors are shown in parentheses alongside the differences.

Table A3: Interest Rate Composition: Pre-Policy

	Proportion of Loans			Proportion of Borrowers		
Interest Rate	Control	Treatment	Difference	Control	Treatment	Difference
9	3.41	3.23	0.18 (0.04)	2.46	3.54	-1.08 (0.01)
10.8	8.57	8.37	0.2 (0.06)	7	8.34	-1.35 (0.02)
12.6	16.83	17.1	-0.27 (0.09)	11.54	13.58	-2.04 (0.02)
14.4	19.97	19.47	0.5 (0.09)	15.82	17.52	-1.71 (0.02)
16.2	16.26	16.61	-0.35 (0.08)	16.96	16.74	0.22 (0.02)
18	11.06	11.27	-0.21 (0.07)	12.04	12.05	-0.01 (0.02)
19.8	6.35	6.53	-0.19 (0.06)	14.39	11.69	2.7 (0.02)
21.6	11.24	11.89	-0.65 (0.07)	14.43	10.97	3.46 (0.02)

This table shows the proportion of observations in each of the selected interest rate groups. Shares are shown for the pre-policy period (week 0 to week 23). For both the proportion of loans, and proportion of panel observations (borrowers x weeks), the industry share for control, treatment, and their differences (control - treatment) are reported; pooled standard errors are shown in parentheses alongside the differences.

Table A4: Welfare Comparison: Letting $\bar{A}$ Vary

	Consumer Surplus		Producer Surplus		Total		□
Case	Per Borrower	% Optimal	Per Borrower	% Optimal	Per Borrower	% Optimal	□
Baseline - 300	28.92	20.68	67.64	48.36	96.56	69.04	-0.0455
Baseline- 200	35.38	25.3	70.97	50.74	106.35	76.04	-0.0237
Baseline - 100	41.37	29.58	72.6	51.91	113.97	81.49	-0.0092
Baseline - 50 (□= 0)	45.66	32.65	72.92	52.14	118.58	84.79	0
Baseline (Empirical)	47.39	33.88	72.87	52.1	120.26	85.99	0.003483
Baseline + 100	53.46	38.22	71.94	51.44	125.4	89.66	0.0148
Baseline + 200	59.52	42.56	69.95	50.02	129.47	92.57	0.025
Baseline + 300	65.6	46.9	66.98	47.89	132.58	94.8	0.0342
Baseline + 400	71.69	51.26	63.14	45.15	134.83	96.41	0.0426
Baseline + 600	83.9	59.99	53.07	37.95	136.97	97.94	0.0577
Baseline + 650 (Optimal)	87.01	62.21	50.05	35.79	137.06	98	0.0612
Baseline + 800	96.15	68.75	40.18	28.73	136.33	97.48	0.0708
Baseline + 1000	108.42	77.52	24.83	17.75	133.25	95.28	0.0825

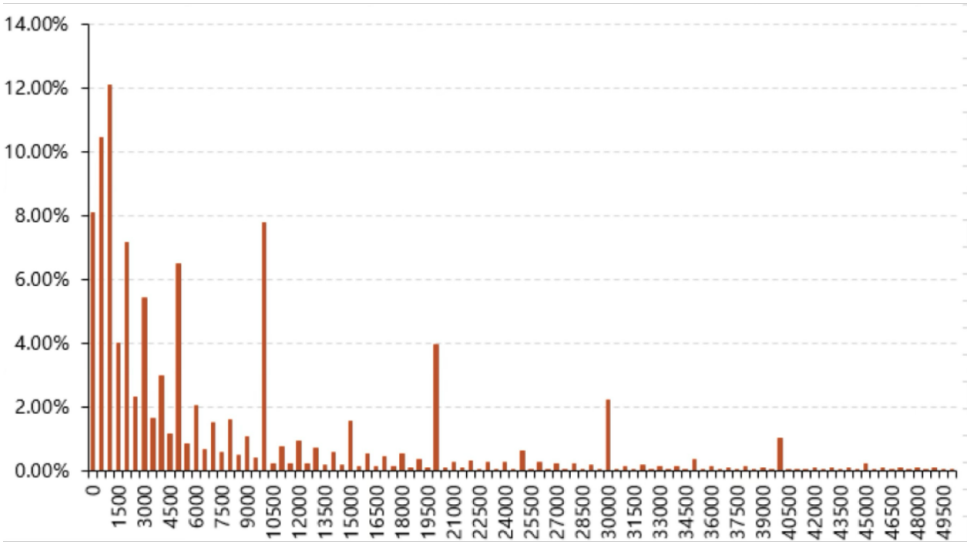
Note: This table shows welfare outcomes under the regulated monopolist situation, letting the repayment amount target, $\bar{A}$, vary. Columns 2, 4, and 6 report per-borrower weekly surplus amounts. Columns 3, 5, and 7 compare those amounts to surplus produced under the optimal pricing scenario. The empirical repayment amount is approximately 816 *yuan*, the amount used in the baseline regulated monopolist case.

B SME loan and institutional background

The typical online borrowers are SME owners who operate small stores, food delivery services, or other establishments using the e-commerce platform. The typical offline borrowers are small vendors that use a quick response (QR) code to receive digital payments from customers. The maturity of loans to both is six months to one year, with repayment over six to twelve installments. The borrower can repay the loans

in advance, and if so, she only needs to pay interest over the period the credit was drawn. This flexibility fulfills the needs of small sellers. The average loan size has grown somewhat over time, to RMB 12,000 for (ŨSD 1,880) for offline borrowers in 2021. The median less than RMB 5,000 (USD 785), showing the skew of distribution toward small loans. On average, borrowers repay the loans in one to two months (Figure A3).

Figure A3: Distribution of Loan Amount (yuan)

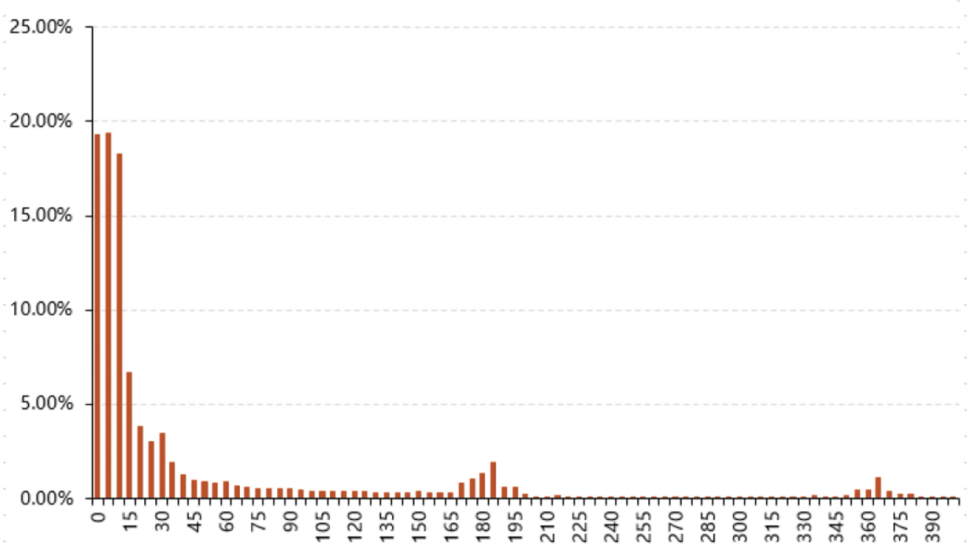


Note: This figure shows the distribution of the loan amount in *yuan*. Loan amounts are predominantly skewed toward smaller sizes, with most under RMB 5,000 (USD 785). The distribution also exhibits clear peaks at round numbers such as 10,000, 15,000, 20,000, etc.

Figure A4 distribution of the total days of loan usage, conditional on the loan are fully repaid. Although the loan terms are six months or one year, most repayments are completed well before the maturity date. More than half of loans are repaid within the first two weeks, emphasizing the short-term nature of borrowing for most clients. After 15 days, the repayment frequency drops substantially, but secondary peaks emerge at six months (180 days) and one year (365 days). The average 90-days delinquency rate of these loans is between 1% to 3%.

In evaluating the effects of the pandemic and lockdown measures, we examined the daily economic activity index across the two groups of cities (Figure A5). This index integrates key economic activities, including the daily total of digital payments in consumption, merchants' sales, and the counts of both active merchants and consumers. The findings demonstrate that while both groups of cities show significant deviations from Wuhan, the differences between the two groups themselves are not significant. During the lockdown period, differences in sales, consumption, and the number of active sellers were not statistically significant across the cities (Figure A6). However, the

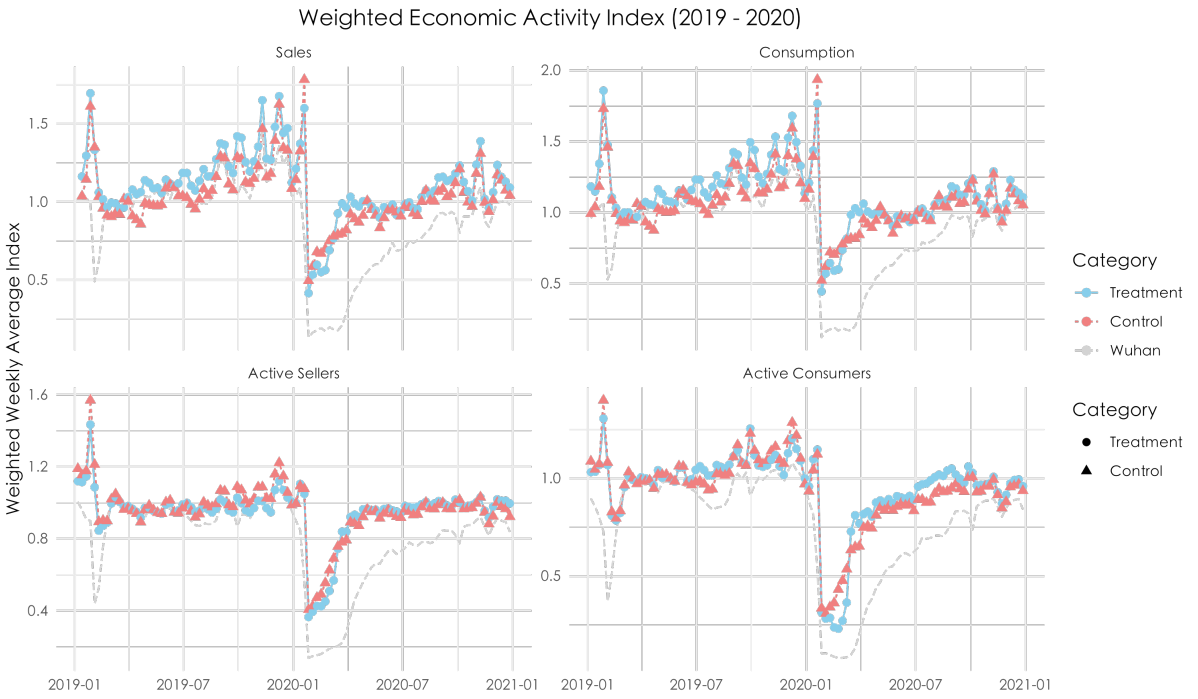
Figure A4: Distribution of Repay Days



Note: This figure shows the distribution of the total days of loan usage, conditional on the loan are fully repaid. Although the loan terms are six months or one year, most repayments are completed well before the maturity date. More than half of loans are repaid within the first two weeks, emphasizing the short-term nature of borrowing for most clients. After 15 days, the repayment frequency drops substantially, but secondary peaks emerge at six months (165 days) and one year (360 days).

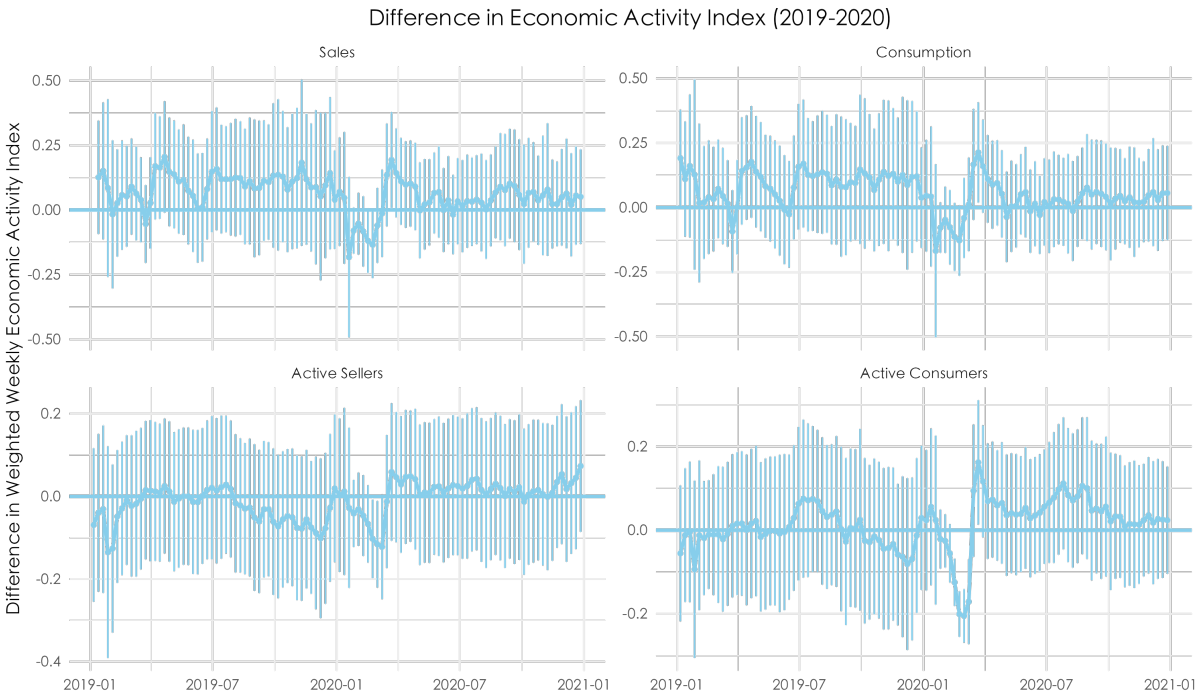
number of active consumers showed a significant difference; the difference likely stems from stricter measures in some cities allowing only one individual per household to go out and purchase groceries on each day.

Figure A5: Economic Activity Index



Note: This figure shows the economic activity index of two groups and Wuhan from 2019 to 2020 at the week level. It includes indexes on sales, consumption, and share of active sellers and buyers. For each index, we normalize the 2019 average in the city as 1. For each group, the index is weighted by the cities' GDP in 2019. Sales indicate the sales of offline merchants in the city, including face-to-face transactions or local purchases. Consumption includes both face-to-face transactions and online commerce transactions. The plots demonstrate that while both groups of cities show differences from Wuhan for the lockdown period (2020/2-4), the differences between the two groups themselves are small.

Figure A6: Distribution of Economic Activity Index



Note: This figure shows the differences in activity index between the two groups from 2019 to 2020 at the week level (treatment - control), with a 95% confidence interval. It includes indexes on sales, consumption, and share of active sellers and buyers. During the lockdown period, differences in sales, consumption, and the number of active sellers were not statistically significant across the cities. However, the number of active consumers showed a significant difference; the difference likely stems from stricter measures in some cities allowing only one individual per household to go out and purchase groceries on each day.

C Robustness Checks

C.1 Repayment Robustness with Random Forests

In our setting, the loan provider has access to a wide range of borrower's personal and financial histories. Our baseline results on the repayment rates reported in Table 3 suppose that the price borrowers face encodes most of that information, and we interpret the negative repayment elasticities as resulting from advantageous selection based on unobservable factors. To better evaluate the extent to which our current results can be explained by change in observable characteristics in the pool of borrowers, we want to control for the wide range borrower histories the platform has access to, without making assumptions on how the platform incorporates and weights this information. In particular, for a function that maps the set of data $X_{i,t}$ the lending platform has access to at the time of loan origination t, we want to evaluate the effects of the policy, conditional on the platform's predicted repayment of loan i, $f(X_{i,t})$, we assume this "predicted repayment" enters linearly into the baseline DiD specification:

$$R_{i,t} = f(X_{i,t}) + \beta_1 Treat_t + \beta_2 H_c + \beta_3 (H_c \times Treat_t) + \epsilon_{i,c,t} \tag{30}$$

To flexibly control for observables, we implement the non-parametric regression forests of Athey et al. (2019). To incorporate these platform-predicted repayment rates into the specification of 30, we proceed with the estimation in two steps, as in Robinson (1988): first, we train a random forest to predict repayment rates:

$$R_{i,t} = f(X_{i,t}) + r_{i,t} \tag{31}$$

$$\tag{32}$$

Our main outcome of interest then becomes the errors in predicted repayment, $\hat{r}_{i,t}$. We additionally residualize each of the difference-in-differences dummies based on their own random forests. For the case where we train each random forest based on the full 70 week sample of the loan, this equates to a re-balancing of the treatment and control groups based on observables:

$$H_c = g_1(X_{i,t}) + h_{i,t} \tag{33}$$

$$Treat_t = g_2(X_{i,t}) + t_{i,t} \tag{34}$$

$$(H_c \times Treat_t) = g_3(X_{i,t}) + v_{i,t} \tag{35}$$

Taking the errors of these equations, we finally estimate the difference-in-differences specification fully controlling for observable characteristics, with our treatment effect

provided by β_3.

$$\hat{r}_{i,t} = \beta_1 \hat{h}_{i,t} + \beta_2 \hat{t}_{i,t} + \beta_3 \hat{v}_{i,t} + \epsilon_{i,c,t} \tag{36}$$

We train random forests f_1, g_1, g_2, g_3 and estimate 36 separately for each interest rate group. We repeat this exercise for two cases of training samples for the random forest: 1) a pre-policy (23-week) in-sample, and 2) the full 70-week sample. In case 1, the random forest predictions for week during and after the policy period are out-of-sample predictions; while in case 2, all predictions are in-sample. While we limit the set of information for a given loan to that accessible to the platform at the loan's origination, training the random forest on the full sample equates to us allowing the platform to account for the changing relationship between observables and repayment rates during COVID and the interest rate policy.

The results for the pre-policy in-sample and full in-sample are reported in A5 and A6, respectively. Compared to our baseline results, the estimated elasticities are attenuated, though we still see significant results indicative of the presence of advantageous selection.

Table A5: Repayment Rates: Across Credit Categories, Week 23 In-Sample Random Forest

Estimate	9	10.8	12.6	14.4	16.2	18	19.8	21.6
				A. Repay Rate				
Baseline	98.182	98.451	98.076	97.678	97.665	97.483	97.491	98.172
	(11.694)	(9.926)	(11.221)	(12.001)	(11.907)	(12.524)	(11.979)	(10.739)
DiD Slope	0.319	-0.522	-0.272	-0.009	-0.417	-0.736	-0.249	0.466
	(0.289)	(0.188)	(0.167)	(0.133)	(0.141)	(0.174)	(0.179)	(0.152)
Elasticity	0.193	-0.267	-0.121	-0.004	-0.149	-0.24	-0.075	0.129
	(0.18)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Obs	26110	83996	140498	191767	174606	135648	114336	116410

Note: Table reports regression results across the 8 borrower groups as defined by the origination interest rate. The specification flexibly controls for borrower and loan observables with a random forest estimated for each interest group, trained with pre-policy loans. The first row reports the pre-policy average repayment rate across all cities, with the sample standard error given in parentheses. The second row reports the DiD coefficient, with robust standard errors reported in parentheses. The third row reports the repayment elasticity with respect to the interest rate discount, standard errors are calculated with the sub-sample (using a fixed proportion $p = .25$) bootstrap (200 replications).

Table A6: Repayment Rates: Across Credit Categories, Week 70 In-Sample Random Forest

Estimate	9	10.8	12.6	14.4	16.2	18	19.8	21.6
				A. Repay Rate				
Baseline	98.182	98.451	98.076	97.678	97.665	97.483	97.491	98.172
	(11.694)	(9.926)	(11.221)	(12.001)	(11.907)	(12.524)	(11.979)	(10.739)
DiD Slope	0.473	-0.375	-0.174	-0.082	-0.451	-0.625	-0.162	0.09
	(0.294)	(0.193)	(0.157)	(0.128)	(0.134)	(0.165)	(0.17)	(0.151)
Elasticity	0.287	-0.192	-0.077	-0.033	-0.161	-0.204	-0.049	0.025
	(0.17)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Obs	26110	83996	140498	191767	174606	135648	114336	116410

Note: Table reports regression results across the 8 borrower groups as defined by the origination interest rate. The specification flexibly controls for borrower and loan observables with a random forest estimated for each interest group, trained on the full sample of loans. The first row reports the pre-policy average repayment rate across all cities, with the sample standard error given in parentheses. The second row reports the DiD coefficient, with robust standard errors reported in parentheses. The third row reports the repayment elasticity with respect to the interest rate discount, standard errors are calculated with the sub-sample (using a fixed proportion $p = .25$) bootstrap (200 replications).

To directly show the ability of the random forests to predict repayment across the sample period specifications, Table A7 shows the root mean squared errors (RMSE) of the models during the pre- and post-policy periods. As can be seen, pre-policy (weeks 1-23) model performance is comparable, with errors ranging from 5% to 8%, with errors typically lower for lower interest rate borrowers. For the post-policy period, both models deteriorate in performance. Importantly, for Model 2, which allows the random forests to incorporate the policy period in its training, there is still a marked decrease in performance. We can interpret this worsened performance as evidence that observable borrower characteristics being less predictive of loan repayment, as a result of advantageous selection of the borrower pool based on unobservable.

Table A7: Random Forest Performance: RMSE of Repayment Rates

Weeks	9	10.8	12.6	14.4	16.2	18	19.8	21.6
Model 1: Pre-Policy In-Sample								
Weeks 1-23	4.95	5.33	6.43	7.03	7.79	7.56	7.34	7.24
Weeks 23-70	12.45	11.02	12.58	12.6	12.48	13.42	12.42	11.32
Model 2: Full In-Sample								
Weeks 1-23	5.1	5.43	6.52	7.14	7.93	7.64	7.44	7.27
Weeks 23-70	8.38	7.79	8.65	8.62	8.69	9.03	8.6	7.91

This table displays the Root Mean Squared Error (RMSE) for the random forest predicted repayment rates. The top panel shows model performance for the random forest for repayment with the pre-policy 23-week training sample; the bottom panel shows performance with the full 70-week training sample. For the top panel, weeks 1-23 show the in-sample error, and weeks 23-70 are out-of-sample. For the bottom, both weeks 1-23 and weeks 23-70 are in-sample predictions. Repayment rates are measured in percentage point terms, ranging from 0 to 100.

Table A8: Variable Importance: Week 23 In-Sample

Variable	9	10.8	12.6	14.4	16.2	18	19.8	21.6	Overall
Current Credit Utilization	46.32	48.02	47.48	49.27	44.63	50.28	46.67	49.91	47.8225
Cumulative Net Balance	19.95	10.98	10.86	9.02	18.54	17.32	21.16	10.35	14.7725
Total Loans Repaid in Week	0.27	2.41	0.69	13.39	8.7	10.23	4.23	15.34	6.9075
Value of New Loans Originated in Week	2.32	4.29	17.33	8.61	10.66	4.31	5.46	1.04	6.7525
Cumulative Number of Loans	4.92	1.86	4.96	0.92	5.34	2.6	3.53	12.62	4.59375
Weekly Net Balance	3.15	13.15	4.74	6.2	2.62	1.1	3.46	1.55	4.49625
Cumulative Loan Amount	3.24	4.78	7.91	4.74	2.98	2.85	4.46	1.51	4.05875
Effective Consumer Credit	4.47	2.06	2.45	2.17	1.25	2.72	2.51	0.78	2.30125
Total Repayment Value in Week	0.69	0.54	0.61	3.73	2.01	2.13	0.7	2.2	1.57625
Age	4.2	0.7	0.44	0.37	0.86	0.23	1.43	0.58	1.10125
Weekly Sales (Unique Number)	3.62	0.43	0.39	0.26	0.13	1.23	0.92	0.34	0.915
Weekly Purchases (Yuan)	2.25	1.78	0.43	0.32	0.58	0.49	0.85	0.57	0.90875
Native Flag	0.18	6.26	0.08	0.04	0.1	0.06	0.46	0.08	0.9075
Number of New Loans Originated in Week	0.38	0.63	0.6	0.34	0.51	0.96	1.02	1.8	0.78
Weekly Purchases (Unique Number)	2.61	0.51	0.26	0.2	0.23	0.34	0.73	0.54	0.6775
Weekly Sales (Yuan)	0.56	0.65	0.33	0.17	0.23	1.49	1.31	0.39	0.64125
Occupation: Unknown	0.12	0.08	0.13	0.04	0.19	0.92	0.3	0.07	0.23125
Occupation: Blue Collar	0.07	0.07	0.09	0.02	0.21	0.49	0.07	0.19	0.15125
New Borrower Flag	0.34	0.23	0.07	0.02	0.19	0.08	0.08	0.04	0.13125
Gender: Male	0.12	0.18	0.04	0.05	0.01	0.1	0.47	0.06	0.12875
Gender: Female	0.13	0.35	0.06	0.07	0.01	0.05	0.17	0.02	0.1075
Occupation: White Collar	0.05	0.05	0.04	0.03	0.02	0	0	0	0.02375
Education: High School or Below	0.02	0	0	0	0	0	0	0	0.0025
Occupation: Civil Servant	0	0	0	0	0	0	0	0	0
Occupation: College Student	0	0	0	0	0	0	0	0	0
Occupation: Listed Company Employee	0	0	0	0	0	0	0	0	0
Occupation: Medical Staff	0	0	0	0	0	0	0	0	0
Occupation: Renowned Company Employee	0	0	0	0	0	0	0	0	0
Occupation: Listed Company Executive	0	0	0	0	0	0	0	0	0
Occupation: Teaching Staff	0	0	0	0	0	0	0	0	0
Education: Bachelors	0	0	0	0	0	0	0	0	0
Education: Doctorate	0	0	0	0	0	0	0	0	0
Education: Junior College	0	0	0	0	0	0	0	0	0
Education: Masters	0	0	0	0	0	0	0	0	0
Education: Unknown	0	0	0	0	0	0	0	0	0
Gender: Unknown	0	0	0	0	0	0	0	0	0
Sesame Credit Access	0	0	0	0	0	0	0	0	0

Note: This table shows a measure of variable importance captured in the random forest estimation for repayment rates, for the model with a 23-week in-sample. For each random forest, variable importance is the proportion of splits performed at each variable. A higher number indicates including that feature typically reduces model error when it is included. For each model (one per interest rate group), the total variable importance sums to 100. The "Overall" importance is a simple mean of importance across the models.

Table A9: Variable Importance: Week 70 In-Sample

Variable	9	10.8	12.6	14.4	16.2	18	19.8	21.6	Overall
Current Credit Utilization	45.73	45.39	48.59	48.81	46.74	45.71	46.2	48.39	46.945
Cumulative Net Balance	13.24	12.43	18.65	12.77	17.73	14.15	20.42	8.19	14.6975
Total Loans Repaid in Week	3.99	15.78	6.78	11.92	11.32	18.74	10.36	20.45	12.4175
Total Repayment Value in Week	18.15	9.71	10.05	12.7	7.37	6.5	5.69	7.41	9.6975
Cumulative Loan Amount	2.34	4	7.07	4.97	3.79	4.33	1.4	0.83	3.59125
Weekly Net Balance	1.98	2.61	1.8	1.37	3.97	3.27	8.01	3.01	3.2525
Value of New Loans Originated in Week	0.91	4.66	2.27	4.4	3.8	4.68	2.34	1.32	3.0475
Cumulative Number of Loans	4.22	2.7	2.91	1.03	3.19	0.54	0.66	6.48	2.71625
Effective Consumer Credit	2.06	0.64	0.49	0.66	0.3	0.75	2.24	0.52	0.9575
Number of New Loans Originated in Week	0.82	0.38	0.47	0.48	0.84	0.34	0.34	1.76	0.67875
Age	2.07	0.21	0.19	0.1	0.1	0.07	0.88	0.56	0.5225
Weekly Purchases (Yuan)	0.85	0.31	0.13	0.28	0.34	0.11	0.49	0.31	0.3525
Weekly Sales (Unique Number)	1.37	0.29	0.22	0.03	0.14	0.09	0.14	0.05	0.29125
Weekly Sales (Yuan)	0.69	0.37	0.14	0.14	0.11	0.13	0.19	0.08	0.23125
Weekly Purchases (Unique Number)	0.35	0.3	0.1	0.03	0.07	0.11	0.24	0.31	0.18875
Occupation: Unknown	0.27	0.04	0.02	0.02	0.02	0.17	0.15	0.13	0.1025
Occupation: White Collar	0.51	0.01	0	0	0.01	0	0	0	0.06625
Native Flag	0.37	0.05	0.01	0.01	0.02	0.02	0.01	0.04	0.06625
New Borrower Flag	0.03	0.02	0.08	0.08	0.09	0.16	0.03	0.04	0.06625
Occupation: Blue Collar	0	0.01	0	0.08	0.04	0.11	0.15	0.04	0.05375
Gender: Male	0.03	0.04	0.02	0.12	0	0.01	0.03	0.05	0.0375
Gender: Female	0.02	0.04	0.01	0.01	0	0	0.02	0.01	0.01375
Occupation: Civil Servant	0	0	0	0	0	0	0	0	0
Occupation: College Student	0	0	0	0	0	0	0	0	0
Occupation: Listed Company Employee	0	0	0	0	0	0	0	0	0
Occupation: Medical Staff	0	0	0	0	0	0	0	0	0
Occupation: Renowned Company Employee	0	0	0	0	0	0	0	0	0
Occupation: Listed Company Executive	0	0	0	0	0	0	0	0	0
Occupation: Teaching Staff	0	0	0	0	0	0	0	0	0
Education: Bachelors	0	0	0	0	0	0	0	0	0
Education: Doctorate	0	0	0	0	0	0	0	0	0
Education: High School or Below	0	0	0	0	0	0	0	0	0
Education: Junior College	0	0	0	0	0	0	0	0	0
Education: Masters	0	0	0	0	0	0	0	0	0
Education: Unknown	0	0	0	0	0	0	0	0	0
Gender: Unknown	0	0	0	0	0	0	0	0	0
Sesame Credit Access	0	0	0	0	0	0	0	0	0

Note: This table shows a measure of variable importance captured in the random forest estimation for repayment rates, for the model with the full 70-week in-sample. For each random forest, variable importance is the proportion of splits performed at each variable. A higher number indicates including that feature typically reduces model error when it is included. For each model (one per interest rate group), the total variable importance sums to 100. The "Overall" importance is a simple mean of importance across the models.

C.2 Short Term Repayment

The repayment results shown thus far are of final loan repayment rates, i.e. long-term measures of loan repayment. Over the sample period, this long-term repayment measure may include responses to COVID or strategic loan repayment, not just the lending platform's policy. To get a short-term repayment measure we define the k-week repayment as the share of an outstanding loan's value paid in the first k weeks

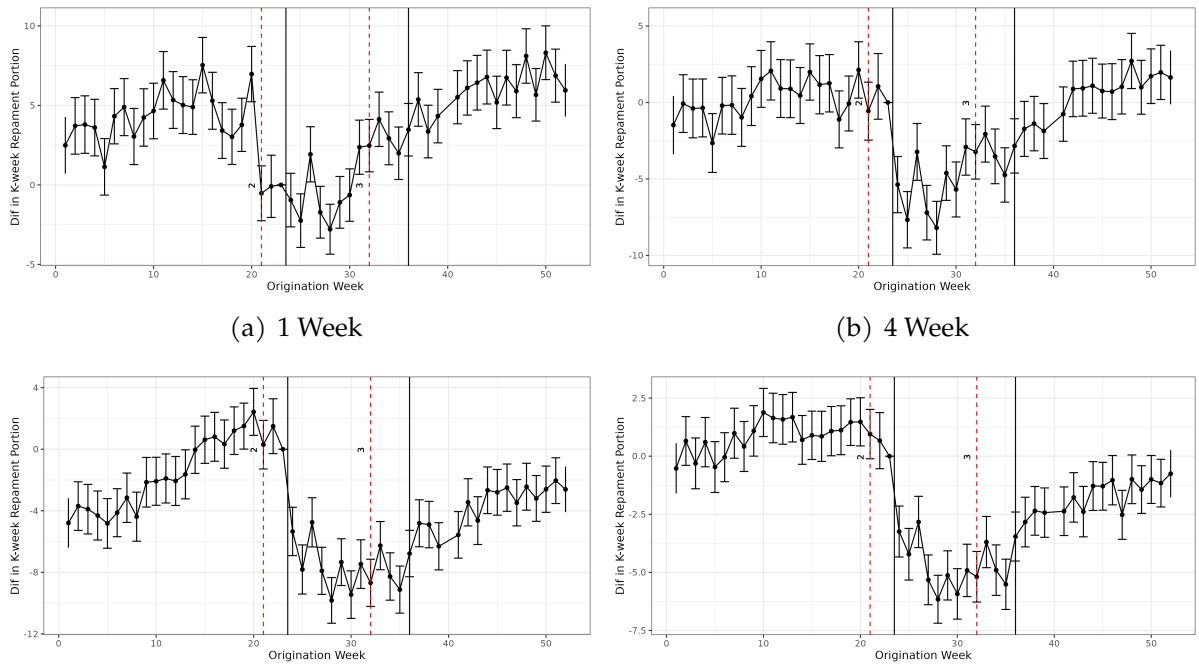
for a loan i originated at time t

$$R^k_{i,t} = \frac{\sum_{j=0}^{k} Repayment_{i,t+j}}{Balance_i}$$

Where $Balance_i$ is the original value of the loan, and $Repayment_{i,t+j}$ is the value of payments made towards the loan in week $t+j$. To get a sense of how short-term repayment depends on other shocks over origination weeks t, we estimate for a given horizon k the following specification:

$$R^k_{i,c,t} = \sum_{t=1;t\neq 23}^{70} \beta^k_t (H_i \times T_t) + I_{c,t} + \mu_c + \epsilon_{i,c,t} \tag{37}$$

We plot the time-varying coefficients β^k_t for different choices of k in A7



(a) 1 Week (b) 4 Week

(c) 12 Week (d) 26 Week

Figure A7: Short Term Repayment for Select Industries, Selected Horizons k

Note: This figure shows β^k_t for different choices of k. Week 0 corresponds to the week of September 1st, 2019. Policy starts week 24 and persists until week 36 (shown by the vertical bars). Important Covid events are shown by the dotted red lines: line 1 shows the start of the Wuhan lockdown (January 23rd 2020), line 2 shows the end of the lockdown (April 8th, 2020)

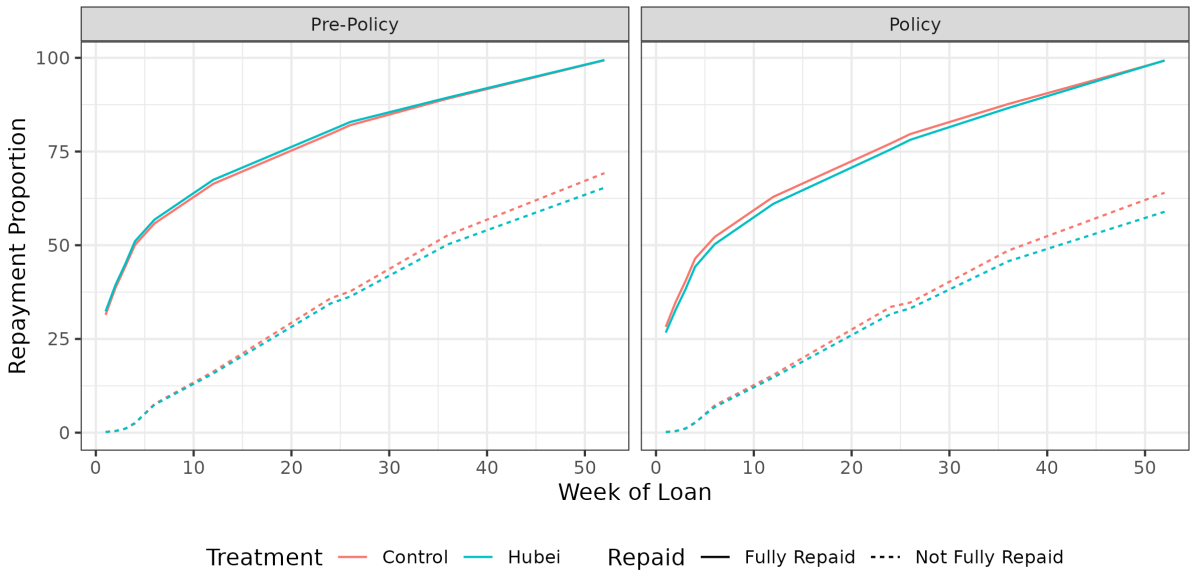


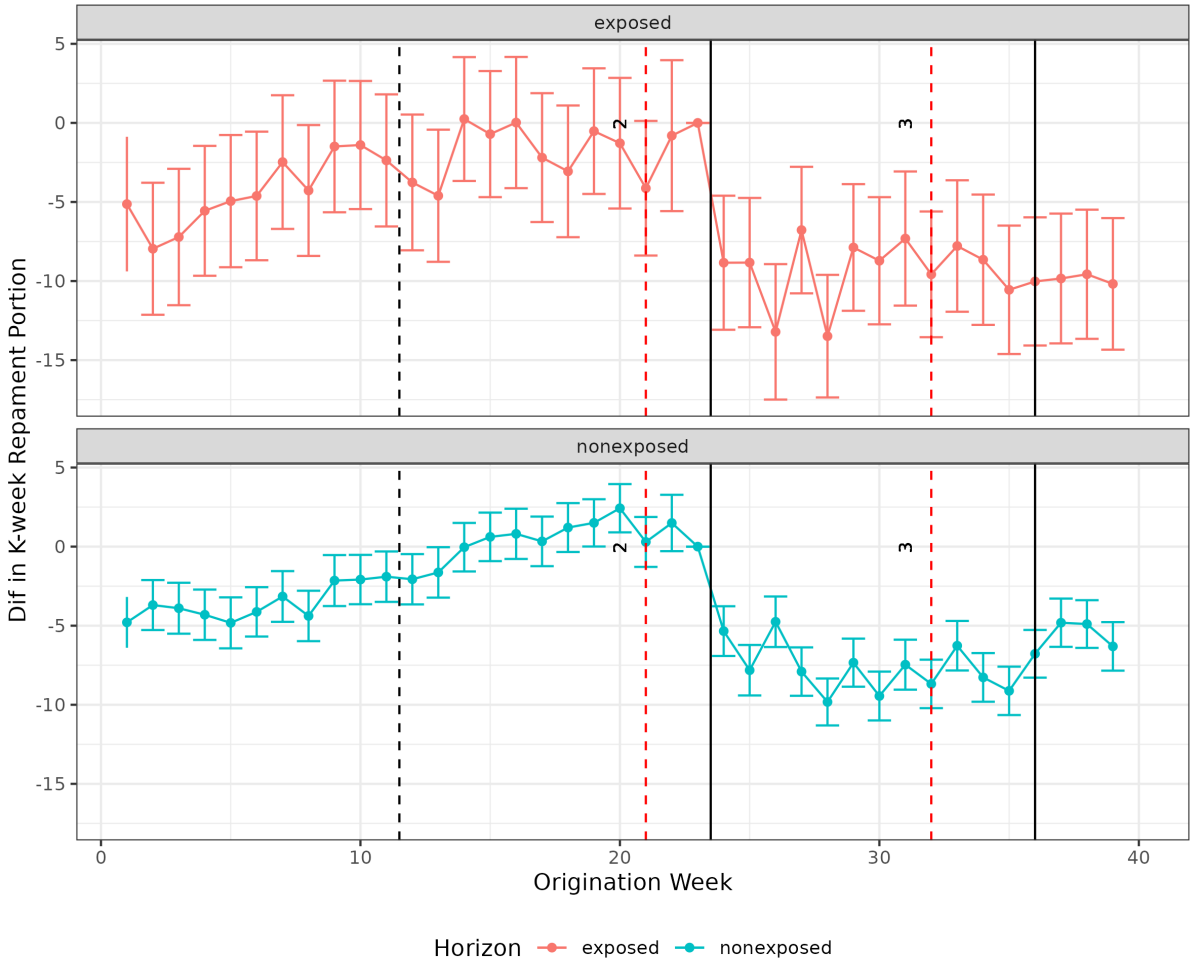
Figure A8: Repayment Patterns over Life of loan

C.3 Pre-trends: Long-term repayment; Effects of Sample Updating

Let $R_{j,t,c}$ be the repayment rate on the loan as defined in the loan dataset, i.e., $R = (amt - bal)/amt$ as of the last update. Then substitute in the maximum of repayment at maturity $\{R^{26}, R^{52}\}$ depending on the terms.

$$R_{j,t,c} = \beta_0 + \sum_{j=0;j\neq 23}^{40} \beta_j^k (\mathbb{1}\{t = j\} \times \mathbb{1}\{c \in Hubei\}) + \mu_c + \epsilon$$

Figure A9: Estimate for non-exposed industries



Note: This figure shows the DID estimate of repayment rates in 12 weeks for non-exposed industries and other industries. Non-exposed industries include restaurants, grocery stores, and clothing shops, consisting of 40% of total borrowers. Exposed groups indicate borrowers from all other industries or unknown industries.

D Extension of Two Dimensional Types

In this section, we look at the borrowers' type in two dimensions: interest rate and the credit line. The idea is that, from the view of a Fintech lender, the interest rate reflects the borrower's willingness to repay, while the credit limit primarily considers their ability to repay. Both aspects of information are valuable for the lender, as its goal is to earn interest while controlling risk. In the data, for each interest rate group, the distribution of credit limits across intervals appears relatively independent. This supports the validity of considering a two-dimensional type space for analyzing borrower characteristics.

For each interest rate group, we divide borrowers into 4 types based on their credit line, with cutoffs of quantiles 12,000, 38,000, and 69,000. Then we estimate the elasticity of repayment rate, loan amount, and probability of application as we did in the benchmark counterfactual analysis. The average values used in calibration is same as in Table 6 and the share for each group is in **??**.

Table A10: Credit Group Heterogeneity: Elasticities

Credit Group	9	10.8	12.6	14.4	16.2	18	19.8	21.6
I. Repay Rate								
(0,5e+03]	4.16	-1.78	1.97	-0.37	0.27	-0.54	-0.41	-0.14
	(1.7)	(0.64)	(0.41)	(0.28)	(0.18)	(0.14)	(0.09)	(0.09)
(5e+03,3.87e+04]	-0.2	-1.44	-0.3	-0.6	-0.64	-0.62	-0.53	-0.22
	(0.71)	(0.3)	(0.2)	(0.1)	(0.1)	(0.11)	(0.11)	(0.09)
(3.87e+04,6e+04]	-0.7	0.22	-0.87	-0.49	-0.76	-0.31	-0.12	0.07
	(0.61)	(0.25)	(0.18)	(0.13)	(0.12)	(0.15)	(0.22)	(0.13)
(6e+04,2.2e+05]	-0.84	-0.44	-0.65	-0.41	-0.2	-0.58	-0.28	-0.39
	(0.22)	(0.16)	(0.13)	(0.11)	(0.11)	(0.12)	(0.21)	(0.13)
II. New Loan Amount								
(0,5e+03]	23.51	29.49	24.55	23.16	23.51	8.67	17.39	7.67
	(32.32)	(7.89)	(5.61)	(3.86)	(3.05)	(2.29)	(1.28)	(1.07)
(5e+03,3.87e+04]	65.74	32.34	19.23	16.87	17.4	14.43	14.21	5.36
	(11.4)	(4.59)	(2.58)	(1.7)	(1.36)	(1.33)	(1.29)	(1.14)
(3.87e+04,6e+04]	34.23	22.14	15.66	11.77	5.94	5.93	0.32	3.57
	(8.27)	(4.16)	(2.24)	(1.62)	(1.34)	(2.01)	(2.64)	(2.31)
(6e+04,2.2e+05]	9.42	3.64	5.81	4.22	6.35	-2.08	-0.5	-3.74
	(3.82)	(2.7)	(1.56)	(1.82)	(1.78)	(2.54)	(3.93)	(2.89)
III. I(New Loan)								
(0,5e+03]	32.34	24.04	14.41	18.6	19.09	14.3	19.82	8.47
	(13.35)	(3.29)	(2.56)	(1.49)	(1.29)	(0.92)	(0.69)	(0.67)
(5e+03,3.87e+04]	26.18	19.8	11.47	14.87	12.46	12.67	11.69	4.9
	(5.78)	(2.49)	(1.32)	(1.15)	(0.75)	(0.77)	(0.87)	(0.68)
(3.87e+04,6e+04]	20.68	15.37	8.42	6.08	4.84	4.35	3.15	0.97
	(4.63)	(2.07)	(1.53)	(0.99)	(0.98)	(1.27)	(1.38)	(1.36)
(6e+04,2.2e+05]	2.22	0.41	0.64	2.02	3	2.48	1.2	-3.52
	(2.41)	(1.44)	(1.01)	(0.86)	(1.05)	(1.37)	(2.18)	(1.76)

Table A11: Credit Group Heterogeneity: Baselines

Credit Group	9	10.8	12.6	14.4	16.2	18	19.8	21.6
I. Repay Rate								
(0,5e+03]	97.88	97.18	95.99	95.69	96.78	97.39	97.58	98.2
	(11.63)	(12.9)	(15.67)	(16.57)	(13.81)	(12.45)	(11.44)	(10.2)
(5e+03,3.87e+04]	98.24	97.94	97.83	97.49	97.17	97.12	96.99	98.08
	(10.64)	(11.01)	(11.4)	(12.01)	(13.03)	(13.18)	(13.37)	(11.02)
(3.87e+04,6e+04]	97.84	98.47	97.81	97.7	97.85	97.47	97.92	98.28
	(12.86)	(9.91)	(12)	(11.94)	(11.54)	(12.89)	(10.98)	(11.32)
(6e+04,2.2e+05]	98.34	98.65	98.44	98.24	98.24	98.11	98.35	98.55
	(11.43)	(9.52)	(10.43)	(10.75)	(10.72)	(11.75)	(10.88)	(10.47)
II. New Loan Amount								
(0,5e+03]	79.61	85.75	99.91	92.23	83.07	82.52	61.91	65.69
	(981.55)	(1069.19)	(1180.92)	(1864.73)	(1247.53)	(1651.13)	(528.95)	(469.91)
(5e+03,3.87e+04]	484.81	537.48	529.47	566.21	550.12	550.32	482.38	478.02
	(2991.35)	(3108.68)	(3194.56)	(3090.76)	(2992.87)	(2812.03)	(2548.41)	(2560.36)
(3.87e+04,6e+04]	1017.74	1210.25	1250.67	1555.78	1562.03	1707.11	1659.07	1344.1
	(5726.31)	(6759.7)	(6009.62)	(6373.06)	(6332.24)	(6818.89)	(6663.89)	(6115.26)
(6e+04,2.2e+05]	2863.9	2939.12	2870.42	3281.79	3020.65	2976.98	2917.92	2445.19
	(14261.47)	(13494.45)	(12444.71)	(13770.31)	(12626.45)	(12244.59)	(11585.3)	(10938.5)
III. I(New Loan)								
(0,5e+03]	0.03	0.04	0.04	0.04	0.04	0.05	0.05	0.05
	(0.16)	(0.19)	(0.19)	(0.2)	(0.2)	(0.22)	(0.21)	(0.22)
(5e+03,3.87e+04]	0.06	0.08	0.08	0.09	0.09	0.09	0.09	0.09
	(0.24)	(0.27)	(0.27)	(0.28)	(0.29)	(0.29)	(0.28)	(0.28)
(3.87e+04,6e+04]	0.08	0.1	0.11	0.13	0.14	0.14	0.14	0.11
	(0.27)	(0.29)	(0.31)	(0.34)	(0.34)	(0.35)	(0.35)	(0.32)
(6e+04,2.2e+05]	0.11	0.12	0.14	0.15	0.15	0.14	0.15	0.12
	(0.31)	(0.33)	(0.34)	(0.36)	(0.36)	(0.35)	(0.35)	(0.32)

Table A12: Welfare Comparison: Compared to Optimal

	Consumer Surplus		Producer Surplus		Per Bororwer Outcomes		Total Surplus	
Case	Per Borrower	% Hubei GDP	Per Borrower	% Hubei GDP	Total Quantities	Average Price	% Opt	% Hubei GDP
Monopsony, Information	61.200	0.009	87	0.013	657.050	0.840	83.845	0.022
Competition, Information	164	0.024	0	0	1,552.740	0.930	92.784	0.024
Competition, No Information	157.500	0.023	0	0	1,458.760	0.920	89.107	0.023
Monopsony, No Information	74.900	0.011	74.100	0.011	574.390	0.830	84.298	0.022
Regulated VaR Monopsony, Information	69	0.010	86.100	0.013	769.700	0.850	87.749	0.023

Note: This table compares welfare across the counterfactual scenarios. Columns 1 and 3 shows weekly consumer and producer surplus per borrower; for information cases, this is the sum of surpluses across the 9 interest rate markets, for no-information cases it is the surplus in the one combined market. Columns 2,4, and 9 scales the per-borrower surplus by the number of borrowers in the sample used for parameter estimation (152,819), and by 52 weeks; this quantity is then divided by Hubei's regional GDP. Columns 6 and 7 show the total weekly credit quantity and borrower-weighted average prices. Column 8 shows the total surplus as a proportion of the welfare produced in the optimal pricing scenario.

Figure A10: Affine Relationships: Competition

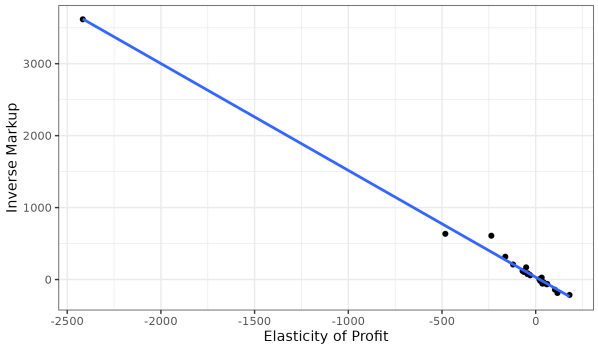
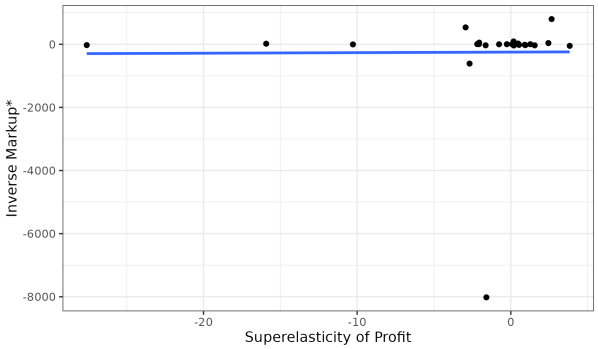


Figure A11: Affine Relationships: monopolist



E Omitted proofs and claims

E.1 Adverse Selection

In this section we argue that with exogenous variation in prices, assuming there is no moral hazard, we can non-parametric identify the degree of adverse or advantageous selection.

Formally, let p be the bond's price (the inverse of the interest rate). Let $b(\omega)$ be the reservation value for the borrower. That is, if (and only if) $p > (\omega)$, the ω sells the bond (of fixed size) to the lender. Let $a(\omega)$ be the value for the borrower of the bond that is being sold. Let $z(\omega)$ be the signal observed by the lender. Finally, let $F(\omega)$ be the distribution of types, which induces a joint distribution $G(a, b, z)$.

By varying prices, we can non-parametrically identify the demand function for each signal realization z, that is, $Q(p|z) = \int \mathbb{1}(b \leq p) dG(a, b|z) = G_b(p|z)$, where $G_b(p|z) \equiv \int_{a \in A} G(a, b|z)$, is the marginal distribution of b, conditional on z. Moreover, by varying prices, we can non-parametrically identify the average repayment rates, denoted $R(p|z)$. Under the assumption that the lender is risk-neutral the average value for the lender is just the product of the repayment rate and a discount rate, that is $Avg_V(p|z) = \delta R(p|z) = \int a \mathbb{1}(b \leq p) dG(a, b|z) = \mathbb{E}[a|b \geq p, z]$.

E.2 Moral Hazard

In this section, we show that moral hazard is unlikely to explain a negative relationship between interest rates and default probabilities.

Consider the following setup. The borrower chooses a hidden action $a \in A$ after signing a borrowing contract. The distribution of returns $x \sim F_a$ for the project depends on that action a. The borrower receives a payment today (which we normalize to one), in exchange for a promise to repay R in the future. The borrower repays if the project return minus the repayment is positive, that is if $x - R \geq 0$. The borrower's expected payoff is $V = \max_{a \in A} \mathbb{E}[max(x - R, 0)|a]$.

Claim: if $R' \geq R$, then $\mathbb{P}(x - R \geq 0|a^*(R)) \geq \mathbb{P}(x - R' \geq 0|a^*(R'))$.

Proof. Suppose that a project a' is safer when the interest rate is R' than a when interest rate is R, that is $\mathbb{P}(x \geq R'|a') > P(x \geq R|a)$. We are going to show that it cannot be that a' is chosen when interest rate is R' while a is chosen when interest rate is R.

For a to become more attractive after the interest rate decreases from R' to R, the change in payoff from selecting a, $\Delta\pi(a)$ should be higher than the change in payoff from selecting a', $\Delta\pi(a')$.

The change in payoff from selecting a' can be bounded below by the change in interest charges in the region above R', that is, $\Delta\pi(a') \geq P(x \geq R'|a')(R' - R)$.

The change in payoff from selecting a can be decomposed into two parts: how profits are different when x is above R', and how profits are different when x is higher than R', and how they are now different when x is between R and R'. The change in the first part is equal to $P(x \geq R'|a)(R' - R)$, while the change in the second part is at most $P(R \leq x < R'|a')(R' - R)$. Thus the change in payoff from selecting a is at most $P(x \geq R|a)(R' - R)$.

Therefore, since $\mathbb{P}(x \geq R'|a') > P(x \geq R|a)$, a cannot be become more attractive than a' when interest rates decrease from R' to R. □

□

E.3 Information policy with adverse selection

Setup Consider two signals z_1 and z_2. Suppose the equilibrium prices associated with each of them are "close to each other." We are interested in the effects of combining (or separating) the two signals into a single message.

E.3.1 Adverse selection with competition

To first-order , the change in welfare is equal to the change in prices. Taylor expand the profits from each of the signals as

$$\pi_i(p) \approx \pi_i(p_i) + \pi_i'(p_i)(p - p_i) \tag{38}$$

Denoting p_i the full disclosure price, we have that $\pi_i(p_i) = 0$. Now we pick the new price $p = \frac{p_1+p_2}{2}$, where $p_1 < p < p_2$. We combine one unit of $z_1$with α units of z_2. The investor will break even at that new price if

$$\pi_1(p) + \alpha \cdot \pi_2(p) = 0 \tag{39}$$

Combining Equation 38 and Equation 39, we conclude that:

$$\alpha = \frac{\pi_1'(p_1)}{\pi_2'(p_2)} \tag{40}$$

Since $\pi_i'(p_i) = Q_i(p_i)AVG_i'(p_i)$, α can be written as

$$\alpha = \frac{Q_1(p_1)(AVG_1'(p_1) - 1)}{Q_2(p_2)(AVG_2'(p_2) - 1)} \tag{41}$$

And

$$\alpha Q_2(p_2) > Q_1(p_1) \iff AVG_1'(p_1) > AVG_2'(p_2)$$

Which reads as "combining signals is bad if elasticities are increasing across markets

where prices are higher." We conclude from the argument above that the statement in Proposition 1 is true. □

E.3.2 Adverse selection with monopoly power

Now, the investor needs that $\pi'(p) = 0$ at each bundle. Taylor expand this expression for each signal:

$$\pi_i'(p) = \pi_i'(p_i) + \pi_i''(p_i)(p - p_i) \tag{42}$$

We will combine the two signals in a bundle which will have an equilibrium price $p = \frac{p_1+p_2}{2}$. If one unit of $z_1$is combined with α units of z_2, investors indifference will hold if:

$$\pi_1''(p_1) - \alpha\pi_2''(p_2) = 0 \tag{43}$$

And thus:

$$\alpha = \frac{\pi_1''(p_1)}{\pi_2''(p_2)} \;\Rightarrow\; \alpha\frac{Q_2(p_2)}{Q_1(p_1)} = \frac{\pi_1''(p_1)/Q_1(p_1)}{\pi_2''(p_2)/Q_2(p_2)} \tag{44}$$

Which reads as: *combining signals is bad if $\pi_i''(p_i)/Q_i(p_i)$ is increasing across markets where prices are higher.* We conclude from the argument above that the statement in Proposition 2 is true. □

Next, we express it in terms of more conventional sufficient statistics, such as elasticities and markups.

$$\pi_i'(p_i) = Q_i'(p_i)(AVG(p_i) - p_i) + Q_i(p_i)(AVG'(p_i) - 1) \tag{45}$$

$$\pi''(p) = Q''(p)(Avg(p) - p) + 2Q'(p)(Avg'(p) - 1) + Q(p)Avg''(p) \tag{46}$$

Define $MK(p) \equiv Avg(p) - p$, then from $\pi(p) = Q(p) \cdot MK(p)$ we have:

$$\pi'' = \left(\frac{Q''p^2}{Q} + 2\frac{Q'p}{Q}\frac{MK'p}{MK} + \frac{MK''p^2}{MK}\right)\frac{QMK}{p^2} \tag{47}$$

Define the superelasticities for these variables as:

$$\eta^+_{Q,p} \equiv \frac{d\left(\frac{Q'p}{Q}\right)}{d\log p} = \frac{Q''p^2}{Q} + \frac{Q'p}{Q} - \frac{Q'^2p^2}{Q^2} \tag{48}$$

$$\epsilon^+_{Q,p} \equiv \frac{\eta^+_{Q,p}}{\epsilon_{Q,p}} \tag{49}$$

$$\eta^{+}_{MK,p} \equiv \frac{d\left(\frac{MK'p}{MK}\right)}{d\log p} = \frac{MK''p^2}{MK} + \frac{MK'p}{MK} - \frac{MK'^2p^2}{MK^2} \tag{50}$$

$$\epsilon^{+}_{MK,p} \equiv \frac{\eta^{+}_{MK,p}}{\epsilon_{MK,p}} \tag{51}$$

Then, we have that :

$$\pi'' = \left((\epsilon^{+}_{Q,p} - 1 + \epsilon_{Q,p})\epsilon_{Q,p} + 2\epsilon_{Q,p}\epsilon_{MK,p} + \epsilon_{MK,p}(\epsilon^{+}_{MK,p} - 1 + \epsilon_{MK,p})\right)\frac{QMK}{p^2} \tag{52}$$

Next let's expand $\epsilon^{+}_{MK,p}$ as a function of the $\epsilon^{+}_{AVG,p}$, and other elasticities, and we can get:

$$\eta^{+}_{AVG,p} = \frac{MK''p^2}{MK}\frac{MK}{AVG} + \frac{AVG'p}{AVG} - \frac{AVG'^2p^2}{AVG^2} \tag{53}$$

We conclude that the second derivative of profits can be written as:

$$\pi'' = \left(\epsilon^{+}_{Q,p}\epsilon_{Q,p} - \epsilon_{Q,p} + (\epsilon_{Q,p})^2 + 2\epsilon_{Q,p}(\epsilon_{AVG,p} - \frac{p}{AVG}) + (\epsilon^{+}_{AVG,p}\epsilon_{AVG,p} - \epsilon_{AVG,p} + (\epsilon_{AVG,p})^2)\frac{AVG}{MK}\right)\frac{QMK}{p^2} \tag{54}$$

Under the assumption that superelasticities are zero, this simplifies to:

$$\pi'' = ((\epsilon_{Q,p})^2 - \epsilon_{Q,p} + 2\epsilon_{Q,p}\left(\epsilon_{AVG,p} - \frac{p}{AVG}\right))\frac{QMK}{p^2} + \left((\epsilon_{AVG,p})^2 - \epsilon_{AVG,p}\right)\frac{QAVG}{p^2} \tag{55}$$

Alternatively, if we assume linear demand and linear average value curves:

$$\frac{\pi''(p)}{Q(p)} = 2\cdot\epsilon_{Q,p}(p)\cdot\frac{Avg'(p)-1}{Avg(p)-p}p\frac{Avg(p)-p}{p^2} = 2\cdot\epsilon_{Q,p}(p)\cdot\epsilon_{MU,p}\cdot\frac{Avg(p)-p}{p^2} \tag{56}$$

Equation 54 provides an alternative general expression for the second derivatives of profits (as they appear in Proposition 2), written in terms of elasticities of take-up and average value functions. Equation 55 specializes them to the case where superelasticities are zero. Equation 56 specializes them to the case where demand and average value curves are linear.