The Distributional Effects of Data Interoperability: Evidence from the U.S. Residential Mortgage Market*

Dam Linh Nguyen†

November 6, 2025

Preliminary Draft

Abstract

We explore the equilibrium consequences of consumer information-sharing between firms in credit markets. Exploiting a novel borrower-lender search dataset and the entry of an industry-led interoperable data network, we report that in-network firms enjoy a relative (1) increase in applications and (2) decrease in ex-post default among originated mortgages. These patterns suggest two potential mechanisms at play: reduced frictions for borrowers and screening efficiency gains for in-network lenders due to improved information access. We propose and estimate a model of consumer search and firm pricing, which internalizes these economic forces. Estimated extra costs of applying to out-of-network lenders are 35% of search costs. Interoperability enhances screening accuracy by 58% and 70% for borrowers with high- and low-repayment prospects, respectively. Counterfactuals reveal that, while overall surplus rises with an industry-wide data-sharing mandate, the division of welfare benefits may be uneven. In our setting, surplus increases by 8% for less risky borrowers, who account for the majority of the consumer pool, but decreases by 30% for risky types: a result driven by strategic selection into application and uniform screening boost. When lenders differ in their ability to improve screening (e.g., due to technological constraints), data interoperability can reduce profits for less technologically advanced firms.

*I am grateful to Chris Conlon, Quang Vuong, and Dan Waldinger for their vital guidance and unwavering support. Special thanks extend to the nurturing IO group at NYU Stern. This paper benefited from helpful discussions with seminar participants at the NYU Stern IO Workshop and NYU Applied Micro Lunch. Any remaining errors are my own.

†Department of Economics, New York University; n.linh@nyu.edu.

1. Introduction

Emerging research on data interoperability has documented a growing shift toward consumer information standards and the adoption of technologies that facilitate efficient data transmission between firms—most notably in the financial sector.[1] By 2021, at least 80 governments had initiated some form of open banking mandate, including regulations requiring banks to share standardized consumer financial transaction data with third parties or to support the development of application programming interfaces ("API") and mobile applications that enable consumer-to-business information sharing (Babina et al., 2024). While policy interventions aimed at promoting interoperable solutions are becoming increasingly widespread, our understanding of the distributional effects of data-sharing and their determinants remains limited.

We fill this void by characterizing the channels through which alternative interoperable regimes in credit markets influence the division of welfare gains across consumers and firms. Our contribution to the existing literature is fourfold.

First, we fuse data from multiple sources to construct a new borrower-level home mortgage dataset with information on consumer application history, lender identity along the search path, and data on the joint distribution of search, price, and ex-post default for a subset of U.S. residential loans between 2018 and 2021.

Second, we use this novel dataset to descriptively analyze the impact of the entry of an industry-led interoperable network: the Financial Data Exchange ("FDX"). We find that FDX firms experience a relative increase in applications and originations alongside a decline in ex-post borrower default rates—while offering prices comparable to those of non-FDX lenders. These patterns suggest two possible mechanisms at work: (a) reduced *transaction costs* for borrowers seeking mortgages from in-network lenders, e.g., due to lower hassle costs associated with manually preparing and uploading financial records; and (b) increased *screening efficiency* for in-network firms, e.g., because of improved information acquisition enabled by standardized data formats or extended panel of borrowers' financial transaction history.

Third, we develop and estimate a structural model of binary-type consumer search and firm pricing, which embeds these two economic forces. Our framework builds on the model of Agarwal et al. (2024), but importantly introduces features that capture the institutional nuances of our setting and the potential drivers of the observed empirical patterns. In particular, we introduce the notion of (1) borrower home bank, (2) lender identity/type, (3) interoperable network; and

[1] Beyond finance, there is growing regulatory interest in integrating consumer data systems in other industries, including healthcare (e.g., California's Data Exchange Framework, DxF, initiative).

incorporate potential (4) borrower transaction costs, and (5) screening efficiencies for in-network as well as in-house lending.

In the model, borrowers sequentially search for mortgages. Upon observing an offer rate, consumers decide whether to apply for a loan or continue searching in the next period. Submitting an application is cost-free for borrowers matched with their home bank or when the borrower's primary bank and the potential lender can share consumer financial data via API. Otherwise, the borrower incurs an additional transaction cost when applying for a mortgage.

Lenders do not perfectly observe borrowers' repayment prospects, but upon a careful review of application materials, receive a noisy—yet informative—signal of each borrower's risk type. When firms assess the creditworthiness of their own customers or borrowers from affiliated in-network banks, we allow for a type-specific boost in screening accuracy due to improved information access. Each firm sets its offer rate to maximize expected profits.

Our framework generates insights into the inner workings of borrower selection in the presence of data-sharing. Specifically, it clarifies that absent of screening efficiencies, any reduction in transaction frictions leads to an increase in applications. However, if data-sharing enhances screening, then this "application effect" becomes amplified for less risky consumers and dampened for risky types. Indeed, this latter force can discourage riskier borrowers from ever applying—an implication of interoperability that may exacerbate the exclusion of risky consumers. For in-network firms, the combination of the application effect and screening boost results in an overall increase in applications and a higher share of originations associated with less risky borrowers—a benefit for in-network firms which we refer to as the "attraction effect."

On the flip side, in-network firms are more likely to lose their own customers to competing in-network firms ("diversion effect"), as a result of data-sharing. In fact, if interoperability enhances screening, then this diversion effect is mitigated for less risky consumers and exacerbated for risky types. This implies that, while data-sharing may lead to an undesirable reduction in the number of originations from in-network firms' own customers, it could result in a more favorable mixture of borrowers due to the exclusion of riskier in-house applicants. The aggregate and distributional equilibrium implications of the countervailing application, attraction, and diversion effects depend fundamentally on the underlying demand and supply model primitives.

Estimation of the model parameters is facilitated by the relationships between realized prices, ex-post default, and borrower search activity. Identification of transaction costs (resp., screening efficiencies) further leverages the differences in application (resp., default) rates between FDX and non-FDX firms (and in-house vis-à-vis external lending), while exploiting the variation in home bank and interoperable network presence across geographies and years.

Among the key parameters, we estimate transaction costs (i.e., extra costs of applying to out-of-network lenders) at 35% of average search costs. We also find that, in our setting, interoperability improves screening accuracy by 58% and 70% for consumers with high- and low-repayment prospects, respectively.

Finally, to quantify the effects of interoperability, we simulate equilibria under alternative data-sharing and screening efficiency structures. Transitioning from a no-interoperability regime to an industry-wide full-interoperability mandate results in an aggregate welfare gain. However, we find that the distribution of welfare effects can be unequal. Given our estimated parameters, consumer surplus rises by 8% for less risky borrowers, who represent the majority of the consumer pool, but declines by 30% for risky types—an outcome driven by the exclusion of the latter borrowers due to screening efficiencies and a near-doubling of their opt-out rates.

We also consider a counterfactual scenario in which lenders differ in their capacity to improve screening (e.g., due to technological constraints). We find that with this heterogeneity, data interoperability can lead to a decline in profits for less tech-savvy firms—a result that could temper the commonly cited pro-competitive benefits of data-sharing, particularly if the profit drop is substantial enough to induce market exit by relatively low-IT lenders.

Related Literature. This article contributes principally to two strands of literature: (1) data interoperability and portability; and (2) consumer search and firm competition in credit markets.

Interest in the role of information asymmetry and transparency in credit markets has a long tradition in economic and finance research. Early theoretical work in this arena has highlighted (i) the importance of pricing as a screening mechanism in the presence of adverse selection; and (ii) the significance of heterogeneity in the access to consumer data as a determinant of rationing equilibria (e.g., Jaffee and Russell, 1976; Stiglitz and Weiss, 1981). A substantial body of research has also explored how credit registries—i.e., regulatory monitoring and distribution of loan performance data to financial institutions—affects borrower sorting, credit supply, pricing responses, and the resulting welfare consequences (e.g., Pagano and Jappelli, 1993; Hertzberg et al., 2011; Djankov et al., 2007; Brown et al., 2009).

More recently, the nascent literature on data interoperability and portability has explored the interplay of data-sharing, competition, the rise of financial technology ("FinTech") firms, and consumer privacy (e.g., Babina et al., 2024; Yin, 2024; He et al., 2023; De Corniere and Taylor, 2024; Parlour et al., 2022; Acemoglu et al., 2022; Jones and Tonetti, 2020). Leveraging a policy experiment that provided lenders with extended access to borrower information, Yin (2024) finds that the benefits from additional data provision accrue primarily to high-technology firms—largely, due to their superior ability to process information efficiently. Babina et al. (2024) employ a calibrated model of consumer data usage to highlight the pro-competitive effects of open banking—increased competition, FinTech entry—as well as the negative consequences for risky consumers and borrowers who opt out of data-sharing due to privacy concerns.[2] We complement this research by (a) providing new descriptive evidence on the empirical implications of the entry of an industry-led interoperable network; and (b) quantifying the reduction in transaction frictions and the improvement in screening efficiencies to study the distributional consequences of interoperability on heterogeneous consumers and firms.

In doing so, this paper also furthers the extensive literature on consumer search and firm competition in credit markets—including research that explores the role of information and search frictions in shaping consumer selection and firm pricing (e.g., Agarwal et al., 2024; Allen et al., 2019), financial inclusion (e.g., Blattner et al., 2022), competition (e.g., Yannelis and Zhang, 2023), and merger effects (e.g., Allen et al., 2014). Methodologically, our framework most closely relates to the model developed by Agarwal et al. (2024), which they use to show that equilibrium prices rise in response to borrower discrimination and more restrictive lending policies. We extend their original framework by incorporating new features into our model, including: borrower home bank, lender identity, interoperable network, transaction costs, and screening efficiency to study the aggregate and distributional effects of data-sharing. While our framework is developed with the mortgage market in mind, it can be readily adapted to study related questions in other credit markets such as small business credit or automobile financing.

Roadmap. The remainder of this paper proceeds as follows. Section 2 provides institutional background and outlines the procedure used to construct the borrower-lender search dataset. Section 3 presents our reduced-form analysis and documents descriptive evidence on the relationship between data-sharing and prices, applications, originations, as well as ex-post default. Section 4 introduces a structural model of consumer search and firm pricing that captures possible drivers of the reported empirical patterns, including transaction costs and screening efficiencies. Section 5 details the estimation procedure and presents our parameter estimates. Section 6 compares equilibrium outcomes and welfare under alternative counterfactual cases. Section 7 concludes with a discussion of potential directions for future research.

[2] In contrast to Babina et al. (2024), our results indicate that for a subset of borrowers, surplus would decline under an industry-wide data-sharing mandate. Possible explanations of our contrasting conclusions include differences in empirical settings and modeling choices. Regarding the latter, it is worth noting that—for computational tractability—we abstract from endogenizing consumer data-sharing participation. Incorporating this feature into our framework would likely offset some of the (negative) screening effects of interoperability on risky borrowers, as voluntary disclosure would yield at least partial safeguard against low-type revelation during the application review process. Similarly, we assume away endogenous firm entry and exit. A priori, including this dimension could also soften some of the adverse impacts of data-sharing on risky borrowers.

2. Background & Data

We focus on the residential mortgage market as our empirical application for several reasons. First, home lending represents a significant segment of the economy. In the United States, e.g., mortgage originations totaled $2.8 trillion (or 11% of gross domestic product) in 2022.

Second, it provides a uniquely compelling environment for studying the influence of data-sharing on consumers and firms. For borrowers, (re)financing a home purchase can broadly be summarized as a two-step procedure. In the first stage, consumers search for a potential lender, a process that may involve substantial search costs (e.g., Agarwal et al., 2024; Allen et al., 2014).

In the second stage, borrowers undergo a comprehensive application process, which typically involves submitting information about the (re)financed property and documents indicating borrower financial health—including, pay stubs, tax returns, and bank statements. Lack of interoperability—which would otherwise facilitate efficient borrower-to-lender transfer of consumer financial data—can create additional frictions in the form of extra application (or transaction) costs such as hassle costs of manually preparing and uploading required documents. This, in turn, can lead to borrower lock-in effects (e.g., Sharpe, 1990; Kim et al., 2003), whereby consumers are disincentivized from seeking credit outside their home bank and, consequently, increased barriers to entry (e.g., Klemperer, 1995).

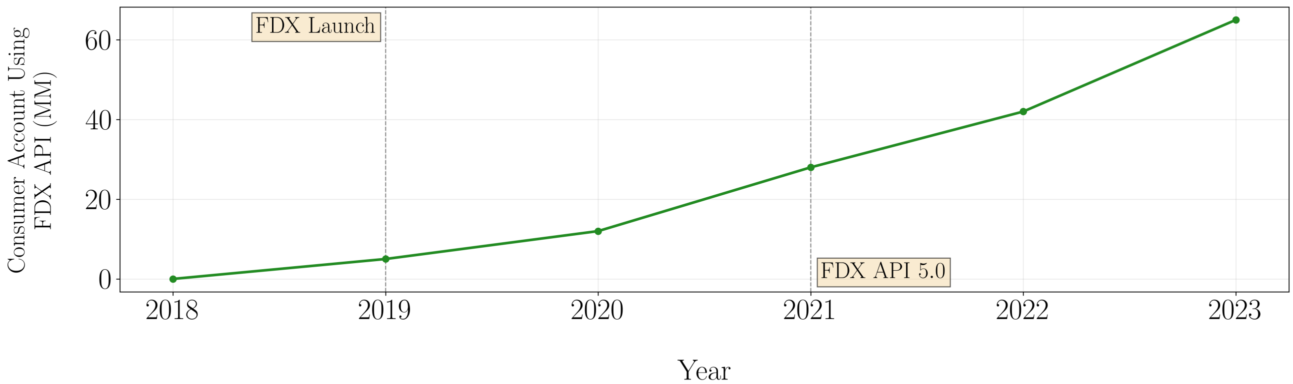
Based on these materials, a loan officer at the lending institution conducts a careful review of the applicant's creditworthiness. This culminates in either a loan rejection or approval. If the loan terms are accepted by the consumer, the mortgage is originated. Interoperability can improve screening accuracy and reduce inefficiencies arising from asymmetric information, if lenders are able to efficiently process additional or enhanced borrower data (e.g., Yin, 2024). Appendix Figure 1 illustrates the potential efficiency gains induced by data-sharing.

2.1 Interoperable Data Network: Financial Data Exchange (FDX)

Data interoperability in our setting is facilitated by FDX, an industry-led nonprofit organization. Launched in 2019, FDX develops, maintains, and improves its API technology for secure access to standardized consumer financial transaction data. By 2023, FDX had amassed north of 200 financial industry members, including depository banks, FinTechs, credit unions, and data aggregators, among others.

The FDX API consumer base has also grown substantially since its inception: from 2 million users in 2019 to over 50 million in 2023—with nearly 3.6 billion API calls per month. Figure 1 plots the evolution of the number of accounts utilizing the FDX API from 2018 to 2023.

Figure 1 – FDX API Consumer Usage over Time



Notes: consumer accounts using FDX API (in millions) from 2018 through 2023.[3]

Since its launch in 2019, the FDX API has undergone a series of advancements. Notably, 2021 marked a major milestone with the introduction of FDX API 5.0, an upgrade that substantially improved the design of the consumer control dashboard, privacy consent mechanisms, and alignment with globally interoperable standards, while enabling "two-way" consumer information-sharing between data providers and third-party FinTechs. In the organization's own words:

> "FDX API 5.0 represents FDX's first major release since early 2020 and significantly expands the standardization of consumer data sharing in the financial industry."[4]

2.2 Data

We combine three publicly available data sources to arrive at our borrower-lender search dataset. Our primary database comes from the Home Mortgage Disclosure Act ("HMDA"). It provides a near-universe of mortgage applications in the United States with information on loan terms (e.g., amount, interest rate); borrower attributes (e.g., income, debt-to-income ratio), property characteristics (e.g., dwelling, occupancy type); lender details (e.g., name, parent organization); and application outcomes (e.g., origination, denial). The HMDA data are available annually at the Census tract level, a small statistical geographic unit.[5]

[3]Last accessed: November 24, 2024. Available at https://www.financialdataexchange.org/FDX/News/Press-Releases/Financial_Data_Exchange__FDX__Reports_76_Million_Consumers_Use_FDX_API.aspx

[4]Last accessed: November 24, 2024. Available at: https://www.financialdataexchange.org/FDX/FDX/News/Press-Releases/Financial_Data_Exchange_Releases_FDX_API_5.0.aspx

[5]The population of each Census tract typically ranges from about 1,200 to 8,000 individuals.

We supplement HMDA with two additional sources. We use the Freddie Mac Single-Family Loan-Level Performance Dataset ("SFLPD"), which provides information on ex-post delinquencies as well as other monthly performance metrics for a subset of originated loans in HMDA (30-year, fully amortizing, single-family, conforming fixed-rate mortgages). We also utilize banking presence—measured by annual branch-level deposits—as a proxy for bank-level consumer data ownership. This information is sourced from the Summary of Deposits ("SOD") data maintained by the Federal Deposit Insurance Corporation ("FDIC").

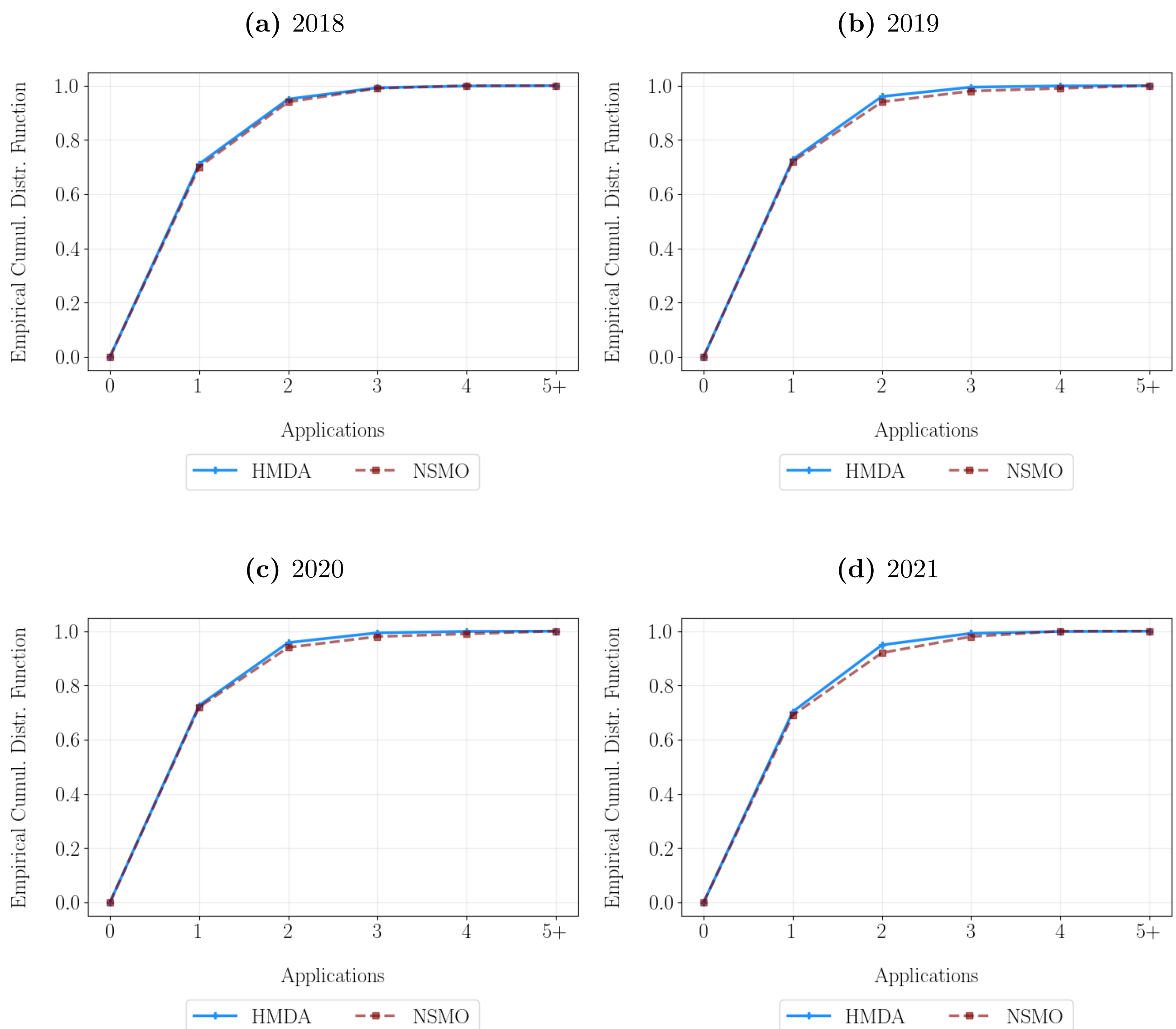
Data Build. Our process of molding these three data sources together proceeds in two steps. A key limitation of the public HMDA data is the absence of a consumer identifier that would directly allow us to link applications and originations. We overcome this restriction by internally matching HMDA applications in the first stage of our data construction. Specifically, we match applications using detailed characteristics (census tract, loan amount, and borrower income). To the best of our knowledge, this is the first paper to construct a borrower-level search dataset within HMDA.[6] In the second step, we match originations identified in the first stage with Freddie Mac SFLPD using information on loan amount, borrower income, location (Metropolitan Statistical Area, "MSA") as well as combined loan-to-value ("CLTV") and borrower debt-to-income ("DTI") ratios.[7] Appendix Section D provides additional data construction details.

To validate our data matching procedure, we compare the resulting borrower-level distribution of applications against that reported in the National Survey of Mortgage Originations ("NSMO"), a quarterly survey of borrowers who originated home loans. Figure 2 shows that the cumulative shares of borrowers (vertical axis) with a given number of applications (horizontal axis) align closely between our matched data and NSMO. Notably, the search distribution is quite stable across years, with about 71% of borrowers applying to only one lender, 23% to two, 5% to three, and less than 1% to more than four lenders.

[6]Other research has combined HMDA with third-party data. For instance, Goodman et al. (2019) merge HMDA with CoreLogic data to analyze trends in the real denial rate over time (i.e., loan rejection rates adjusted for changes in borrower composition).

[7]In a separate context, Liu (2023), for instance, match HMDA with Freddie Mac data to explore the relationship between language barriers and mortgage-related outcomes such as applications, loan terms, and default.

Figure 2 – Data Build Validation: Matched HMDA vs. NSMO



Notes: share of borrowers, who originated mortgages, with a given number of applications in the internally matched/constructed HMDA data (solid blue line) and NSMO dataset (dashed red line).

Analysis Sample. As a result of our data construction procedure, we obtain an analysis sample that includes information on consumer application history, lender identity along the search path, and data on the joint distribution of search, price, and ex-post default from 2018 through 2021.

In our analysis sample, we observe the loan search activity of nearly 750 thousand borrowers, 96% of whom originated mortgages, while the remaining 4% did not. Table 1 reports high-level attributes for each of these two consumer groups. Consistent with our expectations, we find that—compared to borrowers who did not originate loans in Columns (3)–(4)—originating consumers, in Columns (1)–(2), tend to have higher incomes, take out larger loan amounts, both in dollar terms and as share of property value (CLTV), and appear to be less risky based on observable characteristics as indicated by lower DTI ratios. Finally, we find that an average originating borrower, for whom we observe credit score, falls into the "very good" repayment category, according to the FICO classification.

Table 1 – Analysis Sample (2018-2021)

	Originations		Non-originations	
Statistic	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Income ($000s)	105.0	53.2	92.1	55.1
Loan amount ($000s)	270.1	119.0	190.5	123.3
Combined loan-to-value ratio	74.1	15.8	73.4	21.4
Debt-to-income ratio	35.5	8.4	44.8	10.1
Credit score	752.8	126.5	–	–
Borrowers	**714,629 (96%)**		**30,241 (4%)**	

Notes: summary statistics of the analysis sample for borrowers who originated, Columns (1)–(2), and those who did not originate mortgages, Columns (3)–(4).

In Table 2 we zoom into the origination subsample and summarize the average profile of FDX-affiliated banks, Columns (1)–(2), and other depository lending institutions, Columns (3)–(4). We also report the average loan and borrower attributes of consumers who originated loans with each of these two lender types.[8]

FDX banks offering lending services consist of established institutions: Bank of America, JPMorgan Chase Bank, PNC Bank, TD Bank, Truist Bank, US Bank, and Wells Fargo. While there are only seven FDX banks in our analysis sample,[9] they account for a substantial share of originations (38%) and deposits (24%). Across all analysis periods, including pre- and post-FDX years, borrowers originating at FDX banks apply to approximately 1.3 lenders and exhibit a

[8]Appendix Table 1 provides additional summaries of originating borrower characteristics.

[9]Note that the number of FDX firms in our analysis sample is constrained by the limitations of our data. As of 2024, there are 25 mortgage lenders that are FDX members. However, only a subset (i.e., 7) of them originate loans that are backed in the secondary market by Freddie Mac and thus appear in SFLPD. Among FDX non-banks, our analysis sample also includes Rocket Mortgage, a prominent FinTech lender.

relatively lower default rate (2.1% compared to 2.7% for non-FDX banks).[10] They also tend to receive a slightly lower interest rate (3.3% vs. 3.7%). However, after accounting for observable borrower characteristics, the residualized interest rate is slightly higher for consumers borrowing from FDX depository institutions, compared to other banks. To obtain the residualized interest rate, we first regress the raw interest rate on observable attributes,

$$r_{ijgt} = \boldsymbol{\gamma}'\boldsymbol{X}_i + \kappa_j + \mu_g + \alpha_t + \varepsilon_{ijgt},$$

and then compute the difference between the observed rate, r_{ijgt}, and the predicted rate, $\hat{r}_{ijt} \equiv r_{ijgt}(\hat{\boldsymbol{\gamma}})$, controlling for loan amount, term, borrower income, FICO score, CLTV, and DTI, contained in the vector of borrower characteristics, $\boldsymbol{X}_i$; and estimated firm, $\hat{\kappa}_j$, MSA, $\hat{\mu}_g$, and year, $\hat{\alpha}_t$, fixed effects.

Table 2 – Origination Subsample (2018-2021)

	Origination Lender			
	FDX Deposit Banks		Other Deposit Banks	
Statistic	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Applications (Count)	1.3	0.6	1.3	0.5
Interest rate (%)	3.3	0.6	3.7	0.8
Residualized int. rate (%)	0.03	0.37	-0.01	0.35
Default rate (%)	2.1%		2.7%	
Borrowers	**62,763 (38%)**		**101,701 (62%)**	
Lenders	**7 (1.7%)**		**410 (98.3%)**	
Deposit Share/MSA	**24.1%**		**16.7%**	

Notes: summary statistics of the analysis sample for loans originated at FDX depository institutions, Columns (1)–(2), and all other banks, Columns (3)–(4).

In Section 3.2, we dive deeper into the differences between FDX and other deposit banks, while importantly accounting for the timing of FDX entry, its major API upgrade, as well as firm-, MSA-, and year-level fixed effects.

[10] Default rate is computed as the share of originating borrowers who experience 90 or more days of delinquency.

3. Descriptive Patterns

The analysis sample reveals four key data patterns that motivate our econometric model.

3.1 Search, Price, and Default Associations

To study the relationship between prices, default, and search activity, we restrict our attention to the origination subsample, and estimate the following ordinary least squares ("OLS") specification:

$$y_{ijgt} = \sum_{s \geq 2} \beta_s \mathbb{1}[s_i = s] + \gamma' X_i + \kappa_j + \mu_g + \alpha_t + \varepsilon_{ijgt}, \tag{1}$$

where s_i denotes the number of lenders that borrower i applied to; and the outcome variable y_{ijgt} is either (residualized) interest rate or a dummy variable that takes a value of 1 if borrower i—who originated a mortgage at lender j in year t to finance a home purchase in MSA g—defaulted on their interest obligations (had at least one spell of 90+ days of delinquency); and 0 otherwise; κ_j, α_t, and μ_g are lender, year, and MSA, fixed effects, respectively.

Fact 1 (Price Decreases with Search). We find a monotonically decreasing relationship between prices and search. Consumers who consider two lenders tend to originate loans at an interest rate that is 0.01 percentage points lower than the average rate of 3.42% obtained by borrowers who apply only once. This difference widens as consumers increase search activity, as depicted by the coefficient estimates in Panel (b) and the data pattern in Panel (a) of Figure 3. This suggests that, on average, increased search benefits consumers through lower prices.[11]

Fact 2 (Default Increases with Price). We also observe a monotonically increasing positive association between default and interest rates. Based on the following OLS specification, where d_{ijgt} equals 1 if default is realized and 0 otherwise,

$$d_{ijgt} = \beta r_{ijgt} + \gamma' X_i + \kappa_j + \mu_g + \alpha_t + \varepsilon_{ijgt}, \tag{2}$$

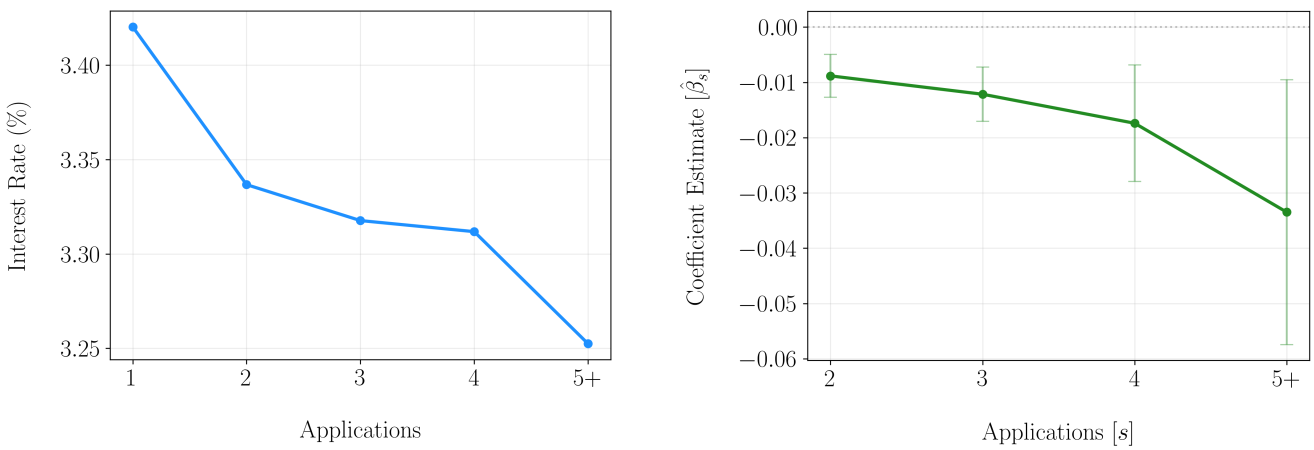
a 1 percentage point increase in the interest rate, is, on average, associated with a 1.74 percentage point increase in the likelihood of default (Figure 4). This is consistent with the interpretation that relatively larger debt obligations are more likely to induce default.[12]

[11] An analogous reduced-form analysis of the relationship between default rates and search reveals little systematic pattern (Appendix Figure 3).

[12] Using residualized interest rates—which control for observed borrower characteristics—uncovers a more nuanced U-shaped relationship (Appendix Figure 4). Specifically, default rates decrease with interest rates at low price points, but then monotonically increase at moderate to high residualized interest rates.

Figure 3 – Interest Rate and Applications

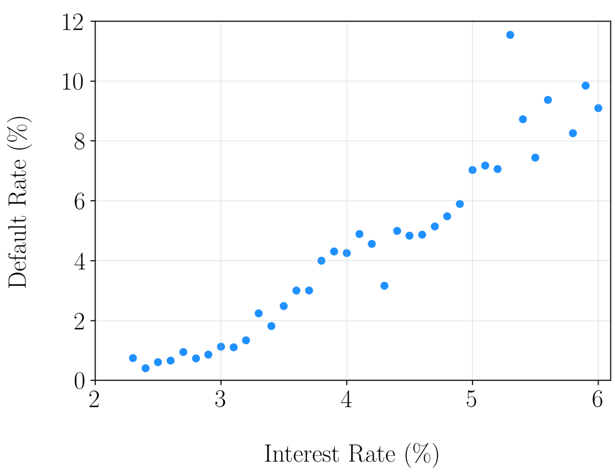
(a) Analysis Data Relationship　　**(b)** OLS Coefficient Estimates



Notes: association between interest rates and applications. Panel (a) plots the average interest rate at each count of applications. Panel (b) shows coefficient estimates from the OLS regression in Equation (1) with interest rate as the outcome variable. Adj. $R^2 = 0.74$; obs. = 714,629.

Figure 4 – Default and Interest Rate

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

	Default Rate (%)
Interest Rate (%)	1.74***
	(0.26)
Year, MSA, Lender FEs	✔
Borrower Char. Controls	✔
Avg. Outcome	2.26
Adj. R^2	0.023
Observations	714, 629

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Notes: association between default and interest rates. Panel (a) plots the average default rate at each interest rate. Panel (b) shows coefficient estimates from the OLS regression in Equation (2).

3.2 Data Interoperability and Search, Price, Default

Shifting gears, in this subsection we explore the relationship between data-sharing and search activity, price, and default—focusing on depository institutions that hold consumer data.

Any effects of interoperability require the presence of at least two FDX firms in a given market. Consequently, for the purposes of the current analysis, we further restrict our analysis sample to year-MSA combinations where at least two FDX firms have a banking presence. Given the FDX timeline—and especially the API 5.0 upgrade in 2021—we also extend our sample period to include data from 2022.[13]

We employ a standard event study design with the entry of FDX in 2019 as our initial point of reference; and the major advancement of FDX's API technology in 2021 as a secondary event,

$$y_{jgt} = \sum_{\tau=0}^{3} \beta_\tau \mathbb{1}[t - 2019 = \tau]\mathbb{1}\{j \in \text{FDX}] + \gamma s^d_{jgt} + \kappa_j + \mu_g + \alpha_t + \varepsilon_{jgt}, \tag{3}$$

where τ denotes the number of years elapsed from the introduction of FDX in 2019; and the outcome variable y_{jgt} is either application, origination count, default rate, or interest rate charged by lender j in year t and MSA g. As before, κ_j, α_t, and μ_g stand for lender, year, and MSA, fixed effects, respectively. Finally, to isolate the contribution of own banking presence on outcome variables, we control for the deposit share, s^d_{jgt}, of lender j in MSA g, and year t.

Identification in our setting relies on the standard assumption in difference-in-differences specifications: absent FDX entry, FDX and other banks would follow "parallel trends." In an ideal data environment, we would test this assumption by establishing a similar evolution of key outcome variables pre-FDX launch. Unfortunately, this direct approach is infeasible due to data constraints: a subset of the key data fields used in our matching procedure (Section 2.2) is unavailable in historical HMDA data—preventing us from extending our analysis sample to encompass years prior to 2018. In lieu of this, we take an indirect approach by analyzing the evolution of consumer deposits of FDX and non-FDX firms in the pre-period. The idea here is that if the outcome variables are correlated with deposits, as we generally find to be the case (Appendix Figures 5–6), then deposits can serve as a proxy for diagnosing pre-trends in our response variables. Indeed, as shown in Appendix Figure 9, we find little evidence of pre-treatment trends. We are now ready to present our two main descriptive results.

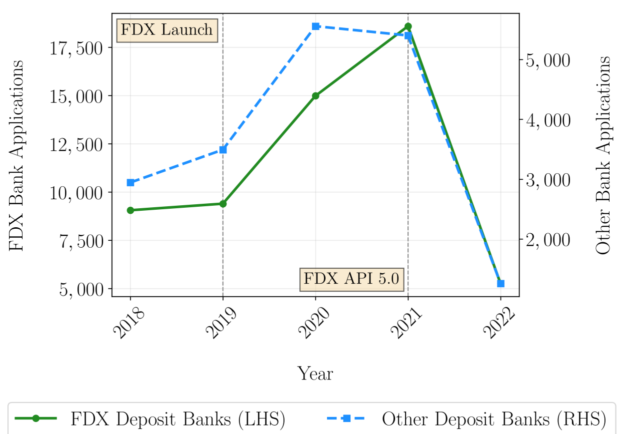
[13] In the first quarter of 2022, the Federal Reserve raised interest rates for the first time since 2018—a potential source of confounding, which we address by excluding 2022 data from all other analyses.

Fact 3 (Applications & Originations Increase with FDX API 5.0 Release). While the patterns during the intermediate period between the FDX launch and the API 5.0 upgrade are somewhat noisy—including, plausibly due to early-stage consumer adoption of new data-sharing technology (Figure 1) and transitional phase of integrating interoperable solutions into firms' existing data systems—we find that post-FDX API 5.0, firms within FDX experience an increase in applications. For instance, reading from Figure 5, we see that FDX data-sharing is associated with an average increase of about two applications per MSA-year for FDX firms (or a 10% rise considering the pre-period average of 20 applications per FDX firm, MSA, and year). Appendix Figure 10 shows a similar pattern for mortgage originations.

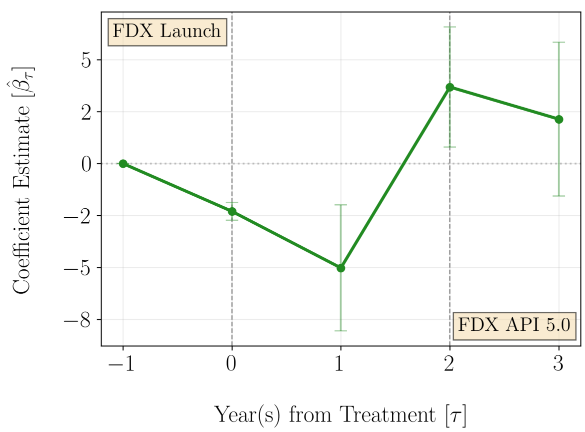
Fact 4 (Default Decreases Post-FDX Launch). We also find that FDX data-sharing is associated with a general decrease in default rates; for instance, a decline of approximately 1 percentage point in 2021—a substantial drop relative to the pre-period average rate of 5%, as shown in Panel (b) of Figure 6.[14]

Figure 5 – Applications: FDX and Other Banks

(a) Analysis Data Relationship



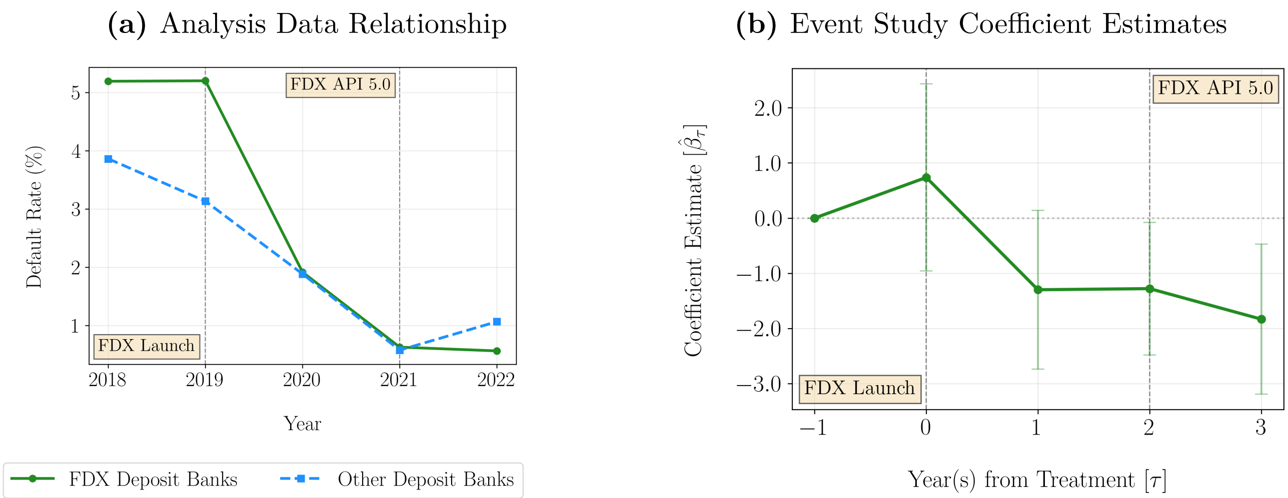
(b) Event Study Coefficient Estimates



Notes: differences in applications between FDX and other banks. Panel (a) plots the number of applications. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with the count of applications as the outcome variable. Adj. R^2 = 0.52; obs. = 3,060.

[14] A similar event study analysis using prices as the outcome variable shows no systematic relationship between (residualized) interest rates and FDX data-sharing (Appendix Figures 11–13).

Figure 6 – Default Rate: FDX and Other Banks



Notes: differences in default rates between FDX and other banks. Panel (a) plots mean default rates. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with default rates as the outcome variable. Adj. $R^2 = 0.068$; obs. = 3,060.

Taking stock, the detailed structure of our borrower-lender search dataset and the general data patterns in Facts 1 and 2, naturally point us toward adopting the model of Agarwal et al. (2024) as a compelling starting point. Fact 3, which shows that FDX firms experience a relative increase in applications and originations, motivates an extension of the original model to allow for a potential reduction in transaction (or extra application) costs in the presence of data-sharing. Finally, Fact 4, which suggests a relative decline in default rates for FDX firms amid no systematic changes in prices, prompts two additional model features: (1) potential screening efficiency gains for in-network lenders due to improved information acquisition; and (2) heterogeneity in borrower sorting, conditional on their home bank's data-sharing affiliation.

4. Model

We build on the framework of Agarwal et al. (2024) and extend it to incorporate model features that reflect (i) the institutional details of our empirical setting; and (ii) the potential determinants of the data patterns documented in Section 3.

Specifically, we layer onto the "search for approval" model of Agarwal et al. (2024) the notions of (1) borrower home bank, (2) lender identity/type, and (3) interoperable network affiliation. We also introduce (4) transaction costs, which are incurred by borrowers who apply for loans outside of their home bank or interoperable network. Finally, we allow for (5) improved screening, which is enjoyed by lenders when evaluating loan applications from consumers whose data are available at their own bank or at affiliated in-network firms.

In what follows, we occasionally refer to loan products as "mortgages," although our framework can be readily applied to other credit market settings (e.g., small business lending or automobile financing).

4.1 The Environment

Borrowers are indexed by $i \in \{1, 2, \ldots, I\} = \mathcal{I}$. Each borrower is exogenously endowed with a home bank $b_i \in \{1, 2, \ldots B\} = \mathcal{B}$ and a repayment prospect $z_i \in \{h, l\} = \mathcal{Z}$. "High" type borrowers are more likely to repay their loans than "low" type borrowers: $x_h > x_l$, where $x_{z_i} \in [0, 1]$ denotes the repayment probability. The share of high (resp., low) types is labeled as λ_h (resp., $\lambda_l = 1 - \lambda_h \in [0, 1]$).

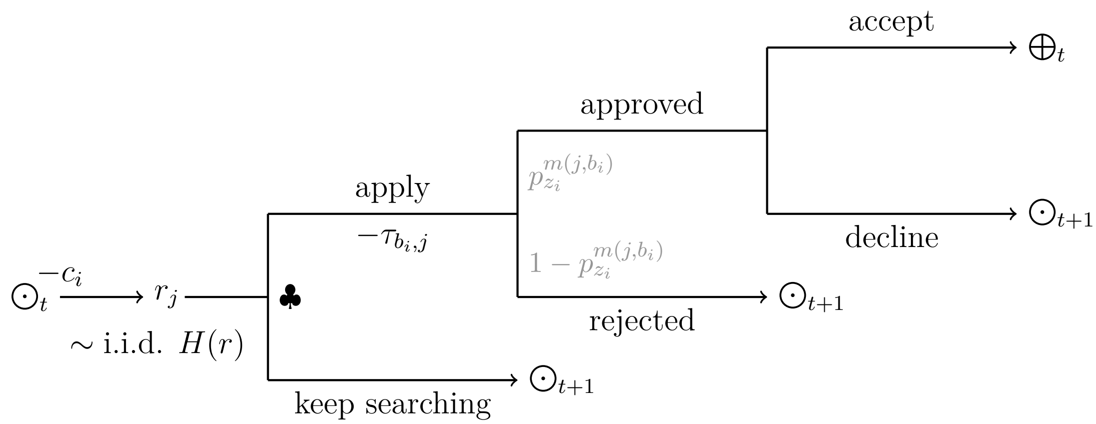
Searching for a mortgage occurs sequentially. Each search iteration t: (a) involves a borrower-specific search cost $c_i > 0$, which is drawn independently from a cumulative distribution function G; (b) leads to a consideration rate r_j, drawn independently from an offer rate distribution H; and (c) results in a (quasi) random match with lender $j \in \mathcal{J}$. Upon observing r_j, the borrower makes a binary decision to apply for the loan or keep searching. Submitting an application is costless for borrower i when matched with their home bank ($j = b_i$); or when i's home bank b_i and lender j are both members of an interoperable network ($b_i, j \in \mathcal{N}$). Otherwise, borrower i incurs an additional "transaction cost" $\tau \geq 0$ when applying for a mortgage; i.e.:

$$\tau_{b_i,j} = \begin{cases} 0 & \text{if } (b_i = j) \text{ or } (b_i \text{ and } j \text{ are interoperable}) \\ \tau & \text{otherwise.} \end{cases}$$

This extra application cost, τ, can be viewed as a hassle cost of, e.g., preparing, scanning, and uploading required financial documents for evaluation at an external lender.

If an application is approved and the borrower accepts the loan terms, then the mortgage is originated, and borrower i receives an indirect utility $u_i(r_j)$. Otherwise, the borrower repeats the search process in the next iteration, $t + 1$. Figure 7 illustrates the borrower search process.

Figure 7 – Borrower Search Process



Notes: illustration of the borrower search process: initial search for lender and offer rate r_j, decision node to apply or keep searching, application approval process, and (non)origination.

Lenders are indexed by $j \in \{1, 2, \ldots, J\} = \mathcal{J}$. Each firm is exogenously characterized by a network affiliation $n_j \in \{0, 1\}$. In-network lenders ($n_j = 1$) maintain standardized consumer data and share an API that facilitates efficient communication of borrower information between firms within the interoperable network $\mathcal{N}$. These benefits of interoperability also extend to in-house lending (i.e., borrowers applying for mortgages at their home bank, $j = b_i$), regardless of the network affiliation status.

We assume that while lenders do not observe each borrower's repayment prospect or search history, they know the distribution of risk types, the identity (k) and consumer data share (s_k^d) of each in-network firm, and the overall size of the interoperable network: $\gamma = \sum_{j \in \mathcal{J}} n_j s_j^d$. In addition, we maintain that lenders observe the identity of each borrower's home bank as part of the application materials.

Screening proceeds as follows. An application submitted by borrower i generates a (noisy) signal of i's risk type ($s_i \in \{h, l\}$). We assume that screening is informative, meaning that a less risky borrower is more likely to send a signal indicating they are a high type compared to a risky borrower. Without interoperability,

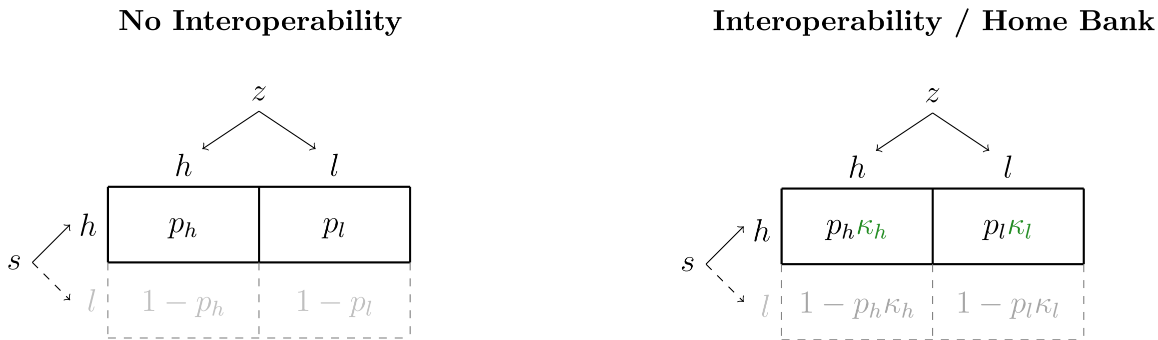
$$p_h \equiv \mathbb{P}[s_i = h | z_i = h] > \mathbb{P}[s_i = h | z_i = l] \equiv p_l.$$

When lenders screen their own customers or borrowers from affiliated banks, we allow for type-specific screening boost, κ_z, due to improved information acquisition (e.g., through standardized data, extended financial information, or instantaneous transmission via API); that is:

$$p_{z_i}^{m(j,b_i)} = p_{z_i} \kappa_{z_i}^{m(j,b_i)},$$

where $m(j, b_i) = \mathbb{1}[j = b_i \text{ or } (n_j, n_{b_i}) = (1, 1)]$ is an in-network/home bank indicator; $\kappa_h^1 \geq 1$; $\kappa_l^1 \leq 1$; and $\kappa_h^0 = \kappa_l^0 = 1$. Figure 8 illustrates the lender screening process.

Figure 8 – Lender Screening Process



Notes: illustration of the lender screening process for lenders without (left panel) and with (right panel) interoperable screening technology; z denotes borrower's true (hidden) type; s labels the signal received by lender during the review process; the top cells show the probability of sending a high signal, $s = h$, given the underlying borrower type $z \in \{h, l\}$.

Once screening is completed, we assume that lenders approve (resp., reject) all mortgage applications that generate a high (resp., low) signal. It follows that, conditional on applying, high (resp., low) type borrowers originate loans at out-of-network/non-home bank lenders with probability $p_h^0 = p_h$ (resp., $p_l^0 = p_l$); and at in-network/in-house lenders with probability $p_h \kappa_h = p_h^1 \geq p_h^0$ (resp., $p_l \kappa_l = p_l^1 \leq p_l^0$). The profit margin earned by lender j, who charges an interest rate r_j is denoted by $\pi(r_j, m_j)$, where m_j encompasses both the screening and origination costs.

4.2 Demand: Borrower Optimal Application Decisions

In each search iteration, t, the borrower's problem boils down to making a binary decision: apply at rate r_j or continue searching in the next iteration, $t + 1$, with the hope of receiving a better offer (♣ in Figure 7 marks the decision node). The optimal choice is determined by comparing the expected utility, net of transaction costs, from applying now against the net expected utility from searching again in a subsequent period. Namely, borrower i will keep searching as long as the latter exceeds the former:

$$\underbrace{-\tau_{b_i,j} + p_{z_i}^{m(j,b_i)} u_i(r_j)}_{\text{net exp. util. from applying now}} \leq \underbrace{-c_i - \mathbb{E}[\tau|n_{b_i}] + \mathbb{E}[p_{z_i}^{m(k,b_i)} u_i(r)|n_{b_i}]}_{\text{net exp. util. from applying next period}},$$

where the calculation of expected transaction costs accounts for the distribution of banking presence as a determinant of a j-specific match,

$$\mathbb{E}[\tau|n_{b_i}] = \begin{cases} -\tau \overbrace{(1 - s_{b_i})}^{\mathbb{P}[j \neq b_i]} & \text{if } n_{b_i} = 0 \\ -\tau \underbrace{(1-\gamma)}_{\mathbb{P}[n_j \neq 1]} & \text{otherwise,} \end{cases}$$

and, analogously, for the computation of expected utility, we have that

$$\mathbb{E}[p_{z_i}^{m(k,b_i)} u_i(r)|n_{b_i}] = \begin{cases} p_{z_i} \int_{\underline{r}}^{r_j} \Big[(1-\gamma)[s_{b_i}^{\text{data}} \kappa_{z_i} + (1 - s_{b_i}^{\text{data}})] + \gamma\Big] u_i(\tilde{r}) dH(\tilde{r}) & \text{if } n_{b_i} = 0 \\ p_{z_i} \int_{\underline{r}}^{r_j} \Big[(1-\gamma) + \gamma \kappa_{z_i}\Big] u_i(\tilde{r}) dH(\tilde{r}) & \text{otherwise.} \end{cases}$$

It follows, by standard results, that the borrower's optimal application decision takes the form of a threshold policy.

Case I ($n_{b_i} = 0$). Borrowers with deposits at out-of-network home banks apply to lender j at rate r_j as long as $r_j \leq r^*$, where the cutoff rate r^* solves

$$\begin{aligned} c_i = &- \Big[\overbrace{(1 - s_{b_i}^d) - (1 - \mathbb{1}[j = b_i])}^{\text{better ``match'' wrt trans. cost}}\Big]\tau \\ &+ p_{z_i} \int_{\underline{r}}^{r^*} \left(\underbrace{\Big[(1-\gamma)[s_{b_i}^d \kappa_{z_i} + (1 - s_{b_i}^d)] + \gamma\Big] u_i(\tilde{r}) - \kappa_{z_i}^{m(j,b_i)} u_i(r^*)}_{\text{better mortgage}} \right) dH(\tilde{r}) \equiv \phi_{z_i}^{0,m(j,b_i)}(r^*). \quad (4) \end{aligned}$$

Case II ($n_{b_i} = 1$). For borrowers with in-network home banks, the threshold rate r^* is determined by a similar relationship:

$$\begin{aligned} c_i = &- \Big[(1-\gamma) - (1 - \mathbb{1}[n_j = 1])\Big]\tau \\ &+ p_{z_i} \int_{\underline{r}}^{r^*} \left(\Big[(1-\gamma) + \gamma \kappa_{z_i}\Big] u_i(\tilde{r}) - \kappa_{z_i}^{m(j,b_i)} u_i(r^*) \right) dH(\tilde{r}) \equiv \phi_{z_i}^{1,m(j,b_i)}(r^*). \quad (5) \end{aligned}$$

These cutoff strategies give rise to four match-specific reservation rate distributions, $F^{n_{b_i},m(j,b_i)}(r^*)$, based on the network affiliation of borrower i's home bank, n_{b_i}, and matched lender, $m(j, b_i)$: $F^{00}(r^*)$ (resp., $F^{01}(r^*)$) for consumers at out-of-network banks facing an offer rate from an external (resp., their own) lender; and $F^{10}(r^*)$ (resp., $F^{11}(r^*)$) for borrowers at interoperable banks with a rate offered by out-of-network (resp., in-network) lender.

In fact, following the observation in Agarwal et al. (2024), if $u_i(r^*)$ is monotonic in r^*, then $\phi(r^*)$ are monotonic in r^*, their inverse functions $\phi^{-1}(c)$ exist and are also monotonic. Consequently, Equations (4)–(5) yield a useful relationship between reservation rates and search costs, expressed in terms of their respective distributions:

$$F_{z_i}^{n_{b_i},m(j,b_i)}(r^*) = G\Big(\underbrace{\phi_{z_i}^{n_{b_i},m(j,b_i)}(r^*)}_{c_i}\Big). \tag{6}$$

4.3 Supply: Lender Optimal Pricing Decisions

Given the demand primitives, lenders compete over prices to maximize their expected profits. We characterize each firm j's optimal price-setting problem as a discrete choice

$$\max_{r_k \in \{r_1,\ldots,r_K\}} \mathbb{E}[\Pi_j(r_k, m_j)] + \varepsilon_{jk},$$

where ε_{jk} is a firm-rate specific idiosyncratic shock drawn independently from a centered distribution with scale σ_ε;

$$\mathbb{E}[\Pi_j(r_k)] = \sum_{z \in \mathcal{Z}} \lambda_z \underbrace{q_{j,z}^{n_j}(r_k)}_{\text{residual demand}} \overbrace{[r_k \tilde{x}_z - m_j]}^{\text{marginal profit}}$$

is the expected profit from offering rate r_k which depends on the marginal profit, $\pi_z(r_k, m_j) = [r_k \tilde{x}_z - m_j]$,[15] and the residual demand captured at that price,

$$q_{j,z}^{n_j}(r_k) = \begin{cases} \gamma q_z^{11}(r_k) + (1-\gamma) q_z^{00}(r_k) & \text{if } n_j = 1 \\ s_j^{\text{data}} q_z^{01}(r_k) + (1 - s_j^{\text{data}} - \gamma) q_z^{00}(r_k) + \gamma q_z^{10}(r_k) & \text{otherwise,} \end{cases}$$

where

$$q_z^{n_{b_i},m(j,b_i)}(r) = \int_r^\infty [H(r^*)]^{-1} dF^{n_{b_i},m(j,b_i)}(r^*).$$

Aggregating the model-implied conditional choice probabilities ("CCPs"), $\mathbb{P}[j \text{ chooses } r_k | m_j; \sigma_\varepsilon]$, across firms yields the offer distribution $H(r)$.

[15] Following Agarwal et al. (2024), the expected repayment can be expressed as $\tilde{x}_z = (x_z - 1)/\log(x_z)$.

4.4 Equilibrium

Equilibrium consists of borrower policies, $\boldsymbol{r}^*(c)$, and lender offer rate distribution, $H(r)$, such that given the demand, $\boldsymbol{\theta}^D \equiv (\boldsymbol{p}, \boldsymbol{x}, \boldsymbol{\beta}_H, \boldsymbol{\beta}_G, \boldsymbol{\kappa}, \lambda_h, \sigma, \tau)'$, and supply model primitives, $\boldsymbol{\theta}^S \equiv (\boldsymbol{m}, \sigma_\varepsilon)'$:

1. $H(r)$ is consistent with the CCPs implied by the lender problem in Section 4.3,
2. $\boldsymbol{r}^*(c)$ is a vector of optimal solutions to the borrower problem in Section 4.2, and
3. $\boldsymbol{q}(r)$ is consistent with lender optimization in Section 4.3,

where $\boldsymbol{\beta}_H$ and $\boldsymbol{\beta}_G$ are parameters governing the offer rate and search cost distributions, resp.; $\boldsymbol{q}(r)$ is the vector of demand captured by r; $\boldsymbol{p} = (p_h, p_l)$; $\boldsymbol{x} = (x_h, x_l)$; $\boldsymbol{\kappa} = (\kappa_h, \kappa_l)$; and $\boldsymbol{m} = (m_j)_{j \in \mathcal{J}}$.

4.5 Model Predictions

Our framework: (i) generates predictions that are consistent with the empirical patterns reported in Section 3; and (ii) elucidates the trade-offs and potential distributional effects of data-sharing on consumers and firms. The following two propositions summarize our key results.

Proposition 1 (Application Effect)**.** *Suppose $\tau > 0$ and that $u_i(r)$ is monotonic in r. Then, under data interoperability ($\tau' = 0$), borrowers at in-network home banks are more likely to apply to affiliated interoperable lenders ("application effect"). Moreover, if data-sharing enhances screening accuracy for high- ($\kappa_h > 1$) and low-type borrowers ($\kappa_l < 1$), then this effect is*

(a) amplified for less risky consumers ($z_i = h$); and

(b) dampened for risky borrowers ($z_i = l$).

Proof. Appendix Section C.1.1. □

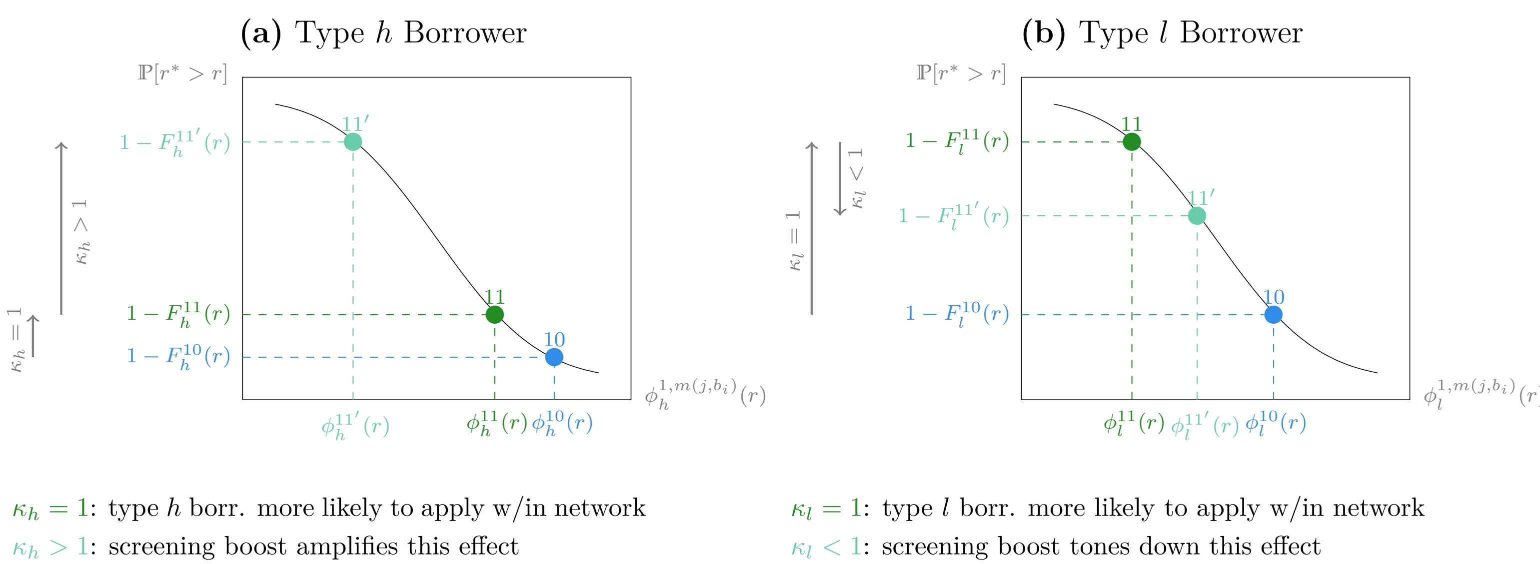
Figure 9 provides intuition for the result in Proposition 1. Take a borrower i with consumer data at an in-network primary bank, $n_{b_i} = 1$. Suppose that i faces an offer r at an out-of-network lender, $m(j, b_i) = 0$, denoted by (10) on the diagram; and the same rate, r, at an in-network lender, $m(j, b_i) = 1$, without screening boost ($\kappa_z = 1$), (11), and with screening improvement ($\kappa_z \neq 1$), (11′). Now, consider two separate cases: when i's repayment prospect is high, Panel (a), and when it is low, Panel (b). In each panel, the vertical axis shows the probability that i applies to each of the three types of lenders; i.e.: the likelihood that i's reservation rate, r^*, exceeds the offer rate r, as derived from the cutoff relations in Equations (4)–(6).

For both types of borrowers, seeking a mortgage within network offers the advantage of reduced transaction costs ($\tau = 0$). This provides an incentive to apply now, rather than delay the search to the next iteration—with the hope of encountering a better offer but at the risk of matching with an out-of-network firm which would entail extra application costs ($\tau > 0$). The resulting borrower sorting into application due to reduced transaction costs benefits in-network firms, as represented in Figure 9 by the vertical gap between nodes (11), (11′), and (10). Indeed, this predicted rise in applications for in-network lenders is consistent with the suggestive evidence in Section 3 showing a relative increase in applications for FDX firms following FDX API 5.0.

When consumer data-sharing improves screening ability for in-network firms ($\kappa_z \neq 1$), the magnitude of this application effect can differ across borrowers. An increase in the approval rate favors less risky consumers, as it increases their chances of originating a mortgage. This, in turn, induces type h borrowers to apply to in-network lenders now, rather than delay their search and gamble on finding a better rate tomorrow, albeit at an out-of-network firm with inferior screening ability. In contrast, risky borrowers are disadvantaged by screening efficiencies, as they are more likely to be revealed as low type and subsequently rejected at in-network firms.

These differing impacts are depicted in Figure 9 by an increase (resp., decrease) in the application probability for high (resp., low) type borrowers as we transition from node (11) to (11′). For in-network firms, this shift in borrower selection due to improved screening can lead to an increase in the relative share of originations associated with high-type borrowers—consistent with FDX firms experiencing a comparatively lower default rate post-launch (Section 3).

Figure 9 – Borrower Selection into Application for In- and Out-of-network Lenders



Notes: application probabilities for high-, Panel (a), and low-type, Panel (b), borrowers at in-network firms who consider the same offer, r, from in- and out-of-network lenders.

As evident from the discussion of Proposition 1, the benefits of data-sharing for in-network firms are linked to its direct impact on borrower selection into application. Proposition 2 balances these advantages against the potential downsides of interoperability for in-network firms.

Proposition 2 (Attraction & Diversion Effect)**.** *Assume $\tau > 0$ and that $u_i(r)$ is monotonic in r. Then, under data interoperability ($\tau' = 0$), in-network firms are more likely to receive applications from consumers at affiliated banks ("attraction effect"). However, they are also more likely to lose their own consumers to competing in-network firms ("diversion effect"). Additionally, if data-sharing enhances screening for high- ($\kappa_h > 1$) and low-type borrowers ($\kappa_l < 1$), then the diversion effect is:*

(a) mitigated for less risky consumers ($z_i = h$); and

(b) exacerbated for risky borrowers ($z_i = l$).

Proof. Appendix Section C.1.1. □

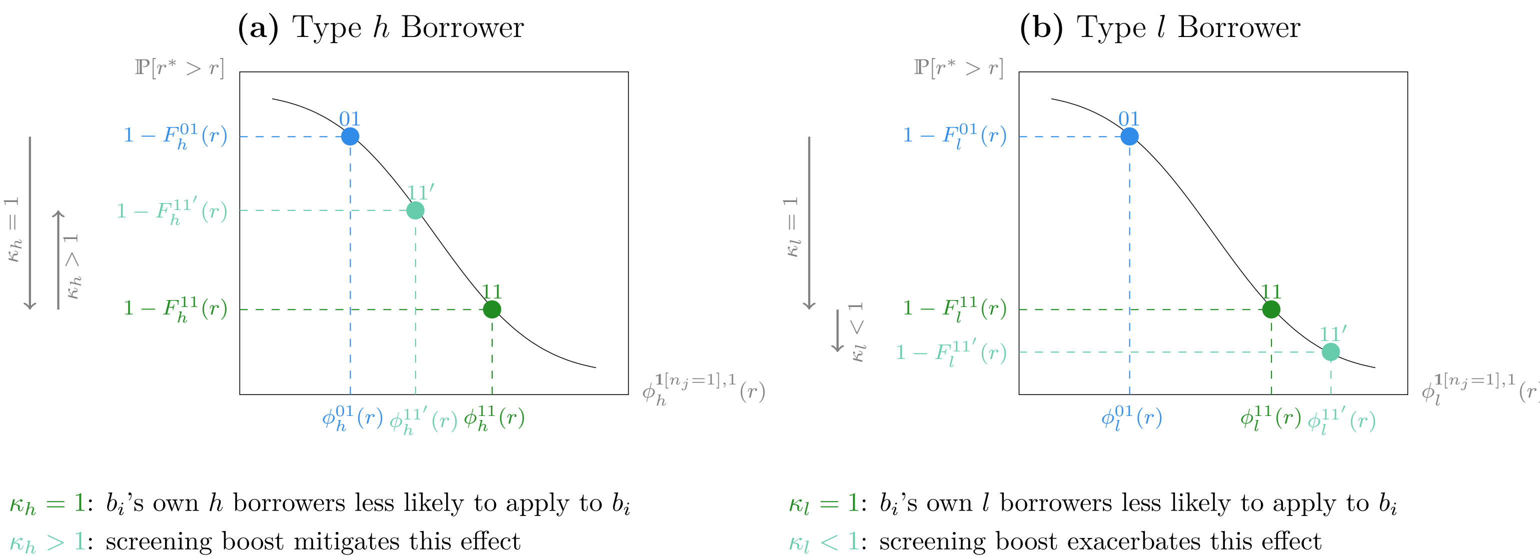
Figure 10 provides intuition underlying the "diversion effect" in Proposition 2. It follows the same layout as Figure 9, except that it considers a borrower i facing an offer r at their home bank b_i when (10) b_i is out-of-network; and when b_i is in-network under two scenarios: (11) with and (11′) without screening improvement.

Absent of interoperability, applying for a mortgage at a home bank carries the benefit of reduced transaction costs. Data-sharing between firms can be viewed as an expansion of the search/application set for consumers. The associated reduction in application frictions for in-network lenders mitigates borrower lock-in effects: borrowers at in-network banks become more likely to apply outside of their primary institutions—a result illustrated by the vertical wedge between nodes (11), (11′), and (10) in Figure 10.

Interestingly, the presence of screening efficiency induces variation in the direction of this diversion effect across borrower types. This is illustrated by an upward (resp., downward) shift from node (11) to (11′) for high (resp., low) type borrowers. To build intuition for this result, recall that screening improvement is particularly valuable to less risky borrowers. As a result, high type consumers become less likely to substitute their home bank for an external lender, as they can benefit equally well from screening efficiencies at both. In contrast, the diversion effect is exacerbated for risky types: because low type borrowers are more likely to be rejected at in-network firms, they are more willing to continue searching until they encounter an out-of-network firm, where they are more likely to be approved for a loan.

This implies that, while the diversion effect may lead to an undesirable quantity reduction in in-network firm's own customer originations, it could result in a more favorable mix of borrowers due to the exclusion of risky (in-house) consumers—consistent with Fact 4 in Section 3.

Figure 10 – Borrower Selection into Application at Home Bank when Primary Bank is In and Out of Interoperable Network



Notes: in-house application probabilities for high-, Panel (a), and low-type, Panel (b), borrowers when home bank is in- and out-of interoperable network.

All in all: the net size, direction, and equilibrium implications of the countervailing application, attraction, and diversion effects will depend on the demand and supply model primitives—an estimation task which we tackle empirically in the next section.

5. Estimation

To make the estimation process computationally tractable, we introduce a few simplifying assumptions à la Agarwal et al. (2024) and parametrize the model as follows.

We assume that consumers prefer lower interest rates, r, and allow for the possibility of type-specific sorting into loans, which is governed by a selection parameter σ.[16] Furthermore, given the rise of non-bank lending during our period of study (2018-2021), we include a FinTech fixed effect, ξ, as a catch-all term for desirable non-bank attributes—including, e.g., faster application

[16] When $\sigma < 0$ (resp., $\sigma > 0$), we are in a world of adverse (resp., advantageous) selection: holding price fixed, high (resp., low) type consumers are more likely to submit an application and originate a mortgage, conditional on being approved.

processing (Fuster et al., 2019). Put together, we parameterize the indirect utility as

$$u_i(r_j, z_i, \xi) = -r_j + \sigma x_{z_i} + \mathbb{1}[s_j^d = 0]\xi.$$

Furthermore, we assume that the (residualized) offer rates are drawn independently from a normal distribution with mean μ_H and standard deviation σ_H,[17] while search costs are i.i.d. following a log-normal distribution centered at μ_G and scaled by σ_G. We also assume that borrowers default at a constant hazard rate. Specifically, for a loan maturing at time T, the probability of observing a non-defaulting borrower t periods after origination is given by

$$\Omega(t|z_i; T) = x_{z_i}^{t/T}.$$

Regarding firm parameters, we assume that idiosyncratic lender shocks are i.i.d. following a centered type-I extreme value distribution with scale factor σ_ε. Moreover, for tractability, we assume that there is no variation in lender costs: all firms incur the same origination cost of m.

Estimation proceeds in two steps. First, we employ a maximum likelihood procedure to recover demand-side parameters, $\boldsymbol{\theta}^D$. Given the estimated model primitives from the first stage, we then use a minimum distance approach to estimate the supply-side parameters, $\boldsymbol{\theta}^S$.

5.1 Demand Estimator

We search for a vector of demand-side parameters, $\boldsymbol{\theta}^D = (\boldsymbol{p}, \boldsymbol{x}, \boldsymbol{\beta}_H, \boldsymbol{\beta}_G, \boldsymbol{\kappa}, \lambda_h, \sigma, \tau)'$, that maximizes the model-implied likelihood of observing each realized mortgage origination, $\mathcal{O}$, and application, $\mathcal{A}$; that is:

$$\max_{\boldsymbol{\theta}^D} \sum_{i \in \mathcal{O}} \left[\log q(\boldsymbol{X}_i) + \log l^{\text{orig}}(R_{ij}, S_i, D_i | t, T; \boldsymbol{\theta}^D) \right] + \sum_{i \in \mathcal{A}} \left[\log[1 - q(\boldsymbol{X}_i)] + \log l^{\text{app}}(S_i | \text{Applied}; \boldsymbol{\theta}^D) \right],$$

where R_{ij} denotes the originated (residualized) interest rate at lender j; S_i labels the observed number of borrower applications; D_i is an indicator variable equal to 1 if borrower i defaulted in period t given loan maturity T and 0 otherwise; finally, $q(\boldsymbol{X}_i)$ is the estimated probability that borrower i, with attributes $\boldsymbol{X}_i$, is observed in the origination dataset. Following Agarwal et al. (2024), we treat the last term as a nuance adjustment. It is included to scale individual likelihood contributions based on observed borrower characteristics, placing more weight on consumers who are more likely to originate loans.

[17] Appendix Figure 13 shows that the distribution of residualized interest rates is approximately normal.

Originations. The likelihood of an origination interacts three components: the probability that borrower i (1) realizes D_i through observed period t given loan maturity T; (2) originates at rate R_{ij} with lender $j \in \mathcal{J}$; and (3) fails to originate a loan at each in-network and out-of-network firm, given the specific borrower-lender search history. We integrate over the distribution of objects that are unobservable to the econometrician: hidden borrower types, z; home banks, b; offer rate distribution, H; and reservation rates, r^*.

To make it as legible as possible, we state the likelihood below for an individual market, defined as a MSA-year pair. In practice, we allow the observed size of the interoperable network, γ, and the distribution of banking presence, s_j^d, to vary across MSAs and years.[18]

$$
\begin{aligned}
l^{\text{orig}}(R_{ij}, S_i, D_i | t, T; \boldsymbol{\theta}^D) = \sum_{z\in\{h,l\}} \lambda_z & \overbrace{\left(D_i(1-x_z^{t/T}) + (1-D_i)x_z^{t/T} \right)}^{\text{pr. } i \text{ realizes default } D_i} \\
\times \Bigg[(1-\gamma)p_z[s_j^{\text{data}}\kappa_z^{1-n_j} + (1-s_j^{\text{data}})]h(R_{ij}) & \int_{R_{ij}}^{\infty} \prod_{k\in\mathcal{S}_i\setminus\{j\}, n_k=1} [1-p_z H(r^*)] f_z^{00}(r^*) \\
\times \prod_{y\in\mathcal{S}_i\setminus\{j\}, n_y=0} & \left[s_y^{\text{data}}[1-p_z\kappa_z H(r^*)] + (1-s_y^{\text{data}})[1-p_z H(r^*)] \right] \left[s_y^{\text{data}} f_z^{01}(r^*) + (1-s_y^{\text{data}}) f_z^{00}(r^*) \right] dr^* \\
+ \gamma \underbrace{p_z\kappa_z^{n_j} h(R_{ij})}_{\text{pr. } i \text{ orig at rate } R_{ij}} \int_{R_{ij}}^{\infty} & \underbrace{\prod_{k\in\mathcal{S}_i\setminus\{j\}, n_k=1} [1-p_z\kappa_z H(r^*)]\, f_z^{11}(r^*)}_{\text{pr. failed orig. at in-network lenders}} \underbrace{\prod_{y\in\mathcal{S}_i\setminus\{j\}, n_y=0} [1-p_z H(r^*)]}_{\text{pr. failed orig. at out-of-network lenders}} f_z^{10}(r^*) dr^* \Bigg],
\end{aligned}
$$

where $\mathcal{S}_i$ denotes the search set, and $f_z^{n_{b_i}, m(j,b_i)}(r^*)$ are the densities of the reservation rate distributions.

Applications. The likelihood contribution of applications (which include both loan originations and non-originations), follows a structure similar to that of originations. Except that here we match the probability of realizing a given number of inquiries, conditional on observing an application. That is,

[18] For brevity, we omit from the above likelihood description the terms that account for differential treatment of non-banks. In model estimation and counterfactuals, we allow for separate reservation rates when borrowers are matched with FinTechs. This expands the number of borrower-lender match cases in Section 4.2 from four to eight. We also incorporate FinTech presence into the determination of all threshold policies.

$$
\begin{aligned}
l^{\text{app}}(S_i|\text{Applied};\boldsymbol{\theta}^D) &= \overbrace{\left[\textstyle\sum_z \lambda_z/p_z\right]^{-1}}^{1/(\text{pr. applied})} \sum_{z_i\in\{h,l\}} \lambda_z \\
&\times \Bigg[(1-\gamma)\int H(r^*) \prod_{k\in\mathcal{S}_i,n_k=1} [1-p_z H(r^*)] f_z^{00}(r^*) \\
&\qquad \times \prod_{y\in\mathcal{S}_i,n_y=0} \left[s_y^{\text{data}}[1-p_z\kappa_z H(r^*)] + (1-s_y^{\text{data}})[1-p_z H(r^*)]\right]\left[s_y^{\text{data}} f_z^{01}(r^*) + (1-s_y^{\text{data}}) f_z^{00}(r^*)\right] dr^* \\
&\qquad + \gamma \int \underbrace{H(r^*)}_{\text{pr. applied}} \underbrace{\prod_{k\in\mathcal{S}_i,n_k=1} [1-p_z\kappa_z H(r^*)]}_{\text{pr. failed orig. at in-network lenders}} f_z^{11}(r^*) \underbrace{\prod_{y\in\mathcal{S}_i,n_y=0} [1-p_z H(r^*)]}_{\text{pr. failed orig. at out-of-network lenders}} f_z^{10}(r^*) dr^* \Bigg].
\end{aligned}
$$

5.2 Supply Estimator

To estimate the supply-side model parameters,

$$
\boldsymbol{\theta}^S = (m, \sigma_\varepsilon)',
$$

we minimize the distance between the parameters governing the offer rate distribution estimated in the first stage, $\hat{\boldsymbol{\beta}}_H = (\hat{\mu}_H, \hat{\sigma}_H)$, and those implied by the CCPs from the supply model, given a choice of marginal cost, m, and standard deviation of the firm-rate idiosyncratic shock, σ_ε. Namely,

$$
\max_{\boldsymbol{\theta}^S} \left(\boldsymbol{\beta}_H(\boldsymbol{\theta}^S) - \hat{\boldsymbol{\beta}}_H^D\right)' \widehat{\boldsymbol{W}} \left(\boldsymbol{\beta}_H(\boldsymbol{\theta}^S) - \hat{\boldsymbol{\beta}}_H^D\right)
$$

Our procedure can be summarized as follows. First, we guess an initial value for $\boldsymbol{\theta}^S$. Next, we compute the supply model-implied CCPs using this guess:

$$
\mathbb{P}[j \text{ chooses } r_k|\boldsymbol{\theta}^S] = \frac{\exp\left(\mathbb{E}\left[\Pi_j(r_k; , m)\right]/\sigma_\varepsilon\right)}{\sum_{a=1}^K \exp\left(\mathbb{E}\left[\Pi_j(r_a; m)\right]/\sigma_\varepsilon\right)}
$$

with

$$
\mathbb{E}[\Pi_j(r_k; m)] = \sum_g M_g \sum_{z\in\{h,l\}} \lambda_z q_{j,z}^{n_j}(r_k)\left[r_k\tilde{x}_z - m\right].
$$

Finally, we compute the distance between the resulting $\boldsymbol{\beta}_H(\boldsymbol{\theta}^S)$ and demand model estimates $\hat{\boldsymbol{\beta}}_H^D$; and iterate until convergence. In practice, we set the weighting matrix $\widehat{\boldsymbol{W}}$ to the identity matrix. This translates to minimizing the sum of squared differences between the mean and standard deviation of the estimated demand and supply model-implied offer rate distributions.

5.3 Identification

As in Agarwal et al. (2024), identification of the model parameters is largely facilitated by the relationships between realized prices, ex-post default, and borrower search activity—including Facts 1–2 in Section 3.1—and the differences in the likelihood contributions of originations and applications. In what follows, we provide intuition behind identification of parameters that are unique to our framework; namely, transaction costs, τ, and type-specific screening boosts, κ_z.

Transaction costs are identified by the differences in application rates between FDX and non-FDX firms (Fact 3 in Section 3.2), as well as in-house vis-à-vis external lending—while exploiting differences in these differences across MSAs and years due, to plausibly exogenous, variation in lender and interoperable network banking presence. Any nonzero valued transaction costs will drive a change in model-implied distribution of reservation rates. This, in turn, impacts borrower selection and the predicted gap in default rates between in-network and out-of-network firms (for visual intuition, see Figures 9–10). Identification of screening improvement parameters then relies on the differences in these model predicted default rates and observed differences (Fact 3 in Section 3.2). The corresponding wedge pins down the estimates of the screening boosts.

5.4 Results

Table 3 presents our parameter estimates.[19] Column (1) reports results from a benchmark model, which excludes transaction costs and screening boost. In the intermediate model in Column (2), we add transaction costs. Finally, in Column (3), we show estimates of the baseline specification, which incorporates both transaction costs and screening boosts.

Focusing on our preferred specification in Column (3), we find that high-type borrowers account for the vast majority, 96%, of consumers in our analysis sample. This is not surprising given the composition of mortgages in our dataset, which consists of relatively homogeneous and low-risk loans (backed by Freddie Mac, a government-sponsored entity). High-type borrowers are likely to repay their loan, while low-type consumers rarely pay back. The size of the estimated selection parameter indicates a strong degree of adverse selection. Borrowers value FinTechs, as evidenced by the magnitude of the non-bank fixed effect relative to the mean offer rate.

Transaction costs are economically meaningful: they amount to 35% of average search costs. Screening is informative, as shown by the sizeable difference in approval rates of high- and low-type borrowers—even in the absence of data-sharing. But it becomes particularly accurate with

[19] In Appendix Table 2, we compare our benchmark model estimates with those of Agarwal et al. (2024).

interoperability, which improves screening by 58% and 70% for high- and low-type borrowers, respectively. Finally, we estimate lender profit margin at 1.61%. This is in line with the 1.77% average rate on 10-year U.S. Treasury securities (2018-2021) and the estimates found in other research (e.g., Agarwal et al., 2024; Allen et al., 2014).

Model fit. Appendix Section B.1 compares key statistics in the data against those generated by each of our models in Columns (1)–(3) of Table 3. Relative to the benchmark and intermediate models, our preferred baseline model produces data patterns that more closely align with those observed in the analysis sample. It performs particularly well in matching origination rates, search distribution, and the relationship between prices and search activity. It also qualitatively replicates differences in applications and default rates between in- and out-of-network lenders.

Table 3 – Parameter Estimates

		Benchmark	**Intermediate**	**Baseline**
		(1)	(2)	(3)
Demand Model Parameters				
Share of high type borr.	λ_h	0.956	0.961	0.964
Repay. prob. of high type borr.	x_h	0.783	0.773	0.924
Repay. prob. of low type borr.	x_l	0.357	0.306	0.001
Approval prob. of high type borr.	p_h	0.555	0.580	0.631
Approval prob. of low type borr.	p_l	0.041	0.038	0.008
Mean of (log) search cost distr.	μ_G	-1.708	-1.341	-1.464
Std. dev. of (log) search cost distr.	σ_G	0.352	0.237	0.086
Mean of (resid.) offer rate distr.	μ_H	-0.089	-0.068	0.011
Std. dev. of (resid.) offer rate distr.	σ_H	0.326	0.320	0.287
Selection parameter	σ	–	-0.246	-9.470
Non-bank fixed effect	ξ	–	0.175	0.040
Transaction cost	τ	–	0.045	0.082
Screening boost for high type borr.	κ_h	–	–	1.584
Screening boost for low type borr.	κ_l	–	–	0.300
Supply Model Parameters				
Origination cost	m	1.645	1.643	1.607
Std. dev. of idiosync. shock	σ_ε	0.431	0.440	0.473
Interoperability & transaction costs		–	✔	✔
Screening boost w/ interoperability		–	–	✔

Notes: parameter estimates from preferred baseline model with both transaction costs and screening boost in Column (3); intermediate model with only transaction costs in Column (2); and benchmark model with neither in Column (1).

6. Counterfactuals

With the estimated model in hand, we now turn to evaluating the aggregate and distributional effects of alternative data-sharing regimes. We explore three scenarios.[20]

No interoperability. As a point of reference for other counterfactuals, we naturally consider the case in which there are no efficiencies stemming from data-sharing. Under this "no interoperability" regime: we (1) assume that all applications outside of home bank involve transaction costs: $\tau > 0$; and (2) shut down non in-house screening efficiencies: $\tilde{\kappa}_h = \tilde{\kappa}_l = 1$.

Full Interoperability. At the opposite end of the spectrum, we study the case of "full interoperability." In this regime, we mandate that all firms share their consumer data. As a result, applications are cost-free. We also assume that all in-network firms are capable of fully exploiting screening efficiencies: $\tilde{\kappa}_l^j = \kappa_l$ and $\tilde{\kappa}_h^j = \kappa_h$.

Full Interoperability with Screening Heterogeneity. Finally, we examine a special case of full interoperability in which firms vary in their ability to internalize screening efficiencies. In this "full interoperability with screening heterogeneity" counterfactual, we partition firms into two equally sized groups: high, $\mathcal{F}_h$, and low-technology, $\mathcal{F}_l$, lenders. More tech-savvy firms are capable of fully internalizing the screening efficiencies from data-sharing. In contrast, we assume that low IT firms only partially reap the benefits of interoperability (e.g., due to technological constraints). We model this heterogeneity through a parameter w, which scales the screening boost of low-tech firms such that, for all $j \in \mathcal{F}_l$,

$$\tilde{\kappa}_h^j = (1 + \kappa_h)w \qquad \text{and} \qquad \tilde{\kappa}_l^j = (1 + \kappa_l)w.$$

In simulations, we let low IT firms be half as capable of integrating screening boost: $w = 0.5$.

As the main basis for comparison, we focus on welfare and investigate the drivers of observed distributional differences across borrower and lender types—including the roles of selection and pricing. For borrowers, we define individual consumer surplus as the difference between the match-specific willingness to pay (reservation rate, r^*_{ij}) and the realized interest rate, R_{ij}. For firms, producer surplus is defined using ex-post profit: the product of realized interest rate, R_{ij} and default, D_i, net of screening/origination costs, m. Aggregating across consumers and firms, total welfare equals:

$$\mathcal{W} = \underbrace{\sum_{i \in \mathcal{I}} \Big(r^*_{ij} - R_{ij} \Big)}_{\text{consumer surplus}} + \underbrace{\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \Big(R_{ij} D_i - m \Big)}_{\text{producer surplus}}$$

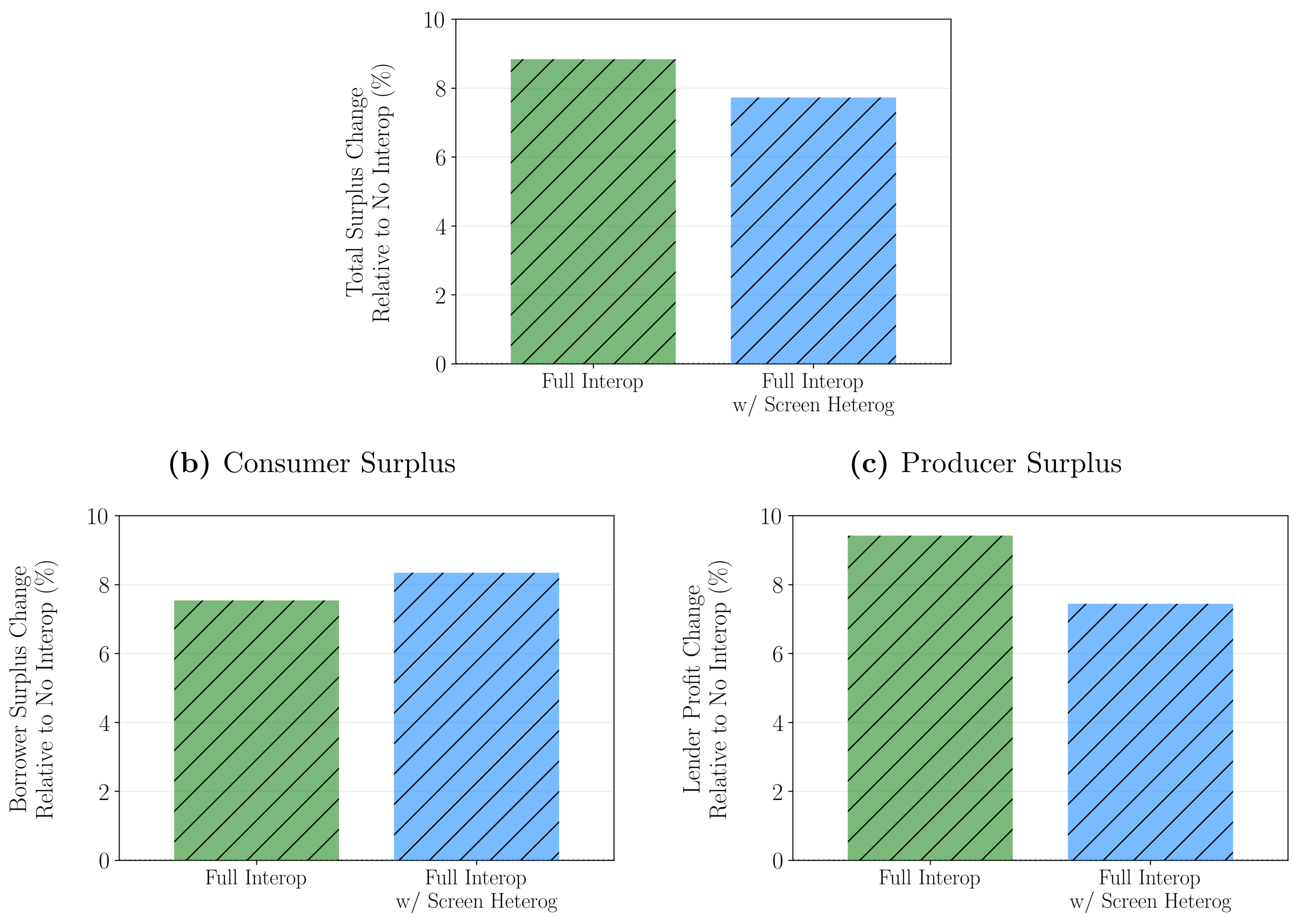
[20] Appendix Section B.3 provides an extension by further evaluating the equilibrium results under the intermediate case of an interoperable network.

6.1 Results

Overall, we find that interoperability is welfare improving. As shown in Figure 11, relative to the no interoperability regime, the full interoperability case without (resp., with) screening heterogeneity leads to an 8.8% (resp., 7.7%) increase in total surplus. Under uniform (resp., heterogeneous) screening improvement, the benefits from interoperability are tilted toward firms (resp., consumers). Intuitively, screening efficiency empowers lenders in approving more desirable low-risk borrowers. This, in turn, leads to higher ex-post profits due to lower default rate within the origination pool.

Figure 11 – Welfare Change Relative to the No Interoperability Regime

(a) Total Surplus

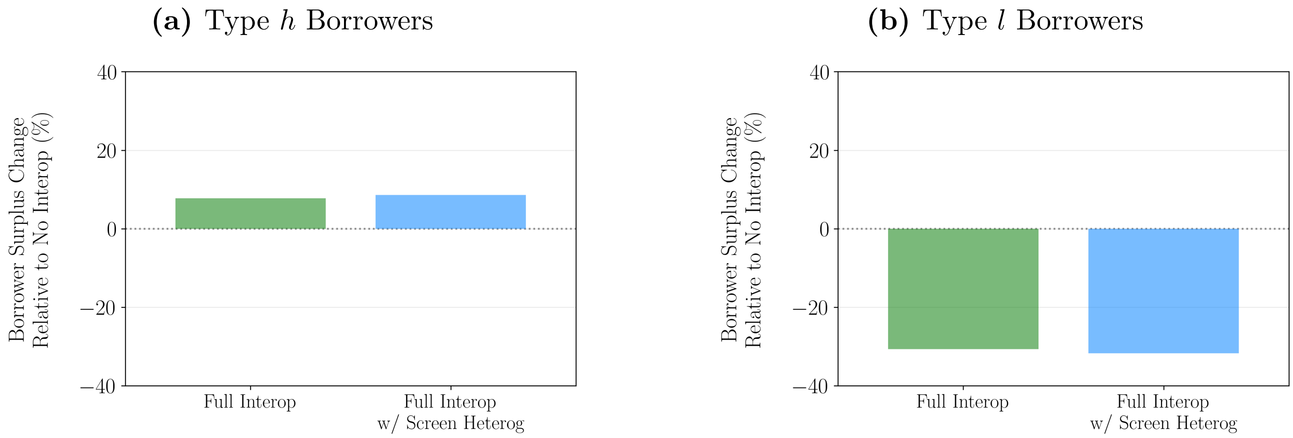


Notes: percentage difference in welfare, relative to the no interoperability regime, in total, Panel (a), consumer, Panel (b), and producer, Panel (c), welfare for the full interoperability regimes with and without heterogeneity in screening boost.

While total surplus increases in the aggregate, we find that—in our setting—the distribution of welfare effects across consumer and firm types is uneven.

Consumers. Starting with borrowers, Figure 12 shows that while high-type borrowers, who account for the majority of the consumer base, benefit from interoperability, low-types are hurt.

Figure 12 – Borrower Welfare Change Relative to the No Interoperability Regime

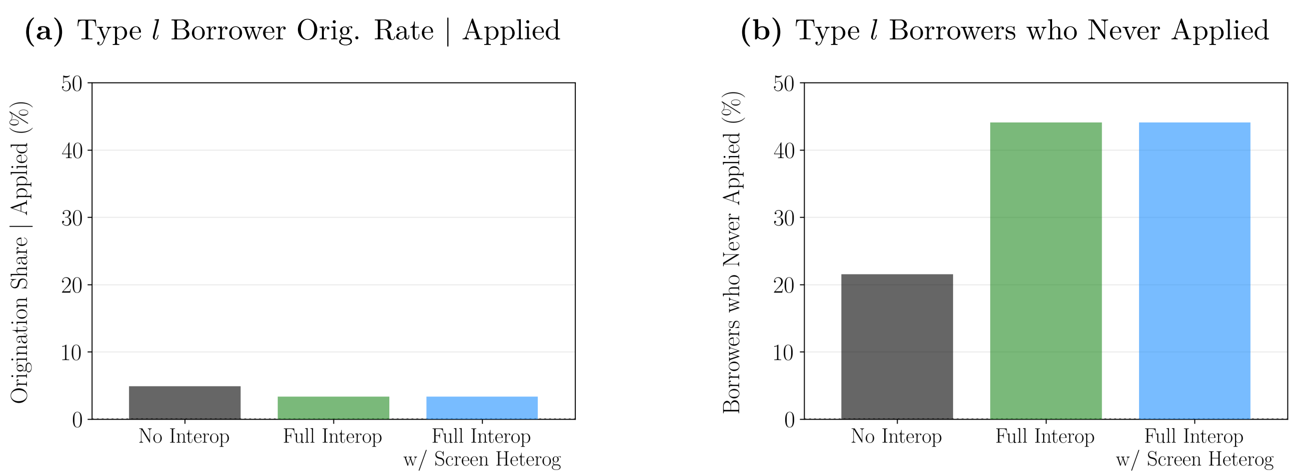


Notes: percentage difference in high, Panel (a), and low type, Panel (b) borrower surplus, when moving from the no interoperability to full interoperability regime with and without screening heterogeneity.

To unpack the main sources of the decline in surplus for low type borrowers, we examine (1) the influence of improved screening, Panel (a) in Figure 13; and (2) changes in borrower selection into application, Panel (b), as we transition away from the no-interoperability regime.

Conditional on applying, low type borrowers are less likely to originate loans—a direct result of enhanced screening. We also find that much of the exclusion of low-type borrowers is driven by the indirect effect of interoperability on borrower selection. Specifically, with data-sharing in place, more than twice as many low type borrowers opt out entirely from seeking credit.

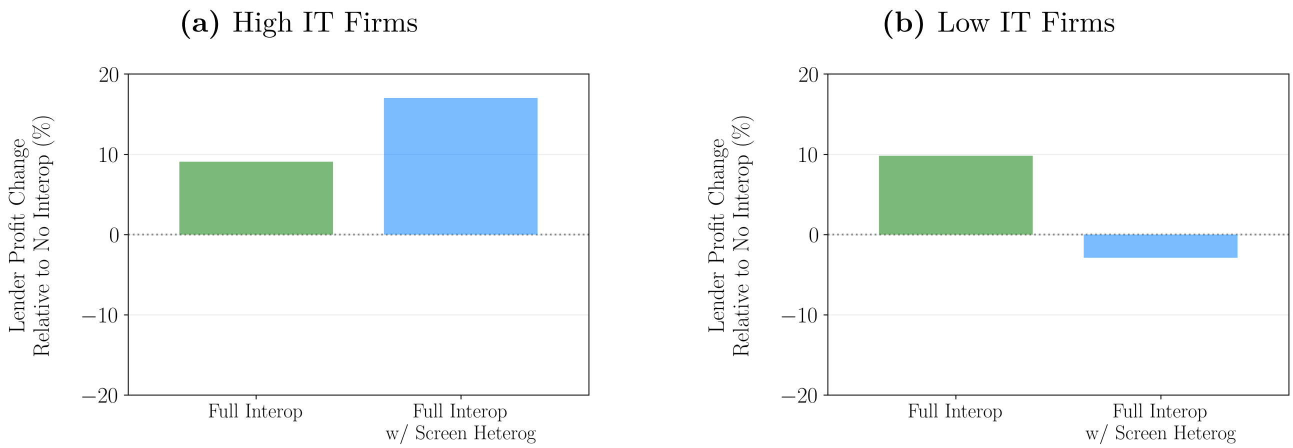
Figure 13 – Determinants of Risky Borrowers' Welfare Loss



Notes: share of low type borrowers, who originate loans, Column (a), and never apply, Column (b).

Firms. Switching gears, we now explore the distributional effects on lenders. When firms share the same screening technology, data-sharing increases profits uniformly. But, in the presence of screening heterogeneity, welfare rises for high- and declines for low-IT firms (Figure 14).

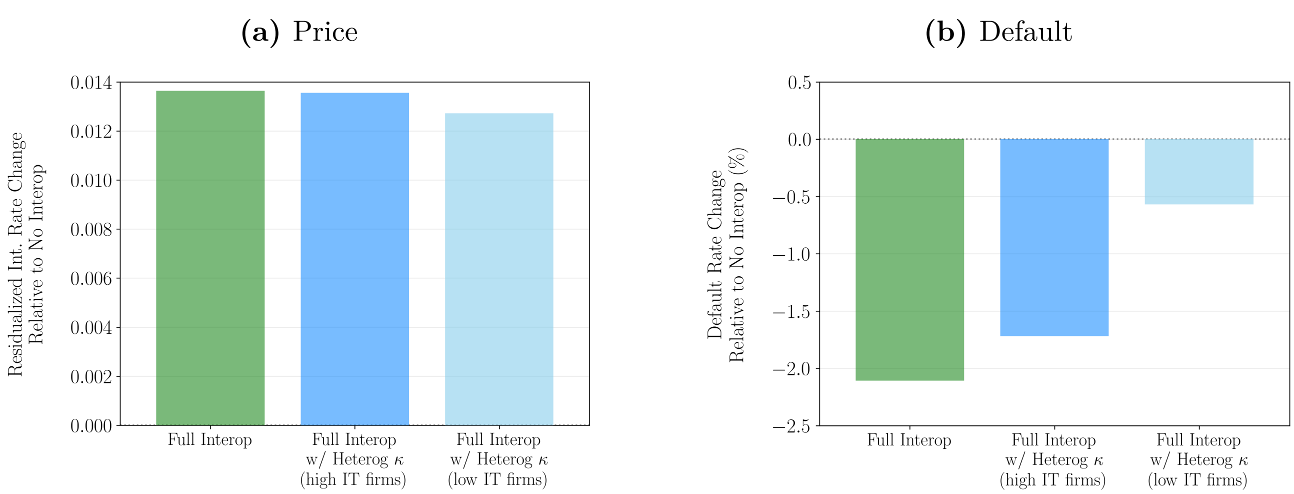
Figure 14 – Lender Welfare Change Relative to the No Interoperability Regime



Notes: percentage difference in high, Panel (a), and low IT, Panel (b) firm profits, when moving from the no interoperability to full interoperability regime with and without screening heterogeneity.

To uncover the determinants of lower profits for less tech-savvy firms, recall that screening accuracy is valuable to high-type borrowers since it increases their likelihood of being approved (Figures 9–10). This, in turn, reduces the ability of low IT firms to extract information rents from low-risk applicants, as indicated by a smaller increase in prices—relative to high IT firms—in Panel (a) of Figure 15. The direct effect of having an inferior screening technology further hurts the composition of low IT firms' origination pool, Panel (b), and, hence, their profits.

Figure 15 – Determinants of Low IT Firms' Welfare Loss



Notes: percentage difference—relative to no interoperability—in interest rates, Column (a), and default rates, Column (b), for the full interoperability regime with and without screening heterogeneity.

7. Conclusion

We use a novel borrower-lender search dataset and leverage the entry of an industry-led interoperable network to study the distributional effects of data interoperability on consumers and firms in the context of the U.S. residential mortgage market between 2018 and 2021.

Descriptive analysis indicates that lenders within the interoperable data network experience a relative rise in applications and decline in ex-post default, while maintaining similar prices compared to other firms. This suggests two types of data-sharing induced efficiencies: reduced transaction costs for borrowers applying to in-network firms; and screening accuracy improvements for in-network lenders due to better information access.

We develop and estimate a model of borrower search and lender pricing, which embeds these economic forces. Estimated additional costs of applying to out-of-network firms are 35% of average search costs. Interoperability improves screening accuracy by 58% and 70% for consumers who are more- and less-likely to repay their loans, respectively.

Counterfactual simulations reveal that an industry-wide data-sharing mandate is overall welfare improving. However, the distribution of equilibrium effects may be uneven. In our setting, we find that consumer information-sharing: (1) alleviates borrower lock-in effects through a cutback on transaction frictions; (2) creates screening efficiency gains, which primarily benefit less risky borrowers and more tech-savvy firms; (3) hurts risky consumers, who strategically opt out from seeking credit or are screened out in the application process; and (4) can generate profit losses for less tech-savvy firms; and potentially contribute to increased lender concentration, if the cost of integrating data-sharing technology is high enough to incentivize market exit.

Future research in this area could benefit from endogenizing firm entry, exit, and strategic network formation to examine the short- and long-term effects of interoperability on market structure—specifically, in the presence of heterogeneity in firms' technological capacity to internalize data-sharing efficiencies. Another promising avenue for further work could entail the analysis of data-sharing effects on multi-service firms, especially in relation to open finance—an initiative that broadens the concept of open banking through the horizontal and vertical integration of consumer data to a wider spectrum of financial services and technology firms.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022): "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 14, 218–256.

Agarwal, S., J. Grigsby, A. Hortaçsu, G. Matvos, A. Seru, and V. Yao (2024): "Searching for approval," *Econometrica*, 92, 1195–1231.

Allen, J., R. Clark, and J.-F. Houde (2014): "The effect of mergers in search markets: Evidence from the Canadian mortgage industry," *American Economic Review*, 104, 3365–3396.

——— (2019): "Search frictions and market power in negotiated-price markets," *Journal of Political Economy*, 127, 1550–1598.

Babina, T., S. A. Bahaj, G. Buchak, F. De Marco, A. K. Foulis, W. Gornall, F. Mazzola, and T. Yu (2024): "Customer data access and fintech entry: Early evidence from open banking," *Journal of Financial Economics*, Forthcoming.

Blattner, L., J. Hartwig, and S. Nelson (2022): "Information design in consumer credit markets," Working Paper.

Brown, M., T. Jappelli, and M. Pagano (2009): "Information sharing and credit: Firm-level evidence from transition countries," *Journal of Financial Intermediation*, 18, 151–172.

De Corniere, A. and G. Taylor (2024): "Data and competition: A simple framework," *RAND Journal of Economics*, Forthcoming.

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

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

Goodman, L. S., B. Bai, and W. Li (2019): "Real denial rates: A new tool to look at who is receiving mortgage credit," *Housing Policy Debate*, 29, 795–819.

He, Z., J. Huang, and J. Zhou (2023): "Open banking: Credit market competition when borrowers own the data," *Journal of Financial Economics*, 147, 449–474.

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

Jaffee, D. M. and T. Russell (1976): "Imperfect information, uncertainty, and credit rationing," *The Quarterly Journal of Economics*, 90, 651–666.

Jones, C. I. and C. Tonetti (2020): "Nonrivalry and the Economics of Data," *American Economic Review*, 110, 2819–2858.

Kim, M., D. Kliger, and B. Vale (2003): "Estimating switching costs: The case of banking," *Journal of Financial Intermediation*, 12, 25–56.

Klemperer, P. (1995): "Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade," *The Review of Economic Studies*, 62, 515–539.

Liu, C. (2023): "Language frictions in consumer credit," Working Paper.

Pagano, M. and T. Jappelli (1993): "Information sharing in credit markets," *The Journal of Finance*, 48, 1693–1718.

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

Sharpe, S. A. (1990): "Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships," *The Journal of Finance*, 45, 1069–1087.

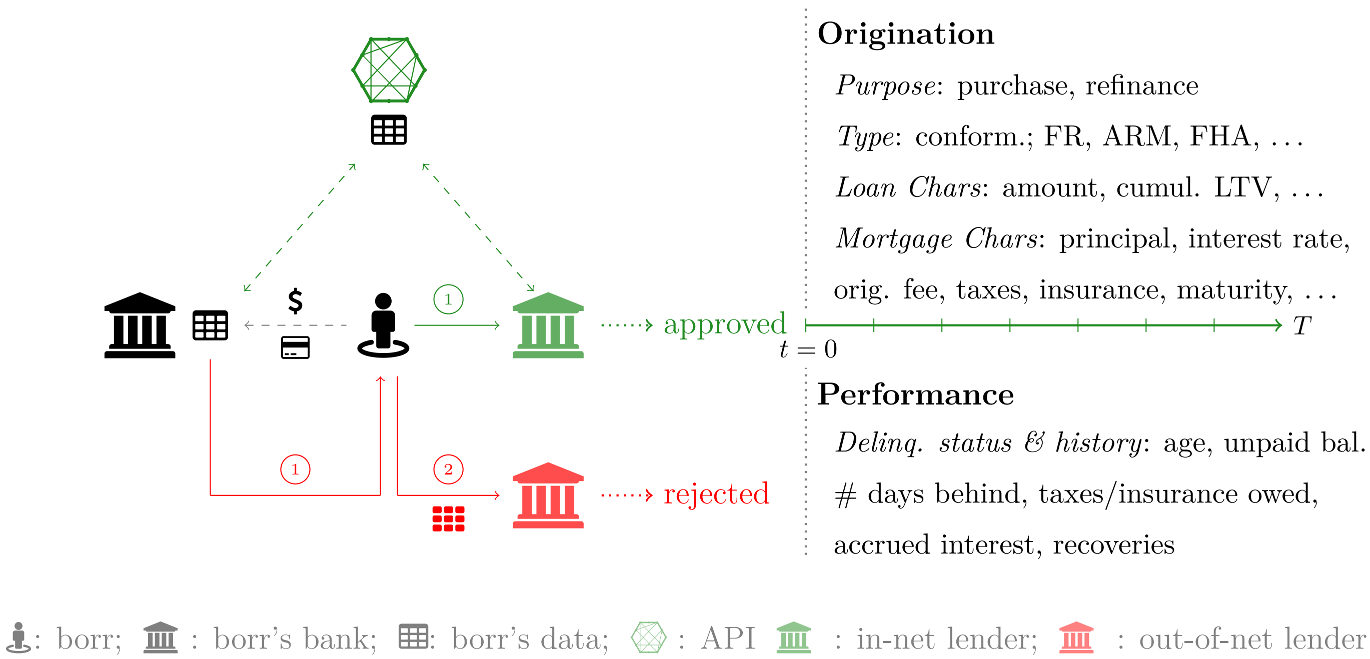
Stiglitz, J. E. and A. Weiss (1981): "Credit rationing in markets with imperfect information," *American Economic Review*, 71, 393–410.

Yannelis, C. and A. L. Zhang (2023): "Competition and selection in credit markets," *Journal of Financial Economics*, 150, 103710.

Yin, X. (2024): "The effects of big data on commercial banks," Working Paper.

Appendix

Appendix Figure 1 – Mortgage Search, Application, and Origination



Notes: illustration of the mortgage search, application, and origination process at an in-house or interoperable lender (shown in green) versus an external lender (in red). With interoperability, a borrower's financial transaction history can be readily transmitted via API (top icon in the diagram); as a result, the application process is simplified and only requires the consumer to submit a loan application. In contrast, without interoperability, the customer needs to additionally retrieve financial data from their primary bank, standardize the information, and submit it to the potential lender. This creates potential extra costs to borrowers when applying to out-of-network lenders. The right panel presents examples of origination and performance data typically tracked for realized mortgages.

Return: Section 2 (Background).

A. Additional Data & Descriptive Patterns

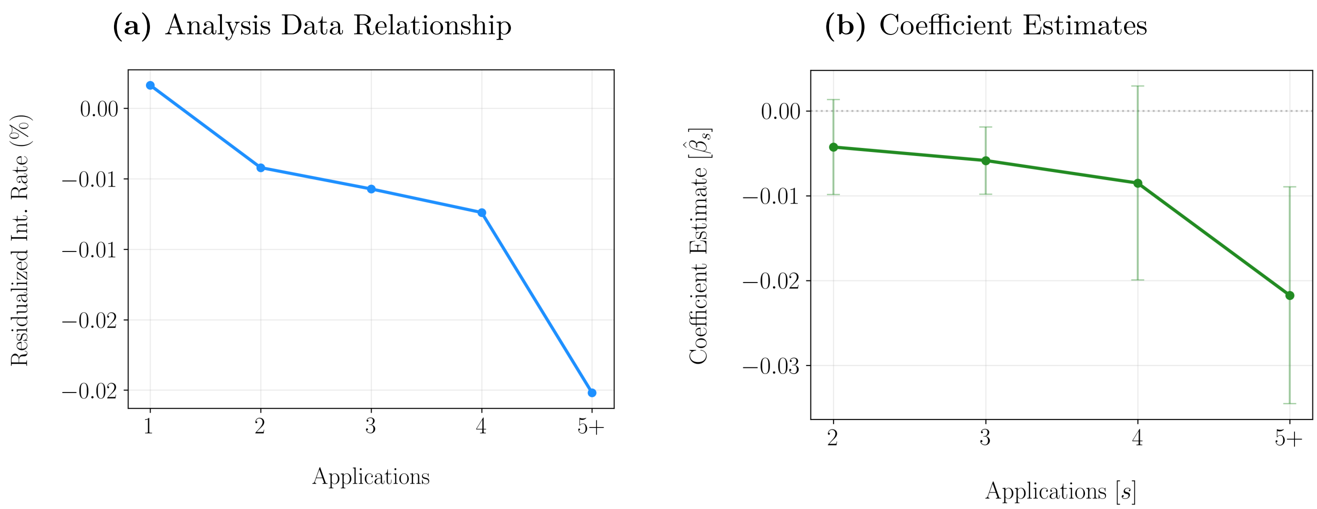
Appendix Table 1 – Origination Subsample (Borrower Characteristics)

	Origination Lender			
	FDX Deposit Banks		Other Deposit Banks	
Statistic	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Income ($000s)	108.4	56.2	103.7	53.4
Loan amount ($000s)	260.3	120.1	242.1	111.1
Combined loan-to-value ratio	71.9	17.2	76.0	15.8
Debt-to-income ratio	35.1	8.4	35.2	8.3
Credit score	757.1	178.4	754.6	151.1
Borrowers	**62,763 (38%)**		**101,701 (62%)**	

Notes: borrower summary statistics of the analysis sample for loans originated at FDX depository institutions, Columns (1)–(2), and all other banks, Columns (3)–(4).

Return: Section 2.2 (Data, Analysis Sample).

Appendix Figure 2 – Residualized Interest Rate and Applications

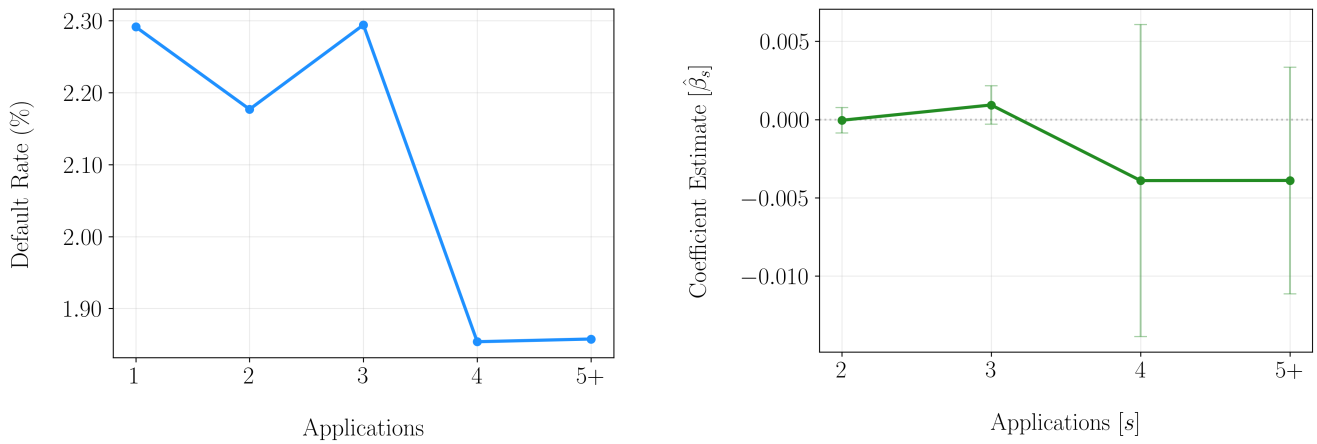


Notes: association between residualized interest rates and applications. Panel (a) plots the average res. int. rate at each count of applications. Panel (b) shows coefficient estimates from the OLS regression in Equation (1) with res. int. rate as the outcome variable. Adj. R^2 = 0.0068; obs. = 714,629.

Return: Section 3.1 (Descriptives, General Patterns).

Appendix Figure 3 – Default and Applications

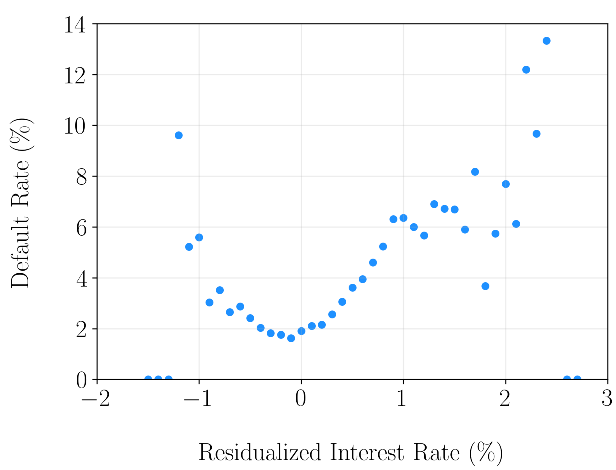
(a) Analysis Data Relationship **(b)** OLS Coefficient Estimates



Notes: association between default rates and applications. Panel (a) plots the average default rate at each count of applications. Panel (b) shows coefficient estimates from the OLS regression in Equation (1) with the outcome variable specified as a dummy equal 1 if default is realized; and 0 otherwise. Adj. $R^2 = 0.021$; obs. = 714,629.

Appendix Figure 4 – Default and Residualized Interest Rate

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

	Default Rate (%)
Resid. Int. Rate (%)	1.43**
	(0.25)
Resid. Int. Rate^2 (%)	1.19***
	(0.00)
Year, MSA, Lender FEs	Yes
Borrower Char. Controls	No
Avg. Outcome	2.26
Adj. R^2	0.018
Observations	714, 629

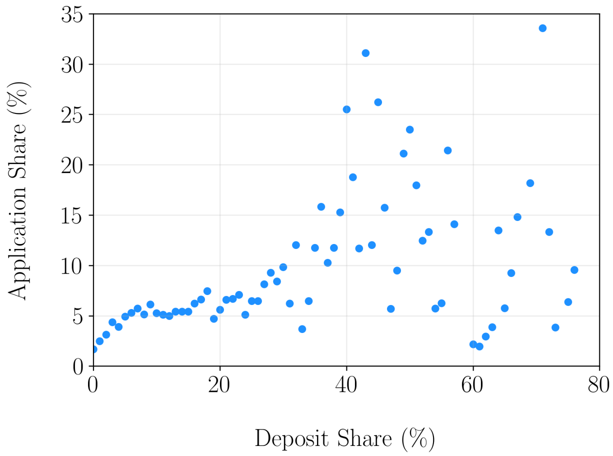
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: association between default and residualized interest rates. Panel (a) plots the average default rate at each res. int. rate. Panel (b) shows coefficient estimates from the OLS regression in Equation (2), but using res. int. rate (and a squared term) instead.

Return: Section 3.1 (Descriptives, General Patterns).

Appendix Figure 5 – Applications and Bank Deposits

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

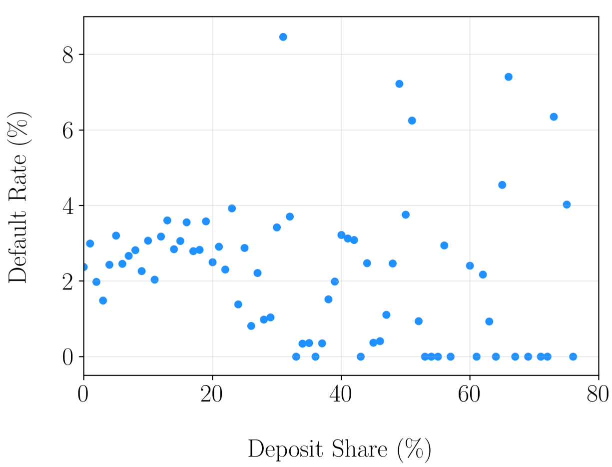
	Apps	Apps Share (%)
Deposit Share (%)	1.55**	0.20***
	(0.31)	(0.02)
Year, MSA, Lender FEs	✔	✔
Avg. Outcome	21.28	2.84
Adj. R^2	0.17	0.49
Observations	19,770	19,770

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: association between application and deposit shares. Panel (a) plots application shares against deposit shares. Panel (b) shows coefficient estimates from an OLS regression at lender-year-MSA level, with outcome variable specified as either application share or count; deposit share as the main covariate; and year, MSA, lender fixed effects.

Appendix Figure 6 – Default and Bank Deposits

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

	Default Rate (%)
Deposit Share (%)	−0.009
	(0.007)
Year, MSA, Lender FEs	✔
Avg. Outcome	2.445
Adj. R^2	0.022
Observations	19,770

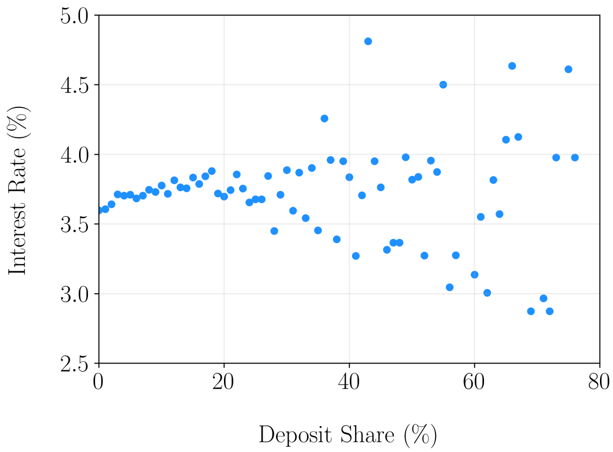
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: association between default rates and deposit shares. Panel (a) plots default rates against deposit shares. Panel (b) shows coefficient estimates from an OLS regression at lender-year-MSA level, with outcome variable specified as the average default rate; deposit share as the main covariate; and year, MSA, lender fixed effects.

Return: Section 3.1 (Descriptives, General Patterns).

Appendix Figure 7 – Interest Rate and Bank Deposits

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

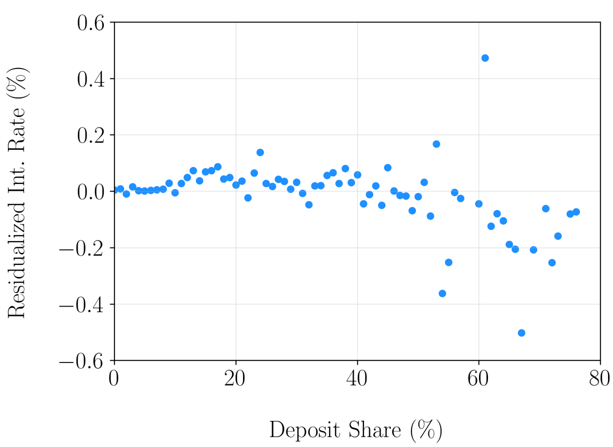
	Interest Rate (%)
Deposit Share (%)	0.0011**
	(0.0003)
Year, MSA, Lender FEs	✔
Avg. Outcome	3.6369
Adj. R^2	0.8137
Observations	19,770

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: association between interest rates and deposit shares. Panel (a) plots interest rates against deposit shares. Panel (b) shows coefficient estimates from an OLS regression at lender-year-MSA level, with outcome variable specified as the average interest rate; deposit share as the main covariate; and year, MSA, lender fixed effects.

Appendix Figure 8 – Residualized Interest Rate and Bank Deposits

(a) Analysis Data Relationship



(b) OLS Coefficient Estimates

	Resid. Int. Rate (%)
Deposit Share (%)	0.0010***
	(0.0002)
Year, MSA, Lender FEs	✔
Avg. Outcome	3.6369
Adj. R^2	0.0273
Observations	19,770

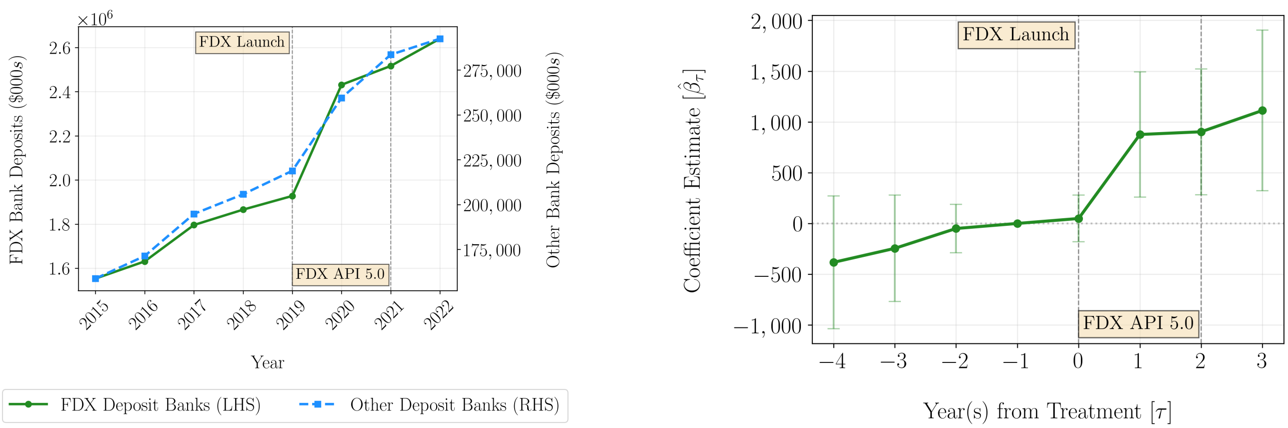
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Notes: association between residualized interest rates and deposit shares. Panel (a) plots resid. int. rates against deposit shares. Panel (b) shows coefficient estimates from an OLS regression at lender-year-MSA level, with outcome variable specified as the average resid. int. rate; deposit share as the main covariate; and year, MSA, lender fixed effects.

Return: Section 3.1 (Descriptives, General Patterns).

Appendix Figure 9 – Deposits: FDX and Other Banks

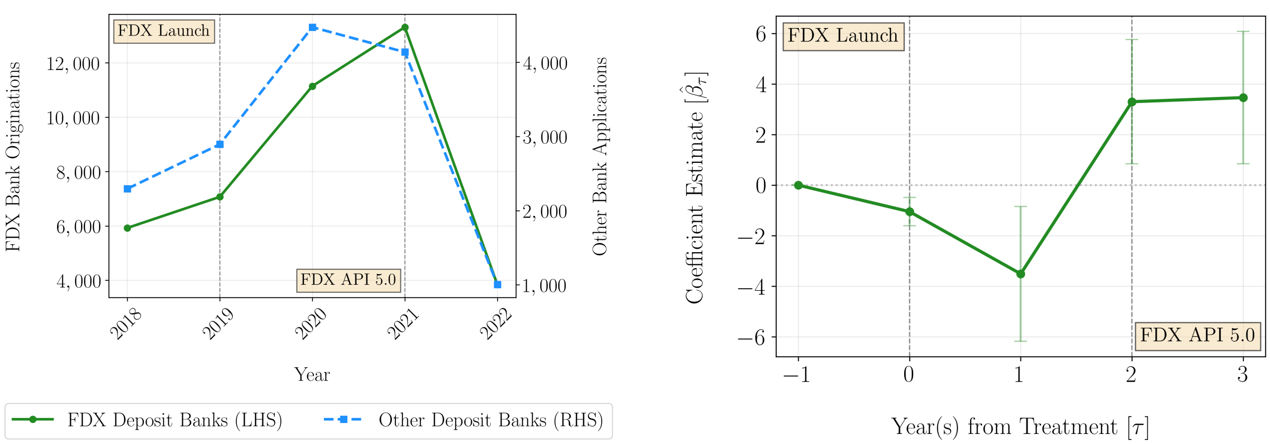
(a) Analysis Data Relationship   **(b)** Event Study Coefficient Estimates



Notes: differences in deposits between FDX and other banks. Panel (a) plots deposits in thousands of dollars. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with deposits as the outcome variable. Adj. $R^2 = 0.41$; obs. = 4,857; avg. deposits at FDX lenders in 2018 ($000s) = 4,056.

Appendix Figure 10 – Originations: FDX and Other Banks

(a) Analysis Data Relationship   **(b)** Event Study Coefficient Estimates

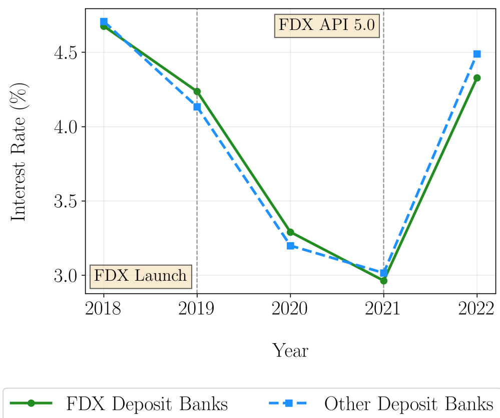


Notes: differences in originations between FDX and other banks. Panel (a) plots the number of originations. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with the count of originations as the outcome variable. Adj. $R^2 = 0.52$; obs. = 3,060; avg. originations at FDX lenders in 2018 = 13.

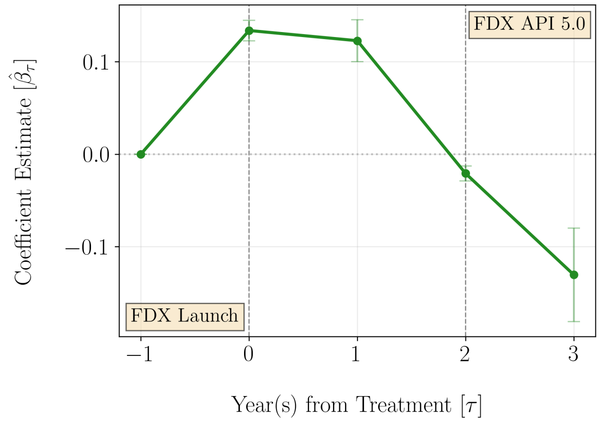
Return: Section 3.2 (Descriptives, FDX Patterns).

Appendix Figure 11 – Interest Rate: FDX and Other Banks

(a) Analysis Data Relationship



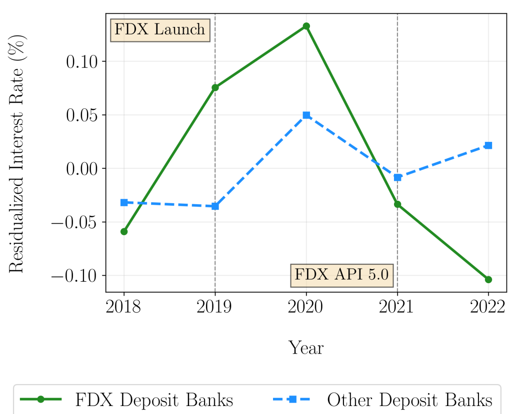
(b) Event Study Coefficient Estimates



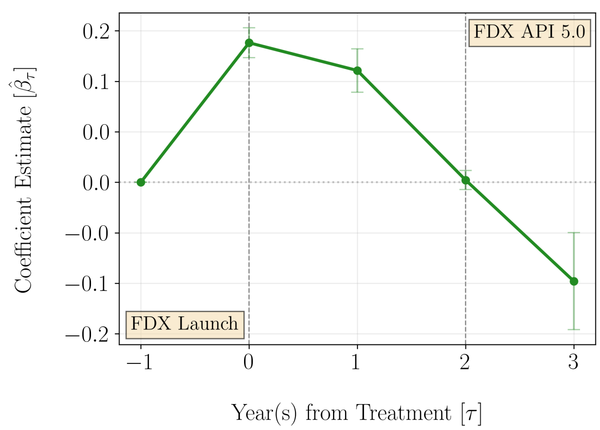
Notes: differences in interest rates between FDX and other banks. Panel (a) plots average interest rates. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with interest rate as the outcome variable. Adj. $R^2 = 0.80$; obs. = 3,060; avg. int. rate at FDX lenders in 2018 (%) = 5.0.

Appendix Figure 12 – Residualized Interest Rate: FDX and Other Banks

(a) Analysis Data Relationship



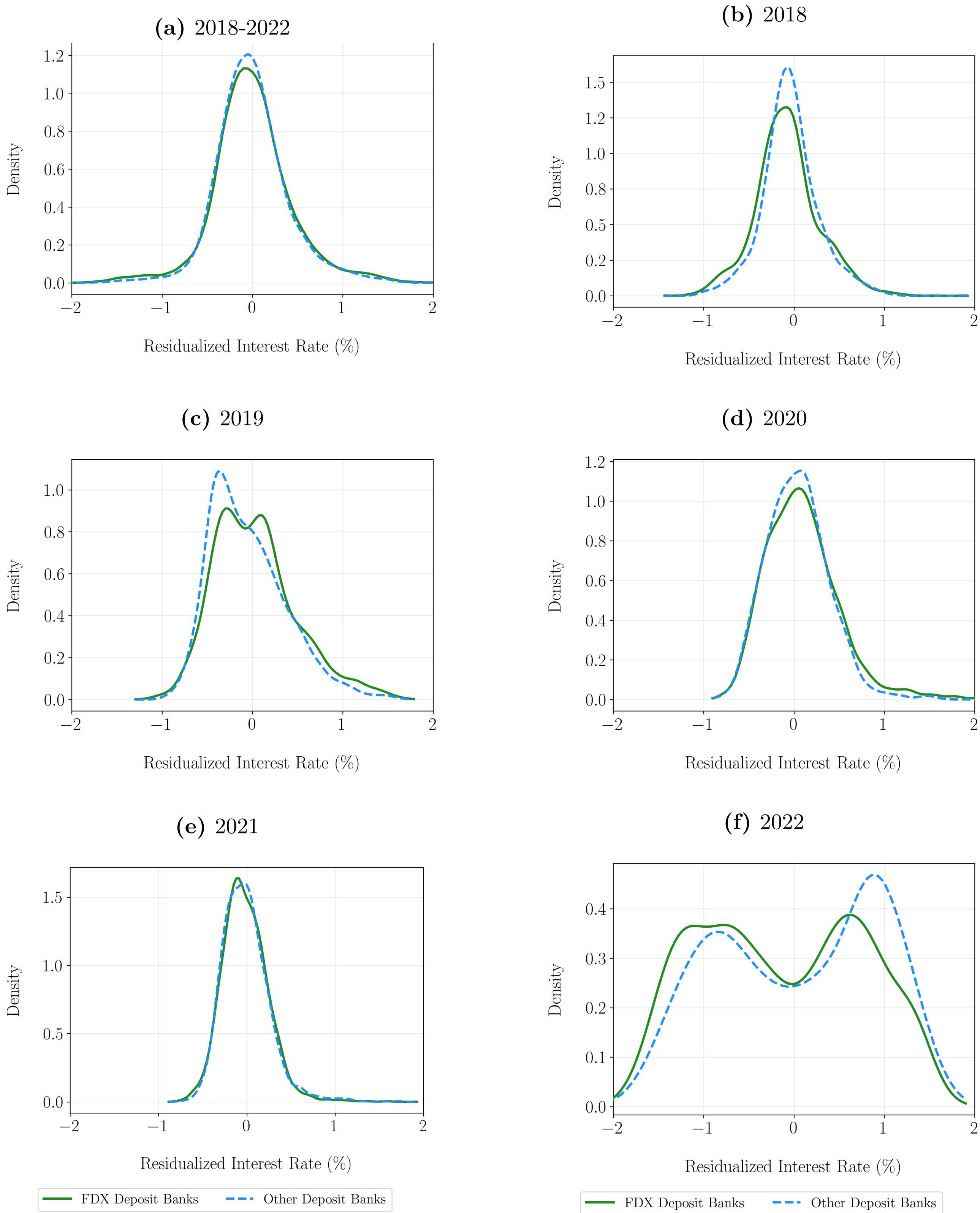
(b) Event Study Coefficient Estimates



Notes: differences in residualized interest rates between FDX and other banks. Panel (a) plots average res. int. rates. Panel (b) shows coefficient estimates from the event study specification in Equation (3) with res. int. rate as the outcome variable. Adj. $R^2 = 0.053$; obs. = 3,060; avg. res. int. rate at FDX lenders in 2018 (%) = -0.1.

Return: Section 3.2 (Descriptives, FDX Patterns).

Appendix Figure 13 – Residualized Interest Rate Distribution



Notes: distribution of residualized interest rates across years 2018-2022 in Panel (a); and for each year from 2018 through 2022 in Panels (b)–(f), respectively.

B. Additional Estimation & Counterfactual Results

Appendix Table 2 – Parameter Estimates: Comparison

		Benchmark	Agarwal et al. (2024)
		(1)	(2)
Demand Model Parameters			
Share of high type borr.	λ_h	0.956	0.268 (0.002)
Repay. prob. of high type borr.	x_h	0.783	1.000 (0.003)
Repay. prob. of low type borr.	x_l	0.357	0.410 (0.001)
Approval prob. of high type borr.	p_h	0.555	1.000 (0.004)
Approval prob. of low type borr.	p_l	0.041	0.193 (0.003)
Mean of (log) search cost distr.	μ_G	-1.708	-1.284 (0.005)
Std. dev. of (log) search cost distr.	σ_G	0.352	0.381 (0.004)
Mean of (resid.) offer rate distr.	μ_H	-0.089	0.142 (0.002)
Std. dev. of (resid.) offer rate distr.	σ_H	0.326	0.547 (0.001)
Selection parameter	σ	–	–
Non-bank fixed effect	ξ	–	–
Transaction cost	τ	–	–
Screening boost for high type borr.	κ_h	–	–
Screening boost for low type borr.	κ_l	–	–
Supply Model Parameters			
Origination cost	m	1.573	1.585 (.)
Std. dev. of idiosync. shock	σ_ε	0.778	0.410 (.)
Interoperability & transaction costs		–	–
Screening boost w/ interoperability		–	–

Notes: parameter estimates from our benchmark model, without transaction costs and screening boost in Column (1); and estimates from Agarwal et al. (2024) in Column (2).

Return: Section 5.4 (Estimation, Results).

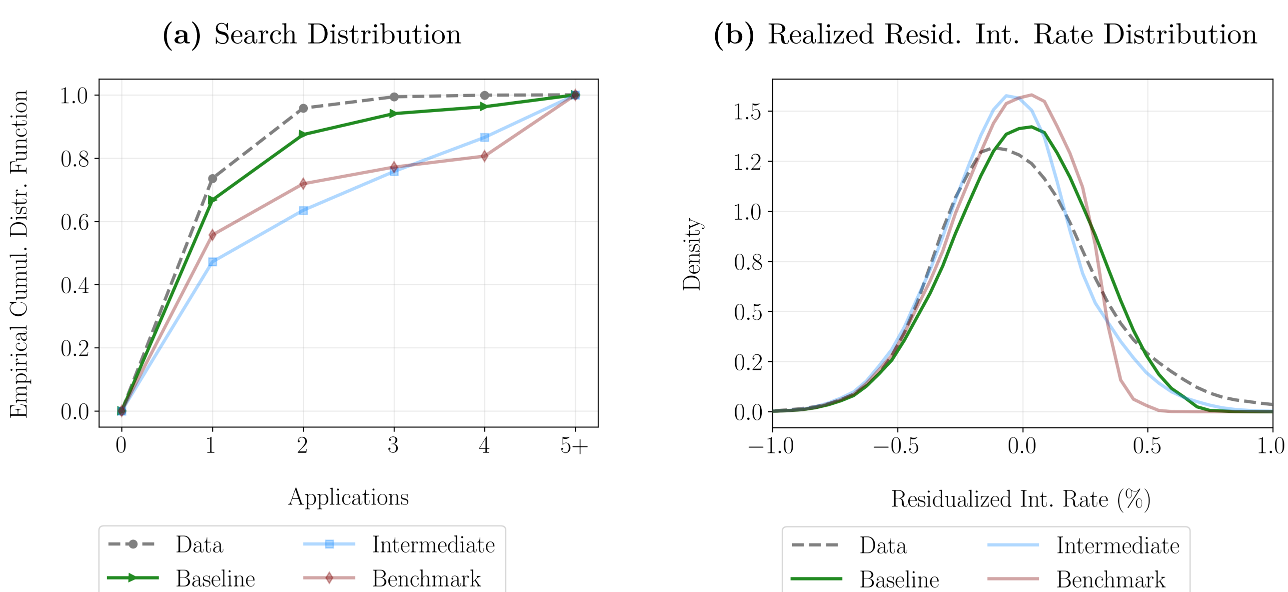
B.1 Model Fit

Appendix Table 3 – Model Fit: Summary

		Model		
	Data	**Baseline**	**Intermediate**	**Benchmark**
Origination share (%)	95.94	97.29	55.15	78.90
Applications / borrower (mean)	1.31	1.55	2.27	2.15
Default rate (%)	2.26	2.81	2.93	2.92
Residualized int. rate (mean)	-0.00001	-0.00656	-0.05264	-0.05379

Notes: comparison of statistics from the analysis sample and those implied by simulations of equilibria under model estimates in Table 3.

Appendix Figure 14 – Model Fit: Search and Realized Price Distribution

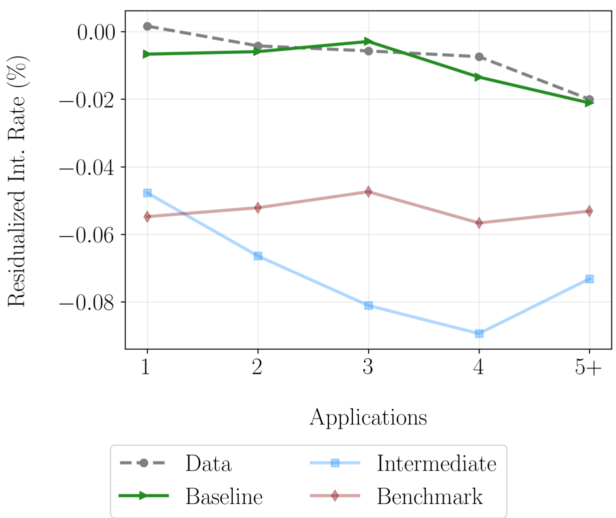


Notes: comparison of search, Panel (a), and realized interest rate, Panel (b), distributions derived from the analysis sample and those implied by simulations of equilibria under model estimates in Table 3.

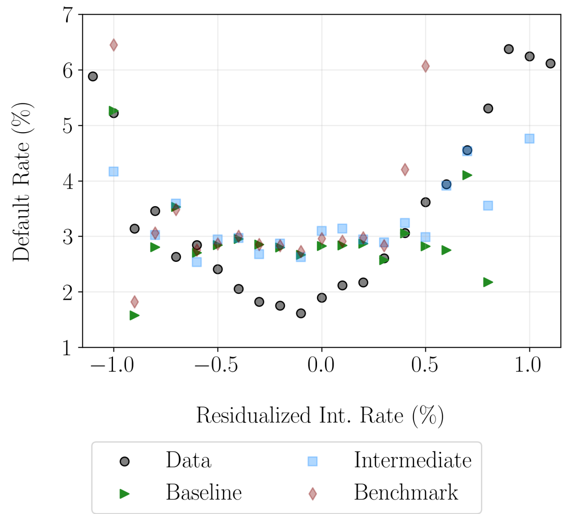
Return: Section 5.4 (Estimation, Results).

Appendix Figure 15 – Model Fit: (Price, Search) & (Default, Price) Associations

(a) Price and Search



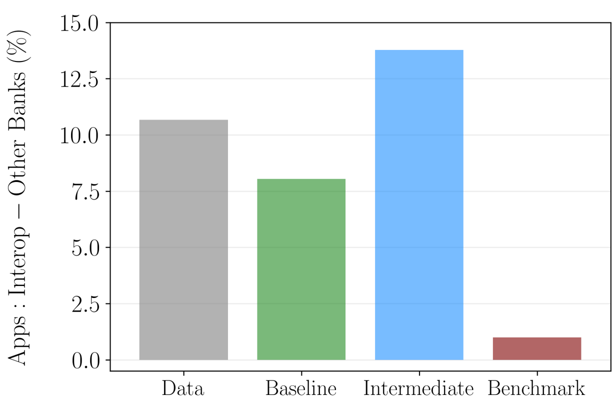
(b) Default and Price



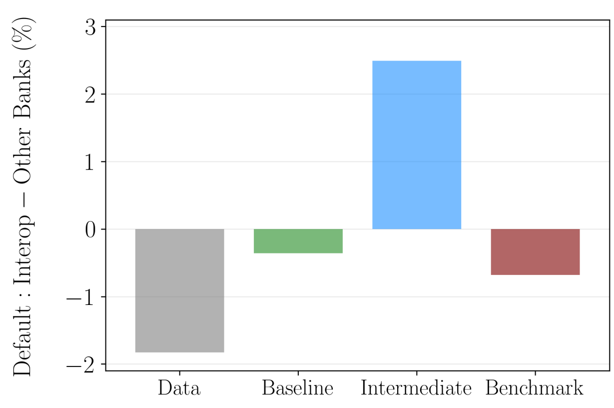
Notes: comparison of the associations between interest rates and applications, Panel (a); and default and interest rates, Panel (b), derived from the analysis sample and those implied by simulations of equilibria under model estimates in Table 3.

Appendix Figure 16 – Model Fit: FDX vs. Other Banks (Applications & Default)

(a) Differences in Applications



(b) Differences in Default Rate

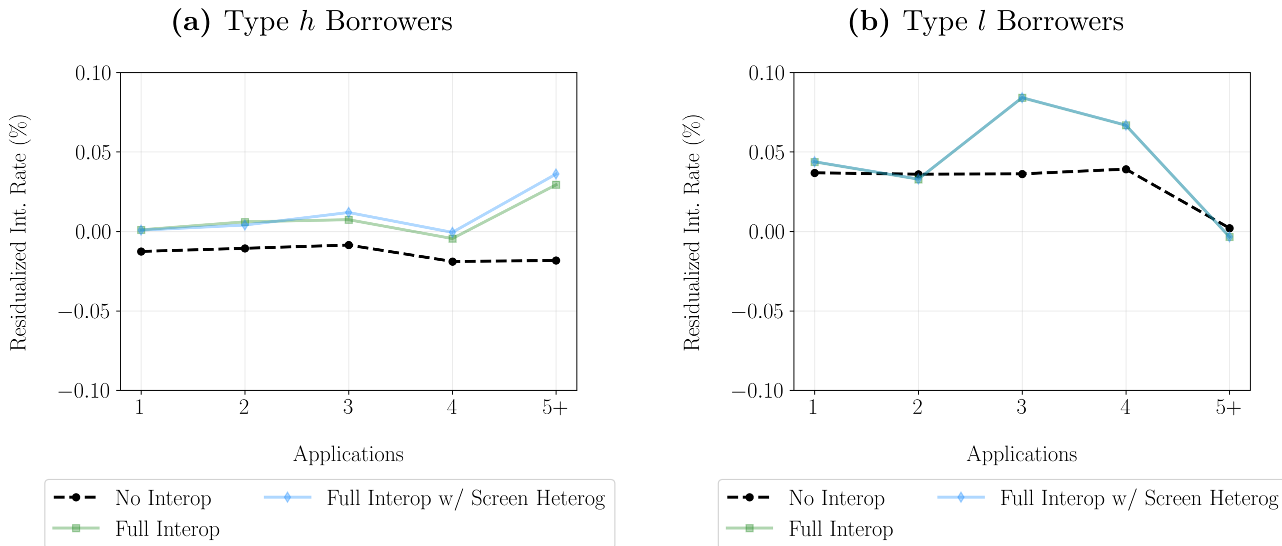


Notes: comparison of differences in applications, Panel (a), and default rates, Panel (b), between FDX and other banks—as derived from the event study of the analysis sample and those implied by simulations of equilibria under model estimates in Table 3.

Return: Section 5.4 (Estimation, Results).

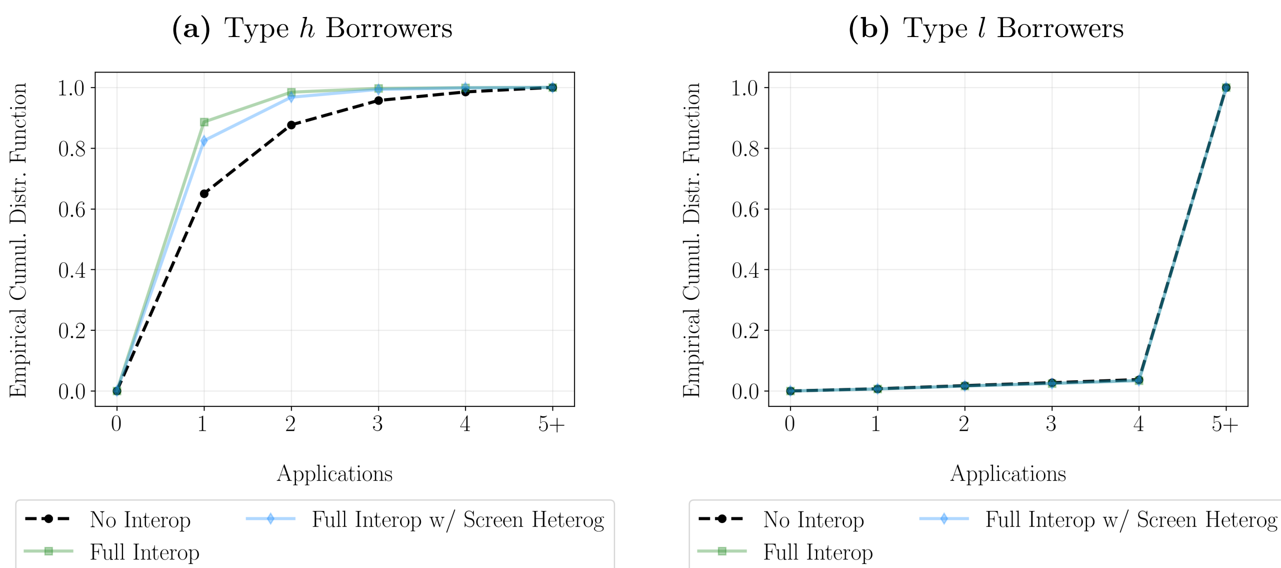
B.2 Counterfactual Results: Additional Exhibits

Appendix Figure 17 – Prices and Search under Counterfactuals



Notes: associations between prices and search across counterfactual scenarios described in Section 6.

Appendix Figure 18 – Search Distribution under Counterfactuals

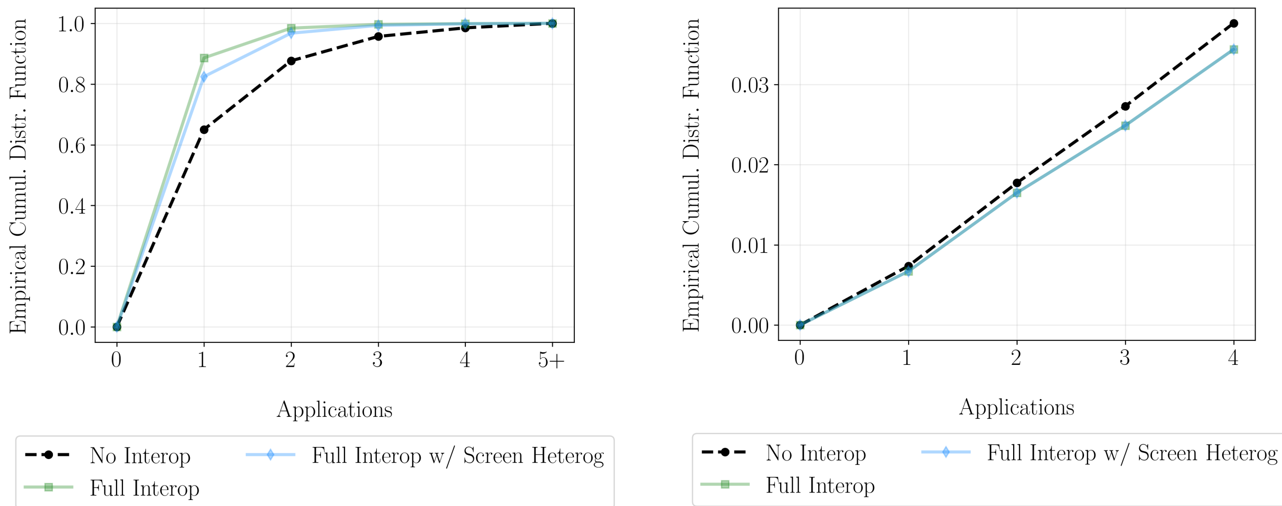


Notes: search distribution across counterfactual scenarios described in Section 6.

Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 19 – Search Distribution under Counterfactuals (Alternative Exhibit)

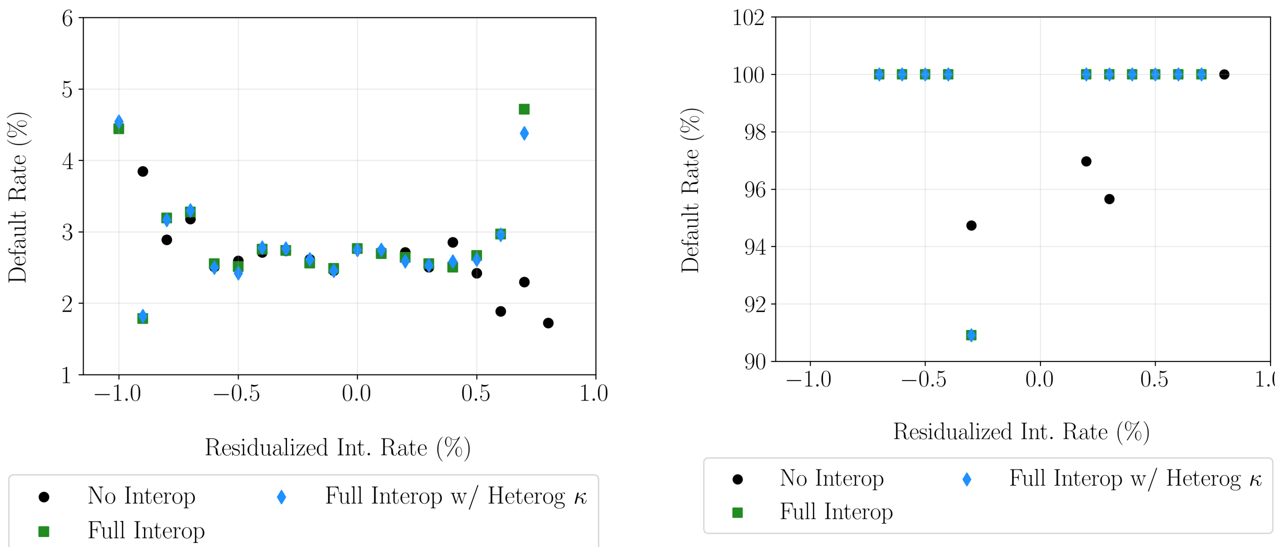
(a) Type h Borrowers **(b)** Type l Borrowers



Notes: search distribution across counterfactual scenarios described in Section 6.

Appendix Figure 20 – Default Rate and Prices under Counterfactuals

(a) Type h Borrowers **(b)** Type l Borrowers

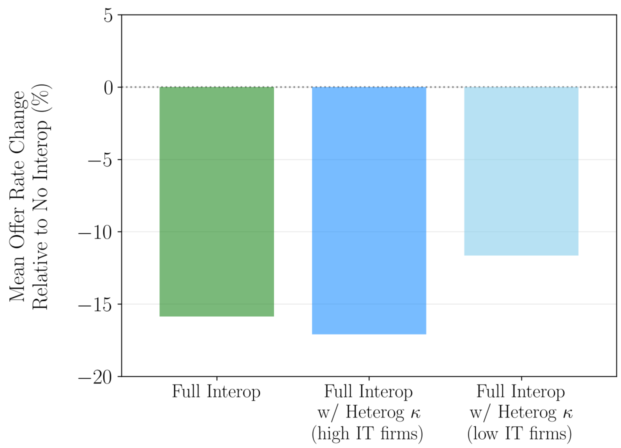


Notes: associations between default and interest rates across counterfactual scenarios described in Section 6.

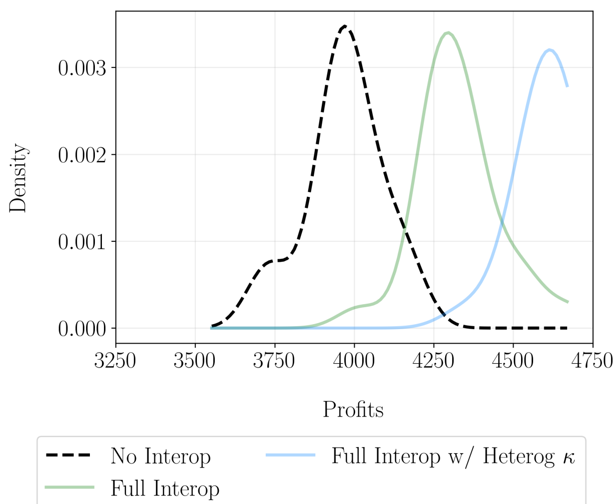
Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 21 – Offer Rate and Profit Distributions under Counterfactuals

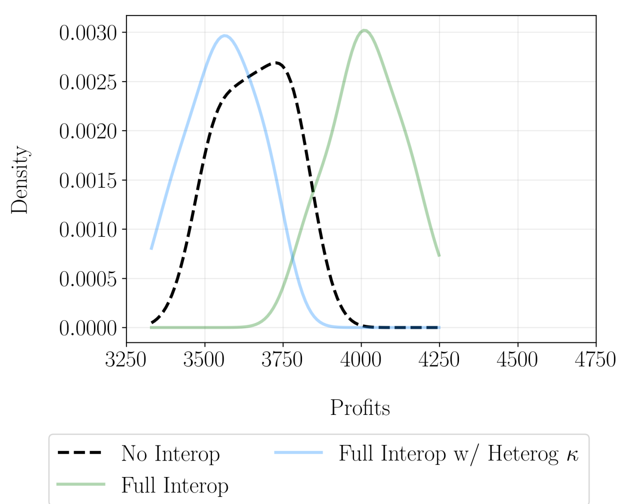
(a) Offer Rate Distribution



(b) High IT Firms - Profits



(c) Low IT Firms - Profits

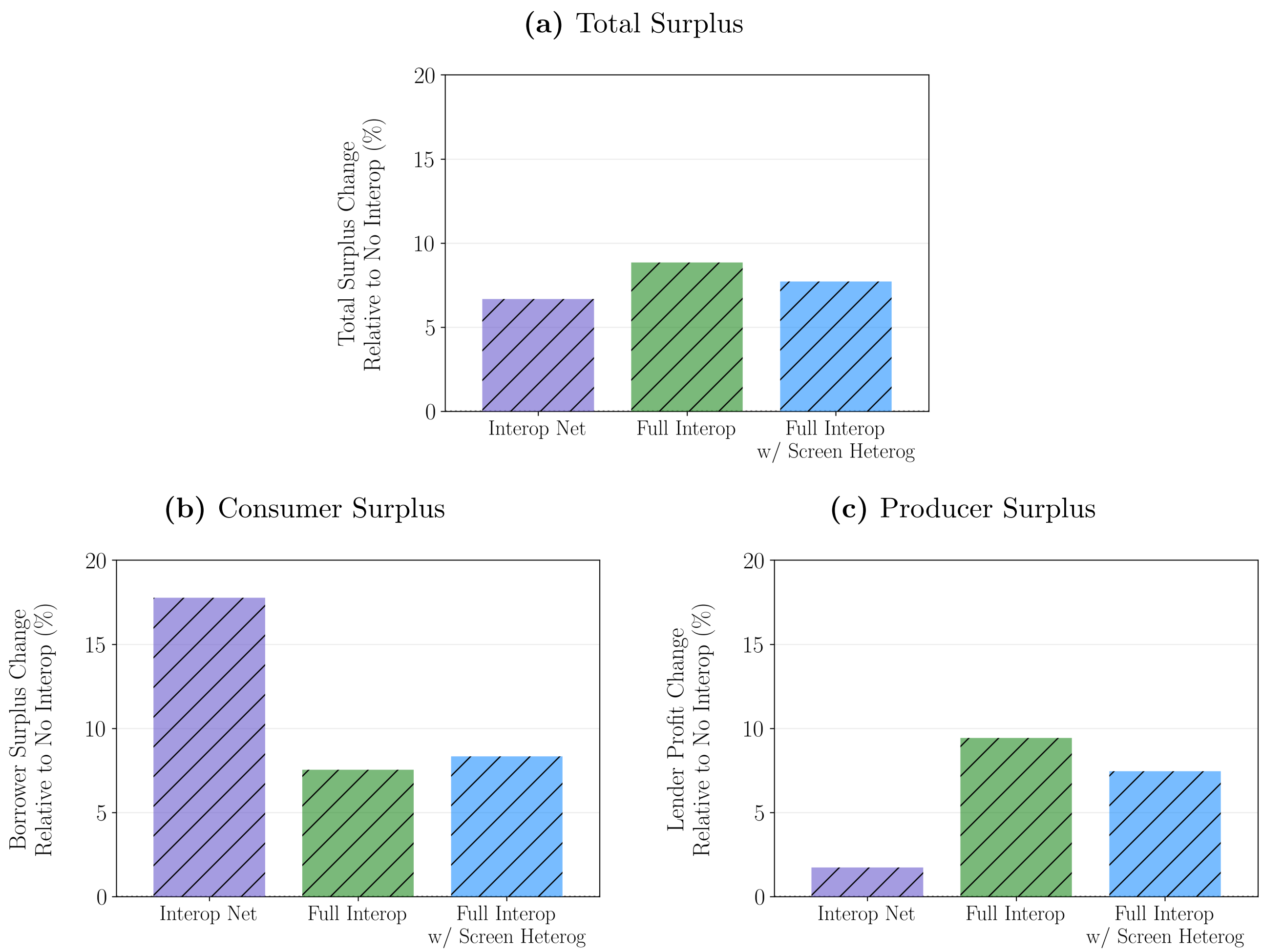


Notes: offer rate change relative to no interoperability, Panel (a), and profit distributions for high IT, Panel (b), and low IT, Panel (c) firms across counterfactual scenarios described in Section 6.

Return: Section 6.1 (Counterfactuals, Results).

B.3 Counterfactuals Extension: The Interoperable Network Regime

Appendix Figure 22 – Welfare Change Relative to the No Interoperability Regime under Extended Counterfactuals

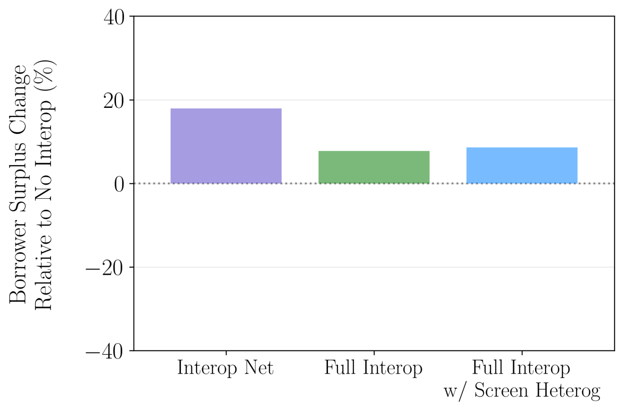


Notes: percentage difference in welfare, relative to the no interoperability regime, in total, Panel (a), consumer, Panel (b), and producer, Panel (c), welfare across alternative counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

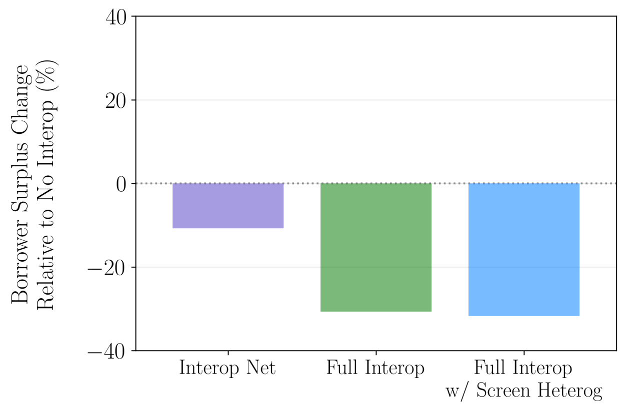
Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 23 – Borrower Welfare Change Relative to the No Interoperability Regime under Extended Counterfactuals

(a) Type h Borrowers



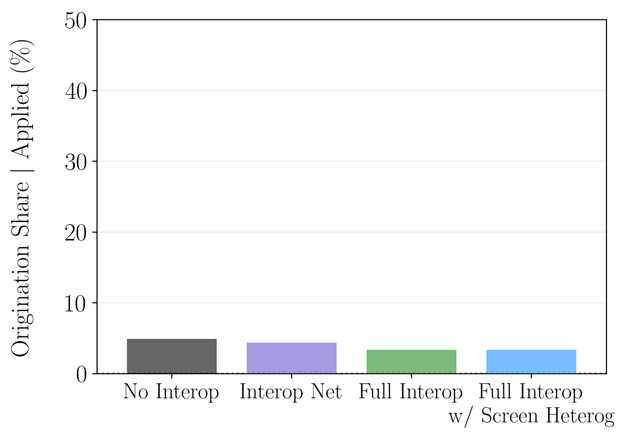
(b) Type l Borrowers



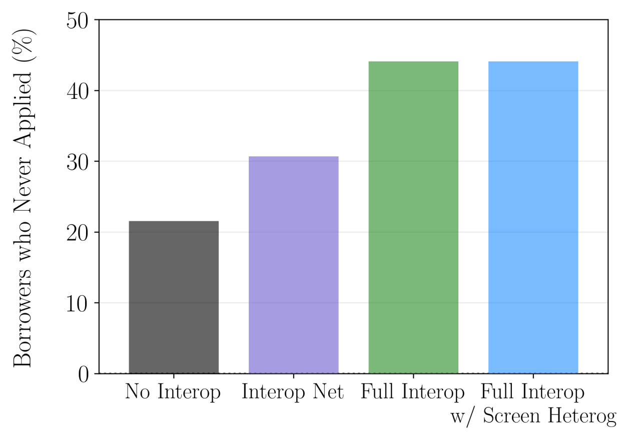
Notes: percentage difference in high, Panel (a), and low type, Panel (b) borrower surplus, when moving from the no interoperability to alternative counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

Appendix Figure 24 – Determinants of Risky Borrowers' Welfare Loss under Extended Counterfactuals

(a) Type l Borrower Origination Rate | Applied



(b) Type l Borrowers who Never Applied

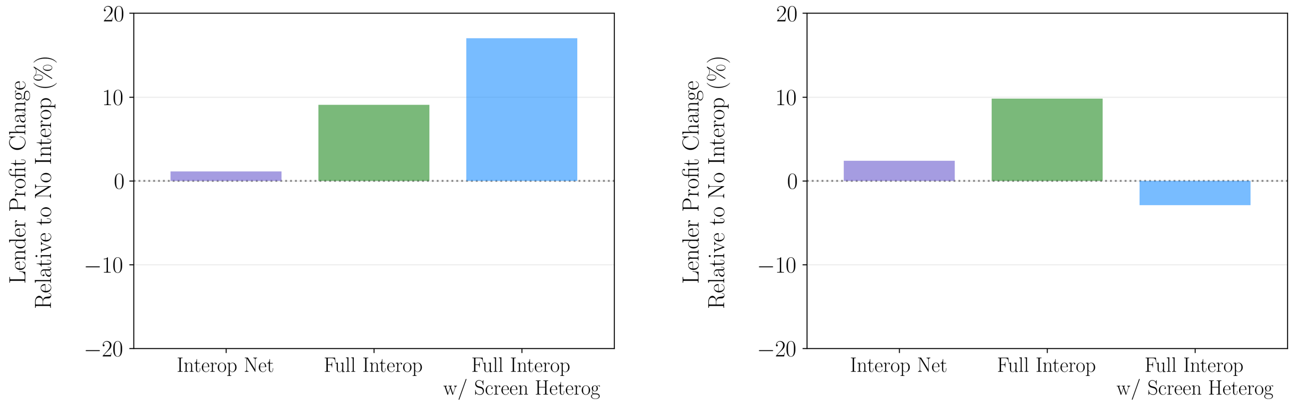


Notes: share of low type borrowers, who originate loans, Column (a), and never apply, Column (b).

Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 25 – Lender Welfare Change Relative to the No Interoperability Regime under Extended Counterfactuals

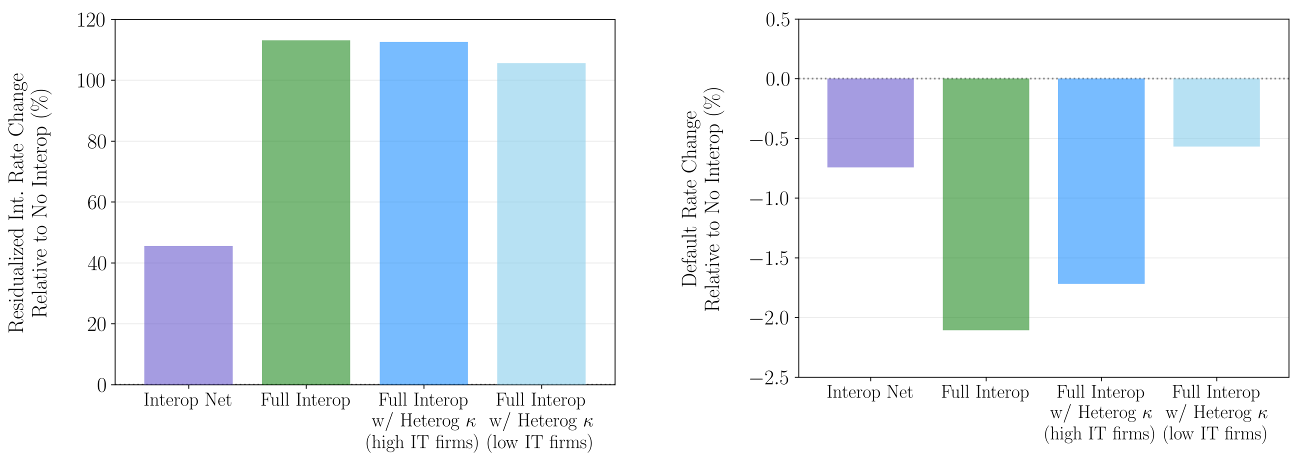
(a) High IT Firms **(b)** Low IT Firms



Notes: percentage difference in high, Panel (a), and low IT, Panel (b) firm profits, when moving from the no interoperability regime to alternative counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

Appendix Figure 26 – Determinants of Low IT Firms' Welfare Loss under Extended Counterfactuals

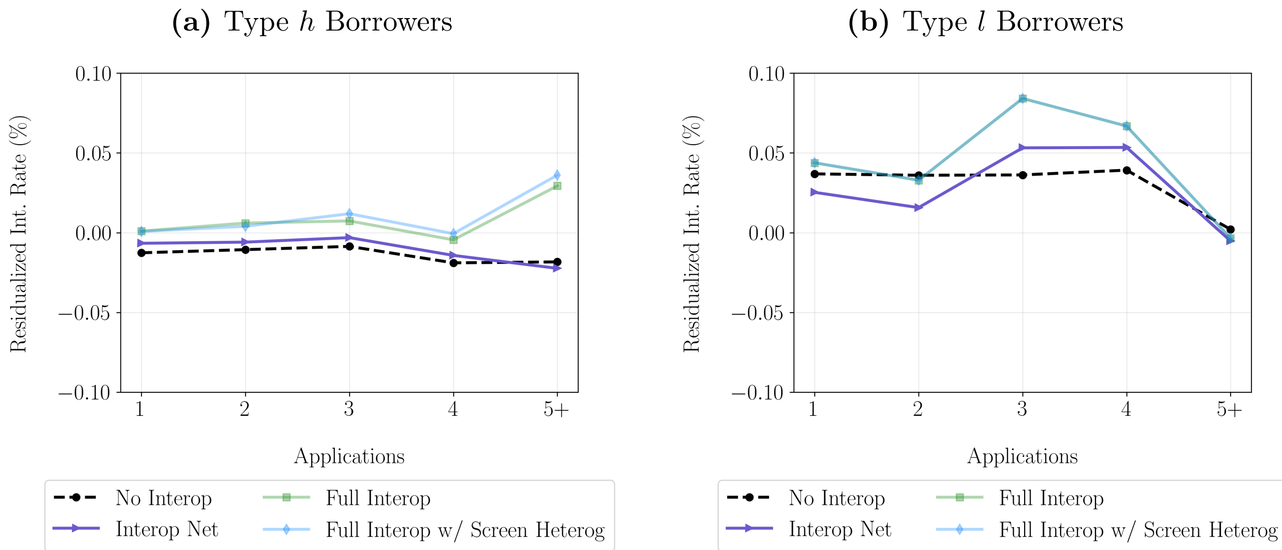
(a) Price **(b)** Default



Notes: percentage difference—relative to no interoperability—in interest rates, Column (a), and default rates, Column (b), across alternative counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

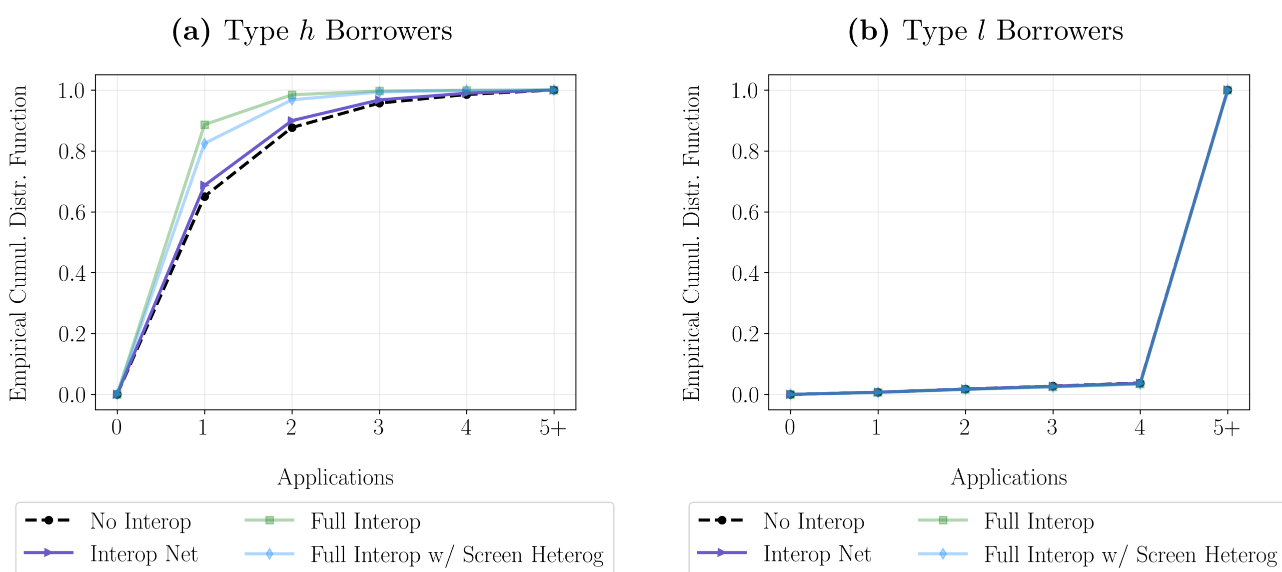
Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 27 – Prices and Search under Extended Counterfactuals



Notes: associations between prices and search across counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

Appendix Figure 28 – Search Distribution under Extended Counterfactuals

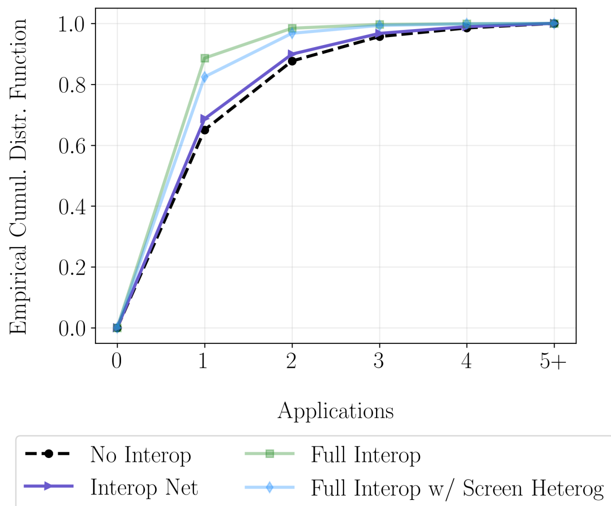


Notes: search distribution across counterfactual scenarios described in Section 6 plus the intermediate case of an interoperable network.

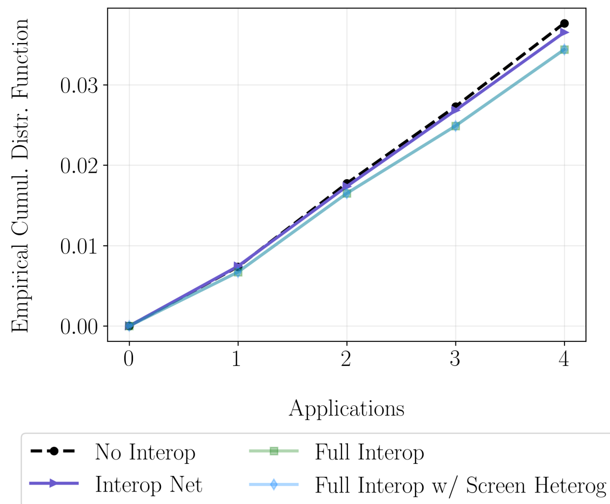
Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 29 – Search Distribution under Extended Counterfactuals (Alternative Exhibit)

(a) Type h Borrowers



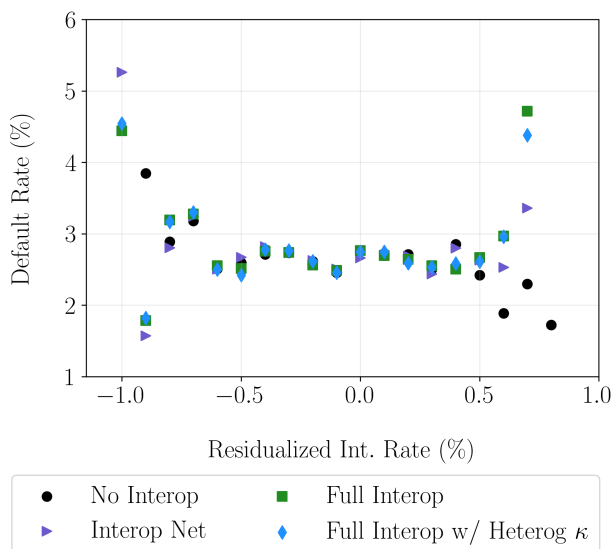
(b) Type l Borrowers



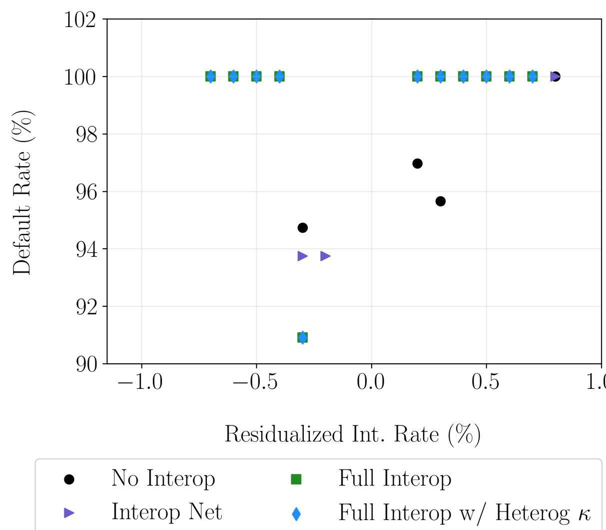
Notes: search distribution across counterfactual scenarios described in Section 6 and and the intermediate case of an interoperable network.

Appendix Figure 30 – Default Rate and Prices under Extended Counterfactuals

(a) Type h Borrowers



(b) Type l Borrowers

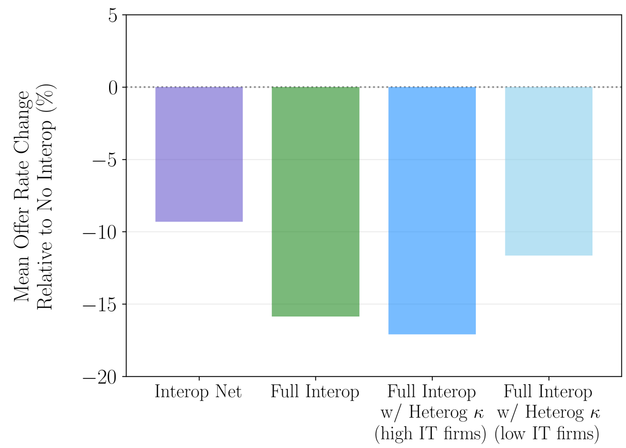


Notes: associations between default and interest rates across counterfactual scenarios described in Section 6 and the intermediate case of an interoperable network.

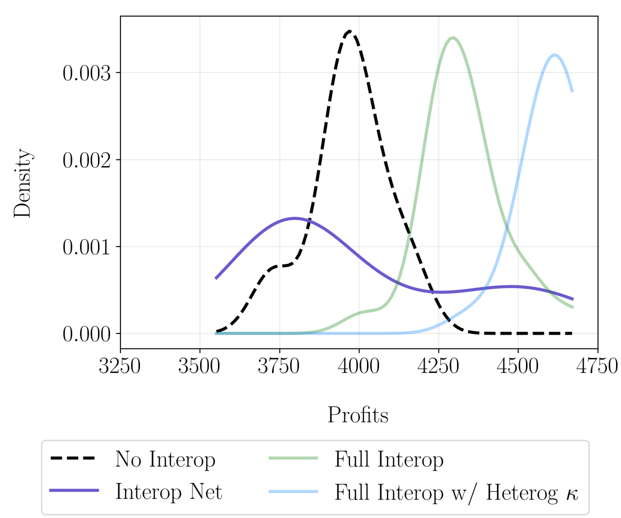
Return: Section 6.1 (Counterfactuals, Results).

Appendix Figure 31 – Offer Rate and Profit Distributions under Extended Counterfactuals

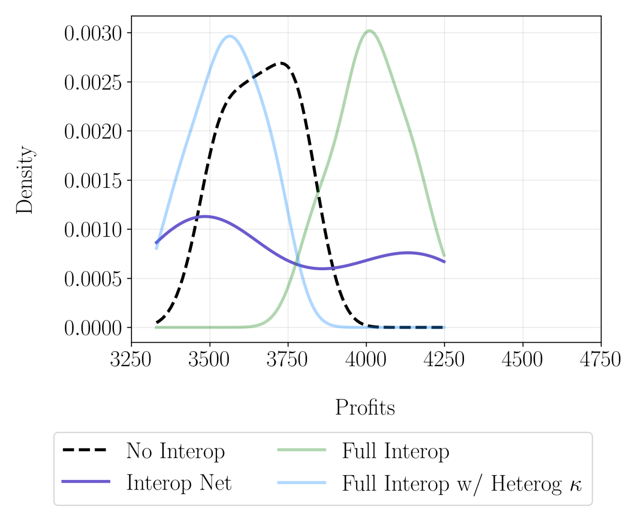
(a) Offer Rate Distribution



(b) High IT Firms - Profits



(c) Low IT Firms - Profits



Notes: offer rate change relative to no interoperability, Panel (a), and profit distributions for high IT, Panel (b), and low IT, Panel (c) firms across counterfactual scenarios described in Section 6 and the intermediate case of an interoperable network.

Return: Section 6.1 (Counterfactuals, Results).

C. Model, Estimation & Counterfactual Details

C.1 Model Details

C.1.1 Proof of Proposition 1

From the associations between search costs and optimal cutoff rates in Equations (4)–(5), we have that

$$\phi_{z_i}^{n_{b_i},1}(r^*) = \phi_{z_i}^{n_{b_i},0}(r^*) - \tau - p_{z_i}(1-\kappa_{z_i})(r^* - \underline{r})u_i(r^*).$$

Since, by assumption, $u_i(r)$ is monotonic in r, we can apply the relationship between the distribution of reservation rates and search costs in Equation (6) to show that, absent of screening boost ($\kappa_h = \kappa_l = 1$),

$$F_{z_i}^{n_{b_i},0}(r^*) = G\Big(\phi_{z_i}^{n_{b_i},0}(r^*)\Big) > G\Big(\phi_{z_i}^{n_{b_i},0}(r^*) - \tau\Big) = F_{z_i}^{n_{b_i},1}(r^*) \qquad \forall z_i \in \{h, l\}.$$

Now, fix r. Then, borrowers at in-network banks, $n_{b_i} = 1$, are more likely to apply within rather than outside of the interoperable network (application effect),

$$\mathbb{P}[r^* > r | m(j, b_i) = 1] = 1 - F_{z_i}^{11}(r) > 1 - F_{z_i}^{10}(r) = \mathbb{P}[r^* > r | m(j, b_i) = 0].$$

By a similar argument, this application effect is amplified for type h borrowers,

$$1 - F_h^{11}(r|\kappa_h > 1) > 1 - F_h^{11}(r|\kappa_h = 1)$$

and dampened for type l borrowers,

$$1 - F_h^{11}(r|\kappa_l < 1) < 1 - F_h^{11}(r|\kappa_h = 1). \qquad \square$$

C.1.2 Proof of Proposition 2

The first part of the result stated in Proposition 2, i.e. the attraction effect, follows directly from Proposition 1. It remains to show that the diversion effect holds.

With a little bit of algebra, one can show, using the relationships between optimal reservation rates and search costs in Equations (4)–(5), that

$$\phi^{11}(r^*) = \phi^{01}(r^*) + \tau(\gamma - s_j) + p_{z_i}(\kappa_{z_i} - 1)[\gamma(1 + s_j) - s_j] \int_{\underline{r}}^{r^*} u_i(\tilde{r}) dH(\tilde{r}).$$

Now, fix r. Then, absent of screening efficiencies ($\kappa_h = \kappa_l = 1$), b_i's own consumers are less likely to apply to b_i when b_i is in-network compared to when b_i is out-of-the-network,

$$\mathbb{P}[r^* > r | n_{b_i} = 1, m(j, b_i) = 1] = 1 - F_{z_i}^{11}(r) < 1 - F_{z_i}^{01}(r) = \mathbb{P}[r^* > r | n_{b_i} = 0, m(j, b_i) = 1].$$

Following an analogous argument, this diversion effect is mitigated for type h borrowers,

$$1 - F_h^{11}(r | \kappa_h > 1) > 1 - F_h^{11}(r | \kappa_h = 1)$$

and exacerbated for type l borrowers,

$$1 - F_h^{11}(r | \kappa_l < 1) < 1 - F_h^{11}(r | \kappa_h = 1). \qquad \square$$

Return: Section 4 (Model).

C.2 Estimation Details

We perform the demand and supply model estimation in `Python` using high performance computing hosted at NYU. In particular, we employ the `trust-constr` solver from the `optimize.minimize` routine of the `SciPy` library.[21] For numerical integration, we use Gauss quadrature approximation.

Return: Section 5 (Estimation).

C.3 Counterfactual Details

In our counterfactual (and model fit) exercises, we simulate equilibria for 1,600 markets: 400 MSAs across 4 years. We consider 100 lenders, 50 of which are depository banks. In each MSA-year pair, there are 100 consumers, who search up to 7 times for 30-year fixed maturity mortgages. Conditional on originating a loan, default is observed after 10 years. Table 4 summarizes our simulation parameters.

In each MSA-year combination, we simulate the banking (i.e., consumer data) market structure by (1) randomly drawing a subset of depository institutions, and (2) simulating their deposit market shares. We allow for the presence of banks that do not offer residential mortgages. Home banks are determined by a random match between each of the 100 consumers and one of the banks operating within a given MSA-year pair, weighted appropriately to match the simulated

[21] Last Accessed: November 26, 2024. Documentation available at: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html.

deposit shares. FinTechs can originate loans to any consumer, while depository institutions are restricted to offering loans in markets where they have banking presence.

Simulating borrower search, application, screening, and origination in each market proceeds as follows. For each borrower i, we draw a search cost, c_i, from $G(\hat{\boldsymbol{\beta}})$ and repayment prospect z_i given the estimated distribution of types $\hat{\lambda}$. Given c_i, we back out each consumer's match-specific reservation rates using the threshold relationships in Equations (4)–(5). To that end, we employ the `optimize.root` routine within the `SciPy` library in `Python`. Next, we simulate search sequentially. In each of borrower i's search iteration, we draw an offer rate, R_{ij}, from $H(\hat{\boldsymbol{\beta}}_H)$ and a lender j, from the set of firms offering mortgages in the given market. Application is realized if the offer rate, R_{ij}, is below consumer i's reservation rate $\hat{r}^*_{ij}$, which accounts for the (network) affiliation between i's home bank, b_i, and firm j. Given an application, we simulate approval according to the match-specific screening accuracy, $\hat{p}^{n_{b_i},m(j,b_i)}$, which considers the relation between b_i and j. If an application is approved, then the loan is originated and search concludes. Finally, we simulate default for originations using the probability that i survives through the 10th year without any delinquencies; i.e.: $\widehat{\boldsymbol{\Omega}}(t|z_i;T) = \hat{x}_{z_i}^{t/T}$.

Appendix Table 4 – Simulation Parameters

Parameter	Value
Markets	
# Years	4
# MSAs	400
Consumers	
# Borrowers in each MSA-year	100
Origination mortgage term	30
Period of observed (non)default	10
Firms	
# Lenders in each MSA-year	100
# Depository Institutions	50

Notes: parameters used in simulating equilibria for model fit and counterfactual analyses.

Return: Section 6 (Counterfactuals).

D. Data Construction Details

D.1 Matching Applications in HMDA Data

The application matching procedure within HMDA proceeds as follows. We begin by collapsing applications to the Census tract, consumer income, and loan amount level. We then merge Census tracts exactly, and allow for a margin of error in income (ω_I) and loan amount (ω_L)—we do so to account for minor differences in these characteristics such as reporting errors, changes in income, and qualifying mortgage terms, which may vary across lenders due to idiosyncratic loan evaluation processes. We keep all unique matches. For non-unique merges, we only select ones with the smallest differences in income and loan amount. We choose margins of error, ω_I and ω_L, that minimize the sum of squared differences in the shares of borrowers with a given number of applications $(1, 2, ..., 5+)$ between those (a) implied by our data build process, and (2) reported in NSMO. Using the optimal ω's, which are generally 3-5%, generates search distributions in our dataset that closely match with those documented in NMSO (Figure 2).

D.2 Matching Originations in HMDA and Freddie Mac Data

We take the internally matched HMDA dataset from Appendix Section D.2, and further match it with Freddie Mac performance data. We do so in two steps. First, subset our HMDA dataset to originations secured by Freddie Mac, as classified in the raw HMDA data. Next, we collapse both datasets to a common set of variables: MSA, consumer income, and loan amount. We employ the same matching procedure as in the internal HMDA merge (Appendix Section D.2), with the same optimal margins of error, and fine-tune the precision of matches by further merging on CLTV and DTI. This final step yields our analysis sample.

D.3 Miscellaneous

FDX lenders. We flag firms in our analysis sample as FDX if they are listed in any FDX press release from 2019 to the present.[22]

FDIC SOD matching. We merge firms in our analysis sample with the FDIC Summary of Deposits using the Replication Server System Database ID ("RSSD ID"), a unique identifier

[22] Last accessed: November 27, 2024. Available at: https://financialdataexchange.org/FDX/FDX/News/Press-Releases-home.aspx?hkey=4120f9d9-6265-4ab7-a19b-9e6ace6d36ab.

assigned to each financial institution by the Federal Reserve. We classify firms as non-banks if they do not appear in the FDIC SOD during our study period (2018-–2021).

Default and delinquencies. For each loan origination, Freddie Mac's Single-Family Loan-Level Performance Dataset tracks the number of delinquent days since loan inception. We classify a borrower as being in default—as of the last reported period—if they experienced at least one monthly spell of 90 or more days of delinquency.

Return: Section 2.2 (Data).