Open Banking and Customer Data Sharing: Implications for FinTech Borrowers

Rachel J. Nam*

First draft: November 2022
Current Version: July 2023

Latest version here

Abstract

Open banking allows loan applicants to easily share payment data with prospective lenders during loan applications. In theory, this could broaden credit access by reducing information asymmetry but may also lead to first-degree price discrimination by exploiting individuals' preferences and behavioral traits. This paper studies the impact of open banking on borrowers and lends empirical support to the sizable benefits of data-sharing driven by improved inferences about borrower credit quality. Using granular loan application data from a leading German FinTech lender in consumer credit, I show that applicants with lower credit ratings are more likely to share data. Exploiting the variation of data sharing choice from observably similar applicants, I document that data sharing increases loan approval rates, reduces interest rates, and is associated with lower *ex-post* default rates. These findings are consistent with models of asymmetric information in which increased transparency and access to borrower data leads to a more efficient allocation of credit and reduced credit risk.

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 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 which provides consumers with enhanced control over their financial data sharing. In the European Union, regulatory mandates require banks and other financial entities to grant third-party providers (TPPs) access to customer data, contingent upon explicit consent from the customers. This has particular relevance in credit markets, where potential borrowers can now share their detailed transaction data seamlessly and securely during loan applications.

To the best of my knowledge, this study is the first to provide empirical evidence of open banking and consent-driven data sharing in the consumer credit market. Using a rich set of granular loan application data from the largest German FinTech lender, I study who shares data and investigate the consequences of such decisions on loan application outcomes. I show that lower credit-rating borrowers are more likely to share data. Importantly, loan applicants benefit from data sharing with a substantial increase in the probability of obtaining a loan and a reduction in the interest rate. Borrowers who share data also exhibit lower *ex-post* default rates, results consistent with the model of information asymmetry in which data sharing can serve as a signal of unobserved borrower quality.

A priori, the determinants of data sharing decision and their subsequent impact are unclear. Users with high privacy concerns may opt-out, yet potential gains from revealing one's private information might in fact encourage some consumers to opt-in (Tang, 2019b). On the one hand, access to granular consumer 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). On the other hand, it can also hurt borrowers if data is used mainly 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 transaction details from their bank account during the loan application, I show that borrowers from the lowest credit score group are 3.9 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). Credit scores, while commonly used, may not always provide a comprehensive or accurate representation of a borrower's actual credit risk. These inaccuracies can arise from the limited availability of relevant information, particularly for borrowers with lower credit ratings. I show that the precision of credit scores in predicting defaults diminishes as credit scores decrease. In this scenario, detailed financial data allows lenders to gain insights into a borrower's behavior and financial habits that might not be apparent from a traditional credit report, in turn, helping lenders make better inferences about borrowers' true credit quality. In these circumstances, borrowers with the highest discrepancy between the reported and true credit worthiness stand to gain the most from sharing data.

Overall, data sharing increases the probability of loan approval by up to 11 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. Data sharing benefits borrowers across all credit risk groups, but lower-rating borrowers benefit more from sharing data on the extensive margin (i.e. they enjoy a larger increase in the chance of getting a loan) relative to high-rated borrowers. High credit score borrowers have *ex-ante* a sufficiently high probability of obtaining a loan, thus, data sharing decisions affect the loan approval decisions to a lesser degree.

Interestingly, however, the effect on the interest rate (intensive margin) is larger for high-score borrowers. I use data to empirically test the two channels through which data could affect the loan price– 1) the risk-reducing effect and 2) the information-revealing effect (Ghosh et al., 2021), I show that the information-revealing channel is relatively stronger for the higher-rating borrowers. This means that their data contains more positive information, thus, leading to a larger improvement in the lender's prior, which justifies a larger reduction in the interest rate after data sharing.

It is important to note that the borrower's true type is unknown and will only be observable *ex-post*. Detailed payment data can alleviate information asymmetry since it can reveal the borrower's traits that are *unobservable* to the lender *ex-ante* but still relevant for credit risk. For instance, cash flows and consumption behaviors are unlikely to be reflected in one's credit score. Access to detailed 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 if data sharing can serve as a signal of unobserved borrower quality, I use two measures, 1) *ex-post* platform-provided scores which will reflect such information often unobserved in traditional risk proxies, and 2) borrowers' loan payment status after getting a loan and provide evidence that data sharing is associated with lower *ex-post* default rates conditional on observable risk (credit ratings). These findings, therefore, confirm 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), thus data sharing can serve as an effective signal.

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. However, when customer financial data are shared, these standard variables explain much less variation in loan application outcomes. 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 11 percentage points increase in loan approval, in reality, represents a significant number of borrowers who would have never received a loan without data sharing. The positive effects consistent across all borrower types suggest that open banking has the potential to 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 in the consumer loan market. As consumer data becomes increasingly more valuable, more firms will want access to it. Simultaneously, this trend is accompanied by growing awareness on consumer privacy and government regulations around data rights. Thus, customer consent will be a critical component in a data-driven economy. Therefore, the implications of this paper can also be applied beyond open banking.

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–transactions, income, spending, consumption behavior, etc, and open banking gives the customer the option to sign up. While credit registers are centralized databases which only cover consumers with credit products above a certain threshold, open banking data is available for anyone with a payment account. Importantly, this type of data updated in real-time. For instance, during a period of personal financial distress, a couple of negative shocks can impact one's credit history substantially. These records can stay within credit registries for a long period of time. Therefore, credit mistakes can be costly, especially for financially constrained borrowers. Transaction data, on the other hand, could provide a more up-to-date representation of consumer financial behavior, and ultimately alleviate financing constraints for marginal borrowers. Most importantly, open banking is a way for consumers to share their financial data with third-party providers for a range of purposes that may extend beyond lending.

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., 2022 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. Goldstein et al., 2022 compare closed banking and open banking in banking competition, resource allocation, and borrower welfare and show that while open banking can improve borrower welfare as competition drives down interest rates, it can also lead to inefficient allocation of resources due to positive non-participation by banks to avoid the winner's curse. 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

[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

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 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,

[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.

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 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, Section 5 reports the empirical results, and Section 6 provides some robustness checks and sensitivity analyses. Then, I provide potential avenues for future research and conclude in Section 7.

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 of open banking regulations vary substantially. While some countries mandate data sharing, others only recommend or facilitate by providing technical standards or infrastructure

[3]Babina, Buchak, and Gornall (2022) provide an excellent description of the status of open banking worldwide. In the United States, the responsibility of drafting regulations for open banking was given to the Consumer Financial Protection Bureau (CFPB) under Section 1033 of the Dodd-Frank Act of 2010. As part of this mandate, the Bureau announced in October 2022 their intention to finalize the governing rules by 2024, after which the implementation phase would begin.

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 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 [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. As a FinTech platform, *Auxmoney* is not a licensed bank, thus not subject to banking regulations. To issue loans, it partners with a fully

[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 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/

licensed credit institution.

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 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

[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).

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

loan contract agreement, loans are usually paid out within a matter of days.

Funding come from both individual and institutional investors. 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.

3.1.3 Descriptive statistics

As shown in Figure 1, 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. 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

[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.

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 1]

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 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 2 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 3). 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]

[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.

[Figure 2 and Figure 3]

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 4 shows the application process on the platform. During the application process, loan applicants are given the option to share 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. Figure 5 is what is shown to the user during the application process. If the customer chooses to share data, the platform will have access to the most recent four months of transaction data. The page also includes a message describing the potential benefit of providing bank data (an average discount on a loan), which is shown to everyone. This message may vary over time but the change happens infrequently (i.e. approximately once a year). Users are also provided with information on how the data is being used, and an explanation for a potential refusal or an increase in the interest rate that data sharing may cause. If the applicant shares data, the platform will use this data along with other credit bureau, application 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 4, and Figure 5]

Figure 6 provides a first glimpse of evidence of open banking. It shows the simple average of loan acceptance rates by data sharing sign-up decisions 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 on 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, on 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 characteristics of the users who share data 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, I employ a Hybrid Matching method to address potential selection bias and ensure comparability between our treatment and control groups. This approach combines two matching techniques to achieve optimal balance on observed covariates: Exact Matching and Propensity Score Matching (PSM). Given that borrower traits may differ substantially across access channels as shown in Table 3 and the data sharing trend fluctuates over-time, Exact Matching is applied to variables *Access channel* and *Loan application year*, ensuring that these categorical covariates are precisely matched between the treated and untreated groups. On the other hand, PSM is used for *Age*, *Income decile*, and *Credit score*. The latter technique allows for a degree of flexibility, creating matches based on the similarity of propensity scores, which are computed through logistic regression using the three aforementioned variables as predictors. Table 5 presents matching results. I test including 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 data sharing on the probability of loan approval using a probit model.

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

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). It should be noted that propensity score matching methods do not account for any unobserved characteristics which may simultaneously determine the selection into treatment and the outcome variable. The omission of such variables may result in endogeneity bias. To address this issue, I conduct further robustness checks in Section 6.

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 characterizes borrowers who 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 3.9 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.

Credit scores are often imperfect measures of the borrower's true credit risk and the signaling power of credit scores may be weaker for low-rating borrowers. In other words, borrowers with lower credit ratings not only have lower credit quality on average but also a more variable credit quality. I use the mean squared error, $MSE = E[(Z - E(Z|X))^2]$, to test the inference quality of credit scores in predicting defaults. A lower mean implies a smaller error in the prediction model[12].

[Table 8]

As shown in Table 8, credit scores exhibit a higher imprecision in predicting default risk among borrowers with low ratings compared to those with higher ratings. Thus, it is conceivable that individuals with lower credit ratings could benefit more from information disclosure. It is, however, important to highlight that data sharing is not without costs. The decision to share data in this context is determined not only by the expected gains from revealing information but also by factors such as privacy concerns and the cost of exposing adverse information, which could vary substantially across different agents. For example, it is plausible that the cost of exposing adverse information could substantially vary depending on the range of one's outside options. This implies that borrowers with limited alternative options may have stronger incentives to share their data.

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 a higher number of outstanding consumer loans and fully repaid past loans exhibit are more likely to share. This could potentially be an indicator that these borrowers have reached their maximum debt capacity and are thus facing more financial constraints. Building on the primary analysis conducted with a probit model, I also

[12] To discard the possibility that the credit score simply captures a mix of other factors that affect defaults, I control for the following borrower and loan characteristics: age, income decile, loan amount, loan duration, female (dummy), main earner (dummy), number of current and past loans, home owner (dummy), car owner (dummy), and access channel (categorical variable).

implemented a linear probability model as a supplemental analysis. The results are consistent, both qualitatively and quantitatively, with those of the main methodology.

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

The results from the previous section demonstrate that the probability of data sharing is higher for *ex-ante* observably lower-rating users. However, the true borrower type will only be known *ex-post*. Existing theories would suggest that, when there is a possibility of signaling, good types will opt in more to send a positive signal. Several papers, therefore, predict that open banking and consumer data provision may benefit good types while bad types will opt out in order to benefit from single pooling (He et al., 2020; Parlour et al., 2022; Babina et al., 2022). To test this, I first divide the sample into different credit score categories and test the signaling effect conditional on observable credit risk. Here, I use two measures, 1) *ex-post* platform scores and 2) loan payment status to infer *good-type*[13]. The platform score can be a better proxy of the true borrower type since it not only uses credit bureau data but also additional information derived from digital footprints and payment data (if provided) which potentially capture traits that are *unobservable* but still relevant for credit risk.

[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. 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

[13] 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/

an 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 9 shows that for the highest credit category, on average, it is 12 percentage points more likely that the good type signs up, and the effect for the second best group is even larger with 14.5 percentage points. The magnitude, however, becomes significantly attenuated for the riskiest borrowers with only 5 percentage points increase in the likelihood with less explanatory power.

[Table 9]

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 10]

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 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. 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 10 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 becomes smaller and insignificant for the lowest-rating borrowers. Overall, these results partially confirm the existing theoretical predictions which claim that voluntary signup will eventually lead only good types to opt-in (He et al., 2020).

5.2 The effect of customer data sharing on loan approval

Table 11 provides regression results of customer-driven data sharing on loan approval from equation (2). Sharing data improves the probability of loan approval for borrowers across all credit risk categories, but the greatest benefit is observed for borrowers from the second-lowest credit group (H-K), with an 11 percentage-point increase in the likelihood of loan acceptance. This effect is both statistically significant and economically sizable. The effect is much smaller for the highest (A-D) rated borrowers (a 2 percentage-point increase)[14]. Intuitively, high credit scores imply *ex-ante* a sufficiently high probability of obtaining a loan, data disclosure affects loan application outcomes to a lesser degree. The results indicate a hump-shaped relationship between the impact of data sharing and credit scores. Data sharing has the strongest effect on borrowers with mid-low tier credit ratings and a relatively weaker effect on those with higher and lower credit ratings. This suggests that mid-low-tier borrowers may be subject to greater credit quality variance, leading to more inaccuracies in their credit ratings. Detailed payment data can be used to reduce such credit-rating errors, thus the relative impact of open banking is expected to be positively correlated with the credit quality variance across credit ratings.

[Table 11]

The greater improvement in loan application outcomes for borrowers with lower credit scores suggests a higher incidence of credit rating errors among this group. In other words, the signal-to-noise ratio of credit ratings may be relatively high for the highest credit scores but lower for the rest. Therefore, for lower-rating borrowers, extra pieces of information could increase credit-quality accuracy to a larger extent, resulting in more loan approvals. This evidence has far-reaching policy implications. These 11 percentage points, in reality, represent a significant portion of borrowers who otherwise would have never been given a loan without open banking. It is, however, true

[14] It is important to note that estimating the marginal effect of interaction terms in a non-linear model is not straightforward (see Ai and Norton, 2003). To compute the effect of data sharing by credit score group, I obtain margins with nested designs. In this case, credit score group is treated as a nesting variable over which margins of data sharing are estimated.

that banks can voluntarily participate in data sharing on their own record and this is already being implemented in jurisdictions without or more lenient open banking regimes where banks are simply *recommended* to share data when customers consent. Nonetheless, empirical evidence from this exercise suggests that mandated data sharing by banks will be of the greatest value to consumers by giving them greater control of their own data. The positive effects of bank data sharing across all borrower types suggest that open banking and customer-directed data sharing can deepen credit markets by extending credit to those who, even with advanced algorithms employed by non-traditional lenders, would not have access to credit.

It should be noted that propensity score matching methods do not account for any unobserved characteristics which may simultaneously determine the selection into treatment and the outcome variable. Therefore, the omission of such variables may result in biased results, giving rise to the endogeneity problem. To address this issue, I conduct robustness checks in Section 6, showing that the results are qualitatively and quantitatively consistent under different specifications which subsumes away unobserved individual characteristics.

5.3 The effect of customer data sharing on loan interest rates

In this section, I test the effect of open banking on the intensive margin. How does consent-based data sharing affect the interest rate conditional on being granted access to credit? Table 12 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 12]

Similar to the findings presented in the previous section, data sharing leads to lower borrowing costs for borrowers across all risk groups. The effect on the intensive margin, however, is the largest for the safest borrowers (A-D) with a 2.14 percentage point reduction in the interest rate, 1.96 (E-G), 1.24 (H-K), and a 0.35 percentage point reduction for the riskiest 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 has selected loans with interest rates lower than the average interest rate of the population and the unselected loans would have been charged higher interest rates. Being a homeowner 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 on 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 high rating borrowers enjoy a bigger reduction in interest rates when sharing data. To understand the differential impact with respect to the interest rate, I explore the mechanisms through which data sharing affects loan application outcomes in the next section.

5.4 The channels through which data sharing affects loan application outcomes

Findings from the previous sections show that consent-driven data sharing on average leads to positive loan application outcomes both on the extensive and intensive margin, yet to varying degrees for borrowers of different risk levels. That is, borrowers with lower credit scores benefit more on the extensive margin whereas those with higher scores enjoy a larger reduction in the interest rate. To understand the source of this heterogeneity, I empirically test the mechanisms through which data disclosure affects loan application outcomes put forth by Ghosh et al., 2021.

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^{-1} is the variance around the lender's prior μ. In this setting, the risk-averse financier observes the information content of detailed payment data and updates her beliefs about the borrower. In this process, two complementary informational effects are at play through which data sharing affects financing outcomes: *risk-reducing effect* and *information-revealing effect.* The *risk-reducing effect* is driven by the fact that observing transaction data, regardless of the informational content (i.e., independent of borrower type), directly reduces the uncertainty faced by the financier. 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. Thus, the variation across borrowers regarding data disclosure is driven by the consumer's binary choice of data sharing which allows me to parsimoniously capture the essence of tranasaction details in determining financing outcomes.

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 investigate these mechanisms empirically, I start by calculating the change in the platform score resulting from data sharing decisions. The platform score is the internal credit rating assigned by the FinTech lender, and is provided only upon completion of the application process. If an applicant chooses to share their transaction data during the loan application, this information, among other data points, contributes to the calculation of the platform score. Consequently, if the transaction details yield more positive insights, the lender is likely to adjust its prior upwards, which could result in lower loan prices. To test this, I use a matched sample again, consisting of two groups that are similar in observable characteristics but differ in their data sharing decisions. From the lender's perspective, these two groups of borrowers are essentially identical based on observable characteristics; as such, their financing outcomes, ex-ante, should not exhibit substantial disparities. However, if the group opting to disclose information obtains markedly higher platform scores compared to the group choosing not to share their data, it implies an improvement in the lender's prior due to data sharing and/or that the disclosed data helps mitigate uncertainty for the lender.

[Table 13]

As depicted in Table 13, the magnitude of improvement in the platform score is not uniform, with a notably larger increase observed for high-rating applicants. The increase in the internal credit rating can be attributed to both an improvement in the lender's prior and mitigation of uncertainty. To single out the risk-reducing effects of data sharing, I measure the quality of platform scores in predicting actual defaults and evaluate the extent of risk (uncertainty) reduction.

[Table 14]

Table 14 indicates that the accuracy of platform scores in predicting actual defaults improves when data is shared. Particularly, for borrowers with high ratings, data sharing leads to a greater decrease in the variability in the prediction errors, thus predictions become more consistent and less uncertain. In other words, data sharing aids the lender in reducing default prediction uncertainty more effectively for these borrowers, thus providing a rationale for the larger impact of data sharing on loan pricing.

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 on 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 15 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 16. Overall, these results indicate that open banking can facilitate non-bank lenders to extend credit to those who are considered risky with thin credit files, such as low credit score, short credit history and contribute to a fairer and more democratic access to finance.

[Table 15 and Table 16]

6 Robustness checks

6.1 Using individual fixed effects to eliminate unobserved characteristics

Throughout this study, the effects of data sharing are estimated using 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 to minimize the omitted variable bias. However, the principal limitation of propensity score matching methods is that they do not account for potential selection on unobservables. In other words, the treatment and the control groups may still differ in unobservable characteristics that may simultaneously determine the selection into treatment and the outcome variables. Such circumstances can lead to biased results. To tackle this potential issue, I exploit individual-day fixed effects to test the robustness of the main findings. On the platform, applicants often file multiple applications on the same day to compare different offers. During this process, a user may first apply without data sharing, and then change her mind and decide to share data. Since there is no change in borrower characteristics within the same day, the variation in the user's data sharing decisions within a day allows me to employ stringent individual-day fixed effects. By subsuming away unobserved individual characteristics that may jointly determine the selection into treatment and the outcome variable, I test the robustness of the effect of data sharing. The sample consists of 34,610 applications from 6,380 users.

[Table 17 and Table 18]

Table 17 and 18 show the robustness test both on the probability of loan approval and the interest rate. The results are both qualitatively and quantitatively similar to the main results. Compared to prime borrowers, riskier borrowers enjoy a higher increase in the probability of loan approval, with middle-tier borrowers benefiting the most. The effects are smaller for the highest- and lowest-rating borrowers, which is in line with the hump-shaped relationship evidenced from the main results. The magnitude is marginally higher for the first and second highest credit score groups (A-D and E-G) by approximately 1-2 p.p. compared to the main results, and is slightly attenuated for the two lowest credit score groups (H-K, and L-M) by about 2 p.p. The effects on the interest rate are also robust quantitatively and qualitatively. Data sharing leads to a larger reduction in the interest rate for high-rated borrowers and the effect decreases for riskier borrowers. Compared to the main results, the magnitude of the reduction in the interest rate is slightly lower for the first three groups (A-D, E-G, H-K) and higher for the lowest rating group (L-M).

6.2 Selection on unobserved variables: *Rosenbaum* sensitivity analysis

To further address hidden bias from unobserved variables that may affect both the assignment to treatment and the outcome variable, I follow the method proposed by Rosenbaum, 2002 and test the size of the quantitative deviation from a random assignment which would result in a statistically insignificant treatment effect. It is a useful tool to test the sensitivity of causal inferences by allowing researchers to quantify how severe unmeasured confounding variables must be between the treated and control units to nullify the treatment effect. *Rosenbaum bounds* explicitly allow the odds of treatment to vary between the treated and control individuals by a parameter, $\Gamma \geq 1$, when the two groups have similar observable characteristics $X_t = X_c$; that is,

$$\frac{1}{\Gamma} \quad \leq \quad \frac{\dfrac{\pi_t}{(1-\pi_t)}}{\dfrac{\pi_c}{(1-\pi_c)}} \quad \leq \quad \Gamma \qquad \text{when} \quad X_t = X_c \tag{8}$$

where $\pi_i = Pr(D_i = 1|X_i) = F(\beta x_i + \gamma u_i)$ is the probability of data sharing which can be expressed is a logistic function F where x_i and u_i is the observable and unobservable variable, respectively. The i's odds of data sharing are $\frac{\pi_i}{1-\pi_i} = e^{\beta x_i + \gamma u_i}$. If $\Gamma = 1$, $\pi_t = \pi_c$ which means the odds of treatment (sharing data) are the same for the treatment and control groups who share similar observable characteristics. By setting the value of Γ larger than one, one can vary the degree of hidden bias. If $\Gamma = 2$, the treatment group is twice as likely as the control group to share data due to unobservable differences.

To do this, I first match individuals who share data on all observable characteristics to create a control group and examine the bounds at which the treatment effect becomes no longer significant. Table 19 reports the bounds parameter Γ. The statistically significant effect of data sharing on the extensive margin will be challenged only if the unobserved biased selection into sharing would have to be so high to cause the odds ratio of data sharing to differ between the two groups by around 5 times for the highest credit rating bracket (A-D). While the results for the second-highest credit score group (E-G) are less pronounced, the selection on unobservables would still have to be more than 50% as high. For the rest of the groups, the effect is statistically significant at all levels of the sensitivity parameter gamma. This means that even if there is a large amount of hidden bias due to unobserved covariates (i.e. twenty times larger odds ratio), data sharing still has a statistically significant effect. Overall, this evidence suggests that selection on unobservables would have to be very large to eliminate the effects of data sharing.

7 Conclusion

This paper provides empirical evidence of open banking, a policy that empowers consumers by giving them greater control of their own data and discretion over sharing their financial 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 (data sharing) is higher among riskier (lower credit rating) borrowers, but conditional on *observable* risk, those who share data have *ex-post* lower default rates. I also provide evidence that customer-directed data sharing can benefit all borrowers both on the extensive and intensive margin. The effect, however, varies across different credit risk levels. Applicants who share data benefit substantially from increased loan approval rates and reduced interest rates. In terms of economic magnitude, lower credit-rating borrowers gain the most on the extensive margin, while high-rating borrowers obtain 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 data sharing is becoming more common, and consumers' decision to share financial data is bringing substantial benefits to loan applicants.

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 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

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics letters*, *80*(1), 123–129.

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., Huang, C., & Yang, L. (2022). Open banking under maturity transformation. *Available at SSRN*.

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*.

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

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

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.

Rosenbaum, P. R. (2002). *Overt bias in observational studies*. Springer.

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
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	2380.33	0.00	1,450.00	1,950.00	2,610.00	50,000.00
Monthly net salary income	2484979	2,593.52	1874.71	0.00	1,300.00	1,800.00	2,370.00	36,000.00
Child support income	2484979	127.21	208.08	0.00	0.00	0.00	204.00	1,638.00
Other income	2484979	194.48	553.84	0.00	0.00	0.00	0.00	10,000.00
Total expenses	2484987	740.65	617.69	0.00	304.00	590.00	933.00	5980.00
Housing related expenses	2484069	481.37	386.32	0.00	180.00	415.00	645.00	3500.00
Credit installments expenses	2484069	166.37	331.21	0.00	0.00	0.00	216.00	3685.00
Other expenses	2484069	23.88	123.50	0.00	0.00	0.00	0.00	2000.00
Insurance expenses	2484069	49.77	153.69	0.00	0.00	0.00	0.00	1707.00
Child support expenses	2484069	19.18	103.03	0.00	0.00	0.00	0.00	1500.00
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
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	33.778	33.762	0.654
Credit score	3.0552	3.0579	0.306
Income decile	5.3476	5.3476	0.997
Access channel		——— exact matching ———	
Application year		——— exact matching ———	

This table shows t-tests for the null hypothesis of equal means for the treated and control groups. This sample is used to compute the effect of data sharing on the probability of loan approval (equation 2). Each of the 188,453 applicants that shared data is matched one-to-one with the closest propensity score to create a control group that did not share data but is observably similar. Matching is done using age, credit score, and income decile, access channel and loan application year. Exact matching is used for the access channel and loan application year. The final sample includes 376,852 loan applications (365,852 unique applicants).

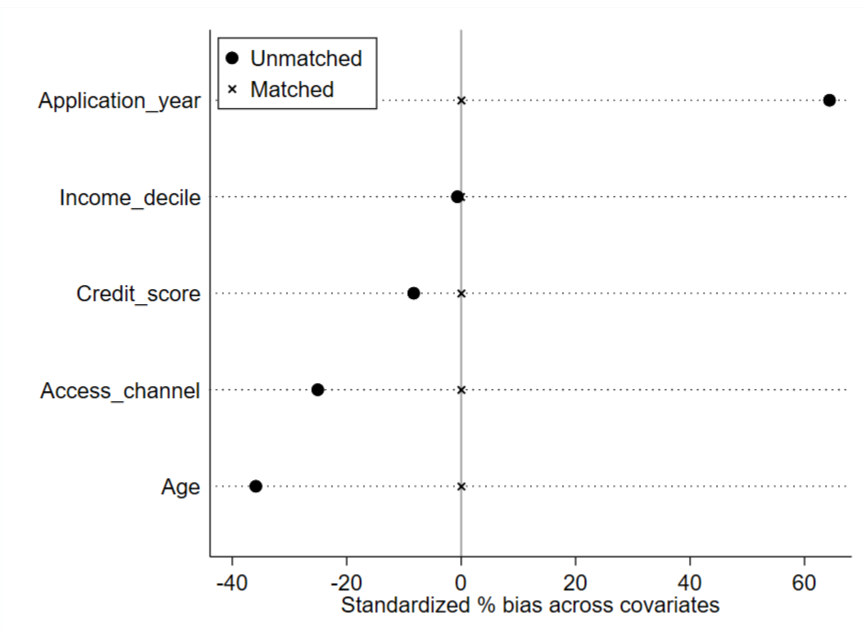


Table 6: **Matched variables and matching results**

	Mean Treated	Mean Control	Mean diff p-value
Age	36.496	36.477	0.662
Credit score	3.3081	3.3081	1.000
Income decile	5.8796	5.8908	0.331
Access channel		——— exact matching ———	
Application year		——— exact matching ———	

This table shows t-tests for the null hypothesis of equal means for the treated and control groups. This sample is used to compute the effect of data sharing on the interest rate (equation 3). Each of the 125,889 *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") to create a control group that did not share data but is observably similar. Matching is done using age, credit score, and income decile, access channel and loan application year. Exact matching is used for the access channel and loan application year. The final sample includes 249,240 loan applications (249,240 unique loan applicants).

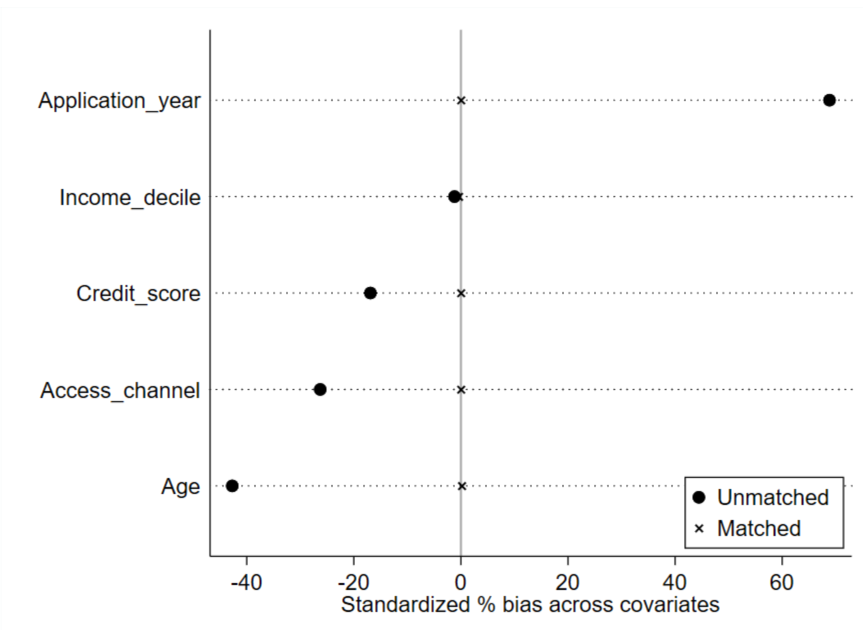


Table 7: **What characterizes borrowers who share data?**

	Probit (marginal effects)				LPM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age ('0)		-0.020***	-0.018***	-0.019***		-0.019***	-0.017***	-0.018***
		(0.0001)	(0.0002)	(0.0002)		(0.0002)	(0.0002)	(0.0002)
Income decile			0.001***	0.000			0.001***	-0.000
			(0.0001)	(0.0001)			(0.0001)	(0.0001)
Credit score (A-D) (base)								
Credit score (E-G)	0.026***		0.013***	0.009***	0.027***		0.014***	0.009***
	(0.0004)		(0.0004)	(0.0004)	(0.0004)		(0.0004)	(0.0004)
Credit score (H-K)	0.039***		0.019***	0.018***	0.038***		0.019***	0.016***
	(0.0006)		(0.0006)	(0.0006)	(0.0006)		(0.0006)	(0.0006)
Credit score (L-M)	0.039***		0.015***	0.021***	0.034***		0.010***	0.013***
	(0.0011)		(0.0010)	(0.0010)	(0.0008)		(0.0009)	(0.0009)
Loan amount requested (ln)				-0.011***				-0.011***
				(0.0002)				(0.0002)
Loan duration (ln)				-0.003***				-0.006***
				(0.0004)				(0.0005)
Female				-0.004***				-0.005***
				(0.0004)				(0.0004)
Main earner				0.009***				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)
Homeowner				-0.008***				-0.008***
				(0.0005)				(0.0004)
Car owner				0.012***				0.009***
				(0.0004)				(0.0004)
Access channel=Homepage (base)								
Access channel=Repeat				0.115***				0.070***
				(0.0031)				(0.0025)
Access channel=Price comp. website				-0.077***				-0.067***
				(0.0012)				(0.0011)
Access channel=Broker				-0.108***				-0.098***
				(0.0014)				(0.0013)
Access channel=Bank				-0.127***				-0.131***
				(0.0015)				(0.0017)
constant					0.020***	0.114***	0.094***	0.261***
					(0.0004)	(0.0007)	(0.0009)	(0.0023)
Dummy	Year	Year	Year	Year	Year	Year	Year	Year
Cluster (Zipcode-Year)	X	X	X	X	X	X	X	X
N	2,237,677	2,237,677	2,237,677	2,237,677	2,237,677	2,237,677	2,237,677	2,237,677
R2	0.0640	0.0724	0.0738	0.1055	0.036	0.040	0.041	0.058

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: **How good are credit scores in predicting defaults? (Mean Squared Error)**

(1) Credit score group	(2) N	(3) MSE	(4) Std. Dev.	(5) Min	(6) Max
A-D	26,871	0.0376	0.1705	5.7e-07	1
E-G	22,496	0.0608	0.2029	1.4e-06	.99
H-K	5,562	0.0818	0.2226	8.9e-06	.97
L-M	422	0.1065	0.2173	2.4e-07	.98

This table presents the imprecision of credit scores in predicting defaults. The imprecision of inference is measured using the mean squared error, denoted as $MSE = E[(Z - E(Z|X))^2]$, wherein Z represents a binary variable that assumes a value of 1 if the loan is in default status (delinquency extending beyond 90 days). A probit model has been used to estimate the default probability using credit scores. A lower MSE value suggests a smaller error in the predictive model. Results are presented by credit score groups, with groups A-D representing the highest and groups L-M denoting the lowest credit score categories. The sample includes loans that are securitized via "Fortuna Consumer Loan ABS (2021, 2022, 2023)", with a total volume of EUR 850 million. Data is sourced from the European Data Warehouse (EDW) which discloses loan-level data in European asset-backed securities. The final sample includes 55,351 loans. To discard the possibility that the credit score simply captures a mix of other factors that affect defaults, I control for the following borrower and loan characteristics: age, income decile, loan amount, loan duration, female (dummy), main earner (dummy), number of current and past loans, home owner (dummy), car owner (dummy), and access channel (categorical variable).

Table 9: **Are good types more likely to share data? (using platform scores)**

	DV = 1 if data is shared Matched sample			
Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
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 10: **Are good types more likely to share data? (using loan payment status)**

	DV = 1 if data is shared			
Credit score group	(A-D)	(E-G)	(H-K)	(L-M)
Goodtype (=1 if always paid on time)	0.072***	0.055***	0.052***	-0.009
	(0.011)	(0.010)	(0.017)	(0.050)
Controls	Y	Y	Y	Y
Cluster (Zipcode-Year)	Y	Y	Y	Y
N	22,447	20,085	5,004	410
Pseudo R2	0.0728	0.0814	0.0718	0.0845

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 credit score group with (A-D) being the highest and (L-M) being the lowest credit score group. The sample in this regression includes loans that are securitized via "Fortuna Consumer Loan ABS (2021, 2022, 2023)", with a total volume of EUR 850 million. Data is sourced from the European Data Warehouse (EDW) which discloses loan-level data in asset-backed securities in Europe. The final sample includes 43,515 loans. Control variables include age, income decile, loan amount, loan duration, female (dummy), main earner (dummy), number of current and past loans, home owner (dummy), car owner (dummy), and access channel (categorical variable).

Table 11: **The effect of data sharing signup decision on loan approval**

	Probit (marginal effects)			LPM		
	(1)	(2)	(3)	(4)	(5)	(6)
Signup	0.063***	0.064***	0.020***	0.033***	0.033***	0.008***
	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
Signup × Credit score (A-D)* (Base)						
Signup × Credit score (E-G)	0.066***	0.078***	0.058***	0.101***	0.102***	0.085***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Signup × Credit score (H-K)	0.101***	0.119***	0.091***	0.132***	0.130***	0.123***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Signup × Credit score (L-M)	0.062***	0.077***	0.047***	0.055***	0.053***	0.048***
	(0.009)	(0.010)	(0.009)	(0.006)	(0.006)	(0.006)
Credit score (A-D) (Base)						
Credit score (E-G)	-0.289***	-0.242***	-0.157***	-0.328***	-0.256***	-0.222***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Credit score (H-K)	-0.058***	-0.550***	-0.460***	-0.621***	-0.523***	-0.480***
	(0.003)	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)
Credit score (L-M)	-0.073***	-0.760***	-0.734***	-0.811***	-0.707***	-0.634***
	(0.005)	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)
Age		0.009***	0.006***		0.007***	0.005***
		(0.000)	(0.000)		(0.000)	(0.000)
Income decile		0.028***	0.015***		0.022***	0.013***
		(0.000)	(0.000)		(0.000)	(0.000)
Loan amount requested (ln)			0.013***			0.015***
			(0.001)			(0.001)
Loan duration (ln)			-0.112***			-0.101***
			(0.002)			(0.002)
Marital status			-0.001			0.001
			(0.001)			(0.001)
Female			0.033***			0.034***
			(0.002)			(0.001)
Main earner			0.028***			0.018***
			(0.002)			(0.001)
Home owner			0.061***			0.044***
			(0.002)			(0.002)
Car owner			0.067***			0.069***
			(0.002)			(0.001)
No. current loan demand			0.019***			0.020***
			(0.001)			(0.001)
No. past loan demand			0.006***			0.006***
			(0.000)			(0.000)
Access channel=Homepage (Base)						
Access channel=Repeat			0.000			-0.075***
			(0.000)			(0.003)
Access channel=Price comp. website			-0.269***			-0.289***
			(0.001)			(0.002)
Access channel=Broker			-0.555***			-0.505***
			(0.005)			(0.004)
Access channel=Bank			-0.478***			-0.452***
			(0.012)			(0.007)
constant				0.656***	0.642***	1.155***
				(0.002)	(0.004)	(0.006)
Dummy	Year	Year	Year	Year	Year	Year
Cluster (Zipcode-Year)	X	X	X	X	X	X
N	376,852	376,852	376,852	376,852	376,852	376,852
R2	0.1721	0.2545	0.3499	0.2461	0.295	0.352

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. It is important to note that estimating the marginal effect of interaction terms in a non-linear model is not straightforward (see Ai and Norton, 2003). To compute the effect of data sharing by credit score group, I obtain margins with nested designs. In this case, credit score group is treated as a nesting variable over which margins of data sharing are estimated.

Table 12: **The effect of data sharing signup decision on interest rates**

	Matched sample	
	(1)	(2)
Inverse mill's ratio		**-0.0083*****
		(0.0009)
Signup	-0.0205***	-0.0214***
	(0.0002)	(0.0002)
Signup × Credit score (A-D) (Base)		
Signup × Credit score (E-G)	0.0026***	0.0018***
	(0.0003)	(0.0003)
Signup × Credit score (H-K)	0.0104***	0.0090***
	(0.0004)	(0.0005)
Signup × Credit score (L-M)	0.0197***	0.0179***
	(0.0010)	(0.0010)
Age ('0)	-0.0157***	-0.0172***
	(0.0004)	(0.0005)
Age ('0) squared	0.0009***	0.0010***
	(0.0001)	(0.0001)
Credit score (A-D) (Base)		
Credit score (E-G)	0.0207***	0.0232***
	(0.0002)	(0.0004)
Credit score (H-K)	0.0310***	0.0369***
	(0.0003)	(0.0007)
Credit score (L-M)	0.0372***	0.0473***
	(0.0007)	(0.0013)
Income decile	-0.0019***	-0.0019***
	(0.0000)	(0.0000)
Loan amount requested (ln)	0.0098***	0.0091***
	(0.0001)	(0.0001)
Loan duration (ln)	0.0046***	0.0068***
	(0.0002)	(0.0003)
Marital status	0.0009***	0.0009***
	(0.0001)	(0.0001)
Female	-0.0025***	-0.0025***
	(0.0002)	(0.0002)
Main earner	-0.0032***	-0.0028***
	(0.0002)	(0.0002)
Home owner	-0.0188***	-0.0202***
	(0.0002)	(0.0002)
Car owner	-0.0046***	-0.0046***
	(0.0002)	(0.0002)
Credit card owner	-0.0063***	-0.0062***
	(0.0002)	(0.0002)
Checking account owner	-0.0018***	-0.0018***
	(0.0004)	(0.0004)
No. current loan demand	-0.0009***	-0.0009***
	(0.0001)	(0.0001)
No. past loan demand	-0.0001**	-0.0001**
	(0.0000)	(0.0000)
Access channel=Homepage		
Access channelr=Repeat	-0.0256***	-0.0256***
	(0.0005)	(0.0005)
Access channel=Price comp. website	0.0063***	0.0063***
	(0.0003)	(0.0003)
Access channel=Broker	0.0189***	0.0188***
	(0.0005)	(0.0005)
Access channel=Bank	0.0217***	0.0215***
	(0.0012)	(0.0012)
Constant	0.0764***	0.0788***
	(0.0011)	(0.0011)
Dummy	Year	Year
Cluster (Zipcode-Year)	X	X
N	249,240	249,240
Adjusted R2	0.4460	0.4462

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

	Dep. var = Platform score			
	Credit score			
	(A-D)	(E-G)	(H-K)	(L-M)
Signup	0.7732***	0.6619***	0.3798***	0.0999***
	(0.0088)	(0.0072)	(0.0077)	(0.0094)
Dummy	Year	Year	Year	Year
Controls	X	X	X	X
Cluster (Zipcode-Year)	X	X	X	X
N	122,905	155,869	64,180	14,362
Adjusted R2	0.3449	0.3223	0.3333	0.4658

This table reports the results of data sharing (*Signup*) on the change in the platform score by credit score group, using the matched sample. Control variables include age, income decile, loan amount, loan duration, female (dummy), main earner (dummy), number of current and past loans, home owner (dummy), car owner (dummy), and access channel (categorical variable). Platform score ranges from 7 (highest) to 1 (lowest and rejected).

Table 14: **Data as uncertainty mitigator**

	Share	Not share	Share	Not share	Share	Not share
Credit score group	A-D		E-G		H-M	
N	7,733	7,733	7,563	7,563	1,628	1,628
MSE	0.033	0.0413	0.0529	0.066	0.0757	0.094
Std.Dev	0.1553	0.1761	0.1914	0.2087	0.2131	0.2257
reduction in Std.Dev	13.39 %		9.03 %		5.91 %	

This table presents the imprecision of platform scores in predicting defaults and the reduction in variance by credit score group. The imprecision of inference is measured using the mean squared error, denoted as $MSE = E[(Z - E(Z|X))^2]$, wherein Z represents a binary variable that assumes a value of 1 if the loan is in default status (delinquency extending beyond 90 days). A probit model has been used to estimate the default probability using platform scores. A lower MSE value suggests a smaller error in the predictive model. Results are presented by credit score groups, with groups A-D representing the highest and groups H-M denoting the lowest credit score categories (due to insufficient observations of the previous denoted L-M group, H-K and L-M are combined for this analysis). The sample includes loans that are securitized via "Fortuna Consumer Loan ABS (2021, 2022, 2023)", with a total volume of EUR 850 million. Data is sourced from the European Data Warehouse (EDW) which discloses loan-level data in European asset-backed securities. A matched sample is used for the analysis to ensure that the two groups are observably similar and comparable but differ by data sharing choices. The final sample includes 33,838 loans. To discard the possibility that the credit score simply captures a mix of other factors that affect defaults, I control for the following borrower and loan characteristics: age, income decile, loan amount, loan duration, female (dummy), main earner (dummy), number of current and past loans, home owner (dummy), car owner (dummy), and access channel (categorical variable).

Table 15: **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 16: **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	Year	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	Year	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	Year	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	Year	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 17: **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. The sample includes multiple applications filed by the same individuals on the same day. These individuals do not share their data in some applications but do share in others. Given that there is within-individual variation in data sharing decisions but borrower characteristics do not change within a day, I employ individual-day fixed effects to test the effect of data sharing on loan approval as a robustness check. Control variables include loan amount and loan duration. The dependent variable is a dummy variable which takes a value of 1 if the loan application is approved and 0 otherwise. *Signup* is a dummy variable which takes a value of 1 if the applicant shared data and 0 otherwise.

Table 18: **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. The sample includes multiple applications filed by the same individuals on the same day. These individuals do not share their data in some applications but do share in others. Given that there is within-individual variation in data sharing decisions but borrower characteristics do not change within a day, I employ individual-day fixed effects to test the effect of data sharing on loan approval as a robustness check. Control variables include loan amount and loan duration. The dependent variable is the loan interest rate conditional on the application being approved. *Signup* is a dummy variable which takes a value of 1 if the applicant shared data and 0 otherwise.

Table 19: **Rosenbaum bounds sensitivity analysis**

	Credit score groups			
	(A-D)	(E-G)	(H-K)	(L-M)
$\Gamma_{p>.01}$	4.98	1.53	20^{+}	20^{+}
$\Gamma_{p>.05}$	5.04	1.54	20^{+}	20^{+}
$\Gamma_{p>.10}$	5.08	1.55	20^{+}	20^{+}

This table shows how much higher the odds of data sharing would need to be for the treated group compared to the control group, based on unobservables, in order for the treatment effect to be insignificant at the 1%, 5%, and 10% level. Data sharers are matched with a pair using all observable variables.

Figures

Figure 1: Number of applications (Monthly)

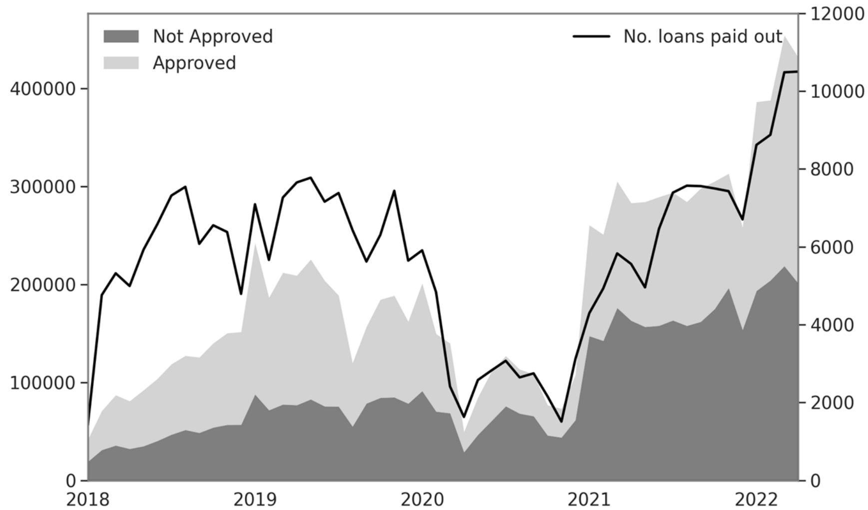
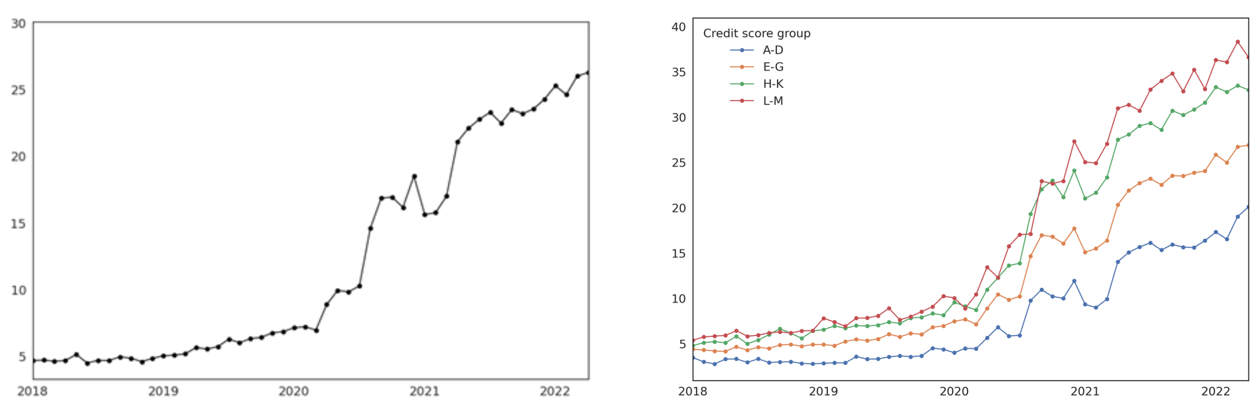
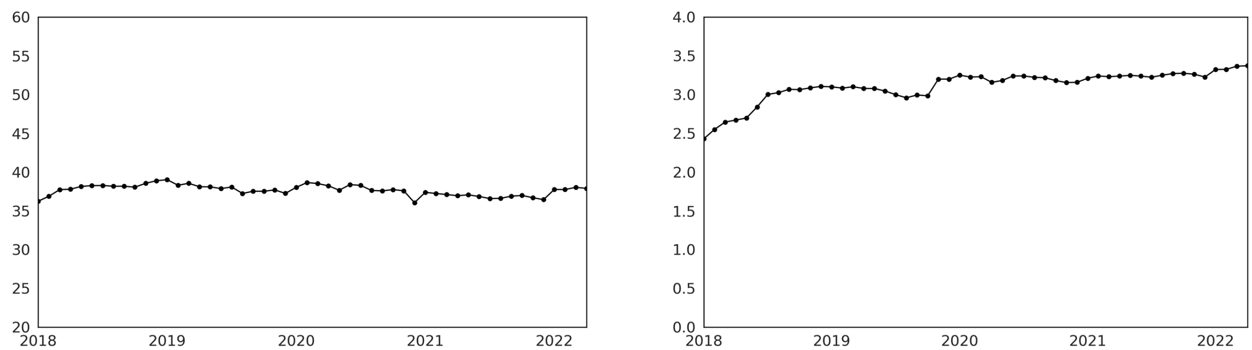


Figure 2: Open banking adoption over time (overall vs. by credit score, monthly)



Notes: The left panel shows the percentage of loan applications in which applicants opt to share their data, calculated as a fraction of the total number of loan applications. The right panel illustrates these percentages in relation to applicants' credit scores (A-D: highest, L-M: lowest).

Figure 3: Average age and Credit score



Notes: The left panel shows the average age of the applicants on the platform. The right panel represents the average credit score of the applicants on the platform.

Figure 4: Loan application, decision, and payout process

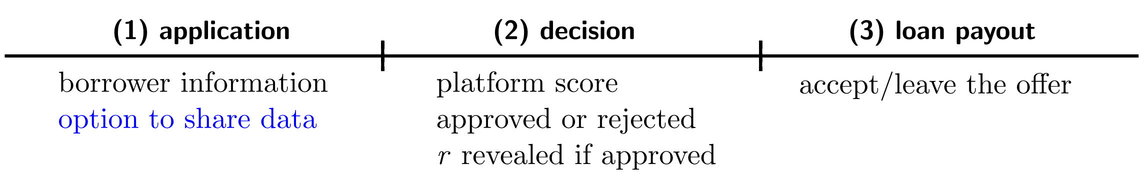


Figure 5: Data sharing during the application

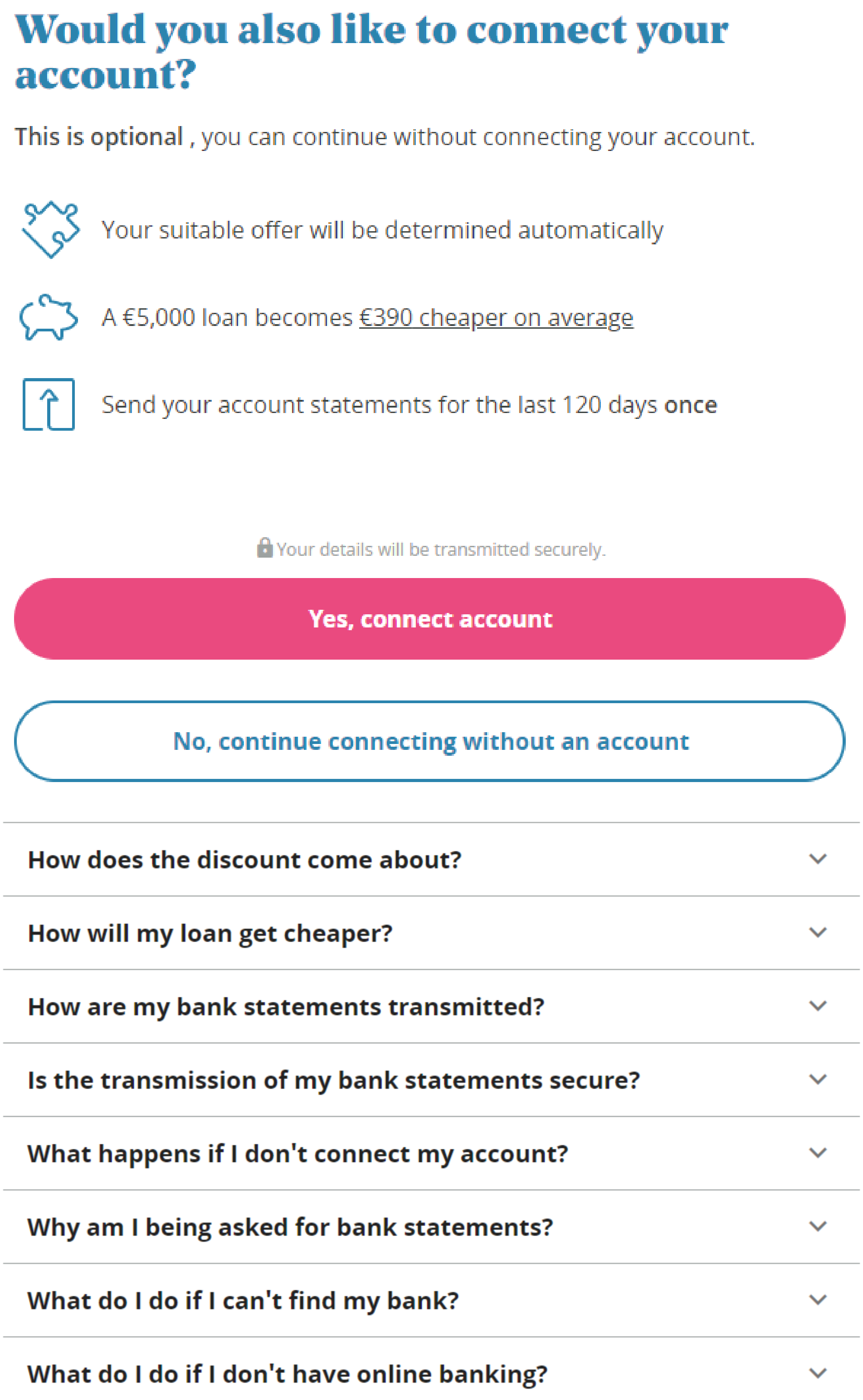


Figure 6: Loan acceptance rate by data sharing signup decision

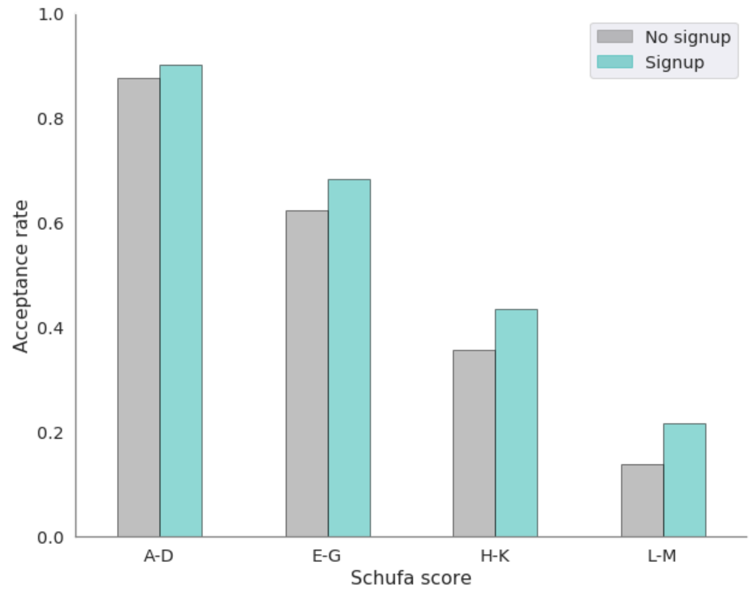


Figure 7: Loan interest rate by data sharing signup decision

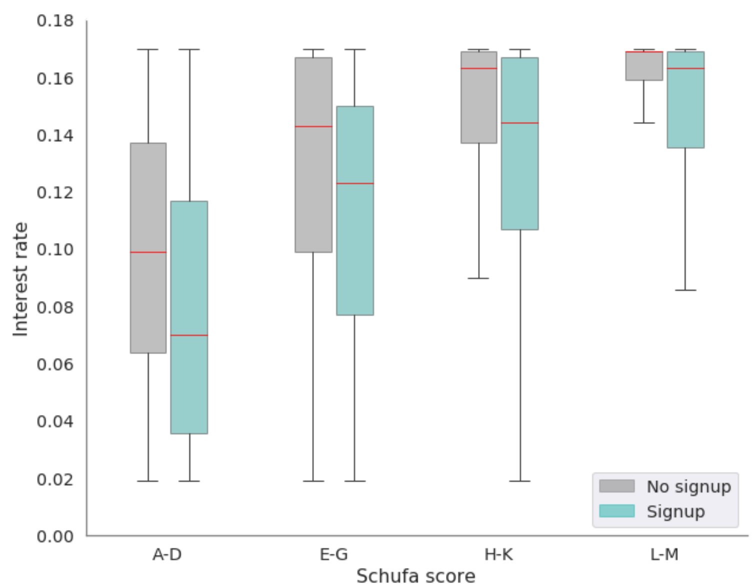
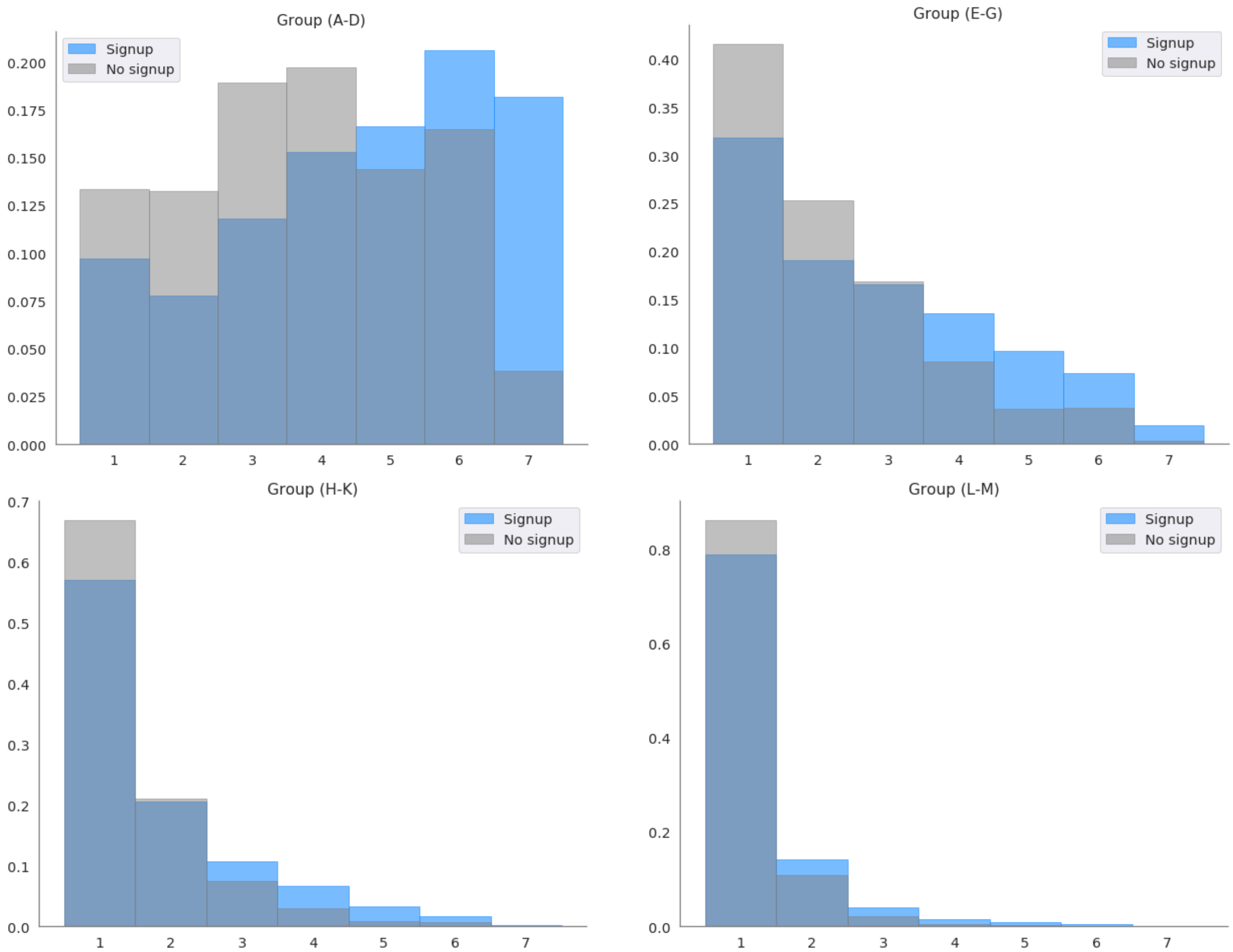
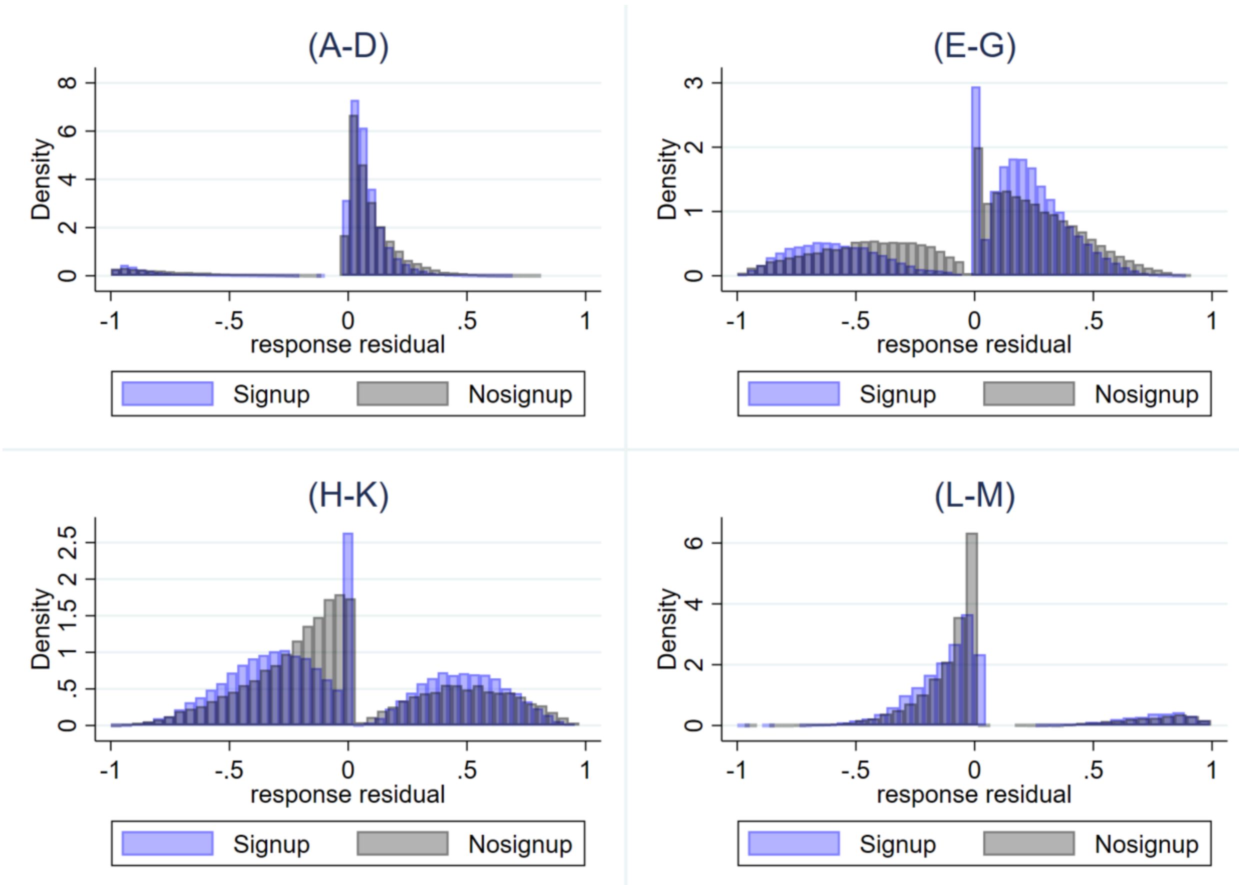


Figure 8: Distribution of platform-provided credit score by signup decision (using the matched sample)



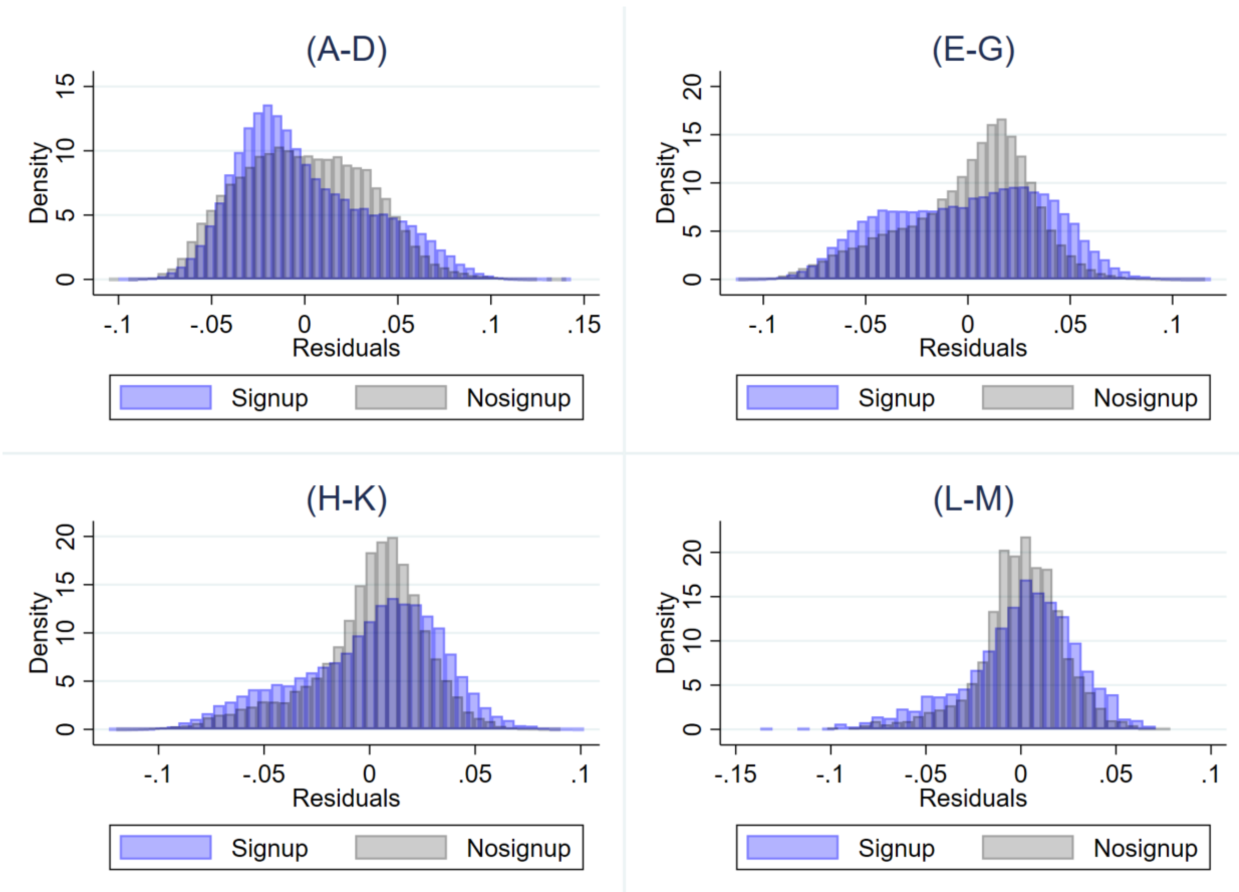
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes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

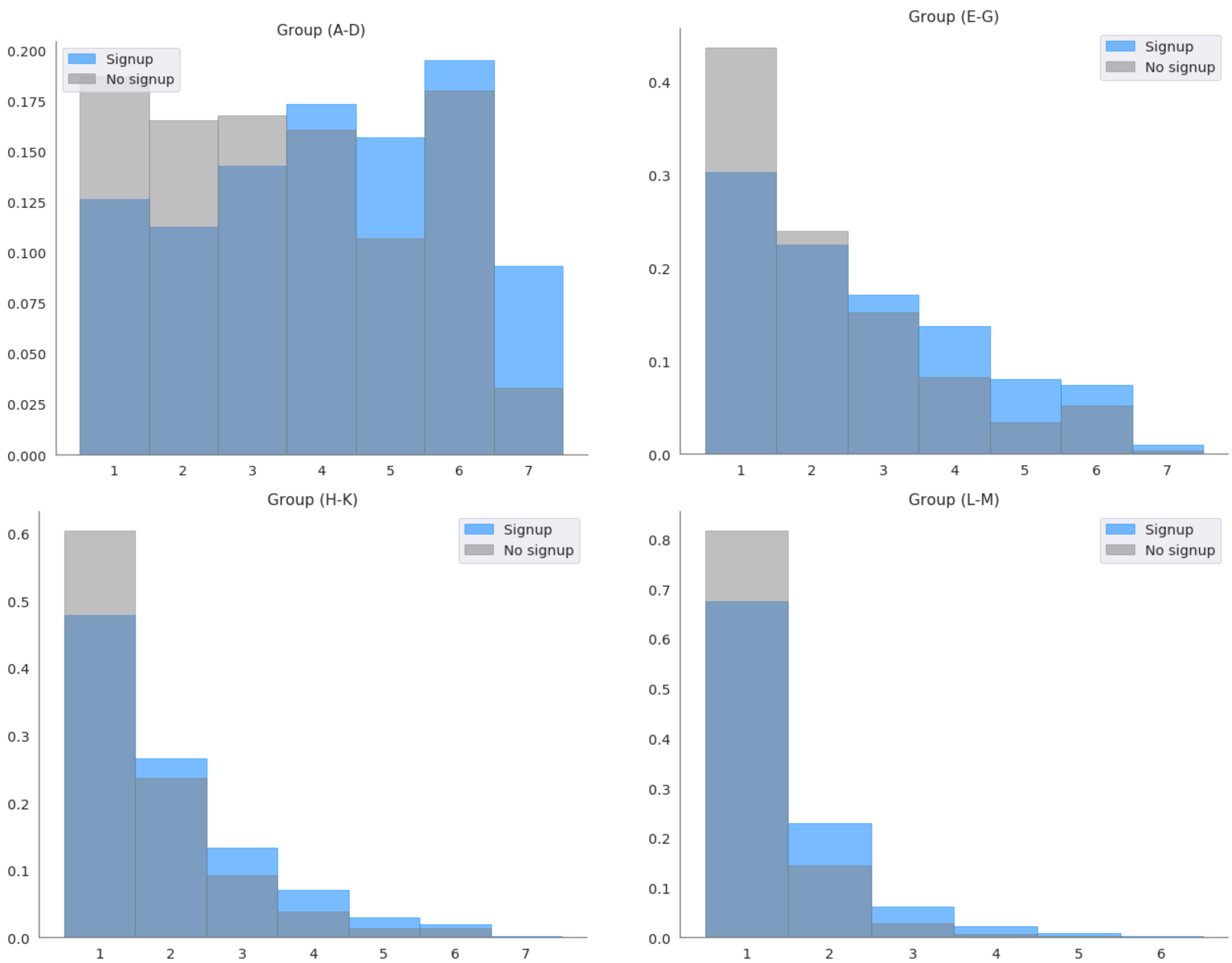


Note: Each panel represents the distribution of residuals from equation (3) (the effect of open banking on interest rate) for each credit score group.

Appendix

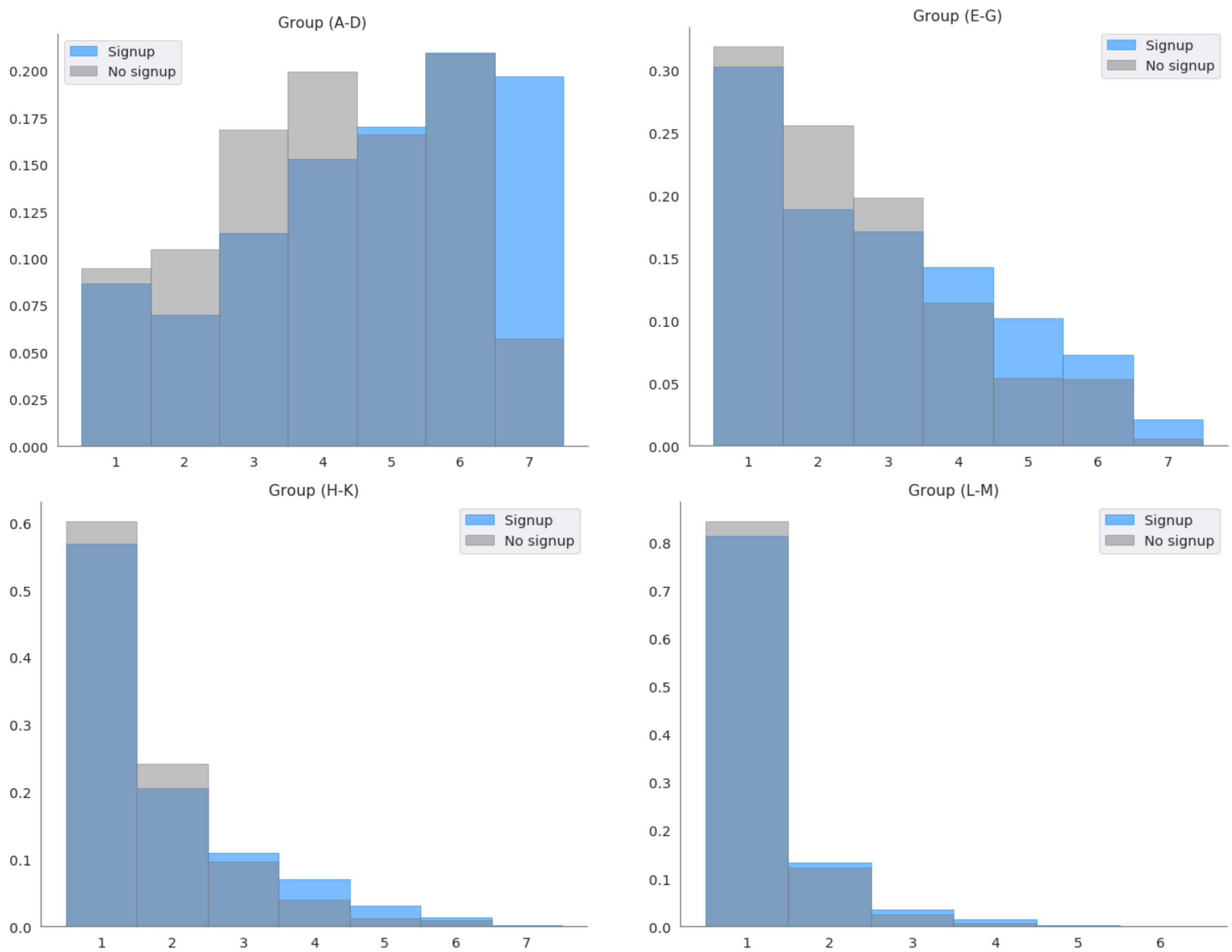
A Figures

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



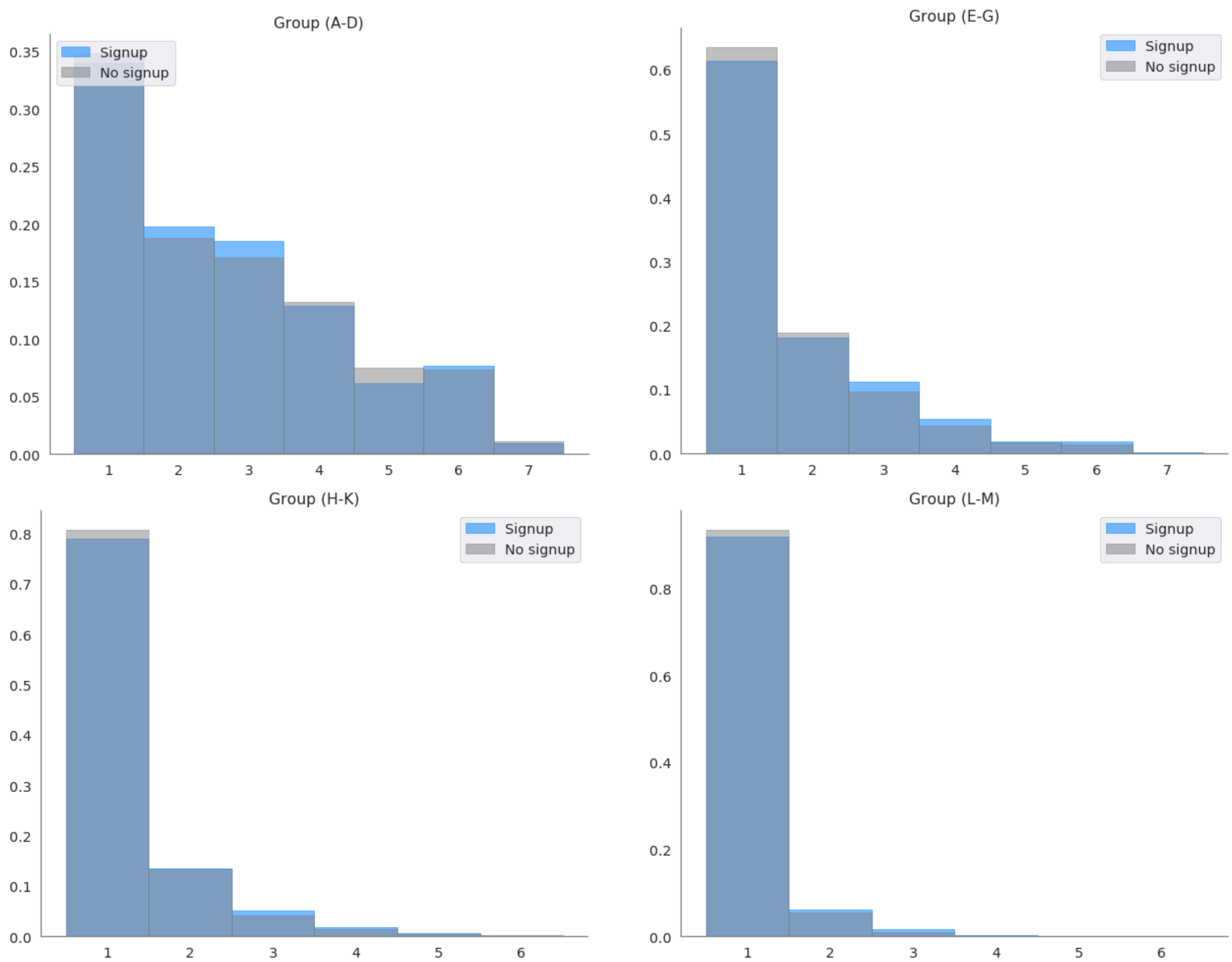
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes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

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



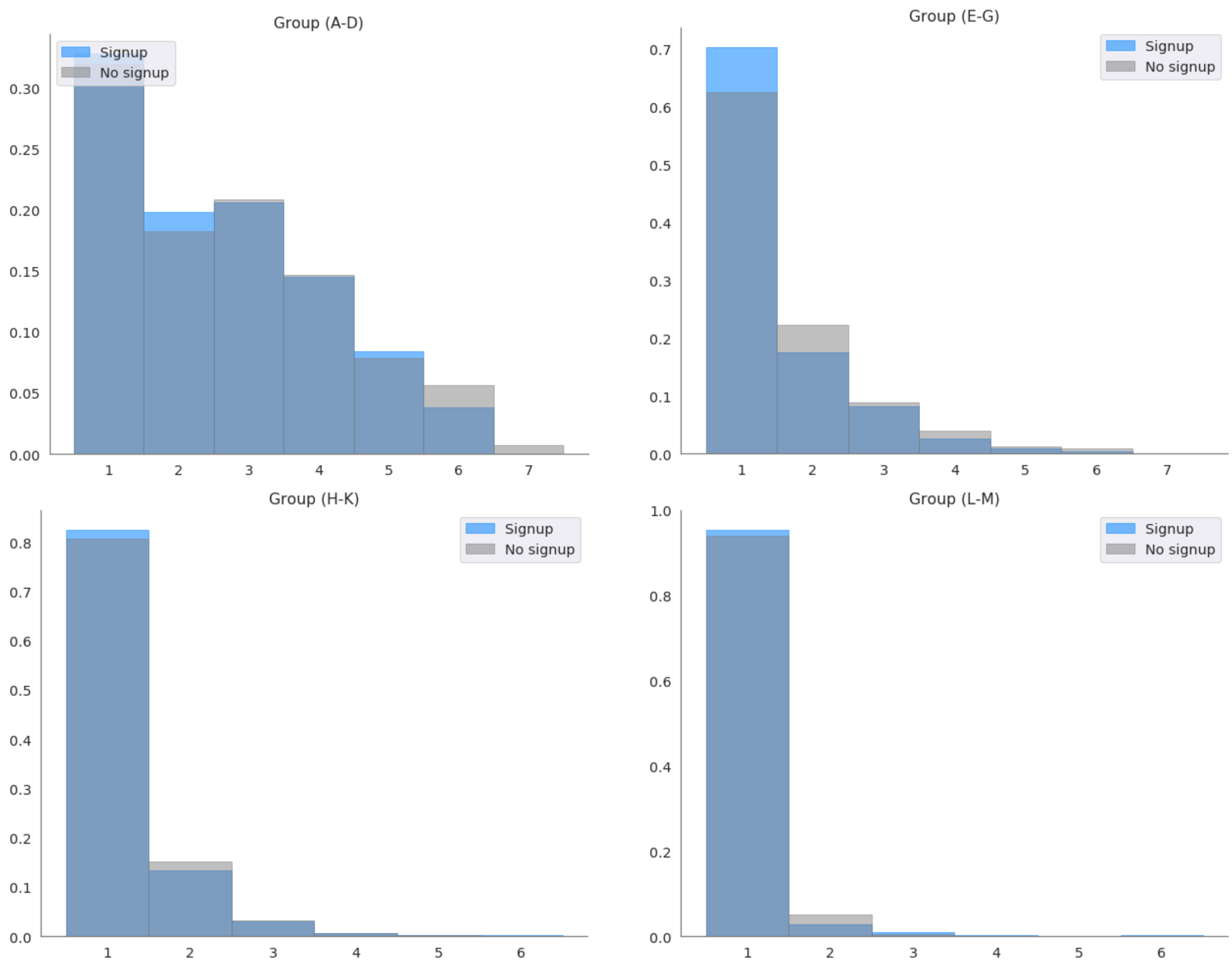
Note: X-axis is the range of scores provided by the platform upon completion of the application, 7 being the highest and 1 being the lowest (rejected). The applicant decides whether to share the data before receiving the loan approval decision, the platform score, and the interest rate. Y-axis indexes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

	Homepage		Repeat		Price comp		Broker		Others(banks)	
variable	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
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.