Open Banking and Customer Data Sharing: Implications for FinTech Borrowers

Rachel J. Nam*

First draft: Nov 30, 2022

Latest version here

Abstract

With open banking, consumers take greater control over their own financial data and share it at their discretion. Using a rich set of loan application data from the largest German FinTech lender in consumer credit, this paper studies what characterizes borrowers who share data and assesses its impact on loan application outcomes. I show that riskier borrowers share data more readily, which subsequently leads to an increase in the probability of loan approval and a reduction in interest rates. The effects hold across all credit risk profiles but are the most pronounced for borrowers with lower credit scores (a higher increase in loan approval rate) and higher credit scores (a larger reduction in interest rate). I also find that standard variables used in credit scoring explain substantially less variation in loan application outcomes when customers share data. Overall, these findings suggest that open banking has the potential to improve financial inclusion, and also provide policy implications for regulators engaged in the adoption or extension of open banking policies.

Keywords: Open banking, FinTech, Marketplace lending, P2P lending, Big data, Customer data sharing, Data access, Data portability, Digital footprints

*House of Finance, Goethe University Frankfurt & Leibniz Institute for Financial Research SAFE. e.V. Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany. Email: nam@safe-frankfurt.de

1 Introduction

With the rapid pace of digital transformation and technological advancement, consumer financial activities such as payments, lending and trading generate large, diverse (structured and unstructured), high-dimensional, and complex sets of data, often referred to as, *Big Data* (Goldstein et al., 2021). This data can be of tremendous value to both financial and non-financial institutions since it can be used for various different purposes including but not limited to customer behavior prediction, provision of targeted and customized products, robust pricing, and risk management. Thus, the ability to collect a wealth of data from existing and new customers and the capacity to extract meaningful information therefrom can be a significant source of market power (Lambrecht and Tucker, 2015; De Ridder, 2019; Kirpalani and Philippon, 2020; Eeckhout and Veldkamp, 2022). In the financial markets, banks have long enjoyed data monopoly as they are often the sole providers of a range of financial products through which customer data is generated and collected. Consumers, on the other hand, have historically lacked rights to their own financial data and haven't reaped the same benefits.

Against this backdrop, countries worldwide are adopting open banking, a paradigm-shifting policy that redefines who owns the data. By granting data rights back to the customer, open banking enables consumers to take greater control of their own financial data and determine who to share it with. Access to such granular customer financial data can alleviate asymmetric information and adverse selection, thus helping technology-enabled firms such as FinTech lenders to leverage big data-driven algorithms to improve credit quality inference and acceptance rates (Ghosh et al., 2021). In particular, borrowers with low credit scores and/or short credit history who are often credit rationed may benefit tremendously, which can contribute to fairer and more democratic access to finance. Thus, the potential of open banking is enormous, yet its implications for consumers are still relatively less understood. The main reason is data availability. Open banking is still in the early stages and many major economies such as the U.S. still are at the discussion stage, making it more difficult to conduct a widely applicable study.

To the best of my knowledge, this study is the first to provide empirical evidence of open banking and customer-driven data sharing in the consumer credit market. Using a rich set of granular loan application data from the largest German FinTech lender, *Auxmoney*, I test predictors of borrowers' decision to share their account data and investigate the consequences of such decisions on loan application outcomes. A priori, the determinants of data sharing decision and their subsequent impact are unclear. Users with high privacy concerns may opt-out, yet potential trade-offs between credit access and privacy might in fact encourage some consumers to opt-in (Tang, 2019b). Additionally, granular transaction data facilitate better inference of credit quality and may lead to more favorable and targeted loan application outcomes, yet can also hurt borrowers if the main use of data is for price discrimination (Babina et al., 2022). Therefore, the question of who decides to share and whether or not borrowers benefit from doing so is an empirical question.

By exploiting the applicant's option to share their bank account data during the loan application, I show that the riskiest borrowers are 3.8 percentage points more likely to sign up than the safest borrowers. The likelihood of data sharing monotonically decreases as credit score improves. The results are robust to controlling for other factors such as age that might be driving the signup decision and are simultaneously correlated with credit score, but the effect is smaller (2.1 pp). Given that credit score is a risk-proxy, the negative association between credit scores and data sharing decisions is puzzling especially in the context of signaling. If we assume that, on average, higher credit score borrowers have more information that signals higher quality, theories would predict that, all else equal, these borrowers would be more willing to disclose data. However, the results show that borrowers' decision to reveal data is more nuanced. In principle, data-sharing preferences can be driven by *signaling*–the consumer's anticipated economic loss (gain) from revealing her private information–but also by the intrinsic value of privacy which is potentially heterogeneous across agents. Consumers' self-selection into data sharing is shown to depend on the magnitudes and correlation between the two components. This suggests that the data-sharing decision may often deviate from the "low types are more willing to hide" argument (Lin, 2022), which is consistent with some of my findings. This outcome suggests that the underlying mechanisms that determine one's credit score such as time and risk preferences, or impulsivity (Arya et al., 2013) may also be a driving force in one's decision to share data via the heterogeneous preferences on the intrinsic value for privacy.

Credit score is a risk proxy observable to the lender which mostly reflects payment history, yet the borrower's true type is unknown and a borrower may also possess traits that are *unobservable* but still relevant for credit risk. For instance, cash flows and risky consumption behaviors are unlikely to be reflected in one's credit score. Access to transaction data could reveal this type of information, thus some borrowers may decide not to share data in an attempt to hide negative information while borrowers with positive information may opt-in. To test this signaling channel, I divide the sample into different credit score cateogires and make an implicit assumption that the intrinsic value of privacy is relatively consistent across borrowers within each category to tease out the signaling channel of information sharing. To test this, I use two measures, 1) *ex-post* platform-provided scores which will reflect such information often unobserved in traditional risk proxies, and 2) borrowers that have always paid on time after getting a loan, to create an indicator variable, *good type*, and provide evidence that good types are indeed more likely to share data. However, the effects become statistically weaker and even insignificant among the riskiest borrowers. This finding therefore only partially confirms the existing theoretical literature which claims that under open banking high types are more likely to opt-in to send a positive signal (He et al., 2020; Babina et al., 2022).

Data sharing can also boost the probability of loan approval by up to 9 percentage points and leads to lower borrowing costs, with a reduction in the interest rate down to 2 percentage points. The results are economically sizable and statistically significant. Overall, data sharing benefits borrowers across all credit risk groups but the magnitude of the effect varies. Riskier borrowers benefit more from sharing data at the extensive margin (i.e. they enjoy a larger increase in the chance of getting a loan relative to safer borrowers). The interpretation of this result is straightforward. For borrowers who are on the margin, extra pieces of information which enhance the accuracy of consumer behavior predictions may push an application from *rejection* to *approval*. Safer borrowers, on the other hand, already have an *ex-ante* high probability of obtaining a loan, thus, their signup decision affects the loan outcome to a lesser degree (5.7 percentage points). Interestingly, however, the effect on the interest rate (intensive margin) is larger for safer borrowers.

Additionally, in the traditional credit scoring system, standard pricing variables such as credit bureau score, age, and employment status play a crucial role in determining one's creditworthiness. When customer financial data are shared, however, these standard variables explain much less variation in loan application outcomes (i.e. down to 9 percentage points in loan approval). This implies that open banking may be particularly salient for borrowers with unfavorable attributes such as low credit scores, low income, and younger demographics with shorter employment histories.

The results from the study have far-reaching policy implications. For instance, the 9 percentage points increase in loan approval, in reality, represents a significant proportion of borrowers who would have never received a loan without open banking and data sharing. The positive effects consistent across all borrower types suggest that open banking and customer-directed data sharing deepen consumer credit markets by extending credit to those who, without this policy, would not have access to credit on the platform. As Europe is the leading adopter of open banking, a study using data from one of the largest FinTech platforms in continental Europe provides a high level of validity and applicability of the findings. I spearhead this nascent literature and document the exact economic magnitudes of data sharing and open banking of the consumer loan market.

As pointed out by Babina et al., 2022, open banking possesses some similarities to credit registries (Djankov et al., 2007; Hertzberg et al., 2011), yet differs in several respects. Customer financial data often contain a richer set of information, and open banking gives the customer the option to sign up. Importantly, this type of data is more dynamic in nature. For instance, during a period of personal financial distress, a couple of negative shocks+ can impact one's credit history substantially. In credit registries, it may take a while to rebuild one's score. Therefore, credit mistakes can be costly, especially for financially constrained borrowers. Transaction data, on the other hand, could provide a more realistic and up-to-date representation of consumer financial behavior, and ultimately alleviate financing constraints for marginal borrowers.

2 Related Literature

First and foremost, I contribute to the nascent literature surrounding open banking and customer data sharing by providing empirical evidence. To date, existing studies relate mostly to theoretical predictions. Parlour et al., 2020 examines consumer welfare where banks rely on consumers' payment data and Fintech lenders compete to obtain information about their credit quality. Even though the term open banking is not directly mentioned in the paper, this setting closely resembles open banking. Using a simple and stylized model, they show that customer data portability has ambiguous effects on welfare: FinTech competition benefits consumers with weak bank affinity thus improves financial inclusion, but may hurt consumers with strong bank affinity. This study, however, does not model individuals' choice to share data, a critical component in open banking.

He et al., 2020 incorporate consumer privacy choices in their theoretical framework to endogenize the signup decision. They examine credit market competition and consumer welfare when data sharing enables FinTech lenders to better compete with banks. Their findings indicate that open banking could make the entire financial industry better off yet leave all borrowers worse off, even if borrowers could choose whether to share their data. This is because high-type borrowers suffer from exploitative targeted loans when open banking ultimately leads to large lender asymmetry favoring Fintechs, and low-types suffer due to a negative signal of opting out. Babina et al., 2022 are the first to conduct an empirical study, in particular, the role of open banking in driving innovation. Using a novel dataset of open banking policies worldwide, they document a substantial increase in FinTech venture capital investment in countries following the adoption. They also develop a simple quantitative model and demonstrate that consumer welfare depends critically on how the data is used. When customer data is used for price discrimination, it may hurt high-cost borrowers while benefiting low-cost borrowers. When they are used to provide more targeted products, however, all consumers benefit. Additionally, the higher competition banks face can also reduce *ex-ante* information production, highlighting potential policy trade-offs. Brunnermeier, Payne, et al., 2022, using a strategic decision-making model by a two-sided platform, also demonstrate that open banking can limit uncollateralized credit[1].

I build on this mainly theoretical literature in the following ways. I provide empirical evidence of open banking and customer-driven data sharing in the consumer credit market. Europe is the leading adopter of open banking, thus the granular loan application data from one of the largest FinTech platforms in continental Europe provides a high level of validity and applicability of the findings. By exploiting the borrower's explicit choice to link their bank account data during the application process, I spearhead this nascent literature and document the exact economic magnitudes of data sharing and open banking of loan application outcomes. Furthermore, I test some of the predictions from the theoretical literature by examining the probability of signing up both by *observable* risk (*ex-ante* credit scores) as well as by *unobservable* risk using imputations

[1]They model strategic decision making of a two-sided platform that provides three services: matching in the goods market, token money creation, and credit extension. They show that under open banking, agents make the opposite data portability choices to the platform. That is, buyers share their transaction histories since they help the new platform improve its matching technology while sellers do not share contract information since it allows them to more easily default if they move to the entrant platform. Thus, it eventually limits uncollateralized credit since the incumbent platform anticipates this and consequently extends less uncollateralized credit in the first place

from *ex-post* platform scores as well as the borrowers' loan payment status.

Next, I add to the literature examining the role of alternative data such as big data, and payment transactions in improving the screening efficiency and its potential benefits to borrowers. Jagtiani and Lemieux, 2019 show, by comparing loans from Lending Club[2] and banks, the correlation of credit ratings issued by the platform and FICO scores have declined substantially and alternative data-based ratings allowed some borrowers to get lower-priced credit. The work by Berg et al., 2020 using a German e-commerce company shows that information that users leave online by interacting with a website (i.e. the type of mobile device used, the access channel, etc.) can robustly predict consumer default probabilities. Gambacorta et al., 2020 investigate how different forms of credit correlate with local economic activity. Using BigTech and bank credit, they show that the use of alternative data can reduce the importance of collateral and contribute to increasing financial inclusion. Similarly, Di Maggio et al., 2022 highlight the role of alternative data in spotting "invisible primes" in the personal loan space, borrowers with low credit scores and short credit histories, but also a low propensity to default. Ghosh et al., 2021 study the impact of cashless payments by firms on loan application outcomes both at the extensive and intensive margins, using data from a large Indian FinTech lender. They find that a larger use of verifiable cashless payments vis-à-vis cash predicts a higher chance of loan approval, a lower interest rate, and a lower risk-adjusted default rate. In a similar vein, Ouyang, 2021 studies the impact of mobile cashless payment on credit provision to the underprivileged, using a sample of Chinese BigTech *Alipay* users and finds a positive impact of in-person payment flow on credit provision. The work by Ghosh et al., 2021 is the closest to my study in its empirical setting, but is different in three ways. First, I use consumer loan data rather than small business loan data. Second, in their loan application, data sharing is mandatory, thus, it does not allow for examining different characteristics among borrowers who sign up or do not sign up. Last, for the aforementioned reason, their paper does not directly connect to open banking and consumer data rights, but rather closely to the value of customer transaction data.

While there is growing evidence that the use of alternative data complements or even challenges traditional credit scoring models, one of the shortcomings in some of the existing studies is the

[2] A peer-to-peer lending platform founded in 2006 originated more than 75 bn in loans. In 2020, Lending Club acquired Radius Bank and discontinued its services to retail investors.

selection problem. If certain users self-select into the platform and this group possesses traits systematically different from the population, the estimates could suffer from bias. Put differently, it is unclear how alternative data used by platforms may have affected financial outcomes for those who use different platforms or for non-platform users in the presence of self-selection. This study alleviates such selection issues by exploiting the user's decision to share or not to share their banking data. In other words, the role of transaction data and its impact on loan application outcomes are estimated using the variation in data disclosure decisions across borrowers within the platform. Thus, this allows me to assess the differential effects of alternative data by controlling for a host of variables.

Lastly, I contribute to the growing FinTech literature, on FinTech disruption, and financial inclusion. I add to the literature discussing the role of technology in reducing disparities in access to finance. Philippon, 2016 highlights that the cost of financial intermediation by traditional players remained surprisingly expensive despite technological advances and has thus resulted in the emergence of new players. Big data are often key in their business model, and they can reduce the impact of negative prejudice in the credit market (Philippon, 2019), such as racial disparities by automating the lending processes (Howell et al., 2021). FinTech lenders also serve in areas with less bank presence, lower incomes, more minority households (De Roure et al., 2022; Erel and Liebersohn, 2022) and with higher business bankruptcy filings and unemployment rates (Cornelli et al., 2022).

These new players may directly compete with traditional lenders like banks by serving infra-marginal borrowers who value immediacy and have a higher willingness to pay (Buchak et al., 2018; Tang, 2019a) or complement bank lending by absorbing unmet demand (Gopal and Schnabl, 2020; Sheng, 2021; Avramidis et al., 2022; De Roure et al., 2022). FinTech lenders can also benefit consumers via more efficient loan application processing (Fuster et al., 2019). Importantly, FinTech loans can greatly alleviate financing constraints faced by SMEs and further improve access to bank financing by providing uncollateralized loans which can be used to acquire pledgeable assets (Beaumont et al., 2021; Eça et al., 2022). I build on this literature by providing the first empirical evidence of how access to customer bank data enabled by open banking further broadens access to finance.

The rest of the paper is organized as follows. Section 3 describes the data and provides descriptive statistics and preliminary evidence of open banking. Section 4 presents the empirical methodology and Section 5 reports the empirical results. Then, I provide potential avenues for future research and conclude in Section 6.

3 Data

3.1 Institutional setting, descriptive statistics, and evidence of open banking

This section provides the institutional background of open banking and the FinTech lender which provides data for this study, descriptive statistics, and descriptive evidence of open banking.

3.1.1 Open banking regulation

With open banking, data ownership is shifted from the bank to the customer. This allows consumers to easily access and take greater control over their own financial data. Consumers can therefore decide which third parties to share their financial data with. As of October 2021, 80 countries worldwide have at least a nascent government-led open banking effort. Most of them are still in the early-discussion phase and only 32 countries have fully implemented the policy (Babina et al., 2022)[3]. The details on open banking regulations vary substantially. While some countries mandate data sharing, others only recommend or facilitate by providing technical standards or infrastructure for data sharing[4]. The scope of customer financial data covered under open banking also varies ranging from transaction data only to savings accounts, lending, and investment records. The EU and the UK are at the forefront of this movement having fully implemented and also considering the extension of the policy. Under the revised Payment Service Directives 2 (PSD2) Access to Account (XS2A) all institutions in the EU that offer payment accounts must grant third parties (both banks and non-banks) access to the customer's transaction account information when customers consent and should also provide dedicated APIs[5] to facilitate secure access. The law came info force in

[3] Babina, Buchak, and Gornall (2022) provide an excellent description of the status of open banking worldwide.

[4] Countries with mandatory data sharing rules include Australia, Bahrain, Brazil, the EU, and Israel. In contrast, in Singapore, Malaysia, and Russia, banks are recommended to share and regulators facilitate the process by mediating industry discussion, providing technical standards for APIs, or providing infrastructure for data sharing. For more information, see Babina et al., 2022

[5] Short for "Application Programming Interfaces". It is a software intermediary that allows two applications to communicate to each other. By facilitating customer data sharing among different institutions, APIs play a critical role in securely transferring data and simplifying the customer journey, thus encouraging consumer participation in open banking. Before the introduction of open banking, it was possible for customers to share bank details but

January 2016 and had to be transposed into national law by January 2018. This makes Europe an ideal empirical setting to test how customer data sharing affects borrowers and consumer welfare[6]. PSD2 was transposed into German law on January 13, 2018. Therefore, in this study, I only include loan applications from January 13, 2018 to May 22, 2022; that is, open banking-driven customer data sharing is by law implemented throughout the entire sample period.

3.1.2 Description about the platform

The data includes approximately 18 million loan applications from the largest German FinTech lending platform, *Auxmoney*. Founded in 2007, it has originated more than EUR 2.3bn in 319,535 consumer loans between Jan 2018 and May 2022, and more than EUR 3bn since its inception, making it one of the largest consumer credit marketplace lenders in continental Europe. A prospective borrower can register on the website and enter a desired loan amount anywhere between EUR 1,000 and EUR 50,000 and is guided through an application process during which the applicant is asked to provide a set of personal information and loan details including loan purpose, employment status, income and expenses, amongst many others.

Upon completion of the loan request, the platform assesses the creditworthiness of each applicant who will be assigned a platform score class AA, A, B, C, D, E or Z if rejected. During this scoring phase, the platform, just like banks, first obtains information from credit agencies such as *Schufa*, Germany's largest credit rating agency[7]. Unlike banks which tend to filter out specific groups such as students, self-employed or temporary workers who are considered "risky borrowers"[8], the platform does not immediately exclude them, but rather makes a first stage screening decision based on the applicant's past default history. If the applicant passes the initial stage, they will move on to the next step where the use of big data and consumer digital data points are utilized. Using a technology developed by the platform over the years, loan requests are evaluated more precisely. To

without proper technological standards, the cost of data collection was simply too high and cumbersome for many consumers.

[6]In Europe, open banking is promoted by the European Commission as part of a digital agenda to open-up services, provide choice, and foster competition and innovation in the market. For more information, see https://www.openbankingeurope.eu/who-we-are/

[7]In Germany, unlike the U.S., one does not need credit history to obtain a credit score. With a simple checking account or utility bills, one will already be provided a credit score somewhere in the middle range.

[8]Under stricter banking regulations such as risk-weighted capital requirements, it is costlier to extend credit to high-risk borrowers since more capital buffer has to be set aside to service them. This can result in banks reducing lending to high-risk borrowers (Berger and Udell, 1994; Kashyap, Stein, et al., 2004; Roulet, 2018; Popov and Udell, 2012; Benetton et al., 2021).

this end, thousands of borrower characteristics as well as combinations of data points are analyzed to deliver a platform score[9]. This platform score is drawn primarily from five different sources: registration details, credit agencies, behavioral data, web data, and experience data. Registration details refer to the information provided by the applicant during the registration stage. For instance, a breakdown of sources of income and expenses such as rent or loan payments provides useful insights into personal finance planning and management. This is combined with credit agency details such as credit bureau score, and the number or type of credit cards held by a person to assess consumer behavior. In the case of repeat borrowers, past loan payment behavior on the platform is also taken into consideration. Then, the platform will also glean data from its interaction with the consumer. This relates to behavior and web data which is often unstructured yet accumulating at every stage of the application process, providing considerable analytical potential for predicting consumer behavior. One example is the length of time a person takes to reply to the confirmation email, or the browser used to access the web page. Such information extracted from a digital footprint–that is, the information generated simply by accessing or registering on a website–is shown to predict consumer risk and default probabilities and can complement existing credit bureau information (Berg et al., 2020). The entire process is automated and in the case of a successful application and loan contract agreement, loans are usually paid out within a matter of days.

Initially, the platform employed a pure peer-to-peer (P2P) lending model where the investor and the borrower are directly matched. In this form of disintermediated lending, lenders pick the individual loans they fund and the platform bears neither maturity transformation nor information collection costs. With the increasing involvement of institutional investors, many FinTech lenders including *Auxmoney* have moved towards the *marketplace* model, where the crowdfunding platform engages in information collection by assessing borrower risks to address information asymmetry among different types of lenders (retail and institutional investors) and sells diversified loan portfolios to investors (Balyuk and Davydenko, 2019; Vallee and Zeng, 2019; Braggion et al., 2020). A significant portion of these loans are now securitized[10]. The platform has been providing platform-generated scores, so-called *Auxmoney scores*, since 2013 and the scoring system has been updated around the end of 2017.

[9]For more information: https://www.auxmoney.com/faq/auxmoney-score

[10]Auxmoney has issued two asset-backed security transactions named "Fortuna Consumer Loan ABS", of about 25,000 loans with a volume of EUR 225 million in 2022 and 30,000 loans with a volume of EUR 250 million in 2021.

3.1.3 Descriptive statistics

As shown in Figure 2, the number of applications on the platform increased steadily over time except for a noticeable slow-down in 2020. Since the beginning of 2021, loan demand on the platform has experienced an uptake, reaching its peak at the end of the sample period. The number of paid-out loans (loan offers accepted by the borrowers) follows a similar trend. The number of applications are far larger than the number of paid-out loans because not all accepted loan offers are taken up by the applicant. A borrower may also file multiple applications Figure 3. There are many explanations for this, but primarily, successful applicants may do so as to compare the terms among the accepted loans. Rejected applicants, on the other hand, may come back regularly to the platform and continue applying. Thus, including multiple applications from the same applicant may introduce bias by over-weighting these borrowers who may also potentially possess characteristics that are *systematically* different from the ones who are one-time applicants. For this reason, I control for multiple applications by limiting at most one application per applicant and taking the first observation. I also exclude incomplete applications since they lack critical pieces of information necessary for the analysis. Additionally, I drop cases where the information provided is contradictory (i.e. the applicant's employment history is longer than her age). The final sample consists of 2,484,987 completed loan applications between January 13, 2018, and May 15, 2022.

[Figure 2 and Figure 3]

Table 1 provides descriptive statistics for the sample. On average, the platform receives a loan request amount of EUR 13,667 with a duration of 55 months. The average age on the platform is 38 and 65% are male applicants. 66% of these loans are approved by the platform with an average interest rate of 12%. The average credit score obtained from credit registries (*Schufa* score) is 3.12 on a scale of 4-1[11]. A median applicant has a monthly income of EUR 1,950 out of which EUR 590 is spent. A majority (93%) own a checking account(s), 63% have one or more credit card(s). 24% are homeowners and 55% possess a car(s). *Number of current* and *past loan demand* indicates the

[11] I assign a numerical value to each credit bureau score group with 4(A-D) being the highest followed by 3(E-G), 2(H-K), 1(L-M). Any applications with a score less than M are excluded from the sample. I also observe non-existing credit scores for some applicants. There are several issues with this category. For instance, when a person who recently arrived to Germany for the first time applies for a loan on the platform, her credit score will be marked as non-existent from the credit bureau since this person has no credit file registered in Germany. However, if she tries again to apply for a loan, she might get a credit score since her profile had been registered with the credit agency. This may introduce inconsistency in the data. Therefore, these observations are also dropped.

proxied number of outstanding (past) loans and on average, an applicant has 1.4 outstanding (1 fully paid) consumer loans. The main variable of interest *Signup* indicates that 8% of the applicants during the sample period shared their bank account details. Figure 4 provides a timeline of sign-up rate over time. There is a noticeable increase in open banking participation by borrowers and this is consistent among credit score groups, with the riskiest borrowers sharing more readily. Younger generations tend to be more comfortable interacting with technology which may explain the rise of the OB participation rate. However, this trend is not driven by age or credit risk factors since the average age has stayed fairly constant over time and the average credit score has slightly gone up (Figure 5). Thus, open banking is indeed being more widely adopted by consumers over time, which is suggestive of the potential consumer benefits of open banking. I explore several of these benefits in the next section. The rising trend observed in the data also confirms the theoretical predictions suggesting that the adoption of open banking might grow as the business models of the Fintech lenders improve (He et al., 2020).

[Table 1]

[Figure 4 and Figure 5]

A quick look at descriptive statistics separately for those who sign up and those who do not, seems to indicate that it is on average slightly riskier borrowers who sign up more (Credit score 3.05 vs. 3.12), younger (33 vs. 38) with relatively less income and less home ownership Table 2. The sign-up population has a higher number of outstanding consumer loans (1.56 vs. 1.33), a potential indication of having reached their maximum debt capacity, thus financially more constrained. There are four different access channels through which a new user applies for a loan: directly via the homepage, price comparison websites, brokers, or banks. These are cooperation partners of the platform. Borrowers who have taken out one or more loans from the platform are classified as repeat borrowers. As shown in Table 3, borrower characteristics differ across access channels. Applicants coming from price comparison websites have the highest average credit score (3.15) followed by repeat borrowers (3.15), directly via the homepage (2.87), brokers (2.85), and banks (2.49). Statistics for each of the five channels by the signup decision can be found in Table B.1 in Appendix B. Table 4 displays pairwise correlations of variables that will be used in the estimation.

Further descriptive evidence of open banking on loan application outcomes is shown in the next section.

[Table 2, Table 3, and Table 4]

3.1.4 Descriptive evidence of open banking on loan application outcomes

With the introduction of open banking in the EU, all financial institutions which provide payment accounts are now obliged to grant access to customer transaction account data to other banks or non-banks such as FinTech firms when the customer consents. This is a paradigm-shifting policy that redefines who owns the data. In general, banks generate customer data by offering various financial products to its customers and bear the costs of information collection (Diamond, 1984). The data are then often owned and controlled by the same institutions who then enjoy data monopoly, a source of increased market power in the digital economy (Lambrecht and Tucker, 2015; De Ridder, 2019; Kirpalani and Philippon, 2020; Eeckhout and Veldkamp, 2022). Open banking changes this data ownership dynamics and thus can promote competition by encouraging new players to enter the market with lower entry costs (He et al., 2020; Babina et al., 2022).

In particular, technology-enhanced firms equipped with tools to leverage big data can benefit significantly from open banking regimes as access to customer financial data enables them to offer more targeted products, and better assess and manage risk by predicting consumer behavior. *Auxmoney* is no exception. When the customer provides her transaction data, the platform can use a wider range of information to correctly assess borrower risk and reduce *ex-ante* the variance when the lender infers the borrower type (Ghosh et al., 2021). Thus, open banking and wider data sharing may lead to improved efficiency by non-traditional lenders such as FinTech platforms and eventually benefit consumers with lower borrowing costs and/or increased access to loans.

Figure 1: Loan application, decision, and payout process

(1) application	**(2) decision**	**(3) loan payout**
borrower information	platform score	accept/leave the offer
option to share data	approved or rejected	
	r revealed if approved	

Figure 1 shows how open banking is implemented on the platform. During the application process, loan applicants are given an option to sign-up to provide their bank account data. The process is simple as it only requires the customer to log in to their bank securely via an API (application programming interface) enabled by a third party provider. If the customer signs up, the platform will have access to the latest few months of transaction data. The platform, then, use this data along with other borrower provided data and digital footprints to calculate the *Auxmoney score*, a platform provided credit score. The platform then notifies the applicant of the loan approval decision and will be provided with an interest rate if the application is successful. In the last stage, the applicant can convert the loan or decides not to take the offer.

Figure 6 provides a first glimpse of evidence of open banking. It shows the simple average of loan acceptance rate by data sharing sign-up decision across different credit score brackets. Borrowers from the lowest credit score group (L-M) appear to benefit most from sharing data by boosting their chance of loan acceptance by 50% (from 13.8% to 21.6%). This difference is relatively small for high quality borrowers (A-D) with a 2.6 percentage point difference (from 87.6% to 90.2%). This preliminary evidence is quite intuitive as borrowers with good credit standing are mostly infra-marginal borrowers, such that an extra set of information to infer credit risk is unlikely to affect the loan approval outcome at the extensive margin. Data sharing also leads to a reduction in interest rates across borrowers of all credit scores Figure 7. Interestingly, the reduction in the interest rate is the largest for the high quality borrowers (a median value reduction from 10% to 7.8%) and for those from the lowest credit score group, the difference is smaller (median value from 15.7% to 14.6%), which indicates that, at the intensive margin, high quality borrowers benefit most from data sharing.

[Figure 6 and Figure 7]

It is important to note that borrowers who decide to share their bank account as opposed to those who do not are not randomly assigned. In fact, even within the same credit risk category, characteristics of borrowers who sign up may be *systematically* different from those who do not sign up. In the next step, therefore, I match the signup borrowers on several observable characteristics to create a comparable group, only differentiated by the signup decision to quantify the effect of open banking on loan outcomes.

4 Methodology

This section provides the regression models used for the analysis, matching methods and results, and selection bias corrections.

4.1 The drivers of the signup decision using a probit model

To estimate the determinants of open banking participation, I use a probit model as a main estimation method.

$$Sign\,up_i = X_i'\beta + G_i'\gamma + Access\,channel + Year + \epsilon_i \tag{1}$$

where i indexes an individual and $Sign\,up_i$ is an indicator variable equal to one if the person participates in open banking by signing up to share bank account data, and 0 otherwise. $Access\,channel$ and $Year$ are access channel and year dummies. X_i are borrower characteristics which include age, credit score, income, dummy variables indicating gender, main earner, homeowner, car owner, and the number of outstanding loans, as well as fully paid loans. G_i are loan characteristics such as loan amount, and loan duration. I am mainly interested in the coefficient β which measures the change in the likelihood of sharing data across different borrower traits. In particular, the main question is how one's credit risk affects the probability to share data. In other words, is it riskier borrowers or safer borrowers who share data more? To this end, the coefficient for each credit score group are of main interest. I also explore the same question for unobservable risk using the distributions of *ex-post* platform scores, conditional on observable risk (credit score). Standard errors are clustered at the individual-year level.

4.2 Matching on observables

In the next step, the effect of open banking participation on loan approval and interest rate is examined. It is important to note that the borrowers who share data may be *systematically* different from those who do not share. Therefore, using the full sample to estimate the effect of $Sign\,up$ on the probability of loan approval or the interest rate may be biased. To address this issue, a propensity score matching method (PSM) is employed to match the treated individuals (those who sign up) to a control group that are similar on observable characteristics. Age, credit bureau score, and income decile are used as matching variables. Table 5 presents matching results. I later include further matching variables such as loan amount and loan duration but the results are both quantitatively and qualitatively similar.

[Table 5]

I use the matched sample to estimate the effect of open banking on the probability of loan approval using a probit model.

$$\begin{aligned} Approved_i = \rho Sign\,up_i + \sigma_k(Sign\,up_i \times Credit\,bureau\,score_i) \\ + X_i'\beta + G_i'\gamma + Access\,channel + Year + \epsilon_i \end{aligned} \tag{2}$$

where $Approved_i$ is an indicator variable equal to 1 if the loan application is approved, and 0 otherwise. $Signup_i$ is an indicator variable equal to one if the person participates in open banking by signing up to share bank account data, and 0 otherwise. To examine if data provision has different effects across credit risk groups, I include an interaction term $Sign\,up_i \times Credit\,bureau\,score_i$. The other variables are the same as in equation (1). Main coefficients of interest are ρ and σ_k which measures, respectively, the change in the likelihood of loan approval by data sharing decision $Sign\,up_i$, and the differential effect across different credit risk categories $k = 4, 3, 2, 1$ (4 being the highest).

4.3 Heckman's two-stage correction to address selection bias

I further explore how data sharing affects the loan interest rate using the matched sample Table 6. However, interest rate is observed only if a loan is approved. In other words, for rejected loans it is unknown how the customer's decision to share data would have affected the interest rate. Since approved borrowers are not randomly selected from the population, estimating the effect of the main variables on the interest rate from the subset of loans (only *approved* loans) may introduce bias. To tackle this issue, I use the Heckman correction model to address omitted variable bias stemming from this specific sample selection problem (Heckman, 1976; 1979).

[Table 6]

Let the loan approval and interest rate functions be given by,

$$L_i^* = Z_i'\gamma + \epsilon_i,$$

$$r_i = X_i'\beta + u_i,$$

First, I introduced the basic Heckman model in a first stage, and estimate the probability of being accepted for all applicants.

$$\begin{aligned} Prob(L_i^* > 0|Z) &= Prob(\epsilon_i > -Z_i'\gamma) \\ &= \Phi(Z_i\gamma), \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (CDF) with the variable of ϵ normalized to 1. Interest rates are observed for those whose $L_i^* > 0$, so that the expected interest rate of a borrower is given by,

$$\begin{aligned} E(r_i|L_i^* > 0, Z) &= X_i'\beta + E(u_i|\epsilon_i > -Z_i'\gamma) \\ &= X_i'\beta + \theta\lambda_i, \end{aligned}$$

where $\theta = \rho\sigma_u, \lambda_i = \frac{\phi(Z_i'\gamma)}{\Phi(Z_i'\gamma)}$ and $\phi(\cdot)$ is the standard normal density function (pdf). In the second stage, the interest rate equation for those who are accepted can then be expressed as the following

$$r_i|L_i^* > 0 = X_i'\beta + \theta\hat{\lambda}_i + e_i$$

where $\theta\hat{\lambda}_i = \rho\sigma_u\hat{\lambda}_i$ represents the correction term. Here, ρ is the correlation between unobserved determinants of probability of being accepted ϵ and unobserved determinants of interest rate u, σ_u is the standard deviation of u, and $\hat{\lambda}$ is the inverse Mills ratio evaluated at $Z_i\gamma$.

More specifically, I use the following equation to estimate the effect of data sharing on the interest rate.

$$\begin{aligned} r_i = \theta\hat{\lambda}_i + \rho Sign\ up_i + \sigma_k(Sign\ up_i \times Credit\ bureau\ score_i) + X_i'\beta + G_i'\gamma \\ + Access\ channel + Year + \epsilon_i \end{aligned} \tag{3}$$

where r_i indexes interest rate and $\hat{\lambda}_i$ is the inverse Mills ratio. The other variables are the same as in equation (2). Main coefficients of interest are ρ and σ_k which measures, respectively, the change in the interest rate by data sharing decision $Signup_i$, and the differential effect across various credit risk profiles $k = 4, 3, 2, 1$ (4 being the highest). A negative θ implies a negative correlation between the error terms and proves the presence of downward selection bias. In other words, borrowers with a below average interest rate (thus safer) are selected into the approved pool of applicants. A priori, the sign of θ is unclear. The platform may prefer borrowers with high interest rates so as to maximize its returns or contrarily select relatively safer borrowers. All the other variables are the same as in equation (1).

5 Results

In this section, I present the results from equation (1), (2), (3); that is, the determinants of open banking participation by borrowers, and how this decision affect the probability of loan approval, interest rate, and suggest potential explanations for the outcomes. Then, I provide evidence of market competition and price transparency in the context of open banking. Lastly, I conduct an additional analysis on whether open banking and big-data driven consumer insights could replace traditional scoring models, and its implications

5.1 What drives borrowers to share data?

Table 7 reports the results of equation (1). Column (1) only includes credit score variables, column (2) only age, both in column (3), and column (4) reports all estimates including all borrower and loan characteristics and access channel and year dummies. In column (5)-(8), I also report OLS regression estimates.

[Table 7]

The results highlight that riskier borrowers are more likely to sign up and share their bank account data relative to safer borrowers. In economic terms, the riskiest borrower (L-M) is on average 3.8 percentage points more likely to share data than the safest borrower (A-D) (column (1)), the likelihood of data sharing monotonically decreases as credit worthiness improves. In other words, safer borrowers are more reluctant to provide account information. Often, younger borrowers adopt technology more readily than older borrowers, and the younger a borrower is, the more likely that she has short credit history, which translates into lower credit score. Thus, the higher share of open banking participation from the higher credit risk categories could be driven by age. Thus, I control for factors that might be driving the outcome variable and are simultaneously correlated with credit score in column (4). The magnitude of the main coefficients are slightly lower but still statistically significant with 2.1 percentage points difference between the riskiest and the safest borrowers. Given that credit score is a risk-proxy, the negative association between credit scores and data sharing decisions is puzzling especially in the context of signaling. If we assume that, on average, higher credit score borrowers have more information that signals higher quality, theories would predict that, all else equal, these borrowers would be more willing to disclose data. However, the data proves that borrowers' decision to reveal data is more nuanced. Consumers' self-selection into data sharing is shown to depend on the magnitudes and correlation between the two preference components– intrinsic preferences for privacy and the economic losses (or gains) from revealing the information. This suggests that the data-sharing decision may often deviate from the "low types are more willing to hide" argument (Lin, 2022), which is consistent with some of my findings. In particular, if the intrinsic value of privacy is heterogeneous across borrowers with different credit scores, a situation may arise where high credit score borrowers are more likely to self-select out of data sharing. In other words, the underlying mechanisms that determine one's credit score such as time and risk preferences, or impulsivity (Arya et al., 2013) may also be driving forces behind one's decision to share data via heterogeneous preferences on the intrinsic value for privacy apart from potential losses or gains from information revealing. Thus, a consumer's data-sharing decision can arise endogenously.

The results also highlight the negative association between female and older borrowers. Female borrowers are 0.4 percentage point less likely to sign up than male borrowers, and all else equal, a 48 year-old is 2 percentage points less likely to share data than a 38 year-old. This is in line with previous studies providing evidence that women and older individuals are more concerned with privacy issues (Goldfarb and Tucker, 2012). Individuals with more outstanding consumer loans and fully paid past loans tend to sign up more, which may indicate these are borrower who have reached their maximum debt capacity, thus financially more constraint. Additional analysis using OLS regressions exhibits highly similar estimation results both qualitatively and quantitatively, proving robustness of the findings.

5.1.1 Are good-type borrowers more likely to share data?

In the above analysis, the probability of signing up is estimated against *observable* risks such as credit scores. However, the true borrower type is unknown. Credit score is a risk proxy that is observable to the lender, yet a borrower may also possess traits that are *unobservable* but still relevant for credit risk. Existing theories would suggest that, when there is a possibility of signaling, good types will opt in more to send a positive signal. In this case, open banking and consumer data provision will benefit good types while bad types will opt out in order to benefit from single pooling (He et al., 2020; Parlour et al., 2020; Babina et al., 2022). To test this, I first divide the sample into different credit score categories to tease out the signaling channel. As discussed in the previous section, heterogeneous preferences for the intrinsic value of privacy may pose a challenge in identifying the signaling channel of data sharing. Here, I make an implicit assumption that the intrinsic value of privacy amongst borrowers with similar credit scores is fairly consistent. Under these circumstances, the variation in data sharing decisions will be primarily driven by the signaling motive. Here, I use two measures, 1) *ex-post* platform scores and 2) loan payment status to infer *good-type*[12].

[Figure 8]

First, the distribution of *ex-post* platform-provided scores for each credit score group is shown in Figure 8. If borrowers who disclose data are truly the good types conditional on credit score, I expect to see a rightward shift of the distribution for those who sign-up as the platform score is provided *ex-post* data sharing. The critical assumption here, however, is that the signup and no-signup population have *ex-ante* an identical distribution. Thus, I used the matched sample for the distribution plot. A quick look at the graphs indicates that the distribution of *ex-post* risk score indeed shifted towards the right. The distributional shift becomes less obvious for the riskier credit groups (H-K) and (L-M). There are, however, limitations to using platform scores as proxies for borrower type. For instance, the signing up decision itself may lead to a better score regardless of the borrower's true type and the information content of the data. Thus, there is a possibility

[12] Data on the payment status of the loans come from the European Data Warehouse (EDW), a Securitisation Repository designated by both the European Securities and Markets Authority (ESMA) and the Financial Conduct Authority (FCA). It was established in 2012 as the first Securitisation Repository in Europe to facilitate the collection, validation, and download of standardized loan-level data for Asset-Backed Securities and private whole loan portfolios. For more information, https://eurodw.eu/

that the rightward shift of the distribution may be partially driven by the signup decision itself, not because of positive information content which signals good-type. Considering that the *ex-post* platform scores may contain positive sign-up bias, the little change in the distributions for groups (H-K) and (L-M) indicate that there may be almost no difference in borrower type between those who share and those who do not. To test this, I regress the data sharing decision (*sign-up)* on a dummy variable *Good type* equal to 1 for the platform score 7,6,5,4,3 and 0 for 2. Platform score 1 (rejected) is excluded. Assuming both the signup and no-signup population, *ex-ante*, have the identical distribution, if the decision to signup was truly random, it is expected that there be no significant shift in the distribution. I estimate this for each credit risk group. Table 8 shows that for the highest credit category, on average, it is 12 percentage points more likely that a good type signs up, and the effect for the second best group is even larger with 14.5 percentage points. The dominance of the signaling channel additionally suggests that, within each credit score category, intrinsic preferences for privacy may be relatively constant across borrowers. The magnitude, however, becomes significantly attenuated for the riskiest borrowers with only 5 percentage points increase in the likelihood with less explanatory power.

[Table 8]

Interestingly, borrower traits differ significantly depending on the access channel. Borrowers who come via brokers or banks are substantially riskier even conditional on credit score Figure A.1-A.4 in Appendix A). This may introduce heterogeneity in the degree to which good types opt-in even within the same risk category. As a robustness check, I run two separate regressions using two samples, 1) applications via homepage and price comparison websites and 2) via brokers and banks, as they are *ex-ante* observably similar in characteristics. The results are both quantitatively and qualitatively similar for sample group 1) (Table B.2 in Appendix B). However, signup decisions by those who come via banks and brokers appear to be random and there seems to be no difference in *unobservable* risk given no statistical significance of the coefficients.

[Table 9]

As an additional test, I look at borrowers' loan payment status and redefine *Goodtype* as those who have always paid on time after getting a loan. Using this as a proxy for the borrower type, I rerun the same analysis as above. Table 9 confirms that indeed it is the good-type borrowers who are more likely to share data. In other words, users with a higher propensity to share banking data are also more likely to pay off their loans on time. However, the effect is only significant for higher credit score borrowers.

Overall, these results only partially confirm the existing theoretical predictions which claim that voluntary signup will lead only good types to opt in (He et al., 2020). Good types are more likely to sign up, but this is the case for *observably* safer borrowers. Unconditionally, it is in fact riskier borrowers who sign up more, thus, challenging the existing theory.

5.2 The effect of customer data sharing on loan approval

Table 10 provides regression results of customer-driven data sharing on loan approval from equation (2). Data sharing improves the probability of loan approval across all credit risk categories but it is the marginal borrowers (H-K and L-M) who benefit the most with about a 9 percentage-point increase in the likelihood of loan acceptance. The effect is highly statistically significant and economically sizable. Safer borrowers (A-D) can also boost their loan acceptance probability by 5.7 percentage points, which is relatively less than the other groups but still a economically meaningful magnitude.

[Table 10]

The increasing likelihood of a favorable loan outcome among high-cost borrowers can be intuitively interpreted. For instance, the chance that a person with great credit history will get a loan approved is *ex-ante* already substantially high such that the marginal benefit of bank account data for the platform might be relatively narrower than a borrower who is on the margin. For marginal borrowers, extra pieces of information, especially when it is as valuable as bank transaction data which enhance the accuracy of consumer behavior predictions, may be just what is needed to cross the threshold at the extensive margin from *rejection* to *approval.* This evidence has far reaching policy implications. These 9 percentage points among the lowest credit score group, in reality, represent a significant portion of borrowers who otherwise would have never been given a loan without open banking. Of course, volunteer participation of banks on their own record is also possible and is in fact being implemented even in jurisdictions without open banking regimes, and/or more lenient open banking regimes where banks are *recommended* to share data when customers consent. However, empirical evidence from this exercise suggests mandated data sharing by banks will be of greatest value to consumers by giving them greater control of their own data. The positive effects of bank data sharing consistent across all borrower types suggest that open banking and customer-directed data sharing are deepening credit markets by extending credit to those who, even with advanced algorithms employed by non-traditional lenders, would not have access to credit.

5.3 The effect of customer data sharing on loan interest rates

In this section, I show the effect of open banking at the intensive margin. That is, how does consumer-directed data sharing affect the interest rate conditional of being granted access to credit? Table 11 reports the results from equation (3). Column (1) reports the baseline results and column (2) presents estimates after correcting for selection bias using the Heckman two-stage selection model.

[Table 11]

Similar to the findings presented in the previous section, data sharing leads to lower borrowing costs for borrowers across all risk groups. The effect at the intensive margin, however, is the largest for the safest borrowers (A-D) with close to a 2 percentage point reduction in the interest rate, and a 0.7 percentage point reduction for the riskier borrowers (L-M). On the platform, the applicant finds out about the interest rate only if the loan request is successful. This means, the sample used for the interest rate equation (3) includes only approved loans, thus not randomly selected from the population. The negative and significant coefficient of the inverse Mill's ratio suggests that there is a downward selection bias with respect to the interest rate. That is, the platform have selected loans with interest rates lower than an average interest rate of the population and the unselected loans would have been charged higher interest rates. Being a home owner is associated with a reduction in interest rates of 2 percentage points, a magnitude similar to the value of data sharing by the top credit category even after controlling for income. In spite of the fact that these are unsecured consumer loans, this finding indicates the information content from bank account details reveals not only consumer behavior insights but also potential collateral owned by the borrower. In a nutshell, open banking is shown to deliver additional value to consumers both at the extensive and intensive margin over and above the alternative data and digital footprints used by the platform. Importantly, open banking benefits marginal borrowers more in the form of a larger increase in the probability of loan approval, yet prime borrowers enjoy a bigger reduction in interest rates when sharing data. The latter finding is somewhat surprising since there is more room for reduction for marginal borrowers given the high level of interest rate. One potential explanation comes from different mechanisms through which data sharing affects loan application outcomes and the strength of each mechanism for different borrowers. I explore this possibility in the next section.

5.4 The channels through which data sharing affects loan application outcomes

Findings from the previous sections show that consumer-driven data sharing can on average lead to positive loan application outcomes both at the extensive and intensive margin, yet to varying degrees for borrowers of different risk levels. That is, borrowers with lower credit scores benefit more at the extensive margin whereas those with higher scores enjoy a larger reduction in the interest rate. To understand the source of this heterogeneity, I explore potential mechanisms through which data disclosure affects loan application outcomes and assess the strength of each mechanism for borrowers of different risk levels. To answer this question, I adapt the model by Ghosh et al., 2021 who examine the impact of firms' verifiable cashless payments on financing outcomes. In this setting, the risk-averse financier observes the actual information content of cashless payments and updates her beliefs about the borrower. In this process, there are two complementary informational effects at play through which verifiable cashless payments affect financing outcomes: *risk-reducing effect* and *information-revealing effect*. The first mechanism is the *risk-reducing effect* which is driven by the fact that observing transaction data, regardless of the informational content (i.e., independent of borrower type), directly reduces the ex-ante risk faced by the financier by shrinking the variance of borrower type inference. The second channel, the *information-revealing effect*, comes from the content of the transaction records, which is informative about the borrower allowing the lender's posterior belief to move closer to the true borrower type. Considering that they also focus on transaction-level payment data, the logic of the model can be effectively applied to this study. The major difference, however, is that in their setting, data sharing is mandatory. Thus, the variation across borrowers regarding data disclosure comes from the degree of verifiable cashless payments vis-à-vis cash while my study features a more parsimonious setting in which the essence of verifiable cashless payment is captured by the consumer's binary choice or sharing or not sharing.

A risk-averse lender has a CARA utility function with absolute risk aversion ρ. The lender's risk-aversion allows the role of data in reducing the uncertainty of borrower type inference (Farboodi and Veldkamp, 2020). Since the lender does not know the borrower type, this must be inferred based on its prior and the data provided by the borrower. The financing price bid by the competitive lender for the borrower is

$$p = E[z|\mathcal{I}] - \frac{\rho}{2} Var[z|\mathcal{I}] \tag{4}$$

where $\mathcal{I}$ is the lender's information set. The price quoted by the lender can be interpreted as loan approval or interest rate in the consumer credit market setting. For a borrower type z, the expected informed financing price becomes

$$p(z) = -\underbrace{\frac{\rho}{2}\frac{1}{\tau_z + D}}_{\text{Risk reducing}} + \underbrace{\frac{D}{\tau_z + D} z}_{\text{Information revealing}} + \frac{\tau_z}{\tau_z + D}\mu \tag{5}$$

where $D = 0$ or 1 is determined by the binary data sharing choice by the borrower. τ_z is the lender's prior. Contrast this to the uninformed financing price

$$p_\mu = \mu - \frac{\rho}{2}\frac{1}{\tau_z} \tag{6}$$

which is the counterfactual price that the lender offers if no borrower shares data. The last two terms in (5) is a weighted average of the borrower's true type z and the lender's prior μ, compared to the simple prior (6). The expected price improvement (= improvement in financing outcomes) from data sharing for borrower type z becomes

$$\Delta p(z) = p(z) - p_\mu \tag{7}$$

To test the strength of each mechanism, I first compute the change in the platform score driven by data sharing decisions. Since the platform score is given only after data sharing, I create a control group similar in observable characteristics who choose not to share. If data sharing leads to an increase in the platform score, this can be the result of a combination of the two effects. As shown in the second panel of Table B.2 and Figure A.3, A.4, borrowers who come via brokers and banks are among the riskiest borrowers referred by brokers and banks. In such cases, it is much less likely that the transaction data reveals positive information about the borrower. Importantly, for this group of borrowers, the dispersion of the platform score distribution is highly similar for both share and non-sharers. This is another indication of the dominance of the risk-reducing effect for these borrowers since the risk-reducing effect is always positive and independent of type. In other words, the information-revealing channel is fairly weak for these borrowers, and any improvement in the platform score is likely to be driven by the *risk-reducing effect*. On the contrast, those who come via price comparison websites and directly via the homepage are among safer borrowers so data sharing is likely to lead to the combination of the two effects. I contrast the results from these two groups in Table 12 and Table 13 and show that on average, between 19% to 27% of the improvement in the platform score stem from the risk-reducing effect.

5.5 Will open banking and big data-driven consumer insights replace traditional credit scoring models?

Big data and algorithm-driven lending exploits different information in addition to standard pricing variables. Buchak et al., 2018 show that standard variables for predicting interest rates, such as FICO and loan-to-value ratio (LTV), explain substantially less variation in interest rates of FinTech lenders relative to non-FinTech lenders. Even within FinTech loans, access to customer financial data may further reduce the explicability of traditional pricing variables in loan application outcomes. In other words, if consumer data indeed provide valuable information in predicting the borrower's credit risk, it is expected that standard variables used in traditional credit scoring models such as credit bureau score, age, income, or employment status, among others, would play less of a role in determining one's creditworthiness. Thus, I expect that the variation in the probability of loan approval and the interest rate would be less explained by these standard variables.

Residual distribution plots from equation (2) and (3) display the dispersion of what is not explained by the model for the probability of loan approval (Figure 9) and the interest rate (Figure 10) by credit score. At first glance, as expected, the dispersion is more apparent for the signup population. It is noteworthy that the dispersion is more pronounced for the riskier groups, which confirms the results from the earlier sections. Prime borrowers already possess desirable traits that are deemed creditworthy and they are reflected in the standard variables. Extra information obtained from account data, thus, is less likely to change the loan application outcome at the extensive margin. The visually apparent bunching around -0.02 from the safest credit profile (A-D) in Figure 9 suggests that prime borrowers benefit substantially more from alternative data at the intensive margin.

[Figure 9 and Figure 10]

To translate this into economic terms, I extend the logic by Buchak et al., 2018 and compare the R^2 between regressions where *Signup* takes a value or 1 and 0 in the other. Table 14 shows that indeed standard pricing variables explain the chance of loan approval substantially less when the customer share data. The economic magnitude is quite sizable, and it can be down to about 8 percentage points. This result is encouraging especially for those who are traditionally considered risky by banks; that is, younger, asset-light, low-income with shorter employment history. The difference in R^2 is also present for borrowing costs but relatively smaller, 2.7 percentage points Table 15. Overall, these results indicate that open banking can facilitate non-bank lenders to extend credit to those who are considered risky with thin credit profiles, such as low credit score, short credit history and contribute to a fairer and more democratic access to finance.

[Table 14 and Table 15]

6 Robustness checks

Throughout this study, the effects of data sharing are estimated using the matched samples. That is, the applicants who share their data are matched to a group of individuals who do not share data who are otherwise similar in observable characteristics so as to minimize the omitted variable bias. However, these two groups may still possess unobservable characteristics that are fundamentally different from one another, leading to biased results. To tackle this potential issue, I exploit the within individual variation in the data sharing decision and test the robustness of the main findings. On the platform, applicants often file multiple applications to compare the offers. For instance, a borrower can test out different loan amounts or maturities, which will result in varying probabilities of loan approval and interest rates. During this course, the borrower may decide to 1) never share, 2) always share or 3) sometimes share their bank account data. The robustness of the main findings is tested using the third case where the applicant do not share their data in some applications, but do share in the others. To ensure that borrower characteristics that may affect the loan applications outcomes do not vary, individual-day fixed effect are employed. The sample consists of 34,610 applications from 6,380 users.

[Table 16 and Table 17]

Table 16 and 17 show the robustness test both on the probability of loan approval and the interest rate. The results are qualitatively similar to the main results. Compared to prime borrowers, riskier borrowers enjoy a higher increase in the probability of loan approval with the middle-tier borrowers benefiting the most with over a 9 p.p. jump in the loan approval chance. For the lowest credit score profile, however, the effect is smaller compared to the main results. The effects on the interest rate are highly robust to different specifications both quantitatively and qualitatively.

7 Conclusion

This paper provides empirical evidence of open banking, a policy that empowers consumers by giving them greater control of their own data, for FinTech borrowers in the consumer credit market. Leveraging highly granular loan application-level data from the largest German online lender, I show that the rate of open banking participation is higher among riskier borrowers, but conditional on *observable* risk, good types share more. The signup rate among good types, however, differs depending on the level of *observable* risk. I also provide evidence that customer-directed data sharing can benefit all borrowers both at the extensive and intensive margin. The effect, however, varies across different credit risk levels. Borrowers enjoy a higher chance of loan approval and riskier borrowers gain most in this regard. Borrowers also receive lower interest rates, and safer borrowers get the largest reduction in the interest rate. Notably, with customer data, standard pricing variables such as credit score, age, and income explain loan application outcomes substantially less. Overall, this study shows that open banking is indeed being more widely adopted, and is bringing substantial benefits to consumers via big data and algorithm-driven FinTech lenders.

There are a few issues I leave for future research. Open banking may generate unintended consequences as it limits banks' ability to extract rent from the customer data. As open banking is still relatively a new policy, future research may empirically test these predictions, that is, the second-order effects of open banking via its impact on incumbents' profitability and its interactions with consumers over time. Additionally, this study is related to the effects of open banking and customer-driven data sharing in the lending market. The implications of open banking, however, may be markedly different across various financial services, which need to be taken into consideration to assess the aggregate impact.

References

Arya, S., Eckel, C., & Wichman, C. (2013). Anatomy of the credit score. *Journal of Economic Behavior & Organization*, *95*, 175–185.

Avramidis, P., Mylonopoulos, N., & Pennacchi, G. G. (2022). The role of marketplace lending in credit markets: Evidence from bank mergers. *Management Science*, *68*(4), 3090–3111.

Babina, T., Buchak, G., & Gornall, W. (2022). Customer data access and fintech entry: Early evidence from open banking. *Available at SSRN.*

Balyuk, T., & Davydenko, S. (2019). Reintermediation in fintech: Evidence from online lending. *Michael J. Brennan Irish Finance Working Paper Series Research Paper*, (18-17).

Beaumont, P., Tang, H., & Vansteenberghe, E. (2021). *The role of fintech in small business lending.* Working paper.

Benetton, M., Eckley, P., Garbarino, N., Kirwin, L., & Latsi, G. (2021). Capital requirements and mortgage pricing: Evidence from basel II. *Journal of Financial Intermediation*, *48*, 100883.

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, *33*(7), 2845–2897.

Berger, A. N., & Udell, G. F. (1994). Did risk-based capital allocate bank credit and cause a "credit crunch" in the united states? *Journal of Money, credit and Banking*, *26*(3), 585–628.

Braggion, F., Manconi, A., Pavanini, N., Zhu, H., et al. (2020). The value of "new" and "old" intermediation in online debt crowdfunding.

Brunnermeier, M., Payne, J. et al. (2022). Platforms, tokens, and interoperability. *Unpublished Working Paper.*

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

Cornelli, G., Frost, J., Gambacorta, L., & Jagtiani, J. (2022). The impact of fintech lending on credit access for us small businesses.

De Ridder, M. (2019). Market power and innovation in the intangible economy.

De Roure, C., Pelizzon, L., & Thakor, A. (2022). P2P lenders versus banks: Cream skimming or bottom fishing? *The Review of Corporate Finance Studies*, *11*(2), 213–262.

Di Maggio, M., Ratnadiwakara, D., & Carmichael, D. (2022). *Invisible primes: Fintech lending with alternative data.* National Bureau of Economic Research.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, *51*(3), 393–414.

Djankov, S., McLiesh, C., & Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, *84*(2), 299–329.

Eça, A., Ferreira, M. A., Prado, M. P., & Rizzo, A. E. (2022). The real effects of fintech lending on smes: Evidence from loan applications.

Eeckhout, J., & Veldkamp, L. (2022). *Data and market power*. National Bureau of Economic Research.

Erel, I., & Liebersohn, J. (2022). Can fintech reduce disparities in access to finance? evidence from the paycheck protection program. *Journal of Financial Economics*, *146*(1), 90–118.

Farboodi, M., & Veldkamp, L. (2020). Long-run growth of financial data technology. *American Economic Review*, *110*(8), 2485–2523.

Fuster, A., Plosser, M., Schnabl, P., & Vickery, J. (2019). The role of technology in mortgage lending. *The Review of Financial Studies*, *32*(5), 1854–1899.

Gambacorta, L., Huang, Y., Li, Z., Qiu, H., & Chen, S. (2020). Data vs collateral.

Ghosh, P., Vallee, B., & Zeng, Y. (2021). Fintech lending and cashless payments. *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Goldfarb, A., & Tucker, C. (2012). Shifts in privacy concerns. *American Economic Review*, *102*(3), 349–53.

Goldstein, I., Spatt, C. S., & Ye, M. (2021). Big data in finance. *The Review of Financial Studies*, *34*(7), 3213–3225.

Gopal, M., & Schnabl, P. (2020). The rise of finance companies and fintech lenders in small business lending. *NYU Stern School of Business*.

He, Z., Huang, J., & Zhou, J. (2020). *Open banking: Credit market competition when borrowers own the data*. National Bureau of Economic Research.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, 153–161.

Hertzberg, A., Liberti, J. M., & Paravisini, D. (2011). Public information and coordination: Evidence from a credit registry expansion. *The Journal of Finance*, *66*(2), 379–412.

Howell, S. T., Kuchler, T., Snitkof, D., Stroebel, J., & Wong, J. (2021). *Racial disparities in access to small business credit: Evidence from the paycheck protection program*. National Bureau of Economic Research.

Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the lendingclub consumer platform. *Financial Management*, *48*(4), 1009–1029.

Kashyap, A. K., Stein, J. C. et al. (2004). Cyclical implications of the basel II capital standards. *Economic Perspectives-Federal Reserve Bank Of Chicago*, *28*(1), 18–33.

Kirpalani, R., & Philippon, T. (2020). *Data sharing and market power with two-sided platforms.* National Bureau of Economic Research.

Lambrecht, A., & Tucker, C. E. (2015). Can big data protect a firm from competition? *Available at SSRN 2705530*.

Lin, T. (2022). Valuing intrinsic and instrumental preferences for privacy. *Marketing Science*, *41*(4), 663–681.

Ouyang, S. (2021). Cashless payment and financial inclusion. *Available at SSRN 3948925*.

Parlour, C. A., Rajan, U., & Zhu, H. (2020). When fintech competes for payment flows. *Available at SSRN 3544981*.

Philippon, T. (2016). *The fintech opportunity*. National Bureau of Economic Research.

Philippon, T. (2019). *On fintech and financial inclusion*. National Bureau of Economic Research.

Popov, A., & Udell, G. F. (2012). Cross-border banking, credit access, and the financial crisis. *Journal of International Economics*, *87*(1), 147–161.

Roulet, C. (2018). Basel iii: Effects of capital and liquidity regulations on european bank lending. *Journal of Economics and Business*, *95*, 26–46.

Sheng, T. (2021). The effect of fintech on banks' credit provision to smes: Evidence from china. *Finance Research Letters*, *39*, 101558.

Tang, H. (2019a). Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, *32*(5), 1900–1938.

Tang, H. (2019b). The value of privacy: Evidence from online borrowers. *Available at SSRN 3880119*.

Vallee, B., & Zeng, Y. (2019). Marketplace lending: A new banking paradigm? *The Review of Financial Studies*, *32*(5), 1939–1982.

Tables

Table 1: **Summary Statistics**

Variable	N	Mean	S.D.	Min	p25	p50	p75	Max
LOAN INFORMATION:								
Credit requested	2484987	13,669.71	12,979.14	1,000.00	4,000.00	10,000.00	20,000.00	50,000.00
Credit offered	1630862	12,143.43	10,631.26	1,000.00	4,000.00	9,500.00	18,000.00	50,000.00
Interest rate	1630862	0.12	0.04	0.00	0.08	0.13	0.15	0.20
Platform score (max 7, min 1)	2484987	2.81	1.81	1.00	1.00	2.00	4.00	7.00
Credit score group (max 4, min 1)	2484987	3.12	0.85	1.00	3.00	3.00	4.00	4.00
Loan duration	2484987	55.10	24.33	0.00	36.00	60.00	84.00	84.00
Application accepted (D)	2484987	0.66	0.47	0.00	0.00	1.00	1.00	1.00
Flagged for quality check (D)	2484987	0.28	0.45	0.00	0.00	0.00	1.00	1.00
Bank account detail shared (D)	2484987	0.08	0.26	0.00	0.00	0.00	0.00	1.00
BORROWER CHARACTERISTICS:								
Age	2484981	37.74	12.62	18.00	27.00	36.00	47.00	69.00
Female	2484987	0.35	0.48	0.00	0.00	0.00	1.00	1.00
Main earner (if married) (D)	2484987	0.62	0.49	0.00	0.00	1.00	1.00	1.00
No. current loan demand	2308526	1.35	1.47	0.00	0.00	1.00	2.00	68.00
No. past loan demand	2308526	1.04	1.78	0.00	0.00	0.00	1.00	76.00
No. months of employment	2326408	78.03	100.23	0.00	11.00	37.00	103.00	839.00
No. months of living in current place	2399902	90.26	112.39	0.00	15.00	49.00	122.00	839.00
INCOMES AND EXPENSES:								
Total income	2484987	3,053.48	129,527.70	0.00	1,450.00	1,950.00	2,610.00	42,949,673.00
Monthly net salary income	2484979	2,593.52	101,982.50	0.00	1,300.00	1,800.00	2,370.00	42,949,673.00
Child support income	2484979	127.21	6,476.70	0.00	0.00	0.00	204.00	8,000,000.00
Other income	2484979	194.48	17,219.65	0.00	0.00	0.00	0.00	18,801,800.00
Total expenses	2484987	740.65	30,550.15	0.00	304.00	590.00	933.00	42,949,673.00
Housing related expenses	2484069	481.37	29,188.60	0.00	180.00	415.00	645.00	42,949,673.00
Credit installments expenses	2484069	166.37	3,428.36	0.00	0.00	0.00	216.00	5,000,000.00
Other expenses	2484069	23.88	938.69	0.00	0.00	0.00	0.00	1,000,000.00
Insurance expenses	2484069	49.77	599.19	0.00	0.00	0.00	0.00	545,454.00
Child support expenses	2484069	19.18	566.81	0.00	0.00	0.00	0.00	600,000.00
ASSETS:								
Credit-card owner	2484987	0.63	0.48	0.00	0.00	1.00	1.00	1.00
EC-card owner	2484987	0.93	0.25	0.00	1.00	1.00	1.00	1.00
Home-owner	2484987	0.24	0.43	0.00	0.00	0.00	0.00	1.00
Car-owner	2484987	0.55	0.50	0.00	0.00	1.00	1.00	1.00

* conditional on being accepted. This table presents summary statistics for the sample. The sample period ranges from Jan 13, 2018 to May 22, 2022. (D) = dummy variable.

Table 2: **Data sharing signup vs. no-signup applicants**

variable	signup	no signup	pvalue
Credit requested	12,608.19	13,756.30	0.00
Credit offered	10,823.04	12,257.02	0.00
Interest rate	0.10	0.12	0.00
Platform score (max 7, min 1)	2.84	2.44	0.00
Credit score group (max 4, min 1)	3.05	3.12	0.00
Loan duration	52.47	55.32	0.00
Application accepted (D)	0.69	0.65	0.00
Flagged for quality check (D)	0.25	0.28	0.00
Bank account detail shared (D)	1.00	0.00	0.00
Age	33.78	38.07	0.00
Female	0.34	0.35	0.00
Main earner (if married) (D)	0.62	0.62	0.00
No. months of employment	62.32	79.36	0.00
No. months of living in current place	81.61	90.98	0.00
No. current loan demand	1.56	1.33	0.00
No. past loan demand	1.28	1.02	0.00
Total income	2,658.31	3,086.17	0.17
Total expenses	728.44	741.79	0.86
Credit-card owner	0.78	0.62	0.00
EC-card owner	0.96	0.93	0.00
Home-owner	0.19	0.25	0.00
Car-owner	0.60	0.55	0.00

This table presents summary statistics separately for the borrowers who share data, *Signup*, and for those who opt out, *No signup*. Monetary unit in EUR. (D) = Dummy variable.

Table 3: **Descriptive statistics by access channels**

	Access channel				
Variable	Directly via homepage	Repeat Borrower	Price comp. website	Broker	Bank
Credit requested	8,280.71	11,331.79	14,887.52	11,437.16	4,772.35
Credit offered	7,094.55	10,957.68	12,762.22	11,269.84	4,718.49
Interest rate	0.12	0.09	0.11	0.14	0.13
Platform score (max 7, min 1)	2.39	4.62	3.04	1.79	1.55
Credit score group (max 4, min 1)	2.87	3.15	3.21	2.85	2.49
Loan duration	26.14	53.61	56.68	62.89	68.68
Application accepted (D)	0.57	0.97	0.72	0.37	0.31
Flagged for quality check (D)	0.26	0.39	0.30	0.21	0.14
Bank account detail shared (D)	0.11	0.17	0.08	0.03	0.03
Age	34.23	43.17	38.35	37.06	29.54
Female	0.38	0.40	0.34	0.38	0.22
Main earner (if married) (D)	0.12	0.32	0.66	0.69	0.87
No. months of employment	59.06	115.18	81.49	69.39	31.28
No. months of living in current place	87.76	127.98	92.71	80.75	16.02
No. current loan demand	1.20	1.84	1.39	1.23	0.73
No. past loan demand	1.02	1.81	1.03	1.12	0.53
Total income	1,700.00	2,001.50	2,000.00	1,750.00	1,832.00
Total expenses	660.00	903.50	600.00	450.00	786.04
Credit-card owner	0.40	0.65	0.69	0.38	0.93
EC-card owner	0.83	0.97	0.96	0.82	0.99
Home-owner	0.16	0.30	0.27	0.15	0.16
Car-owner	0.49	0.67	0.61	0.26	0.37

This table presents summary statistics separately by access channel. (D) = Dummy variable.

Table 4: **Pairwise correlation**

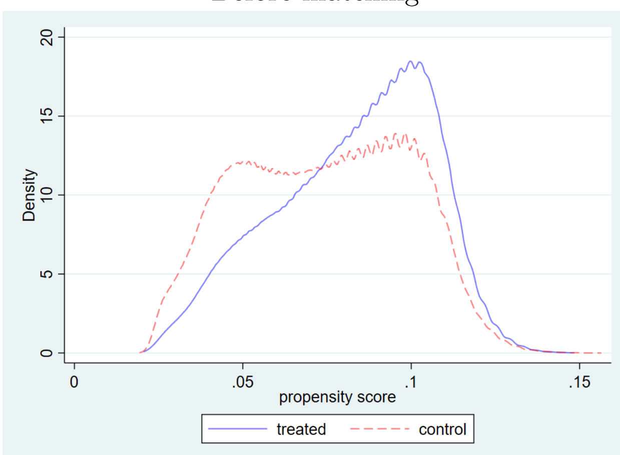
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) Age	1.000															
(2) Signup	-0.089	1.000														
(3) Credit score	0.366	-0.020	1.000													
(4) Income decile	0.259	-0.001	0.157	1.000												
(5) Interest rate	-0.369	-0.099	-0.399	-0.240	1.000											
(6) Loan amount requested ('000)	0.152	-0.022	0.105	0.332	0.003	1.000										
(7) Loan duration (yr)	0.112	-0.029	0.001	0.086	0.040	0.414	1.000									
(8) Marital status	0.428	-0.037	0.127	0.115	-0.115	0.071	0.066	1.000								
(9) Female	0.024	-0.007	0.021	-0.271	0.009	-0.105	-0.007	0.104	1.000							
(10) Main earner	-0.067	0.003	-0.062	-0.041	0.069	-0.010	0.109	-0.018	-0.017	1.000						
(11) Home owner	0.323	-0.036	0.222	0.280	-0.349	0.145	0.062	0.147	-0.054	-0.055	1.000					
(12) Car owner	0.169	0.028	0.162	0.242	-0.185	0.122	0.019	0.098	-0.044	0.041	0.194	1.000				
(13) Credit card owner	0.051	0.086	0.124	0.164	-0.150	0.095	0.051	0.009	-0.045	0.309	0.084	0.199	1.000			
(14) Checking account owner	0.048	0.035	0.092	0.086	-0.083	0.048	0.065	0.016	0.003	-0.003	0.062	0.162	0.249	1.000		
(15) No. current loan demand	0.144	0.041	-0.065	0.248	-0.068	0.115	0.084	0.083	-0.026	0.029	0.127	0.148	0.068	0.057	1.000	
(16) No. past loan demand	0.147	0.037	-0.065	0.191	-0.047	0.100	0.074	0.084	-0.029	0.009	0.084	0.081	0.019	0.024	0.310	1.000

Table 5: **Matched variables and matching results**

	Mean Treated	Mean Control	Mean diff p-value
Age	33.755	33.755	1.000
Credit score	3.0501	3.0501	1.000
Income decile	5.3667	5.3667	1.000

This table shows t-tests for the null hypothesis of equal means for the treated and control groups. Each of the 188,149 applications is matched one-to-one with the closest propensity score ("control"), using the borrower characteristics variables, age, credit score, and income decile as predictors of signing up to share the bank account information.The sample includes 338,924 loan applications.

Before matching



After matching

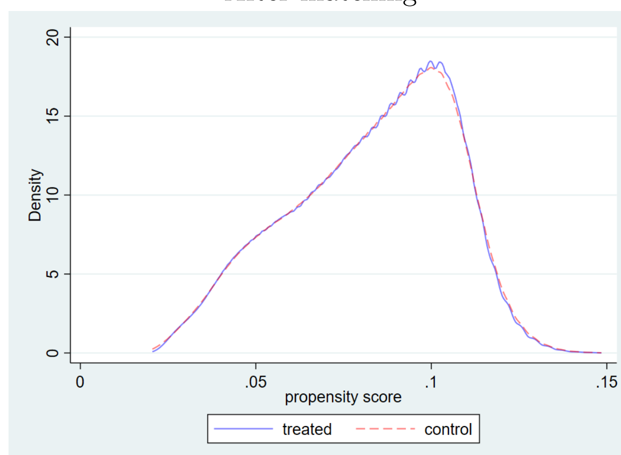
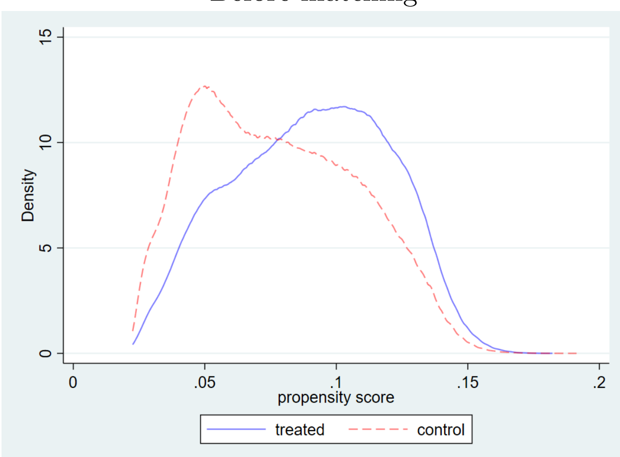


Table 6: **Matched variables and matching results**

	Mean Treated	Mean Control	Mean diff p-value
Age	36.089	36.089	0.999
Credit score	3.2837	3.2837	0.994
Income decile	5.8076	5.8076	1.000

This table shows t-tests for the null hypothesis of equal means for the treated and control groups. Each of the 129,628 *approved* applicants is matched one-to-one with *approved* applicants (to ensure interest rate information is available for all units) using the closest propensity score ("control"), using the borrower characteristics variables, age, credit score, and income decile as predictors of signing up to share the bank account information. The sample includes 259,256 loan applications.

Before matching



After matching

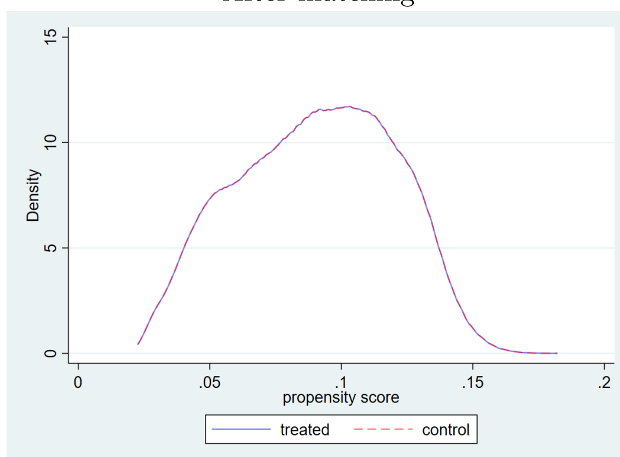


Table 7: **What characterizes borrowers who share data?**

	Probit (marginal effects)				LPM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age ('0)		-0.019***	-0.017***	-0.019***		-0.018***	-0.016***	-0.018***
		(0.0001)	(0.0002)	(0.0002)		(0.0001)	(0.0002)	(0.0002)
Income decile			0.001***	0.000			0.000***	-0.000
			(0.0001)	(0.0001)			(0.0001)	(0.0001)
Credit score (A-D) (base)								
Credit score (E-G)	0.022***		0.009***	0.009***	0.023***		0.010***	0.009***
	(0.0004)		(0.0004)	(0.0004)	(0.0004)		(0.0004)	(0.0004)
Credit score (H-K)	0.038***		0.019***	0.018***	0.037***		0.018***	0.015***
	(0.0005)		(0.0005)	(0.0006)	(0.0005)		(0.0005)	(0.0006)
Credit score (L-M)	0.038***		0.015***	0.021***	0.034***		0.010***	0.013***
	(0.0010)		(0.0009)	(0.0010)	(0.0008)		(0.0008)	(0.0009)
Loan amount requested (ln)				-0.011***				-0.011***
				(0.0002)				(0.0002)
Loan duration (ln)				-0.003***				-0.006***
				(0.0004)				(0.0004)
Female				-0.004***				-0.005***
				(0.0004)				(0.0004)
Main earner				0.008***				0.010***
				(0.0005)				(0.0005)
No. current loan demand				0.006***				0.008***
				(0.0001)				(0.0002)
No. past loan demand				0.005***				0.006***
				(0.0001)				(0.0001)
Home owner				-0.008***				-0.008***
				(0.0005)				(0.0004)
Car owner				0.012***				0.009***
				(0.0004)				(0.0004)
Cooperation partner=Homepage (base)								
Cooperation partner=Repeat				0.113***				0.068***
				(0.0032)				(0.0026)
Cooperation partner=Price comp. website				-0.074***				-0.063***
				(0.0012)				(0.0011)
Cooperation partner=Broker				-0.103***				-0.093***
				(0.0013)				(0.0012)
Cooperation partner=Bank				-0.123***				-0.126***
				(0.0014)				(0.0017)
constant					0.024***	0.111***	0.095***	0.252***
					(0.0010)	(0.0011)	(0.0012)	(0.0024)
Dummy	Year	Year	Year	Year	Year	Year	Year	Year
Cluster (Zipcode-Year)	X	X	X	X	X	X	X	X
N	2,512,185	2,524,191	2,512,178	2,261,576	2,512,185	2,524,191	2,512,178	2,261,576
R2	0.0640	0.0731	0.0739	0.1040	0.035	0.040	0.040	0.057

This table reports the results of probit and LPM regressions modeling the probability that a borrower shares bank data using the full sample in equation (1). The coefficients (1-3) are marginal effects at means. Clustered standard errors are in parentheses. Column (1)-(3) reports pseudo R2 and (4)-(6) adjusted R2.

Table 8: **Are good types more likely to share data? (using platform scores)**

	Matched sample			
Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
	(1)	(2)	(3)	(4)
Goodtype (=1 if platform score 7-3)	0.121***	0.145***	0.111***	0.051**
	(0.006)	(0.004)	(0.007)	(0.021)
Controls	Y	Y	Y	Y
Cluster (Zipcode-Year)	Y	Y	Y	Y
N	109,781	104,889	30,168	3,794
Pseudo R2	0.1628	0.1561	0.1899	0.1547

This table shows the probability of data sharing among *Goodtype* borrowers (a dummy variable equal to 1 if the platform score is 7, 6, 5, 4, or 3, = 0 for 2). A probit model with the matched sample is used for the analysis. Each column represents a risk group with (A-D) being the highest and (L-M) being the lowest credit score group.

Table 9: **Are good types more likely to share data? (using loan payment status)**

Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
	(1)	(2)	(3)	(4)
Goodtype (=1 if never been in arrears)	0.061***	0.070***	0.056**	0.046
	(0.016)	(0.014)	(0.026)	(0.079)
Controls	Y	Y	Y	Y
Cluster (Zipcode-Year)	Y	Y	Y	Y
N	11,345	10,040	2,417	174
Pseudo R2	0.0707	0.0805	0.0873	0.1527

This table shows the probability of data sharing among *Goodtype* borrowers (a dummy variable equal to 1 if the loan has never been in arrears). A probit model is used for the analysis. Each column represents a risk group with (A-D) being the highest and (L-M) being the lowest credit score group. The sample in this regression includes 23,976 loans from the asset-backed security transaction, "Fortuna Consumer Loan ABS (2021)", with a volume of EUR 225 million.

Table 10: **The effect of sign-up decision on loan approval using matched sample**

	Probit (marginal effects)			LPM		
	(1)	(2)	(3)	(4)	(5)	(6)
Signup	0.090***	0.136***	0.057***	0.145***	0.072***	0.024***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
Signup × Credit score (A-D)* (Base)						
Signup × Credit score (E-G)		0.040***	0.007		0.065***	0.051***
		(0.004)	(0.008)		(0.003)	(0.003)
Signup × Credit score (H-K)		0.047***	0.031***		0.069***	0.058***
		(0.005)	(0.004)		(0.004)	(0.004)
Signup × Credit score (L-M)		0.044***	0.029***		0.044***	0.031***
		(0.005)	(0.004)		(0.006)	(0.006)
Credit score (A-D) (Base)						
Credit score (E-G)		-0.204***	-0.136***		-0.226***	-0.194***
		(0.002)	(0.003)		(0.002)	(0.002)
Credit score (H-K)		-0.471***	-0.388***		-0.471***	-0.422***
		(0.002)	(0.005)		(0.003)	(0.003)
Credit score (L-M)		-0.732***	-0.711***		-0.706***	-0.624***
		(0.003)	(0.005)		(0.004)	(0.004)
Age		0.007***	0.005***		0.006***	0.004***
		(0.000)	(0.000)		(0.000)	(0.000)
Income decile		0.027***	0.013***		0.022***	0.011***
		(0.000)	(0.000)		(0.000)	(0.000)
Loan amount requested (ln)			0.023***			0.025***
			(0.001)			(0.001)
Loan duration (ln)			-0.123***			-0.113***
			(0.002)			(0.002)
Marital status			-0.000			0.002**
			(0.001)			(0.001)
Female			0.029***			0.029***
			(0.002)			(0.001)
Main earner			0.026***			0.018***
			(0.002)			(0.001)
Home owner			0.061***			0.044***
			(0.002)			(0.002)
Car owner			0.063***			0.065***
			(0.002)			(0.001)
No. current loan demand			0.021***			0.021***
			(0.001)			(0.001)
No. past loan demand			0.006***			0.006***
			(0.000)			(0.000)
Cooperation partner=Homepage (Base)						
Cooperation partner=Repeat			-0.000			-0.092***
			(0.000)			(0.003)
Cooperation partner=Price comp. website			-0.253***			-0.283***
			(0.001)			(0.002)
Cooperation partner=Broker			-0.514***			-0.495***
			(0.004)			(0.003)
Cooperation partner=Bank			-0.426***			-0.434***
			(0.008)			(0.005)
constant				0.621***	-0.244***	0.336***
				(0.006)	(0.007)	(0.008)
Dummy	Year	Year	Year	Year	Year	Year
Cluster (Zipcode-Year)	X	X	X	X	X	X
N	376298	376298	338924	376298	376298	338924
R2	0.0207	0.2309	0.3317	0.027	0.268	0.338

This table reports the results from equation (2), the effect of customer's decision to share bank account data (Signup) on the probability of loan approval using the matched sample.
*The coefficients (1-3) show marginal effects at means. Clustered standard errors are in parentheses. Column reports pseudo R2 and (4)-(6) adjusted R2. Marginal effects for interaction terms for the non-linear model are calculated manually by estimating the change in the dependent variable evaluated at each credit score.

Table 11: **The effect of sign-up decision on interest rates (accounting for the selection bias)**

	Matched sample	
	(1)	(2)
Inverse mill's ratio		**-0.0172***** (0.0008)
Signup	-0.0165*** (0.0002)	**-0.0184***** (0.0003)
Signup × Credit score (A-D) (Base)		
Signup × Credit score (E-G)	0.0036*** (0.0003)	0.0028*** (0.0003)
Signup × Credit score (H-K)	0.0084*** (0.0004)	0.0070*** (0.0004)
Signup × Credit score (L-M)	0.0141*** (0.0009)	0.0118*** (0.0009)
Age ('0)	-0.0162*** (0.0004)	-0.0188*** (0.0005)
Age ('0) squared	0.0010*** (0.0001)	0.0012*** (0.0001)
Credit score (A-D) (Base)		
Credit score (E-G)	0.0199*** (0.0002)	0.0242*** (0.0003)
Credit score (H-K)	0.0318*** (0.0003)	0.0426*** (0.0006)
Credit score (L-M)	0.0408*** (0.0006)	0.0606*** (0.0011)
Income decile	-0.0018*** (0.0000)	-0.0022*** (0.0000)
Loan amount requested (ln)	0.0072*** (0.0001)	0.0063*** (0.0001)
Loan duration (ln)	0.0053*** (0.0002)	0.0091*** (0.0003)
Marital status	0.0007*** (0.0001)	0.0008*** (0.0001)
Female	-0.0026*** (0.0002)	-0.0034*** (0.0002)
Main earner	-0.0023*** (0.0002)	-0.0014*** (0.0002)
Home owner	-0.0180*** (0.0002)	-0.0202*** (0.0002)
Car owner	-0.0046*** (0.0002)	-0.0047*** (0.0002)
Credit card owner	-0.0058*** (0.0002)	-0.0057*** (0.0002)
Checking account owner	-0.0022*** (0.0004)	-0.0022*** (0.0004)
No. current loan demand	-0.0012*** (0.0001)	-0.0013*** (0.0001)
No. past loan demand	-0.0001*** (0.0000)	-0.0001*** (0.0000)
Cooperation partner=Homepage		
Cooperation partner=Repeat	-0.0215*** (0.0006)	-0.0215*** (0.0006)
Cooperation partner=Price comp. website	0.0057*** (0.0003)	0.0057*** (0.0003)
Cooperation partner=Broker	0.0181*** (0.0004)	0.0179*** (0.0004)
Cooperation partner=Bank	0.0171*** (0.0008)	0.0167*** (0.0008)
constant	0.0916*** (0.0012)	0.0944*** (0.0012)
Dummy	Year	Year
Cluster (Zipcode-Year)	X	X
N	248,631	248,631
Adjusted R2	0.4320	0.4330

This table reports the results of equation (3) which explores the effect of customer's decision to share bank account data (*Signup*) on the interest rate conditional on loan approval, using the matched sample. Column (2) shows the results after correcting for selection bias.

Table 12: **The change in the platform score after data sharing by credit score group: higher risk group**

	Credit score			
	(A-D)	(E-G)	(H-K)	(L-M)
Signup	0.2294***	0.1613***	0.0863***	0.0342***
	(0.0417)	(0.0184)	(0.0141)	(0.0129)
Dummy	Year	Year	Year	Year
Controls	X	X	X	X
Cluster (Zipcode-Year)	X	X	X	X
N	4322	10463	7704	3231
Adjusted R2	0.2780	0.3158	0.2306	0.0721

This table reports the results of data sharing (*Signup*) on the change in the platform score by credit score group, using the matched sample of the borrowers who come via brokers and banks who are among the riskiest borrowers.

Table 13: **The change in the platform score after data sharing by credit score group: lower risk group**

	Credit score			
	(A-D)	(E-G)	(H-K)	(L-M)
Signup	0.8517***	0.7835***	0.4563***	0.1574***
	(0.0105)	(0.0081)	(0.0086)	(0.0115)
Dummy	Year	Year	Year	Year
Controls	X	X	X	X
Cluster (Zipcode-Year)	X	X	X	X
N	109108	133047	52515	10717
Adjusted R2	0.3144	0.2936	0.2840	0.4092

This table reports the results of data sharing (*Signup*) on the change in the platform score by credit score group, using the matched sample of the borrowers who come directly via the homepage or via price comparison websites who are among the safer borrowers.

Table 14: **R2 of different specifications explaining loan approval: signup vs. no signup**

		Specification				Matched sample		**Signup R2 - No signup R2**
	Controls	Dummy	Cluster	Model	Zip-Year FE	Signup	No signup	
(1)	Schufa	Year	Zip-Year	Probit	N	0.1715	0.1971	-0.0256
	Schufa	N	Zip-Year	LPM	Y	0.1801	0.2172	-0.0371
(2)	(1) + loan amount, duration	Year	Zip-Year	Probit	N	0.1770	0.2291	-0.0521
	(1) + loan amount, duration	N	Zip-Year	LPM	Y	0.1831	0.2414	-0.0583
(3)	(2) + age, income, marital status gender, main earner	Year	Zip-Year	Probit	N	0.2259	0.2733	-0.0474
	(2) + age, income, marital status gender, main earner	N	Zip-Year	LPM	Y	0.2198	0.2795	-0.0597
(4)	All	Year	Zip-Year	Probit	N	0.2327	0.3022	-0.0695
	All	N	Zip-Year	LPM	Y	0.2269	0.3049	-0.078

This table shows R2 using different specifications for equation (2) using the matched sample from Table (5).

Table 15: **R2 of different specifications explaining interest rate: signup vs. no signup**

		Specification				Matched sample		**Signup R2 - No signup R2**
	Controls	Dummy	Cluster	Model	Zip-Year FE	Signup	No signup	
(1)	Schufa	Access channel	Zip-Year	LPM	N	0.1832	0.2248	-0.0416
	Schufa	N	Zip-Year	LPM	Y	0.1974	0.2275	-0.0301
(2)	(1) + loan amount, duration	Access channel	Zip-Year	LPM	N	0.2062	0.2773	-0.0711
	(1) + loan amount, duration	N	Zip-Year	LPM	Y	0.2396	0.2410	-0.0014
(3)	(2) + age, income, marital status gender, main earner	Access channel	Zip-Year	LPM	N	0.2963	0.3113	-0.015
	(2) + age, income, marital status gender, main earner	N	Zip-Year	LPM	Y	0.3047	0.3004	0.0043
(4)	All	Access channel	Zip-Year	LPM	N	0.3512	0.3784	-0.0272
	All	N	Zip-Year	LPM	Y	0.3344	0.3473	-0.0129

This table shows R2 using different specifications for equation (3) using the matched sample from Table (6).

Table 16: **The effect of data sharing decision on loan approval (Robustness)**

	Credit score			
	(A-D)	(E-G)	(H-K)	(L-M)
Signup	0.035***	0.094***	0.092***	0.042***
	(0.012)	(0.008)	(0.008)	(0.011)
Controls	Y	Y	Y	Y
Individual-day FE	Y	Y	Y	Y
N	4766	15922	11313	2609
Adjusted R2	0.068	0.077	0.089	0.080

This table shows the effect of data sharing on the probability of loan approval. I exploit the within variation in the data sharing decision from borrowers who file multiple applications on the same day.

Table 17: **The effect of data sharing decision on interest rate (Robustness)**

	Credit score			
	(A-D)	(E-G)	(H-K)	(L-M)
Signup	-0.017***	-0.014***	-0.007***	-0.007**
	(0.001)	(0.001)	(0.001)	(0.003)
Controls	Y	Y	Y	Y
Individual-day FE	Y	Y	Y	Y
N	3523	5625	1580	135
Adjusted R2	0.217	0.181	0.098	0.068

This table shows the effect of data sharing on the interest rate. I exploit the within variation in the data sharing decision from borrowers who file multiple applications on the same day.

Figures

Figure 2: Number of applications (14-D moving average)

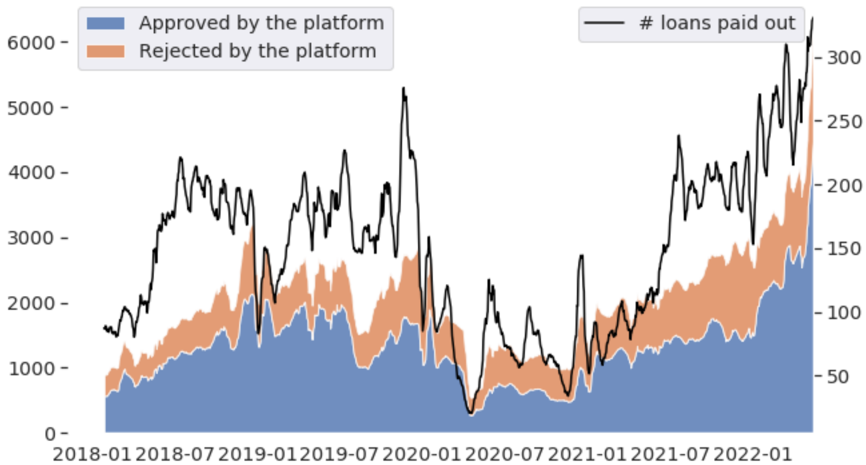


Figure 3: Number of applications per unique borrower

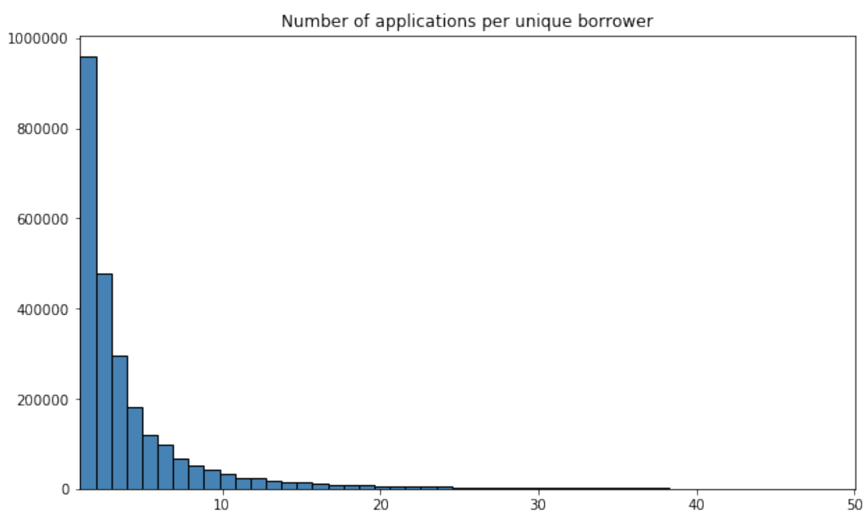
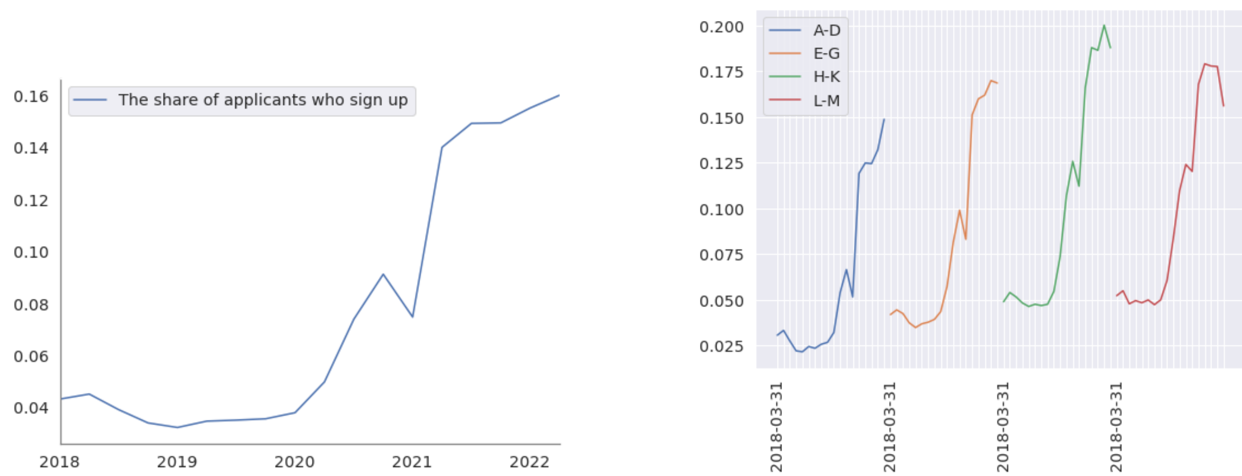
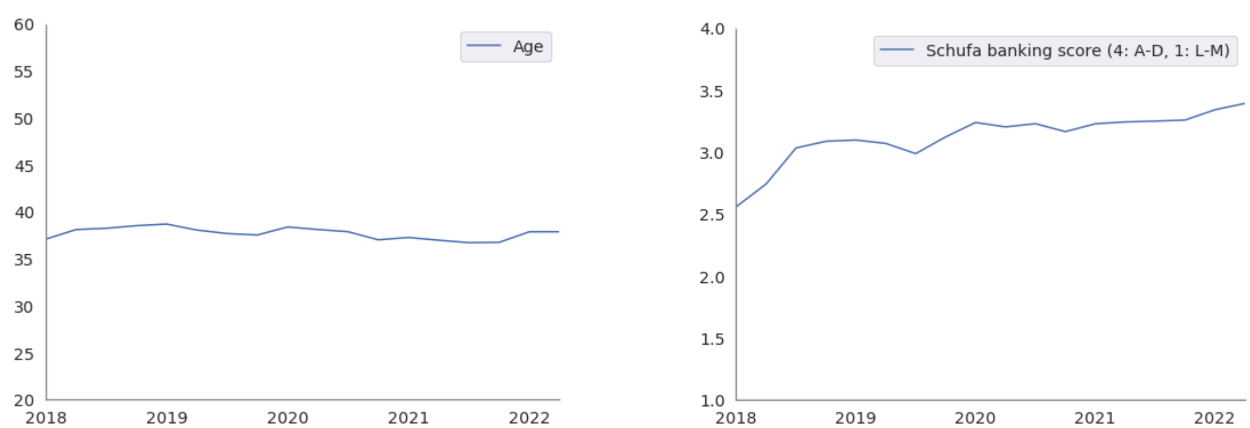


Figure 4: Open banking adoption over time (overall vs. by credit score)



Notes: The left panel shows the overall share of applicants who sign up. The right panel represents the share by credit score over time.

Figure 5: Average age and Credit score



Notes: The left panel shows the average age. The right panel represents the average credit score.

Figure 6: Loan acceptance rate by data sharing signup decision

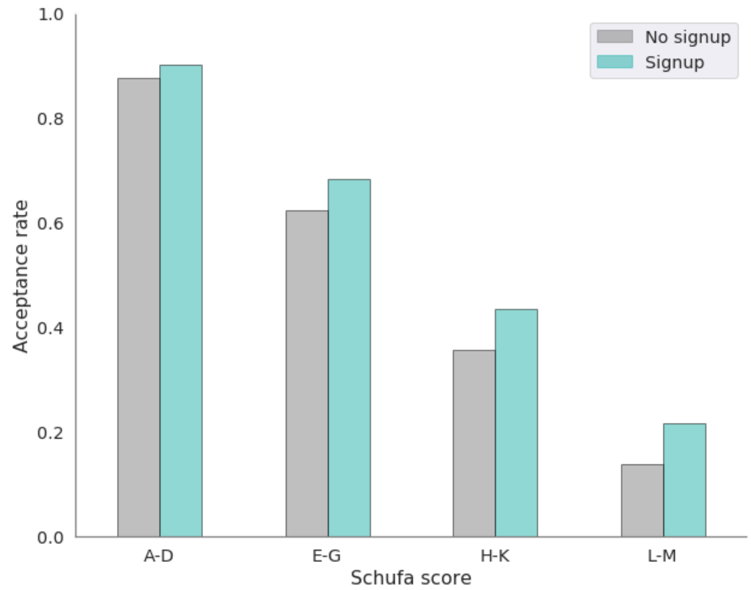


Figure 7: Loan interest rate by data sharing signup decision

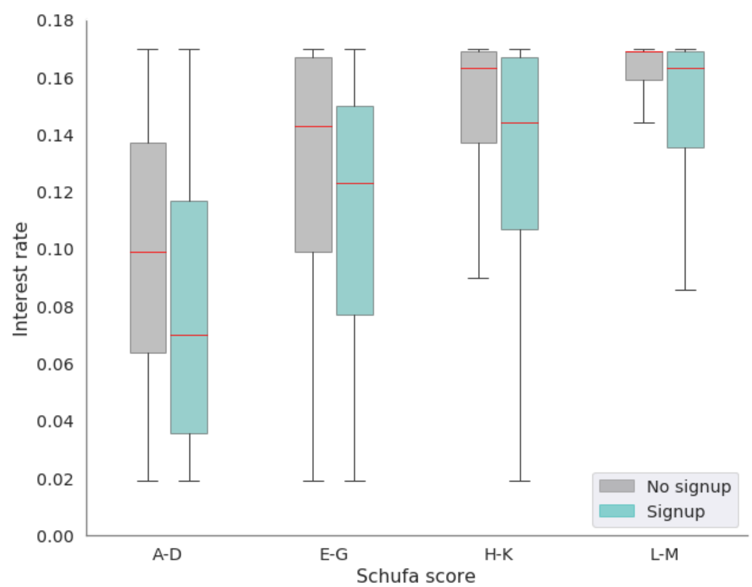
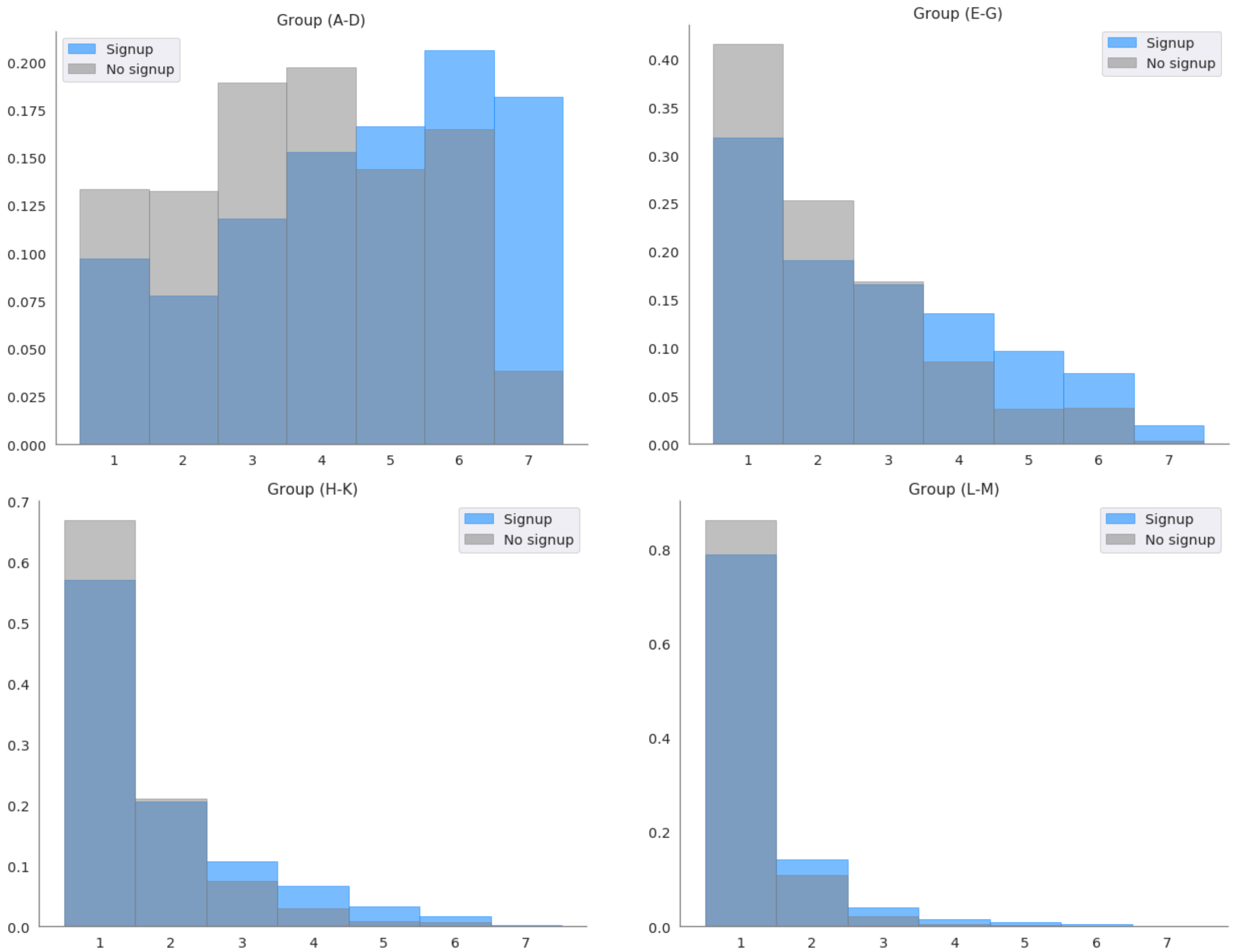
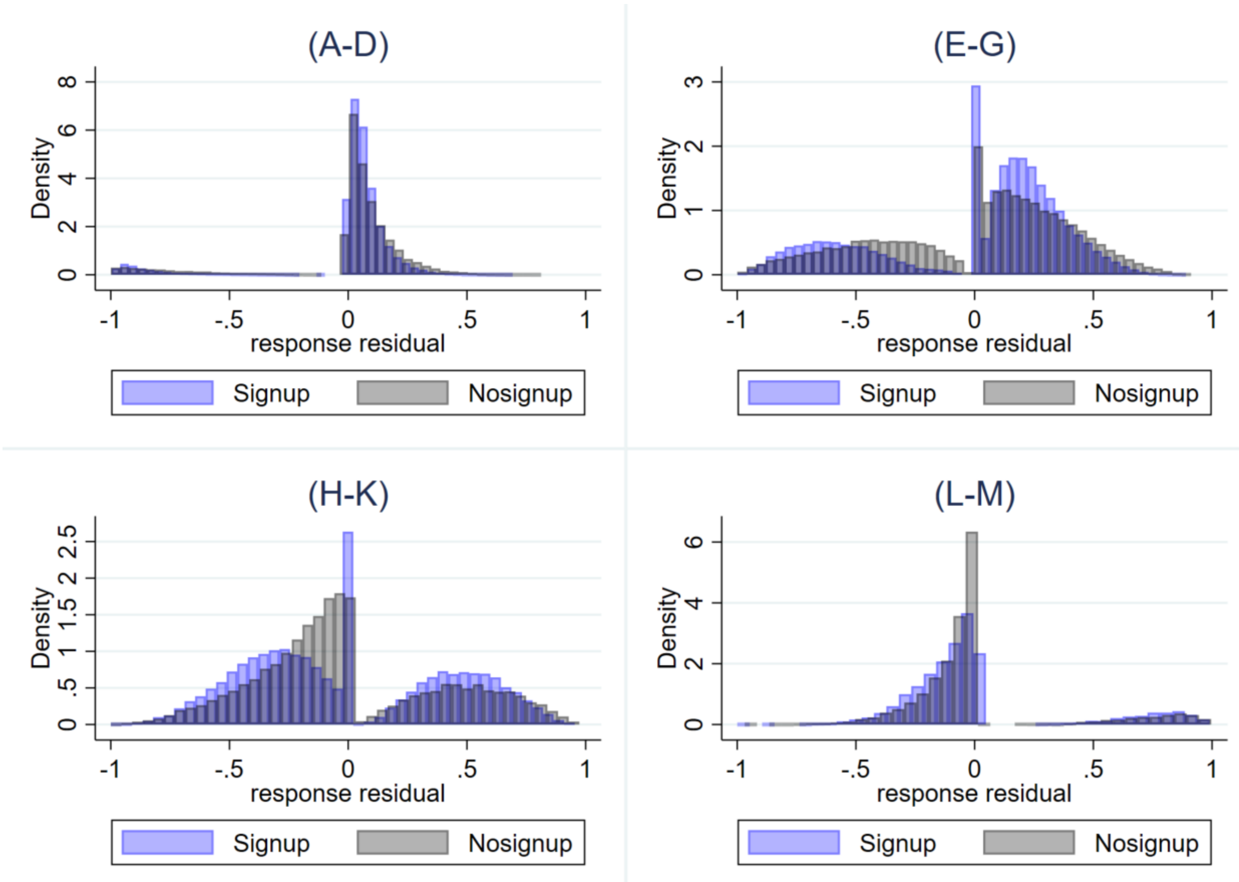


Figure 8: Distribution of platform-provided credit score by signup decision (using the matched sample)



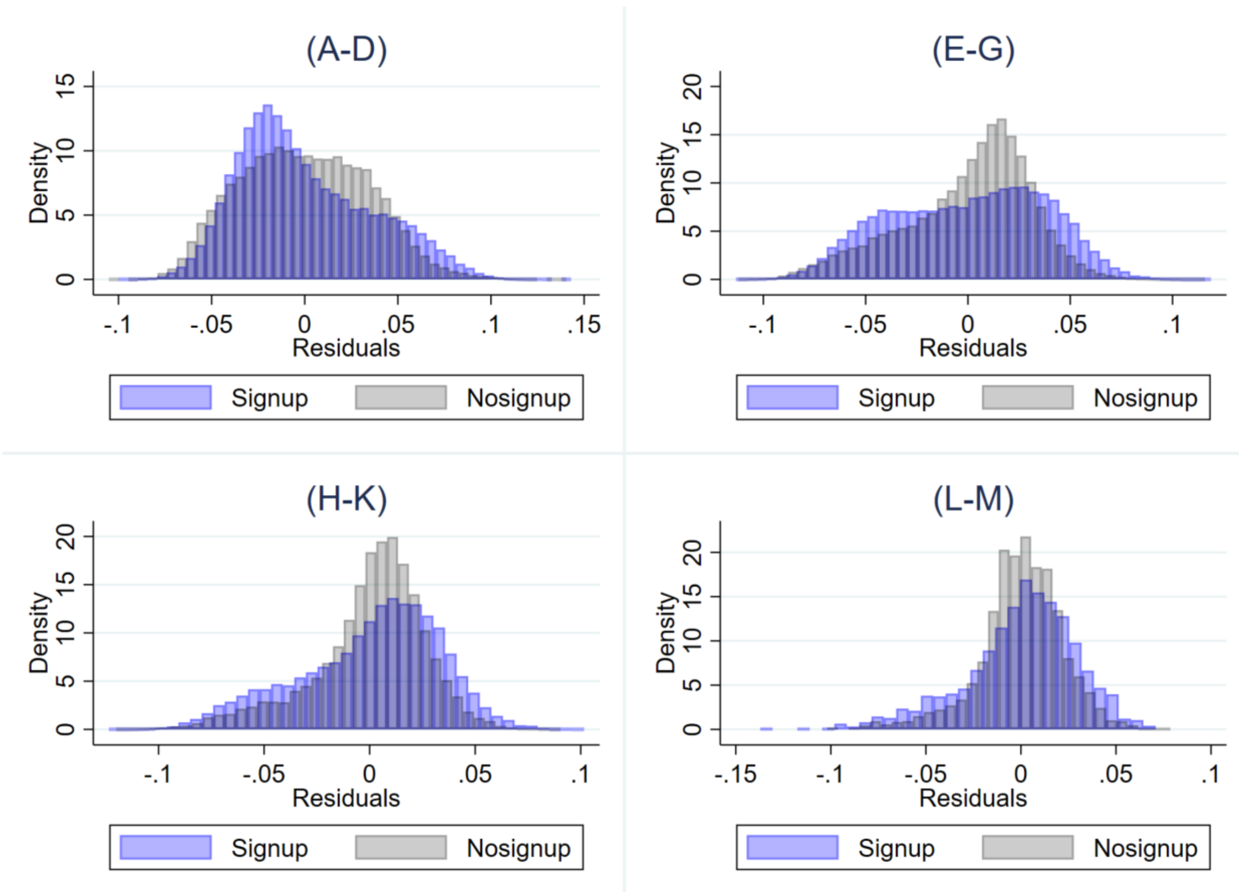
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes the share of applicants.

Figure 9: Residual plots for loan approval by credit score



Note: Each panel represents the distribution of residuals from equation (2) (the effect of open banking on loan approval) for each credit score group. Residuals are computed by estimating the model using the generalized linear model (GLM) with family binomial and probit link.

Figure 10: Residual plots for loan interest rate regression by credit score

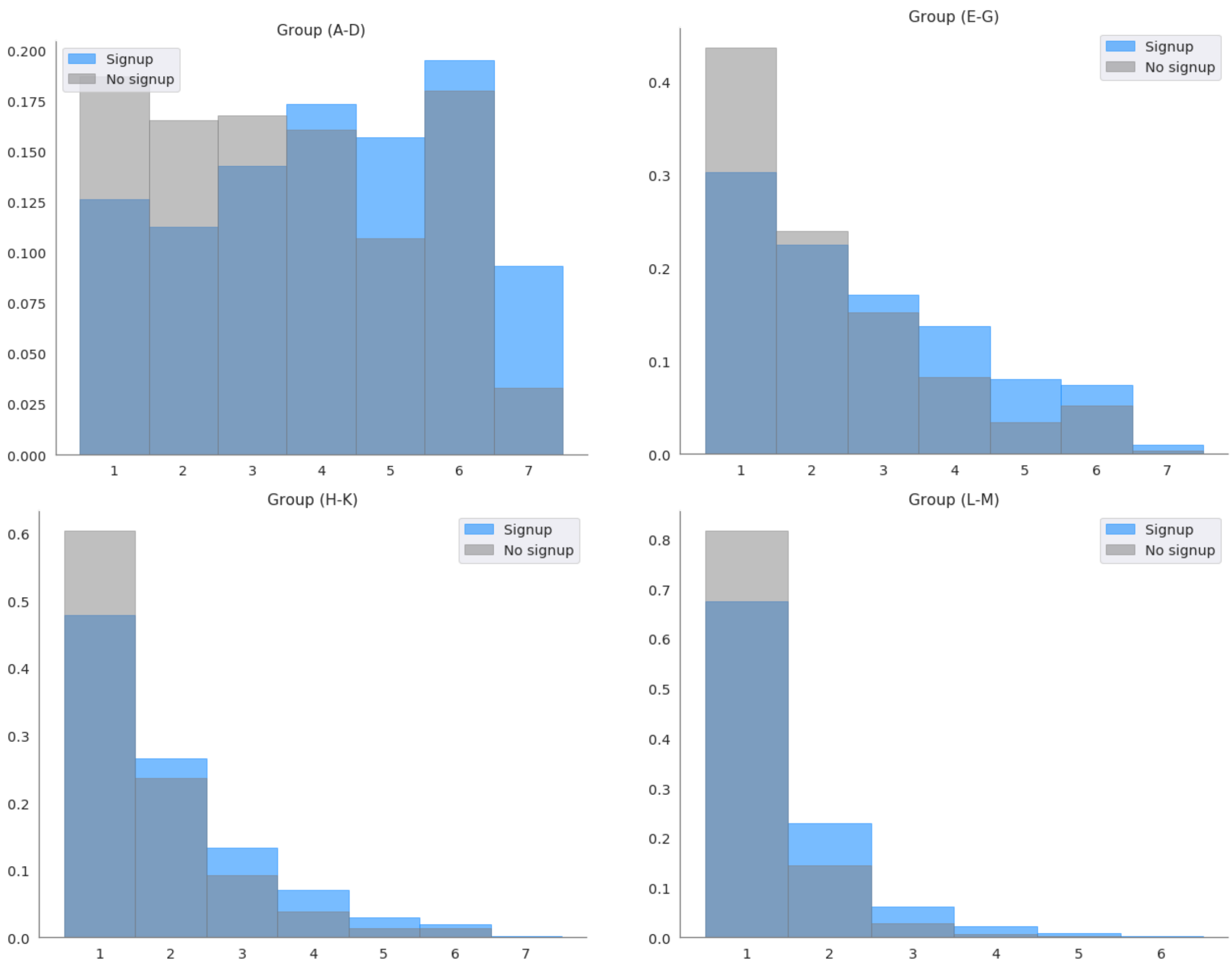


Note: Each panel represents the distribution of residuals from equation (3) (the effect of open banking on interest rate) for each credit score group.

Appendix

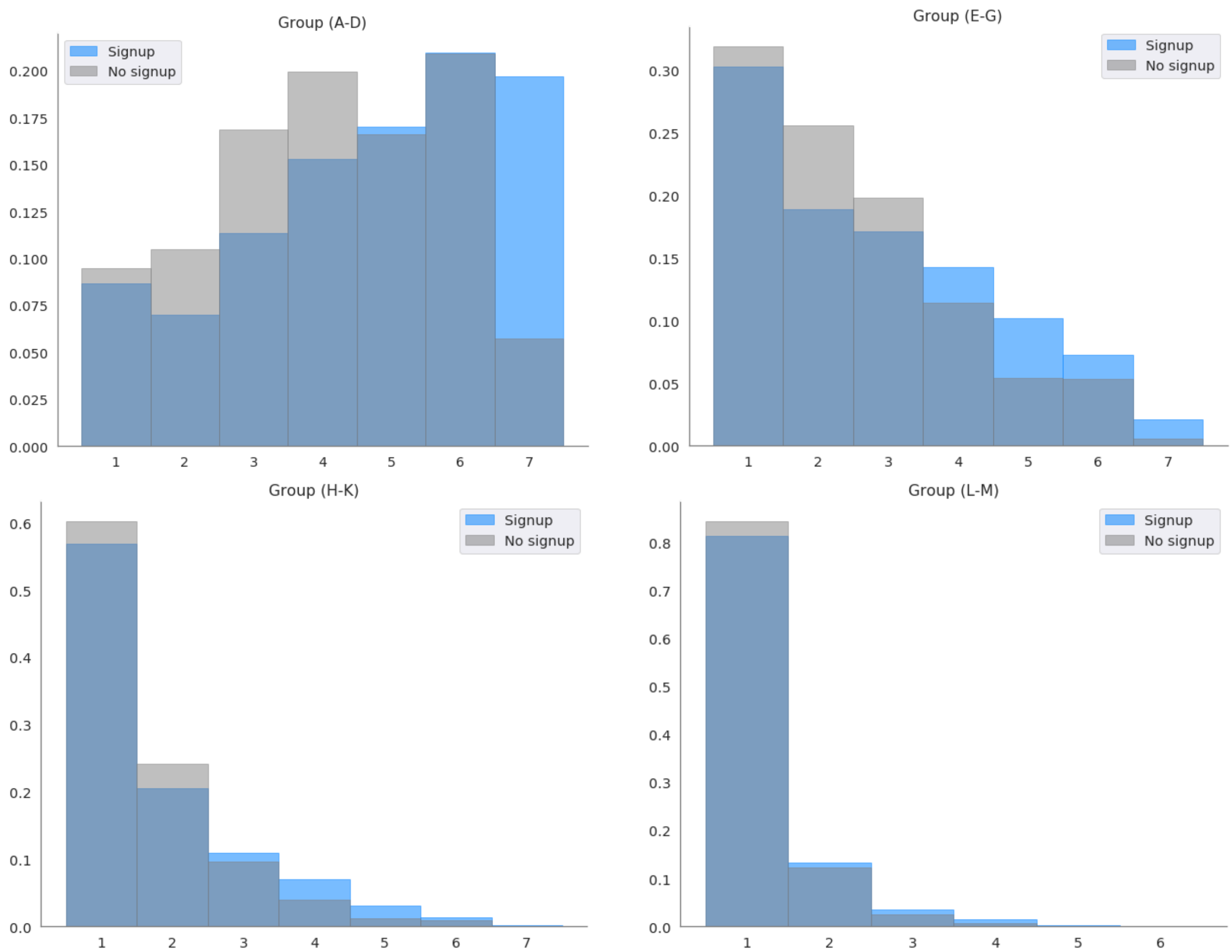
A Figures

Figure A.1: Distribution of ex-post platform-provided credit score by signup decision (Access directly via homepage)



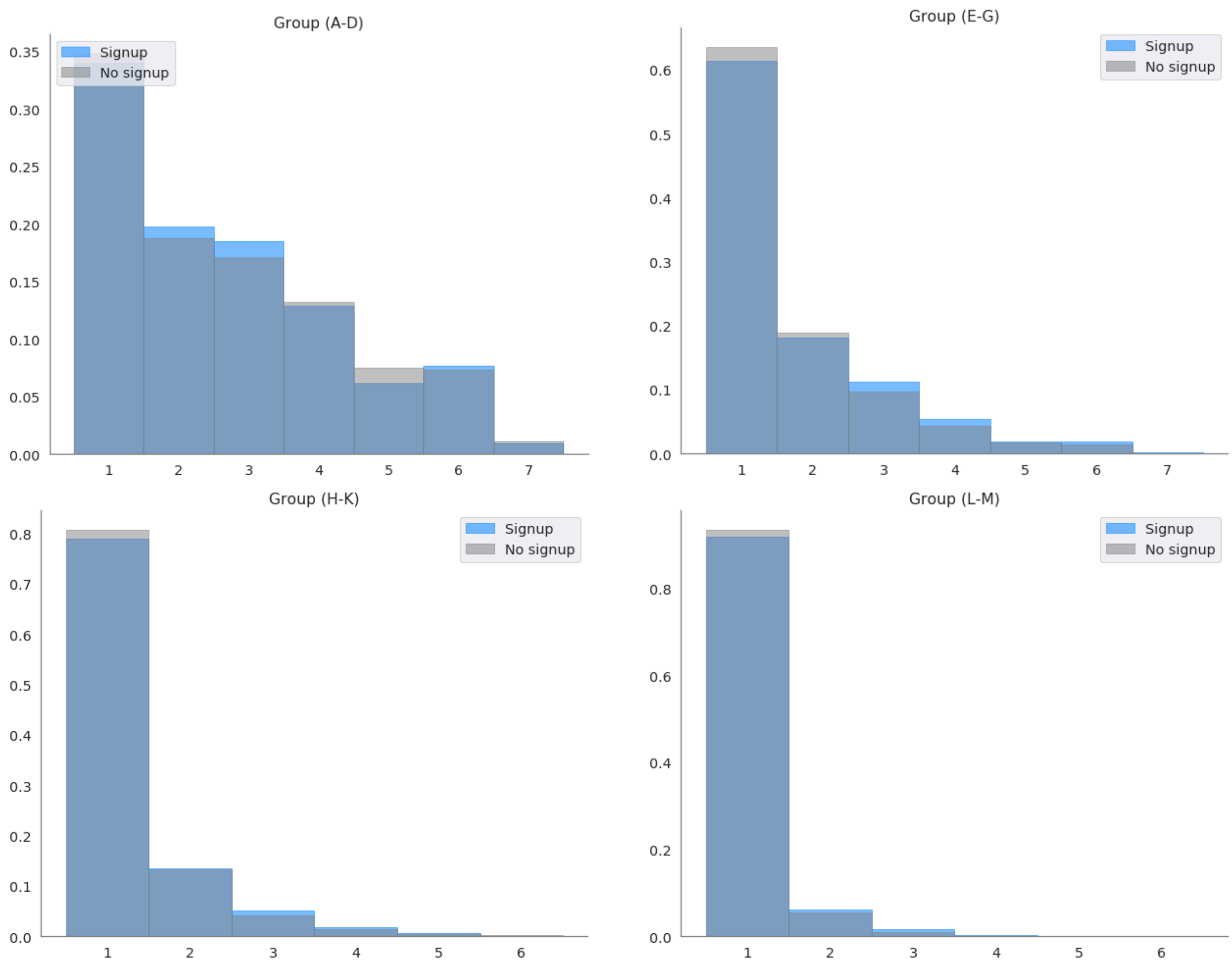
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes the share of applicants.

Figure A.2: Distribution of ex-post platform-provided credit score by signup decision (Access directly via price comparison websites)



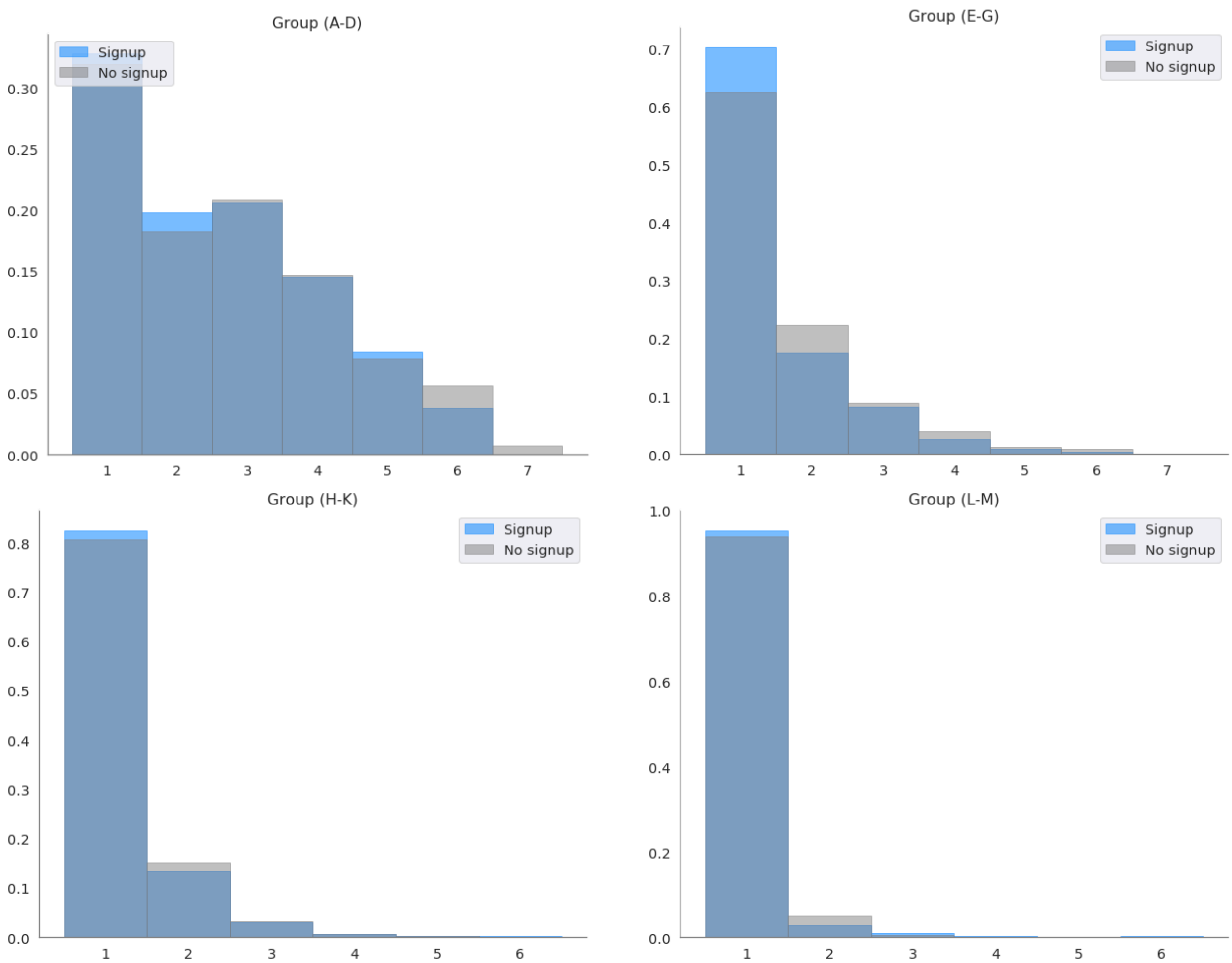
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes the share of applicants.

Figure A.3: Distribution of ex-post platform-provided credit score by signup decision (Access directly via brokers)



Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes the share of applicants.

Figure A.4: Distribution of ex-post platform-provided credit score by signup decision (Access directly via banks)



Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes the share of applicants.

B Tables

Table B.1: **Signup (Y) and no-signup (N) by access channels**

variable	Homepage		Repeat		Price comp		Broker		Others(banks)	
	Y	N	Y	N	Y	N	Y	N	Y	N
Credit requested	6,438.84	8,498.69	9,392.38	11,737.78	13,946.39	14,968.39	9,514.25	11,505.87	4,135.40	4,795.06
Credit offered	5,619.24	7,304.20	9,015.38	11,362.87	11,715.86	12,851.39	9,213.56	11,338.57	3,670.59	4,747.21
Interest rate	0.11	0.12	0.09	0.09	0.10	0.11	0.14	0.14	0.14	0.13
Auxmoney score (max 7, min 1)	2.79	2.35	4.59	4.63	3.28	3.02	1.71	1.79	1.41	1.56
Credit score (max 4, min 1)	2.87	2.88	3.01	3.19	3.12	3.21	2.66	2.86	2.32	2.49
Loan duration	30.16	25.67	53.04	53.73	54.86	56.83	61.83	62.93	71.99	68.56
Application accepted (Dummy)	0.67	0.56	0.96	0.97	0.71	0.72	0.35	0.37	0.24	0.31
Flagged for quality check (Dummy)	0.23	0.26	0.38	0.39	0.25	0.30	0.18	0.21	0.11	0.14
Bank account detail shared (Dummy)	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Age	31.79	34.52	39.81	43.88	33.94	38.72	33.65	37.19	27.21	29.62
Female	0.39	0.38	0.43	0.39	0.32	0.34	0.37	0.38	0.17	0.22
Main earner (if married) (Dummy)	0.11	0.12	0.31	0.32	0.71	0.66	0.63	0.69	0.68	0.88
No. months of employment	52.50	59.82	101.47	118.08	62.61	83.16	62.87	69.63	27.66	31.41
No. months of living in current place	81.16	88.57	116.77	130.35	81.21	93.71	78.73	80.83	14.52	16.07
Total income (median)	1,700.00	1,706.00	1,950.00	2,049.00	2,000.00	2,000.00	1,740.00	1,750.00	1,622.50	1,850.00
Total expenses (median)	600.00	670.00	838.50	920.50	615.00	600.00	450.00	450.00	660.00	794.00
Credit-card owner	0.45	0.39	0.64	0.65	0.84	0.68	0.64	0.37	0.84	0.94
EC-card owner	0.86	0.82	0.98	0.97	0.98	0.96	0.94	0.81	0.96	0.99
Home-owner	0.14	0.16	0.27	0.31	0.20	0.28	0.14	0.15	0.09	0.17
Car-owner	0.49	0.49	0.67	0.67	0.63	0.61	0.47	0.25	0.39	0.37
No. current loan demand	1.37	1.18	2.02	1.80	1.58	1.37	1.54	1.21	0.96	0.72
No. past loan demand	1.34	0.97	2.20	1.73	1.21	1.02	1.69	1.10	0.89	0.52

Table B.2: **Probability of data sharing conditional on observable risk**

Loans accessed via homepage and price comparison website

	Matched sample			
Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
	(1)	(2)	(3)	(4)
Goodtype (=1 if platform score 7-3)	0.126***	0.153***	0.116***	0.054**
	(0.006)	(0.004)	(0.007)	(0.024)
Controls	Y	Y	Y	Y
Cluster (Zipcode-Year)	Y	Y	Y	Y
N	102838	95814	26425	3052
Pseudo R2	0.1409	0.1478	0.1961	0.1600

Loans accessed via broker and bank

	Matched sample			
Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
	(1)	(2)	(3)	(4)
Goodtype (=1 if platform score 7-3)	-0.007	0.018	0.068**	0.069
	(0.015)	(0.013)	(0.022)	(0.067)
Controls	Y	Y	Y	Y
Cluster (Zipcode-Year)	Y	Y	Y	Y
N	5214	6648	2606	391
Pseudo R2	0.0940	0.0968	0.1280	0.0877

The above two tables show the marginal effects of the signup probability from the good type, using a probit model with the matched sample. Each regression includes a dummy variable *Goodtype* equal to 1 if the platform score is 7,6,5, 4, or 3.