Open Banking and Customer Data Sharing: Implications for FinTech Borrowers*

Rachel J. Nam†

Abstract

Open banking enables loan applicants to easily and securely share financial data with lenders. In theory, this can expand credit access by reducing information asymmetry, but it may also facilitate price discrimination. Using data from Germany's largest fintech lender in consumer credit, I examine who shares data and how this affects lending outcomes. Observably riskier applicants for whom information asymmetry is greater are more likely to share. Data sharing improves approval and pricing through better risk assessment and is associated with lower ex post defaults. However, non-disclosure leads to negative inferences, with implications for privacy-conscious applicants.

Keywords: Open banking, FinTech, Marketplace lending, Big Data, Data sharing

---

*I am deeply indebted to my thesis advisor, Loriana Pelizzon, for her invaluable guidance and support. I am grateful to Matthias Efing, Andreas Fuster, Luca Gemmi, Pulak Ghosh, Will Gornall, Jacob Hartwig, Jing Huang, Mark Jansen, Xiang Li, Felix Martini, Christine Parlour, and Uday Rajan for their thoughtful discussions. I also thank Ashwini Agrawal, Tania Babina, Konstantin Egorov, Andrew Ellul, Michael Haliassos, Zhiguo He, Vasso Ioannidou, Christine Laudenbach, Igor Makarov, Antoinette Schoar, Daniel Streitz, Huan Tang, Anthony Lee Zhang, and the organizers and participants of the following conferences and seminars for valuable comments: NBER Innovative Data in Household Finance 2022; Bank of England Annual Research Conference 2023; Swiss Winter Conference on Financial Intermediation 2023; Future of Financial Information 2023; SFS Cavalcade NA 2023; FIRS 2023; SAFE Household Finance Workshop 2023; Bank of Canada Payments and Securities Settlement Workshop 2023; German Finance Association Annual Conference 2023; Federal Reserve Bank of Philadelphia Conference on New Perspectives on Consumer Behavior 2023; OFR Rising Scholars Conference 2024; WFA 2024; NBER SI 2024; EFA 2024; Bayes Business School; BI Norwegian Business School; Rotterdam School of Management; Riksbank; Nova SBE; LSE Finance; IESE Business School; CEMFI; USI Lugano; Federal Reserve Board; UBC Sauder School of Business; Carnegie Mellon Tepper School of Business; University of Naples Federico II; Indian School of Business, Hyderabad. I also thank Auxmoney for providing the data and the Leibniz Institute for Financial Research SAFE for a stimulating research environment.

This paper was awarded the *Brattle Group Ph.D. Candidate Awards For Outstanding Research*, WFA, 2024; *JFI Ph.D. Student Paper Award*, FIRS, 2023; *Best Ph.D. Paper Award*, the Future of Financial Information Conference 2023 (sponsored by BlackRock); *Best Paper Award in Digital Finance*, the DGF German Finance Association 2023; and was selected as *finalist* for the ECB Young Economist Prize 2023.

†Universita della Svizzera italiana (USI Lugano) and Swiss Finance Institute. Via Giuseppe Buffi 6, 6900 Lugano, Switzerland. Email: rachel.nam@usi.ch

1 Introduction

Open banking aims to expand consumer control over personal financial data through data portability and informed consent. In the European Union, the regulatory framework requires banks to provide standardized access to bank transaction data when consumers give explicit consent. This allows prospective borrowers to easily and securely share their financial records with any potential lenders during the loan application process.

However, it is not obvious whether voluntary data disclosure will benefit borrowers. First, the impact critically depends on who chooses to share. In principle, individuals should form rational expectations about their type relative to the population and the quality of their data and disclose only if they expect the data to improve outcomes. In practice, individuals may lack the information or sophistication to make such strategic decisions, and disclosure choices may reflect noise, uncertainty, or concerns about privacy. Second, the impact depends on the informativeness of the data. In an advanced economy with extensive credit bureau coverage, shared transaction data may offer little incremental value and, even when informative, its impact on outcomes may be marginal. Third, it matters how the data are used. If the lender primarily uses the data to improve credit risk assessment, disclosure can reduce information asymmetry and improve borrower outcomes. Alternatively, the lender may use the data to infer borrower-specific preferences, such as willingness to pay or price sensitivity, which may result in higher prices, a mechanism consistent with recent theoretical work on digital price discrimination (Bonatti and Cisternas 2020; Ichihashi 2020; He, Huang, and Zhou 2023; Liu, Sockin, and Xiong 2023).

This paper provides the first empirical analysis of voluntary data disclosure in open banking within consumer credit markets. Using loan application data from Germany's largest fintech consumer lender covering more than 2 million completed applications between 2018 and 2022, I micro-found the key forces shaping the disclosure decision and its consequences. Specifically, I examine the factors influencing the willingness to share, estimate its impact on loan approvals and interest rates, and test whether the disclosure patterns are consistent with rational expectations, namely, whether latent good type borrowers self-select into sharing, and whether non-disclosure leads to adverse inference as predicted by theory.

The findings reveal four main patterns. First, observably riskier applicants (those with lower credit scores, non-homeowners, or high debt burdens) are more likely to share. These are precisely the applicants for whom information asymmetry is expected to be greater. Second, data sharing improves approval rates and pricing, but the effects are highly heterogeneous and vary systematically with applicants' observable credit quality. I show that these improvements are driven by the lender's use of the shared data to update beliefs about borrower risk, rather than treating disclosure as a blanket positive signal or simply rewarding applicants for sharing. Third, data sharing is associated with lower ex post default, suggesting that borrowers who disclose are indeed of higher quality, consistent with rational expectations and strategic disclosure. Finally, non-disclosure leads to negative inference, which may disadvantage privacy-conscious individuals despite the voluntary nature of the regime.

The literature on the disclosure of personal data distinguishes between intrinsic preferences, a general reluctance to reveal private information (Warren and Brandeis 1989), and instrumental preferences, where individuals weigh the economic benefits of the disclosure (Stigler 1980; Posner 1981). Previous empirical studies explore these motivations (Goldfarb and Tucker 2012; Acquisti, John, and Loewenstein 2013; Athey, Catalini, and Tucker 2017; Tang 2019b; Lin 2022; Armantier, Doerr, Frost, Fuster, and Shue 2024) but typically focus on non-financial data, rely on survey responses, or are conducted in low-stakes experimental settings. The setting in this paper offers two advantages. First, I observe real disclosure choices of bank transaction data linked to actual lending outcomes. Second, the dataset includes a rich set of variables such as demographics, standard risk measures, and eventual lending outcomes, allowing me to jointly examine intrinsic and instrumental motives behind disclosure.

The results show that instrumental incentives significantly drive disclosure decisions. Applicants with the lowest credit scores are about 30 % more likely to consent to data sharing than those with the highest scores, and the willingness to share decreases monotonically with credit scores. Non-homeowners and borrowers with more past or outstanding loans also exhibit a substantially higher propensity to share.

Observably riskier applicants, such as those with lower credit scores, no homeownership, or high debt burdens, who are otherwise creditworthy have stronger incentives to share data, driven by expectations that it may improve their approval odds (extensive margin) and loan pricing (intensive margin). Since securing a loan is not certain ex ante, sharing becomes a strategic tool to show creditworthiness. By contrast, applicants with strong observable traits are already likely to be approved, so their primary incentive is to secure better terms rather than access to credit. Moreover, applicants may also withhold data due to model uncertainty, concerns over how the lender interpret transaction data, especially when existing credit metrics already portray them favorably (Harbaugh and To 2020; Bond and Zeng 2022; Quigley and Walther 2024).

Intrinsic preferences also play a meaningful role. Female applicants are less likely to share data than male applicants, with this gender gap widening over time, highlighting systematic differences in privacy concerns or risk perception between genders. Younger applicants exhibit a higher willingness to share data, indicating generational differences in attitudes toward privacy and trust in new technologies.

Having established who shares data and why, I next examine how data sharing impacts lending outcomes. Since data sharing under open banking is voluntary, applicants who disclose likely possess favorable private information. Therefore, the impact of disclosure in this setting cannot and should not be interpreted as a causal average treatment effect (ATE), which would require random assignment of disclosure. Instead, I estimate an average treatment effect on the treated (ATT), focusing on those who voluntarily opt in. Thus, the question I ask is: Given that applicants who share data do so because they possess favorable private information, to what extent does disclosure impact loan approval and loan rates?

To estimate this effect, I construct a counterfactual by matching applicants who share data to non-sharers who are observably identical on credit scores, demographic attributes, loan characteristics, and timing of applications. The key assumption is that, absent data sharing, two identical applicants should receive the same lending decisions. Since the lender makes decisions only based on the information available to them, unobservable applicant traits, while potentially influencing disclosure, cannot directly influence lending outcomes. Additionally, the automated online lending context eliminates the role of soft information from in-person interactions, further mitigating selection concerns about unobserved traits biasing outcomes.

The results show that data sharing increases loan approvals, but the effect varies substantially across applicants. Moderate-to-low score applicants benefit the most, with approval rates rising by over 43 %, suggesting that they are near the lender's approval threshold where even slight improvements in perceived risk can tip the decision toward approval. By contrast, high score applicants gain little (1.7 %) since their approval odds are ex ante high, leaving little room for improvement on the extensive margin.

The value of data sharing also varies by other applicant characteristics, notably homeownership. Even for unsecured loans, homeowners have a 4.3 percentage points (7.3 %) higher likelihood of loan approval compared to non-homeowners, suggesting homeownership may act as a signal for greater financial stability, or an implicit fallback in case of defaults. However, non-homeowners benefit comparably from data sharing, achieving larger increases in approval rates. This indicates that providing banking transaction data helps bridge informational gaps for those without traditional assets, reducing the lender's reliance on collateral-like signals.

Data sharing also reduces loan rates, but the effects vary widely. High score applicants benefit the most, with rate reductions of over two percentage points (20%) compared to observably identical non-sharers. These gains are concentrated among applicants who were already likely to be approved. By contrast, low score applicants see only modest reductions of around 0.2 percentage points (4%). Similarly, for non-homeowners, data sharing offsets only about one-fifth of the rate advantage enjoyed by homeowners.

To understand the source of this heterogeneity and, more broadly, how the lender uses the shared data, I examine changes in *Auxscore*, the lender's proprietary internal risk score that reflects its updated assessment of applicant creditworthiness after incorporating shared transaction data. If the lender were merely using data sharing as a positive signal or engaging in a barter-like exchange where applicants receive benefits simply for giving data, those who share should always be weakly better off. However, my findings challenge this view. In about 11 % of the cases, applicants who share data receive a lower *Auxscore* compared to observably identical non-sharers, some resulting in higher loan rates or even credit denial. This result rules out the possibility that the lender uses disclosure purely as a positive signal or as an exchange mechanism.

The heterogeneity in pricing effects mirrors this pattern. High score applicants whose approval is ex ante likely see substantial rate reductions, as the lender uses the additional data to fine-tune pricing decisions. By contrast, low score applicants remain classified as higher risk even after disclosure, which limits improvements in loan terms. For non-homeowners, data sharing does not fully substitute for the rate advantage provided by homeownership. I show that these effects grow more pronounced over time, consistent with platform learning and model refinement.

Even after applying matching techniques, certain applicant characteristics, digital footprints from platform engagement, are available to the lender but not to the econometrician. If these digital traces are correlated with both the decision to share data and lending outcomes, they could introduce bias in the estimated effects. To address this concern, I implement robustness checks using individual fixed effects, focusing on applicants who submit multiple loan applications within the same day—first without data sharing, then with it. Given the narrow time window, borrower characteristics remain constant, isolating the impact of the data sharing decision. Results remain robust across different model specifications.

Theories of asymmetric information predict that latent higher quality borrowers should have stronger incentives to disclose private information, as doing so allows them to distinguish themselves from riskier applicants (Viscusi 1978; Grossman 1981; Jovanovic 1982). However, in practice, individuals face numerous frictions, such as uncertainty about their's own type, the quality of their data, behavioral biases, or privacy concerns, which all may complicate this decision, challenging the simple "nothing to hide" logic. To test whether disclosure patterns are consistent with rational expectations, I use ex post defaults as a proxy for latent borrower quality. Given that data sharing influences the interest rate, which in turn can impact ex post defaults, I conduct a causal mediation analysis to isolate the role of the latent borrower type from both the effect of pricing on defaults and the platform's improved screening ability. The results show that sharers have approximately 16 % lower ex post defaults compared to observally identical non-sharers, consistent with strategic disclosure by higher quality borrowers. However, this effect weakens among lower score borrowers, who may share simply because they have little to lose. This pattern may also reflect differences in financial sophistication, as lower score applicants might not fully grasp their own risk profile.

While latent good type borrowers tend to disclose their data, a key question arises regarding its equilibrium inferences for non-disclosers (Akerlof 1970). If the lender infer that non-disclosure is a signal of lower quality, their loan approval odds may decline. To test this, I exploit variation in data sharing rates over time across observably similar borrower pools to examine whether non-disclosure leads to adverse outcomes as the share of disclosers increases. The findings indicate a negative yet modest effect on loan approval rates for non-disclosers, suggesting that while some adverse inference occurs, its magnitude is so far limited. This suggests that the perverse effects of non-disclosure may not fully materialize in the presence of high quality yet privacy-conscious borrowers who also choose not to disclose (He, Huang, and Zhou 2023).

These findings have important policy implications. The pronounced positive effects of data sharing on loan approvals for those with lower credit scores, without traditional collateral, such as houses, suggest that open banking can be particularly beneficial for asset-light and thin credit file borrowers who are otherwise creditworthy. This may give borrowers more choice and flexibility in selecting financial products and could help address hold-up challenges tied to information asymmetry or limited credit avenues by lowering search and switching costs (Argyle, Nadauld, and Palmer 2023).

As financial data become more valuable, institutions will increasingly seek access, underscoring the need for clear regulations on consumer consent, data security, and standardized access. Since the trade-offs of data sharing vary across individuals, maintaining it as a voluntary choice is essential, and transparency about data use is crucial for consumers to make informed decisions. Open banking frameworks, particularly in light of GDPR and global data protection laws, can guide this process. As data sharing expands, policymakers must balance consumer choice with protection to ensure privacy-conscious borrowers are not disadvantaged.

This paper contributes to the growing literature on open banking and data portability. It complements recent theoretical work showing that the welfare effects of open banking depend on how it alters competition between FinTech lenders and traditional banks (He, Huang, and Zhou 2023) and on heterogeneity in borrower-bank affinity (Parlour, Rajan, and Zhu 2022). These models diverge in their predictions: Parlour, Rajan, and Zhu 2022 anticipate full unraveling, while He, Huang, and Zhou 2023 show that privacy-conscious borrowers can sustain partial disclosure equilibria. Babina, Bahaj, Buchak, De Marco, Foulis, Gornall, Mazzola, and Yu Forthcoming further emphasize that welfare implications depend critically on how shared data are used. While data access enhances consumer welfare when used to deliver financial advice, its use for credit decisions may produce mixed effects, as improved screening can be accompanied by higher prices for some borrowers. Other theoretical contributions highlight additional risks, including allocative inefficiencies arising from winner's curse dynamics among banks (Goldstein, Huang, and Yang 2022) and shifts in market power toward dominant digital platforms, which may enable rent extraction and distortions in adjacent markets (Brunnermeier and Payne 2024). Together, these studies offer rich theoretical insights, but empirical evidence at the individual level remains scarce. This paper contributes by jointly examining the strategic and privacy considerations behind data disclosure, quantifying the benefits of data sharing, and empirically assessing potential adverse effects for non-disclosers.

This paper also relates to the literature on alternative data in credit markets. Empirical studies show that alternative data can improve credit access and pricing efficiency (Jagtiani and Lemieux 2019; Berg, Burg, Gombović, and Puri 2020). Payment and digital footprints have been found to outperform traditional credit scores (Gambacorta, Huang, Li, Qiu, and Chen 2020; Di Maggio, Ratnadiwakara, and Carmichael 2022; Rishabh 2024). Building digital transaction histories can enhance financing outcomes for SMEs and consumers, especially in countries with less robust financial markets (Ghosh, Vallée, and Zeng Forthcoming; Ouyang 2021; Alok, Ghosh, Kulkarni, and Puri 2024). While prior work focuses on how lenders evaluate alternative data when it is universally available, I examine how voluntary disclosure—through an opt-in choice—shapes borrower behavior, lender inferences, and lending outcomes for both sharers and non-sharers.

Finally, this study adds to the literature on technology's role in reducing market inefficiencies and disparities in credit access. FinTech lenders have been shown to expand small business credit in underserved areas, particularly in regions with fewer bank branches, lower incomes, and a higher share of minority households (Erel and Liebersohn 2022) as well as in areas with higher bankruptcy filings and unemployment (Cornelli, Frost, Gambacorta, and Jagtiani 2022). FinTech loans can ease financing constraints for SMEs and expand access to bank credit by enabling borrowers to acquire pledgeable assets (Beaumont, Tang, and Vansteenberghe 2022; Eça, Ferreira, Prado, and Rizzo 2022). The use of big data and algorithmic decision-making can mitigate human biases in credit markets (Philippon 2019), helping to reduce racial disparities in small business credit (Howell, Kuchler, Snitkof, Stroebel, and Wong 2021), improve loan processing efficiency in mortgage lending (Fuster, Plosser, Schnabl, and Vickery 2019), reduce agency conflicts and enhance underwriting efficiency (Jansen, Nguyen, and Shams 2025). These fintech lenders interact with traditional banks in different ways: they may compete by serving borrowers willing to pay a premium for immediacy (Buchak, Matvos, Piskorski, and Seru 2018; Tang 2019a) or complement bank lending by absorbing unmet demand (Sheng 2021; Avramidis, Mylonopoulos, and Pennacchi 2022; de Roure, Pelizzon, and Thakor 2022; Gopal and Schnabl 2022).

2 Institutional Setting and Data

This section provides the institutional background of open banking and the lender, descriptive statistics, and descriptive evidence of open banking.

2.1 Open Banking Regulation

Open banking aims to grant consumers greater control over their financial information by allowing them to decide what data to share, with whom, and for what purpose. As of October 2021, 80 countries have undertaken government-led initiatives to promote open banking (Babina, Bahaj, Buchak, De Marco, Foulis, Gornall, Mazzola, and Yu Forthcoming).[1] While some jurisdictions mandate bank participation under consumer consent, others take a market-driven approach, encouraging voluntary data sharing and

[1] Open banking continues to expand worldwide. In the U.S., the Consumer Financial Protection Bureau finalized its Personal Financial Data Rights rule in 2024, establishing a regulatory framework for open banking.

setting technical standards to facilitate the process.[2]

The revised Payment Services Directive (PSD2) established the EU's open banking framework, which requires all financial institutions offering payment accounts to grant third parties, both banks and non-banks, access to consumers' payment account data. This includes transaction history and account balances from checking accounts, current accounts, and credit card accounts, provided consumers give explicit consent.[3] To facilitate secure data exchange, these institutions are also required to implement dedicated application programming interfaces (APIs).[4]

This regulatory framework makes the EU a suitable setting to study the impact of consumer-permissioned financial data sharing. Germany, the focus of this study, incorporated the revised PSD2 into its national legal framework on January 13, 2018, establishing a uniform mandate for open banking-driven data sharing. Accordingly, this study examines loan applications submitted between January 13, 2018, and May 22, 2022, a period of consistent regulatory enforcement. As an advanced economy with extensive credit bureau coverage and robust financial infrastructure, Germany offers a relevant point of comparison for other advanced economies pursuing similar reforms. For example, the U.S. Consumer Financial Protection Bureau (CFPB) finalized its rule on personal financial data rights in October 2024, establishing a formal open banking regime.

While open banking shares some similarities with credit bureaus and registries (Djankov, McLiesh, and Shleifer 2007; Hertzberg, Liberti, and Paravisini 2011), it differs in several important respects. In Germany, the credit information infrastructure includes both a public credit registry and private credit bureaus. The public registry, administered by the Bundesbank, covers exposures above EUR 1.5 million and collects data on loan amounts, credit exposures, collateral, loan type, maturity, and borrower identifiers. Its primary purpose is macroprudential oversight and systemic risk. By contrast, private credit bureaus such as *Schufa* compile individual-level data on credit products, outstanding and

[2]Countries with mandatory data-sharing rules include the European Union, Australia, Brazil, Israel, and the United States. By contrast, jurisdictions such as Singapore and Malaysia adopt a voluntary approach, encouraging banks to share consumer data and focusing on the development of technical infrastructure and industry coordination (Babina, Bahaj, Buchak, De Marco, Foulis, Gornall, Mazzola, and Yu Forthcoming).

[3]The European Commission is now expanding this framework through the proposed Financial Data Access (FiDA) framework, which would extend data-sharing beyond payment accounts to include savings, investments, pensions, and insurance.

[4]APIs serve as standardized digital interfaces that enable secure and automated data transfers between financial institutions and third-party providers. By reducing reliance on outdated methods such as screen scraping, APIs lower costs and enhance security, thereby improving consumer participation in open banking.

past debts, and repayment history. These bureaus are the primary source of credit information used in evaluating retail borrowers.

2.2 Description of the Platform and Data Sharing Process

This study is based on loan application data from Auxmoney, Germany's largest FinTech lending platform. Founded in 2007, Auxmoney operates entirely online and has originated over EUR 2.3 billion in 319,535 consumer loans between January 2018 and May 2022, and over EUR 3 billion since inception, making it one of the largest marketplace lenders in continental Europe (Figure 1). Unlike traditional banks, Auxmoney is not a licensed financial institution and operates by partnering with a fully licensed credit institution to issue loans.

[Figure 1]

The financing for loans comes from both individual and institutional investors. Auxmoney initially operated as a decentralized peer-to-peer (P2P) lending platform, where retail investors directly selected loans to fund. Over time, it transitioned to a marketplace model, where investors were given the option to invest in a diversified loan portfolio. As the platform evolved, institutional investors became the dominant funding source (Balyuk and Davydenko 2019; Vallee and Zeng 2019; Braggion, Manconi, Pavanini, Zhu, et al. 2020).[5]

The loan application process consists of three key stages: (i) application and data sharing (ii) decision, and (iii) loan payout and repayment.

In the application stage, applicants specify the desired loan amount (EUR 1,000–EUR 50,000), loan purpose, and provide personal and financial details, including monthly income, recurring expenses, and assets. Following this, applicants are presented with the option to share their bank transaction data to support their application. This process is facilitated through a secure API interface, allowing applicants to log into their bank accounts and grant access to the last four months of banking transactions.[6] A standardized message is displayed outlining the potential benefits of data sharing, while also disclosing that it may have negative consequences, such as higher interest rates or loan rejection.[7]

[5] Auxmoney started issuing asset-backed securities (ABS) in recent years (Fortuna Consumer Loan ABS transactions: 48,000 loans (EUR 350M) in 2023, 25,000 loans (EUR 225M) in 2022, and 30,000 loans (EUR 250M) in 2021.)

[6] The platform ensures security through a three-factor authentication process. Bank login credentials remain confidential and are never visible to the platform.

[7] See Figure A.1.

Importantly, data sharing is a one-time consent-based process during the loan application stage and is not used to track ongoing financial activity after the loan is issued. This ensures that data sharing remains strictly limited to the underwriting phase and does not extend into post-loan monitoring.

Once an application is submitted, Auxmoney evaluates the applicant's creditworthiness using four primary data sources: (1) application data, including personal and financial details provided by the applicant; (2) credit bureau data, primarily Schufa credit scores and additional financial history[8]; (3) digital footprint data, leveraging online behavioral patterns to enhance risk assessment; and (4) bank transaction data, if the applicant consents to sharing.

Unlike traditional banks, which often exclude certain applicant groups such as students, self-employed individuals, or temporary workers,[9] Auxmoney does not impose blanket exclusions (except in the case of past default history), relying instead on a broader credit assessment framework that incorporates real-time financial behavior and alternative data sources. By doing so, the platform claims to conduct a more flexible evaluation of applicants' creditworthiness.

In the decision phase, each applicant receives either loan approval or rejection based on the internal proprietary risk score, *Auxscore* (AA, A, B, C, D, E, or Z), where AA represents the highest credit quality and Z indicates ineligibility for a loan.[10] The score reflects the lender's overall risk assessment, incorporating all available information, including transaction data for applicants who choose to share it. Approved applicants are also assigned an interest rate.

In the final, loan payout and repayment phase, the applicant decides to accept or decline the loan offer, leading to either the disbursement of funds or termination of the process. If the applicant accepts the loan, she will proceed to either repay it or default.

[8] *Schufa* scores, generated by Schufa Holding AG, are Germany's equivalent to the US FICO score. They differ by using a discrete scale from A (best) to M (worst). Unlike in the U.S., German credit scores are assigned even without extensive borrowing history, as basic financial activities, such as maintaining a checking account contribute to the score.

[9] Traditional banks operate under stricter capital regulations, increasing the cost of lending to riskier borrowers (Berger and Udell 1994; Kashyap, Stein, et al. 2004; Popov and Udell 2012; Roulet 2018; Benetton, Eckley, Garbarino, Kirwin, and Latsi 2021).

[10] For analysis, I convert these categories to a numerical scale from 7 (AA) to 1 (Z).

2.3 Descriptive Statistics

The share of loan applications in which banking transaction data was shared increased steadily over time (Figure 2). This upward trend is evident across all credit score groups, with a higher incidence of data sharing among lower score applications. The steady rise in data sharing rates aligns with expectations that open banking adoption would grow as FinTech lenders refine their business models (He, Huang, and Zhou 2023). This trend is not driven by compositional shifts in applicant age, as the age distribution remains stable over time (Figure A.2).

[Figure 2]

On the platform, applicants may submit multiple loan applications over time. Including multiple applications from the same applicant could lead to overweighting this subgroup. Moreover, this study aims to examine both the unconditional probability of data sharing (how applicants make decisions without prior information) and its subsequent consequences. To ensure consistency in the analysis, the sample is restricted to only the first application per borrower within the sample period. The final sample consists of 2,309,359 completed loan applications submitted between January 13, 2018, and May 15, 2022.

Table 1 presents summary statistics for the final sample. The average requested loan amount is EUR 13,876, with an average loan term of 55 months. The mean applicant age is 38, and 65% of applicants are male. 39% of applicants are married, and 62% are the primary earners in their households. The platform has approved approximately 68% of loan applications, with an average interest rate of 12%. The average credit score is 3.13 on a 4–1 scale (where 4 represents the highest credit quality).[11] The average platform score, which is based on the lender's internal credit risk model, is 2.89 (where 7 represents the highest quality). The median applicant has a monthly net income of EUR 1,800 and monthly expenses of EUR 600. Most applicants (94%) have at least one checking account, and 64% hold at least one credit card. Homeownership is reported by 25% of applicants, while 57% own at least one automobile. The variables *No. of current loan demands* and *No. of past loan demands* proxy for outstanding and previously held loans, respectively. On average, applicants have 1.35 active consumer loans and a history of approximately one fully repaid loan.

[11] Credit scores are categorized numerically such that higher values correspond to stronger credit quality: 4 for scores A–D (highest), 3 for scores E–G, 2 for scores H–K, and 1 for scores L–M (lowest).

[Table 1]

The primary variable of interest, *Signup*, is an indicator variable equal to one if the applicant shared bank transaction data during the loan application process. The data sharing rate in the final sample is 8%. This figure varies significantly over time, exceeding 25 in the unrestricted sample toward the end of the observation period. Descriptive statistics by data sharing decision are reported in Table A.1.

[Figure 3]

Panel A in Figure 3 presents the unconditional mean loan acceptance rates by data sharing status across different credit score groups. The approval rate is higher for applicants who share data, with the difference more pronounced among lower score applicants. Panel B in Figure 3 shows that data sharing is also associated with lower interest rates across all credit score groups. The difference in interest rates is largest among the highest score applicants, while it appears more modest for those with lower scores.

3 Factors Influencing Willingness to Share Data

Disclosure decisions reflect trade-offs between economic benefits (instrumental motivations) and costs (intrinsic preferences such as privacy concerns or uncertainty in lender interpretation). Thus, analyzing who discloses data, and why, is essential to assess open banking's overall impact. In this analysis, I focus on *observable* factors.

To estimate the factors influencing one's willingness to share data, I use a probit model and estimate the following:

$$\Pr\ (Sign\,up_i = 1) \ = \ \Phi\ (X_i'\beta \ + \ G_i'\gamma \ + \ Time), \tag{1}$$

where i indexes an applicant and $Sign\,up_i$ is an indicator variable equal to one if the applicant shares data and zero otherwise. X_i includes applicant characteristics such as age, income, credit score, gender, marriage status, main earner status, homeownership, car ownership, number of outstanding loans, and number of fully repaid loans. G_i captures loan characteristics, including loan amount, loan duration, and loan application channel.[12] *Time* includes month-year dummies, and Φ is the standard normal cumulative

[12]Loan access channel is a categorical variable that indicates the channel through which the user applies for a loan. There are five such channels: (1) directly via the Auxmoney homepage, (2) repeat loan, (3) price comparison websites, (4) brokers, and (5) banks.

distribution function. The main coefficient of interest is β, which captures how different applicant characteristics influence the likelihood of sharing data.

Table 2 reports the estimation results of equation (1). Column (1)–(3) report marginal effects using probit, and Column (5)–(8) report ordinary least squares estimates.

[Table 2]

The results indicate that observably riskier applicants—those with lower credit scores, no homeownership, or high debt burdens—are significantly more likely to share their bank transaction data. An applicant in the lowest credit score category is, on average, 3.9 percentage points more likely to share data than an applicant in the highest score category, as shown in column (1), and this likelihood decreases with better credit scores. After controlling for additional applicant characteristics in column (4), the difference remains 2.2 percentage points, implying a 33% higher likelihood of data sharing for the lowest credit score group relative to the highest.[13]

For low score applicants who are otherwise creditworthy, securing a loan is not guaranteed ex ante. Thus, their economic incentives for data sharing are primarily driven by increasing loan approval chances (extensive margin) and improving loan pricing (intensive margin). In this context, data sharing serves as a strategic tool to show creditworthiness. By contrast, high score applicants face a lower rejection, so their primary motivation for sharing data is to secure better loan terms rather than access to credit.

Other instrumental motivations are also associated with higher data sharing rates. Non-homeowners are 0.84 percentage points (14.2%) more likely to share data than homeowners, suggesting that applicants without collateral-like assets may use transaction data to demonstrate their creditworthiness. Likewise, applicants with more outstanding loans and a history of fully repaid past loans show a higher propensity to share, indicating that individuals with greater debt burdens may rely on data sharing to improve lending outcomes.

Income does not appear to be a significant determinant of data-sharing decisions. This finding complements Tang 2019b, who documents no heterogeneity across income levels in willingness to disclose non-financial data, such as social network IDs and employer contacts. My analysis extends this insight to financial data.

[13] Even though the overall rate of data sharing may appear low in the main sample due to restricting the sample to one application per borrower in the case of multiple applications, participation has increased substantially over time, rising from 4% in 2018 to around 15% in 2021 and exceeding 25% in the unrestricted sample. See Figure 2 for details.

Beyond economic factors, uncertainty about how the shared data are interpreted can also influence disclosure decisions (Harbaugh and To 2020; Bond and Zeng 2022; Quigley and Walther 2024). Applicants who are deemed creditworthy based on outside information such as standard risk metrics may withhold private data even if it can be beneficial if they are unsure how the lender will interpret it, whereas those with weaker observable credit signals may be more willing to tolerate this uncertainty, as sharing represents a potential opportunity to demonstrate creditworthiness. In other words, those with weaker credit signals may perceive greater potential upside, while those with stronger signals may see limited benefits or even downside risks, leading to a differential willingness to share data.

Intrinsic preferences further shape data sharing choices. Female applicants are 0.4 percentage points (6%) less likely to share data than male applicants, and older applicants exhibit lower willingness to share, with a 48-year-old applicant being 2 percentage points (25%) less likely to share than a 38-year-old applicant. These findings align with prior research indicating that women and older individuals tend to have stronger privacy concerns, leading to a general reluctance to disclose personal financial information (Goldfarb and Tucker 2012). Importantly, despite the potential economic gains from disclosure in my setting—conditions under which the "privacy paradox" arises wherein individuals often relinquish stated privacy preferences for tangible benefits (Athey, Catalini, and Tucker 2017)—the demographic heterogeneity in intrinsic preferences persists. Notably, the gender gap in willingness to share has widened over time, reaching 1.9 percentage points (10%) by the end of the sample period (Figure A.3).

4 Data Sharing on Lending Outcomes

Building on the analysis of consumers' willingness to share bank transaction data, the next step is to examine its impact on lending outcomes. Specifically, the analysis focuses on how data sharing influences loan approval and interest rates, assessing whether bank transaction data provide incremental information beyond the traditional credit metrics already available to the lender. Quantifying the direction and magnitude of these effects helps determine whether disclosure benefits applicants and provides insight into the subsequent analysis on how the lender processes and utilizes the shared information.

To estimate the impact of data sharing on loan approvals and interest rates, I focus on the average treatment effect on the treated (ATT) rather than the average treatment effect (ATE). Identifying the ATE would require randomly assigning applicants to share or withhold their banking data, which is inconsistent with the open banking framework. Because participation is voluntary, those who share likely possess favorable information.

Instead, ATT provides the appropriate causal parameter by estimating the effect of disclosure for those who actively opt in. This approach is widely used in economics when treatment is self-selected, particularly in policy evaluations where interventions are designed for a specific subset of individuals rather than the entire population (Heckman and Vytlacil 2005; Imbens and Wooldridge 2009).

4.1 Matching on Observables

To address differences in observed characteristics between applicants who share data and those who do not, I use a hybrid matching approach to construct a counterfactual—estimating what would have happened to data sharing applicants had they not shared. This method combines exact matching and propensity score matching (PSM) to ensure comparability between the two groups.

A key identifying assumption is that, in the absence of data sharing, otherwise identical applicants should receive the same lending outcomes. Classic selection on unobservables arises when the same unobserved factors influence both treatment assignment and outcomes, but in this setting, loan decisions are based on lender-observed information. While unobserved traits may affect an applicant's decision to share data, they cannot directly impact loan approval or pricing. Moreover, the fully automated online lending process eliminates the role of soft information from in-person interactions, further mitigating concerns about selection bias.

Given that applicant characteristics vary across multiple dimensions and that data sharing rates and macroeconomic conditions fluctuate over time, exact matching is applied to *Credit score*, *Homeowner*, *Female*, *Access channel*, and *Loan application month-year* to ensure precise balance on these categorical variables. PSM is then used for *Age*, *Income decile* and *Loan amount requested*, allowing for greater flexibility by matching applicants with similar propensity scores, estimated via probit regression. The goodness of the matching procedure is assessed with t-tests for the null hypothesis of equal means for both sharing and non-sharing groups. Detailed matching results are reported in Table A.2 and indicate that the matching is successful.

4.2 The Effect of Data Sharing on Loan Approvals

I use the matched sample to estimate the effect of data sharing on the probability of loan approval using a probit model,

$$\begin{aligned} \Pr(Approved_i = 1) = \Phi(&\rho Signup_i + \sigma_k(Signup_i \times Credit\ score_i) \\ &+ \delta(Signup_i \times Homeowner_i) + X_i'\beta + G_i'\gamma + Time), \end{aligned} \tag{2}$$

where $Approved_i$ is an indicator variable that takes a value of one if the loan application is approved and zero otherwise. To examine whether data disclosure has different effects across applicants, I include the interaction terms. The other variables are the same as in equation (1). The main coefficients of interest are ρ and σ_k which measure the change in the likelihood of loan approval by data sharing decision $Sign\ up_i$, and the differential effect across different credit score $k = 4, 3, 2, 1$.[14] δ measures the impact of data sharing by homeownership. In nonlinear models such as probit, the coefficient on an interaction term does not directly capture the interaction effect on the probability scale. Following Ai and Norton 2003, the marginal effect of data sharing, *signup*, is estimated on the predicted probability of loan approval at different levels of the interacting variables, *Credit Score* and *Homeowner*. These marginal effects are defined as the partial derivative of the predicted probability with respect to *Signup*, and are estimated using the delta method. The effects are averaged over the observed covariate distribution. Table 3 reports estimates using both probit and linear probability models.

[Table 3]

Data sharing, on average, improves loan approval rates, but the effects are highly heterogeneous across the credit score distribution. The largest gains are observed among applicants with lower scores. For example, those in the second-lowest group (H–K) experience a 12.8 percentage point increase in approval probability—a 39.5% improvement relative to non-sharing applicants in the same group. Applicants in the lowest group (L–M) see a 29% increase (a 4 p.p.), while those in the second-highest group (E–G) gain around 15% (9.5 p.p.). The effect is smallest for the highest score group (A–D), at just 1.5 percentage points (1.7%).

These results indicate a non-linear pattern. While data sharing is more prevalent among applicants with lower credit scores, the approval gains do not increase monoton-

[14] Numerical values are assigned to the credit bureau score (*Schufa*) categories such that high scores correspond to higher implied credit quality: 4 for scores A–D (highest), 3 for scores E–G, 2 for scores H–K, and 1 for scores L–M (lowest).

ically with decreasing scores. One interpretation is that the additional information provided through data sharing is most consequential for applicants whose baseline approval probability is neither too high nor too low—where incremental data may meaningfully affect the lender's assessment. To assess the degree of information asymmetry, I estimate a probit model of ex post default on observables separately by credit score group. The explanatory power of these observables is lowest for the second lowest credit score group (H–K) group ($R^2 = 0.075$), compared to 0.102 for the highest (A–D), 0.087 for E–G, and 0.157 for L–M. This suggests that data sharing adds the most value in assessing approval decisions precisely where standard risk metrics are least informative.

By contrast, the lowest score group (L–M) may be too risky ex ante for shared data to meaningfully alter the outcome. Moreover, applicants in this group may be less strategic in their disclosure—more likely to share opportunistically if they believe they have little to lose, a hypothesis explored later. For the highest score group, where approval is already nearly assured (88% among non-sharers), the scope for improvement is naturally limited.

The marginal benefit of data sharing is also significantly larger for applicants without homeownership. While homeownership is positively associated with loan approval-likely reflecting its role as a proxy for financial stability or implicit collateral in case of defaults-the interaction between data sharing and homeownership is negative and statistically significant (-9 p.p.). This suggests that data sharing plays a compensatory role for non-homeowners, helping to offset the informational disadvantage associated with lacking collateral-like assets. The approval gain from data sharing among non-homeowners exceeds the stand-alone effect of being a homeowner (4 p.p.), highlighting the potential of financial data to serve as a substitute for standard asset-based signals in credit evaluation.

Overall, the results indicate that data sharing under open banking can improve credit access for applicants otherwise disadvantaged in standard scoring models–such as lower credit scores or the absence of homeownership.

4.3 The Effect of Data Sharing on Interest Rate

As a next step, I examine the effect of data sharing on interest rates using the matched sample.

$$\begin{aligned} r_i = \rho Signup_i + \sigma_k(Signup_i \times Credit\ score_i) + \delta(Signup_i \times Homeowner_i) \\ + X_i'\beta + G_i'\gamma + Time + \epsilon_i, \end{aligned} \tag{3}$$

where r_i indexes the interest rate. The other variables are the same as in equation (1). The main coefficients of interest are ρ and σ_k which capture the change in the interest rate by data sharing decision $Sign\ up_i$, and the differential effect across different credit score groups $k = 4, 3, 2, 1$ conditional on loan eligibility. δ measures the impact of data sharing by homeownership.

Table 4 reports the results for the main analysis in Equation (3).

[Table 4]

The results indicate that data sharing is associated with lower interest rates across all credit score groups, though the magnitude of the reduction varies substantially with applicant characteristics. Table 4 shows that the largest rate reductions are observed among applicants with the highest credit scores (A–D), who experience a 2.1 percentage point decrease—representing a 21% reduction relative to the average interest rate of non-sharers 9.9% in this group. This is followed by groups E–G and H–K, who see reductions of 1.8 and 1.1 percentage points (14% and 7.3%), respectively. By contrast, the lowest score group (L–M) benefits the least, with a reduction of just 0.15 percentage point (1%).

This heterogeneity confirms that applicants with stronger observable characteristics—particularly those whose approval is already likely—stand to benefit the most from sharing data in terms of pricing. Because most of these applicants are unlikely to be denied credit, any favorable private information revealed through transaction data allows the lender to refine its risk assessment and reduce the interest rate accordingly. By contrast, for applicants with weaker credit scores, approval is already less likely, and even when approved, their pricing remains high due to the lender's continued perception of risk. As a result, the marginal impact of data sharing on interest rates is smaller.

The pricing benefits of data sharing also vary with applicant characteristics, particularly homeownership status. While homeowners receive, on average, a 1.6 percentage point lower interest rate, the marginal benefit of data sharing is larger for those without homeownership. The interaction between data sharing and homeownership is negative and statistically significant (-0.5 p.p.), indicating that non-homeowners derive greater value from sharing their financial data. However, unlike in the approval decision, where data sharing more than compensates for the lack of homeownership, the pricing advantage it confers is more limited. This suggests that while bank transaction data improve pricing for asset-light borrowers, it does not fully offset traditional signals of creditworthiness for loan pricing.

The limited interest rate reduction for lower score applicants, despite their greater willingness to share and larger gains on the extensive margin, reflects the lender's asymmetric response to data across the credit score spectrum. This asymmetry raises the question of how the lender processes and incorporates shared data into its credit assessment.

4.4 Mechanism: How are the Data used?

To understand the mechanism through which data sharing affects lending outcomes, in particular loan pricing, I examine changes in the platform's proprietary internal risk score, *Auxscore*, which reflects the updated risk assessment after observing transaction data.

[Figure 4]

Figure 4 shows the distribution of *Auxscore* by data sharing status across credit score groups using the matched sample. These patterns suggest that shared data meaningfully shift the lender's perception of applicant risk, with differences not only in average scores but also in the skewness of the distribution, indicating heterogeneous updates across the credit score spectrum.

To quantify this adjustment, I repeat the estimation using the matched sample from Equation (3) but *Auxscore* as a dependent variable. Since these groups are observably identical, the *Auxscore* assigned to non-sharers represents the lender's prior belief, while the difference for sharers reflects the lender's posterior belief after incorporating the shared data.

[Table 5]

Table 5 shows that high score applicants (A–D) experience the largest upward adjustment in *Auxscore*—an average increase of 0.86 points on a scale of 7 (best) to 2 (worst).[15] By contrast, the lowest-score group (L–M) sees an average increase of just 0.07 points. These results suggest that although bank transaction data can be informative across all credit score groups, conditional on approval, the lender's posterior beliefs change the most for applicants with high credit scores.

For high score applicants, data sharing leads to a larger upward revision in the lender's posterior belief about applicant quality, resulting in substantial interest rate reductions.

[15] Because only approved loans are given interest rates, applications receiving a score of 1 (rejection) are excluded from this sample.

For lower score applicants, the revision is smaller conditional on approval, limiting the extent of pricing improvements. This helps reconcile the earlier finding that lower score applicants benefit more on the extensive margin but gain less on the intensive margin.

While this analysis demonstrates that data sharing leads to systematic updates in the lender's internal risk score, one may still question whether these updates reflect a genuine reassessment of applicant risk or simply reward mechanisms tied to the act of sharing. For example, the lender could engage in a barter-like exchange, offering better terms in return for disclosure, or treat sharing as a uniformly positive signal, with the magnitude of the "reward" varying by observed credit characteristics.

If the lender merely rewards disclosure, we would expect sharers to always be weakly better off. However, the evidence rejects this hypothesis. As shown in Figure 5, among matched pairs of observably identical applicants, roughly 11% of sharers receive a lower *Auxscore* than their non-sharing counterparts, with the rate rising to over 12% for the lowest score groups (H–M), but lower for higher score groups. To estimate how often data sharing leads to rejection for applicants who would otherwise have been approved, I train a random forest classifier to predict loan approval probabilities based on observable characteristics. The model is trained on non-sharers and evaluated on a 20% hold-out set, achieving 93.6% accuracy and an AUC of 0.98. This allows me to assess outcomes along a continuous probability scale. When applied to sharers, the model reveals that 3.7% of applicants with predicted approval probabilities above 90% are nonetheless rejected (Figure 6). This finding rules out the possibility that the lender treats disclosure as a positive signal per se or engages in uniform exchange-based pricing. Instead, it suggests that the lender systematically interprets the content of the shared financial data and updates its risk assessment accordingly—-even when doing so leads to worse outcomes for the applicant.

[Figure 5, 6 and 7]

Moreover, Figure 7 shows the effects of data sharing on lending outcomes become more pronounced over time, consistent with learning and model refinement. As the lender accumulates more transaction level data and improves its scoring algorithm, the effects on lending outcomes following disclosure grow in magnitude. This temporal pattern reinforces the conclusion that the lender is not simply rewarding disclosure but actively incorporating the information into its credit assessment process.

4.5 Robustness Checks using Fixed Effects

The main analysis relies on a matched sample approach to estimate the effects of data sharing, pairing sharers with observably similar non-sharers to mitigate selection on observables. Even though this strategy is appropriate in a setting where loan decisions are made by the lender based on observable applicant characteristics, matching may still be vulnerable to bias if certain applicant attributes are observed by the lender but not by the econometrician. A main concern is that platform-based digital footprints such as device type, browsing behavior, or time spent on the application may influence both an applicant's decision to share data and the lender's assessment, potentially biasing the estimated effects despite the fact that the direction of this bias is theoretically ambiguous. Such digital traces have been shown to be predictive of creditworthiness (Berg, Burg, Gombović, and Puri 2020), they are not observable in my dataset.

To address this concern, I exploit within-user variation in data sharing decisions by implementing individual-day fixed effects. Although not very common, applicants submit multiple applications on the same day, often to compare different loan offers. A user first applies without sharing data and subsequently chooses to disclose. Because applicant characteristics are kept constant within a given day, any variation in lending outcomes can be attributed to data disclosure. By conditioning on individual-day fixed effects, I effectively control for all time-invariant user-specific characteristics, offering a stringent robustness test.

This analysis uses a subsample of 34,610 applications submitted by 6,380 unique users who submitted at least one application with and without data sharing on the same day.

[Table 6]

Table 6 shows that the results are both qualitatively and quantitatively consistent with the findings of the main specifications. On the extensive margin, data sharing significantly increases the probability of loan approval, particularly for mid-tier credit score groups. The hump-shaped relationship observed earlier is preserved: effects are smaller for the highest (A–D) and lowest (L–M) score groups, and largest for the second-lowest group (H–K).

Results are also robust on the intensive margin. Data sharing leads to statistically significant reductions in interest rates, especially for applicants with stronger observable credit profiles. As in the main analysis, the rate reduction diminishes with decreasing credit scores. Overall, these results lend further support to the baseline findings that data sharing improves lending outcomes, and that the main findings are not driven by what is unobservable to the econometrician but only to the lender.

4.6 Robustness Checks using Subsample Analysis

The baseline estimates pool data from 2018 to 2022, averaging over a period in which the lender's ability to interpret shared transaction data likely evolved. This suggests that the baseline may understate its impact in more recent years. Although the regressions include month-year fixed effects to absorb common shocks, these cannot capture changes in how the effect of data sharing varies over time. To examine this more directly, I re-estimate the main specifications using only the 2021–2022 subsample, where the effects appear strongest. This approach also has the benefit of excluding the early COVID-19 period, when application volumes declined sharply, and applicant behavior may have been atypical.

The results, reported in Tables A.3 and A.4, are qualitatively consistent with the full sample findings. The approval gain from data sharing for high-score applicants (A–D) remains modest (2.35 percentage points, or a 2.6% increase), while the effects for lower-score groups are substantially larger than in the full sample. On the intensive margin, interest rate reductions continue to be concentrated among high-score borrowers, with effects for lower-score groups still modest but larger than in earlier estimates. Finally, the value of data sharing for non-homeowners grows considerably, more than offsetting the baseline approval advantage associated with homeownership. For loan pricing, the interaction effects by homeownership remain both qualitatively and quantitatively consistent with the full-sample results.

5 Voluntary Disclosure and Latent Borrower Type

The previous sections show that data sharing improves lending outcomes, but a question still remains: Do these benefits reflect the revelation of underlying borrower type? In other words, does voluntary disclosure reflect a form of positive selection, whereby applicants with lower inherent risk are more likely to share?

Theoretical models of strategic information disclosure under asymmetric information (Viscusi 1978; Grossman 1981; Jovanovic 1982) predict such behavior: latent high-quality borrowers have stronger incentives to disclose, allowing them to distinguish themselves from lower-quality types. These models often assume that borrowers are fully informed about their own type relative to the population and the quality of their private information and face no uncertainty about how the disclosed data will be interpreted. However, in real-world settings, borrowers may face uncertainty about their own standing, lack financial literacy, or respond passively to interface design—factors that can distort disclosure decisions away from the predictions of these models. If such frictions dominate, one should not expect a strong association between voluntary disclosure and latent borrower quality—making the "nothing to hide" hypothesis less likely to hold empirically.

To assess whether latent borrower quality plays a role, I examine whether data sharing predicts ex post default using the matched sample. A loan is classified as in default if a payment delay exceeds 90 days. If applicants who share data are systematically less likely to default than observably identical non-sharers, this would suggest that voluntary disclosure is driven by underlying borrower type.[16][17]

One empirical challenge in interpreting the relationship between data sharing and default is that part of the observed effect may operate through improved loan terms. As shown earlier, data sharing reduces interest rates, which may in turn reduce borrowers' repayment burdens and lower the likelihood of default—even in the absence of any difference in underlying borrower quality. Since the interest rate is itself a function of the data sharing decision, simply conditioning on it introduces post-treatment bias (e.g., collider bias), distorting the estimated relationship by blocking part of the causal pathway from data sharing to default (Angrist and Pischke 2009). To address this, I implement a causal mediation analysis that decomposes the total effect of data sharing on default into two components: a direct effect, which reflects differences in latent borrower quality, and an indirect effect, which operates through changes in loan pricing.[18] Importantly, since pricing reflects the lender's internal assessment of borrower risk, this approach also helps

[16] Loan performance data are sourced from the European DataWarehouse (EDW), a securitisation repository designated by both the European Securities and Markets Authority and the UK Financial Conduct Authority. Established in 2012, EDW was the first such repository in Europe, providing standardized loan-level data for asset-backed securities and private whole loan portfolios. See `https://eurodw.eu/` for more information.

[17] The EU Securitisation Regulation prohibits cherry-picking and requires that securitized loan pools be representative of the originator's broader portfolio. See: European Banking Authority (2022), Final Draft Regulatory Rechnical Standards on Risk Retention in Securitisation, EBA-RTS-2022-04.

[18] Mediation analysis is a methodological approach used to decompose the total effect of a treatment variable into a direct effect and an indirect effect that operates through a specified mediator. This envolves two steps: (1) estimating the effect of the treatment variable (data sharing) on the mediator (interest rate), and (2) estimating the effect of the mediator on the outcome (default), while controlling for the treatment. Mediation analysis has been used in applied economics to identify causal mechanisms and quantify the contribution of intermediate variables to overall effects (Hicks and Tingley 2011; Heckman and Pinto 2013; Huber 2014; Carpena and Zia 2020).

account for the platform's improved ability to screen—ensuring that the remaining direct effect captures the borrower type rather than the lender's improved risk assessment.

[Table 7]

Table 7 shows that those who share data are significantly less likely to default than observably identical non-sharers. For the highest credit score group (A–D), data sharers exhibit a 1.0 percentage point lower likelihood of default—equivalent to a 16% reduction relative to the identical non-sharer mean of 6.3%. Similarly, in the E–G and H–K groups, ex post defaults are 1.8 and 2.5 percentage points lower for sharers, corresponding to reductions of 16.4% and 15.5%, respectively.

However, the statistical power of these estimates weakens moving down the credit score distribution. For the lowest group (L–M), the effect is no longer statistically significant. This attenuation reflects that, at the very bottom of the score distribution, data sharing is less effective at distinguishing latent borrower type. One potential explanation is that borrowers in this group may be less informed about their type, or they are less strategic in their disclosure decisions—sharing data not necessarily to signal creditworthiness, but simply because they have little to lose. As a result, voluntary disclosure in this segment may not convey much information about underlying risk. This also helps explain the earlier findings that the effects of data sharing on both loan approval and interest rates are muted for this group.

6 Equilibrium Inferences on Non-disclosure

If borrowers who voluntarily share data are less likely to default than observably identical non-sharers, does the lender rationally revise their beliefs about those who withhold information? The rising prevalence of data sharing may cause non-disclosure to be perceived as a negative signal—implying lower borrower quality. This logic echoes the classic unraveling predictions under adverse selection (Milgrom 1981). The key question, then, is: How do approval prospects change for non-sharers as more applicants choose to disclose? Understanding these equilibrium inferences is critical for evaluating the broader consequences of data portability policies, especially for those who refrain from sharing due to reasons other than credit risk.

To test this implication, applicants are grouped into risk clusters using categorical bins for key observable characteristics. Debt-to-income ratio is calculated as the monthly loan payment divided by net monthly income and categorized as: low ($\leq$ 35%), medium

(35–50%), and high ($> 50\%$). Age is grouped into: under 30, 30-50, 50–65, and over 65. Income is divided into terciles (low, medium, high) based on the empirical distribution. Loan size is categorized as small ($\leq$ EUR 5,000), medium (EUR 5,001–EUR 20,000), and large $>$ EUR 20,000). These categories, together with application month-year, credit score, access channel, homeownership status, and gender define the applicant's reference pool. This grouping ensures that the computed disclosure rate reflects the behavior of applicants who are observably similar on dimensions likely to influence both credit outcomes and the lender's inference process. Within each pool, I compute the share of applicants, *Data sharing rate*, who disclosed data. This pool-level disclosure rate serves as the key independent variable in the equilibrium inference analysis. The main regression is estimated on the subsample of non-sharers.

[Table 8]

Table 8 presents the results from three specifications estimating the effect of pool-level data sharing rates on the probability of loan approval for non-sharers. Across all models—Probit, Linear Probability Model (LPM), and pool-level fixed effects (FE)—the coefficient on the share of applicants who disclosed data in an applicant's reference group is negative and statistically significant, ranging from –0.176 to –0.218. This implies that, as the share of disclosers within a comparable applicant pool increases by 10 percentage points, the probability of loan approval for a non-discloser declines by approximately 1.8 to 2.2 percentage points.

While modest in magnitude, these effects are economically meaningful. They suggest that the lender infer higher risk from non-disclosure as data sharing becomes more common—consistent with the unraveling logic of adverse selection. However, the fact that the effects are not larger suggests limits to full unraveling. As long as some withhold data for reasons orthogonal to risk—such as privacy concerns, model uncertainty, or limited digital engagement—the signal from non-disclosure remains noisy. In such settings, the lender cannot fully interpret withholding as evidence of poor creditworthiness (He, Huang, and Zhou 2023).

7 Conclusion

This paper provides the first empirical analysis of voluntary data disclosure under open banking in consumer credit markets. Using granular loan application data from Germany's largest FinTech lender, I examine who chooses to share bank transaction data, how such disclosures affect lending outcomes, and how lenders incorporate the shared data into their credit assessments. I also consider the equilibrium implications of voluntary disclosure, including how non-disclosure is interpreted by the lender.

I find that applicants with weaker observable credit quality such as those with lower credit scores, no homeownership, or higher debt burdens are significantly more likely to share data. These are precisely the applicants for whom information asymmetry is greater. Data sharing is associated with large gains in approval rates for these groups. Rate reductions, in contrast, are concentrated among applicants with strong observable credit profiles. For lower score applicants, pricing adjustments are modest, as their risk remains high even after disclosure.

To understand these patterns and the mechanism, I examine changes in the lender's proprietary internal risk score and show that the shared data is actively used to revise posterior beliefs. I show evidence that disclosure does not universally lead to better outcomes, challenging the notion that it is treated as a positive signal or rewarded through favorable pricing. Over time, as the lender accumulates more transaction-level data and refines its credit model, the effects of sharing become more pronounced, consistent with platform learning.

Sharers are ex post less likely to default than matched non-sharers, suggesting that individuals self-select into disclosure. This points to a degree of rational expectations in how applicants evaluate the consequences of sharing. However, this pattern weakens among the lowest score borrowers, who may be less informed about their own type or act rather opportunistically because they have little to lose. Finally, I find evidence of negative inference on non-sharers who face lower approval probabilities as data sharing becomes more common, highlighting the trade-off between privacy choice and access to credit introduced by voluntary disclosure in an equilibrium setting.

Taken together, these findings offer novel empirical insights into the strategic dimensions of data sharing in credit markets. They suggest that open banking can help improve credit access and pricing especially for underserved borrowers but may also introduce new frictions for privacy-conscious individuals. As regulators and platforms expand data portability regimes, these results underscore the importance of preserving meaningful consent and transparency in how financial data are used.

Several questions remain open for future research. Open banking may generate unintended consequences if it limits banks' ability to extract rents from customer data. As open banking is still a relatively new initiative, future research can empirically test the potential second-order effects of open banking via its impact on incumbents' profitability, and implications for bank-dependent borrowers. Therefore, the findings of this paper should be interpreted with caution in terms of welfare implications, which are not addressed in this study.

Additionally, this study is related to the effects of open banking in the credit market. The implications of open banking, however, may be markedly different across a wider range of financial services, which need to be taken into consideration to assess the aggregate impact.

References

Acquisti, Alessandro, Leslie K John, and George Loewenstein (2013). "What is privacy worth?" *The Journal of Legal Studies* 42.2, pp. 249–274.

Ai, Chunrong and Edward C Norton (2003). "Interaction terms in logit and probit models". *Economics letters* 80.1, pp. 123–129.

Akerlof, George A (1970). "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism". *The Quarterly Journal of Economics* 84.3, pp. 488–500.

Alok, Shashwat, Pulak Ghosh, Nirupama Kulkarni, and Manju Puri (2024). "Open banking and digital payments: Implications for credit access". National Bureau of Economic Research.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). "Mostly harmless econometrics: An empiricist's companion". Princeton university press.

Argyle, Bronson, Taylor Nadauld, and Christopher Palmer (2023). "Real effects of search frictions in consumer credit markets". *The Review of Financial Studies* 36.7, pp. 2685–2720.

Armantier, Olivier, Sebastian Doerr, Jon Frost, Andreas Fuster, and Kelly Shue (2024). "Nothing to hide? Gender and age differences in willingness to share data". *Gender and age differences in willingness to share data (April 26, 2024)*.

Athey, Susan, Christian Catalini, and Catherine Tucker (2017). "The digital privacy paradox: Small money, small costs, small talk". Tech. rep. National Bureau of Economic Research.

Avramidis, Panagiotis, Nikolaos Mylonopoulos, and George G Pennacchi (2022). "The role of marketplace lending in credit markets: Evidence from bank mergers". *Management Science* 68.4, pp. 3090–3111.

Babina, Tania, Saleem A. Bahaj, Greg Buchak, Filippo De Marco, Angus K. Foulis, Will Gornall, Francesco Mazzola, and Tong Yu (Forthcoming). "Customer Data Access and Fintech Entry: Early Evidence from Open Banking". *Journal of Financial Economics.*

Balyuk, Tetyana and Sergei Davydenko (2019). "Reintermediation in FinTech: Evidence from online lending". *Journal of Financial and Quantitative Analysis*, pp. 1–75.

Beaumont, Paul, Huan Tang, and Eric Vansteenberghe (2022). "The role of fintech in small business lending". *Available at SSRN 4260842*.

Benetton, Matteo, Peter Eckley, Nicola Garbarino, Liam Kirwin, and Georgia Latsi (2021). "Capital requirements and mortgage pricing: Evidence from Basel II". *Journal of Financial Intermediation* 48, p. 100883.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri (2020). "On the rise of fintechs: Credit scoring using digital footprints". *The Review of Financial Studies* 33.7, pp. 2845–2897.

Berger, Allen N and Gregory F Udell (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, pp. 585–628.

Bonatti, Alessandro and Gonzalo Cisternas (2020). "Consumer scores and price discrimination". *The Review of Economic Studies* 87.2, pp. 750–791.

Bond, Philip and Yao Zeng (2022). "Silence is safest: Information disclosure when the audience's preferences are uncertain". *Journal of Financial Economics* 145.1, pp. 178–193.

Braggion, Fabio, Alberto Manconi, Nicola Pavanini, Haikun Zhu, et al. (2020). "The Value of "New" and "Old" Intermediation in Online Debt Crowdfunding". Centre for Economic Policy Research.

Brunnermeier, Markus and Jonathan Payne (2024). *FinTech Lending, Banking, and Information Portability.*

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru (2018). "Fintech, regulatory arbitrage, and the rise of shadow banks". *Journal of Financial Economics* 130.3, pp. 453–483.

Carpena, Fenella and Bilal Zia (2020). "The causal mechanism of financial education: Evidence from mediation analysis". *Journal of Economic Behavior & Organization* 177, pp. 143–184.

Cornelli, Giulio, Jon Frost, Leonardo Gambacorta, and Julapa Jagtiani (2022). "The impact of fintech lending on credit access for us small businesses". FRB of Philadelphia Working Paper.

de Roure, Calebe, Loriana Pelizzon, and Anjan Thakor (2022). "P2P lenders versus banks: Cream skimming or bottom fishing?" *The Review of Corporate Finance Studies* 11.2, pp. 213–262.

Di Maggio, Marco, Dimuthu Ratnadiwakara, and Don Carmichael (2022). "Invisible primes: Fintech lending with alternative data". National Bureau of Economic Research.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer (2007). "Private credit in 129 countries". *Journal of Financial Economics* 84.2, pp. 299–329.

Eça, Afonso, Miguel A Ferreira, Melissa Porras Prado, and Antonino Emanuele Rizzo (2022). "The real effects of fintech lending on SMEs: Evidence from loan applications". ECB Working Paper.

Erel, Isil and Jack Liebersohn (2022). "Can FinTech reduce disparities in access to finance? Evidence from the Paycheck Protection Program". *Journal of Financial Economics* 146.1, pp. 90–118.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery (2019). "The role of technology in mortgage lending". *The Review of Financial Studies* 32.5, pp. 1854–1899.

Gambacorta, Leonardo, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen (2020). "Data vs collateral". CEPR Discussion Paper No. DP15262.

Ghosh, Pulak, Boris Vallée, and Yao Zeng (Forthcoming). "FinTech lending and cashless payments". *Journal of Finance*.

Goldfarb, Avi and Catherine Tucker (2012). "Shifts in privacy concerns". *American Economic Review* 102.3, pp. 349–53.

Goldstein, Itay, Chong Huang, and Liyan Yang (2022). "Open Banking under Maturity Transformation". *Available at SSRN*.

Gopal, Manasa and Philipp Schnabl (2022). "The rise of finance companies and fintech lenders in small business lending". *The Review of Financial Studies* 35.11, pp. 4859–4901.

Grossman, Sanford J (1981). "The informational role of warranties and private disclosure about product quality". *The Journal of Law and Economics* 24.3, pp. 461–483.

Harbaugh, Richmond and Theodore To (2020). "False modesty: When disclosing good news looks bad". *Journal of Mathematical Economics* 87, pp. 43–55.

He, Zhiguo, Jing Huang, and Jidong Zhou (2023). "Open banking: Credit market competition when borrowers own the data". *Journal of Financial Economics* 147.2, pp. 449–474.

Heckman, James J and Rodrigo Pinto (2013). "Causal analysis after Haavelmo". *Journal of Econometrics* 175.2, pp. 4–15.

Heckman, James J and Edward Vytlacil (2005). "Structural equations, treatment effects, and econometric policy evaluation 1". *Econometrica* 73.3, pp. 669–738.

Hertzberg, Andrew, Jose Maria Liberti, and Daniel Paravisini (2011). "Public information and coordination: evidence from a credit registry expansion". *The Journal of Finance* 66.2, pp. 379–412.

Hicks, Raymond and Dustin Tingley (2011). “Causal mediation analysis”. *The Stata Journal* 11.4, pp. 605–619.

Howell, Sabrina T, Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong (2021). “Racial disparities in access to small business credit: Evidence from the paycheck protection program”. National Bureau of Economic Research.

Huber, Martin (2014). “Identifying causal mechanisms (primarily) based on inverse probability weighting”. *Journal of Applied Econometrics* 29.6, pp. 920–943.

Ichihashi, Shota (2020). “Online privacy and information disclosure by consumers”. *American Economic Review* 110.2, pp. 569–595.

Imbens, Guido W and Jeffrey M Wooldridge (2009). “Recent developments in the econometrics of program evaluation”. *Journal of Economic Literature* 47.1, pp. 5–86.

Jagtiani, Julapa and Catharine Lemieux (2019). “The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform”. *Financial Management* 48.4, pp. 1009–1029.

Jansen, Mark, Hieu Quang Nguyen, and Amin Shams (2025). “Rise of the machines: The impact of automated underwriting”. *Management Science* 71.2, pp. 955–975.

Jovanovic, Boyan (1982). “Truthful disclosure of information”. *The Bell Journal of Economics*, pp. 36–44.

Kashyap, Anil K, Jeremy C Stein, et al. (2004). “Cyclical implications of the Basel II capital standards”. *Economic Perspectives-Federal Reserve Bank Of Chicago* 28.1, pp. 18–33.

Lin, Tesary (2022). “Valuing intrinsic and instrumental preferences for privacy”. *Marketing Science* 41.4, pp. 663–681.

Liu, Zhuang, Michael Sockin, and Wei Xiong (2023). “Data privacy and algorithmic inequality”. National Bureau of Economic Research.

Milgrom, Paul R (1981). “Good news and bad news: Representation theorems and applications”. *The Bell Journal of Economics*, pp. 380–391.

Ouyang, Shumiao (2021). “Cashless Payment and Financial Inclusion”. *Available at SSRN 3948925.*

Parlour, Christine A, Uday Rajan, and Haoxiang Zhu (2022). “When fintech competes for payment flows”. *The Review of Financial Studies* 35.11, pp. 4985–5024.

Philippon, Thomas (2019). “On fintech and financial inclusion”. National Bureau of Economic Research.

Popov, Alexander and Gregory F Udell (2012). “Cross-border banking, credit access, and the financial crisis”. *Journal of International Economics* 87.1, pp. 147–161.

Posner, Richard A (1981). “The Economics of Privacy”. *The American Economic Review* 71.2, pp. 405–409.

Quigley, Daniel and Ansgar Walther (2024). “Inside and outside information”. *The Journal of Finance* 79.4, pp. 2667–2714.

Rishabh, Kumar (2024). “Beyond the bureau: Interoperable payment data for loan screening and monitoring”. *Available at SSRN 4782597*.

Roulet, Caroline (2018). “Basel III: Effects of capital and liquidity regulations on European bank lending”. *Journal of Economics and Business* 95, pp. 26–46.

Sheng, Tianxiang (2021). “The effect of fintech on banks’ credit provision to SMEs: Evidence from China”. *Finance Research Letters* 39, p. 101558.

Stigler, George J (1980). “An introduction to privacy in economics and politics”. *The Journal of Legal Studies* 9.4, pp. 623–644.

Tang, Huan (2019a). “Peer-to-peer lenders versus banks: substitutes or complements?” *The Review of Financial Studies* 32.5, pp. 1900–1938.

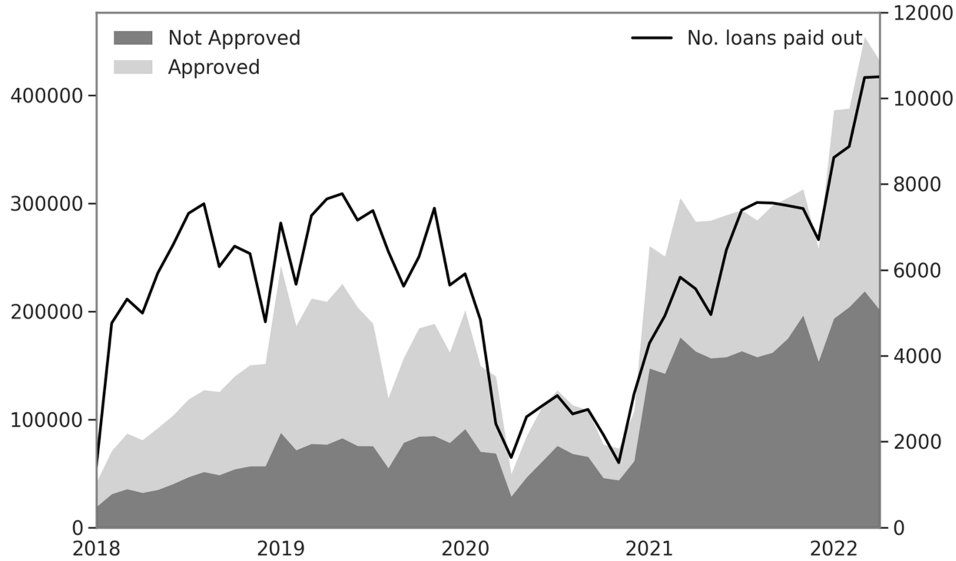
— (2019b). “The value of privacy: Evidence from online borrowers”. *HEC Paris*.

Vallee, Boris and Yao Zeng (2019). “Marketplace lending: A new banking paradigm?” *The Review of Financial Studies* 32.5, pp. 1939–1982.

Viscusi, W Kip (1978). “A note on” lemons” markets with quality certification”. *The Bell Journal of Economics*, pp. 277–279.

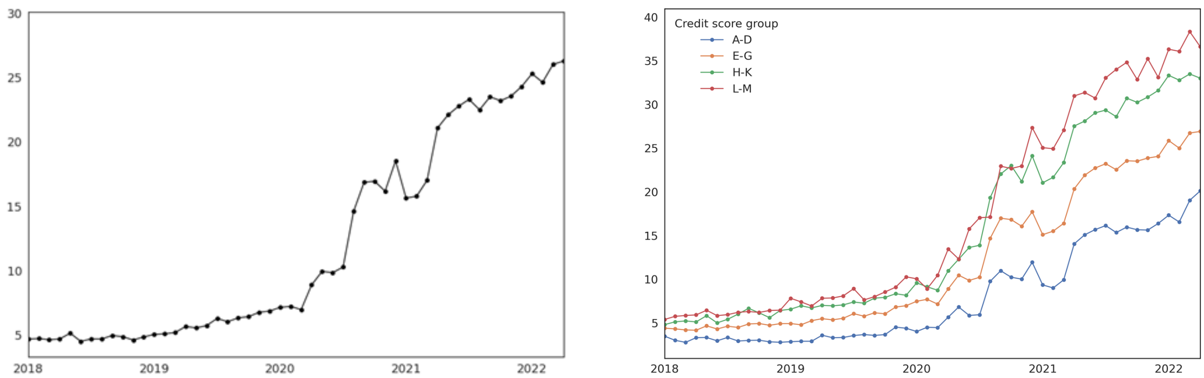
Warren, Samuel and Louis Brandeis (1989). “The right to privacy”. In: *Killing the Messenger: 100 Years of Media Criticism*. Columbia University Press, pp. 1–21.

Figure 1: Number of applications and disbursed loans, monthly



Notes: The figure depicts the monthly count of loan applications, differentiated by approval status. The dark gray bars represent the number of non-approved applications, while the light gray bars indicate approved applications (both plotted on the first y-axis). The second y-axis displays the count of disbursed loans among the approved applications. The sample period is from January 13, 2018 to May 22, 2022.

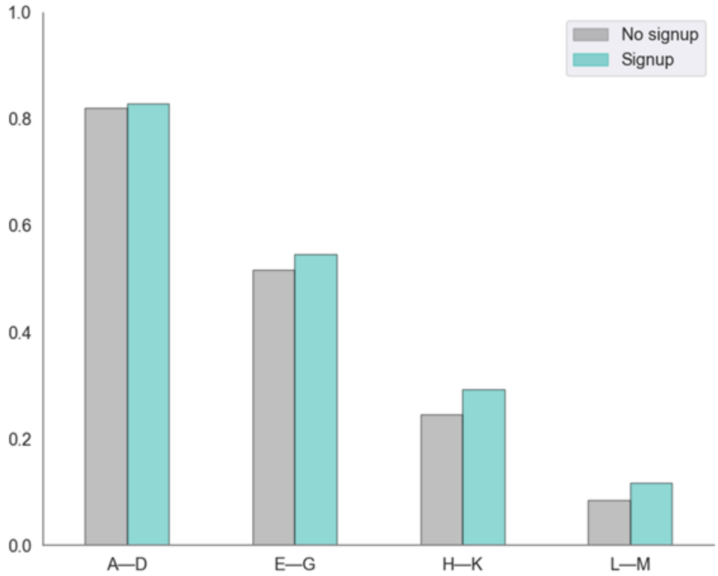
Figure 2: Data sharing over time (overall vs. by credit score), monthly



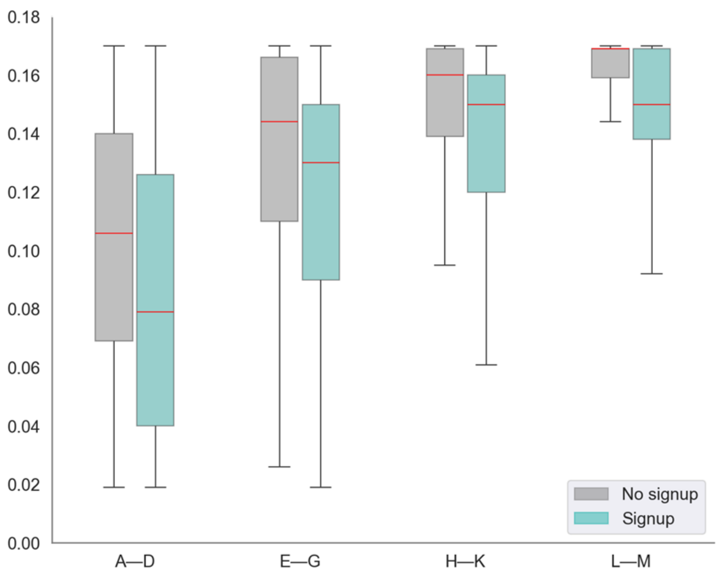
Notes: The left panel shows the percentage of loan applications in which data was shared. The right panel shows these percentages by credit score (A–D: highest, L–M: lowest). The sample period is from January 13, 2018 to May 22, 2022.

Figure 3

A. Loan approval rate by data sharing decision

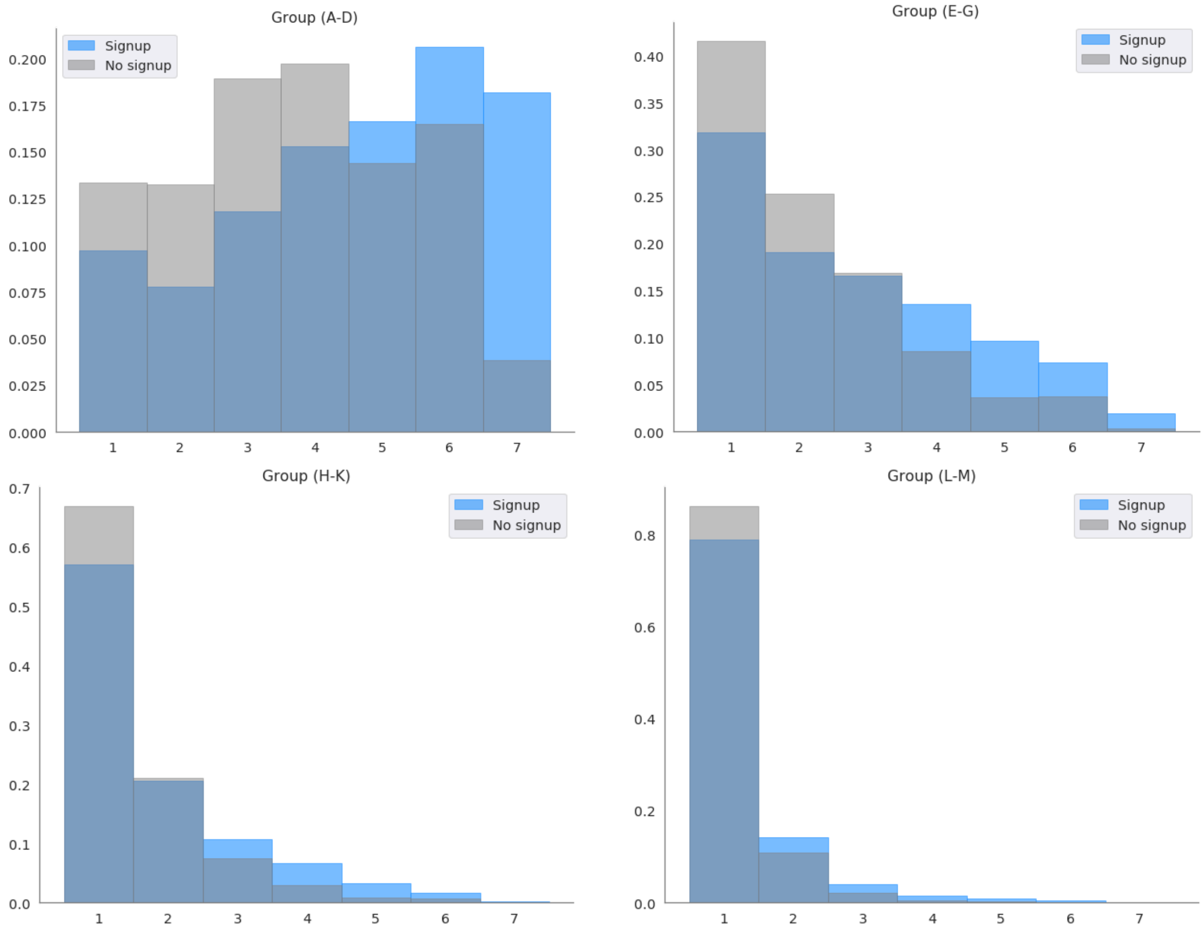


B. Loan interest rate by data sharing decision



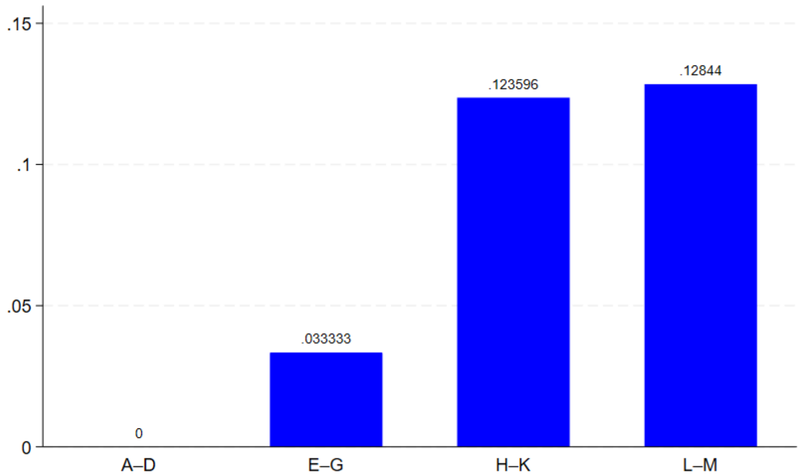
Notes: Panel A displays the unconditional mean of loan approval rate, and panel B shows the inter-quartile range of interest rates. Green bars represent data sharing applicants, and gray bars represent non-sharing applicants across credit score groups from A–D (highest) to L–M (lowest). *Signup* is a dummy variable that takes a value of one if the applicant shared data.

Figure 4: Distribution of platform's proprietary risk scores (Auxscore) by data sharing decision (matched sample)



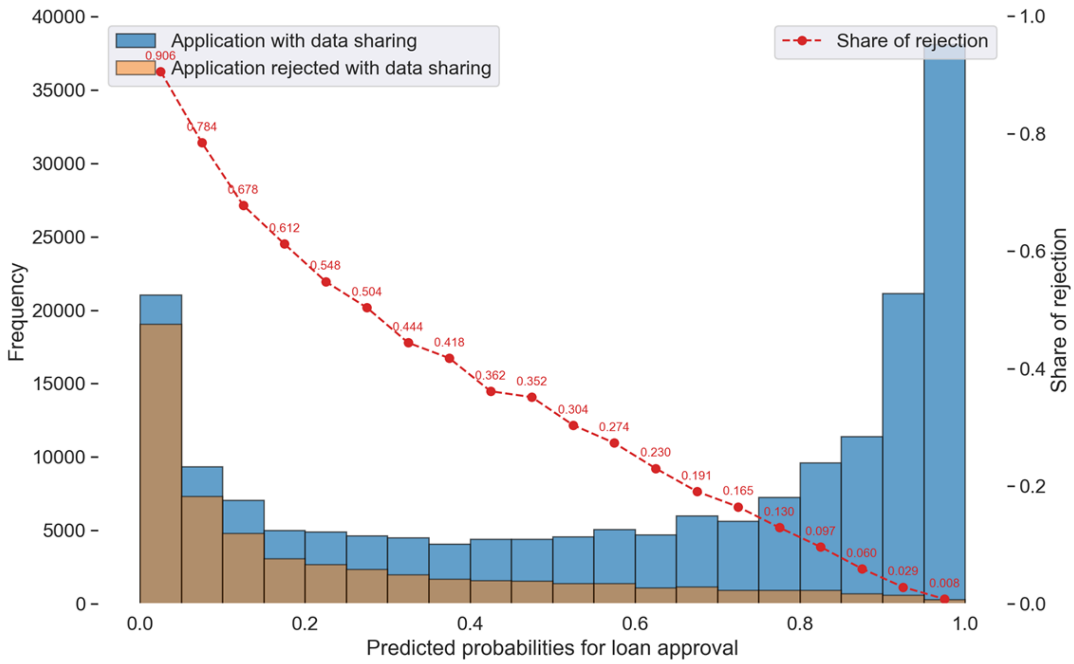
Notes: This figure displays the distribution of the platform's proprietary risk score, *Auxscore*, by credit score (A–D: highest, L–M: lowest) and data sharing status using the matched sample. The x-axis indicates *Auxscore* ranging from 7 (highest) to 1 (rejection) and the y-axis shows the share of applicants. Applicants decide whether to share their bank transaction data prior to receiving *Auxscore*. *Signup* is a dummy variable that takes a value of one if the applicant shared data. Applicants who shared data are matched one-to-one to non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year*.

Figure 5



Notes: This figure shows the share of applicants whose proprietary risk score, *Auxscore*, is lower than that of their matched non-sharing counterpart, plotted by credit score group (A–D = highest, to L–M = lowest). Applicants who shared data are matched one-to-one to non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year*.

Figure 6

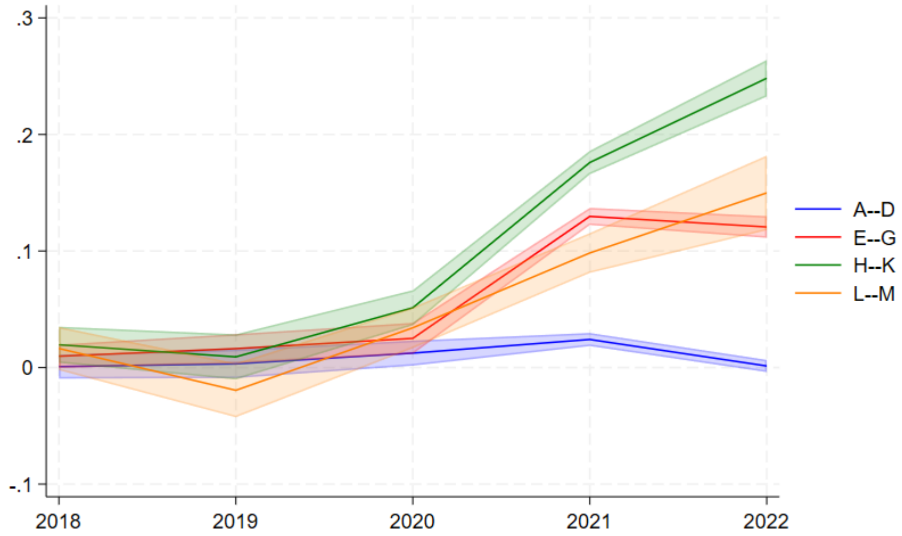


Notes: This figure shows how many applicants who shared data were rejected, broken down by their predicted probability of loan approval. The x-axis displays predicted approval probabilities (binned in intervals of width 0.05). The left y-axis shows the total number of sharers in each bin (blue bars) and, within that, how many were rejected (brown bars). The right y-axis (red line) reports the share of rejected applicants among sharers in each bin.

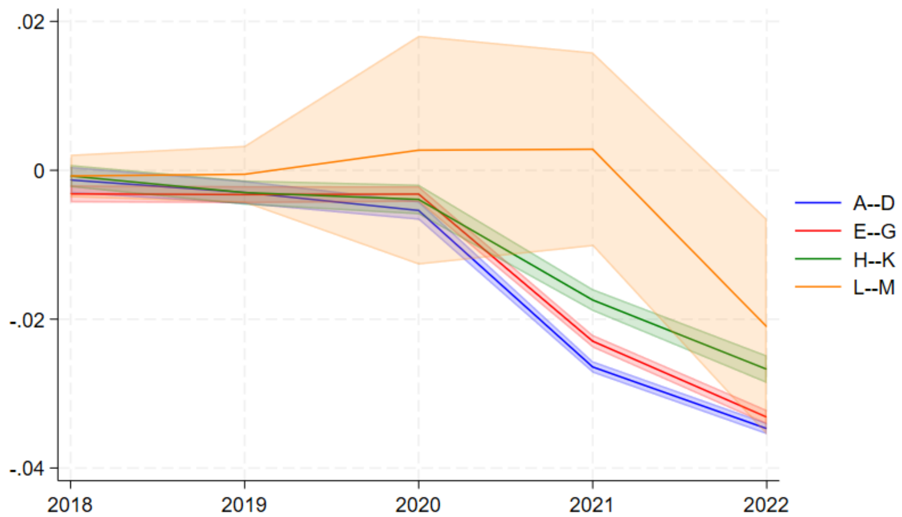
These predicted probabilities are generated from a random forest classifier trained on 80% of applications from non-sharers and evaluated on a 20% hold-out set. The model provides a counterfactual benchmark: what the predicted probability of loan approval would have been for sharers had they not disclosed their data. The model achieves 93.6% classification accuracy and an area under the ROC curve (AUC) of 0.98. Inputs include age, credit score, income, expenses, loan amount requested, homeownership, gender, number of outstanding and past loans, married (D), main earner (D), access channel, and application month-year. In the case of multiple applications per applicant, only the first is included.

Figure 7

A. Data sharing on loan approvals



B. Data sharing on interest rate



Notes: Panel A (Panel B) displays the effect of data sharing on loan approvals (interest rates) from Equation (2) (Equation (3)) over time by credit score group (A–D (highest) to L–M (lowest)).

Table 1: **Summary Statistics**

Variable	N	Mean	S.D.	Min	p25	p50	p75	Max
LOAN INFORMATION								
Loan amount requested	2,309,359	13,876.41	13,068.91	1,000.00	4,000.00	10,000.00	20,000.00	50,000.00
Interest rate*	1,559,902	0.12	0.04	0.00	0.08	0.13	0.15	0.20
Platform score (max 7, min 1)	2,309,359	2.89	1.82	1.00	1.00	2.00	4.00	7.00
Credit score group (max 4, min 1)	2,309,359	3.13	0.87	0.00	3.00	3.00	4.00	4.00
Loan duration	2,309,359	55.51	23.94	0.00	36.00	60.00	84.00	84.00
Application accepted (D)	2,309,359	0.68	0.47	0.00	0.00	1.00	1.00	1.00
Bank account detail shared (D)	2,309,359	0.08	0.27	0.00	0.00	0.00	0.00	1.00
BORROWER CHARACTERISTICS								
Age	2,309,359	38.19	12.51	18.00	28.00	36.00	48.00	69.00
Female (D)	2,309,359	0.35	0.48	0.00	0.00	0.00	1.00	1.00
Married (D)	2,309,359	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Main earner (D)	2,309,359	0.62	0.48	0.00	0.00	1.00	1.00	1.00
No. current loan demand	2,309,359	1.35	1.47	0.00	0.00	1.00	2.00	68.00
No. past loan demand	2,309,359	1.04	1.78	0.00	0.00	0.00	1.00	76.00
INCOME AND EXPENSES								
Total income	2,309,359	2,355.13	1,959.79	0.00	1,500.00	1,998.00	2,650.00	30,388.00
Monthly net salary income	2,309,359	2,085.19	1,658.26	0.00	1,300.00	1,800.00	2,400.00	26,000.00
Child support income	2,309,359	120.30	208.11	0.00	0.00	0.00	204.00	1,513.00
Other income	2,309,359	132.90	489.93	0.00	0.00	0.00	0.00	6,666.70
Total expenses	2,309,359	717.38	613.96	0.00	330.00	600.00	954.00	5,147.00
Housing-related expenses	2,309,359	457.05	383.06	0.00	200.00	425.00	650.00	3,000.00
Credit installments expenses	2,309,359	171.77	330.94	0.00	0.00	0.00	240.00	3,086.00
Other expenses	2,309,359	21.14	114.23	0.00	0.00	0.00	0.00	1,500.00
Insurance expenses	2,309,359	49.87	152.19	0.00	0.00	0.00	0.00	1,420.00
Child support expenses	2,309,359	18.25	98.85	0.00	0.00	0.00	0.00	1,200.00
ASSETS								
Credit card holder (D)	2,309,359	0.64	0.48	0.00	0.00	1.00	1.00	1.00
Checking account owner (D)	2,309,359	0.94	0.24	0.00	1.00	1.00	1.00	1.00
Home owner (D)	2,309,359	0.25	0.44	0.00	0.00	0.00	1.00	1.00
Car owner (D)	2,309,359	0.57	0.50	0.00	0.00	1.00	1.00	1.00

Notes: This table presents summary statistics for the sample. The sample period runs from January 13, 2018, to May 22, 2022. (D) = dummy variable. The monetary unit is EUR. The final sample includes only one application per borrower. In the case of multiple applications, the initial application from each applicant is included. *conditional on loan approval.

Table 2: **Factors influencing Willingness to Share Data**

	Probit (average marginal effects)				Linear Probability Model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age (10 years)		−0.020***	−0.018***	−0.019***		−0.019***	−0.017***	−0.018***
		(0.0002)	(0.0002)	(0.0002)		(0.0002)	(0.0003)	(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***		0.014***	0.010***	0.027***		0.014***	0.009***
	(0.0004)		(0.0004)	(0.0004)	(0.0005)		(0.0004)	(0.0004)
Credit score (H–K)	0.039***		0.019***	0.019***	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***	0.022***	0.035***		0.010***	0.014***
	(0.0011)		(0.0010)	(0.0010)	(0.0009)		(0.0009)	(0.0009)
Loan amount requested (ln)				−0.010***				−0.011***
				(0.0002)				(0.0003)
Loan duration (ln)				−0.003***				−0.006***
				(0.0004)				(0.0005)
Female				−0.004***				−0.005***
				(0.0004)				(0.0004)
Married				0.000				−0.001*
				(0.0004)				(0.0004)
Main earner				0.009***				0.011***
				(0.0009)				(0.0010)
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***
				(0.0005)				(0.0004)
Access channel = Homepage (base)								
Access channel = Repeat				0.115***				0.071***
				(0.0031)				(0.0026)
Access channel = Price comparison website				−0.076***				−0.066***
				(0.0014)				(0.0014)
Access channel = Broker				−0.107***				−0.098***
				(0.0017)				(0.0019)
Access channel = Bank				−0.126***				−0.133***
				(0.0018)				(0.0025)
Constant					0.019***	0.116***	0.094***	0.256***
					(0.0013)	(0.0016)	(0.0017)	(0.0034)
Dummy	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year
Cluster (region-month-year)	X	X	X	X	X	X	X	X
N	2,309,359	2,309,359	2,309,359	2,309,359	2,309,359	2,309,359	2,309,359	2,309,359
R2	0.0706	0.0793	0.0808	0.1124	0.041	0.045	0.046	0.063

Notes: This table reports the results from Equation (1), which estimates the probability that an applicant shares bank transaction data using a probit model. The coefficients (1–3) report average marginal effects. Clustered standard errors are in parentheses. Columns (1)–(3) report pseudo R2 and (4)-(6) adjusted R2. The dependent variable is *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise. *Credit score* range from A–D (highest) to L—M (lowest). In the case of multiple applications, the initial application from each applicant is included.

Table 3: **The effect of data sharing on loan approvals**

	Probit (average marginal effects)			Linear Probability Model		
	(1)	(2)	(3)	(4)	(5)	(6)
Signup	0.020***	0.020***	0.015***	0.020***	0.022***	0.035***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Signup × Credit score (A–D)* (Base)						
Signup × Credit score (E–G)	0.088***	0.088***	0.080***	0.089***	0.087***	0.074***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Signup × Credit score (H–K)	0.131***	0.128***	0.113**	0.129***	0.124***	0.101***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Signup × Credit score (L–M)	0.038***	0.032***	0.025***	0.032***	0.025***	−0.001
	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
Credit score (A–D) (Base)						
Credit score (E–G)	−0.240***	−0.191***	−0.178***	−0.295***	−0.234***	−0.219***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Credit score (H–K)	−0.551***	−0.470***	−0.417***	−0.619***	−0.533***	−0.483***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.003)
Credit score (L–M)	−0.794***	−0.722***	−0.638***	−0.835***	−0.737***	−0.637***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)	(0.005)
Age		0.007***	0.005***		0.006***	0.005***
		(0.000)	(0.000)		(0.000)	(0.000)
Income decile		0.022***	0.014***		0.019***	0.012***
		(0.000)	(0.000)		(0.000)	(0.000)
Homeowner			0.043***			0.068***
			(0.002)			(0.002)
Signup × Homeowner			−0.090***			−0.072***
			(0.002)			(0.003)
Loan amount requested (ln)			0.010***			0.014***
			(0.001)			(0.001)
Loan duration (ln)			−0.093***			−0.096***
			(0.002)			(0.002)
Female			0.030***			0.035***
			(0.001)			(0.001)
Married			0.034***			0.035***
			(0.002)			(0.002)
Main earner			0.027***			0.020***
			(0.001)			(0.002)
Carowner			0.060***			0.072***
			(0.001)			(0.002)
No. current loan demand			0.017***			0.020***
			(0.001)			(0.001)
No. past loan demand			0.005***			0.007***
			(0.000)			(0.000)
Access channel = Homepage (Base)						
Access channel = Repeat			−0.002			−0.118***
			(0.002)			(0.004)
Access channel = Price comp. website			−0.325***			−0.280***
			(0.001)			(0.002)
Access channel = Broker			−0.530***			−0.491***
			(0.004)			(0.004)
Access channel = Bank			−0.460***			−0.425***
			(0.009)			(0.007)
Dummy	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year	Month–Year
Cluster (region-month-year)	X	X	X	X	X	X
N	329,938	329,938	329,938	329,938	329,938	329,938
R2	0.215	0.255	0.347	0.251	0.289	0.356

Notes: This table reports the results from Equation (2) which estimates the effect of data sharing, *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise, on the probability of loan approval using the matched sample. *Credit score* range from A–D (highest) to L—M (lowest). Applicants who shared data are matched one-to-one to non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year.* The final sample includes 329,938 loan applications from 329,938 unique applicants. In the case of multiple applications per applicant, only the first is included.
*The coefficients in columns (1)–(3) show average marginal effects. Clustered standard errors are in parentheses. Columns (1)–(3) report pseudo R2 and columns (4)–(6) report adjusted R2. In the probit model, marginal effects for the interaction are computed as the derivative of the predicted probability with respect to *Signup* at each level of *Credit score* and *Homeowner.* These effects are estimated using the delta method—averaged over the observed covariate values (Ai and Norton 2003).

Table 4: **The effect of data sharing on interest rates**

	Matched sample		
	(1)	(2)	(3)
Signup	−0.0211***	−0.0211***	−0.0212***
	(0.0003)	(0.0003)	(0.0003)
Signup × Credit score (A–D) (Base)			
Signup × Credit score (E–G)	0.0033***	0.0032***	0.0028***
	(0.0004)	(0.0004)	(0.0003)
Signup × Credit score (H–K)	0.0114***	0.0114***	0.0107***
	(0.0005)	(0.0005)	(0.0005)
Signup × Credit score (L–M)	0.0180***	0.0190***	0.0197***
	(0.0012)	(0.0012)	(0.0012)
Credit score (A–D) (Base)			
Credit score (E–G)	0.0292***	0.0227***	0.0211***
	(0.0003)	(0.0003)	(0.0002)
Credit score (H–K)	0.0413***	0.0335***	0.0319***
	(0.0003)	(0.0003)	(0.0003)
Credit score (L–M)	0.0516***	0.0404***	0.0383***
	(0.0008)	(0.0008)	(0.0009)
Age		−0.0009***	−0.0008***
		(0.0000)	(0.0000)
Income decile		−0.0019***	−0.0020***
		(0.0000)	(0.0000)
Homeowner			−0.0156***
			(0.0003)
Signup × Home owner			0.0054***
			(0.0004)
Loan amount requested (ln)			0.0101***
			(0.0002)
Loan duration (ln)			0.0050***
			(0.0002)
Married			−0.0059***
			(0.0002)
Female			−0.0026***
			(0.0002)
Main earner			−0.0043***
			(0.0002)
Car owner			−0.0049***
			(0.0002)
No. current loan demand			−0.0009***
			(0.0001)
No. past loan demand			−0.0002***
			(0.0000)
Access channel = Homepage			
Access channel = Repeat			−0.0299***
			(0.0006)
Access channel = Price comp. website			0.0049***
			(0.0003)
Access channel = Broker			0.0182***
			(0.0005)
Access channel = Bank			0.0182***
			(0.0012)
Dummy	Month–Year	Month–Year	Month–Year
Cluster (region-month-year)	X	X	X
N	242,360	242,360	242,360
Adjusted R2	0.2811	0.3431	0.4765

Notes: This table reports the results of Equation (3), which estimates the effect of data sharing *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise, on interest rates using the matched sample. *Credit score* range from A–D (highest) to L—M (lowest). Interest rates are revealed only for approved applications. Thus, *approved* applicants who shared data are matched one-to-one with *approved* non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year*. The final sample includes 242,360 loan applications from 242,360 unique applicants. In the case of multiple applications per applicant, only the first is included.

Table 5: **Changes in *Auxscore* after data sharing**

	Matched sample
Signup	0.8611***
	(0.0003)
Signup × Credit score (A–D) (Base)	
Signup × Credit score (E–G)	−0.1461***
	(0.0108)
Signup × Credit score (H–K)	−0.4270***
	(0.0151)
Signup × Credit score (L–M)	−0.7932***
	(0.0389)
Controls	X
Dummy	Month–Year
Cluster (region-month-year)	X
N	242,360
Adjusted R2	0.4765

Notes: This table reports the estimates of the effect of data sharing *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise, on *Auxscore*, the platform's proprietary risk score which reflects the lender's updated risk assessment after observing transaction data (7 represents the lowest risk, and 1 is rejection). *Credit score* range from A–D (highest) to L—M (lowest). The purpose of this exercise is to understand the mechanism through which data sharing affects loan pricing. Since interest rates are revealed only for approved applications (therefore, an *Auxscore* of 1 is not included). *Approved* applicants who shared data are matched one-to-one with *approved* non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year.* The final sample includes 242,360 loan applications from 242,360 unique applicants. In the case of multiple applications per applicant, only the first is included.

Table 6: **Robustness checks using fixed effects**

A. Data sharing decision on loan approvals

	Credit score			
	(A–D)	(E–G)	(H–K)	(L–M)
Signup	0.033***	0.089***	0.091***	0.031***
	(0.012)	(0.008)	(0.008)	(0.011)
Controls	X	X	X	X
Individual-day FE	X	X	X	X
Robust cluster error	X	X	X	X
N	4,622	14,160	10,315	2,421
Adjusted R2	0.065	0.104	0.142	0.222

Notes: 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 initial applications but do share it in others. Given that there is within-individual variation in data sharing decisions while applicant characteristics do not change, I employ individual-day fixed effects to test the effect of data sharing on loan approvals as a robustness check. *Signup* is a dummy variable that takes a value of one if the loan application is approved and 0 otherwise. Control variables are the same as in Equation (2). Each column represents a credit score group, with (A–D) the highest and (L–M) the lowest.

B. Data sharing decision on interest rates

	Credit score			
	(A–D)	(E–G)	(H–K)	(L–M)
Signup	−0.016***	−0.014***	−0.008***	−0.006**
	(0.001)	(0.001)	(0.001)	(0.003)
Controls	X	X	X	X
Individual-day FE	X	X	X	X
Robust cluster error	X	X	X	X
N	3,472	5,362	1,566	135
Adjusted R2	0.251	0.190	0.119	0.308

Notes: This table shows the effect of data sharing on interest rates. The sample includes multiple applications filed by the same individuals on the same day. These individuals do not share their data in initial applications but do share it in others. Given that there is within-individual variation in data sharing decisions while applicant characteristics do not change, I employ individual-day fixed effects to test the effect of data sharing on loan approvals as a robustness check. Control variables include loan amount and loan duration. *Signup* is a dummy variable that takes a value of one if the loan application is approved and 0 otherwise. Control variables are the same as in Equation (3). Each column represents a credit score group, with (A–D) the highest and (L–M) the lowest.

Table 7: **Voluntary disclosure and unobserved borrower risk (ex post defaults)**

	Default = 1 if payment is more than 90 days late			
Credit score group	(A–D)	(E–G)	(H–K)	(L–M)
Signup	−0.0104***	−0.0180***	−0.0251**	−0.0642
	(0.004)	(0.004)	(0.0135)	(0.0472)
Controls	Y	Y	Y	Y
Cluster (region-month-year)	Y	Y	Y	Y
N	11,574	11,096	2,479	132
Pseudo R2	0.0421	0.0394	0.0404	0.1263

Notes This table presents the association between voluntary data sharing (*Signup*, a dummy equal to one if the applicant consented to share bank transaction data) and loan default (*Default*, a dummy equal to one if the loan is more than 90 days delinquent). This assesses whether applicants who choose to share data differ in latent borrower quality from observably identical non-sharers. Loan performance data are sourced from the European DataWarehouse (EDW) and are matched to Auxmoney applications using key borrower and loan characteristics (e.g., income, location, loan amount, duration, interest rate, disbursement date, occupation type, loan purpose). The matched sample includes 25,281 loans. Control variables include demographic and loan characteristics: age, income decile, loan amount, loan duration, gender, marital status, main earner, homeownership, car ownership, number of current and past loan inquiries, access channel (categorical), and quarter-year dummies. Standard errors are clustered by region-month-year.

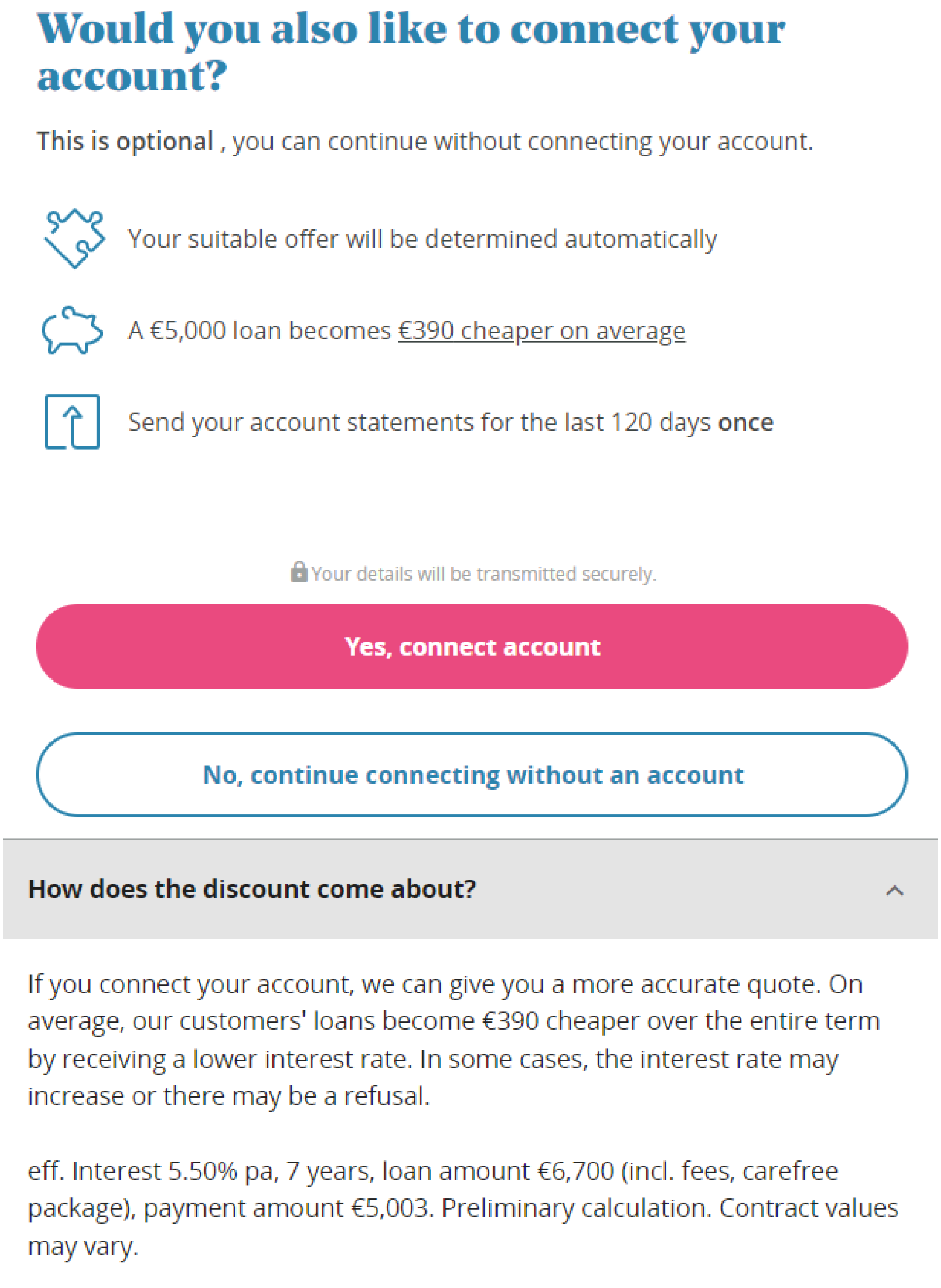
Table 8: **Equilibrium inferences on non-disclosure**

	(1) Probit	(2) LPM	(3) Pool FEs
Data sharing rate	−0.218***	−0.217***	−0.176***
	(0.006)	(0.005)	(0.002)
Controls	X	X	X
Dummy	Month–Year	Month–Year	Month–Year
Cluster (region-month-year)	X	X	X
N	8,159,405	8,159,405	7,905,450
Adjusted R2	0.367	0.398	0.427

Notes: This table shows results from regressions estimating how the probability of loan approval for non-disclosing applicants varies with the share of disclosers in their reference pool. The key independent variable, *Data sharing rate*, measures the share of applicants who share data within each reference pool. Pools are defined by categorical bins for debt-to-income ratio (low: ≤35%, medium: 35-50%, high: >50%), age (under 30, 30-50, 50–65, over 65), income (terciles), and loan size (small: ≤ EUR 5,000; medium: EUR 5,001–20,000; large: >EUR 20,000), along with application month, credit score group, access channel, homeownership status, and gender. The sample is restricted to applicants who did not share data. Column (1) reports the average marginal effect from a probit model. Column (2) estimates a linear probability model with the same controls. Column (3) adds pool fixed effects based on the defined reference pools to absorb unobserved heterogeneity. Robust standard errors are clustered at the region-month-year level.

A Additional Figures and Tables

Figure A.1: Data sharing during the application



Notes: This figure shows the exact manner in which data are shared during loan applications. Loan applicants are also supplied with information regarding data usage, and data security.

Figure A.1: Data sharing during the application (continued)

How does the discount come about?

How will my loan get cheaper?

How are my bank statements transmitted?

Is the transmission of my bank statements secure?

The connection is encrypted, we have no access to your access data and your account. You finally confirm the transmission via 2-factor authentication.

What happens if I don't connect my account?

Why am I being asked for bank statements?

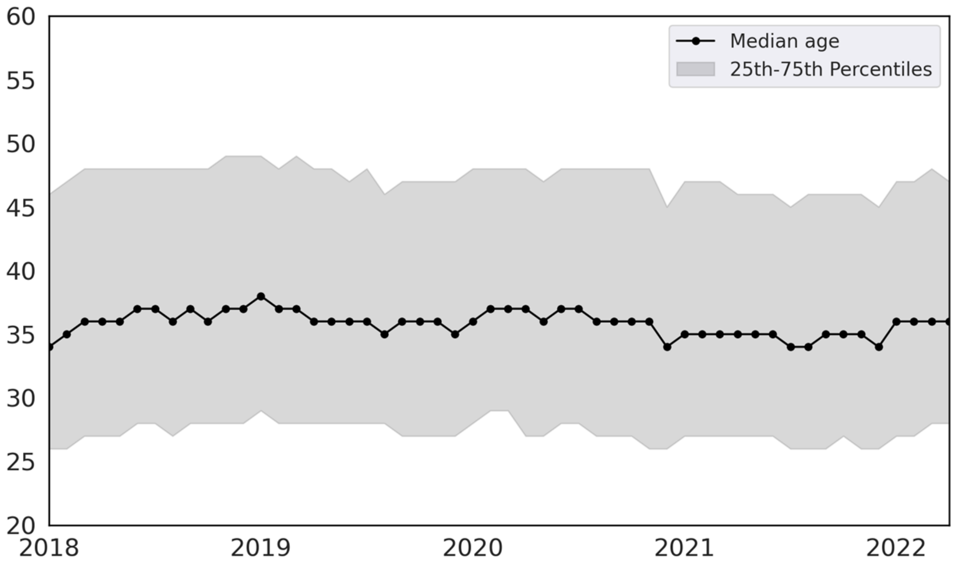
Income and expenses are automatically recognized and analyzed based on account statements. This allows the credit default risk to be assessed more precisely and a suitable offer to be made available. Your offer can improve or deteriorate as a result, and this may also lead to your loan request being rejected.

What do I do if I can't find my bank?

What do I do if I don't have online banking?

Notes: This figure shows the exact manner in which data are shared during loan applications. Loan applicants are also supplied with information regarding data usage, and data security.

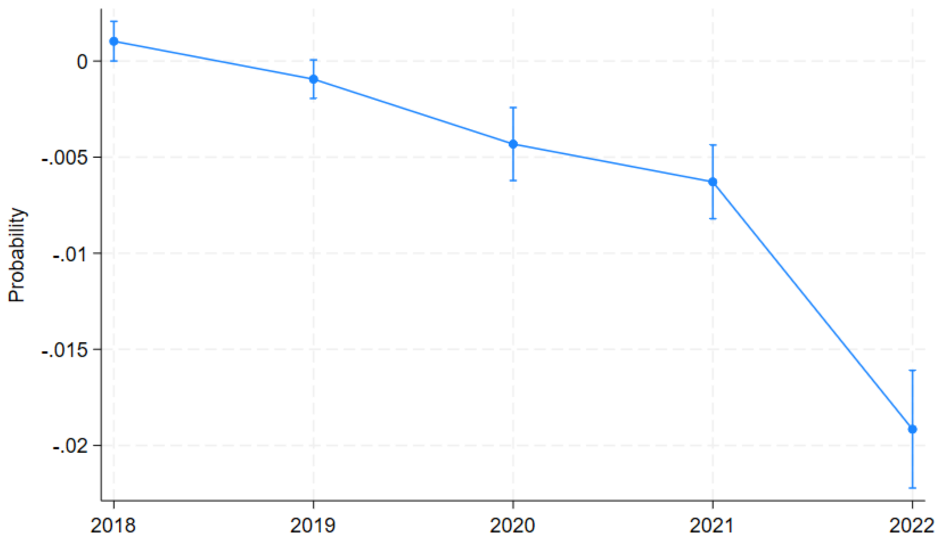
Figure A.2: Age of applicants, monthly



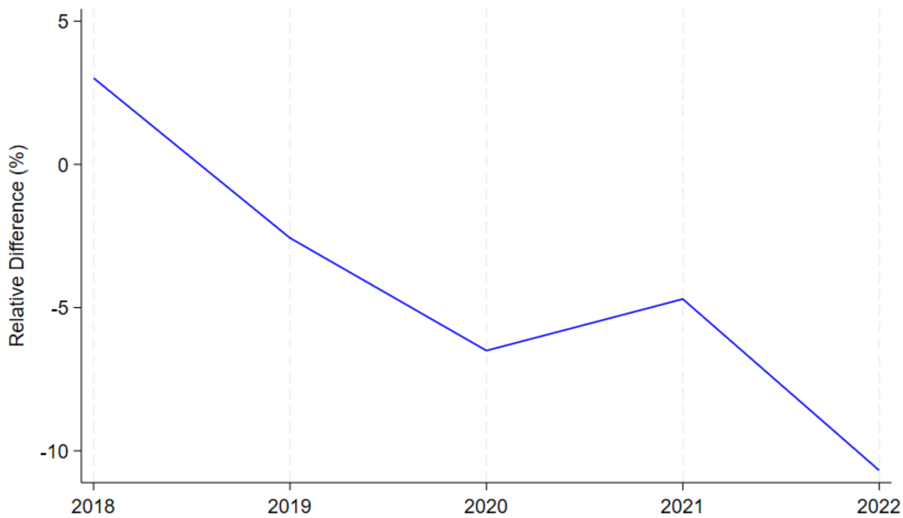
Notes: This figure shows the inter-quartile range of applicant age. The sample period is from January 13, 2018, to May 22, 2022.

Figure A.3

A. Differences in sharing probability by gender



B.Relative differences (%)



Notes: Panel A displays the difference in willingness to share data between female and male applicants over time. The estimates come from a probit model from Equation (1) with an interaction between gender and year, allowing the gender gap to vary across time. Marginal effects measure the difference in the probability of sharing data between female and male applicants in each year, where negative values indicate that female applicants are less willing to share data than male applicants. Panel B shows the relative differences in percent.

Table A.1: **Descriptive statistics by data sharing decision**

Variable	Signup	No signup	p-value
Credit requested	12,841.07	13,963.36	0.00
Interest rate*	0.10	0.12	0.00
Platform score (max 7, min 1)	3.22	2.87	0.00
Credit score group (max 4, min 1)	3.06	3.13	0.00
Loan duration	52.90	55.73	0.00
Application accepted (D)	0.70	0.67	0.00
Bank account detail shared (D)	1.00	0.00	0.00
Age	34.19	38.53	0.00
Female (D)	0.33	0.35	0.00
Main earner (D)	0.63	0.62	0.00
No. current loan demand	1.56	1.33	0.00
No. past loan demand	1.28	1.02	0.00
Total income	2,235.39	2,365.18	0.00
Total expenses	744.41	715.11	0.00
Credit card holder (D)	0.78	0.63	0.00
Checking account holder (D)	0.97	0.94	0.00
Homeowner (D)	0.20	0.26	0.00
Car owner (D)	0.61	0.56	0.00

Notes: This table presents summary statistics separately by data sharing choice, *Signup*, and for those who opt out, *No signup*. (D) = dummy variable. The monetary unit in EUR. The sample period runs from January 13, 2018, to May 22, 2022. The final sample includes only one application per applicant. In the case of multiple applications, the initial application from each applicant is included. *conditional on loan approval.

Table A.2: **Matched variables and matching results**

A. Sample to estimate the effect of data sharing on loan approvals

	Mean Sharers	Mean Non-Sharers	Mean p-value difference
Age	34.4	34.373	0.486
Loan amount requested	13,298	13,355	0.213
Income decile	5.535	5.538	0.805
Credit score group		——— exact matching ———	
Homeowner (Dummy)		——— exact matching ———	
Female (Dummy)		——— exact matching ———	
Access channel		——— exact matching ———	
Application month-year		——— exact matching ———	

Notes: This table shows t-tests for the null hypothesis of equal means for both data sharing and non-sharing applicants. This sample is used to compute the effect of data sharing on the probability of loan approval (Equation (2)). Applicants who shared data are matched one-to-one to non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year.* The final sample includes 329,938 loan applications from 329,938 unique applicants. In the case of multiple applications per applicant, only the first is included.

B. Sample to estimate the effect of data sharing on interest rates

	Mean Sharers	Mean Non-Sharers	Mean p-value difference
Age	36.558	36.541	0.694
Loan amount requested	13,456	13,487	0.566
Income decile	5.898	5.937	0.210
Credit score		——— exact matching ———	
Homeowner (Dummy)		——— exact matching ———	
Female (Dummy)		——— exact matching ———	
Access channel		——— exact matching ———	
Application month-year		——— exact matching ———	

Notes: This table shows t-tests for the null hypothesis of equal means for both data sharing and non-sharing groups. This sample is used to compute the effect of data sharing on interest rates (Equation (3)). Interest rates are revealed only for successful loan applications. Thus, *approved* applicants who shared data are matched one-to-one with *approved* non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year.* The final sample includes 242,360 loan applications from 242,360 unique applicants. In the case of multiple applications per applicant, only the first is included.

Table A.3: **The effect of data sharing on loan approvals (subsample, 2021–2022)**

	Matched sample
Signup	0.0235***
	(0.0020)
Signup × Credit score (A–D) (Base)	
Signup × Credit score (E–G)	0.1305***
	(0.0039)
Signup × Credit score (H–K)	0.2054***
	(0.0055)
Signup × Credit score (L–M)	0.1061***
	(0.0096)
Homeowner	0.0311***
	(0.0028)
Signup × Homeowner	−0.1344***
	(0.0037)
Controls	X
Dummy	Month–Year
Cluster (region-month-year)	X
N	209,290
Pseudo R2	0.3285

Notes: This table reports the results using the subsample (2021–2022) from Equation (2) which estimates the effect of data sharing, *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise, on the probability of loan approval using the matched sample. *Credit score* range from A–D (highest) to L—M (lowest). Applicants who shared data are matched one-to-one to non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score, Homeowner, Female, Access channel and Loan application month-year*. The final sample includes 209,290 loan applications from 209,290 unique applicants. In the case of multiple applications per applicant, only the first is included. The marginal effects for the interaction are computed as the derivative of the predicted probability with respect to *Signup* at each level of *Credit score* and *Homeowner*. These effects are estimated using the delta method—averaged over the observed covariate values (Ai and Norton 2003).

Table A.4: **The effect of data sharing on interest rates (subsample, 2021–2022)**

	Matched sample
Signup	−0.0265***
	(0.0003)
Signup × Credit score (A–D) (Base)	
Signup × Credit score (E–G)	0.0008**
	(.0004)
Signup × Credit score (H–K)	0.0043***
	(0.0055)
Signup × Credit score (L–M)	0.0154***
	(0.0050)
Homeowner	−0.0174***
	(0.0003)
Signup × Homeowner	0.0056***
	(0.0004)
Controls	X
Dummy	Month–Year
Cluster (region-month-year)	X
N	158,652
Adjusted R2	0.5300

Notes: This table reports the results using the subsample (2021–2022) from Equation (3) which estimates the effect of data sharing, *Signup*, a dummy variable that takes a value of one if the applicant shared data and zero otherwise, on interest rates using the matched sample. *Credit score* range from A–D (highest) to L—M (lowest). Interest rates are revealed only for approved applications. Thus, approved applicants who shared data are matched one-to-one with approved non-sharers without replacement. Propensity score matching is used for *Age*, *Loan amount requested*, and *Income decile*, and exact matching is used for *Credit score*, *Homeowner*, *Female*, *Access channel* and *Loan application month-year*. The final sample includes 158,652 loan applications from 158,652 unique applicants. In the case of multiple applications per applicant, only the first is included.