Valuing Reputation in Extralegal Contracting*

Greg Buchak
University of Chicago

October 14, 2016

Abstract

Contracting outside the formal legal system is commonplace, be it among diamond dealers, whalers, or Maghribi traders. Parties' ex-ante willingness to enter into contracts depends crucially on whether the available extralegal consequences for breach are sufficiently severe. While economic and behavioral theories flag several channels through which compliance is enforced, identifying whether these channels are actually active in facilitating transactions is difficult: The canonical examples of extralegal contracting are among small, well-connected, and homogeneous communities, which complicates the separation of purely economic motives from shared moralistic and social motives. This complication limits the inferences that can be drawn about whether extralegal contracting would work in broader or more heterogeneous communities. Drawing on a unique dataset of anonymous and unsecured online peer-to-peer lending, this paper shows that forward-looking, wealth-maximizing rational economic behavior, as opposed to intrinsic or social pressure alone, is clearly at work in facilitating contracting with no formal enforcement mechanism. The broader implication for extralegal contracting is that a small community with homogeneous moral beliefs or strong social connections is not a precondition for these markets to function. Rather, the key requirement is a reliable information transmission mechanism to document past behavior.

*DRAFT. First version, September 2014; Latest Version, October 2016.

1 Introduction

A user identified by the online handle *(borrower1)* posts the following message on an internet forum:

> *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 *(borrower1)* gets a loan and decides not to repay, his hypothetical lender is out of luck.[1] The aggrieved lender will post publicly that someone with the username *(borrower1)* did not pay him back, he will be out \$100, and that will be the end of it. Yet *(lender)*, another user, sends *(borrower1)* \$100. A few weeks later *(lender)* posts reporting that *(borrower1)* has paid him back with interest.

A different user, *(borrower2)* posts this message:

> *Hey everyone, I start my second year of college this Monday and my financial aid will be delayed by a month prior to the beginning of the semester. Need the money for materials required for my architecture classes. Will pay back \$110 on 9/30/16.*

No takers.

What makes *(borrower1)* a good debtor and *(borrower2)* not? They have the same credit history: None; these requests were each borrower's first. The loans are for the same amount and carry the same interest payment. Both have a potentially verifiable source of income—the job to which *(borrower1)* commutes, or the financial aid that *(borrower2)* will receive shortly. There is no reason to think that either one is more honest than the other. The site under consideration is a large site which focuses mostly as a news aggregator and discussion board; users' sitewide positing history is visible—in fact, the lending section of the site is informal, created by users rather than the site administrators, and comprises a vanishlingly small amount of total user activity on site. Could it be that *(borrower1)* is a more active user elsewhere on the site and more bound by social ties? While this paper will

[1] The best case for the lender 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. The lender would likely acquire in a private communication *(borrower1)*'s real name and driver's license, though even if the loans are theoretically enforceable, the small size makes formal enforcement not worthwhile.

find that that more active users tend to have an easier time securing credit, in this case, *(borrower2)* was actually a more active user than *(borrower1)*.

The answer is that *(borrower1)*'s post indicates that he has more to lose than *(borrower2)* if he defaults: *(borrower1)* needs his loan for ordinary, likely-to-be repeated expenses: gas and groceries. He has a job, but he is a little short this week. If *(borrower1)* is a little short this week, he is likely to be a little short next week or next month. *(borrower2)* needs his loan for what looks like a rare or one-time expense: school supplies before he receives a large lump-sum payment. Once *(borrower2)* gets this loan, it is unlikely that he will need another one for a long time, if at all. If the apparently needy *(borrower1)* defaults, his default will be recorded and he will not be able to borrow again even though he is likely to want to; if the apparently independent *(borrower2)* defaults, his default will be recorded, but no big loss to him: borrowing here does not seem to be very important. As *(borrower1)* appears ex-ante more likely to repay, it is easier for him to get a loan.[2]

Transactions like *(borrower1)*'s and *(lender)* are fairly typical. This market, which has been active since mid 2014, sees roughly 700 monthly loan requests and 400 monthly loans filled as of September 2016. In dollar terms, this corresponds to roughly $200,000 in requests and $100,000 in monthly fills. The mean loan is about $250 with a required repayment of $300. The repayment rate is above 90%. Without formal legal system, what forces are at work compelling *(borrower1)* and borrowers like him to repay, making *(lender)* comfortable lending in the first place? Through a quantitative analysis of this niche market, this paper teases out the moral, social, and economic forces that support this market and provides insights into how and why extralegal markets function more generally.

As a long line of examples have shown,[3] formal legal institutions are not necessary to exert social control over an agent with the opportunity to misbehave. Ellickson (2016) points out that motivations for ex-ante self-control come from three sources: (1) First-party control though the potential bad actor's own internalized norms or ethics, (2) second-party control through the potential bad actor's victim engaging in self-help, and (3) third-party control, organized variously either through (a) through diffuse social norms, (b) formal but

[2] To conclude the story of *(borrower1)* and *(borrower2)*: Needy *(borrower1)* in fact requests and repays several more loans in the future. *(borrower2)*, after trying a few more times, eventually mentions that he needs money for recurring transportation and food expenses. He subsequently receives his loan, which is still outstanding.

[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.

non-governmental hierarchical organizations,[4] or (c) governments. Legal scholarship traditionally focuses narrowly on point (3c), third-party control through governments—formal law. However, scholars have long recognized the importance and power of (3a), informal diffuse social control.

Without state power of coercion, systems of informal social control must rely on alternative mechanisms to enforce good behavior when agents will be tempted to misbehave. Case studies of diamond merchants (Bernstein (1992)), whalers, (Ellickson (1989)), wheat dealers (Bernstein (1996)), and many others flag a number of characteristics of these systems that support enforcement. Typically, the focus is on channels for informal third-party enforcement, as opposed to first- or second-party enforcement, and the question is whether third-party agents can credibly threaten to inflict sufficiently harsh punishment in order to discourage ex-ante misbehavior.[5]

Broadly, the punishment mechanism available to third-parties in these systems is *reputational*, *social*, or a combination of both. The *reputational* mechanism is the standard rational economic explanation that agents with bad reputation are excluded from future transactions with the community, and therefore agents' honesty is collateralized with a so-called "reputation bond,"[6] whose value is equal to the present value of future transactions within the community less the forward-looking value of current dishonest behavior. In essence, an agent can commit to behave well so long as the present value of future interaction is greater than the one-time value of cheating. This is critically dependent on expected future interaction with members of the community who can observe past misbehavior.[7] The social mechanism[8] represents the direct loss of social ties, forgoing of social interactions with the community, or other harm inflicted through social stigma, which constitutes a direct, non-economic harm on the misbehaving agent. For instance, a misbehaving agent may lose his friends or the community at large may shun him.[9] Aside from these diffuse third-party enforcement mech-

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

[5]This paper will focus on the potential severity of punishment rather than whether it is credible. 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. The method of punishment examined here will closely resemble the reputational channel in Greif (1993).

[6]See Bernstein (1992)

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

[8]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.

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

anisms, there is also a first-party mechanism in that agents simply like to, have a moral compulsion to, or are habituated to acting pro-socially.

Separating and quantifying these various forms of social control is a subtle theoretical and empirical problem, even when outcomes are clear and quantification is possible. The essential problem is that in a given setting, these distinct informal modes of social control would are almost always all potentially in play Moreover, a theoretical model featuring one or some of these forces will generate almost identical empirical data within a particular market in terms of measurable behavior. Consider three models of informal social control, each relying on different mechanisms. In the *homo moralis*[10] model, some agents behave because they have an moralistic motivation to behave. In the *homo socalis* model, agents behave not for moralistic or economic reasons but because direct social censure for misbehavior is painful.[11] In the *homo economicus* model, agents have no moral compulsions but behave because their discounted utility from behaving is greater than their discounted utility from misbehaving, as mediated through the loss of their reputational capital. These models are difficult to distinguish through both qualitative and quantitative study: Each of these models is an effective means to support an informal market;[12] each model predicts that individuals with a longer history of successful transactions are more likely to cooperate;[13] each model predicts that individuals with a longer history of successful transactions are more likely to find willing business partners;[14] each model is likely to flag institutions that facilitate information transfer between agents as critically important.[15][16]

The fact that these models tend to generate similar predictions and normative conclusions *within* markets will generally limit what market-by-market qualitative case studies can

[10]Alger and Weibull (2013) call this *always does the right thing* agent *homo kantiensis*.

[11]See Ambrus et al. (2014) for such a model.

[12]*Homo moralis*—if everyone has strong moral compulsions to not cheat, they will not cheat. *Homo socalis*—if social stigmas or loss of friends for cheating are especially harsh, to avoid these punishments, they will not cheat. *Homo economicus*—if the value of the revenue bond is so high, it is not worthwhile to sacrifice it for a one-time gain.

[13]*Homo moralis*—a longer repayment history signifies stricter morals. *Homo socalis*—a longer repayment history signifies greater sensitivity to social sigma. *Homo economicus*—a longer repayment history signifies a more valuable reputation bond.

[14]See above.

[15]*Homo moralis*—communication helps agents learn who is moral from past conduct. *Homo socalis*—communication helps agents learn who is sensitive to social stigma from past conduct and enforce social stigma. *Homo economicus*—communication is a means to enforce exclusion after default and to signal the value of one's reputation bond.

[16]Even with a lot of data, identifying dynamically optimal behavior from myopic behavior poses a significant technical challenge and is often theoretically impossible. See Rust (1994) and Magnac and Thesmar (2002).

predict if the institutional details were to change,[17] or whether informal markets would be viable in entirely new settings. Do the diamond dealers, who tended to come from an ethnically and religiously homogeneous community all share similar moral values that enable their extra-legal contracting to work; does the fact that their community is closed make social punishments particularly damaging; or does a more general and less community-specific mechanism like the reputation bond keep the market functioning? Do the Nantucket whalers, who also come from an ethnically and religiously homogeneous community similarly share common moral compulsions; does the fact that they live next to one another on a small island make social pressures particularly strong, or is the economic consequences of reputation loss doing the work? It is difficult to tell, and consequently hard to predict how well a market lacking one or some of these features will function in their absence. Critically, as more and more commerce takes place over a decentralized Internet, it is not clear whether informal markets that were once dominated by face-to-face social interaction will continue to function. If the underlying mechanisms that support these existing markets are moralistic or social, and these behaviors are tied heavily to small, closed, or homogeneous communities, expansion and growth could spell doom for these markets. On the other hand, if these markets are supported by the rationalist *homo economicus* behavior, technologies that aid in broadening and spreading information about these markets will make them work better.

This paper's unique empirical setting allows for a partial, though not complete, disentangling of these forces, and hence provide some quantitative lessons that can be applied to other settings. First, the setting provides quantitative outcome variables with direct economic consequences. Success can be measured as whether a loan is made, the size or interest rate of the loan, and whether the loan is repaid, all in dollar terms. Contrast this with previous case studies of extralegal contracting, or recent studies of online social networks, e.g., Twitter (e.g., Bakshy et al. (2011), Bakshy et al. (2012), or Rost et al. (2016)) whose outcome variables are either not measured or in the case of the recent online studies, whose outcome variables, like re-tweets or twitter followers, have a less-direct economic interpretation. More importantly, the institutional setting rules out second-party self-help and third-party formal governmental or organizational enforcement. This leaves, like many existing studies of extralegal contracting, competing *homo moralis*, *homo economicus*, and *homo socalis* explanations. Unlike previous studies, however, the particular anonymous setting drastically reduces the possible force of direct social stigma, and while the paper cannot *rule out* a *homo moralis* mechanism, can confidently *rule in* a *homo economicus* mechanism by exploiting ex-ante variation in the value of borrowers' reputation bonds and showing that

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

those potential borrowers who appear ex-ante to suffer more from market exclusion are more likely to receive loans and by implication, more likely to repay.

In particular, this paper obtains ex-ante variation in the value of borrower's reputation bonds and shows that borrowers with more valuable reputation bonds are more likely to receive loans. The inference is that they are more likely to receive loans because they are more likely to choose to pay them back. Variation in reputation bond value should only matter if there is an element in *homo economicus* behavior present, and so on the basis of this evidence, the paper concludes that *homo economicus* behavior is indeed present. The source of exogenous variation used is novel: As described earlier, borrowers typically put the reason for their borrowing in their initial loan request, and moreover, this information is typically verifiable by the lender.[18] Using textual machine learning techniques, this paper forms a prediction on the basis of the borrower's request that he will make a subsequent request in the future, which is an estimate intended to approximate what a human lender would believe. Borrowers who are more likely to request loans in the future have more valuable reputation bonds because they derive more expected value from their ability to borrow on the site. Using this measure to obtain variation in the value of borrowers' reputation bonds, this paper shows that it is a statistically significant determinant of whether a borrower obtains a loan, which confirms that a *homo economicus* mechanism is active.

In addition to ruling in the reputation bond mechanism as having quantitative importance, this market also offers a case study of how new entry into community based on extra-legal enforcement works. Whatever the mechanism—*homo moralis*, *homo economicus*, or *homo socalis*—that incentizes repayment, the important criterion for member ship in an extralegal contracting community is that the potential member be sensitive to whatever mechanism enforces compliance. A borrower may be insensitive to punishment for two reasons: First, he may simply not be bothered by any of the enforcement mechanisms above. Indeed, by showing that that borrowers with a high likelihood of returning are more likely to receive loans, this confirms that borrowers with a low likelihood of returning are more insensitive to market exclusion. Second, if market exclusion is the punishment, the market needs to *actually be able to exclude defaulting borrowers from the market.* This sounds obvious but in this context it presents a particularly difficult problem.[19] In particular, exclusion here happens on the *account* rather than the *actual borrower* level: The borrower's observ-

[18]For bills, the site notes that best practices are for the borrower to show the lender a copy of the bill.

[19]See Posner (2009) for a more traditional example of inability to exact punishment against particular individuals. He has in mind non-punishable "high-status" agents. For example, a local monopolist cannot be excluded because those who would punish the monopolist have no other good options with whom to transact.

able accounts to which the social record keeping applies cannot be persistently linked to the decision-making individual behind the account. Like a traveling merchant who can leave his past behind, the human borrower can leave her defaulting account behind and create a new account for free. If scammers—even scammers who genuinely benefit from receiving loans on the site—with no intention of repaying can easily create credible accounts, a small number of active scammers can destroy the market by creating accounts, obtaining funds, discarding the old accounts upon default, repeating the process, and eventually overwhelming the honest borrowers.

It is the borrower's *outside social behavior* that addresses the joint problems of asymmetric information and non-enforceability of sanctions. The paper shows that borrowers cannot get their first loan if they have an insufficiently low engagement in outside social behavior, but that once they have gained access to the community, their outside social behavior ceases to matter and enforcement takes place through the repeated interactions dynamic discussed above. Building a history of outside social participation takes time and effort, and this paper argues that the level of engagement necessary to get a loan is not worthwhile unless the borrower plans to access the lending market repeatedly. Because engagement is only justified for borrowers who need to borrow repeatedly, a history of social activity show that market exclusion is a punishing threat to the borrower, thus addressing the the asymmetric information problem. Next, the cost of the necessary level of engagement is high enough that it does not justify the benefit from receiving and defaulting on a single loan. This serves to make the strategy of repeated account creation and default unprofitable. The paper shows quantitatively that outside social activity is critically important in receiving the *first* loan but not in receiving subsequent loans. This suggests that outside social engagement plays a signaling, rather than a collateralizing role as it does inAmbrus et al. (2014).

Finally, while the paper focuses on the role of social capital and norms in extralegal contracting generally, using this particular setting as a source of data, the paper also offers insights into more traditional finance questions. The traditional financial sector, among other things, connects borrowers and lenders. While the internet can greatly aid in matching borrower and lender, it is unclear how the internet can fulfill other financial roles—screening and enforcement in particular. To date, most "mainstream" online lending sites such as Prosper.com or LendingClub lean on existing real-world institutions to screen and enforce. They collect and verify real-world information about borrowers, and the loans they facilitate come with the usual legal protections.[20] The online interaction in these platforms appears to improve outcomes: Iyer et al. (2015), for example, show that online lenders are particularly

[20] `https://www.prosper.com/policies/borrowers-bankruptcy.aspx`

effective in using "soft" borrower characteristics that traditional lenders have more trouble evaluating. Freedman and Jin (2011), find that online lenders learn from past mistakes quickly. While the online element improves information gathering, it is entirely tethered to real-world institutions. Therefore, it is not clear whether the online interaction can help replace, or merely augment, financial functions beyond matching. This paper's setting is devoid of these existing legal institutions, and shows that indeed, informal reputational and social mechanisms can fill gaps left by a lack of traditional formal institutions, so long as the proper institutional features are present.

Surprisingly, many features of these these tiny, anonymous loans closely resemble sovereign debt.[21] With most private loans, borrowers can either post collateral or commit to punishing legal sanctions. Sovereign borrowers, on the other hand, do not post collateral nor is there any direct punishment available to the aggrieved lender. Similarly, the anonymous nature of this setting means borrowers cannot post real collateral, nor can lenders threaten any formal legal sanctions against defaulters. A large literature on sovereign default, e.g., Eaton and Gersovitz (1981), recognizes that debtor countries usually engage in repeated borrowing, and so the threat of exclusion from borrowing in the future incentivizes repayment today.[22] An open question in this literature is whether market exclusion alone is enough, or that other mechanisms, like direct social costs[23] are sufficient.[24] This paper's setting rules out direct social sanctions, meaning compliance must come either from first-person self-control or through reputation costs, and I am able to show that the latter forward-looking economic considerations are quantitatively important even for the non-expert agents transacting in this market. An implication for sovereign borrowing from this paper is that countries whose decision markers have either greater investment in the value of future borrowing or who directly have longer tenures in office should have greater borrowing capacities, because the value of their reputation bonds is greater.

The paper proceeds as follows. Section 2 lays out the institutional details and presents summary data and some stylized facts. Section 3 provides reduced form evidence of what supports lending and repayment, with a particular focus on ruling in a *homo economicus* revenue-bond mechanism as at least partially facilitating repayment. Section 4 discusses the implications of the findings and concludes.

[21] For a broad overview of this subject, see Aguiar and Amador (2013).

[22] First, give some more citations and talk about this more. Second, an interesting thing to think about: these guys assume exclusion but can that actually survive in equilibrium?

[23] E.g., reputation costs in other relationships, Cole and Kehoe (1998).

[24] Borensztein and Panizza (2009) examines the economic costs of sovereign default.

2 Market Overview

This section provides an overview of the market under study. Section 2.1 details the market structure and data collection; Section 2.2 provides high-level descriptive statistics.

2.1 Institutional Details and Data Collection

A robot collects 10,000 loans from the lending site and a matching engine joins requests to fills and outcome data. The time period is from Mid 2014 through September 2016.

2.2 Descriptive Statistics

This section provides some basic facts to give a flavor for the market environment. Figure 1 gives aggregate time series market statistics since March 2015. As of September 2016, there have been roughly 700 requests and 400 filled loans per month. In dollar terms, this is roughly \$200 thousand worth of monthly requests and \$100 thousand of fills. This corresponds to a mean filled loan size of roughly \$250. The market has doubled in size since March 2015, though most of the growth in fills happened in 2015, and the number and size of fills has been roughly constant in 2016.

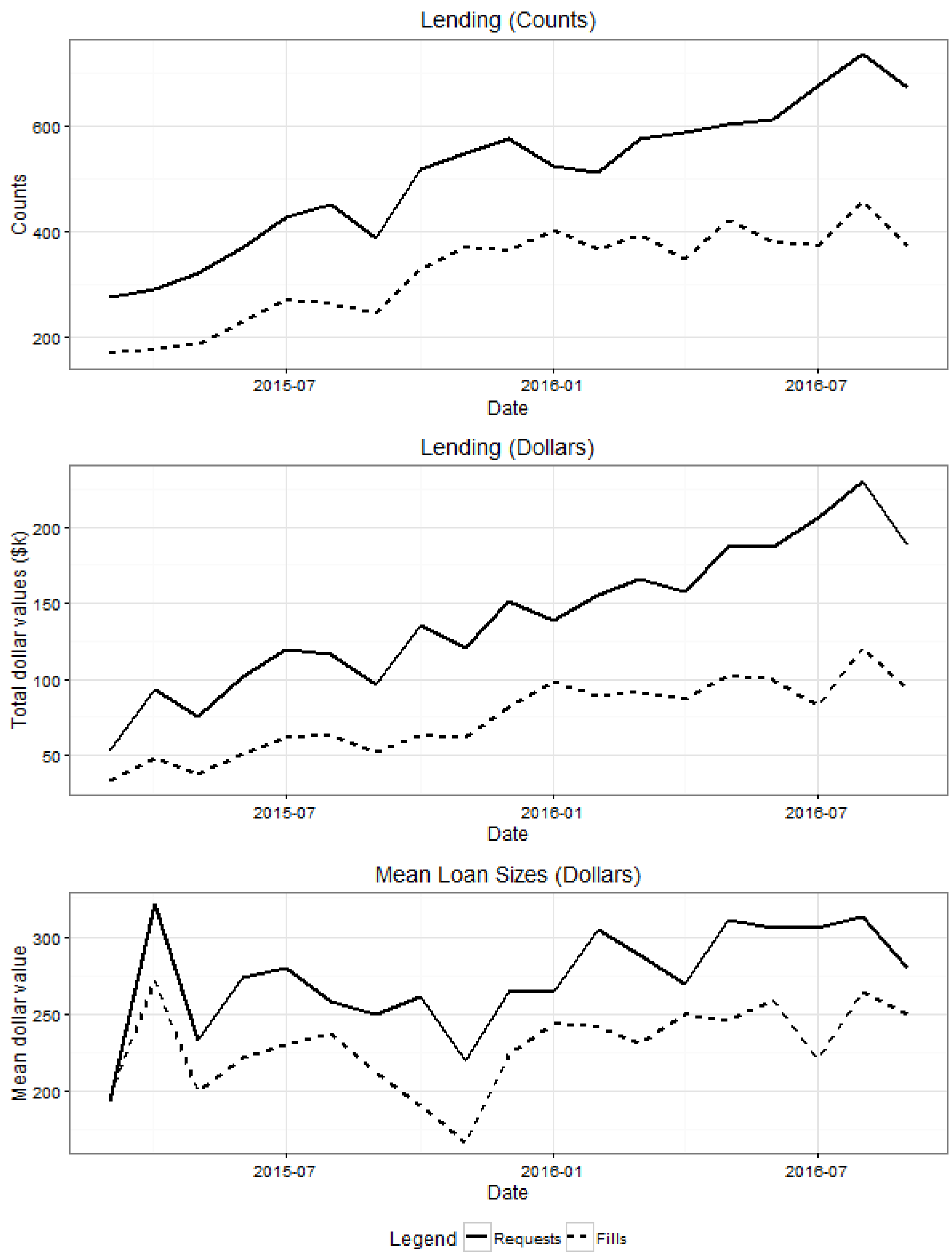

Figure 1: Monthly time series trends of lending and loan sizes from March 2015 to September 2016. The solid line is all requested loans; the dashed line is all filled loans. The top panel shows the counts of requests and fills; the middle panel shows the dollar values (in thousands of dollars) of requested and filled loans; the bottom panel shows the mean loan size of requested and filled loans.

Interest rates on loans, defined here simply as

$$R \equiv \frac{\text{Promised Repayment}}{\text{Lent Amount}} - 1$$

i.e., without any time adjustment[25] is fairly standardized and stable over time. Figure 2 gives the time series of interest rates between March 2015 and September 2016. Note that the 10% quantile interest tends to be exactly 10% and that the median interest rate tends to be exactly 20%. There appears to be a convention to offer repayment amounts in these round percentiles—the median \$100 loan will offer \$120 as repayment; the median \$200 loan will offer \$240 as repayment. This suggests that interest rate is not doing much work to clear markets; rather credit is rationed on the extensive margin. Why this is so will be clarified in a simple model to be presented in Section **??**.[26]

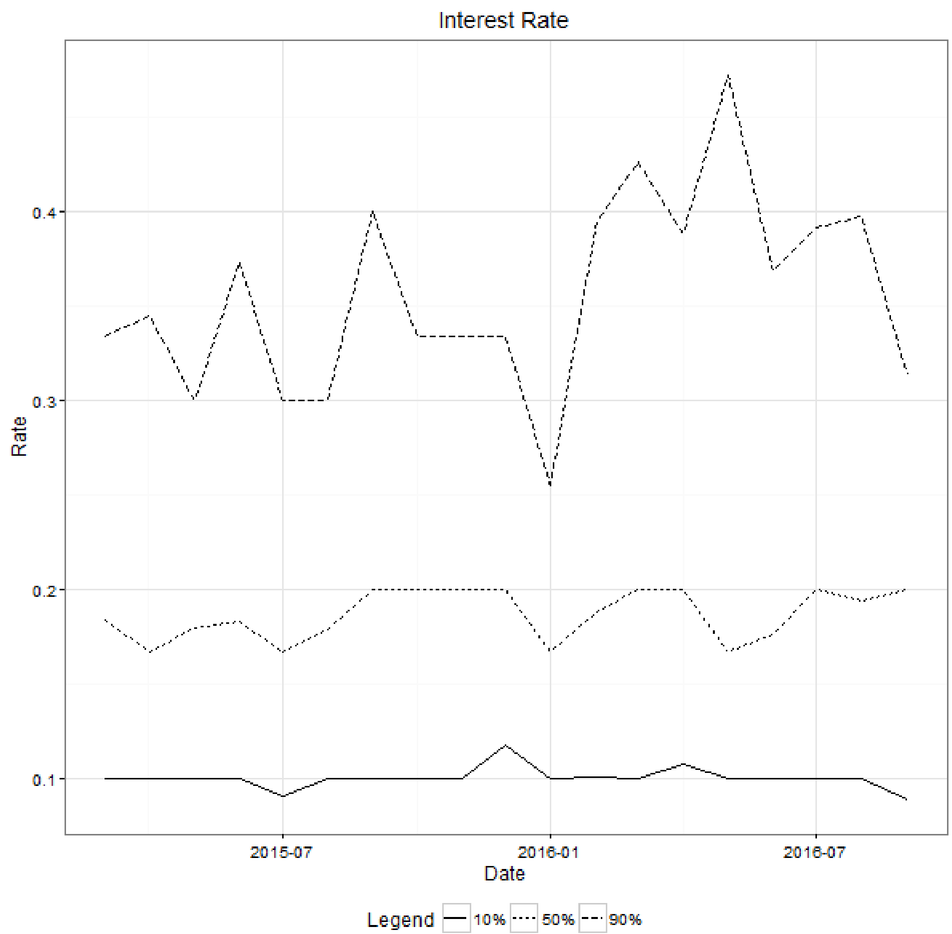

Figure 2: Interest rates of filled loans over time. The solid line is the 10% quantile interest rate; the light dashed-line is the median interest rate; the heavy dashed-line is the 90% quantile interest rate.

[25]In essence, the risk of default dominates so greatly any time variation in loans as to render the duration of the loan nearly irrelevant. From a data perspective, I also often lack the proposed loan term.

[26]In brief—in this moral hazard setup, a borrower cannot get himself a loan simply by offering a higher interest rate, because borrowers who do not plan to repay are *exactly* the kind of borrowers who would like to make this offer. A borrower who will not repay does not care about the interest he promises because he will not pay it. A borrower who will repay does. If a lender sees a borrower offering a very high interest rate, he should think "he is offering that high interest rate because he does not intend to pay it."

Borrowers consistently repay their loans. Figure 3 shows loans, dollars, and percentage of loans repaid since March 2015. The dollar value of repaid loans is the principal amount, not the total amount repaid, so to convert these numbers to dollars received by lenders, a 1.20 gross-up factor is approximately correct. The repayment rate as measured by how many loans are repaid versus how many loans have any reported outcome is roughly 92.5%. Taking these numbers and the interest rate numbers at face value, the back-of-the-envelope implication is that on a \$100 loan, a lender expects to make roughly \$11 figuring a \$20 premium and a repayment rate of 92.5%. An conservative estimate comparing the amount recorded as lent between June, July, and August 2016, equal to \$302,030 to the amount recorded as paid back grossed up by the typical interest rate of 20% between July, August, and September 2016, equal to \$313,283, yields an expected return of roughly \$3.70 per \$100 lent.

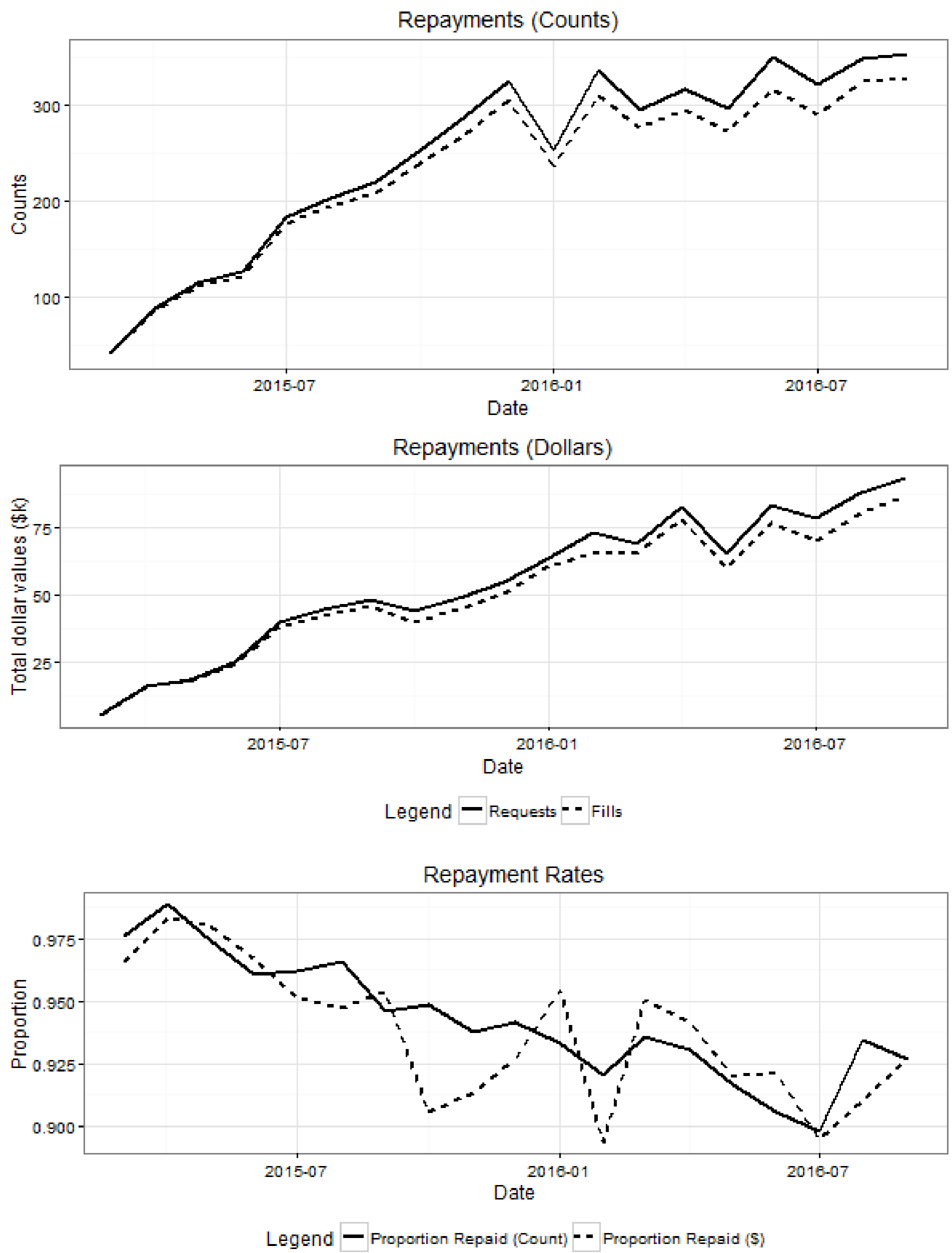

Figure 3: Trends in repayment over time. The solid line is all loans for which there is an outcome (repayment or default) reported; the dashed line is all repayments; the gap represents defaults. The top panel shows the number of loans with outcomes and repayments; the middle panel shows dollar values of these loans (principal, not face value); the bottom panel shows the percentage of outcomes that tare repayments, the solid line being in terms of numbers of loans and the dashed line being in terms of dollars.

The differences in these $11 and $3.70 estimates are explained by noise in the data collection process,[27] Due to this noise, this paper will not push or investigate time-series

[27] See Section 2.1 for details, but in particular, posts flagging a loan as made and post flagging a loan as repaid are separate and sometimes impossible to match. This leaves some made loans without corresponding

differences in repayment rates as it is impossible to determine whether the apparent decrease in repayment rates shown in the bottom panel of Figure 3 is an actual shift or the result of better reporting of defaults in the later period.

The core of this market are the borrowers and lenders. Figure 4 shows the time series of unique borrowers and lenders per month. First, these trends show a constantly growing number of borrowers seeking loans, although the number of borrowers actually receiving them has been roughly constant since the beginning of 2016. Second, the number of active lenders has been essentially constant since mid-2015. Third, the number of borrowers receiving loans is far greater than the number of lenders making them, which means that the structure of the market is that a few lenders make many loans.

"repaid" or "defaulted" outcomes. Assuming all of these unmatched loans are defaulted yields the conservative \$3.70 estimate; assuming that the unreported loans come in the same proportion of repayments to defaults as the reported loans yields the \$11 estimate.

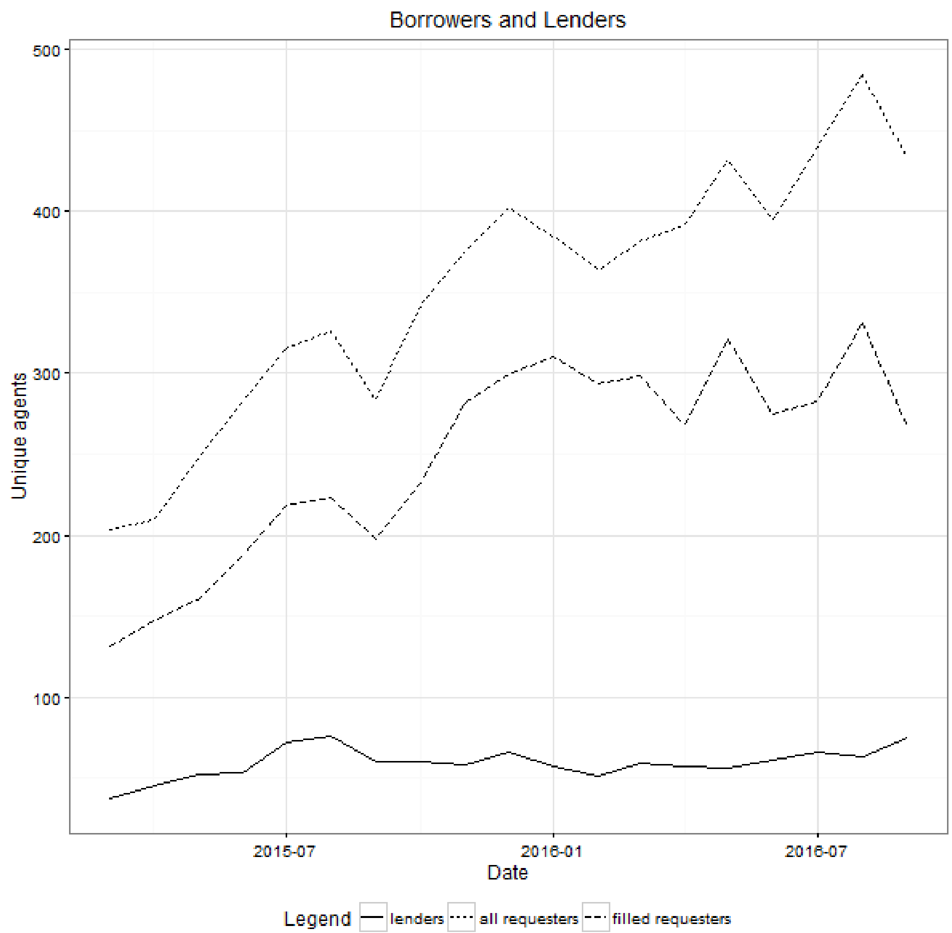

Figure 4: Unique borrowers and lenders per month. The solid line is the number of unique lenders making active loans per month; the light dashed line is the number of unique borrowers *requesting* loans per month; the heavy dashed line is the number of unique borrowers *receiving* loans per month.

Figure 5 breaks out the number of loans and dollar value of loans that typical lenders make per month. The figure shows wide disparity in lending activity on the site. The bottom 10% quantile of lender makes one loan per month with principal between \$30 and \$100. The median lender makes between 1 and 3 loans per month with total principal around \$300. The largest lenders, at the 90% quantile, are very active relative to the smaller lenders, making around 10 loans per month with a total lending amount consistently greater than \$1,000 per month.

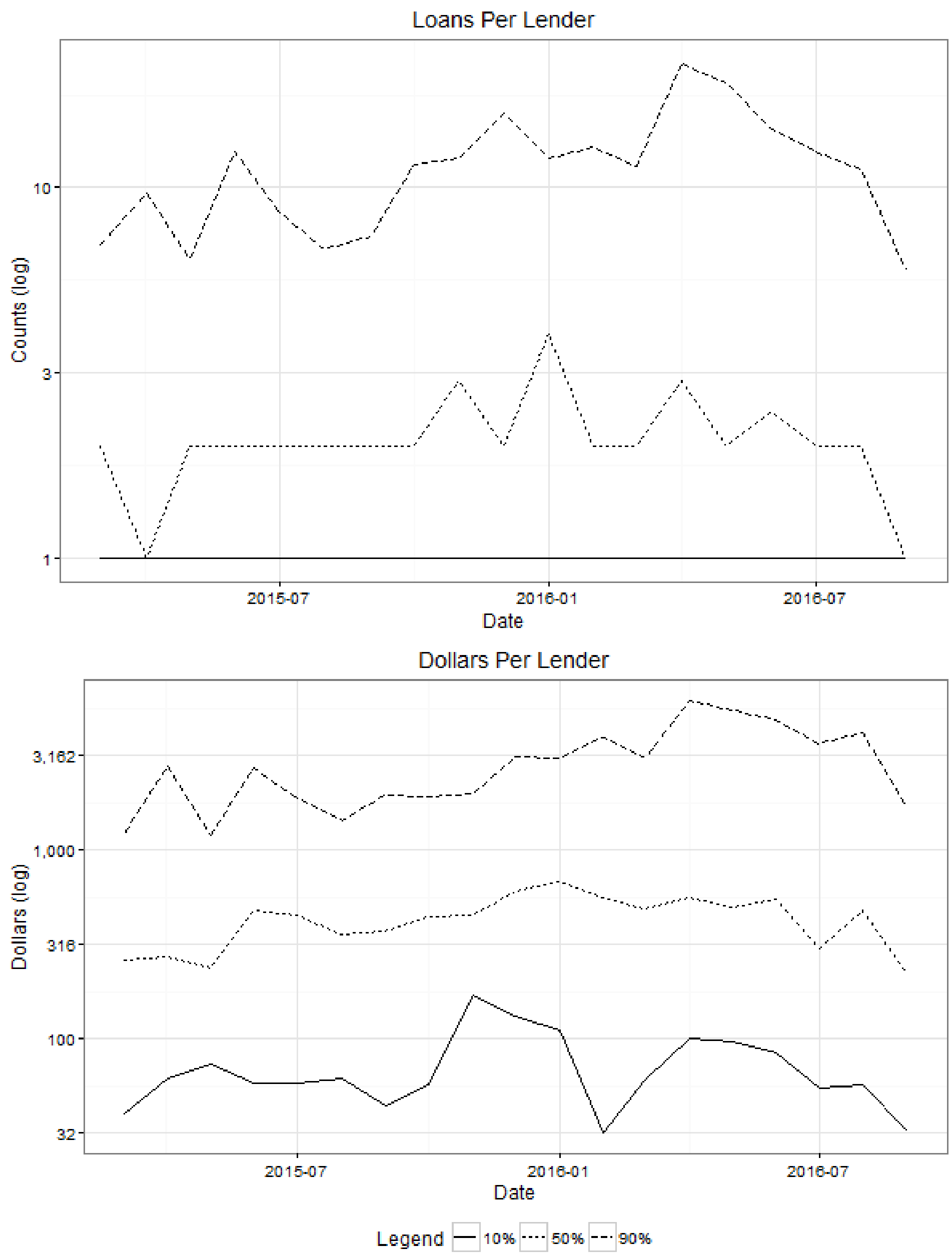

Figure 5: Lender activity per month. The solid line is the bottom 10% quantile of lenders; the light dashed line is the median lender; the heavy dashed line is the 90% quantile of lenders. The top panel shows how many loans these lenders make per month; the bottom panel shows the dollar value of loans these lenders make per month. Scales are log.

Turning to borrowers, Figure 6 breaks out the number previous loans and dollar value of previous loans that typical borrowers have in their borrowing history at the time of receiving a new loan. The figure again shows wide disparity in borrowing activity. The bottom 10% is not visible on the graph because the bottom quantile of borrow has no loan history at the time of receiving a loan. Early in the sample, the median funded borrower had received one or two loans totaling up to between \$100 and \$300 of borrowing; as the sample progresses, borrowers develop a history and that number rises to a median of four previous loans totaling to approximately \$500 of borrowing. The borrows in the 90% quantile have, by September 2016, borrowed about 16 times for a total value of over \$3000.

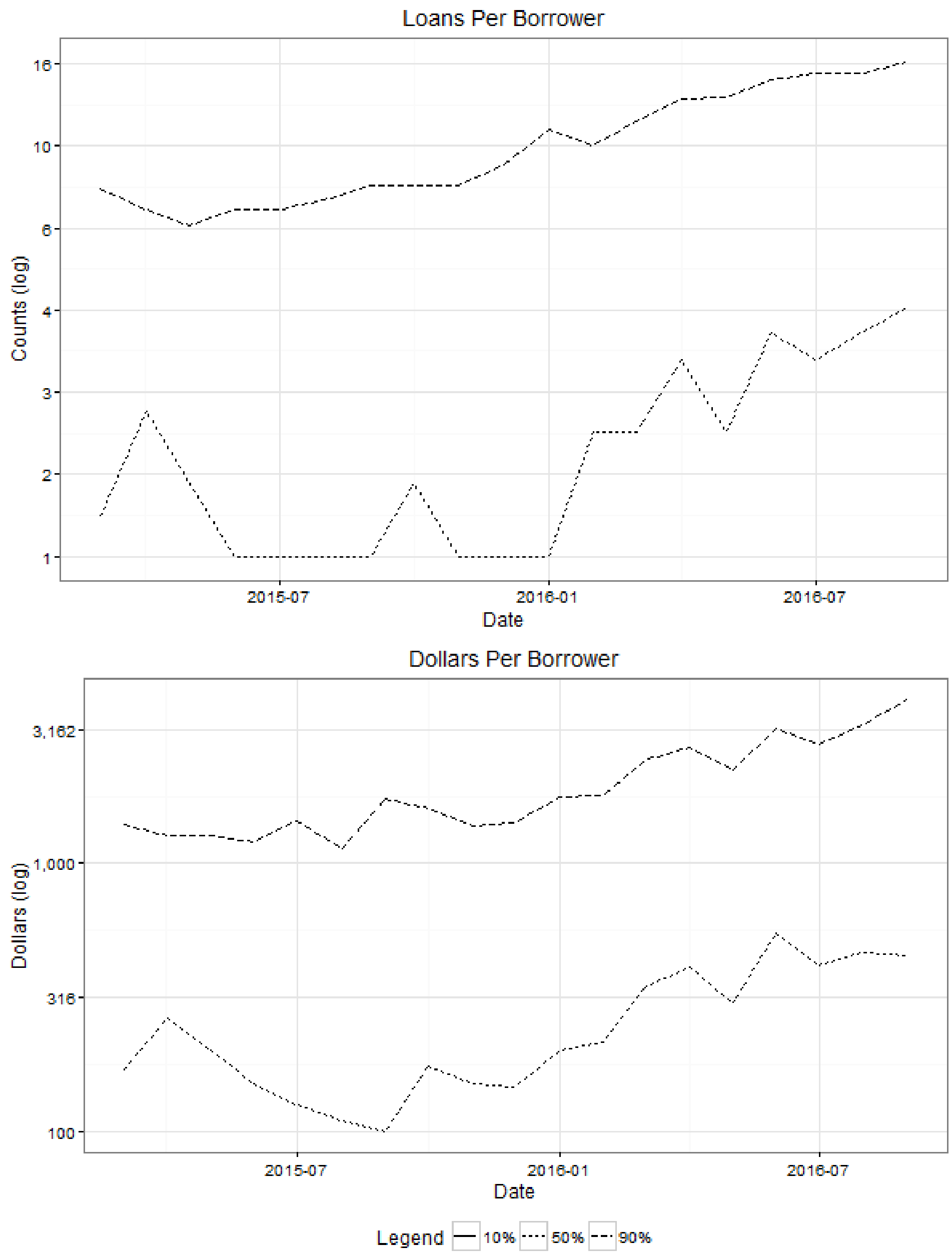

Figure 6: Borrower activity per month. The solid line is the bottom 10% quantile of borrowers by their borrowing history, but it is not visible here because it is always zero; the light dashed line is the median lender; the heavy dashed line is the 90% quantile of lenders. The top panel shows how many loans borrowers have in their history when receiving a new loan; the bottom panel shows how many dollars borrowed a borrower has in its history when receiving a new loan. Scales are log.

Finally, not only are there repeat borrowers and repeat lenders, but repeat borrower-lender pairs. Figure 7 breaks out the number of times and dollars previously transacted in realized borrower-lender pairs. The bottom 10% and median of these values are nearly always zero, but among the 90% quantile, there is significant and growing repeated borrower-lender pairing in lending.

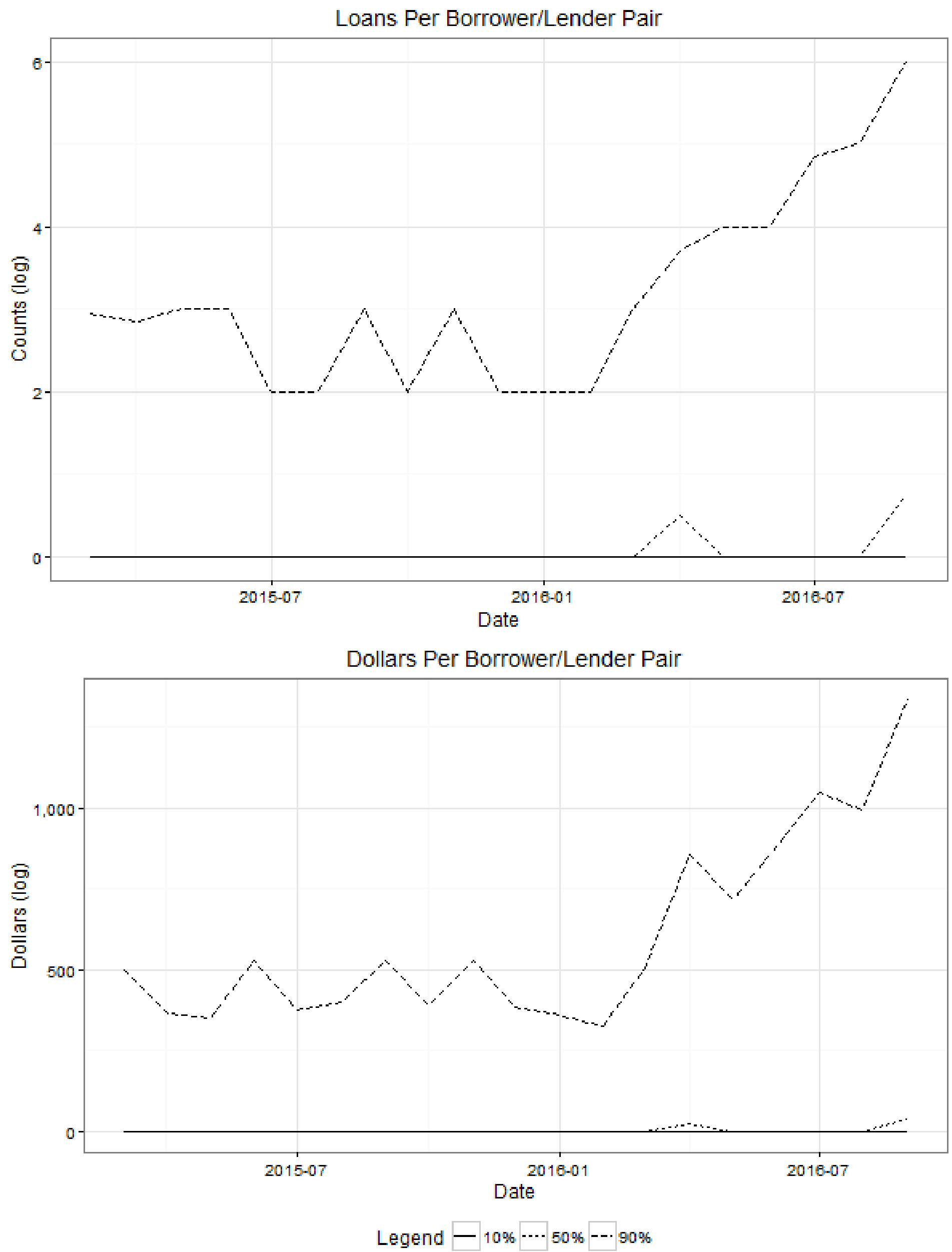

Figure 7: Borrower activity per month. The solid line is the bottom 10% quantile of borrowers by their borrowing history, but it is not visible here because it is always zero; the light dashed line is the median lender; the heavy dashed line is the 90% quantile of lenders. The top panel shows how many loans borrowers have in their history when receiving a new loan; the bottom panel shows how many dollars borrowed a borrower has in its history when receiving a new loan. Scales are linear.

The upshots from this descriptive analysis are as follows: First, the market has grown significantly since its inception and is now presently stable or slightly growing. Currently, roughly \$400 loans are made per month totaling to roughly \$100,000, which gives a mean loan size of about \$250. Second, the typical loan requires an interest rate of %20. Third, loans are repaid more than 90% of the time, and lenders on average appear to net somewhere between approximately \$3 and \$11 per \$100 lent in expectation. Most lenders in the market make just a few loans per month, but there are some very active lenders making more than 10 loans per month corresponding to more than \$3,000 lent. If these lenders are roughly as skilled as other lenders, back-of-the-envelope calculations suggest the most active lenders net somewhere between \$1,400 and \$3,900 per year with about as much working capital. Not nearly enough to live on but perhaps a nice source of supplemental income from a few hours working at home.

For the questions this paper is interested in, namely, why do borrowers repay, a few things are important: First, there are both new and returning borrowers. This allows us to study how the quasi-anonymous borrowers can gain entry into the lending community in the first place—while many borrowers are repeat borrowers, there is still a significant amount of entry from first-timers. Second, there is significant heterogeneity in how often borrowers return, which will be important in identifying the mechanisms driving borrowers to repay. Third, conditional on receiving a loan, the repayment rate is very high, suggesting that a significant amount of screening of borrower's ability and willingness to repay takes place at the lending decision.

3 Reduced-Form Empirical Analysis

There are two questions to test: First, is there evidence supporting the *homo economicus* view that borrowers repay because the rationally consider the future discounted value of market access? Second, how does borrowers' outside social activity aid in facilitating entry into the market? I address each question in turn.

3.1 Testing for Forward-Looking Borrowers

This section presents reduced-form evidence ruling in the *homo economicus* mechanism for why agents in this market repay. This economic setting is well-described by a *dynamic* *discrete choice* model. This is a very general class of model in which agents make repeated discrete choices where each decision has consequences both for the agent's current-period utility and discounted future utility. Agents make choices optimally taking both current and future values into account.[28] To be concrete, here a borrower chooses to repay or default; this choice has consequences today because a defaulting borrower keeps extra money but also may feel bad or suffer some social consequences, and has consequences for the future because the borrower loses market access for the future.

A general feature of these models is that it is very difficult to distinguish a model of a myopic agent who does not care about the future from a model of a forward-looking agent who does, using only past choice behavior and without making very specific parametric assumptions.[29] In the context of this paper and with the extra-legal contracting literature more generally, a *homo economicus* mechanism cannot be distinguished from a simpler *homo moralis* mechanism on the basis of lending and repayment histories alone.

What would identify such a model, however, is exogenous and ex-ante measurable variation in the future value of the remaining in the market. To illustrate, if agent A has greater future value from having the option to borrow than agent B but they are otherwise identical, then conditional on getting a loan, agent A will be more willing to repay than B *if and only if* agents care about the future. Taking a step back, if lenders can determine that agent A has greater future value from getting the loan than agent B, they will be more likely to lend to A other things equal. What this requires, however, is observable variation in future value. In this case, that variation is the agents' likelihood of needing loans in the future.

Recall Figure 6, which showed that there is considerable variation in the frequency with which borrowers ask for and receive loans. Some borrowers are needy in that they require frequent loans; some borrowers are independent in that they require loans infrequently or only once. A needy borrower is one that has frequent liquidity needs and is prone to come to the market often; an independent borrower does not. A needy borrower, therefore, has a high

[28] The foundational structural empirical estimation of such a model is the so-called "Rust Bus" in Rust (1987), where the behavior of Harold Zurcher, superintendent of maintenance at the Madison Metropolitan Bus Company, in replacing bus engines.

[29] In particular, suppose an agent discounts future utility by a factor $\beta \in [0, 1)$. If k is the choice and x is the state variable, without structural assumptions on the period utility function $u(k; x)$, β—even whether $\beta \neq 0$—cannot be identified from the history of past decisions because the period utility function in a $\beta = 0$ world could exactly coincide with the value function $v(k, x)$ in a $\beta \neq 0$ world and produce the same observed behavior. Even with functional form assumptions on $u(k; x; \theta)$, with parameters θ, identification of β is difficult because there are often θ_m's for which $u(k; x; \theta_m)$ in a myopic model coincides closely with $v(k, x; \theta_d)$ in a dynamic model. Hence, these models typically *assume* a non-zero value for β and do not attempt to estimate it or even show that it must be non-zero. See generally Rust (1994), Magnac and Thesmar (2002), and Chen et al. (2009).

valuation in future access to the market relative to an independent borrower if borrowers are not myopic and exhibit *homo economicus* behavior. In comparing a needy borrower to an independent borrower, *homo economicus* behavior should imply that needy borrowers are more likely to repay than independent borrowers, other things being equal, and to the extent that neediness and independence are observable to lenders ex-ante, needy borrowers should be more likely to get loans.

To illustrate simply how the probability of needing a loan in the future will impact the value of continued market access, consider a simplified model where the borrower receives positive utility flow u each time she is able to access the market and repays, and receives positive utility flow n if she accesses the market but defaults, and discounts the future at rate $\beta \in [0, 1)$, and has the need to access the market each period with probability λ. We compare the value of a borrower's reputation bond if she always repays to the value if she defaults. Before a borrower learns if she needs to access the market *this period*, her *value function*, v—equivalently her *reputation bond*, can be written recursively and solved for algebraically:

$$\begin{aligned} \overbrace{v}^{\text{current value}} &= \overbrace{\underbrace{\lambda u}_{\text{needs access}} + \underbrace{(1-\lambda)0}_{\text{does not need access}}}^{\text{period utility}} + \overbrace{\beta v}^{\text{discounted future value}} \\ v &= \frac{\lambda u}{1-\beta} \end{aligned}$$

Consider the decision of a borrower who has accessed the market but has not decided whether or not to repay. She makes the following choice: Repay today, receive u, and get the discounted value of the revenue bond, βv, or default today, receive n, and forfeit the revenue bond. The value of repaying is $u + \beta v$. The value of defaulting is n, and she repays if:

$$\begin{aligned} n &< u + \beta v \\ &= u + \frac{\beta}{1-\beta}\lambda u \end{aligned}$$

Note that if the agent does not care about the future—if $\beta = 0$—then the present value of the reputation bond is zero and importantly does not depend on λ. If the agent does care about the future—if $\beta > 0$—then the value of the reputation bond is positive and strictly increasing in λ. This is the mathematical intuition for why variation in λ will give variation in behavior if and only if borrowers make decisions caring about the future.

Of course, the challenge is to identify high λ (needy) and low λ (independent) borrowers ex-ante. Fortunately, the market structure provides a convenient way to estimate this which is observable to both lender and econometrician: the text of the borrower's loan request. The details of this procedure are given in the Appendix, Section 5.1, but the output of this is an ex-ante measure of $\lambda(X_{it})$, the probability that a borrower will request a loan in the future as a function of observables—in particular, the text of the borrower's post.

3.1.1 Estimation Strategy

I estimate the following reduced-form linear probability model relating whether a borrower's loan request is funded to observables about the borrower, including, importantly, an ex-ante measure of whether the borrower will request another loan in the future:

$$\text{Loan Made}_{it} = X_{it}'\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \underbrace{\hat{\gamma} \times \hat{\rho}(X_{it})}_{\text{somtimes}} + \underbrace{\phi_i}_{\text{sometimes}} + \xi_t + e_{it} \tag{1}$$

Where X_{it}' is a vector of observables about the borrower and $\hat{\lambda}(X_{it})$ is an ex-ante estimate of $\lambda(X_{it})$, the probability that the borrower will request a loan in the future, and $\hat{\rho}(X_{it})$, included as a control in some specifications, is a similarly constructed ex-ante estimate that the borrower will repay a loan conditional on observables about the borrower. ϕ_i is a borrower-specific fixed effect and ξ_t is a year-month fixed effect. Including ϕ_i identifies the impact of variation in $\hat{\lambda}(X_{it})$ within a borrower across her posts, as opposed to across borrowers and posts. ξ_t absorbs time-variation in market conditions, which is not the object of study here. e_{it} is a vector of error terms. which includes

The justification for this specification is given in the Appendix, Section 5.2. The upshot, given more technically in the Appendix, is that an estimate of $\hat{\delta}$ statistically greater than zero implies that borrowers are making choices that are in part led by *homo economicus* forward-looking planning. That is, the reputation bond has value and this channel is active in inducing repayment. The intuition is straightforward in that borrowers who are more likely to need to borrower in the future value continued market access more. If agents choose to repay in part because they do not wish to lose future market access, those agents who value market access more will be more likely to repay. Because agents who need market access more frequently are more likely to repay, lenders are more likely to lend to them ex-ante.[30]

[30]Measuring ex-ante likelihood of returning is important as compared to an ex-post measure, for instance, the number of repayments a borrower has made, because in addition to conveying information about the borrower's need to access the market in the future, the borrower's past choices to repay could be correlated with other non-dynamic reasons for repaying like intrinsic honesty or susceptibility to social pressures.

In order to draw the above inference, there are a number of identifying assumptions that need to be made regarding e_{it}. e_{it} includes (1) measurement error in the assumed probability of asking for a subsequent loan, (2) random uncontrolled-for variation in the borrower's ability to repay, (3) random uncontrolled-for variation in the borrower's choice to repay, and (4) random uncontrolled-for variation in the lenders's choice to lend. Measurement error in $\hat{\lambda}(X_{it})$ will induce unavoidable attenuation bias towards zero in $\hat{\delta}$ which will make it more difficult to reject the null hypothesis that $\delta = 0$, essentially making identification more difficult by increasing the probability of a false negative. A measured value of $\hat{\delta}$ statistically significantly greater than zero in the presence of this measurement error implies that it would be statistically significantly greater than zero without it. For problems (2), (3), and (4), the necessary identifying assumptions are that $\hat{\delta}$ is uncorrelated with unobservable variation in the borrower's ability to repay, unobservable variation in the borrower's choice to repay, and unobservable variation in the lender's choice to lend, after controlling for other observables. On (4), by the construction of the model given in the Appendix, unobservable variation in the lender's choice to lend is orthogonal to everything in the model and systematic components are in effect pushed into problems (2) and (3). On (3), a fair assumption is that the probability that a borrower is prone to receiving liquidity shocks in uncorrelated to borrower-specific or borrower-request specific non-dynamic reasons for not repaying. That is, the problem (3) would bring to estimation would require that, for example, intrinsic moral compulsion and intrinsic need for lending are positively correlated, and one would not expect this ex-ante. As a precaution, however, including lender fixed-effects, as some specifications do, will address this problem.

Identification problem (2), that the ex-ante need for a loan is correlated with unobservable "fundamental" reasons for default, i.e., for unobservable reasons that a borrower *cannot* repay is the most threatening. For instance, borrowers that need loans frequently may need them because they are always a few weeks behind on the bills and can pay as soon as the next paycheck comes. These kinds of borrowers are not "fundamentally" risky because they have a regular source of income and there is merely a timing mismatch. To handle this, first, within-borrower estimation with borrower fixed effects will address this problem to the extent that it is a non-time-varying borrower-specific problem. With these fixed effects, the identifying variation comes from differences in how a single user expresses her request across multiple requests. Second, to enable cross-borrower estimation, some specifications include a best-guess of the borrower's ex-ante "fundamental" ability to repay, $\hat{\rho}(X_{it})$, which is constructed similarly to how $\hat{\lambda}(X_{it})$ is constructed. In particular, looking at funded loans, a machine learning algorithm attempts to use borrower observables, in particular, the text of the borrower's post, to estimate whether the borrower will repay or not. The assumption here is that lenders are good at screening out borrowers that *choose* not to repay, and that in equilibrium the only borrowers that default do so because they *cannot* repay. Including $\hat{\rho}(X_{it})$ allows the specification to control for variation in the *ability* to repay that may be correlated with the paper's estimate of a borrower's probability of receiving liquidity shocks in the future.

3.1.2 Results

The results of the regression specification given in Equation (1) are given in Table 1. The null hypothesis, that forward-looking borrower behavior is inoperative in incentivizing repayment implies that $\delta = 0$. As the table shows, this hypothesis is soundly rejected across all specifications.

Table 1: Results of Loan Made$_{it}$ $= X_{it}'\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \hat{\gamma} \times \hat{\rho}(X_{it}) + \phi_i + \xi_t + e_{it}$. The parameter of interest is $\hat{\delta}$, whose estimate is given in the first row of the table. The null hypothesis is that agents are not forward-looking, which gives $\delta_0 = 0$. Columns (1)-(3) do not include the non-linear repayment control $\hat{\rho}(X_{it})$; columns (4)-(6) do. Columns (1) and (4) use all loan requests without borrower fixed effects. Columns (2) and (5) use first-time requests only without borrower fixed effects. Columns (3) and (6) use all loan requests with borrower fixed effects. $\hat{\lambda}(X_{it})$ and $\hat{\rho}(X_{it})$ are ex-ante estimates of subsequent requests and repayment, respectively. Standard errors are clustered at the borrower level. In the interest of brevity, I omit the estimates for borrower controls, which are found in full in Table 6 in the Appendix.

	Dependent variable:					
	Loan Made					
	(1–All)	(2–New)	(3–FE)	(4–All)	(5–New)	(6–FE)
$\hat{\lambda}(X_{it})$	0.419***	0.871***	0.478***	0.397***	0.848***	0.463***
	(0.077)	(0.141)	(0.162)	(0.077)	(0.141)	(0.162)
$\hat{\rho}(X_{it})$	-	-	-	0.352***	0.398	0.272
	-	-	-	(0.128)	(0.246)	(0.210)
Borrower Controls	Y	Y	Y	Y	Y	Y
Year-Month FE	Y	Y	Y	Y	Y	Y
Borrower FE	N	N	Y	N	N	Y
Observations	4,174	1,493	4,174	4,174	1,493	4,174
R^2	0.281	0.147	0.731	0.282	0.148	0.731
Note:			Borrower Clusters; *p<0.1; **p<0.05; ***p<0.01			

Across all specifications the general patterns are as follows: First, ex-ante likelihood of requesting another loan is highly significant in whether the borrower receives a loan. When included, ex-ante (non-linear) likelihood of repaying the loan conditional on receiving it also positively predicts receiving a loan, which is not surprising. Outside measures the borrower's cumulative activity elsewhere on the site at the time of the request by counting the log number of posts and submissions. The effect is consistently positive and significant, meaning that borrowers with more outside activity are more likely to receive loans. I return to this in the next section. Across all specifications, the size of the loan negatively is negatively associated with whether the loan is made. This is unsurprising, as a higher loan size makes default more desirable. Borrower credit history, as captured by whether the borrower has repaid or defaulted in the past contributes in the expected way: a repayment makes the borrower more likely to receive a loan; more defaults a default makes the borrower less likely to receive a loan. There are so few borrowers that request loans after defaulting, however, that this number is not significant. Measures of the borrower's "network" also matter for his getting a loan. log(Rel Size) totals the number of loans the borrowers' past lenders have made—it measures whether borrowers are connected to large or small lenders, and it appears that being connected to large lenders weakly aids in getting a subsequent loan. log(Rel Out) measures the number of outstanding loans the borrower's past lenders currently have. It is not clear which way this should go ex-ante, as if the borrower's usual suppliers are busy, it could either mean that they lack capacity to lend, or that they are currently in an active phase of lending.

The coefficient of interest, however, is $\hat{\delta}$, the coefficient on $\hat{\lambda}(X_{it})$. Consider first columns (1) and (4), which is a regression on the whole (test) sample of borrowers not including and including the non-linear estimate of repayment probability, respecively. To the extent that the exclusion restrictions are satisfied, the interpretation is that appearing ex-ante more likely to need a loan in the future, controling for other observables, makes a borrower more likely to receive a loan. Columns (2) and (5) focus only on first-time borrowers, and find that the effect is even stronger there. This is to be expected: For first time borrowers, lenders have no credit history on which to rely and therefore must place more weight on textual clues about the borrower's likelihood of needing a loan in the future. Columns (3) and (6) are among all borrowers but using borrower fixed effects. This specification compares requests within a borrower and shows that even looking within a single borrower, those posts that indicate the borrower is likely to return in the future are more likely to be funded. Including borrower fixed effects removes concerns about correlations of $\hat{\lambda}(X_{it})$ to non-time varying borrower unobservables, presumably things like honesty, fundamental creditworthiness, and so on.

In sum, this section presents evidence that a borrower's perceived likelihood of returning for another loan is significant in whether the borrower receives the loan in the first place. This makes sense, because borroweres who are likely to need more loans have better incentives to repay them, assuming that borrowers and lenders engage in forward-looking economic planning.

3.2 Outside Social Participation and Entry

Earlier discussion showed that the manner in which new borrowers enter is a critical point of potential market failure. In particular, this setting is subject to the danger that a scammer can repeatedly create a new account after defaulting on a previous loan, ad infinitum. Here, *social capital*, in the form of difficult-to-fake outside social activity that is tied to an account is critical in mediating entry. In particular, lenders will only lend to borrowers with an established outside social presence, and this norm prevents the feared defaulting and account recreation behavior.

3.2.1 Estimation Strategy

I estimate the following reduced-form linear probability model relating whether a borrower's loan request is funded to observables about the borrower, with a specific focus on the borrower's outside social activity:

$$\text{Loan}_{it} = \beta \times \text{Out}_{it} + \psi \times \text{Out}_{it} \times \text{Old}_{it} + X_{it}'\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \hat{\gamma} \times \hat{\rho}(X_{it}) + \xi_t + e_{it} \quad (2)$$

This is the same specification used in the previous section with the added interaction term, $\text{Out}_{it} \times \text{Old}_{it}$, the borrower's outside social interaction times an indicator for whether the borrower has repaid a loan in the past. A *social capital as a screening mechanism* story predicts that outside social capital is important *for the borrower's first loan* but not for subsequent loans. A *social capital as collateral for repayment* predicts that outside social capital should continue to matter for subsequent loans as well. Both theories predict a positive value for β, meaning that social capital matters, but the screening story predicts further that ψ, the coefficient on the interraction term, should be negative and exactly offset β. The results follow.

3.2.2 Results

The results of the regression specification given in Equation (2) are given in Table 2. A *social capital as screening mechanism* model predicts that $\beta > 0$ and $\beta + \psi = 0$; a *social capital as collateral mechanism* model predicts that $\beta > 0$ and $\psi = 0$. The results are consistent with the *social capital as screening* mechanism.

Table 2: Results of $\text{Loan}_{it} = \beta \times \text{Out}_{it} + \psi \times \text{Out}_{it} \times \text{Old}_{it} + X'_{it}\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \hat{\gamma} \times \hat{\rho}(X_{it}) + \xi_t + e_{it}$. See description of Table 1 for variable descriptions. The only addition here is the interaction term, outside activity times whether the borrower has repaid a loan. Columns (1) and (2) are over the testing sample used above; Column (3) uses the entire sample, including the half used to fit the machine learning estimates, and omits these variables. Standard errors are clustered by borrower. In the interest of brevity I omit the full covariates, whose coefficients are reported in 7 in the Appendix.

	Dependent variable:		
	Loan Made		
	(1)	(2)	(3)
Out	0.015**	0.022**	0.032***
	(0.007)	(0.009)	(0.006)
Out × Repaid	-	−0.023*	−0.038***
	-	(0.013)	(0.009)
Borrower Controls	Y	Y	Y
Year-Month FE	Y	Y	Y
Subsample	Test	Test	Full
Observations	4,174	4,174	10,021
R^2	0.282	0.283	0.267
Note:	*p<0.1; **p<0.05; ***p<0.01		

The covariates of interest are Out and Out × Repaid. In column (1), there is no interaction term and the coefficient on Out means that unconditionally, having more outside social engagement on the site is associated with being more likely to receive a loan. This, however, does not distinguish between social interaction being a screening mechanism to prevent entry from scamming borrowers and social interaction being a form of valuable collateral. Columns (2) and (3) introduce the interaction term. The difference between columns (2) and (3) is the sample used: Column (2) uses the restricted out-of-sample testing observations set-aside when calculating $\hat{\lambda}(X_{it})$ and $\hat{\rho}(X_{it})$. Column (3) uses the entire sample and omits those variables. The results in columns (2) and (3) are consistent with social interaction being a screening mechanism. In particular, the results show that once a borrower has repaid his loan, social capital is no longer important in securing additional loans. This is consistent with screening rather than collateral because the screening function is complete after the first loan, whereas the collateral function should still be in effect after the first loan.

4 Discussion and Conclusion

I examined the mechanisms enforcing extralegal contracting and the role of social capital in facilitating new entry. The conclusions support a model of forward-looking optimizing behavior driving repeated borrowers to repay, and an interpretation of outside social capital as filling a screening, rather than a social capital role. Further research will have a structural model to facilitate counter-factual analysis of institutional features.

References

Aguiar, M. and Amador, M. (2013). Sovereign debt: a review. Technical report, National Bureau of Economic Research.

Alger, I. and Weibull, J. W. (2013). Homo moralis-preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302.

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.

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*.

Borensztein, E. and Panizza, U. (2009). The costs of sovereign default. *IMF Economic Review*, 56(4):683–741.

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

Chen, L.-Y. et al. (2009). Identification of structural dynamic discrete choice models. In *2009 Far East and South Asia Meeting of the Econometric Society (Tokyo, Japan: University of Tokyo)*.

Cole, H. L. and Kehoe, P. J. (1998). Models of sovereign debt: Partial versus general reputations. *International Economic Review*, pages 55–70.

Eaton, J. and Gersovitz, M. (1981). Debt with potential repudiation: Theoretical and empirical analysis. *The Review of Economic Studies*, 48(2):289–309.

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.

Freedman, S. M. and Jin, G. Z. (2011). Learning by doing with asymmetric information: evidence from prosper. com. Technical report, National Bureau of Economic Research.

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

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.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2015). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Magnac, T. and Thesmar, D. (2002). Identifying dynamic discrete decision processes. *Econometrica*, 70(2):801–816.

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.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033.

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143.

Stokey, N. L. and Lucas, R. (1989). *Recursive methods in economic dynamics*. Harvard University Press.

5 Appendix

5.1 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.[31]

5.1.1 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.

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

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

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.

5.1.2 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 3 shows the regression. Table 4 shows the Partial F test. The results are highly significant in the desired direction.

Table 3: Regression of $subsequentrequest = \hat{\lambda}(X_{it} + X'_{it}\Gamma + \epsilon_{it}$

	Dependent variable:			
	subsequentPost			
	(1)	(2)	(3)	(4)
$\hat{\lambda}(X_{it})$	1.018***	1.026***	0.859***	0.862***
	(0.092)	(0.094)	(0.115)	(0.116)
$\hat{\rho}(X_{it})$		−0.126		−0.053
		(0.237)		(0.237)
Outside			−0.028**	−0.027**
			(0.013)	(0.013)
log(principal)			−0.037***	−0.037***
			(0.013)	(0.013)
Has Repaid			0.068**	0.069**
			(0.034)	(0.034)
Constant	−0.014	0.099	0.393***	0.438*
	(0.060)	(0.219)	(0.126)	(0.239)
Observations	1,649	1,649	1,649	1,649
R^2	0.078	0.079	0.089	0.089
Note:				*p<0.1; **p<0.05; ***p<0.01

Finally, Table 5 shows predictions of repayment conditional on being funded. This table addresses two 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

Table 4: Regression of $subsequentrequest = \hat{\lambda}(X_{it}) + X_{it}'\Gamma + \epsilon_{it}$, partial F tests.

	Dependent variable:				
	subsequentPost				
	DF	Sum Sq.	Mean Sq.	F	P(>F)
$\hat{\lambda}(X_{it})$	1	30.08	30.0791	141.3016	$<$ 2.2e-16***
$\hat{\rho}(X_{it})$	1	0.07	0.0654	0.3074	0.579328
Outside	1	0.79	0.7852	3.6887	0.054956
log(principal)	1	2.10	2.1023	9.8760	0.001704**
Has Repaid	1	1.04	1.0447	4.9076	0.026875*
Residuals	1643	349.75	0.2129		
Note:			*p<0.05; **p<0.01; ***p<0.001		

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 (1) and the identifying assumptions needed there. The lack of significance here is reassuring.

Table 5: Regression of $repay = \hat{\lambda}(X_{it}) + X_{it}'\Gamma + \epsilon_{it}$ among funded loans in the testing sample.

	Dependent variable:
	Repaid
$\hat{\lambda}(X_{it})$	0.057
	(0.051)
$\hat{\rho}(X_{it})$	−0.019
	(0.104)
Outside	0.013*
	(0.007)
log(principal)	−0.011*
	(0.006)
Has Repaid	0.098***
	(0.018)
Constant	0.831***
	(0.113)
Observations	1,768
R^2	0.050
Note:	*p<0.1; **p<0.05; ***p<0.01

5.2 Reduced Form Specification

In this section, I write down the simple reduced form model to inform the proper regression specification and clarify the necessary identifying assumptions. A borrower repays if (1) the borrower can repay and (2) the borrower chooses to repay. Let X_{it} be a vector of observables for borrower i at time t. Denote by $p(\text{can}_{it}|X_{it})$ the probability that the borrower can repay given her observables X_{it}. Denote by $p(\text{does}_{it}|\text{can}_{it}, X_{it})$ the probability that a borrower chooses to repay given her observables X_{it} and given that she can repay. We are interested in studying $p(\text{does}_{it}|\text{can}_{it}, X_{it})$.

Assume lenders are risk neutral and hence lend if the following inequality is satisfied:

$$\begin{aligned} c &< e^{z_{it}}(1 + R_{it})p(\text{can}_{it}|X_{it})p(\text{does}|\text{can}_{it}, X_{it}) \\ z^{it} &\sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{3}$$

z_{it} captures unobserved lender reasons for making or not making the loan that are independent of the loan's expected value. For instance, lenders may be unavailable for exogenous reasons, lenders may be busy with other loans, or lenders may have suffered defaults on their other loans. Therefore, z_{it} is assumed to be independent from other variables in the model. Note that Equation (3) is an expected return threshold requirement.[32] Take logs, and with no loss of generality let $c = \log c$. Equation (3) is satisfied if:

$$c < \log(1 + R_{it}) + \log p(\text{can}_{it}|X_{it}) + \log p(\text{does}|\text{can}_{it}, X_{it}) + z_{it} \tag{4}$$

I assume that we can estimate $\log p(\text{can}_{it}|X_{it})$ with the following linear model:

$$\log p(\text{can}_{it}, X_{it}) = X_{it}'\Gamma + \phi_i^1 + \eta_{it} \tag{5}$$

Where ϕ_i^1 is a borrower-specific quantity observable to the lender but not the econometrician, for instance, something related to the borrower's intrinsic creditworthiness communicated credibly over private borrower-lender chat, and η_{it} is an ideosyncratic component specific to this particular loan request.

The standard dynamic discrete choice setup posits that agents choose the greater of two options taking into account the period utility flow from each choice and the dynamic

[32]Explain why ER is correct and risk aversion doesn't matter. It unambiguously correct although I expect complaints about this.

consequences of their choice.[33] In this setting, the borrower's choice is whether to repay or default; a *homo economicus* agent will care about the future value function, a myopic agent will not. The choice problem in full generality[34] is:

$$\text{decision} = \arg\max_{k\in\{r,d\}} \left\{ \underbrace{u_i(X_{it}|r) + \beta E v_i(X_{it}|r)}_{\text{repay}}, \underbrace{u_i(X_{it}|d) + \beta E v_i(X_{it}|d)}_{\text{default}} \right\}$$

$u(\cdot|\cdot)$ is the the in-the-moment utility; $v(\cdot|\cdot)$ is the future utility. The quantity $\Delta u_i(X_{it}) \equiv u_i(X_{it}|r) - u_i(X_{it}|d)$ is the period utility difference between repayment and default. In *homo moralis* and *homo socalis* models, this difference is what mediates lending; the *homo moralis* agent dislikes defaulting because she suffers negative utility from violating her ethics; the *homo socalis* agent dislikes defaulting because she is harmed socially from defaulting. The quantity $\Delta v_i(X_{it}) \equiv v_i(X_{it}|r) - v_i(X_{it}|d)$ is the lifetime future utility difference between repaying and defaulting today. In the *homo economicus* model this differnce is what mediates lending; defaulting kicks the borrower out of the market in the future, and the forward-looking *homo economicus* cares about this. This forward-looking agent overlooks the short-term benefits of default in favor of the long-term benefits of retaining market access.

This leads us to assume the following reduced form relationship:

$$\log p(\text{does}|\text{can}, X_{it}) = \beta_0 \underbrace{\Delta u_i(X_{it})}_{\text{myopic}} + \beta_1 \underbrace{\Delta v_i(X_{it})}_{\text{dynamic}} + \epsilon_{it}$$

The question is whether $\beta_1 > 0$, i.e., whether borrowers take the forward-looking dynamic effects of their decisions into account. An immediate identification problem presents itself because neither $\Delta u_i(X_{it})$ nor $\Delta v_i(X_{it})$ are observed, in that to identify β_1 separtaely from β_0, we must find variation in $\Delta v_i(X_{it})$ that does not impact $\Delta u_i(X_{it})$. We look for a proxy for $\Delta v_i(X_{it})$ that is uncorrelated with $\Delta u_i(X_{it})$. We find such exogeneaty as follows: assume that we can measure λ_{it}, the probability that the borrower will need another loan as a function of her observables. The simple insight is that a borrower who is more likely to need a loan in the future gains more from retaining market access. That is, if λ_{it} is the probability that a borrower will require funding in the future, then

$$\frac{\partial \Delta v_i(X_{it})}{\partial \lambda_{it}} > 0$$

[33]See Stokey and Lucas (1989) for general treatment of dynamic programming in Economics. See Rust (1987) for a specific application in dynamic discrete choice.

[34]Assuming additively separable utility across periods

The borrower's probably of returning for another loan does not have any bearing on the borrower's different flow utilities $\Delta u_i(X_{it})$, so this λ_{it} satisfies the necessary exclusion restriction.

Suppose that lenders form an estimate of the borrower's chances of repayment from her observables, $\lambda(X_{it})$. Assuming we can estimate $\lambda(X_{it})$ and proxy the period utility and rest of the value function differences as $X_{it}'\Omega$, we write:

$$\log p(\text{does}_{it}|\text{can}_{it}, X_{it}) = X_{it}'\Omega + \delta\lambda(X_{it}) + \phi_i^2 + \epsilon_{it} \tag{6}$$

Where ϕ_i^2 is a borrower-specific quantity known to the lender but not the econometrician and ϵ_{it} is a random shock observed to the borrower and not the econometrician. ϕ_i^2 could represent intrinsic honesty, for example. The parameter of interest in this model is δ. If agents are fully incentivized by moralistic or social pressures to repay their loans and that the threat of market exclusion does not matter, then the probability of getting a loan in the future should not impact their decision to repay, and $\delta = 0$. If, on the other hand, agents are not myopic and have *homo economicus* behavior, then agents who are more likely to need loans in the future will be more likely to repay loans today, and $\delta > 0$. Inject Equations (5) and (6) into Equation (4) to obtain the following equation which determines when lending occurs:

$$c < log(1 + R_{it}) + X_{it}'\Delta + \delta\lambda(X_{it}) + \phi_i + \eta_{it} + \epsilon_{it} + z_{it} \tag{7}$$

Where $\Delta \equiv \Gamma + \Omega$ and $\phi_i \equiv \phi_i^1 + \phi_i^2$. Note that Γ and Ω are not separately identified, so as a causal matter we cannot determine how elements of X_{it} impact separately the ability to repay and the choice to repay conditional on having the ability.

To estimate δ, I first obtain an estimator for $\lambda(X_{it})$ as described in the Appendix[35]. Next, from Figure 2, observe that interest rates are essentially fixed and I replace R_{it} by $\bar{R}$ and absorb it into a constant.[36] The specification to run is:

$$\text{Loan Made}_{it} = X_{it}'\hat{\Delta} + \hat{\delta}\hat{\lambda}(X_{it}) + \underbrace{\hat{\phi}_i}_{\text{sometimes}} + e_{it} \tag{8}$$

First, there is a measurement error problem in estimating λ, i.e., $\hat{\lambda}(X_{it}) = \lambda(X_{it}) + \nu_{it}$. I assume that this is a *classical* measurement error problem in that ν_{it} is independent from

[35]Need to write this

[36]Strictly speaking, I do not need to do this but it reduces th e amount of data I have substantially and makes little different on the point estimates.

X_{it} and everything else in the model, including η_{it}, ϵ_{it}, and z_{it}. Classical measurement error here will cause attenuation bias in the estimate of δ but will not affect the sign; in effect, it will make finding significance harder.[37]

Second, I assume that $\lambda(X_{it})$ is uncorrelated with η_{it}, which is borrower specific error in the probability that the borrower repays. That is, the probability that a borrower requests another loan must be uncorrelated with the probability that the borrower is fundamentally unable to repay conditional on all other observables. A reason that this could be violated is that the reasons for borrowing that tend to be repeated are also easy to repay. For instance, if a borrower is constantly two weeks behind paying her bills but has a source of income, the borrower is likely to ask again and also to have cash when it is necessary to repay. Some regresion specifications include an additional control term, $\hat{\rho}(X_{it})$, which is a function fitted in the same way that $\hat{\lambda}(X_{it})$ is fitted to measure whether a borrower repays conditional on receiving a loan.[38] The idea here is that if borrowers effectively determine who will and will not repay when they can, differences in ex-post repayment are driven only by fundamental inability to repay. Consequently, $\hat{\rho}(X_{it})$ controls for non-linear differences in ability to repay that might be correlated with $\hat{\lambda}(X_{it})$.

Third, $\hat{\lambda}(X_{it})$ must be uncorrelated ϵ_{it}, i.e., with unobserved non-dynamic reasons for defaulting. I think there is no strong reason to believe they would be correlated but to the extent that it is correlated with intrinsic honesty or something, specifications with borrower fixed-effects should take care of this. Fourth, $\hat{\lambda}(X_{it})$ must be uncorelataed with z_{it}, i.e., with unobserved reasons for the borrower not making this loan. By assumption, z_{it} is independent from other variables in the model so this is satisfied. Finally, to the extent that network or past relationships matter, I control for borrowers' past lender relationships in addition to their borrowing and repayment history.

5.3 Full Regression Tables

This appendix contains the regression tables with all covariates from Tables 1 and 2.

[37]Actually, it's probably better to say that the specification is $\delta(v_{it})$ v_{it} is the value function and I'm using $\lambda(X_{it})$ for variation in v_{it}.

[38]See the appendix for this.

5.3.1 *Homo Economicus* Full Covariates

Table 6: Results of Loan $\text{Made}_{it} = X'_{it}\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \hat{\gamma} \times \hat{\rho}(X_{it}) + \phi_i + \xi_t + e_{it}$. $\hat{\lambda}(X_{it})$ and $\hat{\rho}(X_{it})$ are ex-ante estimates of subsequent requests and repayment, respectively. *Outside* is the log sum of all submissions and comments made sitewide at the time of the request. *log*(*Principal*) is the log of the requested amount. *HasRepaid* and *HasDefaulted* are dummy variable indicating whether the borrower has repaid or defaulted in the past. *log*(*RelSize*) is the log of the total number of fills the borrowers' past lenders have made. *log*(*RelOut*) is the log of the number of loans the borrower's past lenders currently have outstanding. Standard errors are clustered at the borrower level.

	Dependent variable:					
	Loan Made					
	(1–All)	(2–New)	(3–FE)	(4–All)	(5–New)	(6–FE)
$\hat{\lambda}(X_{it})$	0.419***	0.871***	0.478***	0.397***	0.848***	0.463***
	(0.077)	(0.141)	(0.162)	(0.077)	(0.141)	(0.162)
$\hat{\rho}(X_{it})$	-	-	-	0.352***	0.398	0.272
	-	-	-	(0.128)	(0.246)	(0.210)
Outside	0.018**	0.046***	0.142**	0.015**	0.041***	0.141**
	(0.007)	(0.011)	(0.059)	(0.007)	(0.011)	(0.058)
log(principal)	−0.089***	−0.075***	−0.088***	−0.091***	−0.077***	−0.088***
	(0.008)	(0.011)	(0.018)	(0.008)	(0.011)	(0.018)
Has Repaid	0.156***	-	−0.013	0.152***	-	−0.019
	(0.035)	-	(0.053)	(0.035)	-	(0.053)
Has Defaulted	−0.033	-	−0.247	0.007	-	−0.214
	(0.115)	-	(0.217)	(0.116)	-	(0.221)
log(Rel Size)	0.028*	-	−0.015	0.026*	-	−0.015
	(0.015)	-	(0.024)	(0.015)	-	(0.024)
log(Rel Out)	0.014	-	−0.003	0.017	-	−0.003
	(0.018)	-	(0.029)	(0.018)	-	(0.029)
Year-Month FE	Y	Y	Y	Y	Y	Y
Borrower FE	N	N	Y	N	N	Y
Observations	4,174	1,493	4,174	4,174	1,493	4,174
R^2	0.281	0.147	0.731	0.282	0.148	0.731
Note:	Borrower Clusters; *p<0.1; **p<0.05; ***p<0.01					

5.3.2 Outside Social Participation Full Covariates

Table 7: Results of $\text{Loan}_{it} = \beta \times \text{Out}_{it} + \psi \times \text{Out}_{it} \times \text{Old}_{it} + X_{it}'\hat{\Delta} + \hat{\delta} \times \hat{\lambda}(X_{it}) + \hat{\gamma} \times \hat{\rho}(X_{it}) + \xi_t + e_{it}$. See description of Table 1 for variable descriptions. The only addition here is the interaction term, outside activity times whether the borrower has repaid a loan. Columns (1) and (2) are over the testing sample used above; Column (3) uses the entire sample, including the half used to fit the machine learning estimates, and omits these variables. Standard errors are clustered by borrower.

	Dependent variable:		
	Loan Made		
	(1)	(2)	(3)
Out	0.015**	0.022**	0.032***
	(0.007)	(0.009)	(0.006)
Out × Repaid	-	−0.023*	−0.038***
	-	(0.013)	(0.009)
$\hat{\lambda}(X_{it})$	0.397***	0.391***	-
	(0.077)	(0.077)	-
$\hat{\rho}(X_{it})$	0.352***	0.342***	-
	(0.128)	(0.128)	-
Has Repaid	0.152***	0.290***	0.368***
	(0.035)	(0.085)	(0.058)
log(principal)	−0.091***	−0.091***	−0.096***
	(0.008)	(0.008)	(0.005)
Has Defaulted	0.007	0.002	−0.131**
	(0.116)	(0.119)	(0.060)
log(Rel Size)	0.026*	0.026*	0.052***
	(0.015)	(0.015)	(0.010)
log(Rel Out)	0.017	0.019	0.001
	(0.018)	(0.018)	(0.011)
Year-Month FE	Y	Y	Y
Observations	4,174	4,174	10,021
R^2	0.282	0.283	0.267
Note:			*p<0.1; **p<0.05; ***p<0.01