Identifying the Mechanisms of Extralegal Enforcement

Greg Buchak*
University of Chicago

June 14, 2017

Abstract

This paper studies the mechanisms by which a community enforces contracts absent a formal legal authority. While many studies of self-ordering and extralegal contracting focus on rule *selection*, there has been less focus on rule *enforcement*. Various mechanisms have been proposed, but there has been little quantitative analysis of whether these mechanisms are actually active and what this implies about the robustness of the community. Drawing on a unique dataset of anonymous and unsecured online peer-to-peer loans, this paper shows that forward-looking economic self-interest is primarily responsible for facilitating repayment. Social ties play an important but replaceable role as a costly signal and do little to directly motivate repayment. The results show that the combination of repeat play, a reliable information transmission mechanism, and costly entry is sufficient for effective extralegal enforcement.

*University of Chicago, emails: `gbuchak@chicagobooth.edu`. I thank Lisa Bernstein, William Hubbard, Luigi Zingales, Raghuram Rajan, Gillian Hadfield, Tom Ginsburg, and seminar participants at the University of Chicago Legal Scholarship Workshop and the Stigler Workshop.

1 Introduction

A user identified by the online handle *trigonoah* posts the following message on an internet forum, reddit.com:[1]

> *Looking to borrow $100 to help with daily commute to work and groceries. Am able to pay back $110 on 12/4. I live in North Plainfield, NJ and accept PayPal.*

If *trigonoah* gets a loan and decides not to repay, there is no formal authority to compel payment or punish the offender. The aggrieved lender, now $100 poorer, will post publicly that *trigonoah* defaulted on the loan, and that will be the end of it. Nevertheless, *IgrewAtomato*, another user, sends *trigonoah* $100. A few weeks later *IgrewAtomato* posts reporting that *trigonoah* has paid him back with interest.[2]

This transaction is fairly typical in this market, which has been active since mid-2014. As of September 2016, the market sees roughly 700 monthly loan requests with roughly 70% of them filled, typically within minutes or hours of the initial request. In dollar terms, this corresponds to roughly $200,000 in requests and $100,000 in monthly fills. Despite the lack of formal enforceability, more than 90% of loans are repaid.

This paper asks two questions: First, what motivates borrower repayment absent formal legal enforcement? Second, how does the community at large understand and use this information to transact? Using transaction-level data, this paper seeks to offer a detailed, empirical answer to these questions and offer generalizable lessons broadly applicable to extralegal contracting and self-ordering.

Motivating Repayment: A long line of examples[3] has shown that formal legal institutions are not necessary for social ordering. The focus in previous literature, however, has tended to be on rule *selection* rather than rule *enforcement*. This is a natural focus because many of these self-ordering examples sit in the shadow of a formal enforcement authority that could, in theory, intervene. Often, the question in these settings relevant to legal scholars is whether the formal enforcement authority would or should adopt the community-selected rules.

[1] `https://www.reddit.com/r/borrow/comments/3ugqur/req_100_to_be_paid_back_1242015_interest/`

[2] `https://www.reddit.com/r/borrow/comments/3veyk2/paid_utrigonoah_100_interest_on_time/`

[3] See the examples of whalers (Ellickson (1989)), diamond merchants (Bernstein (1992)), wheat dealers, Bernstein (1996), the cotton industry (Bernstein (2000)). See Bernstein and Parisi (2014) for a detailed discussion.

When no such authority exists, even in theory, the details of how the community handles enforcement rise to first-order importance. This is particularly true in transactions requiring sequential actions. The typical game involves agents L and B, who can engage in some welfare-improving interaction. The transaction is potentially problematic if it requires L to act, at some cost, before B can commit to cooperate. If cooperation is costly for B, B has a strong ex-post incentive to defect. Knowing this, L declines to engage in the interaction ex-ante, creating a welfare loss.[4] For the market to function, there must be some mechanism to constrain B's ex-post bad behavior that L can rely upon ex-ante. What matters for behavior, and consequently welfare, is not B's ex-ante assurances or promises of cooperation, but L's and B's joint prediction of what will in fact happen if B breaks his promise.[5]

Ellickson (2016) provides a general framework for studying what constrains ex-post bad behavior, which this paper adopts. This framework categorizes repayment mechanisms as coming from three sources: (1) First-party control though the borrower's own internalized norms or ethics; (2) Second-party control through the borrower's potential victim engaging in self-help; (3) Third-party control, enforced either through (a) diffuse social norms, (b) formal but non-governmental hierarchical organizations,[6] or (c) governments. Each source can furnish one or more mechanism within the bounds of institutional or technological feasibility.[7] For example, first-party control may function through norms, ethics, or habits; third-party government control may function through fines or imprisonment. For social ordering to work generally, a community must provide one or a set of mechanisms that provide motivations to cooperate that overcome an individual's motivations to defect.

Legal scholarship traditionally focuses narrowly on cases where the punishing agent is the government. This falls in category (3c) of Ellicksons' categorization. In this paper's setting, however, a government or other formal enforcement authority is off the table, and so the community must bootstrap a set of enforcement mechanisms drawing on the alternatives. This paper aims to identify and quantify these mechanisms, and to explore the extent to which these mechanisms create law in the realist sense.[8]

[4] In this paper's context, the moral hazard problem is particularly stark: The lending transaction is that L transfers B money; at some later date, B transfers the money plus interest back to L.

[5] Holmes (1897)

[6] Say, a firm. See more generally, Burt (2005).

[7] The baseline requirement of institutional or technological feasibility simply means that the punishment must be within the community's ability to enact. For instance, the community in this paper lacks the power to directly take funds from a defaulting borrower's bank account or imprison him. Consequently, these mechanisms are a priori off the table.

[8] For a formal treatment of whether the ex-post punishment can be credibly enacted in equilibrium, see Greif (1993). The question of ex-post enforceability is complex when the punishment relies on social exclusion and it is not clear it is possible without strong behavioral assumptions. This paper tables the

When discussing potential enforcement mechanisms, the prior literature, e.g., case studies of diamond merchants (Bernstein (1992)), whalers (Ellickson (1989)), wheat dealers (Bernstein (1996)), and many others, typically flag two categories of third-party informal punishments: Economic mechanisms and social mechanisms. Economic mechanisms[9] refer to direct economic costs such a loss of money.[10] Lacking the technological capability to impose a direct sanction, the monetary penalty works by restricting market access in the future. The cost of the punishment to the misbehaving agent is the present value of future transactions within the community. If the economic mechanism is the only active channel, then the mechanism will work only if the present value of future transactions exceeds the one-time value of current dishonest behavior. Note that this forward-looking mechanism relies critically on an expectation of repeat interaction and an agent that makes decisions with this forward-looking calculus in mind.[11] For this paper, I refer to this mechanism as the *homo economicus* mechanism.

Social mechanisms[12] as opposed to economic mechanisms, function through the direct loss of social ties, social stigma, or other harm inflicted on the defaulting agent mediated through the agent's social embeddedness within the community. In essence, the agent's embeddedness within the community itself furnishes a potential punishment. For instance, a misbehaving agent may lose his friends or the community at large may shun him.[13] Unlike the economic mechanism, whose potency relies on a high value of future transactions, a social mechanism relies on social embeddedness—the fact that the agent finds his social ties valuable and would suffer from losing them, or some other direct social punishment For this paper, I refer to this mechanism as the *homo socialis* mechanism.

This paper attempts to map the *homo economicus* and *homo socialis* mechanisms to observables in the data in order to determine the quantitative importance they play in the market. Separating and quantifying these various forms of social control requires taking a stand on the way in which they manifest and the ways in which lenders interpret borrower

question of whether third-person punishments are incentive compatible; the empirical evidence is strong that the punishments that are technologically feasible are implemented ex-post, regardless of ex-ante incentive compatibility in a rational mechanisms design sense.

[9] The legal literature often refers to this channel as "reputational." See Bernstein (1992)

[10] Or more broadly, a loss of utility stemming from a loss of consumption. Essentially, anything typically called "non-behavioral."

[11] See Posner (2009) for a general game-theoretic analysis of social norms and repeated interactions.

[12] Bernstein (1992) includes in the *social* mechanism the way that an agent's *reputation* spreads. The importance of this force is supplementary to the reputation force, so for the purposes of this paper I think of it as included there. In a sense this mechanism works trivially in this setting because everything is recorded publicly.

[13] Ambrus et al. (2014) treats this mechanism formally in a social network, where social ties have direct consumption value.

behavior. The essential problem is that in a given setting, these separate modes of social control are both possibly active and at a high level will generate similar observable outcomes. Broadly, these mechanisms both work through *reputation*. That is, community members' future interactions depend on the borrower's past conduct. A robust fact in this market is that borrowers who repay are more likely to get loans in the future and borrowers who default are less likely to get loans in the future—an example of "reputation matters." A large literature has studied online reputation, typically on Ebay (e.g., Cabral and Hortacsu (2010) or Tadelis (2016)) but also on the Internet black markets (e.g., Hardy and Norgaard (2016)). The stylized results are that, as is the case here, reputation is closely connected to outcomes.

However, simply showing that reputation *matters* elides the various mechanisms though which reputation operates. In a *homo socialis* world, reputation matters because borrowers' social ties induced them to behave well in the past and will do so again in the future. In this case, reputation is a stand-in for a deeper unobserved omitted variable—social ties. In a *homo economicus* world, reputation matters because past default leading to punishment and past repayment leading to future lending is the equilibrium outcome in a repeated game. Reputation simply summarizes equilibrium outcomes and marks which borrowers are sensitive to losing economic access. Reputation can have a causative impact in the repayment decision because a good reputation may increase the borrower's forward-looking value function, but this channel is not strictly necessary in a homo *economicus world*. Observing that reputation matters in a broad sense does not explain why it matters, through which mechanism it operates, and whether it is merely the outcome of a deeper causative factor.

The fact that these mechanisms generate similar high-level outcomes *within* markets makes comparative statics across institutional settings or market characteristics difficult.[14] For instance, how important is it that the diamond dealers all came from an ethnically and religiously homogeneous community? If the community grew more diverse, would potential social punishments lose their effect, or is a more general, less community-specific *homo economicus* mechanism doing the heavy lifting? Do the Nantucket whalers rely on the fact that they live next to one another on a small island, making social pressures particularly strong, or is the economic consequences losing future community cooperation doing the work? As more commerce moves online, will a breakdown of social embeddedness make extralegal contracting more difficult, or will future economic opportunities become more valuable, thus making extralegal contracting easier? Without a more detailed quantitative analysis of the mechanism driving repayment, it is difficult to answer these questions.

[14]See Bernstein (1992) discussing the implications of the diamond industry's global expansion.

This paper's unique empirical setting allows for a disentangling of these forces, and hence provide some quantitative lessons that can be applied to other settings. The results are strongly consistent with the *homo economicus* mechanism playing a large role in incentivizing repayment, while the *homo socialis* mechanism is largely inactive at the moment of the repayment decision. Social ties play an important signaling role, but this role could easily be replaced with any other costly and verifiable barrier to entry.

In addition to shedding light on what makes extralegal enforcement tick, the results provide some lessons regarding anonymity and privacy on the internet and suggest ways to harness positive aspects of anonymity while discouraging its more negative aspects.[15] In particular, what appears to be a critical institutional design feature given anonymity is a barrier to entry. That is, anonymity is not a deal breaker in a self-ordering context: The real deal-breaker is free entry. So long as barriers to entry can prevent unlimited free entry by anonymous agents, a system can overcome the negative aspects of anonymity.

Community Understanding of Repayment Mechanisms: After exploring the mechanisms that actually motivate ex-post repayment, a natural question is whether the community—in particular the lenders and its institutions—appreciate and effectively leverage those motivations. This is of interest for two reasons. First, there is a baseline question of (constrained) efficiency: Given the available effective mechanisms of social control, do lenders in fact allocate credit to borrowers who are incentivized to repay and away from borrowers who are not? Second, there is deeper question of rule internalization: Do lenders appear to appreciate the norms and mechanisms operating within their community, and is there evidence of heterogeneity, e.g., do more experienced lenders appear to be more informed?

Taking a broader view, there is ample evidence in settings with formal legal systems that lenders allocate more credit at lower costs to projects where their legal remedies for non-performance are greater. For instance, in the context of secured lending, Benmelech et al. (2005) and Benmelech (2009) show that assets with greater reuse value command more borrower-friendly contract terms than assets with less reuse value. Assets with greater reuse value (say, real estate in an area zoned for general use, rather than residential use) are more valuable to the lender in the event that the borrower defaults and the legal system delivers the securitized asset to the lender. That this leads to different ex-ante credit allocation shows that these lenders both internalize and utilize the mechanisms of the legal system in which they operate. The question in this case is whether the lenders internalize and utilize

[15]See, e.g., many studies of online social networks which often emphasize negative aspects of anonymity. For example, Bakshy et al. (2011), Bakshy et al. (2012), or Rost et al. (2016) often discuss harassment and bullying on Twitter.

the mechanisms of their informal system.

Regarding efficiency, parallel to the traditional question of whether self-ordering rule selection is wealth maximizing in Ellickson (1989), this section asks whether the community adopts norms of lending that efficiently take advantage of borrowers' motivations to repay. In particular, the test is whether lenders choose to lend to borrowers who ex-ante have the traits of a borrower who is most likely to repay ex-post, or whether there appears to be inefficiency in their selection. Regarding the internalization of community norms, the test is whether some lenders—the lenders with the most experience in the lending community—realize better loan performance ex-post, and if so, is it because they possess greater skill in selecting borrowers based on what is shown to predict repayment?

The results show that overall, lenders are very adept in selecting borrowers, with only some systematic mistakes appearing ex-post. This suggests that lenders in the market come close to achieving ex-ante efficiency in their understanding of what drives borrowers to repay ex-post. Moreover, borrowers with more lending experience perform better, appear to make no systematic mistakes, and are more likely to lend to the borrowers demonstrating exactly those traits most associated with ex-post repayment. These results imply that indeed, the community has for the most part adopted lending standards that efficiently take advantage of how repayment is motivated ex-post.

The paper proceeds as follows. Section 2 lays out the institutional details and presents summary data and some stylized facts. Section 3 studies what motivates borrower repayment ex-post. Section 4 studies how lenders make decisions with an eye towards testing whether lenders use information regarding borrower repayment efficiently. Section 5 discusses the implications of the findings and concludes.

2 Institutional Details and Market Description

This section provides brief overview of the data and a descriptive overview of the market. The data are collected via web scraping Reddit.com and a post-processing script that builds the transaction history from raw text. Section 6.1 in the appendix provides a detailed description of this process.

2.1 Institutional Details

Reddit is a large online social networking site that is divided into thousands of user-created sub-forums called subreddits, each of which focus on a particular topic or interest. Subreddits are moderated, meaning a certain group of users (typically the subreddit creator and other volunteers) have the power to delete posts, allow machine-operated robots to post, ban users for certain behaviors, and so on. Users on reddit create persistent accounts that are not associated with any personal information. A users' entire posting history is publically viewable. Account creation and deletion is free; a single user can create as many accounts as he likes, and there is no way to link one account to another account belonging to the same user.[16] This paper studies lending taking place on two subreddits, *loans* and *borrow*. *loans* predates *borrow* by several years; around August 2014 the moderators of *loans* became inactive and the community migrated to *borrow*.[17]. Figure 6 shows the sharp drop-off in activity on *loans* and sharp increase in activity on *borrow* around this time.

The *borrow* moderating team instituted a number of policies, among them standardized reporting and information-production tools. In particular, a computer program—"bot"—monitors loan requests, fills, repayments, and defaults, and automatically posts this information whenever a borrower requests a loan.[18] This service existed in the earlier *loans*, but the lack of standardization meant that it was less reliable. While this standardization of rules likely had a significant impact on lending outcomes in the community,[19] it is not the focus of this paper. Rather, the upshot of these institutional changes for this paper is that the data quality improvements stemming from this standardization make the *borrow* subsample data of higher quality. Consequently the analysis focuses on this data. Results using the combined sample, however, are qualitatively similar.

Users requesting a loan post a request to the site. The request header contains some standardized information, including the loan amount and repayment dates. The request also typically includes a text description of the loan, including why the borrower needs it. See the top panel of Figures 1, 2, and 3. The subreddit's bot that keeps track of credit history

[16]Interestingly, it is possible to purchase "used" accounts—accounts having a history of participation on the site. In an extension I investigate whether the price of purchasing an account is consistent with the potential gains one could obtain by purchasing an account for a one-time use.

[17]`https://www.reddit.com/r/Loans/comments/2hjesn/meta_rloans_has_migrated_to_rborrow_please/`

[18]It is important to note that this credit information is theoretically publicly available without the bot's data collection, and user post histories are recorded and publicly viewable. The bot, however, makes this information very easy to see.

[19]Observe the clear structural increase in lending probability between *loans* and *borrow* in Figure 8

immediately posts that history as a comment in the borrower's submission. See Figure 4. A lender wishing to lend posts a standardized public message. Borrower and lender will often communicate privately, which is not observable, and report exchanging names, Facebook accounts, and so on, although making this personal information public on Reddit violates Reddit's terms of service, leading to a banning from the site. The actual money transfer typically takes place through PayPal. Once it has been received, the the borrower confirms with his own standardized message. See the middle panel of Figures 1, 3. If the borrower repays, the lender creates a new submission with its own standardized header information that the bot records for future credit checks. See the bottom panel of 1.

The key friction in this market is the lack of formal enforceability of repayment. Though PayPal typically handles the money transfer, due to the nature of these transactions, PayPal rarely refunds defaulted loans to the lender.[20] Further, although the lender often privately knows the real identity of the borrower, due the small nature of the loans—typically \$250 to \$300—it is simply impractical to resolve disputes in the formal legal system. Rather, a lender's only recourse is to publicly post that the borrower defaulted. See the bottom panel of Figure 3. As discussed earlier, reddit accounts are free to create and the accounts belonging to a single user are unlinkable. As mentioned earlier, Reddit terms of service disallow the public distribution of personal information, even for defaulting borrowers, and so credit history as reported on Reddit must be at the Reddit user-name level, rather than at the human borrower level. As a descriptive matter, a username associated with a default is essentially excluded from future borrowing,[21] although there is no way to prevent a defaulting user from creating a new account.

[20]The best case scenario for a lender in case of default is to file a PayPal dispute and hope that PayPal does not notice that this transaction violates its terms of service: PayPal's Acceptable Use Policy bans "credit transactions" and transactions in violation of state laws, which these transactions almost certainly do due to their high interest rate. Users frequently discuss whether PayPal will refund the payment if they file a dispute. See, e.g., `https://www.reddit.com/r/borrow/comments/3og6xd/meta_paypal_dispute_for_unpaid_loan_is_asking_for/`. Users report mixed success. The typical advice offered is to tell PayPal that the payment was for the purchase of goods or services and was mistakenly sent to the wrong address, although due to these disputes typically being sent several weeks after the payment, this does not appear to work all the time. Moreover, lenders caution each other to not file too many disputes, or PayPal will become suspicious and freeze the lender's account. See `https://www.reddit.com/r/borrow/comments/2ze6g5/meta_how_many_paypal_disputes_should_i_have_going/`.

[21]Among loans that are made, a borrower who does not repay has a 14% change of requesting another loan versus a 60% chance among those who repay. Conditional on requesting another loan, a borrower who has defaulted has roughly a 30% chance of receiving another loan, putting the total probability of receiving another loan after default below 5%. The 30%, which seems high, is likely due to ongoing lender-borrower relationships where a borrower has successfully repaid in the past and the lender is comfortable with lending again (perhaps because the lender knows the borrower has defaulted for a "fundamental," non-strategic reason).

2.2 Descriptive Statistics and Market Trends

Despite these frictions, the market has survived and even grown, though remained fairly small. As of October 2016, *borrow* receives roughly 600 loan requests per month for nearly \$200 thousand in principal, up significantly from when lending began in late 2011. See Figure 5. Roughly two-thirds of these requests are filled, with 400 monthly fills in October 2016, although the average fill size is slightly smaller than the average request size, with dollars filled being roughly \$100 thousand. See Figure 6. The average request size is roughly \$300; the average fill size is roughly \$250. Both average request and fill sizes have increased significantly, suggesting that the market is able to sustain greater loan sizes than it could previously. See Figure 7.

The mean time between lending and repayment is 26 days; the median time is 16 days; the bottom quartile is 7 days, and the top quartile is 31 days. The loans are repaid with interest, with the typical repayment size being 20% above the principal. This paper does not analyze in detail how interest rates are determined in this market. Appendix Section 6.3 contains a brief analysis of interest rate determination and their impact in the lending decision. The upshot is that interest rates are typically highly standardized and due to the complicated asymmetric information and moral hazard problems surrounding this market, are somewhat but not strongly correlated to perceived default risk and the lender's decision of whether or not to lend.

The market sustains between a 60% and 70% fill rate for loans requests. There is a significant structural increase occurring after the migration from *loans* to *borrow*. See Figure 8. The time between a borrower posting a request and a borrower's request being filled is typically about one hour, as of October 2016. This median time to fill decreased significantly around the inception of the community and has been roughly constant since the migration to *borrow*. For borrowers with an established credit history, the median fill time is roughly thirty minutes. See Figure 11. Finally, the repayment rate is high, with more than 90% of loans repaid. See Figure 8.

A striking and important feature of the market is the repeat nature of many borrowers and lenders. Within a given month, there is significant heterogeneity in how active both borrowers and lenders are. Figure 9 shows the distribution of borrower (top panel) and lender (bottom panel) activity per month. As of October 2016, while the median borrower borrows roughly \$200 per month, more active borrowers in the 90th percentile borrow upwards of \$700. Differences are even more stark for lenders, with the median lender making roughly \$500 in monthly loans, while the most active lenders make between four and six thousand worth of loans per month. I will return to the role that these "expert" lenders play in the market later.

Importantly, as the market has gone on, certain borrowers have been able to establish large credit histories. Figure 10 shows the distribution of past borrower fills as of the time of their latest loan request. As the market has become more mature, the average credit history of a typical borrower has improved. Around the start of 2015, the market passed a critical point where the median borrower requesting a loan has had a loan before; as of October 2016, the median requesting borrower has had one loan in the past. As above, the distribution is highly skewed, with requesters in the 90th percentile having received 14 loans in the past.

To summarize the facts in the market: (1) The overall size is small but growing robustly. (2) Typical fills are for roughly $250 to be repaid in slightly under one month with 20% interest. (3) The fill rate is roughly 60-70%, and the repayment rate is in excess of 90%. (4), borrowers and lenders engage in significant repeat borrowing and lending. With these stylized facts presented, the paper proceeds to investigate how enforcement works, what motivates repayment, and what the implications of this market are for extralegal self-ordering.

3 Economic and Social Forces in Extralegal Enforcement

This section puts two mechanisms of third-party social control in a horse-race: *Homo socialis* and *homo economicus*. I begin by formalizing these mechanisms, proceed to measurement, and then report the empirical results of the horse race and discuss their implications.

3.1 Empirical Framework

There are two reasons a borrower may fail to repay: First, he may lack the funds and simply be unable to repay. Second, though he has the ability to repay, he may choose not to. The objective is to understand the latter—the probability that a borrower repays given his observable attributes X_{it} and given that he can repay. I denote this quantity $Pr(\text{does}|\text{can},\ X_{it})$. In particular, the aim is to separately identify, *homo socialis* and *homo economicus* behavior as they impact $Pr(\text{does}|\text{can},\ X_{it})$.

It is important to emphasize that the quantity of interest, $Pr(\text{does}|\text{can},\ X_{it})$, is quite different from a seemingly related quantity, the probability that a borrower repays conditional being able to, *and additionally conditional on having received a loan in the first place.* Simply regressing whether made loans are repaid will recover properties of the latter, not the former. As will be formalized shortly, the intuition is that ex-ante, lenders select borrowers they expect will repay. Looking at ex-post outcomes on this ex-ante selected sample is informative about lender mistakes rather than borrower choices. Consequently, the analysis begins at a prior step, with the lending decision.

3.1.1 Lending Decision

The critical assumption is that lenders decide whether to lend based only on the loan's expected return. In particular, letting R_{it} be the interest rate, X_{it} be characteristics of the borrower observable to the lender, and ϵ_{it} be a random shock to the lender's desire to lend, the loan is made if and only if

$$(1 + R_{it})Pr(\text{repay}|X_{it}) > \epsilon_{it} \tag{1}$$

$Pr(repay|X_{it})$ is the lender's best guess of the probability that the borrower will repay given everything that the lender knows about the borrower. It can be mechanically decomposed as

$$Pr(\text{repay}|X_{it}) = Pr(\text{can repay}|X_{it}) \times Pr(\text{does repay}|\text{can repay},\ X_{it}) \tag{2}$$

$Pr(\text{can}|X_{it})$ captures the borrower's fundamental *ability* to repay; $Pr(\text{does}|\text{can},\ X_{it})$ captures the borrower's choice to repay given the ability to do so. For the borrower to repay, he must (1) be able to repay, and (2) choose to repay given that he can. If $\log \epsilon_{it} \sim \mathcal{N}(0, \sigma_e^2)$, then plugging (2) into (1) and taking logs yields:

$$P(lend_{it}|X_{it}) = \Phi\left(\frac{\log(1 + R_{it}) + \log Pr(\text{can}|X_{it}) + \log Pr(\text{does}|\text{can},\ X_{it})}{\sigma_\epsilon}\right) \tag{3}$$

Equation (3) implies that a probit regression where the left-hand-side variable is whether the borrower lends, with the right-hand-side offset by $\log(1 + R_{it}) + \log Pr(\text{can}|X_{it})$ will recover the parameters of $\log Pr(\text{does}|\text{can},\ X_{it})$. The parameters of $\log Pr(\text{does}|\text{can},\ X_{it})$, and their interpretation, are discussed below.

3.1.2 Repayment Decision

This section details the decomposition of log $Pr(\text{does}|\text{can},\ X_{it})$ into components that I interpret as *homo socialis* and *homo economicus* behavior. To begin, an important feature of this market is that there are good-faith and bad-faith borrowers.[22] A bad-faith borrower, denoted as type B, is one with no plans whatsoever to repay, and whose sole objective is to take the money and run. A good-faith borrower, denoted as type G, is not guaranteed to repay given the chance, but will do so if institutional penalties are able to sufficiently constrain his behavior. In effect, a good-faith borrower is a *strategic* borrower.[23]

Lenders would like to avoid bad-faith borrowers altogether, and lend to those good-faith borrowers who appear likely to repay. Lenders observe information about borrowers, some of which will be informative regarding what would cause a good borrower to repay and some of which will be informative regarding whether a borrower is a good-faith or a bad-faith borrower. I detail these inferences below.

Bad-faith borrowers with loans by assumption do not repay. This further implies that a history of repayment perfectly separates good types from bad types. Mathematically,

$$Pr(\text{does}|\text{can}, B, X_i) = 0 \qquad \text{(Bad does not repay)}$$

$$Pr(B|X_i, Rep > 0) = 0 \qquad \text{(Any repayments imply not bad)}$$

These assumptions buy considerable identifying power off of the presence or absence of past repayments. In particular, among borrowers with no repayment history, the probability of repayment is the probability the borrower is good times the probability that a good borrower chooses to repay. Among borrowers with a positive repayment history, the probability of repayment is simply the probability that a good borrower chooses to repay, because in this case we are sure the borrower is a good-faith type. Mathematically, conditional on observed repayments,

$$Pr(\text{does}|\text{can}, X_i, Rep) = \begin{cases} Pr(\text{does}|\text{can}, G, X_i) \times Pr(G|\text{can}, X_i, Rep = 0) & Rep = 0 \\ Pr(\text{does}|\text{can}, G, X_i) & Rep > 0 \end{cases} \tag{4}$$

What is left, therefore, is to put structure on the strategic reasons that a good-faith borrower

[22] As will be discussed shortly, the community has in place a number of institutional frameworks designed explicitly to deal with these problems.

[23] There could also be good-faith *non-strategic* borrowers who repay whenever possible with no strategic considerations. The presence (or lack thereof) will be picked up in a constant term in the regression.

repays, $Pr(\text{does}|\text{can}, G, X_i)$, and on the lender's assessment of whether a borrower with no repayments is a good-faith type or not, $Pr(G|\text{can}, X_i, Rep = 0)$.

Good-faith borrowers behave strategically according to the motivations outlined by Ellickson (2016). In particular, they compare the benefit of default to the various costs of default. *Homo socialis* and *homo economics* are characterizations of these costs. Letting $\mathcal{B}$ denote the benefit of default and μ denote a latent (constant) cost,[24] let $\mathcal{S}$ denote the *homo socialis* costs and $\mathcal{E}$ denote the *homo economicus* costs. A good-faith borrower repays if:

$$\mathcal{B} \leq \mu + \mathcal{S} + \mathcal{E} \tag{5}$$

$\mathcal{B}$, the benefits of default are clear: the borrower keeps the face value of the loan, F, that he would otherwise pay. I assume that benefits have the following functional form, although the results are robust to alternate specifications:[25]

$$\hat{\mathcal{B}}(X_i) \equiv \beta_d \log F$$

$\mathcal{S}$, the social costs of default capture the direct utility loss a borrower suffers when defaulting due to his social standing, embeddedness in the community, or his social connections more generally. Empirically, I measure social connections in two ways: (1) the borrower's numerical *karma* on Reddit, which accumulates the times other users have approvingly voted the user's comments or submissions on the site, and (2) the borrower's account's age in years on Reddit. Both quantities are observable to lenders. With k denoting *karma* and a denoting age, I assume that $\mathcal{S}$ has the following functional form:[26]

$$\hat{\mathcal{S}}(X_i) \equiv \beta_k \log k + \beta_a a$$

$\mathcal{E}$, the economic costs of default, capture the forward-looking value of lost access to the site in the case of default. By revealed preference, the prospective borrower is in this market attempting to borrow for a reason; namely, that the benefits of borrowing in the market exceed those of borrowing outside the market. Denote by $u_A > 0$ the borrower's per-period utility from having access to the market. To give a simple economic interpretation, u_A is

[24] Due to say, self-control, which I tentatively label *homo moralis* behavior. Note that in the specification, μ is specified as a constant but is not well identified, and consequently I do not emphasize it, although there is a lot that could be said.

[25] The log specification captures decreasing marginal utility of theft, and also roughly transforms the distribution of request sizes into a roughly normal-shaped distribution.

[26] This particular function form essentially transforms the empirical distribution of karma and age into distributions that are not dominated by outliers. As with the function form for default value, the results are robust to alternate specifications.

the utility gain on top of the best outside option of either not borrowing and being unable to smooth consumption, or borrowing from another source that is either more expensive or undesirable for some other reason such as speed or convenience.

Recall that a salient feature of this market is repeat borrowing. That is, conditional on borrowing once, borrowers are likely to attempt borrowing in the future. Let $\lambda_i \in [0, 1]$ denote the borrower's per-period probability of needing market access. When making loans, borrower and lender communicate privately, with the borrower typically demanding proof of the borrowers bills. This suggests that λ_i and u_A are observable to the lender and difficult to fake.[27] Letting β denote the borrower's subjective discount factor, the borrower's lifetime value from having market access, denoted by $\mathcal{E}(X_{it})$ is given by the following recursion:

$$\begin{aligned} \mathcal{E}(X_i) &= \lambda_i u_A + \beta \mathcal{E}(X_i) \\ &= \frac{1}{1-\beta} \lambda_i u_A \end{aligned} \tag{6}$$

Note that the importance of the *homo economicus* mechanism, if it is active, is linearly increasing in λ_i. Measuring λ_i presents a problem. Ideally, we would simply observe whether a particular borrower in fact returns looking for a another loan in the future, and test whether borrowers who return more in the future are more likely to obtain loans today. This approach suffers from an endogeneaty problem. In particular, whether a borrower returns depends crucially on (1) whether the borrower's current loan is funded and (2) whether the borrower in fact repays. Consequently, there is two-way causality between ex-post subsequent requests and the ex-ante lending decision.[28]

To overcome this problem I exploit the textual richness of the dataset. In particular, at the time of requesting a loan, borrowers typically describe the reasons for their request. Some reasons imply that the borrower is likely to have repeated needs. For instance, loans requested for "bills" or "groceries" suggest that the user is chronically behind on expenses and will need funds again in the future. Conversely, loans requested for "emergencies" or "until I start work" suggest that funds are for a one-time need. To formalize and automate this process, I make use of a text-analysis machine learning algorithm, the details of which are

[27]Lenders often request photographs or other proof of borrowers' bills. How easy this is to fake will show up in differences in how economic factors impact lending for new and repeat borrowers. Empirically, there is little difference between the two, suggesting there is not very much deceptive faking going on.

[28]It is not clear which way this effect will bias estimates: On one hand, borrowers who succeed in obtaining funds and repay are likely to be happy and return. On the other hand, borrowers who fail in obtaining funds may make another request shortly after the first. Empirically, the second effect appears to dominate. See Column (1) of Table 2.

described the Appendix, section 6.2. This appendix also contains a number of out-of-sample tests on the algorithm. The upshot is that I obtain an ex-ante measure of the probability of a subsequent borrower request, $\hat{\lambda}$, which performs well at predicting subsequent requests out of sample and does not suffer from the endogeneaty problem associated with using subsequent requests directly. This gives rise to the empirical measurement of $\mathcal{E}(X_i)$:[29]

$$\hat{\mathcal{E}}(X_i) = \beta_{\mathcal{E}} \hat{\lambda}_i$$

With these assumptions and measurements for *homo socialis* and *homo economicus*, we arrive at the specification for $\log Pr(\text{does}|\text{can}, G, X_i)$:

$$\log \hat{Pr}(\text{does}|\text{can}, G, X_i) = \mu + \beta_k \log k + \beta_a a + \beta_{\mathcal{E}} \lambda_i - \beta_d \log F \tag{7}$$

To reiterate the interpretation, β_k or $\beta_a > 0$ imply that the *homo socialis* channel is active in convincing good-faith borrowers to repay; $\beta_{\mathcal{E}} > 0$ implies that the *homo economicus* channel is active in convincing good-faith borrowers to repay.

Population Dynamics: The other key quantity to define is the probability that a borrower with no repayments is a good-faith type, $Pr(G|\text{can}, X_i, Rep = 0)$. In principal, any borrower observable could signal the borrower's type, including those noted above as relevant for whether a good type repays. The critical difference, however, is that signals regarding borrower type only matter before the first repayment, after which by assumption borrower types *who have repaid* are perfectly separated.[30] Consequently, both $Pr(G|\text{can}, X_i, Rep = 0)$ and $Pr(\text{does}|\text{can}, G, X_i)$ are identified even though they share common covariates; identification comes through their interaction with repayment history. Consequently, I assume that $Pr(G|\text{can}, X_i, Rep = 0)$ has the same functional form as $Pr(\text{does}|\text{can}, G, X_i)$, albeit with different coefficients to be estimated:

$$\log \hat{Pr}(G|\text{can}, X_i, Rep = 0) = \mu^s + \beta^s_k \log k + \beta^s_a a + \beta_{\mathcal{E}} \lambda^s_i \tag{8}$$

[29]In several specifications I parameterize this quantity as

$$\hat{\mathcal{E}}(X_i) = \beta_{\mathcal{E}} \lambda_i \log F$$

I.e., multiplying by the log of the loan size. The assumption here is that borrowers requiring larger loan sizes benefit from from continued access to the site. The results are qualitatively unchanged with this specification, which is reported alongside the main specification.

[30]Borrower types who have defaulted are not separated, because a borrower may have defaulted because he is a bad type or because he is a good type who was either unable to repay or choose not to repay. Empirically borrowers who have defaulted are unlikely to return and their numbers are not empirically relevant.

Measuring repayment ability: Though not the focus of this paper, an important control is the borrower's ability to repay, independent of his choice to do so. That is, I need an estimate of $Pr(\text{can}|X_{it})$. Here I follow a machine learning strategy similar to that for $\hat{\lambda}$. In short, to estimate $Pr(\text{can}|X_{it})$, I condition on loans that have been made, and use all available data regarding the user's post text, borrowing history, and so on, to predict whether the loan is in fact repaid. The assumption is that ex-post, lenders only make loans that default for "fundamental" reasons, i.e., for reasons outside of the borrower's control. It is possible to test whether any of the ex-ante predictors of the *choice* to repay predict loan default ex-post, which they do not.

Putting it together: Combining Equations (3), (4), (5), and (8) suggests that a probit regression of whether the loan is made will recover the coefficients of interest. For robustness I use both a linear probability model and a probit model. I discuss the results and their implications in the following section.

3.2 Results

The OLS estimates are in Table 2. The results of the structural probit specification are shown in Table 3. Column (1) of both tables uses the raw (endogenous and uninstrumented) measure of whether a borrower in fact requests a subsequent loan. Columns (2), (3), and (5) use the reduced form, machine-learned measure $\hat{\lambda}$ directly, and columns (4) and (6) instrument for ex-post requests using $\hat{\lambda}$. Columns (1) and (2) do not separately identify $\hat{Pr}(\text{does}|\text{can}, G, X_i)$ and $\hat{Pr}(G|\text{can}, X_i, Rep = 0)$ using the strategy of conditioning by numbers of repayments; the rest do. Columns (5) and (6) use an alternate specification of the *homo economicus* factor that includes current loan size.

First, as is to be expected, there is a strong negative relationship between loan size and the probability of the loan being filled, a strong positive relationship between the predicted ability of the borrower to repay and the probability of the loan being filled, and a strong positive relationship between whether the borrower has repaid in the past and the probability of the loan being filled.

In Column (2), which does not attempt to separately identify the probability that a good-faith borrower repays and the probability that a borrower is a good-faith type, there is clear evidence in both the linear probability model and the probit specification that the probability of needing a loan in the future is strongly predictive of his loan request being filled. That is, the *homo economicus* channel is clearly active. The results are more nuanced regarding the *homo socialis* channel, however. The linear probability model shows no effect of outside social activity (*karma*) in the lending decision, and only a very weak effect of borrower age. Outside activity appears to matter more in the probit specification, but the effect is still small. While Column (2) is not meant to carefully identify the factors driving ex-post repayment and disentangle actual incentives to repay versus screening, it is worth keeping in mind as a descriptive baseline result. To obtain economically interpretable results, I move on to the other specifications.

Columns (3)-(6) estimate separate coefficients by whether the requesting borrower has previously repaid. As discussed earlier, this allows for identification of the role that borrower observables play in causing good types to repay separately from their role in signaling that borrowers are good types. The coefficients on $X \times$ Has Not Repaid measure the combined effect of observable X in inducing good types to repay and signaling good types; the coefficients on $X \times$ Has Repaid measure only the effect of observable X in inducing good types to repay; consequently, their difference is the signaling effect. To discuss the results, I focus on the linear probability specification column (4), although the results are qualitatively similar in other specifications.

I begin with the factors that predict good-faith borrower repayment, given by the interactions of $X \times$ Has Repaid. TRegarding the *homo economicus* channel, the coefficient on whether the borrower requests a subsequent loan is roughly 0.60, meaning that other things equal, a borrower who will surely request another loan is 60% more likely to receive a loan in the first place. With the interpretation that we are measuring directly a good-faith borrower's propensity to repay, this implies that recurring economic needs are strongly predictive of future repayment. That is, the *homo economicus* channel is an economically significant and meaningful channel.

Regarding the proxies for *homo socialis* channel, there appears to be weak evidence that it is active in encouraging good-faith borrowers to repay. The coefficient on outside social participation is positive but not statistically significant, and the coefficient on borrower age is significant only at the 10% confidence level.[31] These results suggest that a borrower's degree of social ties or embeddedness do little to meaningfully encourage repayment. It is important to note that this does not mean that the borrower's social ties are unimportant for his getting a loan. Rather, it means that they do not impact his ex-post decision to repay. I return to the role of the borrower's social ties shortly.

[31] Note in other specifications it tends to be even closer to zero.

Moving on to the factors that predict whether a borrower is a good-faith borrower. Recall that the difference in coefficients between has-repaid and has-not-repaid borrowers identify these channels. The coefficient on whether the borrower requests a subsequent loan is roughly 0.45 for has-not-repaid borrowers versus 0.60 for has-repaid borrowers, which implies that first-time borrowers who are more likely to request a subsequent loan are less likely to be good-faith types.[32]. Further, it is among first-time borrowers where social ties become important. The coefficient on borrower age becomes larger and highly statistically significant, implying that age signals that the borrower is a good-faith type. The coefficient on borrower social activity also becomes larger and marginally statistically significant.

3.3 Discussion

The results provide robust evidence about the varying roles of *homo economicus* and *homo socialis* channels in facilitating repayment. In particular, economic motivations are central to good-faith borrowers' choices to repay, and social ties play little to no role. Social ties are important in the market functioning, but they manifest in a screening, rather than third-party enforcement role.

Why do good-faith borrowers repay? Good-faith borrowers, which here mean borrowers who are in theory susceptible to some kind of social control, appear to be clearly motivated, at least in part, by matters of pure forward-looking economic rationality. The results show that borrowers who value future market access—borrowers who are likely to ask for loans in the future—are significantly more likely to receive loans and therefore by implication are significantly more likely to repay them ex-post. These are the borrowers for whom market exclusion from the community, a form of bootstrapped third-party punishment, is most potent.

The fact that social ties matter so little for the repayment decision is surprising in one way and unsurprising in another. On one hand, many studies, particular those looking at microlending (see, e.g., Feigenberg et al. (2013)) in developing villages, have shown that social ties can play a very strong role in the ex-post decision of whether to repay microloans. That it has essentially no impact here is a striking difference. On the other hand, given the institutional setting, a community with no face-to-face interaction where repeated social interaction tends to be rare, it is not surprising that this social channel has little bite. What

[32]The difference here is not statistically significant but in most other specifications it is larger and statistically significant

this suggests is that while social collateral *can* be very helpful as an enforcement mechanism, it is far from necessary. Online communities, in particular, do not need to rely on it, and further this suggests that the broadening changes to communities as the move online will not spell the end of extralegal contracting or trust more generally

I argue that this economic form of social control is among the most robust to institutional changes and different institutional settings. The bare-bones requirements for this mechanism to work are (1) a market where participants tend to have future needs and (2) a market where information transmission is possible. This represents exactly the kinds of markets studied in Bernstein (1992) (Diamond Dealers) or Greif (1993) (Medieval Merchants). While these two examples additionally had other elements possibly present, such a social ties or shared morals, the results from this study suggest that while these things may help, they are not absolutely essential. Rather, a group of Diamond Dealers possessing only a benefit from dealing with each other in the future and a method for flagging bad dealing would be sufficient. Similarly, a group of Medieval Merchants with their network structured in such a way as to permit sharing of merchant misconduct (See Burt (2005) and Bernstein (2017) WIP) can function in largely the same way.

A community that relies on economic ties rather than social ties tends to benefit from, rather than be weakened by, technology that makes the community larger. While a larger community can dilute social ties, to the extent that social ties are unimportant, this does not matter. On the other hand, a larger community means there are more potential future partners for transaction, which raises the value of being inside the community, making defaulting more costly.

Finally, it is worth noting that the results in this particular likely underplay the importance of the economic motivation. In this setting, the value of continued market access was u_A, which represented the period utility from being inside the market relative to the next best option. Because there are many substitutes for short-term consumer credit, it is likely that the difference between continued access and the next best alternative is low. Consequently, the estimated importance of these economic drivers is somewhat smaller. In contexts where there are fewer outside options or worse outside substitutes, one would expect this force to be stronger. Despite this, however, the positive results here show that the economic motivation is still quite strong and important for a functioning market, and that agents, both borrowers and lenders, take this economic channel seriously when making decisions.

How does the community identify good-faith borrowers? A critical element in all of these extralegal settings is the ability for a community to screen potential borrowers ex-ante for their susceptibility to ex-post punishment. In the simple model presented in this paper, I simplify the problem to one of identifying bad-faith borrowers, who are assumed to be simply immune to any kind of punishment and will default on any loan they receive, and good-faith borrowers, who behave strategically. The problem of identifying the good from the bad is particularly important in this quasi-anonymous online community because the free-recreation of new anonymous accounts makes it possible for a bad-faith borrower to repeatedly get a loan, default, recreate his account, and repeat the process. It is here that social ties play a large role. Importantly, while a new account is free to recreate, it is costly to acquire social ties.

The costly acquisition of social ties allow them to play the role of screening device. In particular, suppose a bad-faith lender receives benefits $\mathcal{B}$ from defaulting. To prevent the lender from repeatedly borrowing and defaulting, the community must require him to pay some cost $\mathcal{C} > \mathcal{B}$ before having the opportunity to borrow again. If the cost in terms of time and effort of acquiring the requisite (an equilibrium quantity) social ties is sufficiently large, this makes it not worthwhile for the borrower to engage in repeated deception. This system works to screen bad-faith borrowers and let in good-faith borrowers because good-faith borrowers' benefits can be much larger than the one-time benefit of default: They are the lifetime benefits of ongoing economic access $\mathcal{E}$.

Finally, it should be emphasized that while in this community, social ties have been bootstrapped into playing the role of costly entry mechanism, in other contexts this role does not need to be filled by social ties. In fact, they could be replaced with any other costly and verifiable entry cost, such as professional qualifications, specialized education, or simply a membership fee, that is available to the community.

To conclude this section: The setting here provides a unique opportunity to test the varying roles of economic and social forces in identifying which borrowers are susceptible to community enforcement and incentivizing repayment when there is no formal legal system to do so. The fact that borrowers require financing for many different reasons and must provide proof to their lenders of their needs provides significant variation in borrowers' economic motivations to repay. In particular, by comparing borrowers who are likely to need subsequent loans to borrowers who are unlikely to need subsequent loans, I show that borrowers' self-interest in maintaining market access is of first-order importance in incentivizing repayment. Finally, though borrowers differ significantly in their social connections to the site at large, the results are clear that these connections do not play a large role in incentivizing repayment. Rather, they play a gatekeeping screening role.

4 Lender Behavior and Expertise

Having established what drives borrower repayment, a natural question is whether taking these motivations as given, the community at large lends in an efficient way. In other words, given what motivates repayment, do lenders lend to borrowers who exhibit those traits, or do the norms of lending tend to allocate capital in some other way? There are two broad questions: (1) do lenders appear to make mistakes regarding their appraisal of how borrowers make repayment decisions, and (2) are lenders with more lending experience—those more embedded within the community—have fewer defaults ex-post and demonstrate a better understanding of community norms ex-ante? The answer to (1) is particularly important as the assumption that lenders are more likely to make loans with greater expected values, and so this analysis serves the dual purpose of testing this assumption and highlighting if and where it fails.[33] The answer to (1) is that aside from small deviations, lenders appear aware of community norms and make few systematic mistakes, which lends support to the empirical strategy in the earlier part. The answer to (2) is while mistake are indeed rare, more embedded lenders display clear signs of expertise, making fewer mistakes ex-post and making ex-ante lending decisions that are more correlated with the actual mechanisms underlying community norms.

For the entirety of this section, an "experienced" lender is defined as a lender who has had at least 60 loan repayments. This cutoff is chosen so that approximately one-third of loans are made by experienced lenders and two-thirds of loans are made by non-experienced lenders. The first empirical test looks at whether lenders appear to make ex-ante mistakes regarding predicting borrowers' choice to repay, and then ask whether expert lenders—those presumably more likely to understand the factors impacting a borrowers ex-post repayment decision—are less likely to make these mistakes. The second empirical test looks for the presence of experts and roots out the source of the experienced lenders' better performance. The third empirical test directly compares the characteristics of borrowers that experienced lenders lend to to the characteristics of borrowers that inexperienced lenders lend to in order to test whether experienced borrowers in fact select on borrower characteristics shown in earlier parts to be important parts of the mechanisms driving ex-post borrower repayment.

[33]It is important to note that that this test does not depend on the earlier results or their interpretation being true.

4.1 Lender Mistakes

In this section I test whether borrowers appear to make mistakes in evaluating the ex-ante information. In particular, I test whether the characteristics that drive ex-ante loan selection are able to predict ex-post default. This is an important test for two reasons: First, an identifying assumption in identifying the mechanisms driving borrowers to repay was that lenders make minimal mistakes ex-ante. Consequently, this section provides a test for whether lenders are making "type-1" errors—lending when they should not. Second, to the extent that there are errors, it identifies which factors lenders place too much or too little weight on, which sheds light on lenders' ex-ante understanding of community norms and expectations regarding repayment and the mechanisms surrounding it.

The main specification for this section is as follows. Among made loans,

$$Repaid_{it} = X_{it}'\Gamma + \delta_t + \epsilon_{it} \tag{9}$$

Where X_{it} is a vector of ex-ante borrower characteristics. X_{it}' will include $\log Pr(\text{can}|X_{it})$, the ex-ante probability that a borrower *can* repay. We expect a positive coefficient here because even if lenders make no mistakes borrowers with higher probabilities to receive bad "fundamental" shocks ex-ante will be more likely to default ex-post.[34] More importantly, X_{it} includes the characteristics hypothesized to drive the ex-post decision to repay given that the borrower *can* repay. Here the assumption was that lenders could determine whether the borrower could choose to repay if he could, and consequently if coefficients on these variables are non-zero, it suggests that lenders should have made different ex-ante decisions.

The results of Regression (9) are shown in Table 4. All columns include time fixed effects. Columns (1), (2), and (IV) include all lenders; Column (1a) includes only expert lenders. Consistent with expectations, the coefficient on $\log Pr(\text{can}|X_{it})$ is positive and statistically significant.[35] There are two consistent mistakes that lenders appear to make: First, the coefficient on log(loanSize) is negative, suggesting that lenders tend to underestimate the relationship between loan size and borrowers' benefits of default.[36] Second, the coefficient on

[34] In terms of interest rates, Appendix 6.3 shows that this coefficient also predicts lower interest rates, meaning that lenders require higher interest rates as compensation for these riskier loans.

[35] Note, this is essentially the out-of-sample test of the $\log Pr(\text{can}|X_{it})$ machine learning process.

[36] An alternate explanation is as follows: There is little variation in interest rates, because they mostly reflect fixed and administrative costs other than risk. Consequently, larger loans are more profitable to lenders holding risk constant, and consequently lenders would be willing to make riskier larger loans. Consequently, we expect larger loans to have a greater default rate in equilibrium because they are riskier on average. Though the results are not presented in this paper, controlling for either interest rate or the absolute

whether a borrower has previously repaid is positive, suggesting that lenders underestimate the informativeness of having repaid in the past. This can mean two things: Either they overestimate the moral impulses of a new, unproven borrower, or they underestimate the moral impulses of an old, proven borrower.

It is worth noting, however, the magnitude of the coefficients on these variables is quite small relative to their importance in making a loan. Comparing Table 2 to Table 4, the coefficient on log(loanSize) for the initial lending decision is roughly -8, while the coefficient on log(loanSize) for the repayment outcome is roughly -1. Similarly, the coefficient on past repayment for the initial lending decision is roughly 30%, while the coefficient on past repayment for the repayment outcome is roughly 5%. These differences suggest that while borrowers make mistakes in utilizing the information contained in these observables, they are very small relative to the way in which they use these variables for the initial lending decision. This supports the identifying assumptions discussed earlier.

The important exception to these findings come among experts, which appear to not make mistakes. That is, the coefficients on all variables are statistically no different from zero, with the exception of $\log Pr(\text{can}|X_{it})$, which is to be expected. That is, the outcomes on experts' loans line up much more closely as would be expected from their initial lending decision. I now analyze this point in detail.

4.2 Experienced Lender Performance

Motivated by the fact that there are at least some lender mistakes but lenders with more experience appear to make fewer mistakes, this section tests whether experienced lenders' loans perform better ex-post, and if so, what the source of their superior ex-post performance is. I run the following regression on made loans:

$$Repaid_{it} = \beta_0 + \beta_1 Experienced_l + X_{it}'\Gamma + \delta_i + \delta_t + \epsilon_{it} \tag{10}$$

Where $Repaid_{it}$ is whether borrower i's time-t loan is repaid, $Experienced_l$ is whether the lender has had at least 60 loans repaid at the time of lending, X_{it}' is a vector of borrower controls, and δ_i is a borrower fixed effect, δ_t is a time fixed effect. Table 5 Panel A. All columns contain time fixed effects to absorb broad time-variation in market outcomes.[37]

difference between loan amount and required repayment amount does not impact the results.

[37] In particular, a worry would be that mechanically lenders are "experts" later in the sample, where the repayment rate is higher for other reasons.

Additionally, column (1) includes only lender type with no borrower controls or fixed effects. Column (2) includes borrower controls. Column (3) includes borrower controls among new borrowers only. Column (4) contains borrower fixed effects.

The results for Column (1) show that unconditionally, experienced lenders' loans are roughly 6% more likely to be repaid. Given that the unconditional default rate is roughly 10%, this is a large difference. Broadly, there are three reasons that experienced lenders may perform better. First, they may select better borrowers based on commonly observed borrower characteristics—that is, they may use publicly available data better. Second, they may select better borrowers based on private information—that is, by virtue of being experts, they have developed relationships which allow them to choose better borrowers in the future. Third, they may have skill in extracting payment ex-post—that is, conditional on the borrower, the expert may be better at getting the borrower to pay him. The tests in Columns (2)-(4) help evaluate these interpretations.

Column (2), which includes borrower controls, shows that after adjusting for observable borrower characteristics, experienced lenders are roughly 3% more likely to be repaid, compared with 6% before controlling for these characteristics. This difference suggests that indeed, expert borrowers do appear to use publicly available information than inexperienced borrowers, but that this does not entirely explain the difference. I return shortly to how the experienced borrowers use this information differently.

Column (3), which only looks at first-time borrowers and includes borrower controls shows that experienced lenders are roughly 7.5% more likely to be repaid than inexperienced lenders are among first-time borrowers. While this estimate is not statistically distinguishable from the estimates in Columns (1) and (2), it is further evidence that experienced lenders' expertise does not lie entirely in relationship lending or in utilizing observable characteristics better, but that they have expertise that allows them to use unobservable data—perhaps discussions before lending—that does not simply come from having past dealings with the borrower.

Column (4), which now includes a borrower fixed effect, tests whether among the same borrower, a more experienced lender is more likely to be repaid than an inexperienced one. This tests for whether experienced lenders appear to have ex-post skill in extracting payment. The coefficient is roughly 1% but statistically insignificant, meaning that we cannot reject the null hypothesis that experienced lenders do not have any special expertise regarding ex-post payment.

To summarize these results, experienced lenders are significantly more likely to be repaid than inexperienced lenders. Breaking the difference down, it appears that roughly half of the experienced lenders' skill comes from their ability to select borrowers based on observable differences, with the balance coming from their ability to select based on unobservable differences or perhaps better extract payment ex-post. It is their ability to select based on observable characteristics that is most interesting, because it suggests that these experienced lenders may better understand the community norms that drive borrowers to choose to repay.

4.3 Experienced Lender Decision Making

Evidence from the previous two sections points to the fact that experienced borrowers better understand the ex-post repayment decision that borrowers face. To test this directly, I next test whether there are observable differences in borrower characteristics among loans that experienced lenders have made versus loans that inexperienced lenders have made. In particular, among made loans, I regress

$$Characteristic_{it} = \beta_1 Experienced_l + X_{it}'\Gamma + \delta_t + \epsilon_{it}$$

Where $Characteristic_{it}$ is one of the borrower characteristics found to impact repayment: Whether the borrower has repaid in the past, the probability of requiring another loan, the borrower's outside social activity for new borrowers, and the borrower's outside social activity for old borrowers. X_{it}' is a vector of borrower controls, including the characteristics not on the left-hand-side. δ_t is a time fixed effect. The results are shown in Table 5, Panel B. The results are strongly consistent with experienced borrowers selecting loans with characteristics corresponding more strongly to ex-post repayment choices.

An expert's borrower is roughly 25% more likely to have repaid in the past, which results show is strongly predictive of future repayment. Similarly, an expert's borrower has roughly a 2.13% higher ex-ante likelihood of requesting a loan in the future. Moreover, if the borrower is new, an expert's loan typically has greater outside social activity, which again we demonstrated was a good indicator of future repayment among new borrowers. Finally, among old borrowers, expert loans have slightly less outside social activity. This is also consistent with experts using observable information better, as earlier results showed that outside social activity among *old* borrowers was not a reliable indicator of repayment.

To summarize this section's findings: First, while lenders appear to make ex-post mistakes, for the most part their mistakes are small in magnitude. They tend to underestimate the importance of loan size and the importance of past repayment history in the repayment decision. When looking only at the most experienced lenders, evidence for systematic mistakes vanishes. Second, experienced lenders do perform better, with roughly half of their superior performance being attributable to better selection of borrowers based on observable criteria. Third, differences in experienced versus inexperienced lenders' selection tracks very well those characteristics that most predict ex-post repayment. Taken together, these results suggest that the community lenders understand well the borrowers' motivations for repayment and lend based on criteria that are consistent with how enforcement works ex-post.

5 Discussion and Conclusion

5.1 Review of Results

Section 3 studied what motivates ex-post repayment of a loan when there is no formal legal system available for enforcement. In particular, it examined the role of two potential mechanisms: A *homo socialis* mechanism, where borrowers with more social embeddedness in the community face social pressures to repay, and a *homo economicus* mechanism, where borrowers with more to lose economically if they are excluded from the market face economic pressures to repay. Additionally, this section considered the extent to which these forces act as a screening mechanism to separate borrowers who simply possess no intentions of repaying from those who are potentially controllable through these informal modes of social control.

The results show that the most powerful aspect of social control was the threat of future exclusion from the market. This market attracts, likely as an equilibrium outcome, borrowers who are likely to request funding in the future. A robust finding is that borrowers who appear more likely to need funding in the future are more likely to receive funding today meaning that they are seen as being more likely to repay. This supports the notion that borrowers value continued access to the market, because those likely to return are those who value it the most. Borrowers who would surely make another request are roughly 60% more likely to have their requests filled and by implication are roughly 60% more likely to voluntarily repay. Social connections proved to be a weak motivating force for repayment, with age and outside social activity playing little role in incentivizing strategic borrowers to repay. There is an important role for social connections, however, in screening bad-faith borrowers from good-faith borrowers.

Having established the mechanisms that induce borrowers to repay ex-post, the question turned to whether the lending side of the community internalizes these drivers when making lending decisions. Section 4 provided affirmative evidence to that effect. First, while the community appears to make some ex-ante lending mistakes by over-weighting or under-weighting certain borrower characteristics when choosing to lend, those mistakes are small and essentially disappear when focusing only on the most experienced borrowers. Second, there are significant differences in ex-post loan performance depending on the identity of the lender. In particular, the top third of experienced lenders are more likely to have their loans repaid by a significant margin, with roughly half of the margin explained by their ability to select borrowers based on ex-ante observable characteristics.

These two facts motivate a third test, which is to examine in detail the characteristics that loans made by experienced borrowers have in contrast to loans made by inexperienced borrowers. The findings suggest that more experienced lenders tend to fund borrowers with precisely those characteristics associated with inducing repayment ex-post and downplay those borrowers with less positively informative characteristics. This suggests that more experienced borrowers have a clearer understanding of the community norms and expectations regarding repayment.

5.2 Lessons for Institutional Design

As a case study, these findings suggest several lessons for the institutional design of contract enforcement, online markets, and anonymity. First, without any of the quantitive results, the mere continued existence and functioning of this market suggests that contracting entirely outside the formal legal system is not a phenomenon requiring deeply embedded social connections or cultural homogeneity. Unlike the diamond merchants or Nantucket whalers, social ties in this market are extremely weak and play essentially no quantitative role in encouraging repayment. Rather, economic self-interest, an intrinsic motivation to repay, and basic barriers to entry to keep out dedicated cheaters are sufficient for the market to work.

Homo Economicus* beats *Homo socialis: In the horse-race of third-party social control, there was strong evidence that forward-looking, "rational" economic motivations play a large role in repayment, while variation in social ties played essentially no role. This provides guidance on implications for these markets in the future. In particular, as more transactions move online, this movement tends to cut in two ways: On one hand, with more people inside a community, there is a greater economic value to being in the community. An online community naturally attracts more people, thus increasing the value of being inside. On the other hand, a large online community—particularly one with anonymous users—intuitively lacks the social cohesion and embeddedness that a small in-person community has,[38] and so social costs of default are potentially smaller.

That borrower behavior appears so sensitive to these economic motivations suggests that moves online towards larger but less-socially connected settings will not spell doom for extralegal contracting. It is natural to ask why the canonical examples of self-ordering, such as Jewish diamond merchants of Nantucket whalers seem to take place within small, well-connected communities. One explanation was that intimate, deeply embedded social connects are a *per se* requirement for self-ordering because the threat of social punishment is so important for ex-post enforcement. Another explanation was that these deep social connections merely enable the transmission of data regarding community members' behavior, thus making market exclusion an institutionally feasible punishment. While the direct social consequences of misbehavior may have had greater bite in those settings, the results here show that social connections as a mere information transmission mechanism are sufficient and therefore replacing intimate social connections with some other, less-intimate transmission mechanism will not necessarily destroy the market.

The clear limitation here, however, is that this *homo economicus* mechanism only works when the value of future market access is high relative to the value of default. This is clearly not the case for one-off transactions or one-off interactions within the community, and so in those cases, some other mechanism, including perhaps a stronger *homo socialis* mechanism would be necessary. Moreover, it places a severe upper bound on what can be pledged ex-post. In this case, the size of any particular loan that a borrower can commit to repay is limited by the average size of future transactions. If repayment is secured mostly by the relatively modest value of uture loans for groceries or gas, it is unlikely that the threat of losing the ability to obtain those loans could motivate the repayment of a mortgage or auto loan.

Signaling and deterring cheaters: Parallel to the economic and social forces studied in detail here is the question of the average intrinsic motivation to repay, absent economic or social pressures. The results here highlight an extremely important aspect of extralegal communities, which is the question of entry and growth. Many communities studied in the literature have highly restricted entry. Essentially, entry is either restricted outright (e.g.,

[38]Indeed, this appears to be the case here.

limited by birth[39]) or by an expensive entry cost.[40] At the very least prospective community members can never dodge their reputations and start over. On the Internet, however, with no other institutional features, entry is not restricted, and what's more, anonymity makes it trivially easy to lose a bad reputation. In a world where some people—even almost all people—have moral scruples and would not cheat, the presence of even a tiny fraction of cheaters can potentially destroy the entire community unless there are some barriers to entry. In particular, a few bad apples can create multiple accounts for free, thereby pushing the unconditional probability of a repayment for new borrowers to zero. With no way to take on new borrowers, the community could not grow.

This community has therefore levered institutional features of the site to create its own barriers to entry. In particular, it requires borrowers have a demonstrated history of on-site participation. The content of the participation does not matter,[41] rather, all that matters is that it has taken place. This time-consuming participation requirement is sufficient to deter repeat cheaters. The key requirement is that the cost of doing the necessary participation be higher than the value of one-time theft[42] and lower than the value of repeated market access.[43] While social participation is a natural candidate for this costly barrier to entry here, I emphasize that in other contexts, it could be replaced with direct fees, certification, specialized education, or a similar verifiable and costly activity.

Of course, signaling is costly and the requirement of doing so leads to an outcome worse than the unconstrained first-best. Therefore, it is worth considering institutional design changes that could offer welfare improvements. A simple institutional design change that would remove this ability while still maintaining outward anonymity would be to remove this recreation ability though a centralized but still-anonymous database linking user accounts to a real-world observable.[44] The effect would be to eliminate the need for costly signaling through on-site participation, because a new account would not be a clear signal of a scammer; rather, the scammer probability of a new account would approach more clearly the unconditional probability.

This provides a more nuanced view on Internet privacy. There is a sense that total anonymity

[39] Jewish diamond merchants as in Bernstein (1992), Nantucket whalers Ellickson (1989)

[40] Being a remote cattle rancher (Ellickson (1986)) or being a star soccer player (Ellickson (2016)).

[41] The results are not included in this paper, but for example, the participation in political, pornographic, or drug-related communities does not impact borrowers' outcomes.

[42] The incentive compatibility constraint

[43] The individual rationality constraint.

[44] Say, a drivers license number.

on the internet breeds bad behavior,[45] and that therefore removing anonymity and putting a real-world face to online behavior would be beneficial. These results suggest another approach, which is does not go as far: Rather than tying Internet activity to a real-world identity, tie all activity of a single user together, without the last step of the real-world connection. Here, the basic problem with anonymity is the combination of free entry with the existence of a tiny minority of scruples-free borrowers. Account recreation allows these few bad apples to multiply their numbers and overwhelm the market; thus creating the need of a signaling mechanism. By tying users down to a single online persona, even if it remains anonymous, this multiplicity problem is eliminated.

Directions for future research: Two aspects of this market were conspicuously absent from this analysis. First, perhaps more relevant to economics than to law, is how interest rates are determined. The problem is somewhat more complex than a simple risk-return tradeoff due to the ex-post payment choice. Essentially, a borrower cannot simply offer a higher interest rate in exchange for a loan, as the types of borrowers most likely to prefer a higher contracted interest rate are exactly those who will not repay. I conjecture that interest rate setting plays an important signaling role, but this analysis is far from complete.

Second, while I focused on community-level motivations such as social harms or economic exclusions, I put aside mechanisms coming from the particular lender in the borrower-lender relationship. While the lender is also a part of the community in general, his proximity to the transaction gives rise to come important differences. In particular, while default is treated roughly the same by the community and the actual lender, there is evidence that other forms of non-performance, such as late payments, are penalized much more by the specific lender and much less by the community at large. Investigating this difference could yield payoffs in terms of what it means to be "law" when only some parties enforce certain contract terms, why some terms are enforced only by inside parties, and whether improvements in formal enforcement are always positive, especially when they may crowd out the possibility of second-party enforcement of certain contract terms.

Third, by utilizing variation in times between when the loan is requested and when it is filled, it would be possible to study how lenders use different types of information, particularly when some information is presented in a clear and easy to observe way, and when some information is publically available to more difficult to obtain. There are long-standing questions regarding disclosure requirements and how information is presented to consumers, particularly in the context of consumer finance. Studying how lenders, especially those of different experience

[45] See, e.g., Levmore and Nussbaum (2010).

and ability, appear to use easy-to-access and hard-to-access information could shed light on some questions in this debate.

References

Ambrus, A., Mobius, M., and Szeidl, A. (2014). Consumption risk-sharing in social networks. *The American Economic Review*, 104(1):149–182.

Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74. ACM.

Bakshy, E., Rosenn, I., Marlow, C., and Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web*, pages 519–528. ACM.

Benmelech, E. (2009). Asset salability and debt maturity: Evidence from nineteenth-century american railroads. *Review of Financial studies*, 22(4):1545–1584.

Benmelech, E., Garmaise, M. J., and Moskowitz, T. J. (2005). Do liquidation values affect financial contracts? evidence from commercial loan contracts and zoning regulation. *The Quarterly journal of economics*, 120(3):1121–1154.

Bernstein, L. (1992). Opting out of the legal system: Extralegal contractual relations in the diamond industry. *The Journal of Legal Studies*, 21(1):115–157.

Bernstein, L. (1996). Merchant law in a merchant court: Rethinking the code's search for immanent business norms. *University of Pennsylvania Law Review*, 144(5):1765–1821.

Bernstein, L. (2000). Private commercial law in the cotton industry: Creating cooperation through rules, norms, and institutions. *Mich. L. Rev.*, 99:1724.

Bernstein, L. E. and Parisi, F. (2014). Customary law: An introduction. *Customary Law and Economics*.

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Oxford university press.

Cabral, L. and Hortacsu, A. (2010). The dynamics of seller reputation: Evidence from ebay. *The Journal of Industrial Economics*, 58(1):54–78.

Ellickson, R. C. (1986). Of coase and cattle: Dispute resolution among neighbors in shasta county. *Stanford Law Review*, pages 623–687.

Ellickson, R. C. (1989). A hypothesis of wealth-maximizing norms: Evidence from the whaling industry. *Journal of Law, Economics, & Organization*, 5(1):83–97.

Ellickson, R. C. (2016). When civil society uses an iron fist: The roles of private associations in rulemaking and adjudication. *American Law and Economics Review*, page ahw011.

Feigenberg, B., Field, E., and Pande, R. (2013). The economic returns to social interaction: Experimental evidence from microfinance. *The Review of Economic Studies*, page rdt016.

Greif, A. (1993). Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. *The American economic review*, pages 525–548.

Hardy, R. A. and Norgaard, J. R. (2016). Reputation in the internet black market: an empirical and theoretical analysis of the deep web. *Journal of Institutional Economics*, 12(03):515–539.

Ho, T. K. (1995). Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE.

Holmes, O. W. (1897). *The path of the law.* The Floating Press.

Levmore, S. and Nussbaum, M. C. (2010). *The offensive Internet: Privacy, speech, and reputation.* Harvard University Press.

Posner, E. A. (2009). *Law and social norms.* Harvard university press.

Rost, K., Stahel, L., and Frey, B. S. (2016). Digital social norm enforcement: Online firestorms in social media. *PLoS one*, 11(6):e0155923.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, 8:321–340.

6 Appendix

6.1 Data Collection

Data are scraped from Reddit.com using a Python script with a significant amount of post-processing in R. Reddit's API allows users to obtain the entire (non-deleted) history of posts, both for subreddits (like r/loans and r/borrow) and for specific users, like u/trigonoah. The raw scraped data include the main submission as well as subsequent comments added to the submission. For each submission or comment I observe title, text, user, and timestamp. The scraped data are saved for post-processing.

From the raw data I construct a transaction history. Submissions identify by a standardized tag whether they are requests or reports of outcomes like repayment or default notifications. For each request, I parse the request amount, time to be repaid, and interest rate, if one is posted. For purposes of community record keeping, reports of loan fills are recorded as comments on the initial request thread with a standardized syntax: The lender writes "$loan 100" to indicate filling a $100 loan, and the borrower writes "$confirm 100" to confirm that the transaction has been completed. Due to this standardization, creating the data set of loan requests and fills is straightforward for an automated script.

Matching repayments and defaults to filled requests is somewhat more difficult because although there is a standardized system for reporting outcomes, the system does not require referencing the original transaction. Therefore, there are several stages of manual matching that the post-processor attempts. The script proceeds in this order and stops upon success.

1. If the outcome post contains a hyperlink to a known request post, this is the match.

2. If the outcome post contains a hyperlink to an unknown request post (it would be unknown because the original post was deleted, and the initial scrape does not return deleted posts), deleted posts are available via a hard link (i.e., following a hyperlink rather than through reddit's API), which another python scraper follows and records the appropriate information.

3. If the outcome post contains no hyperlink:

   (a) Look for past unmatched request posts that match on borrower, lender, and amount.

(b) If nothing, look for past unmatched request posts that match on borrower and lender, and take the request with the nearest amount, so long as the amount is with 30%.[46]

(c) If nothing, look for past unmatched request posts that match on borrower, and take the request with the nearest amount within the 30% bandwidth.

4. If no match can be found, stop.

This process leaves 1,187 outcome reports unmatched, representing 16% of the total reported outcomes. Typically, matches fail because the borrower has his account, and the outcome report contains no followable hyperlink to the original request. Fortunatley, the unmatched outcome reports have a repayment rate of 92%, roughly the same as the matched outcome reports' rate of 94%, suggesting that there is minimal selection in the matching process that biases the reported repayment rate significantly in one way. User post history is collected in a similar manner: The script collects all unique borrowers and lenders and downloads their on-site posting history.

To arrive at the final data set used for the majority of the tests in the paper, the following exclusions are made.

1. There are 17,070 successfully matched transactions.
2. I exclude transactions lacking a reported loan size, leaving 14,942 transactions.
3. One quarter of these observations are used to train the machine learning algorithms.
4. Transactions lacking post text, or posts reporting to have been prearranged, are excluded.
5. This leaves 9,474 observations.
6. Due to lower data quality with the /r/loans subsample, I use only /r/borrow loans, leaving 6,750 observations.

[46]This bandwith handles cases where one amount includes an interest payment and another does not. For example if B requests 100 and does not report the interest payment, and L fills the loan, if L reports that B repaid 120, this is a match because it is likely that the agreed-on repayment was 100 plus an interest payment of 20.

6.2 Text Analysis and Machine Learning

This section describes how I estimate $\hat{\lambda}(X_{it})$ and $\hat{\rho}(X_{it})$. See Ho (1995) for technical details of the Random Forest, the machine learning algorithm used here. This paper used an R implementation of the Random Forest by Breiman, Cutler, Liaw, and Wiener.[47]

6.2.1 Approach

The objective is to obtain an estimate of $\hat{\lambda}(X_{it})$ that is not biased and is orthogonal to other reasons for repaying. The most obvious candidate to measure how often a borrower gets a liquidity shock is to look at the frequency of his past requests or his future requests. Unfortunately this is extremely endogenous for many reasons: (1) if a borrower's requests are frequently rejected, he will stop asking, so we will observe many past requests for borrowers who tend to be good debtors for unobservable reasons. (2) looking at future requests will bias towards acceptance because a borrower is unlikely to request in the future if the current loan is rejected. (3) looking at past requests will have a similar problem, and (4) looking at past requests will make it impossible to compare outcomes for borrowers' first requests. So we look for something different.

I first remove posts where:

1. The post text suggests the loan was privately pre-arranged between a borrower and a lender, meaning it contains, *arranged*, *already*, *record*, *set up*, *pre*, *arranged.*

2. The post text is "bad" in some way, meaning it is equal to *0*, *#n/a*, *""*, *deleted*, *removed*, *canceled*, *no longer need*, *no longer needed*

Of the remaining posts, I partition the data into a training set and a testing set by randomly selecting one third of the borrowers into the training set and the rest into the testing set. The testing set is set aside.

In the training set, I find loan requests satisfying the following criteria:

1. The post contains valid data for the loan principle amount.

2. The post was made 60 days prior to the last date in the sample.

[47] `https://cran.r-project.org/web/packages/randomForest/randomForest.pdf`

3. The loan was made.

4. The loan was repaid.

For these posts, I determine whether there was in fact a subsequent loan request within 60 days. Criterion (2) is necessary in order to make sure all loans in the training set had the chance to request another loan within 60 days. Criteria (3) and (4) are necessary because we want to make sure all loans in the training set are "eligible" to request again: a borrower will be unlikely to ask again if his loan is not made or if the loan is made and he defaults. If we do not condition on this, the estimation could be identifying which posts are *eligible* to make another request, i.e., which requests are funded and repaid, which would bias the prediction towards loans that are made and repaid for *any* reason, rather than simply whether the borrowers are more or less likely to request a loan.

For the posts selected to be in the training set: I convert the text to lower case. I remove non-alphanumeric characters. I remove English *stop words*, which are syntactic works like *and*, *of*, or *the* that do not have semantics meaning. I strip whitespace. I stem words, meaning *asking* or *asked* become *ask*. Word order is discarded and posts are coded as baskets of words—word counts of each word in the entire corpus.

With this as the raw data, I fit a *Random Forest* model to use post text to predict whether the borrower will make a subsequent post. A *Random Forest* automatically builds a series of tall decision trees optimized to use features of the data to predict outcomes. A decision tree is tall if it has many branches, i.e., uses many features of the data to reach the classification decision. A single tall decision tree has low *bias*, i.e., it tends to classify well in-sample, but high *variance*, i.e., it tends to classify poorly out-of-sample due to overfitting. A feature of the random forest model is that it contains many different trees, and when a new classification problem is given to the forest of trees, by averaging over each tree's classification, a classification with lower variance is produced. Using the training set, the random forest model is fitted.

With the fitted model, I make out-of-sample predictions for the testing data set. The constructed variable of interest is the proportion of trees in the random forest that predict that the borrower will request another loan, which I call $\hat{\lambda}(X_{it})$.

I follow a similar procedure to calculate $\hat{\rho}(X_{it})$, the ex-ante probability of repayment conditional on receiving a loan, although, for obvious reasons, I do not require the loan to have been repaid.

6.2.2 Verification

Out-of-sample of the fitted machine learning model is very important. Because I am using this to construct a proxy for the probability of asking for another loan, the correct criteria to evaluate the fit is (1) a positive coefficient in the regression of prediction on outcome controlling for other variables to be used later in the model and (2) a statistically significant partial R^2. Table 1 shows the results of the regression and the partial F-test. The results are highly significant in the desired direction.

Finally, the results of whether loan repayment is predictable in Table 4 addresses two methodological concerns: First, one concern is that $\hat{\lambda}(X_{it})$ is being accidentally fitted specifically to predict repayment conditional on receiving a loan, which would raise concerns that we are measuring not an exogenous difference in probability of needing a loan but rather some fundamental difference that makes these loans better. Finding no significance in $\hat{\lambda}(X_{it})$, as we do here, allays this fear. Second, the lack of significance in $\hat{\lambda}(X_{it})$ implies that borrowers are correctly taking into account the probability of repayment reflected in the probability of asking for a subsequent loan. If we found significance here, we would wonder why borrowers are not fully incorporating this information and would give reason to doubt specification (3) and the identifying assumptions needed there. The lack of significance here is reassuring.

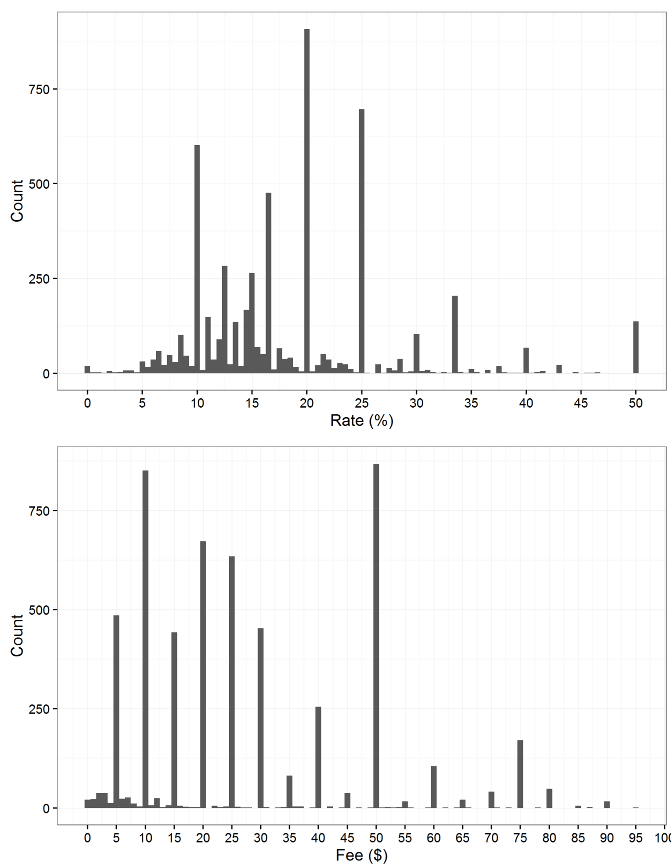
6.3 Interest Rate Determination

This appendix briefly analyzes interest rate determination and its role in the lending decision. I begin by noting that interest rates are excluded from the primary analysis because reliable interest rate measures are missing from a large number of loans. In particular, the calculation of interest rate requires a repayment amount and a loan amount (e.g., a \$120 repayment amount and a \$100 loan amount) and repayment amount is missing, equal to the loan amount for many loans, and implausibly high, likely due to data parsing errors. I study both "interest rates" and "fees," which are defined here as:

$$Rate = \frac{Repayment\ Amount}{Loan\ Amount} - 1$$
$$Fees = Repayment\ Amount - Loan\ Amount$$

Loans with missing repayment amounts, $Rate = 0$, or $Rate > 0.50$ are excluded.

Both fees and rates tend to cluster around whole numbers and multiples of five. The following figure shows the distribution of interest rates on made loans.



This begins to highlight the essential problem of determining interest rates in this market, which is that in the presence of asymmetric information regarding borrower type and the ex-post moral hazard problem, there is not an obvious monotonic relationship between the interest rate offered, the risk surrounding repayment, and whether a lender will be willing to make the loan. When borrowers are not formally bound to repay, a higher interest rate can decrease the expected return of the loan because it makes repayment less ex-post desirable for the borrower. Whats more, when comparing a borrower of a bad type to a good type, a kind of anti-single-crossing property holds, wherein types who will not repay ex-post do not care if they offer a higher interest rate, whereas types who will repay ex-post are harmed.

Moreover, to the extent that community norms imply that interest rates "should" be 20% or 25%, a borrower offering an interest rate outside the community norms requires lenders to infer the borrower's type given this off-equilibrium action. As long as lender's off-equilibrium beliefs are that such interest rate offers imply bad types (a consistent belief), then borrowers have no incentives to deviate from the community norms.

Because doing so lies somewhat outside the purpose of this paper, I do not attempt to fully analyze the interest rate setting game. Rather, I provide some simple reduced form evidence that captures (1) how interest rates are related to observables, and (2) how loan decisions relate to interest rates. The two regressions of interest are:

$$Rate_i = X_i'\Gamma + \epsilon_i$$
$$loanMade = \beta_1 Rate_i + X_i'\Gamma + \epsilon_i$$

Table 6 shows the results of this regression. Column (1) studies what determines interest rates; Columns (2)-(5) look at the impact of interest rates on the lending decision, controlling for successively more borrower and loan characteristics.

The results regarding interest rate determination show a result that is on its face puzzling but makes sense upon consideration: Loan size an interest rates are negatively correlated. While one might expect larger loans to carry a higher premium, especially because borrowers are less likely to repay large loans, this negative relationship highlights the large role of fixed costs interacting with small loan sizes plays. In particular, the overhead involved in making a $5 loan and a $500 loan—contacting the borrower, effecting the fund transfer, monitoring and collecting the loan—are not likely to be vastly different. For the $5 loan to be worth it, the borrower needs to compensate the lender for his time, thus leading to a higher interest rate.

The other important determinants of interest rate are (1) the fundamental risk of the loan, (2) the borrower's credit history, and (3) the borrower's time on reddit. First, borrowers who are more likely to be *able* to repay command lower interest rates, which is consistent with the intuition that the interest rate is a compensation for the risk of the loan. The results are weaker for characteristics shown earlier to relate to the borrower's *choice* to repay. In fact, the only covariates that appear significant are whether the borrower has repaid in the past and the borrower's age on the site. Both of these characteristics were shown earlier to be informative about borrower type. In particular, borrowers who have repaid are assumed to be good-faith types, and borrowers who are older are more likely to be good-faith types. Characteristics that aid in identifying members of the pooled good- and bad-faith types should be related to interest rates because they identify borrower repayment probabilities.

On the other hand, interest rates appear unrelated to whether the borrower will have an economic need in the future. This is also consistent with the model presented because this borrower characteristic is instrumental purely for a good-type's choice to repay, and the prediction on the relationship between this characteristic and interest rate is ambiguous: On one hand, a borrower with a strong economic need to repay is a borrower who has the ability to credibly offer a higher interest rate, because it is more likely for him to repay. However, because the borrower appears to be a better risk, he does not *need* to offer a high interest rate in order to secure the loan.

Columns (2)-(5) analyze the impact of interest rate on lending decision. Column (2) controls only for interest rate and loan size, and in this case, there is a statistically insignificant relationship between rate and whether the loan is made. This is to be expected as per the above discussion: On one hand, conditional on the borrower repaying, of course the lender would like to make a higher interest rate loan. On the other hand, it is precisely the borrowers who are least likely to repay who are most likely to make the higher interest rate loans. It is not until Column (4), after controlling for whether the borrower has repaid in the past, that higher interest rates predict a statistically significantly higher probability of lending. This is the control we would expect to make a difference, because once repaying, the borrower is able to signal that he is not a bad type. The effect is still somewhat weak and borderline significant, however, likely because higher interest rates are still associated with an aggravated moral hazard problem ex-post.

Table 1: Machine Learning Out-of-Sample Verification

Table 1: Panel A reports the out-of-sample verification of the machine learning algorithm. The sample is restricted to requests "eligible" for subsequent requests, i.e., filled and repaid requests. In particular, on the testing sample, we regress Subsequent $\text{Request}_{bt} = \beta\hat{\lambda}_{bt} + X'_{bt}\Gamma + \delta_t + \epsilon_{bt}$. $\hat{\lambda}$ is the prediction of a subsequent request. Controls are log borrower karma, borrower age, log loan size, and borrower credit history. Columns (3)-(4) include these controls. Columns (2) and (3) include the machine-learned probability of repayment. The left-hand-side variable is in percents and its mean is 57.70%. Standard errors are clustered at the borrower level. **Panel B** reports the partial F-tests.

Panel A: Out-of-Sample Validation

	Dependent variable:			
	Subsequent Request Given Fill and Repay			
	(1)	(2)	(3)	(4)
$\hat{\lambda}$	1.122***	1.032***	0.920***	0.829***
	(0.075)	(0.089)	(0.102)	(0.108)
$\hat{f}$	-	0.326**	-	0.483***
	-	(0.166)	-	(0.178)
ML Repay Controls	N	Y	N	Y
Borrower Controls	N	N	Y	Y
Time FE	Y	Y	Y	Y
Observations	3,038	3,038	3,038	3,038
R^2	0.097	0.098	0.107	0.110
Note:				*p<0.1; **p<0.05; ***p<0.01

Panel B: Partial F-Tests

	Dependent variable:				
	subsequentPost				
	DF	Sum Sq.	Mean Sq.	F	P(>F)
$\hat{\lambda}$	1	20.20	20.1986	90.4325	< 2.2e-16***
$\hat{f}$	1	13.55	13.5459	60.6470	9.321e-15***
Outside	2	6.36	3.1816	14.2445	6.963e-07***
Age	1	0.89	0.8942	4.0034	0.0454988*
log(loanSize)	1	3.86	3.8569	17.2678	3.337e-05***
Has Repaid	1	15.16	15.1626	67.8854	2.554e-16***
Has Defauled	1	0.00	0.0019	0.0087	0.925902
Time FE	18	10.88	0.6042	2.7052	0.0001298***
Residuals	3011	672.52	0.2234		
Note:					*p<0.05; **p<0.01; ***p<0.001

Table 2: Probability of Borrower Funding (Reduced Form)

Table 2: Table 2 reports the drivers of whether a loan is filled using the testing sample in a linear probability model. In particular, $Loan\_Made_{bt} = X_{bt}'\Gamma + \delta_t + \epsilon_{bt}$. Column (1) uses subsequent request directly, columns (2), (3), (5) use fitted $\hat{\lambda}$, columns (4) and (6) instrument for subsequent request with $\hat{\lambda}$. Columns (1)-(2) have no repaid-attribute interactions; columns (3)-(6) do. Columns (5)-(6) measure economic needs as size times probability, rather than probability. The left-hand side variable is in decimals; its mean is 0.62. Standard errors clustered by borrower are in parentheses.

	Dependent variable:					
	Loan Made					
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\lambda}$	-0.041^{***}	0.420^{***}	-	-	-	-
	(0.013)	(0.068)	-	-	-	-
Outside	-0.0003	0.005	-	-	-	-
	(0.005)	(0.005)	-	-	-	-
Age	0.009^{**}	0.009^{**}	-	-	-	-
	(0.004)	(0.004)	-	-	-	-
$\hat{\lambda}\times$ Has Not Repaid	-	-	0.289^{**}	0.454^{**}	-	-
	-	-	(0.121)	(0.207)	-	-
$\hat{\lambda}\times$ Has Repaid	-	-	0.501^{***}	0.599^{***}	-	-
	-	-	(0.079)	(0.124)	-	-
log(loanSize)$\times\hat{\lambda}\times$ Has Not Repaid	-	-	-	-	0.040^{**}	0.080^{**}
	-	-	-	-	(0.020)	(0.038)
log(loanSize)$\times\hat{\lambda}\times$ Has Repaid	-	-	-	-	0.145^{***}	0.186^{***}
	-	-	-	-	(0.014)	(0.030)
Outside $\times$ Has Not Repaid	-	-	0.006	0.017^{*}	0.009	0.020^{**}
	-	-	(0.007)	(0.009)	(0.007)	(0.009)
Outside $\times$ Has Repaid	-	-	0.007	0.013	0.011^{*}	0.017
	-	-	(0.006)	(0.009)	(0.006)	(0.012)
Age $\times$ Has Not Repaid	-	-	0.012^{**}	0.022^{***}	0.014^{**}	0.023^{***}
	-	-	(0.006)	(0.008)	(0.006)	(0.008)
Age $\times$ Has Repaid	-	-	0.005	0.017^{*}	0.002	0.018
	-	-	(0.006)	(0.009)	(0.006)	(0.012)
log $Pr(\text{can}\mid X_{it})$	0.459^{***}	0.344^{***}	0.336^{***}	0.246^{**}	0.259^{***}	0.152
	(0.077)	(0.078)	(0.078)	(0.107)	(0.080)	(0.117)
log(loanSize)	-0.095^{***}	-0.086^{***}	-0.086^{***}	-0.090^{***}	-0.132^{***}	-0.155^{***}
	(0.006)	(0.006)	(0.006)	(0.007)	(0.008)	(0.017)
Repaid	0.325^{***}	0.254^{***}	0.156	0.185	-0.003	-0.038
	(0.015)	(0.019)	(0.099)	(0.177)	(0.087)	(0.144)
Time FE	Y	Y	Y	Y	Y	Y
$\hat{\lambda}$ Specification	OLS	RED	RED	IV	RED	IV
Observations	6,750	6,750	6,750	6,750	6,750	6,750
R^2	0.234	0.237	0.238	-0.084	0.245	-0.394

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3: Probability of Borrower Funding (Structural)

Table 3: Table 3 shows the results of regression (3). All columns offset probabilities by $\log(1.2) + \log Pr(Can|X_{it})$. Column (1) uses subsequent request directly, columns (2), (3), (5) use fitted $\hat{\lambda}$, columns (4) and (6) instrument for subsequent request with $\hat{\lambda}$. Columns (3)-(6) have repaid-attribute interactions. Columns (5)-(6) measure economic needs as size times probability, rather than probability. Standard errors are in parentheses.

	Dependent variable:					
	loanMade					
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{\lambda}$	-0.213^{***}	2.730^{***}	-	-	-	-
	(0.058)	(0.348)	-	-	-	-
Outside	0.062^{***}	0.073^{***}	-	-	-	-
	(0.020)	(0.021)	-	-	-	-
Age	0.052^{***}	0.049^{***}	-	-	-	-
	(0.019)	(0.019)	-	-	-	-
$\hat{\lambda}\times$ Has Not Repaid	-	-	0.875^{*}	0.760^{*}	-	-
	-	-	(0.484)	(0.420)	-	-
$\hat{\lambda}\times$ Has Repaid	-	-	4.627^{***}	4.021^{***}	-	-
	-	-	(0.519)	(0.451)	-	-
log(loanSize)$\times\hat{\lambda}\times$ Has Not Repaid	-	-	-	-	0.338^{***}	0.245^{***}
	-	-	-	-	(0.088)	(0.080)
log(loanSize)$\times\hat{\lambda}\times$ Has Repaid	-	-	-	-	0.782^{***}	0.728^{***}
	-	-	-	-	(0.085)	(0.078)
Outside $\times$ Has Not Repaid	-	-	0.093^{***}	0.093^{***}	0.094^{***}	0.095^{***}
	-	-	(0.027)	(0.027)	(0.027)	(0.027)
Outside $\times$ Has Repaid	-	-	0.090^{***}	0.090^{***}	0.076^{**}	0.081^{**}
	-	-	(0.033)	(0.033)	(0.032)	(0.033)
Age $\times$ Has Not Repaid	-	-	0.053^{**}	0.053^{**}	0.062^{**}	0.062^{**}
	-	-	(0.024)	(0.024)	(0.024)	(0.024)
Age $\times$ Has Repaid	-	-	0.031	0.031	0.021	0.020
	-	-	(0.031)	(0.031)	(0.030)	(0.031)
log(loanSize)	-0.541^{***}	-0.472^{***}	-0.470^{***}	-0.470^{***}	-0.713^{***}	-0.663^{***}
	(0.028)	(0.028)	(0.028)	(0.028)	(0.037)	(0.033)
Repaid	1.716^{***}	1.237^{***}	-0.515	-0.171	0.326	0.281
	(0.061)	(0.079)	(0.523)	(0.479)	(0.447)	(0.436)
Constant	1.925^{***}	0.252	0.849^{***}	0.929^{***}	1.711^{***}	1.700^{***}
	(0.212)	(0.284)	(0.326)	(0.300)	(0.263)	(0.260)
Exact Interest Rate	N	N	N	N	N	N
Model-Imposed Offset	Y	Y	Y	Y	Y	Y
$\hat{\lambda}$ Specification	OLS	RED	RED	IV	RED	IV
Observations	6,750	6,750	6,750	6,750	6,750	6,750
Pseudo (CU) R^2	0.252	0.261	0.266	0.266	0.266	0.267
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 4: Probability of Loan Repayment

Table 4: Table 4 reports the drivers of whether a loan is repaid conditional on it being funded, using the testing sample. In particular, $Repaid_{bt} = X'_{bt}\Gamma + \delta_t + \gamma_b + \epsilon_{bt}$. Column (1), (2), and (IV) use all made loans. Column (1a) uses loans made by experienced lenders. The left-hand side variable is in percents; its mean for columns (1)-(3) and (5) is 93.97% and for column (4) is 88.03%. Standard errors clustered at the borrower level are in parentheses.

	Dependent variable:			
	Loan Repaid			
	(1)	(1a)	(2)	(IV)
log $Pr(\text{can}\|X_{it})$	22.788***	37.798**	21.209***	24.167***
	(7.970)	(16.741)	(8.058)	(8.898)
log(loanSize)	−1.225***	−0.671	−1.252***	−1.476**
	(0.427)	(0.545)	(0.428)	(0.589)
$\hat{\lambda}$	−3.099	2.192	−3.301	−0.054
	(3.833)	(3.818)	(3.827)	(0.070)
Repaid	5.890***	1.023	20.051***	6.885***
	(1.260)	(2.080)	(7.481)	(1.852)
Outside	0.196	−0.257	-	0.125
	(0.313)	(0.433)	-	(0.344)
Outside × Has Not Repaid	-	-	−0.276	-
	-	-	(0.213)	-
Outside × Has Repaid	-	-	1.422	-
	-	-	(0.887)	-
Other Loan Controls	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Experienced Lenders Only	N	Y	N	N
Observations	3,314	1,253	3,314	3,314
R^2	0.048	0.039	0.050	0.0002
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 5: Lender Expertise

Table 5: Panel A of Table 5 examines whether lender expertise is associated with a greater probability of repayment. An experienced lender is a lender with at least 60 successful repayments, comprising roughly 30% of made loans. All columns include time fixed effects. Column (2)-(4) include loan controls as in table 4. Column (3) focuses on borrowers with first-time loans only. Column (4) includes borrower fixed effects. Standard errors, clustered at the borrower level, are in parentheses. **Panel B** examines differences in made loan characteristics based on borrower expertise. In particular, among made loans, $Char_{bt} = \beta Expert_{lt} + X'_{bt}\Gamma + \delta_t + \epsilon_{bt}$. Column (1) has whether the borrower has previously repaid as the characteristic; Column (2) uses the predicted probability of a subsequent request as the charcateristic. Columns (3)-(4) use the borrowers' outside site activity, with Column (3) considering new lenders only and Column (4) considering experienced lenders only. Characteristic means are 66.30, 51.68%, 8.32, and 8.48, respectively. All columns control for the characteristics not on the left-hand-side and include time fixed effects. Standard errors, clustered at the lender level, are in parentheses.

Panel A: Lender Expertise and Default

	Dependent variable:			
	Loan Repaid			
	(All)	(All)	(New)	(All)
Experienced Lender	6.120***	3.362***	7.514**	1.272
	(0.876)	(0.867)	(3.011)	(1.062)
Loan Controls	N	Y	Y	Y
Time FE	Y	Y	Y	Y
Borrower FE	N	N	N	Y
Observations	3,308	3,308	949	3,308
R^2	0.022	0.056	0.053	0.779
Note:		*p<0.1; **p<0.05; ***p<0.01		

Panel B: Lender Expertise and Loan Characteristics

	Dependent variable:			
	Has Repaid	$\hat{\lambda}$	Outside (New)	Outside (Old)
	(1)	(2)	(3)	(4)
Experienced Lender	24.884***	2.139***	0.239**	−0.105*
	(2.176)	(0.532)	(0.116)	(0.055)
Other Loan Controls	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Observations	7,181	7,181	1,850	5,331
R^2	0.778	0.089	0.130	0.123
Note:		*p<0.1; **p<0.05; ***p<0.01		

Table 6: Interest Rate Determination

Table 6: Table 6 examines the determination of interest rates and their impact on the lending decision. Column (1) regresses interest rates on loan characteristics among made loans. Columns (2)-(5) regress whether the loan is made on loan characteristics. Standard errors in parentheses are clustered at the borrower level.

	Dependent variable:				
	Rate	loanMade			
	(1)	(2)	(3)	(4)	(5)
Rate	-	-0.102	0.104	0.262**	0.279**
	-	(0.133)	(0.126)	(0.119)	(0.119)
log(loanSize)	-0.038***	-0.107***	-0.081***	-0.082***	-0.077***
	(0.003)	(0.013)	(0.013)	(0.012)	(0.013)
$\log \hat{Pr}(\text{can}\|X_{it})$	-0.112***	-	1.196***	0.465***	0.399***
	(0.039)	-	(0.119)	(0.132)	(0.150)
Has Repaid	-0.016**	-	-	0.288***	0.230***
	(0.007)	-	-	(0.024)	(0.030)
$\hat{\lambda}$	-0.012	-	-	-	0.332***
	(0.027)	-	-	-	(0.106)
Outside	0.001	-	-	-	-0.005
	(0.002)	-	-	-	(0.008)
Age	-0.004**	-	-	-	0.010
	(0.002)	-	-	-	(0.008)
Time FE	Y	Y	Y	Y	Y
Observations	1,570	2,415	2,415	2,415	2,415
R^2	0.220	0.073	0.126	0.195	0.199

Note: *p<0.1; **p<0.05; ***p<0.01

6.4 Figures

Figure 1: : Loan Request, Fill, and Repayment. The beginning-to-end communication of borrower *trigonoah* requesting (top panel), receiving (middle panel), and repaying (bottom panel) a loan.

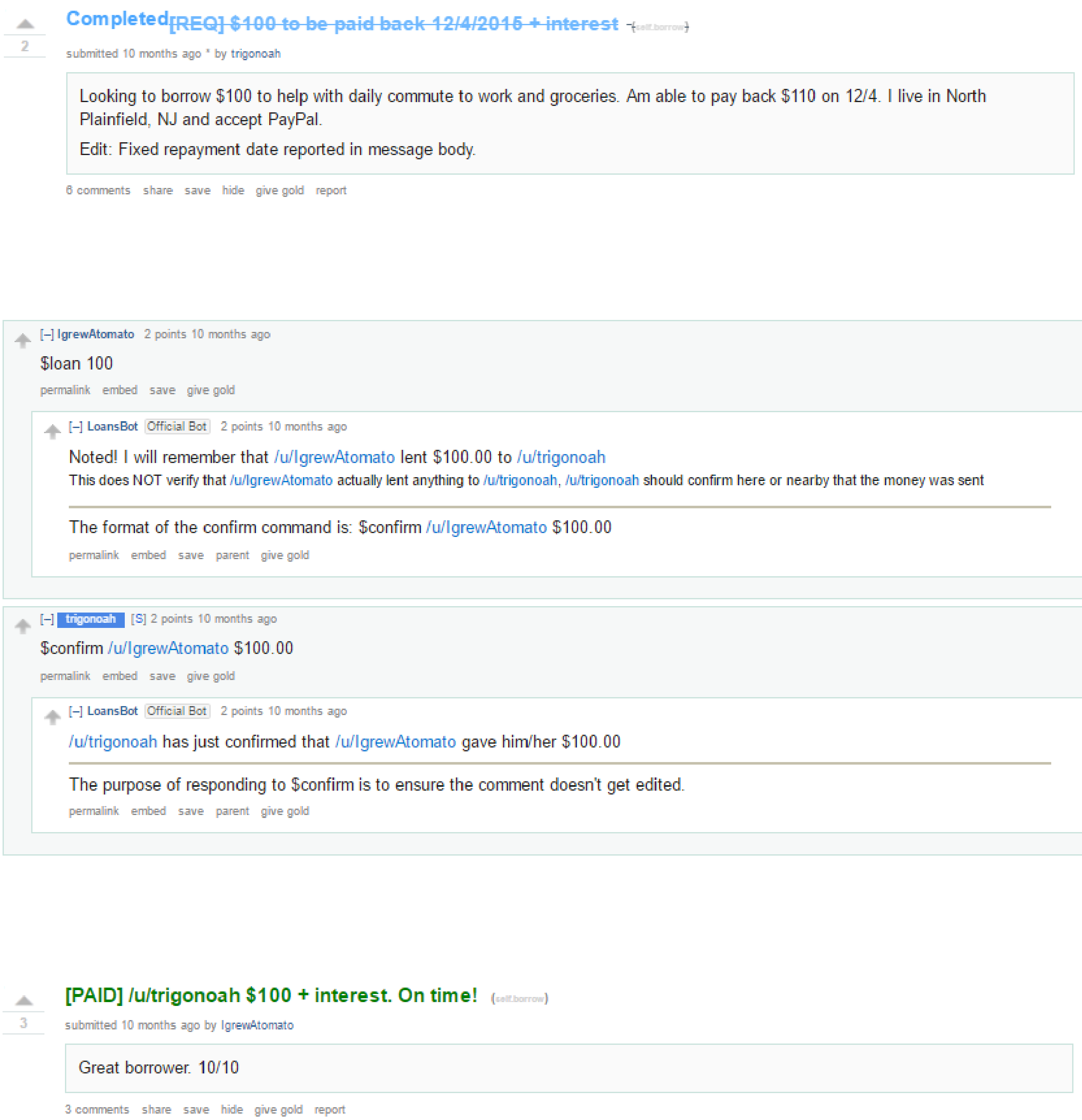


Figure 2: : Loan Request, with no fill. The beginning-to-end communication of borrower *DopaGuru* requesting (top panel) and not receiving (bottom panel) a loan.

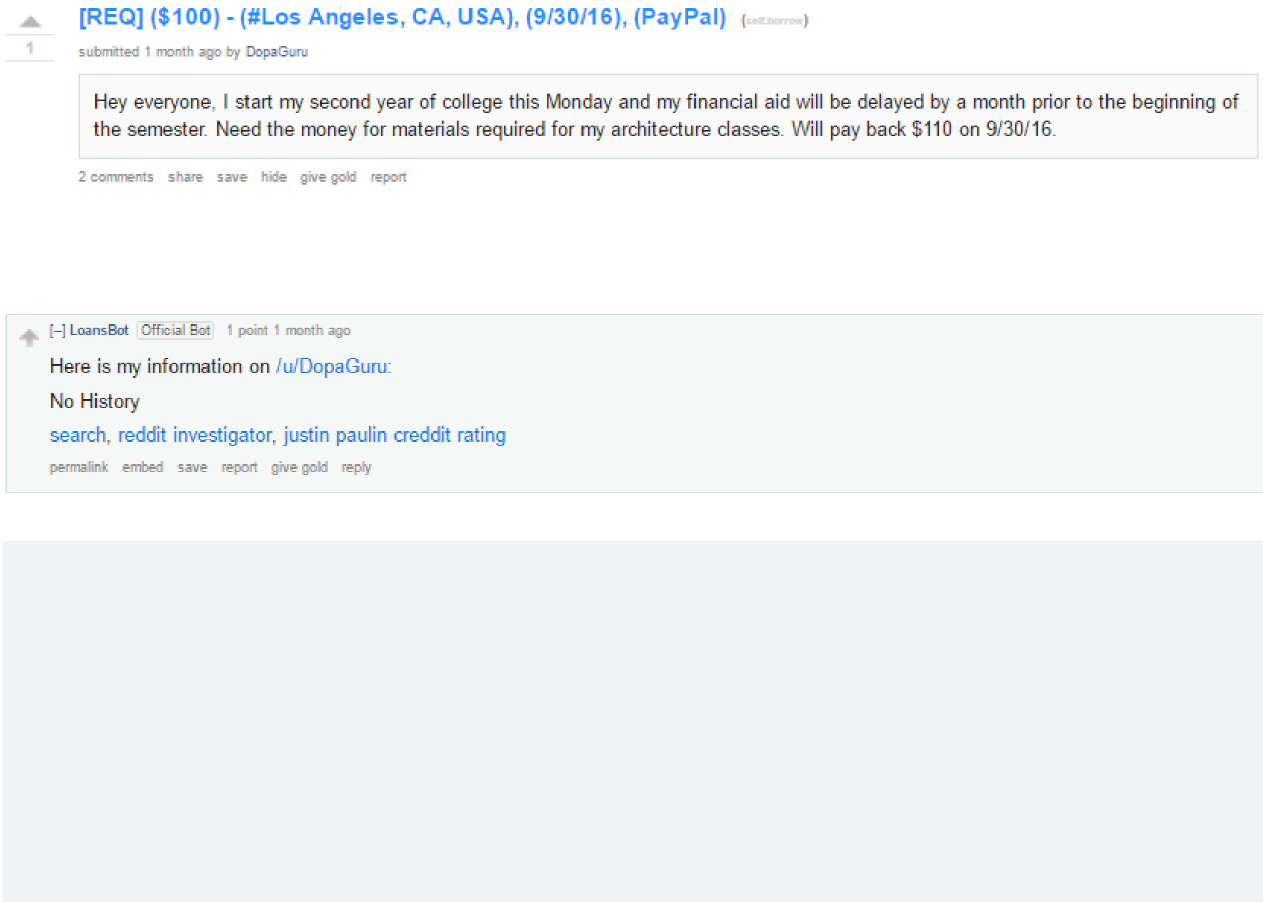


Figure 3: : Loan Request, Fill, and Default. The beginning-to-end communication of borrower *LilWoadie* requesting (top panel), receiving (middle panel), and repaying (bottom panel) a loan.

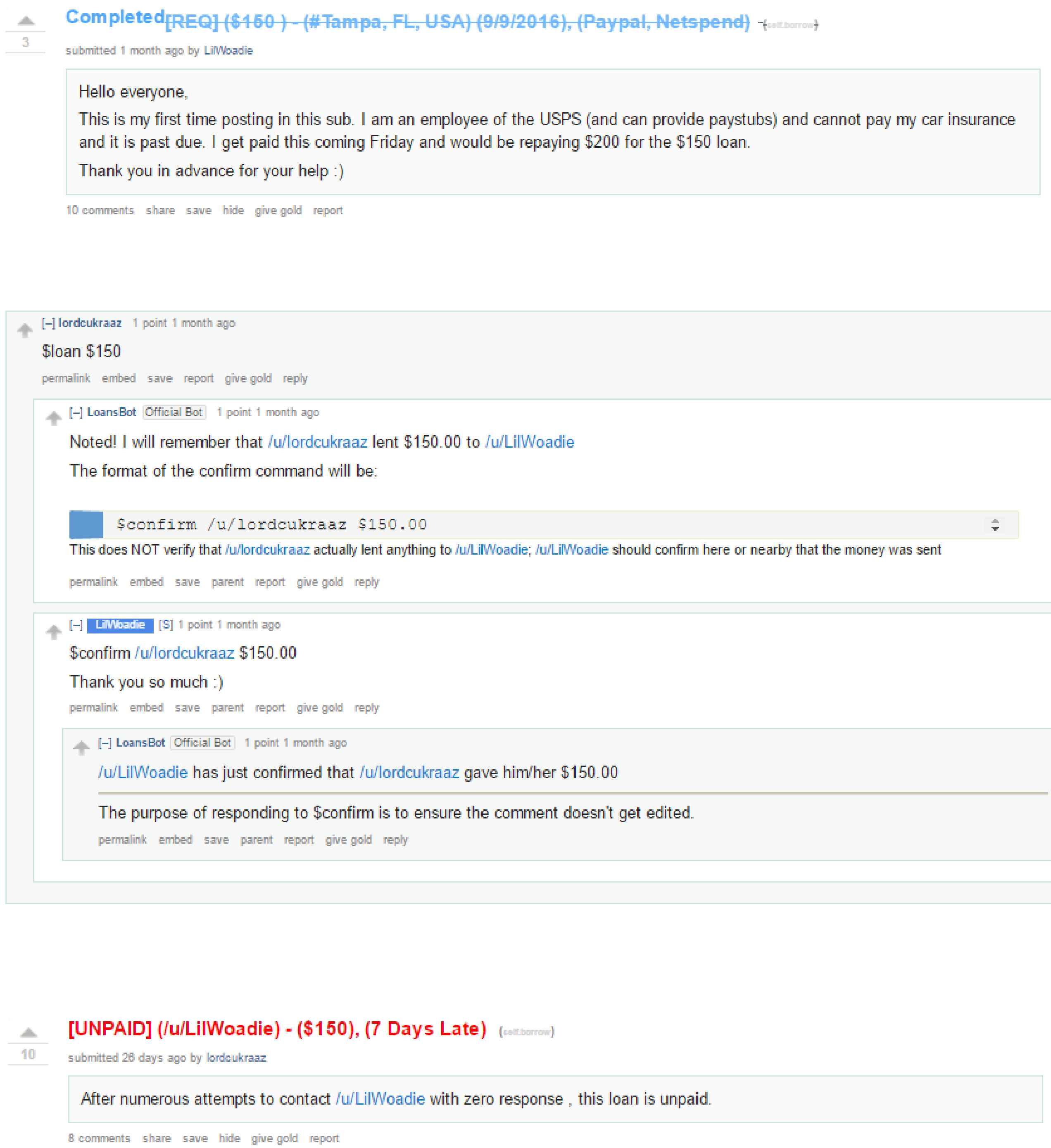
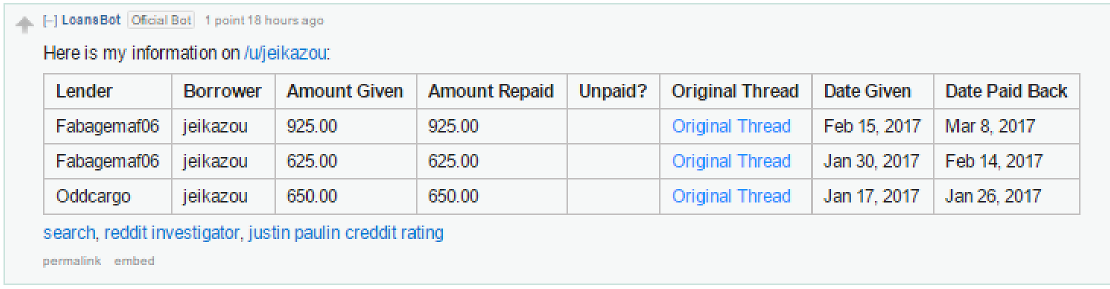


Figure 4: : The automatically generated credit report visible to all potential lenders. This report is automatically generated upon a borrower's loan request.



[–] LoansBot Official Bot 1 point 18 hours ago

Here is my information on /u/jeikazou:

Lender	Borrower	Amount Given	Amount Repaid	Unpaid?	Original Thread	Date Given	Date Paid Back
Fabagemaf06	jeikazou	925.00	925.00		Original Thread	Feb 15, 2017	Mar 8, 2017
Fabagemaf06	jeikazou	625.00	625.00		Original Thread	Jan 30, 2017	Feb 14, 2017
Oddcargo	jeikazou	650.00	650.00		Original Thread	Jan 17, 2017	Jan 26, 2017

search, reddit investigator, justin paulin creddit rating

permalink embed

Figure 5: : Borrower requests, counts (top panel) and dollars (bottom panel). Sample is divided based on whether request is on r/loans or r/borrow.

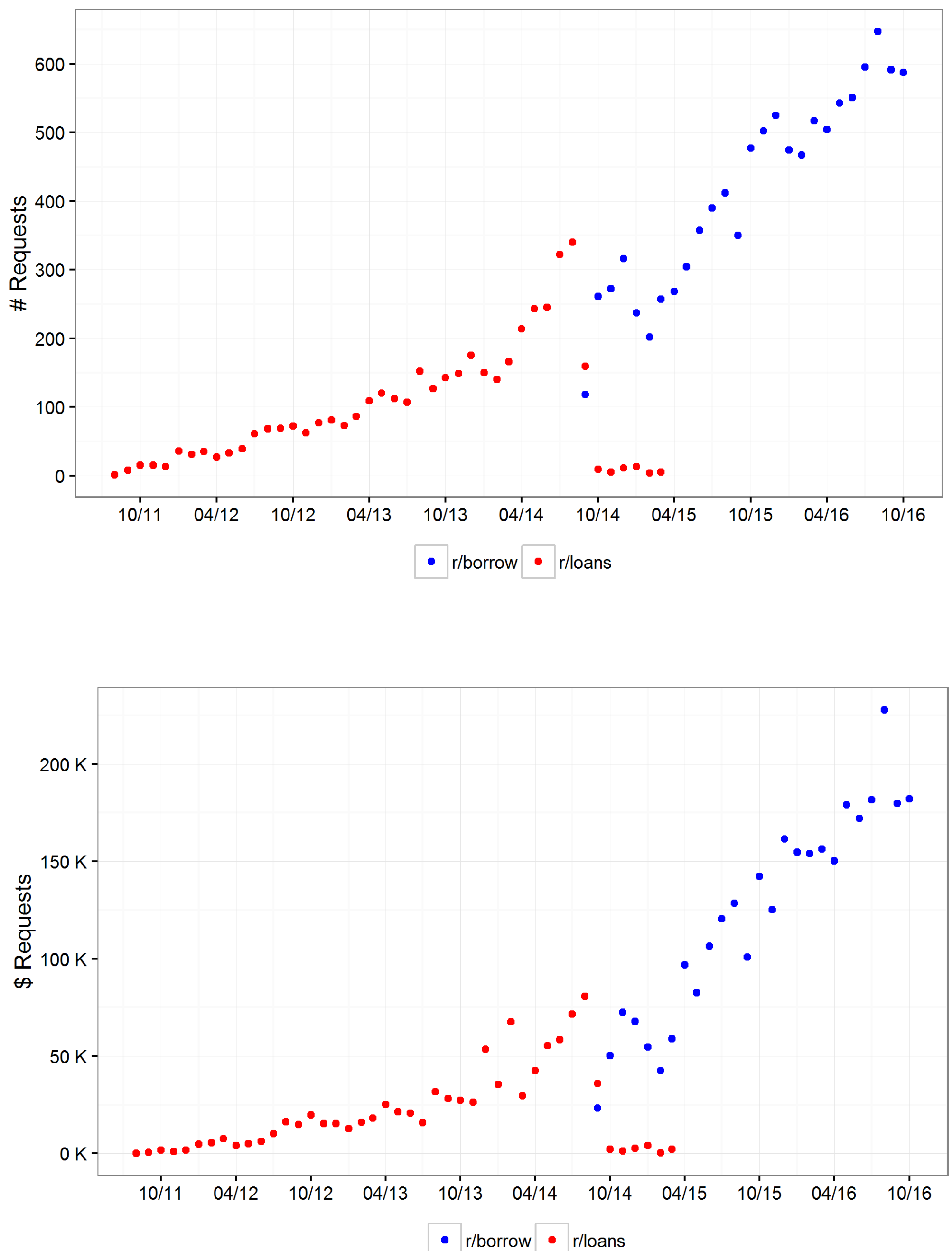


Figure 6: : Fills, counts (top panel) and dollars (bottom panel). Sample is divided based on whether fill is on r/loans or r/borrow.

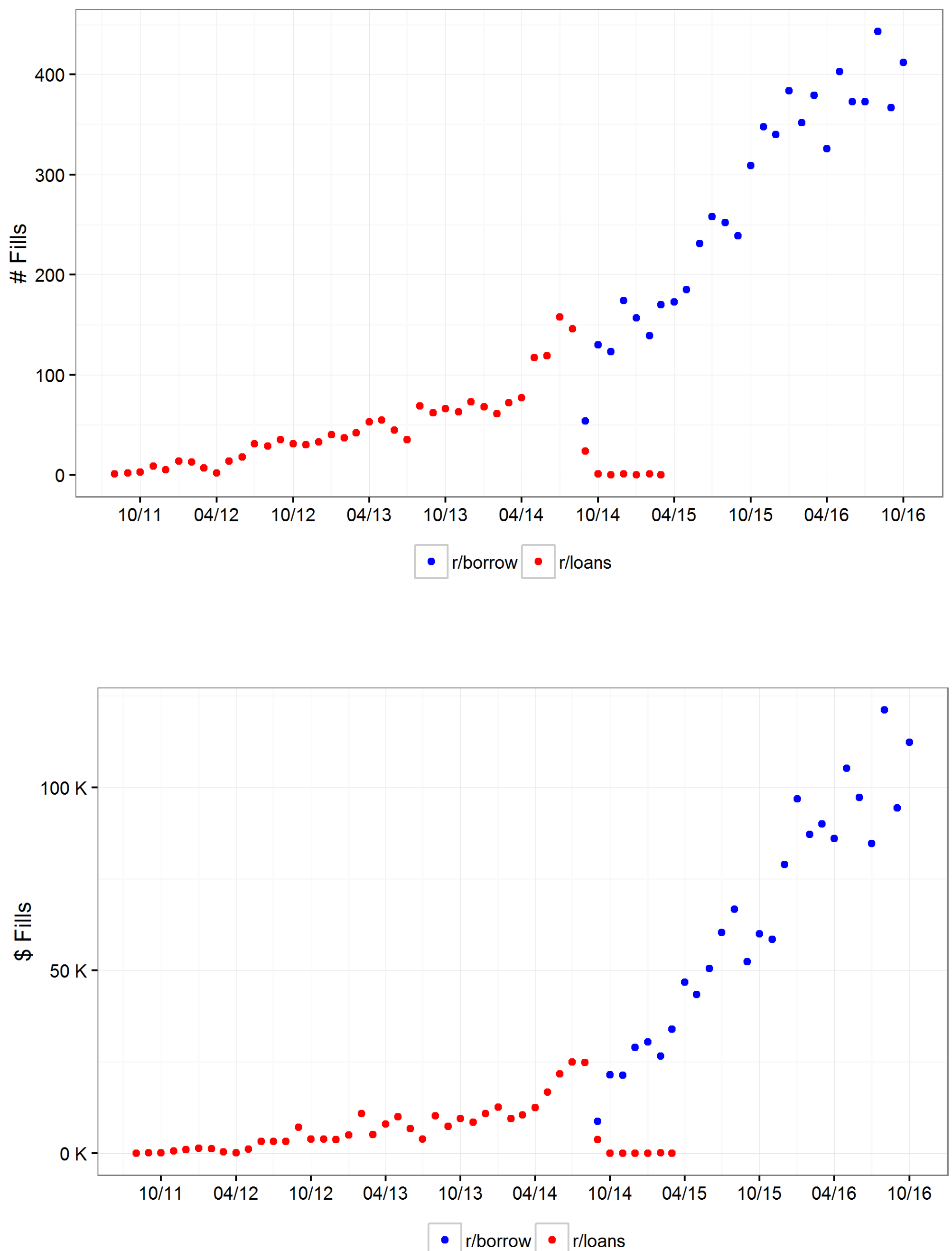


Figure 7: : Loan size per request (top panel) and loan size per fill (bottom panel). Sample is divided based on whether fill is on r/loans or r/borrow.

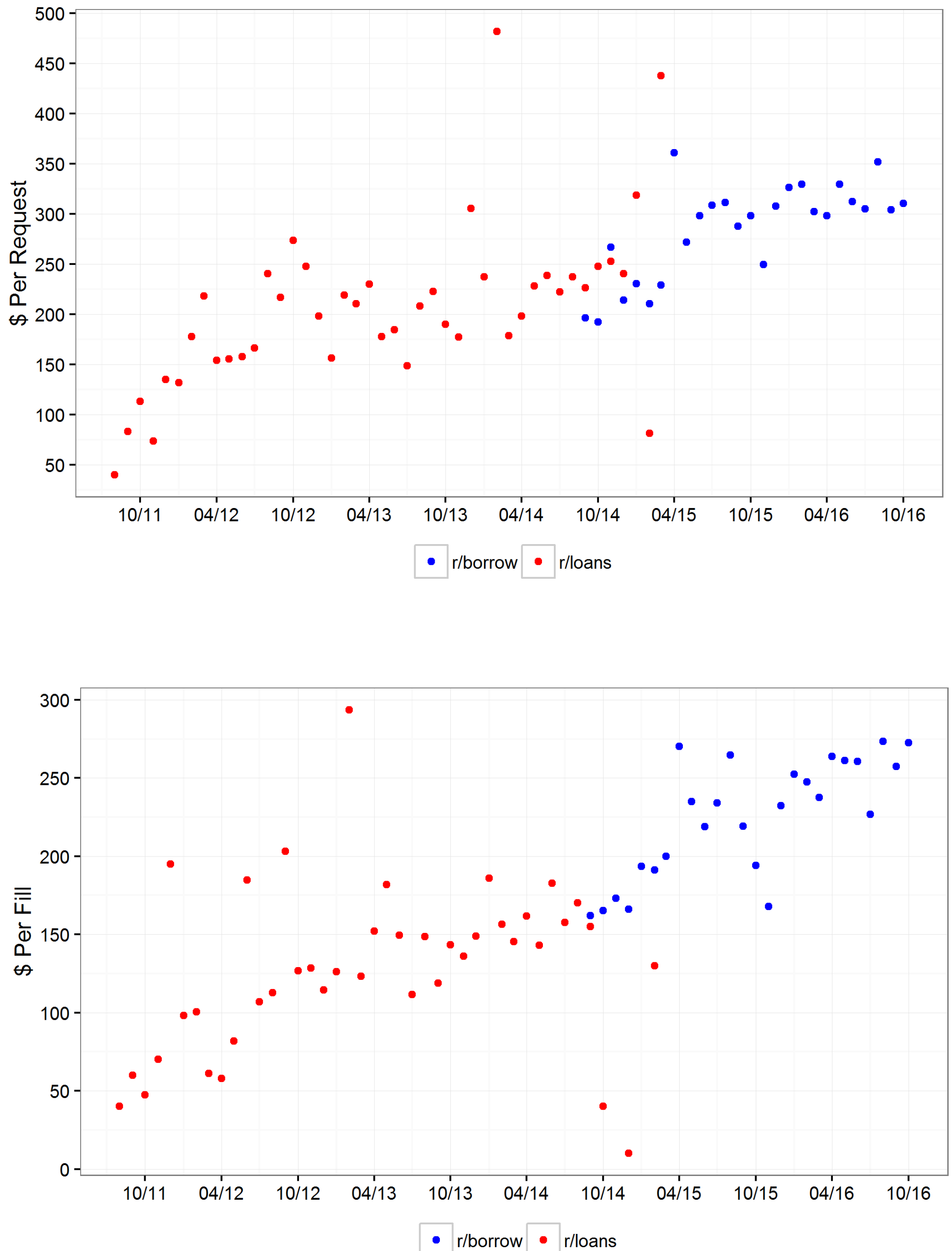
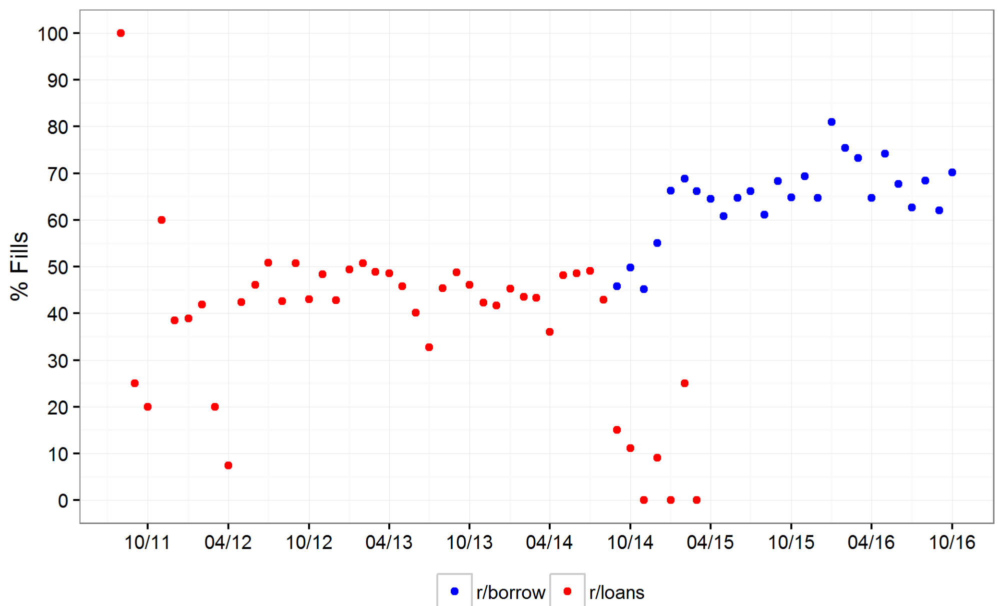


Figure 8: : Percent of requests filled (top panel) and percent of fills repaid (bottom panel). Sample is divided based on whether fill is on r/loans or r/borrow.



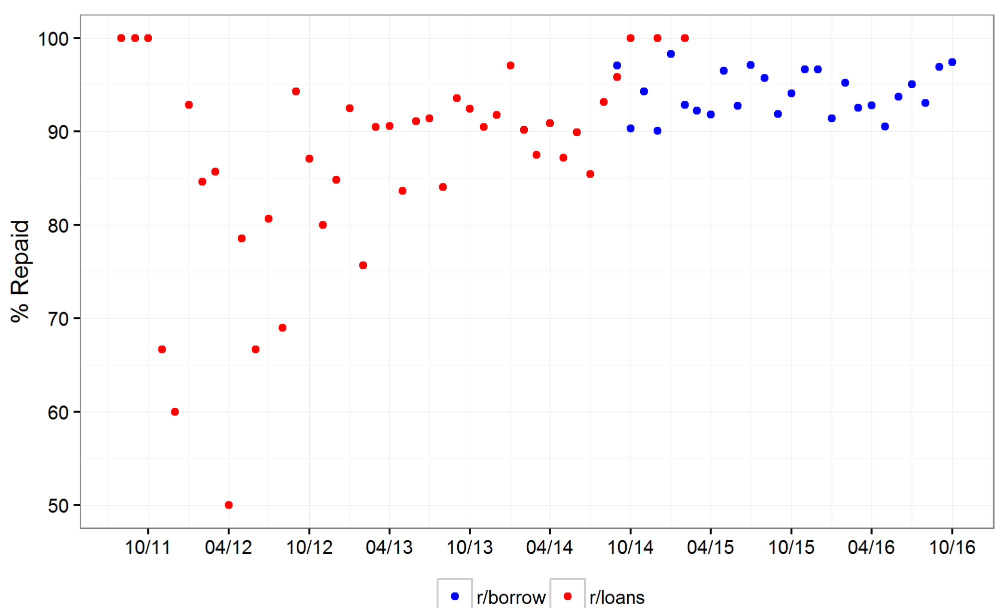
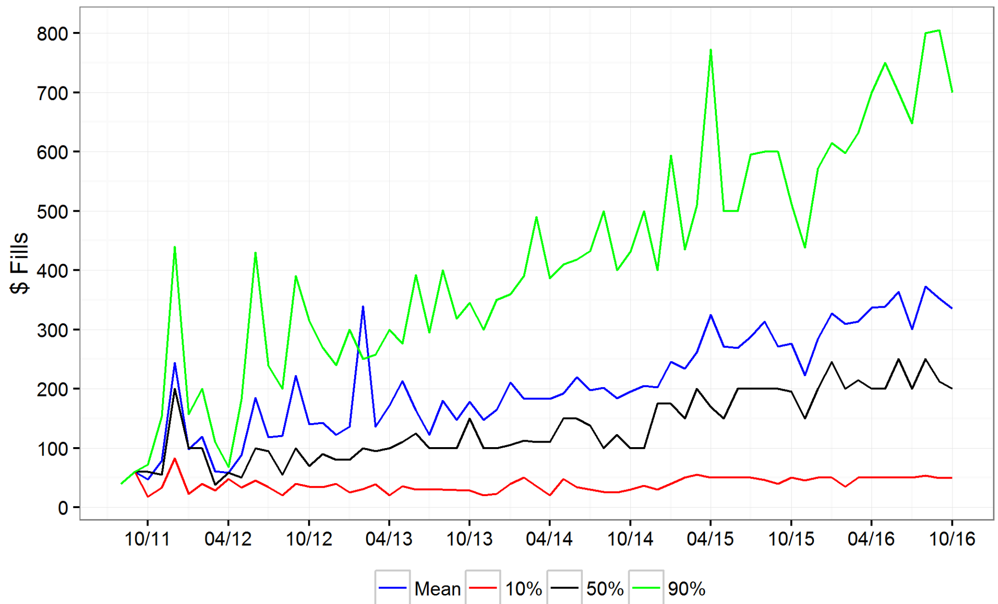


Figure 9: : Distribution of per-borrower monthly borrowing (top panel) and distribution of per-lender monthly lending (bottom panel). Blue line is mean, red is 10% quantile, black is median, green is 90% quantile.



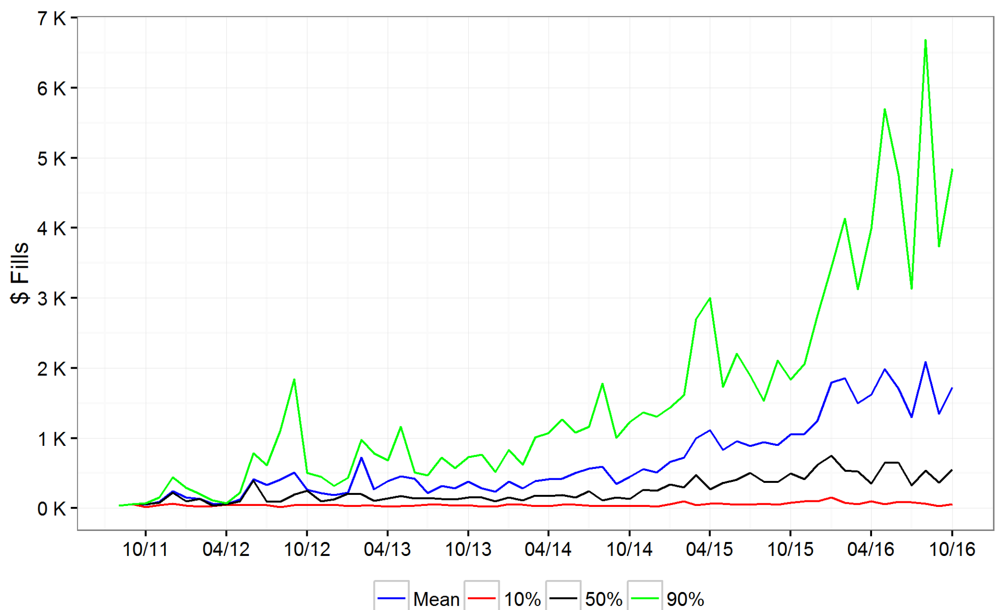


Figure 10: : Borrower history at loan request: How many previous loans have typical requesting borrowers had at the time of request? Blue line is mean, red is 10% quantile, black is median, green is 90% quantile.

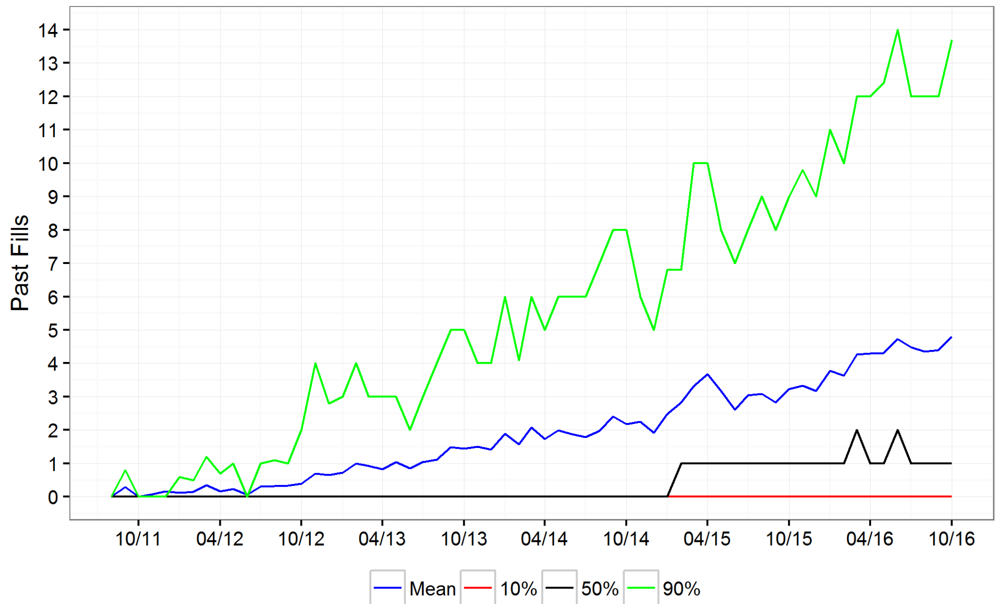


Figure 11: : Median fill times for filled loans in hours. The top panel is split by r/loans and r/borrow. The bottom panel focuses on August 2014 (around the time of the migration to r/borrow) and splits the sample by new borrowers and borrowers with a positive repayment history.