Consumer Credit Without Collateral, Regulation, or Intermediaries*

Filipe Correia † Antonio Martins ‡ Anthony Waikel §

March 5, 2024

Abstract

Using novel data from an online informal credit market, we investigate how consumer credit unfolds without a financial system. We find borrowers display high rates of default and face high credit prices. A minority of consistently successful lenders gain a disproportionately large and profitable market share, while the average lender realizes losses. High-skill lenders achieve better loan outcomes and provide more lenient loan terms. Loans are more likely to be funded and repaid when acquiring information about borrowers is easier for lenders. These findings highlight the role of intermediaries in consumer credit: they bring skill and internalize information acquisition.

Keywords: Consumer Credit, Financial Intermediation, Household Finance, Informal Credit

JEL Classification: O33, G20, G51

*We thank Marco Bonomo, Isaac Hacamo, Pekka Honkanen, Jeff Netter, Annette Poulsen, Bryan Routledge, Rik Sen, Leigha Waikel for valuable comments. Anthony Waikel thanks the University of Georgia Graduate College for a Summer Grant utilized in this project. All errors are our own.

†Corresponding Author. Terry College of Business, University of Georgia, 600 S Lumpkin St, Athens, GA 30605, filipe.correia@uga.edu

‡Lisbon School of Economics and Management, Research Unit on Complexity and Economics (UECE), Research in Economics and Mathematics (REM). Rua do Quelhas 6, 1200-781 Lisboa, acmartins@iseg.ulisboa.pt

§Terry College of Business, University of Georgia, Athens, GA, 30605, waikela@uga.edu

1 Introduction

In the United States, 4.5% of households are unbanked, with the primary reason being "[not] hav[ing] enough money to meet minimum balance requirements" (FDIC 2022). The lack of access to credit and considerable liquidity constraints lead these households to search for alternative forms of consumer credit, outside the financial intermediation ecosystem.

The evolving debate on the role of financial intermediaries[1] is usually centered around firm financing, rather than consumer credit (e.g., Fang, Ivashina and Lerner (2015); Harris, Jenkinson, Kaplan and Stucke (2018); Crawford, Pavanini and Schivardi (2018); Duygun, Hashem and Tanda (2021)) and it is split into two views. On one hand, financial institutions extract rents via spreads and fees, have their own set of incentives, or even engage in predatory practices. On the other hand, financial intermediation lowers transaction and search costs, alleviates information asymmetry, and efficiently pools and allocates capital. We have a rich and vast literature on intermediated consumer credit, but little is known about how it would function in the absence of intermediaries. We study the role of intermediaries in consumer credit, by directly observing a consumer credit market, operating without intermediation.

We contribute to the debate in the context of consumer credit markets, using novel data from an online informal credit market, operating on a Reddit forum[2]. Informal credit is defined as "loans that rely on personal relationships or social sanctions as means of enforcement" (Karaivanov and Kessler 2018). These loans may be obtained from family, friends, neighbors, or strangers with no intermediary between lender and borrower. Whether informal loans have positive real effects for borrowers is very unclear (Banerjee, Karlan and Zinmin 2015; Angelucci, Karlan and Zinmin 2015), mostly because the inner workings of informal

[1]The debate dates back to Pyle (1971); Allen and Santomero (1997); Allen and Gale (1999); Scholtens and Van Wensveen (2000), and has consistently called for more research throughout the decades, being still the focus of the latest Nobel Lecture by Diamond (2023)

[2]Correia, Martins and Waikel (2022) describe the dataset in detail, and study this market from the point of view of market efficiency and find that loan terms reflect the economic context of borrower and lenders, with loan interest rate and requested amounts increasing after a local COVID-19 lockdown was implemented.

credit have not been the target object of study. The nature of non-intermediated loans makes it very difficult to collect data that allows for in-depth studies of the informal credit market, despite its widespread use among consumers. Our hand-collected dataset contains information about 95,000 loans (26 million USD, 2016-2021) between about 3,000 unique lenders and 15,000 unique borrowers, who transact in the absence of a financial system to intermediate, regulate, and collateralize credit.

With surface-level descriptive evidence, we show that borrowers display high rates of default and face high credit prices when compared to intermediated and collateralized subprime credit, such as payday loans. The absence of collateral is key, as collateral mitigates adverse selection and moral hazard (Ioannidou, Pavanini and Peng 2022). In this market, borrowers announce the quantity and price they want to borrow, and lenders decide whether to fund. Thus, adverse selection is prevalent in this market, as borrowers willing to pay such a high rate are announcing their known higher credit risk (Kawai, Onishi and Uetake 2022). Additionally, a small group of sophisticated lenders occupies a disproportionately large and very profitable market share, with low default rates, leaving out unprofitable and riskier loans for unsophisticated lenders. This suggests that some attributes of the lender, information of the borrower, or both determine lender choices. A survey of 100 participants in this market shows lenders to be more financially sophisticated than borrowers, on average, and a positive relationship between financial literacy scores and portfolio profitability.

We build a stylized framework to understand borrowers' and lenders' decision-making in this market. We model decision-making in three steps, very similar to Kawai, Onishi and Uetake (2022) but different from the Prosper setting, where lenders bid. First, the borrower chooses whether to post a loan request promising a given interest rate. This key feature matches our empirical setting. Second, given that rate, the lender chooses whether to fund the loan. Lastly, the borrower chooses whether to repay the loan, based on whether that repayment earns them enough reputation to allow them to borrow in the future. This feature also matches our empirical setting, where reputation is the only device that incentivizes borrowers to repay in informal credit settings (Karaivanov and Kessler 2018; Xin 2023). We assume lender unit costs of origination decrease with acquired skills, like in Besanko, Doraszelski and Kryukov (2014), where marginal cost decreases with the stock of know-how. The reputation gain from loan repayment decreases the lenders' assessment of the proportion of high-risk borrowers in the market. This reputation gain is a reflection of the lender's ability to learn and infer from the borrower's repayment. Our setting generates two predictions. First, higher lender experience/skills should result in higher loan funding likelihood, loans with a lower default rate, and lower prices. Second, less opacity between borrowers and lenders should result in lower default incidence and higher funding likelihood.

With our dataset, we observe loan requests posted by borrowers, containing requested amounts, promised repayment amounts, and maturity. We can use these three quantities to compute a promised monthly interest rate. Additionally, we can match loan requests to actual loans and obtain information on whether the request is funded (i.e., converted into a loan), its repayment status, repayment date, and effective interest earned. We measure lenders' acquired skill by the number of past loans the lender funded at the date of each loan. While the measure of acquired skill is complicated by a wide dispersion of lender activity and endogeneity, we show results hold when changing the acquired skill measure and we use an Instrumental Variable approach to address endogeneity issues.

Consistent with our hypotheses, we find that more skilled lenders are more likely to fund a loan, face lower default, and choose loans with more lenient interest rates. This result suggests they understand moral hazard risks and are more diversified, allowing them to take on more loans and handle losses. This finding speaks to the economic importance of financial intermediation in lowering search and screening costs per dollar originated. More skilled lenders develop tools and procedures to screen borrowers at a lower cost. Intermediaries also benefit from diversification via their larger funding capacity. Likewise, skilled lenders given past successes in lending, gather larger and more diversified portfolios.

This is also consistent with lenders learning how to screen soft information. Soft information contained in loan funding requests enhances funding likelihood, but loans selected based on this information underperform (Dorfleitner, Priberny, Schuster, Stoiber, Weber, de Castro and Kammler 2016; Herzenstein, Sonenshein, Dholakia, Stoiber and Weber 2011). There is, however, information content in narratives. Iyer, Khwaja, Luttmer and Shue (2015) find that lenders achieve 87% of the predictive power of an econometrician who observes all standard financial information about borrowers, and that screening on soft or nonstandard information is relatively more important when evaluating lower-quality borrowers.

Additionally, we test the hypothesis that borrower-lender connections increase the funding likelihood for a loan request. We extend our data to measure the overlap in social footprints of each pair of borrowers and lenders. We find that loans are more likely to be funded by lenders who have more in common with the borrower. Connectedness also increases the probability that the borrower repays the loan, in line with Lin, Nagpurnanand and Viswanathan (2012) who find that consumers on marketplace lending platforms with more online friendships are associated with lower default rates. Similarly, Xin (2023) finds that reputation-based systems disincentivize defaults and expand credit access to borrowers.

This is consistent with the reputation gain being larger for borrower-lender pairs who are more connected. This result holds for lenders of all skill levels, but loan terms do not vary with connection level, also consistent with our stylized framework. Our findings shed light on information asymmetry in credit markets and the importance of established lending relationships to mitigate them (Karolyi 2017). Without a financial system to provide credit scores and identity verification, the only mechanism that enforces loan repayment is reputation. Therefore lenders need to be able to relate and gain familiarity with borrowers, to accurately measure their reputation and form their beliefs.

Much of the literature uses geographic proximity as a proxy for connectedness (Baily, Cao, Kuchler, Stroebel and Wong 2018; Bayer, Mangum and Roberts 2021), Some studies, such as Fracassi (2017) and Lin, Nagpurnanand and Viswanathan (2012) use shared education networks, or online friendships to measure connectedness between two individuals. We add to the social connectedness research by measuring connectedness between people using their observed mutual interests. This has two unique features: (i) geographic location is irrelevant, and (ii) individuals do not necessarily know they are connected. We obtain user activity over our entire sample period, generating variation in connectedness in the cross-section and in the time series.

We also shed light on the importance of platforms and their design in emerging marketplace lending in increasing access to credit for risky borrowers neglected by banks. In these markets interest rates offer adequate returns for lenders, on a risk-adjusted basis (Kraussl, Kraussl, Pollet and Rinne 2022). Berger and Gleisner (2009) find that peer-to-peer lending platforms help reduce information asymmetry. Dinerstein, Einav, Levin and Sundaresan (2018) document a trade-off in online platform design between reducing search frictions and intensifying price competition among sellers, while Einav, Farronato, Levin and Sundaresan (2018) documents a shift from auctions to posted prices in online commerce, both using eBay data. On the credit side, Franks, Serrano-Velarde and Sussman (2021) use data on marketplace loans from the UK to show that credit marketplaces also deviated from the auction platform design, and in turn set prices and allocate credit on their own.

The closest paper to ours is Braggion, Manconi, Pavanini and Zhu (2022), which models the role of platform and lender decision-making on peer-to-peer platforms moving from direct peer-to-peer loans towards an intermediated loan portfolio model. Fitting the model to data from the Chinese platform Renrendai, they find that this transition generates lender surplus, profits for the platform, and enhances credit provision. Without platforms, peer-to-peer credit would resemble our setting, with high prices, and a small share of lenders capturing all the profitable opportunities. The inability to enforce repayment, the lack of identity verification, and the absence of a platform to intermediate and equalize lending opportunities for lenders are key in this off-bank marketplace. We are the first to empirically analyze a pure peer-to-peer (as compared to peer-to-platform-to-peer) market without a platform conducting intermediation. Our findings reconcile the transition in peer-to-peer lending platforms from individual lenders funding individual loans, and instead obtaining bank charters[3], or partnering with banks to originate loans.

Our unique dataset is one of the largest datasets on credit without intermediaries, collateral, or regulation collected to date. Altogether, our findings speak to the role of the financial system in consumer credit, by (i) mitigating information asymmetry: with reduced screening costs, implementing screening systems like credit scores, and building lending relationships (ii) improving access to affordable credit: by generating stable returns from pooling funds and sharing risk, improving liquidity and funding capacity, internalizing the credit pricing process, and implementing mechanisms allowing consumers to post collateral.

The remainder of the paper proceeds as follows. In Section 2 we describe the loan market and our data collection. In Section 3 we provide descriptive evidence on the market. In Section 4 we build our stylized framework and derive hypotheses. In Section 5 we measure skill to study how loan outcomes and terms change with lender skills, and Section 6 develops our measures of user connectedness. We conduct robustness checks of our findings in Section 7. Finally, we conclude in Section 8.

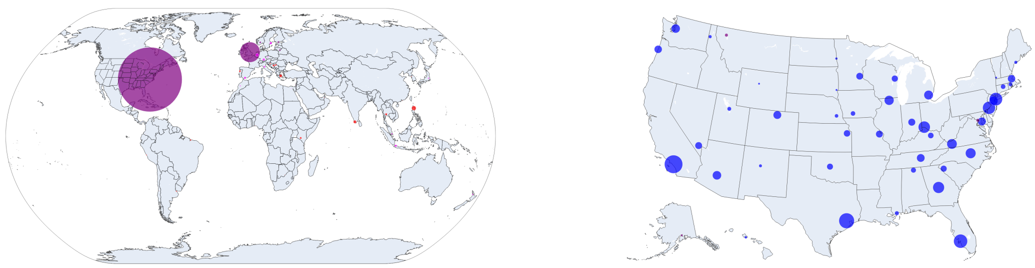
2 Online Informal Credit in a Forum

The informal credit community we study has about 128 thousand users, and its sole purpose is to be an online communication channel on Reddit where borrowers and lenders find each other. The members are identified by a username, which is very rarely identity-revealing. There are some community rules. In order to be a member of this community, an account used must be more than 90 days old, and needs to be active in other channels on the website. Posts requesting a loan are not allowed to be deleted by the borrower, and each user can post a maximum of once every 24 hours. By clicking a user's nickname others can observe the user's post and comment history in that and other communities. The sequential protocol of loan origination is: (1) a borrower requests a loan, (2) a lender chooses to fund

[3] This makes them obey banks' underwriting criteria (Evans 2019).

the loan, and (3) the borrower decides to repay the loan. We explain each step below.

Figure 1: Geographic Distribution of Loan Requests



Note: This figure shows the geographic distribution of loan requests. On the left panel, we show how loan requests are distributed around the world. On the right panel, given the predominance of the United States, we show the state-level geographic distribution across the country. The size of the circles represents the amount of loan requests in a given geography.

2.1 Borrower Requests a Loan

A potential borrower posts on the forum requesting a loan. They typically list the requested loan amount, promised repayment amount, repayment date, location, and a short description of the loan purpose. This post requesting a loan has strict formatting requirements, which allows the community to have an automated moderation tool (or bot), for parsing and checking posts. This bot will automatically delete requests that do not meet the formatting requirements and requests made by users who do not satisfy the membership standards. Figure 1 shows the geographic distribution of loans. Despite being very concentrated in the United States, loan requests are widespread across the different states.

2.2 Lender Funds a Loan

After a loan request is made, a potential lender will then see the post requesting a loan, and arrange the loan with the potential borrower. The lender can ask for any information they deem necessary to vet the borrower, while the borrower is free to choose what information they give or don't give to the lender. The two parties negotiate privately, and if they agree to a loan, the lender will comment on the request post stating they gave a loan to the borrower in the required format. The borrower then confirms they obtained a loan from the lender. From the moment that the two endpoints of the transaction confirm the existence of the loan, and independently disclose matching loan amounts, the loan is outstanding in a public ledger.

2.3 Loan Repayment is Due

Once a loan is due, the borrower is expected to repay it. The lender will return to the forum to report whether the borrower repaid the loan or not. Borrowers are ruled to be banned from the community in the case of no repayment. There are three main reporting tools available for participants in this market, funded on a donation basis. First, there is a list of users that represent scam threats. Users can search this list for a given username, and learn whether it was verified as a potential scammer. Second, a lender legitimacy checking tool, where borrowers can check whether a potential lender is an active lender in the community and whether they are in good standing. Lastly, there is a credit registry where all the loans paid, unpaid, and outstanding are disclosed. This can be used to verify the lending history of lenders, the credit history of borrowers, or even to infer default rates in the market as a whole.

In this market, there is no incentive for compliance with the rules other than user reputation and the ability to access credit via the community in the future. There is also no recourse in the event of default, beyond being banned from the community, and perhaps being reported to the payment processing provider. Since loans do not have any collateral, we assume a loss of 100% in the case of default. This loss given default is consistent with anecdotal evidence from posts and comments about unpaid loans and lenders referring to borrowers "disappearing" or "vanishing".

Figure 2: Example loan request

Posted by [username] x hours ago

[REQ] – ($350) (#St Louis, MO, USA), (Repay $400 by 4/22/2022), (PayPal, CashApp)

Got overzealous with bills and thought I could be all caught up and boy I was very wrong. Thanks

Note: This figure summarizes an example of a loan request in the online forum. We can observe the username of the potential borrower, the date and time of the post, the title of the post obeying strict formatting rules, and the body of the post. In the title, we can identify a flag for loan requests, the loan request amount, the borrower's location, and repayment terms.

2.4 Data Collection

Strict rules on post formatting allow us to collect data and parse needed loan information systematically. An example of a typical loan request on the forum is shown in Figure 2. The bracketed structure of the post title gives us an indicator of a loan request, "[REQ]", the loan requested amount, the borrower's self-reported location, and offered repayment terms. The auto-moderation bot cleans up improperly formatted postings, or posts by users who do not meet the minimum requirements of account history, or activity in other communities.

We collect all posts from this community and filter for all loan requests. For each request, we collect the username of the user requesting the loan, the date of the request, the requested amount, the poster's location, the promised repayment amount, the promised repayment date, the system that is used to transfer money (Paypal, Venmo, etc), and the text of the request. The latter may contain a short narrative about why the loan is needed.

Additionally, we obtain the credit registry database that stores all loans originated in the market. From this database, we can extract the username of the borrower and lender, the amount lent, the date of the loan, and whether a loan was marked as paid or unpaid. We can merge the requests and loans.

Our full dataset contains 168,317 loan requests for the period 2016-2021. Out of these requests, 111,670 are not removed due to the violation of community rules, out of which

51,020 requests have complete information on loan amount, promised repayment amount, and promised maturity, allowing us to compute an interest rate on the loan. Some requests are not converted into loans. This brings our sample to 35,066 loans we can accurately price with promised repayment date and actual repayment date. These loans occur between 1,962 unique lenders and 8,544 unique borrowers, totaling more than 7.5 million dollars in informal credit. Compared to the entire population of loans originated, our sample can capture 65% of the unique lenders, 56% of the unique borrowers, and 30% of the total value of loans.

While the negotiation process between borrower and lender is done in private without any intermediary, there is no reason to believe the lender would receive less in repayment than the offered repayment in the loan request. Likewise, there is no reason for borrowers to repay more than the already high price, mentioned in the request. Most posts have only one lender commenting to let the potential borrower know that they privately messaged them and later confirmed the loan. Therefore, we believe the information contained in loan requests provides a good measure of the credit terms in this marketplace.

Panel A: Loan requests with complete information (N=51,020)

	Mean	**St. Dev.**	**p10**	**p50**	**p90**
Requested Amount ($)	273	401	40	150	600
Promised Maturity (Days)	23	26	4	15	46
Promised Interest (% per month)	107	135	10	45	438
Converted Into Loan	68%				

Panel B: Loans merged with requests (N=35,066)

	Mean	**St. Dev.**	**p10**	**p50**	**p90**
Loan Principal ($)	216	300	35	125	500
Effective Maturity (Days)	25	44	3	13	52
Promised Interest (% per month)	131	160	15	57	519
Default Lower Bound	4.2%				
Default Upper Bound	10.6%				

Table 1: Summary Statistics. This table contains summary statistics for proposed loan terms in loan requests (Panel A), and loan terms and default bounds for funded loans (Panel B). Default lower bound reflects the percentage of loans marked as unpaid. Default upper bound reflects the complement of the percentage of loans marked as paid.

3 Descriptive Analysis

In Table 1, we report summary statistics for loan requests and loans. Out of 51,020 requests with complete and accurate information, we can find 35,066 corresponding loans. The average amount requested is $273, with the median at $150. The mean and median amount lent is less than the mean and median amount requested, suggesting that lenders are not willing to grant high-dollar value loans, some credit rationing, or screening on the extensive margin of credit. Interestingly, promised maturity is lower than realized maturity, due to some late payments. Finally, loans have a slightly higher average and median interest rate than requests. This corroborates screening on the extensive margin, whereby lenders choose to lend to those most profitable to them.

We cannot observe the true default rate in this market. When a loan defaults, a lender is expected to mark the loan as paid or unpaid. However, without an explicit incentive for record-keeping, we do see that lenders do not always mark whether a loan results in repayment or default. We have about 11% of our matched loans without data on whether a loan defaulted or was repaid. To overcome this, we calculate an upper and lower bound of default. The lower bound is defined as the percentage of loans marked as unpaid by the lender. The upper bound of default is one minus the percentage of loans marked as paid, as all the loans not marked as paid could potentially be unpaid loans. Figure 3 plots the upper and lower bounds of default over the sample period and shows that these bounds generally co-move, and the spread between them is small. The difference between the average default rate in the upper bound and the average default rate in the lower bound is about 6%.

Appendix Table A1 shows loan summary statistics broken down by year. Overall, average loan size and average interest rate have increased over time, but the average maturity of a loan has stayed relatively constant. Of note, in 2019 we observe a large spike in the number of loans in our sample. This is due to the community enforcing stricter rules on loan requests, allowing us to successfully parse more loans beginning in 2019.

The average promised interest rate on requests is 107% per month, and the average

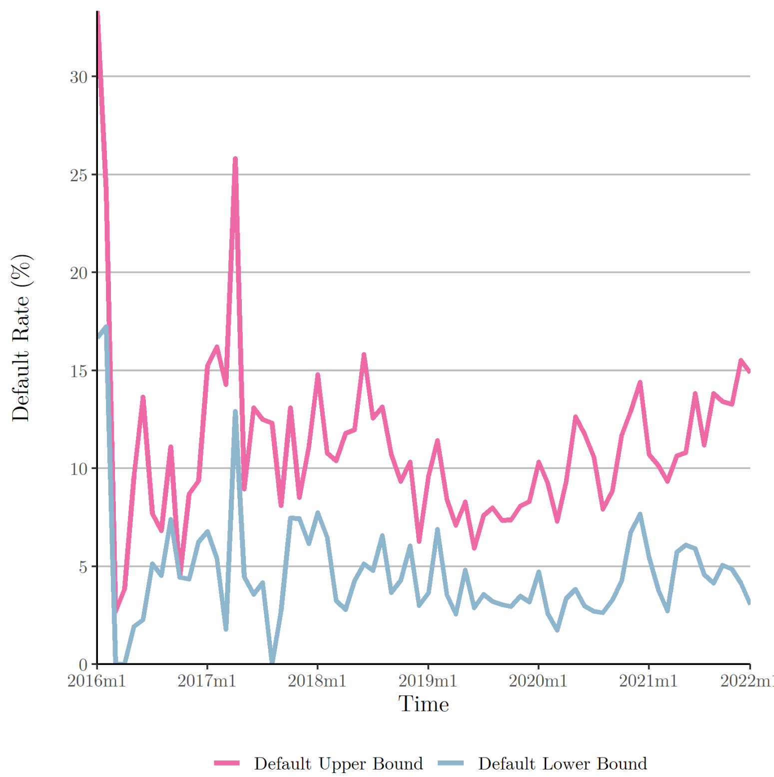


Figure 3: Default Bounds. In this figure, we plot the time series of default rates in this market. In red, we plot the upper bound for default, which includes both those loans marked as unpaid, and those not marked at any time as defaulted loans. In blue, we plot the lower bound for the default rate, which includes only loans marked as unpaid.

interest rate on loans is 131% per month. These values are highly affected by extreme values from very short-maturity loans. However, even when considering the medians of 45% and 57% per month, respectively, these interest rates are very large.

This leads us to the high default rates in this market. With an average maturity of about 25 days, and considering the midpoint of the default rate bounds of 7%, these loans are highly risky. While this is an economically large default rate (68% when annualized), the corresponding break-even interest (15%) rate would be substantially lower than the observed median interest rate of more than 50% per month.

Our evidence suggests that the risk-sharing and loan-pooling capabilities of intermediaries, coupled with the monitoring role of collateral lead to less risky and more accessible consumer credit. For example, despite being a high-seniority claim on a worker's salary, payday loans are some of the most expensive and riskier types of unsecured consumer credit. Even when compared to payday loan interest rates, which range between 300-400% (Correia, Han and Wang 2023) per year, these loans are much more expensive and riskier. As pointed out in Correia, Martins and Waikel (2022), interest rates in this market exhibit an inverted term structure, usually associated with times when financial distress is ahead.

	Mean Borrowers	Mean Lenders	Difference	St. Dev. Borrowers	St. Dev. Lenders	Median Borrowers	Median Lenders
A. User Statistics							
Male	60%	86%					
Age (Median)	25-34	25-34					
College? (Median)	Some	Bachelors					
Income (Median)	\$50-\$60k	\$100k +					
Employment (Median)	Part Time	Full Time					
Has Savings Account	59%	93%	$t = 4.3$				
Has Retirement Account	37%	80%	$t = 4.8$				
Has Investment Account	22%	85%	$t = 8.4$				
Has Mortgage	10%	53%	$t = 4.8$				
Has Personal Loans	42%	14%	$t = 3.1$				
Evaluated Fin. Skills (/4)	3.56	3.94	$t = 3.2$	0.72	0.23	4	4
Self-Rep. Fin. Skills (/7)	3.95	5.52	$t = 5.5$	1.40	1.27	4	6
Financial Satisfaction (/7)	2.77	5.35	$t = 7.3$	1.75	1.57	2	6
Financial Attitude (/7)	5.50	5.89	$t = 1.8$	1.06	1.03	6	6
Financial Skills (/4)	3.01	3.43	$t = 1.9$	1.29	0.93	3	4
B. Loan Characteristics							
Requested Amount	\$169	\$249	$t = 8.8$	\$283	\$378	\$100	\$150
Promised Maturity	18.5	19	$t = 0.9$	22.2	27.2	11	12
Actual Maturity	25	27	$t = 1.9$	41.5	46.9	13	13
Promised Interest (% per month)	131%	122%	$t = 2.0$	157%	154%	57%	52%

Table 2: Borrower and Lender Demographics. This table shows descriptive statistics from a survey conducted on 100 participants in the informal market we study. 53 participants are lenders, and 47 participants are borrowers. We surveyed the participants on their demographics, their financial knowledge, and their perceived financial knowledge. In the third column, we have t-tests for the differences in mean outcomes of continuous variables between borrowers and lenders.

With monetary incentives in place, we also survey 100 participants in this market, of which 53 are lenders and 47 are borrowers, randomly sampled. This allows us to understand better some prevalent demographic patterns of market participants. In Table 2, we report the results, as well as t-tests for the differences between borrowers and lenders.

The most predominant age range is 25-34 years old. On the lender side, the male predominance is much heavier than on the borrower side (86% vs. 60%). The educational attainment is high with the median achievement in both groups being some college attendance. Lenders, though, predominantly have bachelors degrees or higher. While more than half of the lenders are employed full-time, that median is part-time employment for the borrowers. All in all, this results in the median income being significantly higher for lenders (\$100k +) than for borrowers (\$50-60k).

Household balance sheets are also statistically different. Most lenders have savings, retirement, and investment accounts. The same is not true for borrowers: only 6 in 10 borrowers have a savings account, a third has a retirement account, and 1 in 4 reports having an investment account. While more than half of the lenders have a mortgage, only a sixth have personal loans. The relationship flips for borrowers: 10% have a mortgage, and 42% have a personal loan. Together with substantial differences in their income and human capital, these differences in financial outcomes suggest a much higher financial vulnerability for borrowers than for lenders.

That intuition is confirmed using questions from the FINRA National Financial Capability Study survey. When we compare the scores of financial literacy, skills, and attitudes between borrowers and lenders, lenders score significantly higher. They enjoy not only higher levels of financial satisfaction but also perceive themselves to be more financially knowledgeable. Lastly, the economically similar loan terms between the average borrower and the average lender lend credence to the survey in that we believe the two groups are likely to represent the individuals who interact in this market.

Beyond borrower-lender differences, there is heterogeneity in the lender base, illustrated in Figure 4. While the majority of lenders only fund 1 or 2 loans, there is a minority at the upper tail of the distribution, who has originated up to thousands of loans. Moreover, in Panel B of Figure 4, we see that lender profits bunch at zero. Assuming lower bounds of default as actual default, the market is mechanically more profitable, than assuming upper bounds of default. However, in both extremes, the distribution exhibits the same shape. The online informal loan market is profitable on average, but access to those profits has barriers.

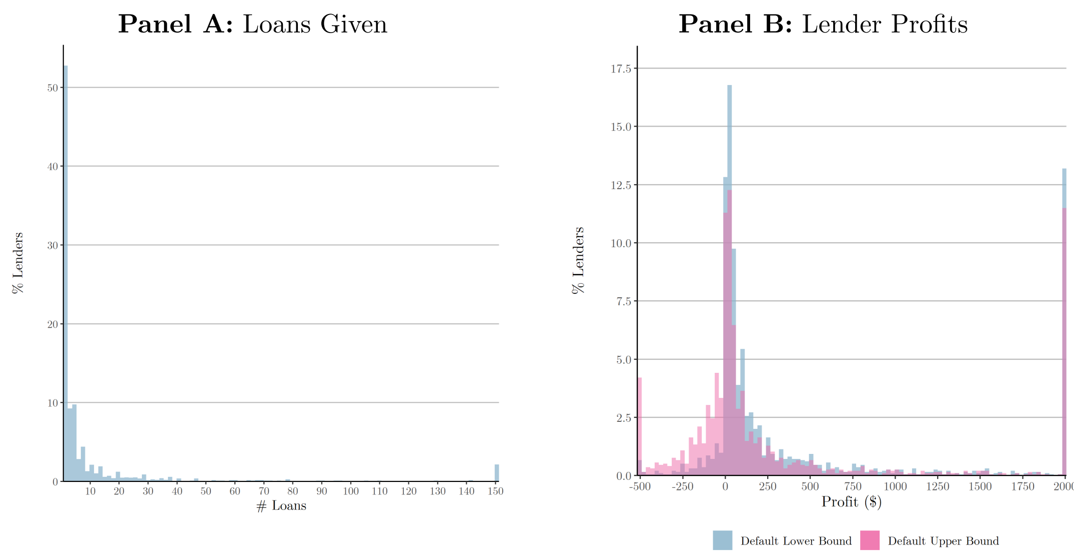


Figure 4: Lender Heterogeneity. In Panel A, we plot the distribution of the number of loans (truncated at 150) given by lenders. In Panel B, in grey, we plot the distribution of profits (truncated at $2000) for lenders, assuming the lower bound of default as the actual default. In red, we plot the distribution of profits (truncated at $2000) for lenders, assuming the upper bound of default as the actual default.

Only a minority of lenders has experienced large profits as shown by the truncated upper tail.

In Figure 5, we study this segment of lenders who appear to be highly present and successful in the market. In Panel A, for every year, we consider the top 10 lenders in the year before, defined based on profitability. We confirm that for the average lender in the market, it is not profitable, as the time series of cumulative returns consistently lies below zero. For the top 10 lenders in the previous year, the cumulative return is very high, consistently trending up, ending at around 500% in 5 years. The plot has two main takeaways: (i) lender performance is persistent; and (ii) our inability to perfectly track default has a minor impact when we are measuring aggregate market outcomes, since time series that assume lower or upper bounds of default exhibit the same behavior. Performance persistence is expected, especially since most lenders who experience a default leave the market, therefore leaving successful lenders operating for a long period.

In Panel B, we plot the evolution in market shares of the all-time top 10 lenders throughout our sample period. Since 2018, at each point in time, the top 10 lenders represent 5-10%

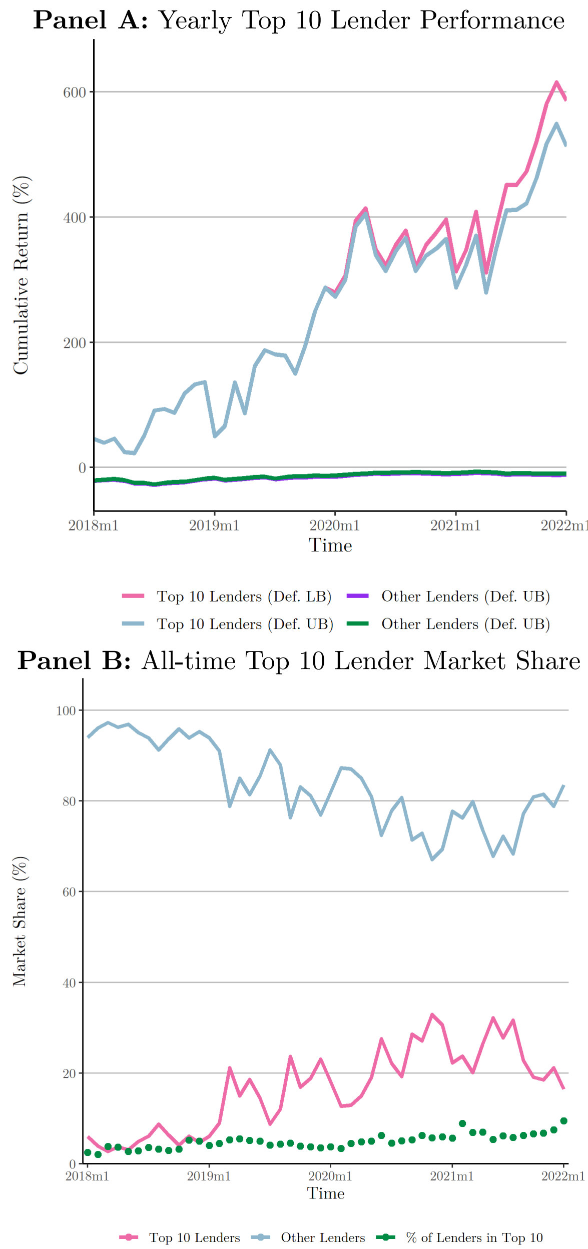


Figure 5: Top 10 Lenders. In Panel A, we plot the cumulative return of the loan portfolio of the top 10 lenders in the market, and for the other lenders in the same time period. Top 10 Lenders are defined in each year, as the 10 lenders who profited the most in the previous year. In Panel B, we establish the comparison for market shares between all-time top 10 lenders, and other lenders.

of the lender base, as illustrated by the green dots. However, they have occupied 20-30% of the market. Altogether the two panels even suggest that this minority of lenders is growing steadily, and picking an extremely profitable part of the market.

The disparity in profitability between the top 10 lenders and the rest of the market can

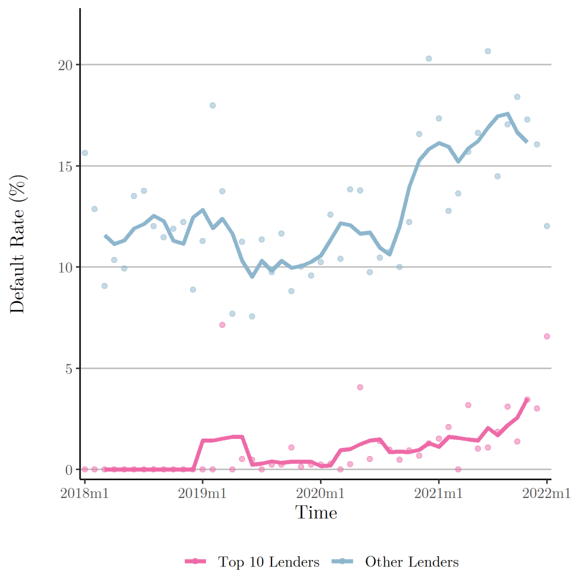


Figure 6: Top 10 Lenders Default Rates. We plot the average default rates return of the loan portfolio of the all-time top 10 lenders in the market, and for the other lenders in the same time period.

have two main drivers. One is that for the same credit risk, top lenders pick loans that compensate them at a higher rate. Another is that they are simply better at managing risk, and picking safer loans. In Figure 6, we plot average default rates (midpoint between the upper bound and lower bound) for each calendar month, for all-time top 10 lenders, and others. We find a clear superiority of top lenders managing default risk, or at least picking safer investments. Towards the end of the sample, default rates are mechanically inflated because of the upper bound of defaults, which assumes that loans (not yet) market as paid or unpaid are defaulted.

Our descriptive evidence suggests that each loan individually is very risky, but pooled, these loans pay a rate substantially higher than the break-even rate. This opens space for intermediation to create value, and efficiently pool risk and allocate financial resources. Additionally, a small and sophisticated fraction of lenders exhibits outstanding performance, stemming from a differential ability for credit risk management and screening. Other lenders do not enjoy the same stability and profitability in their portfolios. The financial system provides screening and risk management tools that individuals in this market do not have access to. Lastly, we can speculatively elaborate on the role of collateral. Intermediaries provide a legal and institutional framework for collateral to be posted. Collateral diminishes losses given default, attenuating credit pricing. It also acts as a monitoring device, lowering default.

4 Stylized Framework

In this section, we aim to understand what mechanisms are at play in this market. It has two features in which it differs from intermediated consumer credit. First, recall that demand is revealed and borrowers name their price. Second, there is no collateral or recourse in case of default. Lenders choose whether to fund a loan, given the terms provided by the borrower. We formalize below the interaction between a representative borrower and a representative lender. We follow the loan origination steps described in Section 2.

4.1 Borrower's Loan Request

The borrower needs an amount A_0 in the current period, and anticipates needing A_1 at a future period. Without an assumed future need for credit, the borrower would have no incentive to repay and maintain a reputation.

The borrower also faces a cost for borrowing outside of this market a_b. This cost can represent both monetary and non-monetary costs. Monetary costs can be the actual interest charged by an outside lender, or the monetary search costs for alternatives. Non-monetary costs can be the stigma stemming from borrowing from family and friends, entering a payday loan shop, or even letting go of property under a title or pawn loan.

As illustrated by Figure 7, the borrower will request a loan offering a rate r_b, if borrowing in this market is more affordable than the alternative cost. Formally, if:

$$r_b < a_b \tag{LR}$$

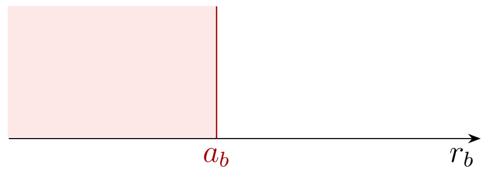


Figure 7: Borrower's Loan Request. In the axis for the interest rates that a borrower can offer, we mark the threshold below which the borrower is willing to request credit in this market. In the shaded area, below a_b, the borrower is willing to request credit in this market.

4.2 Lender's Funding Decision

Lenders face an origination cost of c_l per dollar of credit originated. Like a_b, this cost goes beyond monetary opportunity costs and can represent the time spent doing credit risk analysis, or the aversion to lending in an unsecured, non-intermediated, and unregulated market.

In this market, borrowers are protected by a vast layer of anonymity, making the market highly information asymmetric. We assume that lenders form their beliefs about the market, rather than the individual borrower. They believe there are *good* borrowers who never default, and *bad* borrowers, who default with probability d. Lenders then have their current period perceived proportion of bad borrowers μ_0.

Conditional on the amount requested A_0 and an interest rate r_b offered by the borrower, the lender decides whether to fund the loan. The lender funds the loan if they expect to make a profit on it:

$$A_0 + c_l A_0 < (1 - \mu_0)(1 + r_b)A_0 + \mu_0(1 - d)(1 + r_b)A_0 \tag{1}$$

Re-arranging Equation (1), the lender will fund the loan, if the rate offered by the borrower compensates them enough for their origination costs and perceived degree of credit risk.

$$r_b > \frac{c_l + \mu_0 d}{1 + \mu_0 d} \tag{LF}$$

As illustrated by Figure 8, a loan will exist if there is a rate that is low enough for the

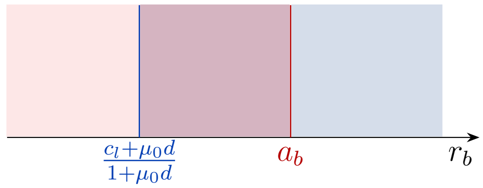


Figure 8: Lender's Funding Decision. In the axis for the interest rates that a borrower can offer, we mark the threshold below which the borrower is willing to request credit in this market. In the shaded area, below a_b, the borrower is willing to request credit in this market. In blue, we add the threshold above which the rate offered by the borrower needs to be so that the lender is willing to fund the loan. For offered rates in the intersection of the two, both the borrower and the lender are willing to contract on a loan.

borrower to initiate a request, but high enough to compensate the lender. This is shown in the figure by the intersection of the shaded blue and orange areas. Moreover, see that if the alternative borrowing market is very low-cost, the lenders perceive very high credit risk, or face high origination costs, the intersection might not exist. A loan exists if a request exists, satisfying (LR) and if it is funded (LF).

4.3 Borrower's Repayment Decision

At maturity, the borrower needs to decide whether to repay the loan. Recall that there is no seizure of collateral or any recourse in case of default. Reputation for future borrowing is the only device that incentivizes the borrower to repay.

The borrower believes that repaying the loan yields a reputation gain. Such reputation gain is reflected by lowering the lender's perception of the proportion of bad borrowers in the market. Using the next loan as a latent reality, the borrower gauges whether repaying the loan yields sufficient reputation gain so that the next loan is likely to exist.

The updated perceived proportion of bad borrowers in the next loan, $\tilde{\mu_1}$ is defined in Equation (2) as:

$$\tilde{\mu_1} = \mu_0 - b_{b,l} \tag{2}$$

where $b_{b,l}$ is the reputation gain upon the perception of the lender, given by borrower repayment. The borrower repays the current loan if the reputation gain creates a feasible interest rate interval for a potential next loan, via lowering the perceived proportion of bad borrowers. Formally, as illustrated in Panel A of Figure 9 the borrower repays the loan if:

$$\frac{c_l + (\mu_0 - b_{b,l})d}{1 + (\mu_0 - b_{b,l})d} < a_b \tag{RL}$$

Note that (RL) refers to a hypothetical future loan, where a_b can change for the borrower, and the lender can be a different lender, with a different c_l. If, like in Panel B of Figure 9, they believe that the reputational gain is not enough to lower the lender's minimum acceptable rate, then the likelihood of a future loan is very low, and the borrower is more likely to default.

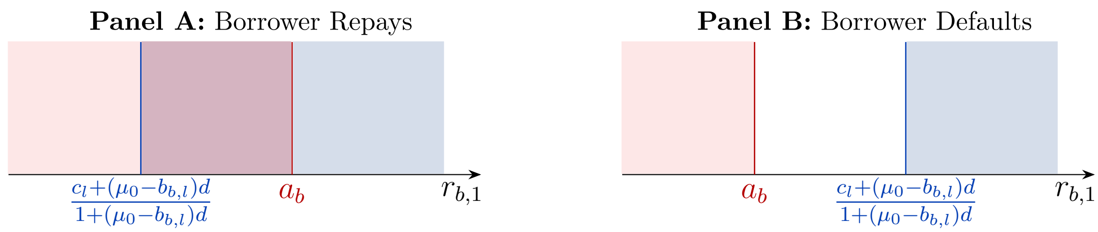


Figure 9: Borrower's Repayment Decision. The borrower repays the current loan if it is likely that a future loan can exist, given the update to the perceived proportion of bad borrowers (Panel A). In turn, if the reputational gain is not enough to create a feasible set of interest rates, the borrower does not have an incentive to repay the loan and, hence, defaults (Panel B).

Figure 9 shows that the loan is repaid (RL) if both (LR) and (LF) are likely to overlap in a potential future loan, incorporating the reputational gain from repaying the current loan.

Using the setup of the model, we can discuss the two previous empirical results. First, the high default rate in the market is justified by the absence of screening and monitoring devices beyond reputation. Similarly, the high cost of credit can be justified simultaneously by (i) rational expectations from lenders about high market credit risk and (ii) high-cost alternative credit solutions for the financially underserved.

Second, we can discuss lender heterogeneity in light of the model. For lenders with high origination costs, it is only worth lending if they can pick high-quality loans or have superior monitoring abilities. As lenders diversify their loan portfolio and repeat procedures, per-dollar origination costs are likely to decrease, and the marginal impact of each default in the dollar value of the portfolio also decreases.

We will make two key assumptions to derive testable hypotheses in the direction discussed above: (i) the origination burden is decreasing both with lender experience and with lender ability, and (ii) as borrowers and lenders are more familiar, borrower loan repayment yields larger reputational gains. We now re-arrange (RL) to relate these two quantities:

$$c_l < [a_b - (1 - a_b)\mu_0 d] + [(1 - a_b)d]\, b_{b,l} \tag{3}$$

This inequality in Equation (3) for the origination cost is linear in the reputational gain. Given an alternative borrowing cost, and the lender's perception of credit risk in the market, it creates a minimum reputational gain that justifies the borrower repaying the loan, for each level of lender origination costs. Figure 10 illustrates this relationship. The larger the origination cost faced by the lender, the larger the reputation gain needs to be so that the borrower repays their loan.

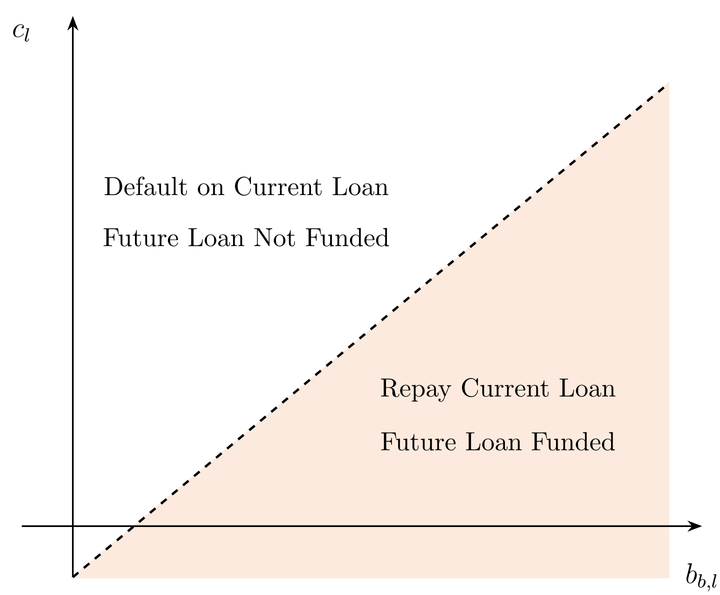


Figure 10: Origination Costs and Reputation Gains. In this graph, we use the overlap in conditions in Panel A of Figure 9 to relate origination costs and reputational gains from loan repayment. For each level of lender origination costs, there is a minimum reputational gain that needs to occur for the borrower to repay the loan. The dashed line and the shaded area under it represent that relationship.

The plot helps us formulate the key dimensions across which lenders or lender-borrower pairs can differ to generate the heterogeneity observed in Section 2. With our two key assumptions, we can derive testable hypotheses to assess whether our stylized framework accurately describes decision-making in the absence of intermediaries:

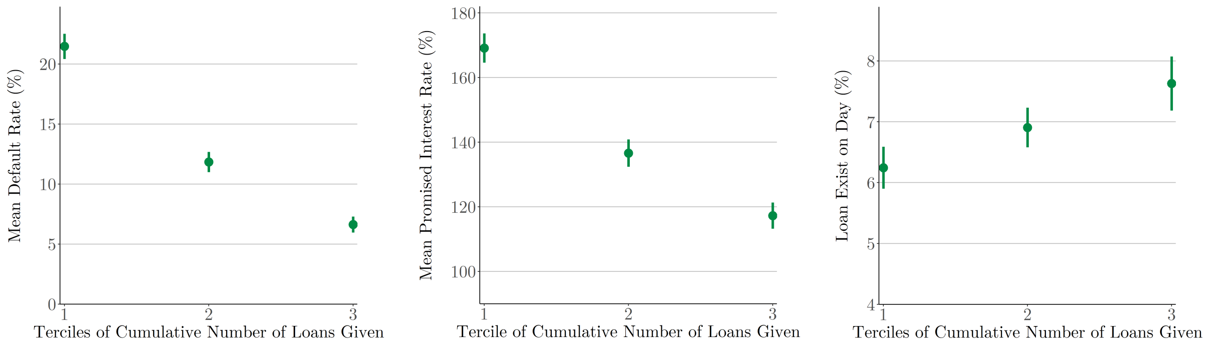
Origination Burden Hypothesis. If lender ability or lender experience decrease unit costs of origination, then (i) we expect to observe lower default, and higher propensity to fund loans; and (ii) the condition (LF) would also allow for lower prices offered by more experienced/skilled lenders.

Borrower-Lender Opacity Hypothesis. If a larger reputation gain arises when there is less opacity between borrowers and lenders, we expect to observe lower default and a larger propensity to fund between pairs of more familiar borrowers and lenders.

5 Origination Costs: Experience vs. Skill

According to our first hypothesis, lenders who experience a lower origination burden would accept loans with lower promised interest rates, experience less default and exhibit a higher propensity to lend, on average. The origination burden, we argue, decreases as the lender becomes more acquainted with the workings of the market, and as they learn the credit risk evaluation and due diligence processes. Hereafter, we refer to this as the experience channel. It can also be a lender-specific trait, whereby they are more skilled or have higher abilities to obtain hard or soft information on borrowers, reducing the burden of lending. For example, it takes them less time to analyze a potential loan. We refer to this second channel as the skill channel.

Our approach to understanding whether and how each channel plays a role is to start with a measure that can proxy for either mechanism and use different econometric specifications to test different mechanisms. Our relevant variable is the total number of loans a lender has funded until the date of each loan they participate in, a lending activity history variable. The underlying assumption is that each past loan contributes to lowering the origination



Panel A: Default (p.p.) Panel B: Interest Rate (p.p.) Panel C: Propensity to Lend (p.p.)

Figure 11: Average Outcomes by Tercile of Lending Activity. In this figure, we plot averages and 95% confidence intervals for default rates (in Panel A), monthly interest rates (in Panel B), and propensities to lend (in Panel C) for lenders in the first, second, and third terciles of past lending activity.

cost of the next loan. This can happen because as a lender lends more they become more experienced, or because more skilled lenders will lend more successfully, and in turn have higher lending activity.

Regardless of the mechanisms at play, for our hypothesis to match empirical evidence, we should observe a negative relationship between past lending activity and default and interest rates, and a positive relationship between past lending activity and propensity to lend. Figure 11 visually confirms that this pattern is present in the data. We estimate Equation (4) to validate our measure, where we can condition on observables:

$$y_{i,l,b,t} = \alpha + \beta CumulativeLoans_{l,t} + \phi X_{b,t} + \theta X_i + \varepsilon_{i,l,b,t} \tag{4}$$

where $y_{i,l,b,t}$ is the outcome of loan i, between lender l and borrower b, originated at time t, $CumulativeLoans_{l,t}$ is the measure of past lending activity for lender l until day t, $X_{b,t}$ is a vector of time-varying controls for borrower characteristics and X_i is a vector of controls for loan proposed terms. Borrower controls include whether it is the first loan for this borrower, whether they have ever defaulted, whether they have ever paid their loans late, their number of loans outstanding, and their total number of past loans. The vector of controls for loan request terms includes maturity, the complexity of text, readability, and misspelling measures about the loan request text, and depending on the dependent variable, other controls such as requested and promised repayment amounts.

	(1)	(2)	(3)	(4)
	Panel A: Default Rate (p.p.)			
Cumulative Loans	-1.066***	-0.854***	-0.683***	-0.554***
	(0.083)	(0.078)	(0.074)	(0.074)
N	17,558	23,864	17,558	23,864
R^2	0.021	0.018	0.076	0.077
	Panel B: Interest Rate (p.p.)			
Cumulative Loans	-3.492***	-3.265***	-1.961***	-1.623***
	(0.572)	(0.539)	(0.442)	(0.419)
N	17,558	23,864	17,558	23,864
R^2	0.010	0.011	0.236	0.233
	Panel C: Propensity to Lend (p.p.)			
Cumulative Loans	0.047**	0.245***	0.122***	0.267***
	(0.016)	(0.026)	(0.010)	(0.019)
N	439,119	411,561	439,119	411,561
R^2	0.000	0.003	0.656	0.581
Sample	Sub	Full	Sub	Full
Controls	No	No	Yes	Yes

Table 3: Past Lender Activity. This table reports regression estimates for the number of previously funded loans by the lender, on a default indicator in Panel A, interest rate in Panel B, and on a dummy that flags request-to-loan conversion. In Panels A and B, the unit of observation is a loan, and in Panel C, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except Panel B), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling, readability, and complexity of the loan request text. The first two columns exclude controls, and odd-numbered columns do not contain loans privately prearranged between borrower and lender. All standard errors are clustered by lender, and presented in parentheses.

In Table 3, we report estimates for the coefficient of interest, β. We run the specification above with and without controls. Our preferred sample does not include loans marked as previously arranged loans, as we do not observe private communications between lender and borrower, which define the loan terms. For robustness, we also present estimates on a sample including prearranged loans. Loan terms for prearranged loans are imputed from previous loan terms between borrower-lender pairs and are by definition less sensitive. Panels A and B confirm that lending activity is associated with lower default and accepting loans with lower rates.

In Panel C, we analyze the propensity to fund a loan, from the lender perspective. We build a sample with all pairs of possible borrower-lender pairs from those who borrow or lend on any given day. For each lender, we consider the set of borrowers requesting loans on a given day to be their choice set on that day. We then use a dummy variable equaling one for the loans that happened on that day, and zero otherwise. We then measure how the propensity for a request to be funded associates with lender experience, conditional on the two vectors of borrower and loan request controls described above. Higher past lending activity is also associated with a higher propensity to lend.

These findings assert that our proxy for monetary and non-monetary origination costs are associated with the three key variables in the way predicted by our stylized framework. More interesting is to understand whether experience and learning or skill and ability are determining this relationship. We start by adding week fixed effects to the previous specification. Doing so eliminates lender-specific time variation in lending activity, shutting down the experience acquisition channel. In this specification, β measures how the outcomes change across individuals with different levels of past lending activity. Table 4 reports the results.

The results suggest that cross-sectional differences in past lending activity drive our results. Our preferred specification is that of column (3) since it contains controls and eliminates some of the omitted variable bias, and it does not contain prearranged loans, for which most of the loan term data is imputed. The cross-sectional results are all statistically significant and stronger in magnitude. Economically, this means that even not allowing for the experience channel to operate, lenders with higher past lending activity in a given week accept lower interest rates, experience lower default, and are more likely to fund a loan, than lenders with lower past lending activity.

Alternatively, to absorb time-invariant differences in lender skill, we add lender fixed effects to our baseline specification. Here, the variation of interest is that of the same lender along the time series. This approach absorbs pre-existing differences in the burden of loan

	(1)	(2)	(3)	(4)
	Panel A: Default Rate (p.p.)			
Cumulative Loans	-1.241***	-0.997***	-0.861***	-0.690***
	(0.090)	(0.070)	(0.076)	(0.064)
N	17,558	23,864	17,558	23,864
R^2	0.057	0.048	0.112	0.107
	Panel B: Interest Rate (p.p.)			
Cumulative Loans	-4.249***	-3.987***	-2.467***	-2.083***
	(0.600)	(0.548)	(0.448)	(0.417)
N	17,558	23,864	17,558	23,864
R^2	0.044	0.038	0.255	0.247
	Panel C: Propensity to Lend (p.p.)			
Cumulative Loans	0.075***	0.267***	0.134***	0.277***
	(0.014)	(0.024)	(0.008)	(0.018)
N	439,119	411,561	439,119	411,561
R^2	0.018	0.021	0.658	0.584
Sample	Sub	Full	Sub	Full
Controls	No	No	Yes	Yes
Week FE	Yes	Yes	Yes	Yes

Table 4: Cross-Lender Variation in Past Lending Activity. This table reports regression estimates for the number of previously funded loans by the lender, on a default indicator in Panel A, interest rate in Panel B, and on a dummy that flags request-to-loan conversion. In Panels A and B, the unit of observation is a loan, and in Panel C, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except Panel B), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling, readability, and complexity of the loan request text. The first two columns exclude controls, and odd-numbered columns do not contain loans privately prearranged between borrower and lender. All columns were estimated using week fixed effects. All standard errors are clustered by lender, and presented in parentheses.

origination. Table 5 reports the results with and without borrower and loan request controls, for the sample without pre-arranged loans. If the origination costs decrease as lenders get more experienced with lending, our hypothesis indicates that we should observe estimates for β with the same sign as in Tables 3 and 4.

We find the role of experience in lowering origination burdens to be undetectable. Focusing on the even-numbered columns of Table 5, where the specifications include controls, we find that as the same individual lends more, they are more likely to encounter default. This is mostly mechanical, as we are using default as the complement of loans marked as paid, and as the pool of loans increases, the percentage of loans that go as unmarked increases. However, it rejects a negative relation between within-lender experience and default. In column (4), we also observe a non-result on the correlation between experience and loan pricing. The only pattern that holds is that of propensity to lend, in column (6). Here, it could be a mechanical relationship or an actual effect of experience acquisition through lending activity, on the extensive margin of lending. Not only we are unable to disentangle a mechanical effect, but also, statistical significance is lower, and the magnitude of the estimate is less than a fifth of the estimate allowing for skill to play a role.

	(1)	(2)	(3)	(4)	(5)	(6)
	Default Rate (p.p.)		Interest Rate (p.p.)		Prop. to Lend (p.p.)	
Cumulative Loans	0.171	0.286**	-2.508**	-1.077	-0.017	0.049**
	(0.095)	(0.094)	(0.773)	(0.642)	(0.041)	(0.019)
N	17,558	17,558	17,558	17,558	439,119	439,119
R^2	0.299	0.326	0.243	0.362	0.012	0.659
Sample	Sub	Sub	Sub	Sub	Sub	Sub
Controls	No	Yes	No	Yes	No	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Time Variation in Past Lending Activity. This table reports regression estimates for the number of previously funded loans by the lender, on a default indicator, interest rate, and on a dummy that flags request-to-loan conversion. In columns 1-4, the unit of observation is a loan, and in columns 5-6, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except columns 3-4), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling, readability, and complexity of the loan request text. Odd-numbered columns exclude controls, and all columns do not include loans privately prearranged between borrower and lender. All columns were estimated using lender fixed effects. All standard errors are clustered by lender, and presented in parentheses.

Thus far, our analysis rules out the experience channel, and highlights the role of skill in managing credit risk, while decreasing the price of credit, and increasing likelihood of credit origination. Skill being a key determinant of loan terms is consistent with our descriptive results presented in Figures 5 and 6. We address the robustness of our result in terms of measurement, sampling, and estimation in Section 7.

The results in this section speak to the role of skilled lending in decreasing default rates, decreasing the price paid for credit, and increasing credit availability. An advantage of intermediated credit, relative to a market like the one we analyze is human capital. Financial intermediaries employ highly skilled and financially literate staff and use complex models to undertake loan origination, credit pricing, and funding decisions. Therefore, another role that intermediaries play in consumer credit is that of bringing skill to the consumer credit market.

6 Reputation Gain: Borrower-Lender Connectedness

According to our second hypothesis, pairs of borrowers and lenders with a thinner layer of opacity are more likely to originate a loan, and that loan is more likely to be repaid. We argue that lenders' ability to learn from borrower repayment leads to a larger reputation gain from borrower repayment. Thus, it incentivizes repayment, and therefore loan origination. This reasoning is very akin to relationship lending.

A substantial amount of literature has shown that firms, particularly small firms, benefit tremendously from a degree of proximity to their lenders. Elyasiani and Goldberg (2004) comprehensively review earlier literature. More recent works (Kysucky and Norden 2016; López-Espinosa, Mayordomo and Moreno 2017; Banerjee, Gambacorta and Sette 2021) use microdata to confirm the intuition that banking relationships reduce opacity between borrowers and lenders and allow for soft information acquisition focusing on small firms. Holmes, Isham, Petersen and Sommers (2007) study relationship lending in consumer credit, in the auto loan space, and posit that it is especially important for low-income and low credit score individuals.

In our setting, we do not rely on the lending history between borrower and lender to proxy for the lender's ability to acquire soft information or for the asymmetry of information on each loan. Instead, we measure how connected each pair of borrower and lender is through their activity in other online communities of the same online platform. If the borrower and lender share a lot of their interests in sports, food, or other hobbies, and coexist prominently in other communities, then their relationship is less opaque.

We measure connectedness in two ways: in scope and intensity. Our scope measure hinges solely on the extensive margin of whether borrower and lender post or comment in a given community and measures the overlap. Intensity uses the percentage of their activity that goes into each community. We exclude any posts or comments in the credit marketplace as that would induce undesirable connectedness in our measure. In Equation (5) we express how we construct each measure:

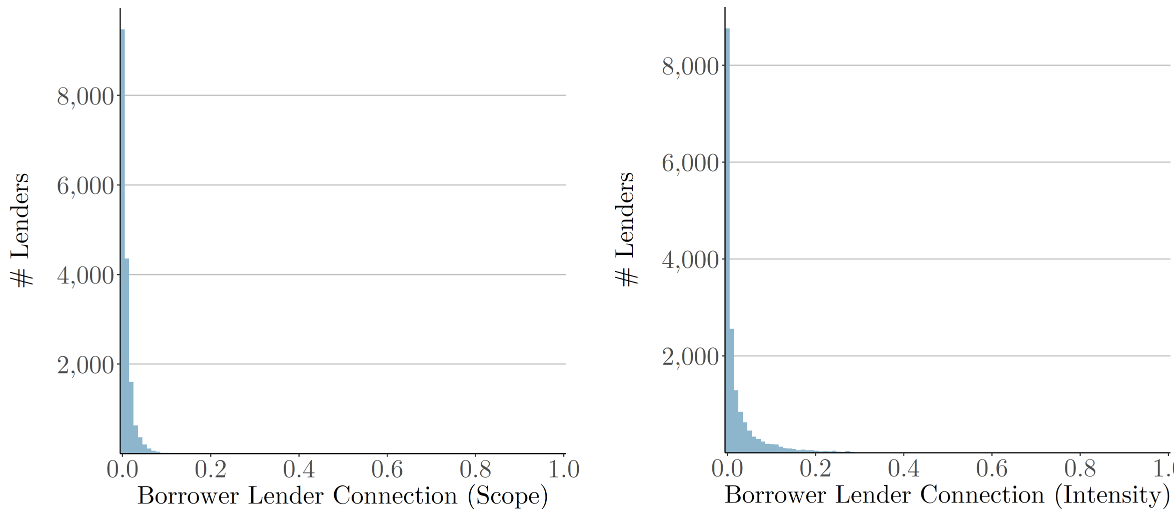
$$Scope_{b,l,t} = \frac{\sum_c I_{c,t}(b,l)}{\sum_c I_{c,t}(b)} \times \frac{\sum_c I_{c,t}(b,l)}{\sum_c I_{c,t}(l)} = \%ScopeBorrower_{b,l,t} \times \%ScopeLender_{b,l,t} \tag{5}$$

where $I_{c,t}(X)$ takes the value of one if the users in the set X have ever posted or commented in the community c as of time t. It measures the percentage of communities visited by the borrower that are also visited by the lender, multiplied by its counterpart. It gives the same weight to all communities, despite of how much time, energy, and participation each user devotes to each community. Our intensity measure in Equation (6) takes that into account:

$$Intensity_{b,l,t} = \frac{\sum_{c \in L} N_{c,t}(b)}{\sum_{c \in B} N_{c,t}(b)} \times \frac{\sum_{c \in B} N_{c,t}(l)}{\sum_{c \in L} N_{c,t}(l)} = \%IntensityBorrower_{b,l,t} \times \%IntensityLender_{b,l,t} \tag{6}$$

where $N_{c,t}(X)$ denotes the number of comments or posts the users in the set X have ever done in the community c as of time t, the sets B and L represent the sets of communities where the borrower and the lender, respectively, have any activity as of time t. It measures the percentage of posts and comments by the borrower in communities where the lender is also active, multiplied by its counterpart. Both measures can vary between zero and one.

In Figure 12, we plot the distribution of lenders' average connectedness across all their loans. While most lenders' activity does not have a substantial overlap with the borrowers they give credit to, we do see some variation in the extent to which their activity overlaps, particularly on the intensity measure. This suggests that lenders who are marginally



Panel A: Connectedness Scope

Panel B: Connectedness Intensity

Figure 12: Histograms of Average Connectedness Measure by Lender. In this figure, we plot histograms of average connectedness scope (in Panel A) and intensity (in Panel B) by lender.

connected with a potential borrower, are at an advantage relative to the majority of lenders.

We use these connectedness measures to test our second hypothesis. We do so using a similar specification as in Section 5.

$$y_{i,l,b,t} = \alpha + \beta Connectedness_{l,b,t} + \phi X_{b,t} + \theta X_i + \varepsilon_{i,l,b,t} \tag{7}$$

where we replace $Connectedness_{l,b,t}$ with our different measures of connectedness, and all else is defined as before. The outcomes of interest, $y_{i,l,b,t}$, are default rates and lender propensity to fund a loan. Table 6 reports the results.

Due to the sensitivity of the estimates to the inclusion of controls, we opt to interpret the most restrictive specification in columns (2) and (4). We neglect the magnitude of coefficients as their interpretation would require the comparison between a zero-overlap pair of borrower and lender, with a perfectly overlapping pair.

A qualitative assessment of our estimates suggests that while the scope of connectedness is associated with higher chances of loan origination, the intensity of connectedness is associated with higher loan repayment. Intuitively, it matches our reasoning: the ability of the lender to collect soft information about the borrower by coexisting in the same online communities

	(1)	(2)	(3)	(4)
	Panel A: Default Rate (p.p.)			
Scope	-20.020	-21.317		
	(17.442)	(14.857)		
Intensity			-11.137	-15.672**
			(5.881)	(4.922)
N	16,918	16,918	16,918	16,918
R^2	0.000	0.067	0.000	0.068
	Panel B: Propensity to Lend (p.p.)			
Scope	34.996***	9.556**		
	(5.164)	(3.216)		
Intensity			6.976***	1.762
			(1.576)	(0.945)
N	438,124	438,124	438,124	438,124
R^2	0.001	0.656	0.000	0.656
Sample	Sub	Sub	Sub	Sub
Controls	No	Yes	No	Yes

Table 6: Borrower-Lender Connectedness. This table reports regression estimates for two borrower-lender connectedness measures, on a default indicator in Panel A, and on a dummy that flags request-to-loan conversion in Panel B. In Panel A, the unit of observation is a loan, and in Panel B, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts, maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling and complexity of the loan request text. The sample does not contain loans privately prearranged between borrower and lender. All standard errors are clustered by lender, and presented in parentheses.

determines that the pair forms and a loan is funded. Only then, the intensity of connection matters to increase the reputational gain of loan repayment, inducing the borrower to repay their loan.

Our interpretation requires that the key variation is the effort the lender devotes to activity in communities where the borrower is active. It could also be that the important factor is that most of the borrower activity is in communities where the lender is present. We separate the variation in our scope and intensity measures to understand which forces are driving the associations we highlight. We re-run the regression models in Table 6, but with the borrower and lender components as separate variables. This test clears whether it is important that most of the lender activity is in spaces where the borrower exists, or whether most of the borrower activity is in communities where the lender is active. Our estimates are reported in Table 7.

Both panels show that the relevant variation is in the portion of lender activity where the borrower is also active. This shows that even if the borrower has a significant portion of their activity in communities where the lender is not active, the fact that the lender devotes a significant portion of their activity to communities where the borrower is active is relevant. This is true both for loan repayment and loan origination decisions and in the directions predicted by our hypothesis.

Because credit prices are set by the borrower before a lender decides to fund the loan, we do not expect any of the connectedness measures to be associated with credit prices. We run that test and report results in Table 8 below.

As expected, we do not find an association of either connectedness scope or intensity between a borrower-lender pair, because credit prices are set by the borrower alone. This result also suggests that users with broader or more intense activity, hence more likely to be connected with a lender, do not systematically promise higher or lower interest rates.

This section highlights the importance of information acquisition, monitoring, and even social collateral (Nguyen and Dang 2020) in consumer credit. In light of the role of financial intermediaries, our results speak to three main aspects. First, financial intermediaries internalize information asymmetries between those with excess capital and those in need of capital, eliminating the need for any connectedness between depositors/investors and borrowers. Second, financial intermediaries allow collateral to be posted, lowering the cost of credit, and raising incentives for repayment. Third, through their establishment network, financial intermediaries can allocate funds from depositors/investors to borrowers active in different locations or communities, without the need for them to have any proximity, geographic or other.

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Default Rate (p.p.)			
Borrower Scope	3.144		3.028			
	(3.319)		(3.551)			
Lender Scope		-9.417***	-9.360***			
		(2.802)	(2.819)			
Borrower Intensity				-0.153		0.500
				(1.724)		(1.832)
Lender Intensity					-5.470**	-5.535**
					(1.778)	(1.794)
N	16,918	16,918	16,918	16,918	16,918	16,918
R^2	0.067	0.068	0.068	0.067	0.068	0.068
			Panel B: Propensity to Lend (p.p.)			
Borrower Scope	0.289		0.312			
	(0.171)		(0.198)			
Lender Scope		1.722***	1.731***			
		(0.508)	(0.506)			
Borrower Intensity				0.288*		0.235
				(0.138)		(0.140)
Lender Intensity					0.588*	0.558
					(0.298)	(0.294)
N	438,124	438,124	438,124	438,124	438,124	438,124
R^2	0.656	0.656	0.656	0.656	0.656	0.656
Sample	Sub	Sub	Sub	Sub	Sub	Sub
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: Separating Borrower and Lender Overlaps. This table reports regression estimates for separate borrower lender community overlap measures, on a default indicator in Panel A, and on a dummy that flags request-to-loan conversion in Panel B. In Panel A, the unit of observation is a loan, and in Panel B, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts, maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling, readability, and complexity of the loan request text. The sample does not contain loans privately prearranged between borrowers and lenders. All standard errors are clustered by lender, and presented in parentheses.

7 Robustness of Results

In this section, we assess whether our results are robust to different measurements, estimation procedures, sample designs, or conceptual approaches.

	(1)	(2)	(3)
	Interest Rate (p.p.)		
Scope	15.663		-17.608
	(88.582)		(101.651)
Intensity		17.185	20.646
		(26.896)	(30.774)
N	16,918	16,918	16,918
R^2	0.203	0.203	0.203
Sample	Sub	Sub	Sub
Controls	Yes	Yes	Yes

Table 8: Borrower-Lender Connectedness and Interest Rate. This table reports regression estimates for two borrower-lender connectedness measures, on interest rates. The unit of observation is a loan. Controls include maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling and complexity of the loan request text. The sample does not contain loans privately prearranged between borrower and lender. All standard errors are clustered by lender, and presented in parentheses.

7.1 Lender Experience

Cross-sectional variation in lender skill is less problematic than its time-series experience counterpart. Even though our goal is to interpret our results as strong descriptive associations, rather than direct causal estimates, they can still be biased. One view of the problem is that the lender chooses whether to lend in the past, and at every point in time, they are choosing, again, whether to lend, what price to accept, and how to manage default.

In our best attempt to isolate some of the time variation in past lending activity, we build an instrument that does not rely solely on lender choices. Every day, for each lender, we compute the number of prior loans given to borrowers contained in that day's choice set. The choice set is comprised of all the borrowers who requested a loan on that day. While lenders determine who to lend to, they do not determine who is part of their choice set on a given day, as measured by our instrument. This instrument is both conceptually and statistically correlated with past lending activity, as shown in Panel A of Table 9. It is expected that the experience a lender acquires in the market as a whole correlates with the experience the lender has, in a random subset of that market. The fit of the first stage is very strong with R^2 at above 0.6 for all three outcomes.

	(1)	(2)	(3)
	Panel A: Cumulative Loans - IV First Stage		
Instrument	3.931***	3.937***	4.737***
	(0.627)	(0.630)	(0.603)
N	17,123	17,123	438,648
R^2	0.627	0.626	0.602
	Panel B: IV Second Stage		
	Default Rate (p.p.)	Interest Rate (p.p.)	Prop. to Lend (p.p.)
Predicted Cumul. Loans	0.812**	-2.242	0.261***
	(0.310)	(1.566)	(0.062)
N	17,123	17,123	438,648
R^2	0.324	0.365	0.659
Sample	Sub	Sub	Sub
Controls	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes

Table 9: Cumulative Loans Instrumental Variable. This table reports instrumental variables regression estimates for the number of previously funded loans by the lender, on a default indicator, interest rate, and on a dummy that flags request-to-loan conversion. In Panel A, we present first-stage results, and in Panel B, we present second-stage results. In columns 1-2, the unit of observation is a loan, and in column 3, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except column 2), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling, readability, and complexity of the loan request text. All columns were estimated using lender fixed effects. All standard errors are clustered by lender, and presented in parentheses.

In Panel B, we instrument the number of past loans of each lender using the predicted value from the estimation in Panel A, and find very similar results to our OLS estimation: a weak positive association of experience with the default measure, a non-result in pricing, and a positive association between experience and propensity to lend. Note that in the cross-section, this approach still bears endogeneity, but in the time series, for a given lender, the choice set available on a given day is taken as given.

An additional concern with the experience analysis is that of sample composition. As discussed in Section 2, a small number of lenders occupies a large market share, and a large number of lenders lend only once or twice. In such an unbalanced sample, it is hard to speak to experience, as most loan observations will belong to highly experienced individuals with hundreds of loans funded, using perhaps very standardized procedures. This does not leave

	(1)	(2)	(3)	(4)	(5)	(6)
	Default Rate (p.p.)		Interest Rate (p.p.)		Prop. to Lend (p.p.)	
Cumulative Loans	0.360	0.413**	-0.426	0.837	0.177***	0.082***
	(0.200)	(0.202)	(0.956)	(0.898)	(0.016)	(0.011)
N	7,020	7,020	7,020	7,020	187,697	187,697
R^2	0.438	0.460	0.404	0.509	0.026	0.794
Sample	First 10	First 10	First 10	First 10	First 10	First 10
Controls	No	Yes	No	Yes	No	Yes
Lender FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 10: Past Lending Activity During First 10 Loans. This table reports regression estimates for the number of previously funded loans by the lender, on a default indicator, interest rate, and a dummy that flags request-to-loan conversion. In columns 1-4, the unit of observation is a loan, and in columns 5-6, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except columns 3-4), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling and complexity of the loan request text. All columns are ran on a subsample of the first 10 loans for all the lenders who funded at least 10 loans. All columns were estimated using lender fixed effects. All standard errors are clustered by lender, and presented in parentheses.

a lot of room for experience to impact the outcomes.

To address that, we build a sample with the first 10 loans of lenders who have at least 10 loans, a sample in which all lenders take the same weight. We report the results with and without controls in Table 10. Our results are not only identical to those with the full sample, but also robust to whether we consider the first 5, 10, or 20 loans for each lender, as shown in Appendix Table A2.

7.2 Lender Skill

To test whether our results are due to mechanical artifacts of our measure of past lending activity, we run our baseline specification on an alternative measure. For each loan, instead of using the number of loans that the lender has funded in the past, we use the number of days since the lender's first loan, as our measure of lending activity. Table 11 reports the results. In all three outcomes the results are qualitatively similar, and so is the fit across specifications and samples, confirming the lender skill mechanism as an important determinant of credit risk, credit price, and loan origination.

	(1)	(2)	(3)	(4)
	Panel A: Default Rate (p.p.)			
Cumulative Days	-0.150***	-0.109***	-0.085***	-0.050**
	(0.020)	(0.020)	(0.016)	(0.015)
N	19,842	33,137	19,842	33,137
R^2	0.016	0.011	0.076	0.079
	Panel B: Interest Rate (p.p.)			
Cumulative Days	-0.646***	-0.572***	-0.365***	-0.334***
	(0.159)	(0.171)	(0.089)	(0.096)
N	19,842	33,137	19,842	33,137
R^2	0.012	0.011	0.206	0.195
	Panel C: Propensity to Lend (p.p.)			
Cumulative Days	0.014	0.080***	0.025***	0.080***
	(0.009)	(0.019)	(0.005)	(0.015)
N	466,347	479,266	466,347	479,266
R^2	0.000	0.007	0.625	0.501
Sample	Sub	Full	Sub	Full
Controls	No	No	Yes	Yes

Table 11: Alternative Measure of Past Lender Activity. This table reports regression estimates for the number of days since the lender's first loan, on a default indicator in Panel A, interest rate in Panel B, and on a dummy that flags request-to-loan conversion. In Panels A and B, the unit of observation is a loan, and in Panel C, the unit of observation is potential loans by combining pairs of borrowers and lenders on the same day. Controls include requested and promised repayment amounts (except Panel B), maturity, dummies for the borrower's first loan, borrower default in the past, and borrower late payment in the past, as well as measures of misspelling and complexity of the loan request text. The first two columns exclude controls, and odd-numbered columns do not contain loans privately prearranged between borrower and lender. All standard errors are clustered by lender, and presented in parentheses.

7.3 Relevance of Connectedness

Throughout our analysis and interpretation of results, we view connectedness in other communities as a proxy for the acquisition of soft information, and for the ability of the lender to monitor the borrower, leading the borrower to repay. An angle we do not directly analyze is that of borrower reputation in other communities and the threat to that reputation. A credible threat of shaming the borrower in communities where the borrower holds a reputation, and the lender knows this, could be the mechanism that links connectedness to repayment, and subsequent willingness to lend.

To test this, we conduct an event study around borrower default, and estimate lender's "abnormal" activity in communities where the borrower is active, *vis-à-vis* their activity in communities where the borrower is not active. We do so by estimating Equation (8):

$$y_{l,\tau,c} = \alpha_l + \sum_{k=-3}^{3} \lambda_k I(\tau = k) + \gamma I(Overlap_c) + \sum_{k=-3}^{3} (\beta_k I(\tau = k) \times I(Overlap_c)) + \varepsilon_{l,\tau,c} \quad (8)$$

where $y_{l,\tau,c}$ denotes the number of posts or comments made by the lender in the community c, τ weeks relative to borrower default. α_l are lender fixed-effects to absorb the lender's average activity across all platforms, $I(\tau = k)$ are dummy variables for different weeks around the default event. The coefficients $\{\lambda_k\}$ absorb lender activity around default on communities where the borrower is not active, and $Overlap_c$ is a dummy equal to one if community c is a community where both borrower and lender are active. The coefficients of interest are $\{\beta_k\}$, and the omitted event time is -1. We plot the coefficients in Figure 13.

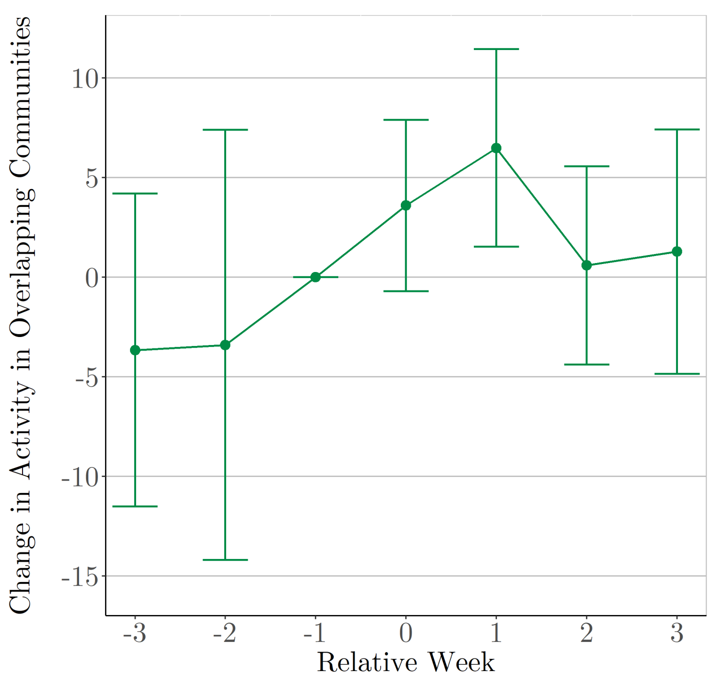


Figure 13: Lender Activity Event-Study Around Default. In this figure we plot coefficients of a differences-in-differences event study around borrower default. We measure lender post and comment activity in all communities except the credit marketplace, and separate communities based on lender and borrower overlap. Therefore, the figure reports estimates for abnormal changes in lender activity in communities where the borrower is active.

We observe abnormal lender activity in overlapping communities in the week of the default and in the following week. We interpret this result as a validation of our measure of connectedness. This activity can have two purposes: (i) monitoring the borrower, and fishing for their activity or (ii) publicly shaming the borrower, affecting their reputation. Only in 57 out of more than 1.8 million posts and comments from lenders, the borrower's username is mentioned. This evidence points towards a monitoring role of connectedness, rather than a credible threat to the borrower's reputation.

8 Conclusion

We collect a comprehensive dataset on an online informal lending forum to study unregulated consumer credit, without any intermediaries. Our first result is that credit prices and default rates are both extremely high, even when compared with subprime loans. Secondly, we find that a small group of lenders gains a large and very profitable market share, while the vast majority realizes losses, on average.

We develop a stylized framework to understand the determinants of prices, default, and loan origination. We bring the model to the data and find that lender skill is associated with a higher likelihood of originating loans, lower credit prices, and lower default. Additionally, we find that monitoring and acquisition of information also play a role in decreasing default and increasing the propensity to originate.

Our study is one of the first to study informal lending and to use that setting to understand the role of financial intermediaries in consumer credit. We speak to the debate on financial disintermediation by showing that beyond efficient allocation of capital, risk sharing, and diversification, intermediaries have two more important roles in consumer credit markets, facilitating access to credit, easing prices, and reducing risk. First, they bring in skill, by employing skilled human capital to make origination and pricing decisions. Second, they eliminate the need for investors to directly acquire information about borrowers.

References

Allen, Franklin and Douglas Gale (1999) "Innovations in Financial Services, Relationships, and Risk Sharing," *Management Science*, 45 (9), 1239–1253.

Allen, Franklin and Anthony M Santomero (1997) "The Theory of Financial Intermediation," *Journal of banking & finance*, 21 (11-12), 1461–1485.

Angelucci, Manuela, Dean Karlan, and Jonathan Zinmin (2015) "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco," *American Economic Journal: Applied Economics*, 7 (1), 151–182.

Baily, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong (2018) "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 32 (3), 259–280.

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinmin (2015) "Six Randomized Evaluations of Microcredit: Introduction and Further Steps," *American Economic Journal: Applied Economics*, 7 (1), 1–21.

Banerjee, Ryan N, Leonardo Gambacorta, and Enrico Sette (2021) "The Real Effects of Relationship Lending," *Journal of Financial Intermediation*, 48, 100923.

Bayer, Patrick, Kyle Mangum, and James W. Roberts (2021) "Speculative Fever: Investor Contagion in the Housing Bubble," *American Economic Review*, 111 (2), 609–651.

Berger, Sven and Fabian Gleisner (2009) "Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending," *Business Research*, 2 (1), 39–65.

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov (2014) "The Economics of Predation: What Drives Pricing When There is Learning-by-doing?" *American Economic Review*, 104 (3), 868–897.

Braggion, Fabio, Alberto Manconi, Nicola Pavanini, and Haikun Zhu (2022) “The Value of Financial Intermediation: Evidence from Online Debt Crowdfunding,” *European Corporate Governance Institute–Finance Working Paper*, 719, 2020.

Correia, Filipe, Peter Han, and Jialan Wang (2023) “The Online Payday Loan Premium,” *Unpublished Working Paper*.

Correia, Filipe, Antonio Martins, and Anthony Waikel (2022) “Online Financing Without FinTech: Evidence From Online Informal Loans,” *Journal of Economics and Business*, 121.

Crawford, Gregory S, Nicola Pavanini, and Fabiano Schivardi (2018) “Asymmetric Information and Imperfect Competition in Lending Markets,” *American Economic Review*, 108 (7), 1659–1701.

Diamond, Douglas W (2023) “Nobel Lecture: Financial Intermediaries and Financial Crises,” *Journal of Political Economy*, 131 (10), 000–000.

Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan (2018) “Consumer Price Search and Platform Design in Internet Commerce,” *American Economic Review*, 108 (7), 1820–1859.

Dorfleitner, Gregor, Christopher Priberny, Stephanie Schuster, Johannes Stoiber, Martina Weber, Ivan de Castro, and Julia Kammler (2016) “Description-text Related Soft Information in Peer-to-Peer Lending – Evidence From Two Leading European Platforms,” *Journal of Banking & Finance*, 64, 169–187.

Duygun, Meryem, Shatha Qamhieh Hashem, and Alessandra Tanda (2021) “Financial Intermediation Versus Disintermediation: Opportunities and Challenges in the FinTech Era,” *Frontiers in Artificial Intelligence*, 3, 629105.

Einav, Liran, Chiara Farronato, Jonathan Levin, and Neel Sundaresan (2018) "Auctions Versus Posted Prices in Online Markets," *Journal of Political Economy*, 126 (1), 178–215.

Elyasiani, Elyas and Lawrence G Goldberg (2004) "Relationship Lending: A Survey of the Literature," *Journal of Economics and Business*, 56 (4), 315–330.

Evans, Lawrance (2019) "Agencies Should Provide Clarification on Lenders' Use of Alternative Data," *United States Government Accountability Office*.

Fang, Lily, Victoria Ivashina, and Josh Lerner (2015) "The Disintermediation of Financial Markets: Direct Investing in Private Equity," *Journal of Financial Economics*, 116 (1), 160–178.

FDIC (2022) "2021 FDIC National Survey of Unbanked and Underbanked Households."

Fracassi, Cesare (2017) "Corporate Finance Policies and Social Networks," *Management Science*, 63 (8), 2420–2438.

Franks, Julian, Nicolas Serrano-Velarde, and Oren Sussman (2021) "Marketplace Lending, Information Aggregation, and Liquidity," *The Review of Financial Studies*, 34 (5), 2318–2361.

Harris, Robert S, Tim Jenkinson, Steven N Kaplan, and Ruediger Stucke (2018) "Financial Intermediation in Private Equity: How Well Do Funds of Funds Perform?" *Journal of Financial Economics*, 129 (2), 287–305.

Herzenstein, Michal, Scott Sonenshein, Utpal Dholakia, Johannes Stoiber, and Mar Weber (2011) "Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-Peer Lending Decisions," *Journal of Marketing Research*, 48, 138–149.

Holmes, Jessica, Jonathan Isham, Ryan Petersen, and Paul M Sommers (2007) "Does Relationship Lending Still Matter in the Consumer Banking Sector? Evidence From the Automobile Loan Market," *Social Science Quarterly*, 88 (2), 585–597.

Ioannidou, Vasso, Nicola Pavanini, and Yushi Peng (2022) "Collateral and Asymmetric Information in Lending Markets," *Journal of Financial Economics*, 144 (1), 93–121.

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F.P. Luttmer, and Kelly Shue (2015) "Screening Peers Softly: Inferring the Quality of Small Borrowers," *Management Science*, 62 (6), 1554–1577.

Karaivanov, Alexander and Anke Kessler (2018) "(Dis)advantages of Informal Loans – Theory and Evidence," *European Economic Review*, 102, 100–128.

Karolyi, Stephen Adam (2017) "Personal Lending Relationships," *The Journal of Finance*, 73 (1), 5–49.

Kawai, Kei, Ken Onishi, and Kosuke Uetake (2022) "Signaling in Online Credit Markets," *Journal of Political Economy*, 130 (6), 1585–1629.

Kraussl, Roman, Zsofia Kraussl, Joshua Pollet, and Kalle Rinne (2022) "The Performance of Marketplace Lenders," *Available at SSRN 3240020*.

Kysucky, Vlado and Lars Norden (2016) "The Benefits of Relationship Lending in a Cross-country Context: A Meta-analysis," *Management Science*, 62 (1), 90–110.

Lin, Mingfeng, Prabhala Nagpurnanand, and Siva Viswanathan (2012) "Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending," *Management Science*, 59 (1), 17–35.

López-Espinosa, Germán, Sergio Mayordomo, and Antonio Moreno (2017) "When Does Relationship Lending Start to Pay?" *Journal of Financial Intermediation*, 31, 16–29.

Nguyen, Ha-Diep and Huong Dang (2020) "Social Collateral," *Unpublished Working Paper*.

Pyle, David H (1971) "On the Theory of Financial Intermediation," *The Journal of Finance*, 26 (3), 737–747.

Scholtens, Bert and Dick Van Wensveen (2000) “A Critique on the Theory of Financial Intermediation,” *Journal of Banking & Finance*, 24 (8), 1243–1251.

Xin, Yi (2023) “Asymmetric Information, Reputation, and Welfare in Online Credit Markets,” *Available at SSRN 3580468*.

Time Period	N. Obs	Requested Amount ($)	Promised Maturity (Days)	Prom. Monthly Interest Rate	Converted Into Loan	Default Upper Bound	Default Lower Bound
Panel A: Loan Requests							
Full Sample	51,020	273	23	107%	68%	-	-
2016	810	268	27	117%	53%	-	-
2017	1,627	275	25	92%	54%	-	-
2018	6,365	273	26	92%	53%	-	-
2019	15,032	276	23	107%	63%	-	-
2020	13,883	271	22	110%	74%	-	-
2021	13,303	272	20	115%	81%	-	-
Panel B: Funded Loans							
Full Sample	35,066	216	25	131%	-	4.2%	10.6%
2016	426	208	22	108%	-	4.7%	14.1%
2017	880	211	19	117%	-	5.3%	18.2%
2018	3,340	204	20	119%	-	4.8%	15.9%
2019	9,426	201	17	135%	-	3.6%	11.6%
2020	10,248	219	18	133%	-	3.9%	14.4%
2021	10,746	231	18	131%	-	4.7%	17.2%

Appendix Table A1: Summary Statistics by Year. This table reports yearly averages for promised loan terms for loan requests in Panel A, and loan terms for funded loans in Panel B.

	(1)	(2)	(3)
	Panel A: Default Rate (p.p.)		
Cumulative Loans	0.645	0.413*	0.064
	(0.517)	(0.204)	(0.091)
N	2,325	2,810	2,860
R^2	0.313	0.189	0.160
Sample	First 5	First 10	First 20
	Panel B: Interest Rate (p.p.)		
Cumulative Loans	1.339	0.837	-1.354*
	(2.123)	(1.042)	(0.592)
N	2,325	2,810	2,860
R^2	0.481	0.365	0.296
Sample	First 5	First 10	First 20
	Panel C: Propensity to Lend (p.p.)		
Cumulative Loans	0.102***	0.082***	0.047***
	(0.025)	(0.013)	(0.007)
N	65,467	74,729	72,503
R^2	0.799	0.759	0.717
Sample	First 5	First 10	First 20
Controls	Yes	Yes	Yes
Lender FE	Yes	Yes	Yes

Appendix Table A2: Sample Composition Robustness. This table reports regression estimates for the number of previously funded loans by the lender, on a default indicator in Panel A, interest rate in Panel B, and a dummy that flags request-to-loan conversion in Panel C. In columns 1, 2 and 3, the sample contains the first 5, 10, and 20 loans for lenders with at least 5, 10, and 20 loans, respectively. All columns were estimated using lender fixed effects. All standard errors are clustered by lender, and presented in parentheses.