Signaling in Online Credit Markets*

Kei Kawai† | Ken Onishi‡ | Kosuke Uetake§

New York University | Northwestern University | Northwestern University

November 2012

Abstract

This paper studies how signaling can facilitate the functioning of a market with classical adverse selection problems. Using data from Prosper.com, an online credit market where loans are funded through auctions, we provide evidence that reserve interest rates that borrowers post can serve as a signaling device. We then develop and estimate a structural model of borrowers and lenders where low reserve interest rates can credibly signal low default risk. Announcing a high reserve interest rate increases the probability of receiving funding at the cost of higher expected interest payments conditional on obtaining a loan. Borrowers regard this trade-off differentially, which results in a separating equilibrium. Using the estimated parameters of the model, we compare the credit supply curve and welfare under three alternative market designs in our counterfactual policy experiment – a market with signaling, a market without signaling, and a market with no asymmetric information. We find that the cost of adverse selection can be as much as 16% of the total surplus created under no asymmetric information, up to 95% of which can be restored with signaling. We also estimate the credit supply curves for each of the three market designs and find backward–bending supply curves for some of the markets, consistent with the prediction of Stiglitz and Weiss (1981).

JEL Code: D82, G14

*We thank Igal Hendel, Hide Ichimura, Alessandro Lizzeri, Aviv Nevo, Isabelle Perrigne, Rob Porter, Jeffery Prince, Quang Vuong, Yasutora Watanabe, and Michael Whinston for their valuable comments and suggestions.

†(Corresponding Author) New York University, Stern School of Business, Henry Kaufman Management Center, Economics Department, #7-82, 44 W 4th St., New York, NY, 10012. Fax: 212-995-4218. Phone: 212-998-0760. Email: kkawai@stern.nyu.edu.

‡Northwesern University: onishi@u.northwestern.edu

§Northwesern University: uetake@u.northwestern.edu

1 Introduction

Inefficiencies arising from adverse selection is a key feature in many markets, with examples ranging from "lemons" in used car markets (Akerlof, 1970) to toxic assets in financial markets (Morris and Shin, 2012). An important source of inefficiency in these markets lies in the inability of agents who are of "good" types (e.g., sellers of high–quality cars) to distinguish themselves from the "bad" (e.g., sellers of low–quality cars), resulting in markets to unravel completely in the worst–case scenario. The key insight of Spence (1973), however, is that when costly signaling devices are available, agents who have different marginal cost of signaling can be induced to take action that reveals their true type in equilibrium. Hence signaling can prevent the market from unraveling, with possibly large welfare implications.

In this paper, we empirically study how signaling affects the functioning of a market for unsecured loans using data from Prosper.com, an online peer-to-peer lending market where potential borrowers are directly matched to potential lenders through auctions. At least since the seminal work of Stiglitz and Weiss (1981), markets for unsecured loans have been considered to be classic examples of markets that suffer from potential adverse selection problems. A key feature of Prosper.com, however, is that each borrower can post a public reserve interest rate – the maximum interest rate that the borrower is willing to accept – when the borrower creates a listing on its Web site. A reserve interest rate is the equivalent of a reserve price in standard auctions; we explore how the borrower's reserve interest rate can signal his creditworthiness in this market and how signaling interacts with lending and repayment of loans.

The idea that the reserve interest rates can signal the borrowers' creditworthiness is quite intuitive in the particular market we study. Consider, for example, a borrower who is posting a high reserve rate – say, higher than the prime rate charged for typical bank loans. Then one may infer that this borrower faces difficulty borrowing from outside sources, which in turn raises concerns about the creditworthiness of the borrower. Of course, this intuition is not a complete explanation of signaling, because there needs to be a countervailing force that induces borrowers to post higher reserve interest rates. In the market we study, the natural countervailing force is the probability of obtaining a loan. If listings with very low reserve rates are funded with very low probability and, moreover, if the funding probability increases as a function of the reserve

rate, then this can counteract the incentive for the borrower to post low reserve rates. These two opposing incentives create different trade-offs for different borrowers, giving rise to the possibility of equilibrium dispersion in the reserve rate.

This rather simple intuition forms the basis of our model of the borrowers. In our model, borrowers are heterogeneous with regard to the cost of borrowing from outside sources and the ability to repay the loan once it has been made. Given a trade-off between higher funding probability and higher interest rate, the heterogeneity among the borrowers regarding the cost of borrowing translates to the single-crossing condition. The low-cost types (e.g., borrowers with easy access to credit from local banks etc., who use Prosper primarily for obtaining more favorable interest rates) value a decrease in the interest rate on the potential loan relatively more than a simple increase in the probability of obtaining a loan from Prosper. Conversely, the high–cost types (e.g., borrowers that do not have access to outside credit) would value an increase in the probability of obtaining a loan more than a decrease in the interest rate. As long as borrowers of low–cost types also tend to have higher ability to pay back loans, a separating equilibrium can be sustained in which the low–cost types have incentives to post low reserve rates (and receive low interest loans with relatively low probability) and the high–cost types have incentives to post high reserve rates (and receive high–interest loans with relatively high probability).

In order to understand the role of the reserve interest rate in this market, we begin our analysis by providing results from a series of regressions that lend support to the view that the reserve rate functions as a signal. In our first set of regressions, we examine the effect of the reserve interest rate on the funding probability and on the actual interest rate conditional on being funded. The results indicate that a lower reserve rate leads to a lower funding probability, but a lower reserve rate leads to a more favorable contract interest rate on average (conditional on receiving a loan). This implies that borrowers indeed face a trade-off between the funding probability and the interest rate in setting the reserve rate. Moreover, this is consistent with the notion that there exists heterogeneity in how borrowers evaluate this trade-off: The considerable dispersion that we observe in the reserve interest rate suggests that those who post high reserve rates care more about the probability of being funded than about what interest they will pay and vice versa.

In our second set of regressions, we examine whether there are any systematic differences between those who post high reserve rates and low reserve rates. We find evidence that those who post

high reserve rates are more likely to default than those who post low reserve rates, even *conditional on the contract interest rate* (the actual interest rate that the borrower pays on the loans). This implies that borrowers who post high reserve rates are riskier in that they are less likely to repay their loans. Taken together, our findings suggest that borrowers are heterogeneous with respect to their repayment ability and also with respect to how they evaluate the trade-off between a decrease in the interest rate and an increase in the funding probability. The results moreover suggest that both dimensions of borrower heterogeneity are closely tied to the reserve interest choice. Anticipating this relationship between borrower type and reserve rate choice, the lenders offer high interest rates for "bad" types and low interest rates for "good" types. In other words, the reserve interest rate signals the type of the borrower, and this information is being used by the lenders.

These descriptive findings motivate us to develop and estimate a structural model of the online credit market with informational asymmetry between the lenders and the borrowers. As explained above, our model of the borrowers allows for heterogeneity regarding creditworthiness and the cost of borrowing, which are privately known to the borrowers. The borrowers choose which interest rate to post, where the choice reveals their types in equilibrium. As for the supply side of the credit market, we model the lenders to be heterogeneous regarding their attitude toward risk. Each lender chooses whether to fund a loan or not, what interest rate to charge, and how much to lend. Once the loan is originated, the borrower faces monthly repayment decisions, which we model as a single–agent dynamic programming problem.

In terms of identification, the key primitives of the model that we wish to identify are the distribution of the borrowers' types and the distribution of the lenders' attitude toward risk. For identifying the borrowers' type distribution, we exploit variation in the borrower's reserve rate and how it is related to the default probability. In particular, we use the fact that the borrower's type and the borrower's reserve rate have a one-to-one mapping in a separating equilibrium. Hence, a borrower who posts a reserve rate corresponding to a particular quantile (say the α – quantile) of the reserve rate distribution can be associated with the α – quantile of the borrower's type distribution. This feature is very useful, because it allows us to condition on a particular quantile of the type distribution by simply conditioning on the reserve rate distribution. Then the observed default probability at each quantile nonparametrically identifies the borrower's type distribution. The distribution of the lenders' attitudes toward risk is also nonparametrically identified by relating

the characteristics of the borrowers to the funding probability.

In our counterfactual experiment, we compare the equilibrium market outcome and welfare under three alternative market designs – a market with signaling, a market without signaling (i.e., pooling) and a market with no information asymmetry between borrowers and lenders. In particular, we simulate the credit supply curve under each or the three market designs by re-computing the lenders' and borrowers' behavior using the estimates we obtain from our structural model. As pointed out by Stiglitz and Weiss (1981), the credit supply curve in the market with adverse selection problems may not be monotonically increasing in the interest rates, and sometimes it becomes backward bending. The results of our counterfactual support their prediction: the credit supply curve becomes more backward bending under pooling when borrowers cannot signal their type with the reserve interest rate. With respect to welfare, we find that the cost of adverse selection can be as much as 16% of the total surplus created under no asymmetric information, that is, the total surplus under pooling can be as little as 84% of the surplus under no asymmetric information. We also find that up to 95% of the difference between the surplus under pooling and no asymmetric information can be restored by signaling.

The organization of the paper is as follows: We review several related literature in the next subsection. Then in section 2, we describe the institutional background and the data we use in the estimation, and in section 3, we show some descriptive evidence of signaling in our data. We then develop our structural model of the borrowers and the lenders in section 4. In section 5, we describe identification, and in section 6, we discuss estimation. We present our results in Section 7 and demonstrates the results of the counterfactual policy experiments in section 8. Section 9 concludes.

1.1 Related Literature

Our paper is related to several strands of the literature. First, our study is related to the literature on adverse selection in credit markets. Since the seminal work of Stiglitz and Weiss (1981), there have been many studies testing for adverse selection in credit markets. Examples include Berger and Udell (1992), Ausubel (1999), Karlan and Zinman (2009), and Freedman and Jin (2010). While testing for adverse selection is important in its own right and is the first step for further analysis, estimating a model that explicitly accounts for information asymmetry among the players

allows the researcher to answer questions regarding welfare and market design.[1] Our paper goes in this direction. In particular, we compare borrower and lender welfare as well as market outcomes under three alternative market designs – a market with signaling, a market without signaling (i.e., pooling), and a market with no information asymmetry. These are questions that have not received much empirical study.

The second strand of the literature to which our paper is related is the theoretical literature on signaling. Starting with the seminal work of Spence (1973), signaling has been applied to a wide range of topics.[2] Even confined to applications in industrial organization, signaling has been applied to advertising (Kihlstrom and Riordan, 1984; Milgrom and Roberts, 1986), entry deterrence (Aghion and Bolton, 1987; Milgrom and Roberts, 1982), and war of attrition (Hörner and Sahuguet, 2011). More directly related to our paper, there is also a small theoretical literature on signaling in auctions, whereby a seller signals her private information through the reserve price (Cai, Riley and Ye, 2007, and Jullien and Mariotti, 2006, for example).

In contrast to the large body of theoretical work, however, the empirical industrial organization literature on signaling is very thin. In fact, this paper is the first structural analysis of signaling in industrial organization to the best of our knowledge. This is because identifying the effect of signaling often requires data on both the transaction and ex-post outcome, something that is hard to come by in industrial organization.[3] Consider, for example, studying whether sellers in used-car auctions can signal the quality of their cars through the reserve price. For this purpose, one would like to correlate the reserve price of the seller with the ex-post performance of the car (e.g., the maintenance cost of the car), conditional on car characteristics that are observable to the buyers. Correlation between the reserve price and the ex-post performance of the car would be indicative of signaling. However, data on the ex-post performance of cars is usually unavailable. In this sense, the data set of Prosper is ideal because it allows us to link the connection between the signal (i.e., the reserve rate) and the outcome (i.e., the default/repayment decision of the borrower). This particular data structure allows us to take a rare look at how signaling affects market outcomes

[1] See Einav, Finkelstein and Levin (2010) for a survey and motivation of recent papers that go beyond testing the existence of information asymmetry.

[2] Examples include corporate finance (Bhattacharya, 1979), political economy (Prat, 2002), and social norms (Austen-Smith and Fryer, 2005).

[3] Outside of industrial organization, there are some empirical papers that examine signaling – for example, papers on the sheepskin effect (e.g., Hungerford and Solon, 1987). However, much of the literature has tended to focus on testing for the existence of signaling (a few exceptions are Gayle and Golan, 2012, and Fang, 2006).

and to quantify the effect of signaling on key market outcomes. Another useful feature of Prosper data is that the researcher has access to almost all the information that the lenders observe. This feature mitigates the risk of confounding the effects of signaling with unobserved heterogeneity.

Our paper is also related to the large empirical literature on screening. In particular, Adams, Jenkins and Levin (2009) and Einav, Jenkins and Levin (2012) are two papers that are closely related to our paper. They consider how an auto insurer can screen borrowers using the down payment. They show that partly because of adverse selection, the lender's expected return on the loan is a non-monotone function of the loan size.[4] A key feature of our paper that is different from theirs is that we incorporate an explicit model of credit supply, which allows us to estimate the credit supply curve and how it is affected by adverse selection. Our paper also examines signaling, while they examine screening.

Finally, there are a number of papers that use data from Prosper.com. Examples include Freedman and Jin (2010), who examine adverse selection and learning; Rigbi (2011), who studies the effect of usury laws on lending; Ravina (2008), who studies the effect of posted pictures on the terms of the contract; and Iyer et al. (2010), who examine the lenders' ability to infer borrowers' creditworthiness. In Iyer et al., the authors find, among other things, that the reserve interest rate affects the contract interest that the lenders receive, and note that signaling can be one interpretation of their finding. Because the focus of their paper is on the determinants of the interest rate, the signaling story is not explored further.[5] We view our paper and their paper as complementary in that we make signaling the focus of our paper and explore the mechanism through which signaling affects the contract interest rate, lending, and repayment.

2 Institutional Background and Data

2.1 Institutional Background

Prosper.com is an online peer-to-peer lending Web site that matches borrowers with lenders and provides loan administrative services for the lenders. Established in 2006, it has become America's

[4] The issue of loan size (i.e., the borrower's request amount) is also an important aspect of our setting. In our structural model, we allow for the possibility that the borrower's unobservable type may be arbitrarily correlated with the amount choice.

[5] For example, they do not test how the reserve interest rate affects the funding probability or how it affects repayment behavior.

largest peer–to–peer lending marketplace, with more than a million members and over $280 million in loans. In this section, we describe how Prosper operates, with a particular emphasis on the auction mechanism used to allocate funds and to determine the interest rate.[6] For details on other aspects of Prosper, see Freedman and Jin (2010).

The sequence of events occurs according to the following timeline:

1. A borrower posts a listing.
2. Lenders bid.
3. Funding decision is made.
4. (If the borrower receives a loan in step 3) The borrower makes monthly loan repayments.

We explain each step in turn.

1. Borrower posts a listing A potential borrower who is interested in obtaining a loan through Prosper must first create an account by providing his social security number, driver's license number, and home address. Prosper then pulls the applicant's credit history from Experian, a third–party credit–scoring agency. If the applicant's credit score exceeds the minimum requirement, then a listing is created, which contains information regarding the borrower's characteristics, the funding option (either "closed" or "open"), the amount of loan requested, and the maximum interest rate (hereafter, *reserve interest rate*) he is willing to accept.[7] There is no fee for posting a listing.[8] The characteristics of the borrower that appear in the listing include credit grade, home–ownership status, debt–to–income ratio, purpose of the loan, as well as any other additional information (text and pictures) that the borrower wishes to post. The credit grade, which corresponds to seven distinct credit score bins (AA, A, B, C, D, E, and HR), and home–ownership status are both

[6] The auction format was used until December 19, 2010. Prosper no longer uses auctions: Instead, each listing has a "posted price," or Prosper–determined pre–set rates, which are based on the borrowers' credit risk. During our sample period (May 2008 to December 2008), the terms of the loan and the match between borrowers and lenders were determined through auctions.

[7] The listing remains active for 7 days. For "closed" listings, the listing becomes inactive after 7 days or after the listing attracts enough lenders to fund the whole loan, whichever occurs first (unless the borrower withdraws). For "open" listings, the listing remains active for 7 days even after the requested amount is fully funded. Since less than one-fourth of the listings are closed listings, we work only with open listings in our sample. We also limit our sample to listings that were not withdrawn.

[8] Prosper charges fees to both borrowers and lenders only if the loan originates. See Freedman and Jin (2010) for details.

verified by Prosper.[9] Other information, such as debt–to–income ratio and purpose of the loan, is provided by the borrower without verification by Prosper. Finally, a feature of the listing that is important for our analysis is the reserve interest rate, which the borrower can choose. The reserve interest rate is the maximum interest rate the borrower is willing to accept, and it plays a similar role to that of the reserve price in regular auctions. The requested loan amount and the reserve interest rate are both variables that the borrower chooses, subject to Prosper's conditions and state usury laws.[10]

2. Lenders Bid Prosper maintains a list of active listings on its Web site for potential lenders. Each listing contains information we described above as well as the active interest rate and the fraction of the loan funded. The active interest rate corresponds to the marginal bid in multi-unit auctions.[11] The fraction of the loan funded is just the ratio of the total amount of submitted bids to the requested loan amount. We will explain what the active interest is in more detail below.

The set of potential lenders are those who have registered with Prosper as lenders. If a lender finds a listing to which she wishes to lend money, she may then submit a bid on the listing, similar to a proxy bid in online auctions. Each bid consists of an amount that the lender is willing to lend (typically a small fraction of the loan amount that the borrower requests), and a minimum interest rate that the lender is willing to accept. The lender can submit a bid with an amount anywhere between \$50 and the borrower's requested amount, but the modal bid amount is \$50. The bidding is similar to other online auctions such as eBay auctions, in the sense that the lender can bid on any active listing at any time.

[9] A credit grade of AA corresponds to a credit score of 760+, a grade of A corresponds to 720–759, B to 680–719, C to 640–679, D to 600–639, E to 540–599, and HR to 540–. The numerical credit score is not listed. These definitions of the credit grades were used throughout our sample period, but Prosper made changes to the definitions of the credit grades in July of 2009.

[10] The minimum loan amount was \$1,000 and the maximum amount was \$25,000. Regarding the interest rate, before April 15, 2008, it was capped by the usury law of the state in which the borrower resided. After April 15, 2008, the interest rate cap was uniformly set at 36% across all states in our sample. See Rigbi (2011) for more information.

[11] For fully funded listings, the active interest rate corresponds to the marginal bid in multi-unit auctions as described in the main text. For listings that have not been fully funded, the active interest corresponds to the borrower's reserve interest rate.

3. Funding Decision The auction used in Prosper is similar to a uniform–price auction with a *public* reserve price. Using an example, we explain below how the terms of the loan are determined and which lenders end up lending. Suppose a borrower creates a listing with a requested amount of \$10,000 and a reserve interest rate of 25%. Then, Prosper adds the listing to the set of currently active listings that are displayed to potential lenders for a period of 7 days. Potential lenders who are interested in lending money can bid on the listing during this period. For simplicity, let us assume that the lenders can submit a bid amount of only \$50. At the time the lender submits her bid, she observes the fraction of the loan funded.[12] For listings that have yet to attract enough bids to reach the requested amount (i.e., less than 200 bids in this example; see left panel of Figure 1) that is all she observes. In particular, she does not observe the interest rate of each bid. As for listings that have already received enough bids to cover the requested amount, (i.e., more than 200 bids, see right panel of Figure 1) the lender observes the active interest rate, which is the interest rate of the marginal bid that brings the supply of money over the requested amount. In our example, this corresponds to the interest rate of the 200th bid when we order the submitted bids according to their interest rate, from the lowest to the highest. Moreover, for fully funded listings, the lender also observes the interest rate of the losing bids, i.e., the interest rate of the 201st bid, 202nd bid, and so on. However, the lender does not observe the interest rate of the bids below the marginal bid.

At the end of the bid submission period, listings that have attracted more bids than is necessary to fund the full requested amount are funded.[13] *However, there are no partial loans for listings that have failed to attract enough bids to fund the total requested amount.* In the first panel of Figure 1, a listing would not be funded even though \$8,000 out of \$10,000 has been funded. As for funded listings, the interest rate on the loan is determined by the marginal bid, and the same interest rate applies to all the lenders. In the second panel of Figure 1, the listing is funded at 24.8% and the same rate applies to all lenders who submitted bids below 24.8%. In this sense, the auction is similar to uniform–price auctions.

[12] More precisely, the lender also has access to the bid amount for each of the submitted bids as well as the number of submitted bids.

[13] A potential borrower may withdraw his listing at any time until the end of the bid submission period. Once the funding decision is made, however, the borrowers may not withdraw.

4. Loan Repayments All loans originated by Prosper are unsecured and have a fixed loan length of 36 months. The borrower pays both the principal and the interest in equal installments over the 36–month period.[14] If a borrower defaults, the default is reported to the credit bureaus, and a third–party collection agency is hired by Prosper to retrieve any money from the borrower. From the perspective of the borrower, defaulting on a loan originated by Prosper is just like defaulting on any other loan, resulting in a damaged credit history.

2.2 Data

The data for our analysis come directly from Prosper.com. The dataset is unique in the sense that virtually all the information available to potential lenders as well as the ex-post performance of the loans are observed to the researcher. We have data on the borrower's credit grade, debt–to–income ratio, home ownership, etc., and additional text information that borrowers provide to lenders.[15]

We retrieved the data from the Web site of Prosper.com in January of 2012. Our data consist of all listings that were created from May to December of 2008 (and the corresponding loan repayment data for funded listings which go until the end of 2011).[16] Note that all loans in our sample have either matured or have ended in default. From this sample, we drop observations that were either withdrawn by the borrower, cancelled by Prosper, or missing parts of the data.[17][18] We also dropped closed listings. We are left with a total of 35,241 listings, of which 5,571 were funded. Below, we report some summary statistics.

[14] There is no penalty for early payment: early repayments go directly into paying off the principal. If a borrower's monthly payment is more than 15 days late, a late fee is charged in addition to the principal and the interest.

[15] The only piece of information missing is the conversation that takes place between borrowers and potential lenders through the Prosper Web site.

[16] We use data from this period because there were substantial institutional differences across states before April 2008, such as interest rate caps. We use observations from April 2008 only, in order to work with a cleaner data set. Second, Prosper entered into a settlement with state securities regulators over sales of unregistered securities on December 1, 2008. As a result, Prosper stopped originating new loans until July 2009. Hence, we have no observations from December 2008 to June 2009. Lastly, Prosper made changes to the minimum bid amount from $50 to $25 and also changed its definition of the credit grades after its relaunch in July 2009. Hence we drop the observations after July 2009.

[17] About 27% of the listings that are created are later withdrawn. Most of the withdrawals occur immediately after the creation of the listing. Conditional on withdrawal, 78% are withdrawn within one day. This is probably due to some mistake the borrower found in the listing, which he subsequently wanted to correct, or to creating a listing and then deleting it just to learn how to use the system. Thus we do not think that withdrawals occur as a response to borrowers seeing an unexpectedly high interest rate just before origination. Moreover, given that submitting a lower reserve interest rate tends to lower the contract interest rate, it is suboptimal for borrowers to submit a reserve rate above the rate at which they are willing to borrow. Hence we do not think that dropping withdrawn listings creates any severe sample–selection issues.

[18] We also dropped listings registered in Texas because a different interest rate cap was used for Texas.

Summary Statistics of Listings Table 1 reports sample statistics of the listings by credit grade. The mean requested amount is reported in the first column of the table, and it ranges from a high of more than $13,000 dollars for AA listings to a low of less than $5,000 for HR listings. The average among the whole sample is $6,603. In columns 2 through 4, we report the average reserve interest rate, the debt-to-income ratio, and the home–ownership status by credit grade. In column 5, we report the bid count, which is the average number of bids submitted to a listing, and in column 6, we report the funding probability. By and large, the characteristics of the loans are related to the credit grade in the expected way.

In Figure 2, we present the distribution of the reserve rate across different credit grades. As expected, the reserve rate is higher for worse credit grades and lower for better credit grades. One important thing to note is that there is a spike at 36% for credit grades B and below. This is because 36% was the usury law maximum for our sample. In particular, note that for credit grades D and below, there is little variation in the reserve rate. As the main focus of our analysis is on the reserve rate and the extent to which it can be used as a signal of the creditworthiness of the borrower, variation in the reserve rate is crucial for our analysis. The fact that there is little variation in the reserve rate of the listings for credit grades D and below implies that listings in these categories are not very informative about the signaling value of the reserve interest rate. As a consequence we focus on the results from the top four credit grades in presenting some of our results below.

Summary Statistics of Bids In Figure 3, we report the distributions of the bid amount, again by credit grade.[19] The fraction of lenders who bid $50 exceeds 70% across all credit grades, and the fraction of lenders who bid $100 is more than 10% in all credit grades. Hence, more than 80% of lenders bid either $50 or $100. We also find that a small fraction of lenders bid $200, but rarely beyond that. These observations motivate us to formulate the potential lenders' amount choice as a discrete–choice problem in our model section, where lenders choose from {$50, $100, and $200} rather than from a continuous set.

[19] The sample of bids used to create Figure 3 consists of only the final bids for each bidder, i.e., if a potential lender bids more than once in a listing, only the last bid is used.

Summary Statistics of Loans (Funded Listings) Table 2 reports sample statistics of listings that were *funded*, which is a subset of the full set of listings. As in Table 1, we report the mean requested amount (loan amount), reserve rate, contract interest rate, debt–to–income ratio, home–ownership status, and bid count in columns 2 through 7 by credit grade. We note that the mean loan amount reported in Table 2 is smaller than the mean requested amount shown in Table 1, which is natural given that smaller listings need to attract a smaller number of bids in order to get funded. Also, note that the average bid count in Table 2 is higher than in Table 1, for the obvious reason that listings need to attract sufficient bids to get funded.

Summary Statistics of Repayments For each loan originated by Prosper, we have monthly data regarding the repayment decisions of the borrower, i.e., we observe whether the borrower repaid the loan or not every month, and whether the borrower defaulted.[20] In Table 3, we report sample statistics regarding the default probability and the timing of default conditional on default. The first column reports the default probability by credit grade. Note that all loans in our sample have either matured or have ended in default. The average default probability is lowest for AA loans at 14.9%, while it is highest for HR loans at 43.9%. The second column reports the average time until default, which is 17.5 months for the full sample. In columns 4 through 8, we report the quantiles.

Summary Statistics of Internal Rate of Return Finally, we report the internal rate of return (IRR) for the loans originated by Prosper in Table 4.[21] The average IRR for all listings is -4.6%, and it is negative in all credit grades except grade E, whose average IRR is 0%. The IRR for our sample period is generally low. These low returns may reflect the fact that default rates on loans were generally very high after the economic downturn during the financial crisis.[22]

[20] Prosper records loans that are more than 4 months late as "charge off." There are exceptions where the loans are kept on Prosper's books even after being late for 4 months. Our definition is the same as Freedman and Jin (2010).

[21] If we denote the (monthly) IRR by R, then R is the interest rate that equalizes the loan amount to the discounted sum of the stream of actual monthly repayments, i.e.,

$$\text{Loan Amount} = \sum_{t=1}^{T} \frac{t\text{-th Monthly Payment}}{(1+R)^t}.$$

In Table 4, we report the annualized IRR.

[22] There is evidence that loans originated after the end of our sample seem to be doing better. Using the subset of loans that originated right after Prosper resumed operation in 2009, we find that the average IRR was 1.1%, which is significantly higher than -4.6%. Moreover, this 1.1% estimate is conservative because some lenders had not finished repaying by the day we retrieved our data.

It may also reflect the fact that lenders were not fully aware of the creditworthiness of the pool of borrowers on Prosper.[23] We will revisit these issues when we discuss our model of lenders in Section 4.2.[24]

The standard error of IRR for grade AA is 0.283, and it monotonically increases as the credit grade becomes worse, i.e., loans to lower credit grade borrowers involves higher risk with a larger variance. Table 4 also reports the quantiles of IRR for each credit grade. Note that the median IRR is very high compared to the mean, rainging from 8.2% (grade AA listings) to 24.9% (for grade E listings).

3 Evidence of Signaling Through the Reserve Rate

In this section, we provide some evidence that the borrower's reserve interest rate serves as a signaling device. In our first set of regressions, we examine the effect of the reserve rate on the funding probability as well as its effect on the contract interest rate, conditional on being funded. In our second set of regressions, we examine whether there are any systematic differences between those who post high reserve rates and those who post low reserve rates.

Funding Probability and Contract Interest Rate In order to analyze the effect of the reserve rate on the funding probability, we run a Probit model as follows:

$$\text{Funded}_j = 1\{\beta_s s_j + \mathbf{x}_j'\boldsymbol{\beta}_x + \varepsilon_j \geq 0\}, \tag{1}$$

where Funded_j is a dummy variable for whether listing j is funded or not, s_j is the reserve rate, $\mathbf{x}_j$ is a vector of controls that include the requested amount, the debt–to–income ratio, dummy variables for home ownership, the credit grade, and other variables,[25] and ε_j is an error term following a standard Normal distribution. In alternative specifications, we included more covariates, such as calendar month, hour of day the listing was created, and other borrower characteristics. The results of these alternative specifications (contained in the Supplementary Material) are broadly consistent

[23] Freedman and Jin (2010) study lender learning where lenders learn about the creditworthiness of borrowers over time.

[24] In our estimation, we impose rational expectations of lenders (i.e., we assume that lenders' beliefs regarding the IRR and the realized IRR coincide on average) but we check the robustness of our results to alternative beliefs.

[25] Other variables include the calendar month and the hour of day the listing was created.

with the results we report below.[26]

The first column of Table 5 reports the results of this regression. The coefficient that we are interested in is the one on the reserve rate. As reported in the first row, the coefficient is estimated to be 2.13, and it is statistically significant. This implies that the higher the posted reserve interest rate, the more likely a listing is to be funded, even after controlling for observed listing characteristics. In the second row, we report the estimated coefficient on the requested amount, which is one of the control variables included in $\mathbf{x}_j$ in (1). The coefficient is negative and statistically significant, which is natural given that listings with a large requested amount must attract many lenders to get funded.[27]

Next, we run the following Tobit regression to examine the effect of the reserve rate on the contract interest rate:

$$\begin{aligned} r_j^* &= \beta_s s_j + \mathbf{x}_j'\boldsymbol{\beta}_x + \varepsilon_j, \\ r_j &= \begin{cases} r_j^* & \text{if } r_j^* \leq s_j \\ \text{missing} & \text{otherwise} \end{cases}. \end{aligned} \tag{2}$$

In this expression, r_j denotes the contract interest rate, r_j^* is the latent contract interest rate, $\mathbf{x}_j$ is the same vector of controls as before, and ε_j is a Normally distributed error term. The first equation relates the latent contract interest rate to the reserve rate and other characteristics. The second equation is the selection equation, which accounts for the fact that the contract interest rate r_j is always less than the reserve rate, s_j. The interpretation of this Tobit specification is that r_j^* is the (latent) interest rate at which the loan is funded in the absence of any censoring. Note that if we were to run an OLS regression of r_j on s_j and $\mathbf{x}_j$, the estimate of β_s would be biased upwards because the mechanical truncation effect would also be captured in β_s.[28]

[26] We also ran the specification separately for each credit grade. The results are also consistent with our findings. The estimates are available upon request.

[27] In a study of subprime lending in used–car markets, Einav, Jenkins, and Levin (2012) find that the loan size can be used to screen the borrower's unobserved type. It is possible that the negative coefficient on the request amount that we find here may reflect an underlying relationship between the requested amount and the creditworthiness of the borrower, as discussed in their paper. While we do not explicitly model the borrowers' choice regarding the request amount and how it is related to the borrowers' type in our structural analysis, our identification and estimation allow for the possibility that the borrowers' type distribution depends on the request amount in an arbitrary manner. We come back to this point below.

[28] Even if s_j had no causal effect on r_j^*, r_j and s_j will have a positive correlation because the contract interest rate is observed only if $r_j^* \leq s_j$. To see this, assume that r_j^* and s_j are independent, and consider any s_j and s_j', $(s_j < s_j')$.

We report the results from this regression in the second column of Table 5. As before, the coefficient that we are interested in is the one on the reserve rate, which measures how the reserve interest rate affects the contract interest. As reported in the first row, this coefficient is estimated to be positive and significant. Hence, posting a lower reserve interest rate leads to a lower contract interest rate conditional on the observable characteristics and net of the censoring effect, which is consistent with our hypothesis. Note that it is possible to give an alternative interpretation of these results from the perspective of the lenders. As we discuss in the next section, borrowers who post high reserve rates are relatively less creditworthy. If we take this as given, then another interpretation of the results of regression (2) is that lenders are charging higher interests to riskier borrowers, who post high reserve rates.

In addition to the Tobit model above, we also estimated a censored quantile regression model (see, e.g., Powell, 1986) using the same specification as equation (2). The quantile regression allows us to test whether a similar relationship between r_j^* and s_j that we find holds for different quantiles. The results of the quantile regressions were qualitatively similar. These results seem to imply that $F(r^*|s)$ first order stochastically dominates $F(r^*|s')$ for $s \geq s'$.[29]

The two types of regressions that we ran suggest that a borrower faces a trade-off in setting the reserve price, i.e., the borrower must trade-off the increase in the probability of acquiring a loan with the possible increase in the contract interest. Given that there exists considerable dispersion in the reserve rate, it is natural to think that there is unobserved borrower heterogeneity that induces borrowers to weigh the trade-off differently. For example, if borrowers are heterogeneous with respect to the cost of obtaining credit from outside sources, borrowers who have low cost will tend to post low reserve rates, while those who have high cost will post high reserve rates, giving rise to dispersion in the reserve rate.

Then we have

$$F_{r_j}(t|s_j) = \frac{F_{r_j}(t|s_j')}{F_{r_j}(s_j|s_j')} \text{ for } t \in [0, s_j]$$

and

$$F_{r_j}(t|s_j) = 1 > F_{r_j}(t|s_j') \text{ for } t \in [s_j, s_j']$$

where $F_{r_j}(\cdot|s_j)$, and $F_{r_j}(\cdot|s_j')$ are the conditional distributions of r_j given s_j and s_j'. This means that $F_{r_j}(\cdot|s_j')$ first order stochastically dominates $F_{r_j}(\cdot|s_j')$, which implies that r_j and s_j have a mechanical positive correlation.

[29] The results are available upon request.

Repayment Behavior and Reserve Interest Rate We now explore the extent to which borrowers who post high reserve rates are similar to or different from those who post low reserve rates in terms of their ability to pay back. In order to do so, we first run a panel Probit of an indicator variable for default on observable characteristics of the loan as well as the reserve rate:

$$\text{Default}_{jt} = 1\{\beta_s s_j + \beta_r r_j + \mathbf{x}_j'\boldsymbol{\beta}_x + \mu_t + \alpha_j + \varepsilon_{jt} \geq 0\}, \tag{3}$$

where Default_{jt} denotes a dummy variable that takes a value of 1 if borrower j defaults on the loan at period t, s_j denotes the reserve rate, r_j denotes the contract interest rate, $\mathbf{x}_j$ is a vector of control variables, μ_t is a period-t dummy, α_j is a borrower random-effect and ε_{jt} is a random error following a Normal distribution. The coefficient β_s captures the relationship between the reserve interest rate and the default probability. Note that because we control for the contract interest rate in the regression as well as other observable loan characteristics, the effect captured by β_s is purely due to selection. In other words, given that the reserve rate does not directly affect the behavior of the borrower once we condition on the contract rate, we do not need to be concerned that β_s is picking up the effect of moral hazard. Thus, β_s captures only the adverse selection effect, i.e., the heterogeneity in the creditworthiness among borrowers who posted different reserve rates.

The parameter estimates obtained from this regression are shown in the first column of Table 6. The coefficient associated with the reserve interest rate is positive and significant, with $\beta_s = 1.54$. This implies that borrowers who post higher reserve interest rates tend to default more often, which is consistent with the notion that the reserve rate reveals the type of the borrowers, i.e., the reserve interest rate can be used as a signal of the creditworthiness of the borrower. On the first column of Table 6, we also report our estimates of the coefficient on the contract interest rate and the coefficient on the requested amount. We find that both coefficients are positive and statistically significant. The positive coefficient on the contract interest rate may be capturing moral hazard – higher interest tends to increase the probability of defaulting. The positive coefficient on the amount can be a result of either adverse selection or moral hazard.[30]

[30] Borrowers who request a bigger loan may be less creditworthy, or a bigger loan may induce borrowers to default more often because of higher interest payments. The former explanation would be consistent with adverse selection, and the latter would be consistent with moral hazard. The borrower's choice of the loan size is an interesting issue, but it is hard to tease out moral hazard and adverse selection. That is one reason why our paper focuses on the borrower's choice of the reserve rate. (Note, however, that we are *not* ruling out the possibility that the loan amount can also be a signal. See sections 4.1 and 5.1 for more details.) For an analysis of the loan size and down payment in

We now wish to examine how the reserve rate relates to the borrower's repayment behavior from the perspective of the lender. In order to do so, we analyze how the IRR is related to the reserve interest rate by estimating the following model:

$$IRR_j = \beta_s s_j + \beta_r r_j + \mathbf{x}_j'\boldsymbol{\beta}_x + \varepsilon_j, \tag{4}$$

where IRR_j is the internal rate of return of loan j and $\mathbf{x}_j$ is the same vector of observable characteristics as before. The coefficient on s_j captures the relationship between the reserve rate and the IRR. As with our discussion of regression (3), the coefficient on s_j captures the selection effect.

The parameter estimates obtained from this regression are shown in the second column of Table 6. As expected, the reserve interest rate has a negative and significant effect on the IRR ($\beta_s = -0.59$), which indicates that on average, lenders make less money on loans that are made to borrowers who posted high reserve interest rates. This is consistent with the results of regression (3), where we examined the relationship between r_j and the default probability.

Finally, we run the following hazard model to examine the effect of the reserve rate on the default timing. Specifically, we estimate the Cox's proportional hazard model as follows:

$$\lambda(t_j|\,(s_j, r_j, \mathbf{x}_j)\,, \boldsymbol{\beta}) = \lambda_0(t_j)\exp(\beta_s s_j + \beta_r r_j + \mathbf{x}_j'\boldsymbol{\beta}_x), \tag{5}$$

where t_j denotes the period at which borrower j defaults, and $\lambda_0(t)$ is the baseline hazard function. As in the regressions above, we control for the same set of observable characteristics of the loan. The third column of Table 6 reports the parameter estimates of the regression. Our estimate of β_s, which captures the relationship between the hazard rate and the reserve rate, is positive and statistically significant. The results of this regression corroborate our previous findings that borrowers with higher reserve rates tend to default more often.

Interpretation of the Results Taken together, our regression results seem to indicate that (1) there is a trade-off in setting the reserve rate, i.e., a trade-off between a larger funding probability and a higher contract interest rate; (2) borrowers are heterogeneous with respect to how they

the context of subprime lending in used–car markets, see Adams, Einav, and Levin (2009) and Einav, Jenkins, and Levin (2012).

evaluate this trade-off; (3) those who post high reserve rates tend to be relatively creditworthy and those who post low reserve rates tend to be relatively less creditworthy; and (4) the lenders anticipate this and charge a higher interest to riskier borrowers who post high reserve rates. These results are informative about how signaling is sustained in equilibrium: "high cost" types, who have high cost of borrowing from outside sources are more willing to sacrifice a favorable interest rate for a bigger probability of being funded, while the opposite is true of the "low cost" types. Because borrowers who post high reserve rates default relatively more often than borrowers who post low reserve rates, "high cost" types are also less creditworthy while "low cost" types are more creditworthy. Hence borrowers who are "low cost" and creditworthy prefer (high interest, high probability of receiving a loan) to (low interest, low probability of receiving a loan), and vice versa for borrowers who are "high cost" and less creditworthy. This prevents "bad" types from mimicking "good" types and sustains separation of types through signaling.

In the supplementary material, we check the robustness of the results. We examine other specifications and add more variables. In particular, we include more variables on the borrower's credit information, such as the number of current delinquencies and the number of current credit lines. Furthermore, we control for other borrower–provided information (i.e., stated purpose of the loan, the title of the listing, description of the borrower, etc.). While the accuracy of some of these additional variables is not verified by Prosper.com, it might transmit some information about the borrower's creditworthiness.[31] Overall, we find that the results we obtain are qualitatively similar to our baseline results.

4 Model

In this section, we develop a model of the borrowers and the lenders who participate in Prosper, which we later take to the data. Our model of the credit market and the estimated primitives allow us to go beyond providing evidence of signaling and answer questions regarding welfare and market design. For example, in our counterfactuals, we compare what the credit supply curve looks like when borrowers can signal, when borrowers cannot signal (i.e., pooling), and when there is no

[31] Ravina (2008) augments the Prosper data with additional data on the perceived attractiveness of the photo of the lender. Iyer et al. (2009) use the actual credit scores in their analysis. They both report a statistically significant effect of the reserve rate on various loan outcomes (see Table IV of Ravina, 2008, and Table 5 of Iyer et al., 2009).

information asymmetry.

Our model has three parts. The first part of our model concerns the reserve interest rate choice of the borrowers. The key element of this part of the model is how borrower heterogeneity interacts with his choice of the reserve interest rate. Each borrower is heterogeneous with respect to the ease with which he can borrow money from alternative sources and also his repayment ability. This heterogeneity among the borrowers interacts with the choice of the reserve rate, because different types differentially evaluate the trade-off between lower interest and higher probability of obtaining a loan.

The second part of our model concerns the lenders' bidding behavior. The allocation and the contract interest rate are determined through an auction that is similar to a uniform–price auction. We model the lenders to be heterogeneous with regard to their attitude toward risk. The lenders decide whether to bid or not and what to bid. A bid consists of an amount and an interest rate, i.e., how much money the lender is willing to lend and at what interest rate.

The third part of our model pertains to the borrowers' repayment behavior. We model the repayment decision as a finite horizon dynamic programming problem. At each period, the borrower chooses whether to pay back the loan or default, depending on whether the disutility from paying back outweighs the disutility from default.

4.1 Borrowers

Borrower Repayment We first describe the repayment stage of the borrower's decision problem and work our way backwards. We model the repayment behavior of the borrower as a sequential decision of 36 ($= T$) months, which is the length of the loans that Prosper originates.[32] We write the terminal decision of the borrower at period T as follows:

$$\begin{cases} \text{full repayment: if } u_T(r) + \varepsilon_T \geq D(\varphi) \\ \text{default: otherwise,} \end{cases} \tag{6}$$

[32] Of all loan repayments, about 4% were "early repayments," in which the borrower paid more than 50% of the regular monthly loan repayments. We abstract from modeling early repayments because they unnecessarily complicate our model. The relevant information in the data that we use for estimating the model of the borrowers is the timing of default. We assume that the borrower has made regular repayments until default.

where $u_T(r) + \varepsilon_T$ denotes the period utility of the borrower if he repays the loan in full and r denotes the interest rate on the loan.[33] We let φ denote the (unobservable) type of the borrower, which shifts the cost of defaulting, and we let $D(\varphi)$ denote the cost of defaulting. We assume φ to be independent of ε_T, condiotional on observables. The conditional independence of ε_T and φ may be a strong assumption, but we come back to this point below. Note that we assume without loss of generality that $D(\varphi)$ is monotonically decreasing in φ, i.e., the disutility of defaulting is larger for borrowers with higher φ.[34] Hence borrowers with high φ are "good" types who value avoiding default and maintaining a good credit history.

Now let V_T denote the expected utility of the borrower at the beginning of the final period T, defined as $V_T(r, \varphi) = E[\max\{u_T(r) + \varepsilon_T, D(\varphi)\}]$. Then, the decision of the borrower at period $t < T$ is as follows:

$$\begin{cases} \text{repayment: if } u_t(r) + \varepsilon_t + \beta V_{t+1}(r, \varphi) \geq D(\varphi) \\ \text{default: otherwise,} \end{cases}$$

where $u_t(r) + \varepsilon_t$ is the period t utility of repaying the loan, β is the discount factor, and $V_{t+1}(r, \varphi)$ is the continuation utility, which can be defined recursively. We assume $\{\varepsilon_t\}$ to be independent and identically distributed across t and instead allow u_t to depend on t to capture any deterministic time dependence. We discuss below the implications of assuming that $\{\varepsilon_t\}$ are independent.

We will now make a few remarks concerning our specification. Our first remark is related to the interpretation of φ. In our specification, the unobservable type of the borrower is modeled as default cost. However, we could write an alternative model that is isomorphic to our current model, where φ has the interpretation of unobserved income/assets of the borrower.

To see this, consider the following alternative specification:

$$\text{full repayment} \Leftrightarrow \tilde{u}_T(\tilde{\varphi} + \omega_T - \text{repayment}) + \tilde{\varepsilon}_T \geq 0,$$

[33] Actually, being behind on loan payments by one month does not automatically imply that Prosper records the borrower as defaulting. The borrower is charged a late fee instead. Usually there is a three–month lag between the first missed payment until the loan is charged off by Prosper as a default. For our estimation, we defined the month of default as the first month of consecutive missed payments which subsequently result in default. While it is possible to consider an alternative model of the borrower that incorporates the number of months the borrower is behind schedule, this comes at a significant increase in computation (we need to track the number of months late as a state variable). Given that once the borrower misses a payment, the probability of ending up in default is considerable (more than 85%), we think that the results will not be affected very much by our simplifying assumption.

[34] We can always redefine φ so that $D(\varphi)$ is monotonically decreasing.

where $\tilde{\varphi} + \omega_T$ is now the (unobserved) income/assets of the borrower with a persistent component, $\tilde{\varphi}$, and a transitory component, ω_T. The problem of the borrower for $t < T$ is defined analogously. Now rearranging terms in the previous expression and using the fact that the repayment is equal to the interest multiplied by the loan amount, $(r * amt)$,[35] we obtain

$$\text{full repayment} \Leftrightarrow -(r * amt) - \tilde{u}_T^{-1}(-\tilde{\varepsilon}_T) + \omega_T \geq -\tilde{\varphi}.$$

Note that if we redefine u_T and D in equation (6) as $u_T(r) = -(r * amt)$, $\varepsilon_T = \omega_T - \tilde{u}_T^{-1}(-\tilde{\varepsilon}_T)$, and $D(\tilde{\varphi}) = -\tilde{\varphi}$, then the two specifications are equivalent. Regardless of the source of unobserved heterogeneity that affects the propensity to default – whether it be default cost, income, or some combination of the two – the resulting specification will be similar and the difference will be only in the interpretation of φ. For our purposes, the source of heterogeneity among the borrowers is not very relevant either, as borrower heterogeneity is structural to our counterfactual policy in our view. This is not to say, however, that the distinction may be very important in other contexts.

Our second remark concerns the independence assumption of ε_t and φ. While independence is a restrictive assumption, we note that mean independence of ε_t, i.e., $E[\varepsilon_t|\varphi] = 0$, is without loss of generality. This is because we can always redefine ε_t as $\varepsilon_t - E[\varepsilon_t|\varphi]$ and $D(\varphi)$ as $D(\varphi) - E[\varepsilon_t|\varphi]$, which will result in $E[\varepsilon_t|\varphi] = 0$. Of course, mean independence is not the same as independence of ε_t and φ, but it does give some credibility to the independence assumption. Relatedly, we only require independence ε_t and φ conditional on listing characteristics. Conditional independence is a weaker assumption because it allows for ε_t and φ to be correlated unconditionally. The independence assumption greatly enhances the tractability of the model.

Our third remark is related to the independence of $\{\varepsilon_t\}$ across t. Note that what we observe in the data are a sequence of binary decisions (repay or default) for each borrower, in which default is an absorbing state: If a borrower defaults, we do not observe any repayment decisions from that point on. Unlike in a situation where there are distinct decisions for each of the T periods (i.e., no absorbing state), our particular data structure precludes us from identifying possible serial correlation in $\{\varepsilon_t\}$. Only the marginals of $\{\varepsilon_t\}$ are relevant for data generation. This implies that there is a model with independent $\{\varepsilon_t\}$ that is observationally equivalent to a model with

[35] The interest rate r used to compute the repayment amount is different from the contract interest rate determined at the end of the auction because of compounding and amortisation.

serially correlated $\{\varepsilon_t\}$. While this may appear to be a limitation, there is a sense in which it is not important for our purposes. This is because the distribution of $\{\varepsilon_t\}$ is exogenous to our counterfactual policy experiments in our view.

Finally, we have presented the model up to now without making explicit the dependence of the primitives of the model on observable borrower/listing characteristics. This is purely for expositional purposes. In our identification and estimation, we let u_t, F_{ε_t}, and F_φ depend on observable characteristics. In particular, we allow F_{ε_t} and F_φ to depend on observable characteristics in an arbitrary manner in our identification. One important consequence of this is that we are allowing φ to be (arbitrarily) correlated with the requested amount. This means that we are allowing for the possibility that there are other mechanisms besides the reserve price through which the borrower can signal their type (e.g., through the requested amount), although we do not explicitly model this.

Borrower Reserve Rate Choice Next we describe our model of the borrower's reserve interest choice. When the borrower determines the reserve interest rate, s, he has to trade off the effect of s on the probability that the loan is funded, and the effect of s on the contract interest rate, r. Recall from the previous section that increasing s tends to increase the funding probability while increasing the contract interest rate. The borrower's problem is then to choose s, subject to the usury law limit, as follows:

$$\max_{s \leq 0.36} V_0(s, \varphi) = \max_{s \leq 0.36} \left[\Pr(s) \int V_1(r, \varphi) f(r|s) dr + (1 - \Pr(s)) \lambda(\varphi) \right], \tag{7}$$

where $\Pr(s)$ is the probability that the loan is funded, $f(r|s)$ is the conditional distribution of the contract interest rate given s, and $\lambda(\varphi)$ is the borrower's utility from the outside option, i.e., the borrower's utility in the event of not obtaining a loan from Prosper. We suppress the dependence of $\Pr(s)$ and $f(r|s)$ on the characteristics of the borrower, e.g., requested amount, and credit grade. Although $\Pr(s)$ and $f(r|s)$ are equilibrium objects, they are known and taken as exogenous by the borrower. The usury law maximum during the sample period was 36%, and this is reflected in the choice set of the borrower. While this restriction is rarely binding for high credit grade borrowers, it is binding for many borrowers in lower credit grades (see Figure 2).

The first term in the bracket in equation (7) captures the borrower's expected utility in the event of obtaining a loan through Prosper: $V_1(r, \varphi)$, which is the value function of the borrower at period $t = 1$ ($V_{t=1}$ – see previous subsection) when the contract interest rate is r, is integrated against the distribution of the contract interest rate $f(r|s)$. The second term in equation (7) captures the utility of the borrower in the event the loan is not funded: $(1 - \Pr(s))$ is the probability that this event occurs, which is multiplied by the utility of the outside option, $\lambda(\varphi)$. We assume that $\lambda(\varphi)$ is increasing in φ, where φ is the private type of the borrower we defined earlier, which shifts the disutility of default. This assumption simply reflects the idea that "good" types (high φ), who value their credit history, for example, have an easier time obtaining a loan from outside sources, such as relatives, friends, and local banks, etc., and hence have a high $\lambda(\varphi)$. On the other hand, "bad" types, with low cost of default, e.g., borrowers who have a damaged credit history or are expecting to default in the future anyway, are likely to have only limited alternative sources of funding, and hence have a low $\lambda(\varphi)$.

Now we make two remarks. In addition to the reserve interest rate, an important variable that the borrower needs to optimize over is the requested amount. We do not explicitly model the amount choice of the borrower and instead focus only on the reserve rate choice. It is worth noting that even when the borrower needs to optimize over the requested amount, borrowers still choose the reserve rate in accordance with equation (7). That is, regardless of the amount the borrower chooses to request, the borrower must still optimize over the reserve interest rate in accordance with equation (7). All of our model and estimation can be described conditional on the requested amount, hence the results we obtain are robust to this abstraction. Our second remark concerns our specification of λ. Our specification above imposes a deterministic relationship between the default cost, $D(\varphi)$, and the outside option, $\lambda(\varphi)$. This is not necessary, however. We can introduce an individual shock by specifying λ as $\lambda(\varphi + \eta)$, where η is independent of φ. As we discuss in our identification section, this alternative model turns out to be isomorphic to our baseline model in which λ depends just on φ.

The first–order condition associated with problem (7) is as follows,

$$\frac{\partial}{\partial s} V_0(s, \varphi) = \frac{\partial}{\partial s} \Pr(s) \left(\int V_1(r, \varphi) f(r|s) dr - \lambda(\varphi) \right) + \Pr(s) \int V_1(r, \varphi) \frac{\partial}{\partial s} f(r|s) dr = 0, \qquad (8)$$

for an interior solution. The first–order condition captures the two trade-offs that the borrower faces in determining the reserve interest. The first term is the incremental utility gain that results from an increase in the funding probability, and the second term is the incremental utility loss resulting from an increase in the contract interest rate.

Recall from the previous section that we found strong evidence that $\Pr(s)$ is increasing in s and that $F(r|s)$ first order stochastically dominates $F(r|s')$ for $s \geq s'$, where $F(r|s)$ is the conditional CDF of r. We note that under these conditions, the single crossing property (SCP) is satisfied for $s < 0.36$, i.e., $\frac{\partial}{\partial \varphi}\left[\frac{\partial}{\partial s}V_0(s,\varphi)\right] < 0$. From the perspective of the borrower, SCP is necessary and sufficient to induce separation. Hence there is no pooling among types below the usury law maximum (i.e., 36%) and pooling occurs only at the maximum. We state this as a proposition below.

Proposition 1 *If* $\frac{\partial}{\partial s}\Pr(s) > 0$ *and* $F(r|s)$ *FOSD* $F(r|s')$ *for* $s' > s$*, then we have SCP, i.e.,*

$$\frac{\partial^2}{\partial s \partial \varphi}V_0(s,\varphi) = \frac{\partial^2}{\partial s \partial \varphi}\left[\Pr(s)\int V_1(r,\varphi)f(r|s)dr + (1-\Pr(s))\lambda(\varphi)\right] < 0.$$

To see the intuition for why SCP holds, consider the marginal utility from increasing s, $\frac{\partial}{\partial s}V_0(s,\varphi)$, for a given type φ. The claim of Proposition 1 is that this marginal utility is a decreasing function of φ. As we explained in our discussion of expression (8), $\frac{\partial}{\partial s}V_0(s,\varphi)$ has two components. One is the incremental utility gain from an increase in the funding probability, and the other is the incremental utility loss resulting from an increase in the contract interest rate. The first component is decreasing in φ, because borrowers with high φ already have a high outside option – these borrowers do not appreciate the increase in the funding probability as much as low–φ types. The second component is also decreasing in φ, because borrowers with high φ are likely to bear the full cost of an increase in r, while borrowers with low φ will not – the low–φ types will default with high probability anyway.[36] A formal proof is contained in the Appendix.

Before turning to the lenders' model, we briefly discuss the optimal reserve rate choice of the borrowers when the usury law limit is binding. Note that the condition in Proposition 1, ($\Pr(s)$ is increasing and $F(r|s)$ FOSD $F(r|s')$ $(s \geq s')$) guarantees that SCP is satisfied for $s < 36\%$; that is, SCP holds for reserve rates strictly below the usury law maximum. Recall from our previous

[36] Conditional on default, the borrower does not have to bear the full cost of a high interest rate.

discussion of Figure 2 that the usury law maximum of 36% is rarely binding for borrowers with a credit grade of AA or A. This means that even though the conditions in the proposition guarantees SCP only for $s < 36\%$, these conditions are sufficient for separation of types for markets with credit grades AA and A, given that there is almost no bunching at 36% for these markets.

In contrast, recall from our previous discussion of Figure 2 that for credit grades B and C, there is a non-negligible mass at exactly 36%, implying that the usury law maximum is a binding constraint for many. For these two credit grades, the pattern in the data seem broadly consistent with partial pooling. For there to be partial pooling (i.e., pooling at 36% and separation occurring anywhere below 36%), we need an extra condition to hold (in addition to the requirements in Proposition 1) that prevents the pooled types from deviating. We describe these conditions in the Appendix. For these two credit grades, we will use them in our estimation accounting for the fact that there is separation of types below 36%, and some pooling at 36%. Finally, as for credit grades D and below, there is little variation in the reserve interest rate, which means that data from these categories are not very informative about the signaling value of the reserve rate. Hence in our estimation, we focus on the analysis only of credit grades AA, A, B, and C.

4.2 Lenders

In this subsection, we describe the model of the lenders. Let N be the (random) number of potential lenders who view a particular listing on Prosper's Web site. We let F_N denote its cumulative distribution function with support $\{0, 1, \cdots, \bar{N}\}$, where $\overline{N}$ is the maximum number of potential lenders. The potential lenders are heterogeneous with regard to their attitude toward risk and with regard to their opportunity cost of lending. Each potential lender who observes a listing on Prosper then decides whether to submit a bid or not and what to bid if she does, where a bid is an interest-amount pair. At the time of bidding, a potential lender observes the active interest rate and the interest rate of the losing bids (see section 2 for details) in addition to various characteristics of the listing, such as the reserve rate.

When the lender determines whether to bid and what to bid, the lender must first form beliefs over the return she will make if she funds part of the loan. Following the standard specification used in the asset pricing literature (see, e.g., Paravisini, Rappoport and Ravina, 2011), we assume that the lender's utility from owning an asset depends on the mean and variance of the return on

the asset. Thus, we specify the utility of lender j who holds an asset with a random return Z, with mean return $E[Z] = \mu(Z)$ and variance $Var[Z^2] = \sigma^2(Z)$, as follows:

$$U_j^L(Z) = \mu(Z) - A_j\sigma^2(Z),$$

where A_j is a lender specific random variable known only to lender j that determines her attitude toward risk. Note that if the lender holds q units of asset Z, then $E[qZ] = q\mu(Z)$ and $Var(qZ) = q^2\sigma^2(Z)$. Hence, lender j's utility of having q units of asset Z can be expressed as follows,

$$U_j^L(qZ) = q\mu(Z) - A_j(q\sigma(Z))^2.$$

In our application, $Z(r)$ denotes the (random) return from the loan when the contract interest is equal to r. Thus the lender's problem is to choose an amount $q_j \in M$ and an interest rate r_j to maximize utility, $U_j^L(q_jZ(r)) - c(q_j)$, where $c(q_j)$ is the cost of committing q_j dollars and M is the feasible set of amount choices. In principle, the lender is free to bid any amount between \$50 and the full amount requested by the borrower, but as we showed in section 2, the vast majority of the bid amounts are either \$50, \$100, or \$200. We therefore proceed with the assumption that $M \equiv \{\$50, \$100, \$200\}$ in what follows.

In order to understand the lender's problem, it is useful to illustrate it graphically. Figure 4 is a graphical representation of the lender's problem in a simplified setting without any amount choice – we first discuss this simplified version before turning to the full model with amount choice. In the left panel of this figure, the horizontal axis is σ^2 and the vertical axis is μ. Note that for each listing and each realization of the contract interest rate, r, we can assign a point on this $\mu - \sigma^2$ plane corresponding to the mean, $\mu(Z(r))$, and variance, $\sigma^2(Z(r))$, of the loan.

Now take some listing and consider mapping this listing to a point on the $\mu - \sigma^2$ plane for different realizations of r. First, suppose that the listing is funded at a contract interest rate equal to the reserve rate, so that $r = s$. At $r = s$, this listing has mean return $\mu(Z(s))$ and variance $\sigma^2(Z(s))$, and the corresponding point (labeled $r = s$) is drawn accordingly in the figure. Now consider plotting $(\mu(Z(r)),\ \sigma^2(Z(r)))$ for different values of r. In the figure, the trajectory of points, $(\mu(Z(r)),\ \sigma^2(Z(r)))$, as r changes is shown as a movement along Curve C in the direction

of the arrows. Note that as the contract rate is bid down from s, and as the corresponding point on the $\mu - \sigma^2$ plane changes, so does the utility from funding the loan. This is shown in the right panel of Figure 4, which shows how the utility of the lender changes as the contract interest rate falls from s.

Note that we have also drawn a dashed line in the left panel of Figure 4. This is the lender's indifference curve, $\{(\mu(Z(r)),\ \sigma^2(Z(r))) : U_j^L(Z(r)) - c(50) = \varepsilon_{0j}\}$, where $c(50)$ is the cost of committing \$50, and ε_{0j} denotes the outside option/opportunity cost of the lender (we will discuss more about the interpretation of ε_{0j} below). Note that this is the set of points (μ, σ^2) that make the lender just indifferent between lending and not lending. As the lender's utility function is linear with respect to μ and σ^2, the indifference curve is a straight line, i.e., $\mu - A_j\sigma^2 - c(50) = \varepsilon_{0j}$. Any point above this line gives the lender a strictly higher utility than the outside option, and vice versa. Now suppose that $(\mu(Z(r^0)),\ \sigma^2(Z(r^0)))$ is the intersection of curve C and the lender's indifference curve; that is, at contract interest $r = r^0$, the utility from lending money to the listing is exactly equal to ε_{0j}.[37] Note that in the right panel of the figure, this is reflected in the fact that $U = U_j^L(Z(r)) - c(50)$ crosses $U = \varepsilon_{0j}$ at r^0. We claim that under the assumption that the lender behaves as if she is not pivotal, there is a (weakly) dominant strategy for the lender, which is to bid r^0.

In order to see that bidding $r_j = r^0$ is a dominant strategy, suppose that the lender bids an interest rate, r_j, that is higher than r^0. If the final contract interest r turns out to be above r_j, then the lender funds a loan at r regardless of whether she bid r^0 or r_j. If the contract interest r turns out to be less than r^0, then the lender does not get to fund the loan, regardless of whether she bid r^0 or r_j. The only circumstance under which bidding r_j or r^0 makes a difference is when the final contract interest rate is between r^0 and r_j. In this case, the lender will be able to lend at a rate equal to r if she bids r^0, while she will not be able to lend if she bids r_j. Since lending at $r \in [r^0, r_j]$ gives the lender higher utility than not funding the loan, setting the rate equal to r^0 weakly dominates setting it to r_j. Likewise, it is also easily shown that submitting a bid that is lower than r^0 is weakly dominated by bidding r^0.

The reason why there exists a dominant strategy in this setting is that the auction used in

[37] Note that as drawn in the figure, Curve C intersects with the lender's indifference curve only once. When it intersects multiple times, the analysis is slightly more complicated (but not much more). Proposition 2 below is general enough to cover the case where there are multiple intersections.

Prosper has the flavor of the second price auction: As long as the lender's bid is not "pivotal" (or marginal) – a pivotal bid is a bid that brings the cumulative amount bid just over the requested amount when we order the outstanding bids by their interest rate – at the end of the bidding period, the contract interest rate is determined by the bid of someone else. Hence, under the assumption that bidders behave as if they will not be pivotal, there is a dominant strategy for lenders.

The preceding argument hinges on the assumption that lenders behave as if they will never be pivotal. We think that this is a reasonable approximation of the lenders' behavior even though there is some probability that a given lender does end up being pivotal in practice.[38] Given that the average requested amount is $\$6,603$ for all listings ($5,821 for funded listings) and that the vast majority of the lenders bid $50, a large number of bids are required to fund a single loan (on average there are about 80 winning bids; see Table 2). Hence the probability of becoming the pivotal bidder is quite low. Moreover, not only is the probability of being the pivotal bidder very low, the possible gain from submitting a bid strategically is also small – the difference between the lowest interest rate among the losing bids and the interest rate of the marginal bid is typically very small, about 0.12%. For these reasons, we assume in what follows that lenders behave as if they will not be pivotal.

Thus far, our discussion has considered the case with no amount choice for the lenders. Now consider the case with amount choice, where the borrower chooses q from the set $M = \{50, 100, 200\}$ or chooses not to bid. When the lender faces an amount choice, she needs to keep track of the utility associated with all possible actions. This is depicted in Figure 5. The three curves in the figure, Curve 50, Curve 100, and Curve 200, are defined as

$$W_q^L(r) \stackrel{DEF}{=} U_j^L(qZ(r)) - c(q) = q\mu(Z(r)) - A_j(q\sigma(Z(r)))^2 - c(q), \text{ for } q \in \{50, 100, 200\} \qquad (9)$$

where $c(q)$ is the cost of committing q dollars as before. Just as before, there is a (weakly) dominant strategy for the lender under the assumption that the bidder is not pivotal. For the case shown in

[38] The case in which the probability of being pivotal is literally zero is if bidders are restricted to bid $50 and the requested amount is in multiples of $50. If the requested amount is $50 \times M$, then M lowest bids win the auction. The interest rate, on the other hand, is determined by the $(M+1)$-th lowest bid. Hence winning bids are never marginal.

Figure 5, a (weakly) dominant strategy can be described by the following bid strategy:

$$\begin{aligned}
&\text{bid amount \$200 and interest } r' \text{ if active interest rate} \in [r', s]\\
&\text{bid amount \$100 and interest } r'' \text{ if active interest rate } \in [r'', r')\\
&\text{bid amount \$50 and interest } r''' \text{ if active interest rate } \in [r''', r'')\\
&\text{do not bid if active interest rate } \in [0, r'''),
\end{aligned}$$

where the active interest rate is understood to be equal to s if the listing has not attracted enough bids to reach the requested amount. It is easy to check that this is a dominant strategy by following the same logic we used to explain that bidding $r_j = r^0$ is the lender's dominant strategy in Figure 4. We now state the previous analysis in the form of a proposition.

Proposition 2 *Define a partition* $I_0 = [0, r_1]$, $I_1 = [r_1, r_2], \cdots$ $I_M = [r_M, s]$, *and a corresponding quantity for each interval,* $q(0)$, $q(1), \cdots$, $q(M)$, *where* $q(k) \in \{\$0, \$50, \$100, \$200\}$, *so that* $W^L_{q(k)}(r) \geq W^L_{q'}(r)$ *for all* q' *and* $r \in I_k$. *Under the assumption that the lender behaves as if she is not pivotal, it is a dominant strategy to bid* $q(k)$ *and interest rate* r_k *when the active interest rate is in* I_k.

Before proceeding to the next section, we make a few remarks about the lenders' model. First, note that the optimal strategy described above *does* require the lenders to submit new bids as the active interest rate changes. In the case depicted in Figure 5, for example, the lender would submit new bids as the active interest rate drops below r', r'', and r'''. This implicitly takes as given that lenders have low cost of revising their bid.[39]

Our second remark concerns ε_{0j}, the (random) utility associated with not investing in Prosper. Note that ε_{0j} is meant to capture the outside option of the lender: For example, ε_{0j} can be the opportunity cost of taking money away from an existing asset in the portfolio and putting it in this listing.[40] If ε_{0j} is bigger than the maximum expected utility from submitting a bid on a listing, the lender does not bid. To the extent that lenders must decide on which listings to bid from a large

[39] Some bidding strategies can be replicated with a one-time proxy bid. For example, one can submit four \$50 bids, two bids with interest rate r', and two others with r' and r'', respectively. This bidding strategy is equivalent to the dominant strategy we described for Figure 5.

[40] In our identification and estimation, we assume that ε_{0j} is i.i.d. across j conditional on a set of time dummies, i.e., we require independence of ε_{0j}, but only net of possible common macro shocks.

pool of Prosper listings, ε_{0j} can also be interpreted as a reduced form way of capturing the value of investing in other listings.[41]

Our third remark concerns how the model of the lenders ties together with our identification and estimation. The lenders in our model form beliefs over the distribution of the return on loans when they make their bidding decision. We assume rational expectations with respect to these beliefs,[42] i.e., the lenders' beliefs coincide with the realized distribution of returns – which can be backed out directly from the data. We will come back to this assumption in our estimation.

Our fourth remark also concerns the relationship between the model and identification. Note that we can identify the mean and variance of a return from funding a listing for each contract interest rate directly from the ex-post borrower repayment data. In particular, we can identify the mean and variance of a return from funding a listing at the reserve interest rate, i.e., we can identify the "starting end point" of Curve C for any listing. This means that for each distribution of A and N (the risk aversion parameter of the lender and the number of potential lenders) the lenders' bidding strategy described above will induce a probability distribution over (i) whether a listing is funded and (ii) the number of lenders who bid \$0, \$50, \$100, and \$200 for listings that are not funded.[43] In the next section, we show that this mapping from the primitives to the probability distribution over (i) and (ii) is actually a one-to-one mapping. Correspondingly, our estimation is based on matching the predicted distribution with the sample distribution.

Lastly, there are other bidding strategies that ensure the same payoffs as the strategy described in the proposition above. The strategy we described requires the least number of bid revisions necessary, but it is only one of potentially many (weakly-) dominant strategies. As we explained in the previous paragraph, our identification and estimation relies on matching (i) and (ii). As will become clearer in the next sections, other possible (weakly-) dominant strategies (beside the one we described above) can also give rise to the same distribution over (i) and (ii). Given that we

[41] To the extent that $\{\varepsilon_{0j}\}$ reflects the attractiveness of other concurrent listings that are available on Prosper, $\{\varepsilon_{0j}\}$ may very well be endogenous. However, given that Prosper maintains a fairly large list of active listings at any given point in time, we think that imposing independence of $\{\varepsilon_{0j}\}$ across listings net of macro shocks is not unreasonable. Also, the same lender typically submits multiple bids across listings, but the lenders in this market are fairly small scale. For this reason, we think that correlation in $\{\varepsilon_{0j}\}$ that comes from the lenders is fairly limited.

[42] More precisely, we assume rational expectations without common shocks.

[43] The distribution over the number of lenders who bid \$0, \$50, \$100, and \$200 for unfunded listings does not rely on any assumptions regarding the timing at which potential lenders arrive. This depends only on the distribution of N. On the other hand, the distribution over the number of lenders who bid \$0, \$50, \$100, and \$200 for funded listings may depend on the timing. Because we do not make any assumptions regarding the timing at which potential lenders arrive, we use the distribution for unfunded listings only for identification and estimation.

are using the implied distribution only over (i) and (ii) (rather than using the full implications of the particular dominant strategy described above), our estimates are robust as long as lenders are using a dominant strategy that induces the same distribution over (i) and (ii).

Equilibrium To close the model, we discuss a few issues related to equilibrium below. First, the existence of a pooling equilibrium is generally guaranteed, while the existence of a separating equilibrium is not. General sufficient conditions for the existence of a separating equilibrium are provided in Mailath (1987). While it is relatively straightforward to check whether the model satisfies the sufficient conditions in Mailath (1987) for a given parameter value, it is not easy to analytically characterize the set of parameters that satisfy these conditions. In what follows, we proceed by estimating the model assuming that the agents are playing a separating equilibrium. Once we have estimated our parameters, we then check whether the sufficient conditions for separation are satisfied at the estimated values.[44] At the estimated parameter values, the conditions seem to generally hold.

As for uniqueness of equilibrium, signaling models generally admit multiple equilibria because there are always pooling equilibria in which no information is transmitted. What we require for our identification and estimation is that the agents play the same separating equilibrium conditional on listing characteristics.[45] This may be a strong assumption if there are many equilibria. It turns out, however, that under a mild assumption on the beliefs over borrower types off the equilibrium path, there is a unique separating equilibrium (see Mailath, 1987). Given our regression results from section 3, assuming that the agents are playing a separating equilibrium is not unreasonable.

[44] The conditions identified in Mailath (1987) is a monotonicity requirement on the borrower's utility function,

$$\frac{\partial}{ds}V_0(s,\varphi,\tilde{\varphi}) \Big/ \frac{\partial}{d\tilde{\varphi}}V_0(s,\varphi,\tilde{\varphi}) \text{ is increasing in } \varphi, \forall\varphi, \forall X,$$

where $V_0(s,\varphi,\tilde{\varphi})$ is the borrower's expected utility from posting a reserve interest rate s, when the borrower is of type φ, and the lenders perceive him to be of type $\tilde{\varphi}$. The reason why we don't include this condition in our estimation routine is because we need to verify whether the monotonicity requirement is satisfied for all X. It would be computationally impossible to include this condition in the estimation routine.

[45] More precisely, we require that agents play the same equilibrium where borrowers separate below 36% and pooling occurs only at 36%.

5 Identification

5.1 Identification of the Borrower's Primitives

The primitives of the borrower that we would like to identify are the period utility function, $u_t(\cdot)$, the distribution of borrower types, $F_{\varphi|X}$, the cost of default, $D(\cdot)$, the utility from the outside option, $\lambda(\cdot)$, and the distribution of ε_t, $F_{\varepsilon_t|X}$. We specify u_t to depend on the repayment amount and a time trend as $u_t(r) = -(r \times x_{amt}) + d_t$, where d_t is a period specific constant term.

We begin with a few remarks. First, note that we allow the distribution of φ, $F_{\varphi|X}$, as well as the distribution of ε_t, $F_{\varepsilon_t|X}$, to depend on borrower characteristics, X. In particular, the distribution of φ can depend on the amount requested. To the extent that there is some signaling value in the requested amount, the conditional distribution of φ will depend on the amount requested. We are allowing for this possibility. Second, note that we can normalize $D(\cdot)$ without loss of generality.[46] Therefore, we normalize $D(\varphi) = -\varphi$.[47] It is also easy to see that we can normalize one of the constants in u_t without loss of generality: Hence we set $d_T = 0$.[48] Also, we can normalize the location of $F_{\varphi|X}$ at one point: Hence, we set $F^{-1}_{\varphi|X^*}(\alpha^*) = 0$ for some $\alpha^* \in (0,1)$ and X^*.[49]

Our identification result relies on the observation that there is a one-to-one (monotonic) mapping of s to φ conditional on X (for the case of no pooling). This means that conditioning on a quantile of s (given X) is equivalent to conditioning on a quantile of φ (given X). If we take observations (loans) in which the reserve rate is equal to the α - quantile of s ($= F^{-1}_{s|X}(\alpha)$) – note that it is possible to do so because s is observable – the borrowers all have φ equal to $F^{-1}_{\varphi|X}(\alpha)$. We use this fact extensively. For expositional simplicity, we explain the identification when the distribution of ε does not depend on X in the main text and discuss in the Appendix the case where the distribution of ε does depend on X. In the Appendix we also discuss the case when there is pooling at $s = 36\%$.

[46] This is because a specification with $\tilde{D}(\varphi) = -\varphi$, $\tilde{F}_{\varphi|X} = F_{\varphi|X} \circ D^{-1}$, and $\tilde{\lambda} = \lambda \circ D^{-1}$ is going to be observationally equivalent to one with D, $F_{\varphi|X}$, and λ. The important component of the model is the distribution of $D(\varphi)$, rather than the distribution of φ or the shape of $D(\cdot)$ per se.

[47] In the model section, we noted that a model in which the cost of the outside option depends on another random variable, as $\lambda(\varphi+\eta)$, is isomorphic to the baseline model in which the outside option is just $\lambda(\varphi)$. To see this, redefine $\tilde{\varphi} = \varphi + \eta$, and $\tilde{\varepsilon}_t = \varepsilon_t - \eta$. This formulation would now induce correlation between $\{\tilde{\varepsilon}_t\}$, but this is not a problem as we discussed in a remark when we presenting the model of borrower repayment in section 4.1.

[48] If we set $\tilde{d}_t = d_t + \kappa$ $(\forall t)$ $\tilde{\varepsilon}_t = \varepsilon_t - \kappa$ $(\forall t)$, it will be observationally equivalent to d_t, $F_{\varepsilon|X}$.

[49] Given that $D(\varphi) = -\varphi$, if we set $\tilde{\varepsilon}_t = \varepsilon_t + \kappa$ $(\forall t)$, $\tilde{F}_{\varphi|X}(h) = F_{\varphi|X}(h+\kappa)$, $\tilde{d}_T = d_T$, and $\tilde{d}_t = d_t - \beta\kappa$ $(t \in \{1, ..., T-1\})$ and $\tilde{\lambda}(\varphi) = \lambda(\varphi) + \beta\kappa$, it will be observationally equivalent to ε_t, d_t, $F_{\varphi|X}$, and λ. This normalization is convenient for proving identification, but we use an equivalent normalization (i.e., $Med[\varepsilon_t|X] = const.$) for our estimation.

We start with our discussion of how to identify the primitives from the last period, $t = T$, and work backwards. Consider the repayment decision of the borrower with $\varphi_{\alpha^*} = F^{-1}_{\varphi|X^*}(\alpha^*)$ at period $t = T$. The borrower's problem is as follows:

$$\begin{cases} \text{repay: if } -(r \times x_{amt}) + \varepsilon_T \geq -F^{-1}_{\varphi|X^*}(\alpha^*) = 0 \\ \text{default: otherwise,} \end{cases}$$

where x_{amt} is the loan size and we have replaced $u_T(r)$ with $-(r \times x_{amt})$, which is the repayment amount of the borrower. This is simply a binary threshold-crossing model; hence using variation in r, we can nonparametrically identify the distribution of ε_T, $F_{\varepsilon|X^*} = F_\varepsilon$. Once F_ε is identified, we can identify $F^{-1}_{\varphi|X}(\alpha)$ for all α and X by conditioning the sample on the α-quantile of s given X (i.e., samples with $s = F^{-1}_{s|X}(\alpha)$).[50] This is because $F^{-1}_{\varphi|X}(\alpha)$ is just a constant term in the binary threshold-crossing model where the distribution of ε_T has already been identified.

Now consider the $t = T - 1$ period problem:

$$\begin{cases} \text{repay: if } -(r \times x_{amt}) + d_{T-1} + \beta V_T(r, F^{-1}_{\varphi|X}(\alpha)) + \varepsilon_{T-1} \geq -F^{-1}_{\varphi|X}(\alpha) \\ \text{default: otherwise} \end{cases}$$

Note that $V_T(r, F^{-1}_{\varphi|X}(\alpha))$ is already identified as well as the distribution of ε_{T-1} (recall that ε_{T-1} and ε_T have the same distribution by assumption).[51] In fact, d_{T-1} and β are the only parameters that are not identified in the expression above. Hence identification of d_{T-1} and β are immediate. It should also be clear that $\{d_t\}_{t \leq T-2}$ can also be identified by looking at the borrower's period t problem and the associated default probability.

Finally, we discuss how to identify $\lambda(\varphi)$. Recall the borrower's FOC in equation (8):

$$\frac{\partial}{\partial s} \Pr(s) \left(\int V_1(r, \varphi) f(r|s) dr - \lambda(\varphi) \right) + \Pr(s) \int V_1(r, \varphi) \frac{\partial}{\partial s} f(r|s) dr = 0.$$

Solving for $\lambda(\varphi)$, we obtain

$$\lambda(\varphi) = \int V_1(r, \varphi) f(r|s) dr + \frac{\Pr(s)}{\frac{\partial}{\partial s} \Pr(s)} \int V_1(r, \varphi) \frac{\partial}{\partial s} f(r|s) dr.$$

[50] Here, we are using the fact that $\varphi \perp \varepsilon$.

[51] The identical distribution of $\{\varepsilon_t\}$ is not crucial. In fact, $\varepsilon_t + d_t$ is nonparametrically identified for each t.

Note that all the terms on the right hand side are identified. First, V_1 is identified given that F_ε, $F_{\varphi|X}$, β, and $\{d_t\}$ have already been identified. Also, we know that lenders of type φ submit a reserve rate equal to $s(\varphi) = F^{-1}_{s|X}(F_{\varphi|X}(\varphi))$. Then evaluating $\Pr(s)$ and $f(r|s)$ – which are both directly observed in the data – at $s(\varphi)$, we can identify the right-hand side of the equation. Hence the previous equation identifies $\lambda(\varphi)$.

5.2 Identification of the Lender's Primitives

The primitives of the model that we need to identify are the distribution of the coefficient of risk, F_A, the distribution of the outside option, F_{ε_0}, the cost of lending, $c(q)$, and the distribution of the number of potential bidders, F_N, which is assumed to have finite support $\{1, .., \overline{N}\}$. We first show how to identify F_A, F_{ε_0}, and $c(q)$ under the assumption that $P_q(\mu, \sigma)$, which we will define below, is known for all values of (μ, σ) and $q \in M \cup \{\$0\} \equiv \{\$0, \$50, \$100, \$200\}$. We will then show how $P_q(\mu, \sigma)$ and F_N are identified. For all $q \in M \cup \{\$0\}$, define $P_q(\mu, \sigma)$ as follows:

$$\begin{aligned} P_q(\mu,\sigma) &= \Pr\left(q\mu - A(q\sigma)^2 - c(q) \geq \max\left\{\varepsilon_0, \max_{q' \in M}\{q'\mu - A(q'\sigma)^2 - c(q')\}\right\}\right) \text{ for } q \in M \text{ and} \\ P_0(\mu,\sigma) &= \Pr(\varepsilon_0 \geq \max_{q' \in M}\{q'\mu - A(q'\sigma)^2 - c(q')\}). \end{aligned}$$

$P_q(\mu, \sigma)$ is just the probability that funding q dollars of a listing whose return is known to have mean and variance equal to μ and σ^2 gives higher utility than funding q' $(q' \neq q)$ dollars. Note that $P_q(\mu, \sigma)$ simply corresponds to the probability that (A, ε_0) lie in the region defined by the inequality inside the brackets on the right-hand side of the expression above. By varying μ and σ, this region changes. Our proposition below claims that with enough variation in μ and σ, we can recover the probability that (A, ε_0) is contained in an arbitrary set, i.e., identify F_A and F_{ε_0}.[52]

Proposition 3 *There is a one-to-one mapping from* $(F_A, F_{\varepsilon_0}, c(\cdot))$ *to* $(P_q(\mu, \sigma), P_0(\mu, \sigma))$.

Proof. See Appendix. ■

This proposition just claims that for $(F_A, F_{\varepsilon_0}, c(\cdot))$ and $(F'_A, F'_{\varepsilon_0}, c'(\cdot))$ that are different, there exists at least one (μ, σ) and $q \in M \cup \{\$0\}$ such that $P_q(\mu, \sigma)$ induced by $(F_A, F_{\varepsilon_0}, c(\cdot))$ and

[52] Our identification strategy is similar to the one taken in Cohen and Einav (2007).

$(F'_A, F'_{\varepsilon_0}, c'(\cdot))$ are different. In other words, $(F_A, F_{\varepsilon_0}, c(\cdot))$ is identified if $P_q(\mu, \sigma)$ are identified.

Proposition 4 *$P_q(\mu, \sigma)$ is identified for all q and (μ, σ) on the support of (μ, σ). F_N is also identified.*

Proof. See Appendix. ■

Here, we briefly discuss the intuition for why F_N and $P_q(\mu, \sigma)$ are identified. We use the fact that while the actual contract interest rate depends on the timing at which lenders arrive, the funding probability itself is not affected by the timing. Consider a listing Z with a requested loan amount equal to $\bar{x}_{amt}$, and whose mean return is $\bar{\mu}$ and variance is $\bar{\sigma}^2$ *if funded at the reserve interest*, s. Under the strategy described in section 4.2, a lender who is faced with a listing that have yet to receive enough bids to cover the full requested amount, bids an amount equal to q_j if and only if lender j's risk aversion parameter and the outside option, (A_j,ε_{0j}), are such that $W^L_{q_j}(s) \geq \max\{\max_{q' \in M} W^L_{q'}(s), \varepsilon_{0j}\}$, where we defined W in expression (9). Given that a listing is funded if and only if there are sufficient number of potential bidders who are willing to fund the listing when the return from the listing is evaluated at s, we can express the probability that a listing is funded, as a function of F_N, $P_{50}(\bar{\mu}, \bar{\sigma})$, $P_{100}(\bar{\mu}, \bar{\sigma})$, and $P_{200}(\bar{\mu}, \bar{\sigma})$. Since the probability that a listing is funded can be identified for all x_{amt}, μ, and σ^2, if we assume that F_N is invariant to x_{amt} and (μ, σ), sufficient variation in x_{amt} and (μ, σ) identifies both F_N and $P_q(\mu, \sigma)$.[53] The proof is in the Appendix.

Note that for our identification of the lenders' primitives, we rely on the fact that when a lender with $(A_j,\ \varepsilon_{0j})$ visits a listing that is still not fully funded, the lender submits a bid with amount q if and only if $W^L_{q_j}(s) \geq \max\{\max_{q' \in M} W^L_{q'_j}(s), \varepsilon_{0j}\}$, where $W^L_{q_j}(\cdot)$ is evaluated at the return from funding the listing at the contract interest rate, s. Note that this strategy is consistent with the dominant strategy we described in section 4.2, but it is can also consistent with other possible dominant strategies.

[53] Assuming that there is rich variation in x_{amt} is a bit problematic because the borrowers cannot request more than $\$25,000$ i.e., $x_{amt} \leq 25000$.

6 Estimation

We estimate our model in three steps. First, we estimate the conditional distribution of the contract interest rate given the reserve rate, $f(r|s)$, and the funding probability, $\Pr(s)$. We estimate these two functions nonparametrically: As $f(r|s)$ and $\Pr(s)$ are both equilibrium objects we estimate them without placing parametric assumptions. The second step involves estimating the primitives of the model of the borrower, and in the last step, we estimate the model of the lender. While our discussion of identification in the previous section focused on nonparametric identification, we place parametric functional forms for some of the model primitives in our estimation, as we will describe below.

6.1 Estimation of $f(r|s)$ and $\Pr(s)$

Our estimation proceeds first by estimating $f(r|s,x)$ and $\Pr(s,x)$, where x is a vector of observable listing characteristics such as the requested amount, debt-to-income ratio, and home ownership. Since we observe the empirical distribution of r and the funding probability, we can nonparametrically estimate these objects. Our estimation of $f(r|s,x)$ is based on Gallant and Nychka (1987), who propose a maximum likelihood estimation with Hermite series approximation.[54] Our estimation of $\Pr(s,x)$ is based on a Probit model with flexible functional forms. The details regarding the estimation are contained in the Appendix.

6.2 Estimation of the Borrower Model

We parameterize the borrower's period t utility function and outside option with parameters θ_B and denote them by $u_t(r, x_{amt}; \theta_B)$ and $\lambda(\varphi; \theta_B)$. The default cost $D(\varphi)$ is normalized as $D(\varphi) = -\varphi$ (see footnote 47).

In order to estimate θ_B, we maximize the likelihood with respect to the repayment behavior of each borrower. Note that for each θ_B, our model of the borrower's repayment behavior generates a probability distribution over sequences of repayment and default decisions for each borrower type φ. Given that we do not observe φ, we cannot use the probability distribution directly to form a likelihood. Recall, however, that there is a monotone relationship between φ and s (conditional on

[54] We use a second-order Hermite series approximation.

x), where this relationship is implicitly defined by the borrower's first-order condition (equation (8)). This means that we can back out the type of the borrower from his choice of s by using the first order condition. Once we can assign a φ for each borrower, we can then compute the likelihood.

The actual computation of the likelihood proceeds as follows: First, note that we can compute $V_1(r, \varphi, x; \theta_B)$ given θ_B. That is, for any value of $\{r, \varphi, x\}$, we can recursively solve the borrower's dynamic problem, and compute the value function, $V_1(r, \varphi, x; \theta_B)$, given θ_B. Second, we can assign a φ for each borrower from the observed reserve rate choice. Recall that the borrower's choice of the reserve rate satisfies the first-order condition;

$$\frac{\partial}{\partial s}\Pr(s,x)\left(\int V_1(r,\varphi,x;\theta_B)f(r|s,x)dr - \lambda(\varphi;\theta_B)\right) + \Pr(s,x)\int V_1(r,\varphi,x;\theta_B)\frac{\partial}{\partial s}f(r|s,x)dr = 0. \tag{10}$$

Given that we observe the reserve rate chosen by each borrower, this equation can be seen as an equation in φ. In other words, the first-order condition reveals, for each choice of s, the type of borrower φ who found it optimal to choose s. Since we have estimated $\Pr(s,x)$ and $f(r|s,x)$ in the first step, we can replace these objects with our nonparametric estimates $\widehat{\Pr}(s,x)$ and $\widehat{f}(r|s,x)$. This allows us to back out the borrower's type, $\widehat{\varphi} \equiv \widehat{\varphi}(s,x;\theta_B)$, for each borrower.[55] Note that Proposition 1 shows that the right-hand side of equation (10) is monotonic in φ, guaranteeing that a unique solution exists given s and x (for unpooled types).[56]

The third step of our procedure is to compute the likelihood for a given sequence of repayment decisions for each borrower i, using $\widehat{\varphi}_i = \widehat{\varphi}(s_i, x_i; \theta_B)$, which we obtained from the first-order

[55] We solve for $\widehat{\varphi}$ by the bisection method. More precisely, we evaluate the left hand side of expression (10) at some value of φ, $\varphi = \varphi_0$. Denote this value as $L(\varphi_0)$. If $L(\varphi_0)$ is positive, we take a larger value of φ, φ_1 ($> \varphi_0$). If $L(\varphi_0)$ is negative, we take a smaller value of φ, φ_1 ($< \varphi_0$). We continue taking larger values of φ as long as $L(\varphi)$ is positive, and similarly, we continute taking smaler values as long as $L(\varphi)$ is negative. If, at some $n \in \mathbb{N}$, the left hand side is positive at φ_n and negative at φ_{n+1} ($\varphi_n < \varphi_{n+1}$), we take φ_{n+2} to be the midpoint of φ_n and φ_{n+1}. Similarly, if the left hand side is negative at φ_n and positive at φ_{n+1} ($\varphi_n > \varphi_{n+1}$), we take φ_{n+2} to be the midpoint of φ_n and φ_{n+1}. We continue until $|\varphi_n - \varphi_{n-1}|$ is less than 10^{-5} in absolute value.

[56] In practice, there are a few borrowers (less than 10% of the sample) for whom we could not solve for $\widehat{\varphi}(s,x;\theta_B)$ even when $s < 36\%$. This would happen if the single-crossing condition is not satisfied for a given (s,x), i.e., $f(r|x,s)$ does not satisfy FOSD or $\Pr(s,x)$ is not increasing at (s,x).

In principle, Mailath (1987) gives conditions under which a separating equilibrium exists (in particular, these conditions imply that the single crossing property for the borrowers is satisfied for equation (10)). We checked whether the conditions in Mailath (1987) are satisfied at the estimated parameters: By-and-large, they seem to be. But for some values of x, the condition fails, and as a result, we cannot solve for $\widehat{\varphi}(s,x;\theta_B)$ for some borrowers (i.e., $L(\varphi)$ – defined in the previous footnote – is positive or negative for all values of φ). When we fail to solve for $\widehat{\varphi}$, we replace $\widehat{\varphi}$ with a large positive number φ^U or a large negative number φ^L. We tried two different values for (φ^U, φ^L) and the results seem to be pretty stable. The results from the different specifications are available on request.

condition. The borrower i's default probability at period t is

$$\Pr(\text{default at } t;\theta_B) = \int 1\left\{-\widehat{\varphi}_i \geq u_t(r_i, x_{i,amt};\theta_B) + d_t + \varepsilon_{it} + \beta V_{t+1}(r_i, \widehat{\varphi}_i)\right\} dF_{\varepsilon|x}. \tag{11}$$

Similarly, the probability of paying back at period t is

$$\Pr(\text{repay at } t;\theta_B) = \int 1\left\{-\widehat{\varphi}_i \leq u_t(r_i, x_{i,amt};\theta_B) + d_t + \varepsilon_{it} + \beta V_{t+1}(r_i, \widehat{\varphi}_i)\right\} dF_{\varepsilon|x}.$$

Let ι_{it} be an indicator variable that is equal to 1 if borrower i defaults at period t, and 0 otherwise. Then, the likelihood of a sequence of repayment decisions, $\{\iota_{it}\}$, is

$$l_i(\theta_B;\widehat{\varphi}_i) = \prod_{t=1}^{T_i} \Pr(\text{default at } t)^{\iota_{it}} \times \Pr(\text{repay at } t)^{(1-\iota_{it})}, \tag{12}$$

where $T_i \equiv \max\{1+\sum_{\tau=1}^{T}\iota_{i\tau}, 36\}$, i.e., the number of periods until default or 36 periods, whichever is smaller.

Finally, the likelihood is written as

$$L(\theta_B) = \prod_{i=1}^{N_L}\left[\prod_{t=1}^{T_i} \Pr(\text{default at } t)^{\iota_{it}} \times \Pr(\text{repay at } t)^{(1-\iota_{it})}\right], \tag{13}$$

where N_L is the number of loans. We obtain our parameter estimates by maximizing the likelihood function.[57]

6.3 Estimation of Lender Side

The last part of the estimation considers the model of the lender's bidding behavior. In particular, we discuss how to estimate the distribution of the number of potential bidders, F_N, the distribution of the lender's risk attitude, F_A, and the lender's cost of bidding, $c(q)$. We parameterize F_N, F_A, F_{ε_0} and $c(q)$ by θ_L, as $F_N(\cdot;\theta_L)$, $F_A(\cdot;\theta_L)$, $F_{\varepsilon_0}(\cdot;\theta_L)$ and $c(q;\theta_L)$.

[57] Up to now, our discussion focused on the case when there is no pooling among the borrowers. Note that even when there is (partial) pooling, we can obtain the same likelihood (expression (12)) for the types that are not being pooled, i.e., borrowers who submit a reserve rate below 36%. For estimating the parameters of the borrowers when there is pooling, we proceed by using just the subsample of borrowers who are not pooled. While this may not be the most efficient way of estimation, our estimates of the parameters are still consistent for all of the borrower primitives except for $F_{\varphi|X}$, for which we will not have a point estimate.

We use a (simulated) method of moments by matching the conditional funding probability and the number of bids in order to estimate θ_L. First, let fd_i be a dummy variable which equals 1 if listing i is funded, and 0 otherwise. Then $\frac{1}{I}\sum_{i=1}^{I} fd_i$ gives the (empirical) probability that a listing is funded, where I is the number of observations. Likewise, let $fd_i(\theta_L)$ ($\equiv fd(x_i, s_i; \theta_L)$) denote a random dummy variable which equals 1 if listing i is funded and 0 otherwise, given parameter θ_L. As we will explain below, $fd(x_i, s_i; \theta_L)$ can be expressed as

$$\begin{aligned} fd(x_i, s_i; \theta_L) &= 1\left\{\sum_{j=1}^{N} q_j^* \geq x_{i,amt}\right\} \text{ and} \\ q_j^* &= \underset{q_j \in M \cup \{0\}}{\arg\max} 1\{q_j \neq 0\} W_{q_j}^L(s) + 1\{q_j = 0\}\varepsilon_{0j}, \end{aligned} \tag{14}$$

where N is the (random) number of potential lenders, $W_{q_j}^L(s)$ is the utility of lending q_j dollars at interest rate s (defined in expression (9)), and $1\{q_j = 0\}$ ($1\{q_j \neq 0\}$) is an indicator function that equals one if $q_j = 0$ ($q_j \neq 0$). Taking this expression as given for now, our objective function minimizes the difference between the sample moments and the model expectation:

$$\frac{1}{I}\sum_{i=1}^{I} fd_i - E[fd_i(\theta_L)].$$

We now explain why $fd_i(\theta_L)$ can be expressed as (14). Suppose that there are $N = \tilde{N}$ potential lenders and their risk attitude and outside option are $(A_j)_{j=1}^{\tilde{N}}$ and $(\varepsilon_{0j})_{j=1}^{\tilde{N}}$. Now, consider what the optimal amount choice for each lender would be if the return from funding the listing were evaluated at *the reserve interest rate*. The optimal choice is given by the second equation in expression (14):

$$q_j^* = \underset{q_j \in M \cup \{0\}}{\arg\max} 1\{q_j \neq 0\} W_{q_j}^L(s) + 1\{q_j = 0\}\varepsilon_{0j},$$

where $W_{q_j}^L(s)$ is computed by imposing rational expectations as we explained at the end of Section 5.2.[58] Now consider the right hand side of the first equation of (14). $x_{i,amt}$ is the loan amount requested by borrower j, and $\sum_{j=1}^{N} q_j^*$ is just the sum of the lenders' bid amount. Assuming that the lenders play the strategy we described in section 4.2, a loan is funded if and only if $\sum_{j=1}^{N} q_j^*$ is bigger than the requested loan amount.

[58] We use the relized distribution of $Z(r)$ to evaluate $W_{q_j}^L(s)$.

In addition to the funding probability, we also use another set of moments to estimate the parameters of the lender's model. For each listing i, let $N_{i,q}$ denote the number of lenders who bid an amount equal to q. Then, $\left[\sum_{i=1}^{I} 1\{fd_i = 0\}\right]^{-1} \times \sum_{i=1}^{I} 1\{fd_i = 0\}N_{i,q}$ gives the expected number of lenders who bid an amount equal to q conditional on a listing being unfunded. Now consider the model counterpart, $\{fd_i(\theta_L) = 0\}N_{i,q}(\theta_L)$, where $N_{i,q}(\theta_L)$ is defined as

$$N_{i,q}(\theta_L) = \left[\sum_{j=1}^{N} 1\left\{q_j^* = q\right\}\right] \text{ (for } q \in \{50, 100, 200\}) \text{ and}$$
$$q_j^* = \arg\max_{q_j \in M \cup \{0\}} 1\{q_j \neq 0\}W_{q_j}^L(s) + 1\{q_j = 0\}\varepsilon_{0j},$$

where $W_{q_j}^L$ is evaluated at $r = s$ as before. This expression corresponds to the number of lenders who bid an amount equal to q for unfunded listings if the lenders play the strategy we described in Proposition 2. Note that this object does not depend on the timing at which the lenders arrive. Conditional on the loan being unfunded, the number of lenders who bid an amount equal to q is invariant to the timing at which the lenders visit the listing. This is in contrast to the actual contract interest rate or the number of lenders who bid an amount equal to q conditional on the listing being funded. Since we do not place any restrictions on the timing, we use the number of lenders who bid amount q only for unfunded listings.[59]

The last set of moments that we use is the fraction of listings that receive no bids. For each listing i, let nb_i denote a dummy variable that equals 1 if listing i receives no bids at all, and 0 otherwise. Let $nb_i(\theta_L)$ denote its model counterpart. Our objective function for estimating the lender's model is thus

$$\begin{aligned} Q_I(\theta_L) &= w_{1,I}\left(\frac{1}{I}\sum_{i=1}^{I} fd_i - E[fd_i(\theta_L)]\right)^2 \\ &\quad + w_{I,2}\left(\frac{1}{\sum_i^I 1\{fd_i = 0\}}\sum_{i=1}^{I} 1\{fd_i = 0\}N_{i,q} - E[1\{fd_i = 0\}N_{i,q}(\theta_L)]\right)^2 \\ &\quad + w_{I,3}\left(\frac{1}{I}\sum_{i=1}^{I} nb_i - E[nb_i(\theta_L)]\right)^2, \end{aligned}$$

where $w_{1,I}$, $w_{I,2}$, $w_{I,3}$ are weights given by the inverse of the variance of the sample moments.

[59] We need additional assumptions on the timing of lender arrival in order to obtain a distribution over the contract interest rate or the distribution over the number of lenders who bid different amounts.

Because the model expectations in the objective function are difficult to analytically compute, we use simulation in practice. We also note that while we have suppressed the conditioning variables in our exposition, we have a set of moment conditions for each conditioning variable.[60]

7 Results

The exact specification we use to estimate the model of the borrower is as follows:

$$
\begin{aligned}
&u_t(r_j;\theta_B) = -r \times x_{amt} + \theta_t \times 1\{d_t = 1\},\\
&D(\varphi) = -\varphi,\\
&F_{\varepsilon|X}(\varepsilon) = F_{\varepsilon}(\frac{\varepsilon}{\sigma_{\varepsilon}}),\\
&\lambda(\varphi, x;\theta_B) = \lambda_{x_{gr}}\varphi,\\
&F_{\varphi|X}(\varphi)
\end{aligned}
$$

where $\lambda_{x_{gr}}$ is a grade-specific constant and $\{\theta_t\}$ ($t \in \{1, 2, ..., 35\}$) are time dummies.[61] We assume that F_{ε} corresponds to the CDF of a Type I extremum value distribution and σ_{ε} is the standard error of ε. We nonparametrically estimate the distribution of φ for each credit grade. The outside option $\lambda(\varphi)$ is specified as a linear function with a credit grade specific slope. The discount factor, β, is set at $0.95^{1/12}$.

As for the lenders' side, we estimated the lender's utility function, the distribution of potential lenders, F_N, the distribution of ε_{0j}, and the costs of bidding for each amount choice, $\{c_{100}, c_{200}\}$ assuming rational expectations.[62] In our estimation, we specified F_N to follow a log normal distribution with parameters μ_N and σ_N^2. Moreover, we specified the distribution of both risk attitude

[60] For our first two moments (fd_i and $1\{fd_i = 0\}N_{i,q}$), we compute the moments for each credit grade, each quantile of the debt-to-income ratio and each quantile of the amount requested. For our last moment (nb_i), we just compute one moment for each credit grade. We then sum the moment conditions for each credit grade.

[61] In practice, we impose $\theta_t = \theta_{t+1} = \theta_{t+2}$ for $t = 3N + 1$ ($N \in \{0, ..., 11\}$) and normalize one of them, estimating 11 time dummies for each credit grade.

[62] In particular, rational expectations without a common shock. Given that the mean return from investing in Prosper was very low during the time period we study, imposing rational expectations without a common shock may not be attractive. We re-estimated the model by perturbing the beliefs of the lenders. In particular, we re-estimated the model using distributions with a 1% higher expected return and a 2% higher expected return, as well as distributions with a 5% increase in the variance and a 5% decrease in the variance. The credit supply curves we obtain by simulating the equilibrium outcome under the estimated parameters were all qualitatively similar. The estimation results are available upon request.

F_{A_j} and the distribution of ε_{0j}, F_{ε_0}, to be Normally distributed with $N(\mu_A, \sigma_A^2)$ and $N(\mu_{\varepsilon_0}, \sigma_{\varepsilon_0}^2)$.

We report the estimation results in Table 7 and Table 8 (except for the time dummies $\{\theta_t\}$, which we suppress). In the first two columns of Table 7, we report the parameter estimates of the borrowers' model and in the rest of the table, we present the estimation results of the lenders' model. In Table 8, we report the distribution of the default cost of the borrower, φ.

Recall that λ is a parameter that measures the relationship between the default cost of the borrower (φ) and the cost of borrowing from outside sources ($\lambda\varphi$). Our estimates for λ, reported in Table 7, indicate that λ is smaller for high credit grades ($\lambda_{AA} = 3.1$ and $\lambda_A = 4.3$) and becomes larger for low credit grades ($\lambda_B = 9.5$ and $\lambda_C = 8.8$). Given that the interquartile range of the distribution of φ are 3.0 for credit grade AA and 1.8 for credit grade A, the interquartile range of the outside option are about $\$9,300$ for credit grade AA and $\$7,900$ for credit grade A. The interquartile range of the outside option are $\$10,100$ and $\$8,600$ for credit grade B and C, respectively. In Table 8, we report the quantile of the default cost (φ) for each credit grade. The median default costs are estimated to be around $\$9,400$ for credit grade AA, $\$6,800$ for credit grade A, $\$2,800$ for credit grade B and $\$3,000$ for credit grade C.[63]

Columns 3 through 7 of Table 7 report parameter estimates for the lenders. We estimated a log Normal distribution for the number of potential bidders. The parameter estimates reported in the table translate to a mean number of potential lenders of about 127.1, 92.1, 142.3, and 55.8 for each of the four credit grades. Our estimates of the lenders' risk aversion parameter range from 1.91×10^{-2} to 3.67×10^{-2}. One way to interpret these numbers is to consider a lottery that yields \$10 with probability 0.5 and yields $-\$x$ with probability 0.5 and ask, "At what amount is the lender indifferent between participating in the lottery and not participating?" It turns out that x equals \$9.63, \$9.57, \$9.32, \$9.35 for each of the credit grades.

Finally, in order to understand how the type of the borrower is correlated with observable characteristics, we regressed the estimated φ on observable characteristics using a Tobit.[64] The results are reported in Table 9. We find that the coefficient on log requested amount is negative, which means that a borrower who requests a larger loan amount tends to be of a worse type. This is consistent with evidence found in Adams, Einav, and Levin (2009), in which the authors find

[63] For borrowers who posted a reserve rate equal to 36%, we do not have a point estimate of their types. The quantiles are not affected by this however.

[64] We use a Tobit because we do not have point estimates for borrowers who submitted a reserve rate equal to 36%.

evidence of bad types selecting into larger loans. Our estimated coefficients show that the debt-to-income ratio is negatively correlated with φ and that the home ownership is positively correlated with φ.

8 Counterfactual Experiment

In our counterfactual experiment, we compare the equilibrium market outcome and welfare under three alternative market designs – a market with signaling, a market without signaling (i.e., pooling) and a market with no information asymmetry between borrowers and lenders. This counterfactual is interesting because it allows us to empirically quantify the extent to which credit markets suffer from adverse selection and the extent to which signaling can improve market conditions. In particular, the question of how adverse selection affects credit supply goes back to Stiglitz and Weiss (1981) but few empirical attempts that study the effect have been made.[65]

In Figure 6, we present the credit supply curve we estimated for each of the four credit grades. The horizontal axis in the Figure corresponds to the average supply of credit and the vertical axis corresponds to the interest rate. The scale of the horizontal axis is different for each of the four panels reflecting the fact that the amount of credit supplied varies considerably from credit grade to credit grade. The thick dotted curve in each of the panels represents the credit supply curve under no signaling (i.e., pooling). The credit supply curve under pooling is computed for the counterfactual scenario in which borrowers submit a *secret* reserve rate. Under this counterfactual scenario, each borrower posts a secret reserve rate, and if the contract interest at the end of the bid closing period is less than the secret reserve rate, the borrower takes out a loan. This market design would induce pooling of types, i.e., at a given interest rate, r, there would be a mix of different borrowers who take out the loan, and the lenders have no way of differentiating among them. More precisely, the mix of borrowers who take out the loan at r are borrowers whose outside options are below a certain threshold, i.e., borrowers whose type is low enough that he would rather borrow at r than not borrow at all. Note that under this market design, it is a dominant strategy for each borrower to submit a (secret) reserve rate equal to the interest rate at which the borrower is

[65] We treat F_N, the distribution of the number of lenders as exogenous in simulating our counterfactual. We also take F_φ, the borrower's type distribution, as exogenous. We acknowledge that these are potential limitations of our counterfactual experiment.

indifferent between borrowing and not borrowing. (For a more detailed computational procedure for obtaining the credit supply curve under pooling as well as the supply curve under signaling and under no asymmetric information, see Appendix 10.4)

In Figure 6, the solid curve corresponds to the credit supply curve under signaling and the dotted line that lies on top of it corresponds to the supply curve under no asymmetric information. Note that these two supply curves correspond to the supply for a fixed φ: In the Figure, the type is set to the type of the median borrower.[66] The credit supply curve under signaling corresponds to the average amount of credit that potential lenders are willing to supply under the mechanism used by Prosper, because we assume that the data is generated from a signaling equilibrium. Note also that the supply curve for signaling is truncated above – at the reserve interest rate. Under the signaling equilibrium, the borrower does not have access to credit above the reserve rate.

The credit supply under no asymmetric information is computed under the counterfactual scenario in which the lenders have perfect knowledge of φ. The dotted line on top of the solid curve is the credit supply curve of the median borrower under no asymmetric information. This credit supply curve extends beyond the reserve rate, but it is otherwise the same as the one for signaling. This is because under both the signaling equilibrium and no asymmetric information, the lenders have perfect knowledge of borrower type in equilibrium: The lenders know that they are lending to the borrower with a particular type. Hence, the two credit supply curves partly coincide. The only difference between the two is that under no asymmetric information, the borrowers can borrow at rates that are higher than the reserve rate that they would post under the signaling equilibrium. This means that the credit supply curve for no asymmetric information extends beyond the reserve rate all the way until the point at which the borrower is indifferent between borrowing and not borrowing. The truncated supply curve under the signaling equilibrium can be viewed as capturing the cost that borrowers must pay (or the surplus that has to be burned) in order to differentiate himself from lower types.

The Figure makes clear the role of adverse selection and moral hazard in credit markets. First, note that the credit supply curve under signaling and under no asymmetric information for grades B and C are backward bending. This is a result of moral hazard. As borrowers are charged a

[66] The two credit supply curves reported in the Figure (Supply curve for signaling and no asymmetric information) correspond to the supply for a fixed listing characteristic, $x_{amt} = 10,000$, $x_{dti} = 0.2$, $x_{ho} = 1$. The type of the borrower corresponds to the type of the borrower who posts a median reserve interest rate.

higher interest, it increases the likelihood that they will default. Above a certain interest rate, the marginal increase in revenue from a higher interest rate is overwhelmed by the loss from increased incidence of default. As a result, the supply of credit starts to decrease at a certain point.

On the other hand, the shape of the supply curves under pooling reflects both moral hazard and adverse selection. Both adverse selection and moral hazard combine to suppress the supply of credit at higher interest rates. The borrowers who are willing to take out a loan at high interest rates tend to be of low types who are likely to default to begin with. Moreover, the borrowers that take out the loan are likely to default because of high interest. This is the reason why the supply curves for pooling start to bend backwards sooner (i.e., at lower interest rates) than the supply curves under signaling and under no asymmetric information. While we have abstracted from the demand of credit, if the demand for credit is large enough, there could be credit rationing, as demonstrated by Stiglitz and Weiss (1981).

The Figure is also informative about the severity of adverse selection for different credit grades. There are substantial differences between the credit supply under pooling and the supply curve under no asymmetric information for credit grades B and C. This is indicative that adverse selection in these credit categories is relatively more severe. This is broadly consistent with the findings in Iyer et al. (2010) where they find that conditional on the credit grade, the borrowers' credit score had a statistically significant effect on the default rate in grade C, but not in higher credit grades.[67]

Finally, we examine the welfare implications of signaling and information asymmetry. In Table 10, we report the expected surplus of the lenders and the borrowers per listing for each of the three different market designs we consider. We compute the expected and the median surplus by simulating the model using the estimates we obtained from our structural model. Details are discussed in the supplementary material. Listing characteristics such as the amount, debt-to-income ratio, and home ownership are set to their median values, as before. In the first three columns, we report the expected surplus averaged over the borrower's type distribution. In the last three columns, we report the expected surplus that corresponds to a listing with a median borrower type. First, consider the welfare of the borrowers reported in the first column. Comparing the borrower welfare under pooling and under no asymmetric information, we find that the welfare loss from information asymmetry is relatively modest in credit grade A ($142.7 under pooling and $143.3

[67] See Table 2, (p.34) of Iyer et al. (2010).

under no asymmetric information) while the welfare loss is relatively more severe in credit grades B and C ($343.1 under pooling $398.4 under no asymmetric information for credit grade B, and $374.7 under pooling and $425.5 under no asymmetric information for credit grade C), confirming our earlier remark that adverse selection is more serious in credit grades B and C. Comparing these numbers to the welfare results under signaling, we find that welfare improves relative to pooling in all credit grades except for credit grade AA. In particular, for credit grades B and C, signaling restores most of the welfare loss caused by adverse selection. For credit grade AA, the borrower welfare under pooling is higher than the borrower welfare under signaling. This happens because the surplus that must be burned (i.e., the transactions that must be foregone by submitting a lower reserve interest rate under signaling) in order to maintain a separating equilibrium is sufficiently costly. This off-sets any benefits gained by reducing information asymmetry between the lenders and the borrowers. Note that in general, it is not possible to Pareto-rank equilibrium under pooling and signaling.

Second, consider the welfare of the lenders reported in the second column. Comparing the welfare of the lenders under pooling and under no asymmetric information, we find that welfare decreases considerably under pooling in all credit grades except for credit grade A, where welfare of the lenders is slightly higher under pooling. Similar to what we found for the case of borrowers, we find that signaling improves welfare in credit grades B and C, but not in credit grade AA (and A). Again, the reason for this is that for credit grade AA, there is a net decrease in the listings that are funded as a result of low reserve rates. This is in contrast to credit grades B and C where the increased credit supply from reducing information asymmetry outweighs the reduction in transactions that result from lower reserve interest rates.

Finally, the third column of Table 10 is informative about the cost of adverse selection, as well as the extent to which welfare can be restored through signaling. Comparing the total surplus under pooling and no asymmetric information, we find that the cost of adverse selection can be quite large, with a 16% ($157.5) decrease in total surplus for credit grade B and a 13% ($124.2) decrease in total surplus for credit grade C. We also find that in some instances, signaling can restore a large fraction of the potential welfare loss from adverse selection, with 95% and 87% of the welfare loss avoided through signaling in credit grades B and C.

9 Conclusion

In this paper, we study how signaling can restore some of the inefficiencies arising from adverse selection using the data from an online peer-to-peer lending market, Prosper.com. We first provide some evidence showing that the reserve interest rate posted by potential borrowers work as a signaling device. Based on this evidence, we then develop and estimate a structural model of borrowers and lenders, where low reserve interest rate can credibly signal low default risk. In our counterfactual, we compare the credit supply curve and welfare under three different market designs: a market with signaling, a market without signaling, and a market with no asymmetric information. We find that in one of the credit grades, signaling exacerbates the welfare cost of adverse selection, but we also find that signaling can restore much of the welfare losses that result from adverse selection in other credit grades.

Our paper is the first structural analysis of signaling in industrial organization to the best of our knowledge, and it is also the first attempt at estimating the credit supply curve, as far as we are aware. We also believe that the methods developed in the paper can be applied to other settings in which signaling is important (e.g., auctions and reservation price). For future research, we think that it is important to study other types of credit markets in order to understand more fully the costs of adverse selection and the benefits of signaling. Given that our analysis in this paper is limited to a particular P2P lending platform, we think that studying other credit markets is important for a general understanding of credit supply and the role of signaling.

Stern School of Business, New York University
Northwestern University
Northwestern University

A Appendix

A.1 Proof of Proposition 1

We provide a proof of Proposition 1. We do so by first proving the following lemma.

Lemma 1 $\frac{\partial}{\partial \varphi} V_1(r, \varphi)$ *is non-increasing in* r.

Proof. The proof is by induction. We first show that $\frac{\partial}{\partial r\partial\varphi}V_T(r,\varphi) \leq 0$, $\frac{\partial}{\partial r}V_T(r,\varphi) \leq 0$, and $D'(\varphi) \leq \beta\frac{\partial}{\partial\varphi}V_T(r,\varphi) < 0$. We then show that if $\frac{\partial}{\partial r\partial\varphi}V_\tau(r,\varphi) \leq 0$ and $D'(\varphi) \leq \frac{\partial}{\partial\varphi}V_\tau(r,\varphi) < 0$ hold for some $\tau \leq T$, then the same conditions hold for $\tau - 1$. First, for $t = T$,

$$\begin{aligned}
\frac{\partial}{\partial\varphi}V_T(r,\varphi) &= \frac{\partial}{\partial\varphi}\int \max\{u_T(r)+\varepsilon_T, D(\varphi)\}dF_{\varepsilon_T}(\varepsilon_T) \\
&= \int \left(0\times 1\{u_T(r)+\varepsilon_T \geq D(\varphi)\} + D'(\varphi)1\{u_T(r)+\varepsilon_T < D(\varphi)\}\right) dF_{\varepsilon_T}(\varepsilon_T) \\
&= D'(\varphi)\Pr_T(r,\varphi),
\end{aligned}$$

where $1\{\cdot\}$ is an indicator function and $\Pr_T(r,\varphi) = \Pr(u_T(r)+\varepsilon_T < D(\varphi))$. It is easy to see that $D'(\varphi) \leq \frac{\partial}{\partial\varphi}V_T(r,\varphi) < \beta\frac{\partial}{\partial\varphi}V_T(r,\varphi) < 0$ because $D'(\varphi) < 0$, by assumption and $\Pr(u_T(r)+\varepsilon_T < D(\varphi)) \in (0,1)$. Also, note that $\frac{\partial}{\partial r}u_T(r) < 0$ implies $\frac{\partial}{\partial r}\Pr(u_T(r)+\varepsilon_T < D(\varphi)) > 0$, which means that $\frac{\partial}{\partial r\partial\varphi}V_T(r,\varphi) \leq 0$. It is also easy to see that $\frac{\partial}{\partial r}V_T(r,\varphi) < 0$.

Now, assume $\frac{\partial}{\partial r\partial\varphi}V_{t+1}(r,\varphi) \leq 0$, $\frac{\partial}{\partial r}V_{t+1}(r,\varphi) \leq 0$, and $D'(\varphi) \leq \beta\frac{\partial}{\partial\varphi}V_{t+1}(r,\varphi) < 0$ for some t. Then,

$$\begin{aligned}
\frac{\partial}{\partial\varphi}V_t(r,\varphi) &= \frac{\partial}{\partial\varphi}\int \max\{u_t(r)+\varepsilon_t+\beta V_{t+1}(r,\varphi), D(\varphi)\}dF_{\varepsilon_t}(\varepsilon_t) \\
&= \int \left(\begin{array}{c} \frac{\partial}{\partial\varphi}\beta V_{t+1}(r,\varphi)1\{u_t(r)+\varepsilon_t+\beta V_{t+1}(r,\varphi) \geq D(\varphi)\}+ \\ D'(\varphi)1\{u_t(r)+\varepsilon_t+\beta V_{t+1}(r,\varphi) < D(\varphi)\} \end{array}\right) dF_{\varepsilon_t}(\varepsilon_t) \\
&= \begin{array}{c} \frac{\partial}{\partial\varphi}\beta V_{t+1}(r,\varphi)(1-\Pr_t(r,\varphi))+ \\ D'(\varphi)\Pr_t(r,\varphi) \end{array} \\
&\geq D'(\varphi),
\end{aligned}$$

where $\Pr_t(r,\varphi) = \Pr(u_t(r)+\varepsilon_t+\beta V_{t+1}(r,\varphi) < D(\varphi))$. The last inequality holds since $\frac{\partial}{\partial\varphi}V_{t+1}(r,\varphi) \geq D'(\varphi)$. Again, it is easy to see $\frac{\partial}{\partial\varphi}V_t(r,\varphi) < 0$, and $\frac{\partial}{\partial r}V_t(r,\varphi) \leq 0$. To see that $\frac{\partial}{\partial r\partial\varphi}V_t(r,\varphi) \leq 0$,

note that

$$\begin{aligned}
&\frac{\partial}{\partial r \partial \varphi} V_t(r,\varphi) \\
= \; &\frac{\partial}{\partial r}\left[\begin{array}{c} \frac{\partial}{\partial \varphi}\beta V_{t+1}(r,\varphi)(1-\Pr_t(r,\varphi))+ \\ D'(\varphi)\Pr_t(r,\varphi) \end{array}\right] \\
= \; &\left[\begin{array}{c} \frac{\partial^2}{\partial r \partial \varphi}\beta V_{t+1}(r,\varphi)(1-\Pr_t(r,\varphi))+ \\ \frac{\partial}{\partial r}\Pr_t(r,\varphi) \times (D'(\varphi) - \frac{\partial}{\partial \varphi}\beta V_{t+1}(r,\varphi)) \end{array}\right] \leq 0.
\end{aligned}$$

By induction we conclude that $\frac{\partial}{\partial r \partial \varphi} V_1(r,\varphi) \leq 0$. ■

Proposition 1 *If $\frac{\partial}{\partial s}\Pr(s) > 0$ and $F(r|s)$ FOSD $F(r|s')$ for $s' > s$, then we have SCP, i.e.,*

$$\frac{\partial^2}{\partial s \partial \varphi} V_0(s,\varphi) = \frac{\partial^2}{\partial s \partial \varphi}\left[\Pr(s)\int V_1(r,\varphi) f(r|s) dr + (1-\Pr(s))\lambda(\varphi)\right] < 0.$$

Proof. First, let us consider the second term. Note that $\frac{\partial^2}{\partial s \partial \varphi}(1-\Pr(s))\lambda(\varphi) = -\Pr'(s)\lambda'(\varphi) < 0$. This is because $\Pr'(s) > 0$ and $\lambda'(\varphi) > 0$ by assumption. Second, we consider the first term. Note that for $s_0 < s_1$, $F(r|s_1)$ first-order stochastically dominates $F(r|s_0)$. Hence if $\frac{\partial}{\partial \varphi} V_1(r,\varphi)$ is non-increasing in r, then $\int \frac{\partial}{\partial \varphi} V_1(r,\varphi) dF(r|s_0) \geq \int \frac{\partial}{\partial \varphi} V_1(r,\varphi) dF(r|s_1)$ for any s_0 and s_1 s.t. $s_0 < s_1$. This implies that $\frac{\partial}{\partial s \partial \varphi}\Pr(s)\int V_1(r,\varphi) dF(r|s) \leq 0$. Thus, we complete the proof. ■

A.2 Identification of the Model of the Borrowers When F_ε depends on X

We discuss the identification of the borrower's model when F_ε depends on X. The only difference between the case in which F_ε depends on X and the case described in the main text is that $F_{\varepsilon|X^*} \neq F_{\varepsilon|X}$ in the former. Hence even when F_ε depends on X, it is easy to see that $F_{\varphi|X^*}$ and $F_{\varepsilon|X^*}$ are both identified, following the same argument as the proof in Section 4.1. Also, β, d_t and $\lambda(\varphi)$ can also be identified just from the observations with $X = X^*$. To see that $F_{\varphi|X}$ and $F_{\varepsilon|X}$ are identified for any X, note that it is enough to identify $F_{\varphi|X}(0)$ – if $F_{\varphi|X}(0)$ is identified, we can follow the proof in Section 4.1 to identify $F_{\varphi|X}$ and $F_{\varepsilon|X}$. Note that if we set $\tilde{F}_{\varepsilon|X}(h) = F_{\varepsilon|X}(h-\kappa)$, $\tilde{F}_{\varphi|X}(h) = F_{\varphi|X}(h+\kappa)$, $\tilde{d}_T = d_T$, $\tilde{d}_t = d_t - \beta\kappa$ $(t < T)$, and $\tilde{\lambda}(\varphi) = \lambda(\varphi) + \beta\kappa$, $(\tilde{F}_{\varepsilon|X}, \tilde{F}_{\varphi|X}, \tilde{\lambda}, \tilde{d}_t)$ will be observationally equivalent to $(F_{\varepsilon|X}, F_{\varphi|X}, \lambda, d_t)$. Moreover, it should be easy to see that for each given profile, $(F^*_{\varepsilon|X}, F^*_{\varphi|X}, \lambda^*, d^*_t)$, the set of profiles that are observationally equivalent to

$(F^*_{\varepsilon|X}, F^*_{\varphi|X}, \lambda^*, d^*_t)$ are given by $\{(F_{\varepsilon|X}, F_{\varphi|X}, \lambda, d_t) : F_{\varepsilon|X}(h) = F^*_{\varepsilon|X}(h-\kappa),\ F_{\varphi|X}(h) = F^*_{\varphi|X}(h+\kappa),$ $d_T = d^*_T,\ d_t = d^*_t - \beta\kappa\ (t < T),\ \lambda(\varphi) = \lambda^*(\varphi) + \beta\kappa\}$. Given that we have already identified $\lambda(\varphi)$, we can identify $F_{\varphi|X}(0)$.

A.3 Identification of the Model of Lenders

In this section, we prove identification of the primitives of the model of the lenders. First we provide a proof of Proposition 3.

Identification of $c(\cdot)$ Recall from section 5.2 that $P_q(\mu, \sigma)$ corresponds to the probability that (A, ε_0) falls into a region defined by inequalities. Fix a particular value of ε_0, and $P_q(\mu, \sigma)$ can be considered as defining a region for A. The region of A that corresponds to $P_q(\mu, \sigma)$ is defined by the intersection of straight lines $U_q = U_q(A) \equiv q\mu - A(q\sigma)^2 - c(q)$ for $q =$ \$50, \$100, and \$200 (U_{50}, U_{100}, U_{200}). Figure 1 illustrates this for the case of $\frac{100\mu - c(200) + c(100)}{30000\sigma^2} < \frac{c(50) - c(100) + 50\mu}{7500\sigma^2}$ ($\Leftrightarrow \mu > \frac{-c(200) + 5c(100) - 4c(50)}{100}$) (which ensures that the intersection between U_{200} and U_{100} is to the left of the intersection between U_{100} and U_{50}).

Note first that it is possible to assume $c(50) = 0$ without loss of generality.[68] We also assume that $c(200) > 3c(100)$ for our proof below. This restriction is just for exposition: Identification for $c(200) < 3c(100)$ can be shown analogously. Now, consider $P_{200}(\mu, \sigma)$. Given μ and σ, bidding \$200 is optimal if the risk parameter A_j is sufficiently small and the outside option ε_0 is also sufficiently small. Hence, $P_{200}(\mu, \sigma)$ can be expressed as follows,

$$\begin{aligned} P_{200}(\mu, \sigma) &= \Pr(\{U_{200}(\mu, \sigma) > \max\{U_{50}(\mu, \sigma),\ U_{100}(\mu, \sigma)\}\} \cap \{\varepsilon_0 < U_{200}(\mu, \sigma)\}) \\ &= \Pr(A_j < \overline{A}(\mu, \sigma) \wedge \varepsilon_0 < 200\mu - A_j(200\sigma)^2 - c(200)), \end{aligned}$$

where $\overline{A}(\mu, \sigma) = \frac{c(100) - c(200) + 100\mu}{30000\sigma^2}$.[69]

Observe that if $\overline{A}(\mu, \sigma) < 0$, then as we let $\sigma \to 0$ (while keeping μ fixed), $P_{200}(\mu, \sigma)$ would tend to 0.[70] However, if $\overline{A}(\mu, \sigma) = 0$, then as $\sigma \to 0$, $P_{200}(\mu, \sigma)$ would converge to a positive number,

[68] We can add a constant to $c(50)$, $c(100)$, $c(200)$ and shift the distribution of ε to the right without changing the distribution of outcomes.

[69] This is true as long as μ is "big" enough, i.e., $\frac{-c(100) + 50\mu}{7500\sigma^2} > \frac{100\mu - c(200) + c(100)}{30000\sigma^2}$ ($\Leftrightarrow \mu > \frac{-c(200) + 5c(100)}{100}$).

[70] This is because $\overline{A}(\mu, \sigma)$ $\left(= \frac{c(100) - c(200) + 100\mu}{30000\sigma^2}\right)$ tends to $-\infty$ as $\sigma \to 0$ (while keeping μ fixed).

i.e.,

$$\lim_{\sigma \to 0} P_{200}(\mu, \sigma) = \Pr(A_j < 0 \wedge \varepsilon_0 < c(200) - 2c(200)),$$

where we have used the fact $\overline{A}(\mu, \sigma) = 0 \Longleftrightarrow c(100) - c(200) + 100\mu = 0$ and $A_j(200\sigma)^2 \to 0$.[71] Let us define μ^* as

$$\mu^* = \sup_{\mu} \{ \lim_{\sigma \to 0} P_{200}(\mu, \sigma) = 0 \}.$$

Then, μ^* is identified because everything in the right hand side of this expression is identified. Since μ^* solves $c(100) - c(200) + 100\mu^* = 0$, we can identify $c(100) - c(200)$. Similarly, working with the intersection between U_{50} and U_{100}, we can identify $c(100)$.

Identification of F_A Now we consider identification of F_A, given that $c(\cdot)$ has already been identified. Again note that

$$P_{200}(\mu, \sigma) = \Pr(A_j < \overline{A}(\mu, \sigma) \wedge \varepsilon_0 < 200\mu - A_j(200\sigma)^2 - c(200)).$$

Now take μ and σ so that $\overline{A}(\mu, \sigma) = \delta^+$, or equivalently, $\mu = \frac{c(200) - c(100) + 30000\sigma^2\delta^+}{100}$, where δ^+ is some positive number.[72] Then consider keeping $\overline{A}(\mu, \sigma)$ fixed at δ^+, but moving $200\mu - A_j(200\sigma)^2 - c(200)$ by changing both μ and σ. In particular, as $\sigma \to 0$, we have

$$\begin{aligned} \mu &\to \frac{c(200) - c(100)}{100} \text{ and} \\ P_{200}(\mu, \sigma) &\to \Pr(A_j < \delta^+ \wedge \varepsilon_0 < c(200) - 2c(100)) \\ &= \Pr(A_j < \delta^+) \Pr(\varepsilon_0 < c(200) - 2c(100)), \end{aligned}$$

where we have used the independence assumption between A_j and ε_0 for going from the second line to the third line. By varying δ^+ (> 0), we can identify $\Pr(A_j < t) \Pr(\varepsilon_0 < c(200) - 2c(100))$ for all $t > 0$. Similarly, by taking μ and σ such that $\overline{A}(\mu, \sigma) = \delta^-$ for some negative constant, we

[71] Recall that the expression for P_{200} takes the form in the text only if $\mu > \frac{-c(200) + 5c(100)}{100}$. Hence implicitly, we are assuming that the value of μ which solves $\frac{c(100) - c(200) + 100\mu}{30000\sigma^2} = 0$ ($\Leftrightarrow \mu = \frac{c(200) - c(100)}{100}$) satisfies this restriction, i.e. $\frac{c(200) - c(100)}{100} > \frac{-c(200) + 5c(100)}{100} \Leftrightarrow c(200) > 3c(100)$.

[72] As before, we need μ to satisfy $\mu > \frac{-c(200) + 5c(100)}{100}$. This means that $\frac{c(200) - c(100) + 30000\sigma^2\delta^+}{100} > \frac{-c(200) + 5c(100)}{100} \Leftrightarrow c(200) > 3c(100) - 15000\sigma^2\delta^+$. If $c(200) > 3c(100)$, this restriction will be satisfied for all σ and δ^+.

can identify $\Pr(A_j < t)\Pr(\varepsilon_0 < c(200) - 2c(100))$ for all $t < 0$.[73][74] Combining these two results together, F_A is identified.

Identification of F_{ε_0} We now discuss identification of F_{ε_0} given that F_A and $c(\cdot)$ have been identified. Recall that $P_{200}(\mu, \sigma)$ can be expressed as follows,

$$P_{200}(\mu, \sigma) = \Pr(A_j < \overline{A}(\mu, \sigma) \wedge \varepsilon_0 < 200\mu - A_j(200\sigma)^2 - c(200)).^{75}$$

Suppose we take a μ so that $c(100) - c(200) + 100\mu > 0$ ($\Leftrightarrow \mu > \frac{c(200)-c(100)}{100}$). Now consider holding μ constant and taking the limit as $\sigma \to 0$. Then $P_{200}(\mu, \sigma) \to \Pr(\varepsilon_0 < 200\mu - c(200))$. Because we can move μ in the region $\mu > \frac{c(200)-c(100)}{100}$,[76] the distribution of ε_0 is identified for all $t > c(200) - 2c(100)$.

Now consider $P_{100}(\mu, \alpha)$, which is expressed as follows,

$$P_{100}(\mu, \sigma) = \Pr(\overline{A}(\mu, \sigma) < A_j < \frac{-c(100) + 50\mu}{7500\sigma^2} \wedge \varepsilon_0 < 100\mu - A_j(100\sigma)^2 - c(100)).^{77}$$

Again, take a μ so that $-c(100) + 50\mu > 0$ and $c(100) - c(200) + 100\mu < 0$ ($\Leftrightarrow \frac{c(100)}{50} < \mu < \frac{c(200)-c(100)}{100}$). As before, we take $\sigma \to 0$, while holding μ constant. Then $P_{100}(\mu, \sigma) \to \Pr(\varepsilon_0 < 100\mu - c(100))$. Because we can move μ in the region $\frac{c(100)}{50}, < \mu < \frac{c(200)-c(100)}{100}$[78] the distribution of ε_0 is identified for all $t \in [c(100), c(200) - 2c(100)]$.

Likewise, consider $P_{50}(\mu, \alpha)$,

$$P_{50}(\mu, \sigma) = \Pr(A_j > \frac{-c(100) + 50\mu}{7500\sigma^2} \wedge \varepsilon_0 < 50\mu - A_j(50\sigma)^2).^{79}$$

[73] We can apply the analogous argument here. We first fix $\overline{A}(\mu, \sigma)$ at some negative constant δ^-, but move $200\mu - A_j(200\sigma)^2 - c(200)$ by changing both μ and σ. Then considering $\sigma \to 0$, we obtain $\mu \to \frac{c(200)-c(100)}{100}$, and $P_{200}(\mu, \sigma) \to \Pr(A_j < \delta^- \wedge \varepsilon_0 < c(200) - 2c(100) = \Pr(A_j > \delta^-)\Pr(\varepsilon_0 < c(100) - 2c(200))$. Hence, by moving δ^- appropriately, we identify $\Pr(A_j < t)\Pr(\varepsilon_0 < c(200) - 2c(100))$ for all $t < 0$.

[74] We need μ to satisfy $\mu > \frac{-c(200)+5c(100)}{100}$. This means that $\frac{c(200)-c(100)+30000\sigma^2\delta^-}{100} > \frac{-c(200)+5c(100)}{100} \Leftrightarrow c(200) > 3c(100) - 15000\sigma^2\delta^-$. If $c(200) > 3c(100)$, for each δ^-, there will be some interval $(0, \epsilon_{\delta^-})$ such that for any $\sigma \in (0, \epsilon_{\delta^-})$ this restriction is satisfied.

[75] This is true as long as μ is "big" enough, i.e. $\frac{c(50)-c(100)+50\mu}{7500\sigma^2} > \frac{100\mu-c(200)+c(100)}{30000\sigma^2}$ ($\Leftrightarrow \mu > \frac{-c(200)+5c(100)}{100}$).

[76] Note that $c(200) > 3c(100)$ implies $\frac{c(200)-c(100)}{100} > \frac{-c(200)+5c(100)}{100}$.

[77] This is true as long as $\frac{c(50)-c(100)+50\mu}{7500\sigma^2} > \overline{A}(\mu, \sigma)$ ($\Leftrightarrow \mu > \frac{-c(200)+5c(100)}{100}$).

[78] Note that $c(200) > 3c(100)$ implies $\frac{c(200)-c(100)}{100} > \frac{c(100)}{50} > \frac{-c(200)+5c(100)}{100}$.

[79] This is true as long as $\frac{c(50)-c(100)+50\mu}{7500\sigma^2} > \frac{100\mu-c(200)+c(100)}{30000\sigma^2}$ ($\Leftrightarrow \mu > \frac{-c(200)+5c(100)}{100}$).

As before, take a μ so that $-c(100) + 50\mu < 0$ ($\Leftrightarrow \mu < \frac{c(100)}{50}$). Then $P_{50}(\mu, \sigma) \to \Pr(\varepsilon_0 < 50\mu)$. Because we can move μ in the region $\frac{c(100)}{50} > \mu$ ($> \frac{-c(200)+5c(100)}{100}$), the distribution of ε_0 is identified for at all $\frac{-c(200)+5c(100)}{2} < t < c(100)$.

Lastly, consider $P_{50}(\mu, \alpha)$, when $\mu < \frac{-c(200)+5c(100)}{100}$:[80]

$$P_{50}(\mu, \sigma) = \Pr(A_j > \frac{150\mu - c(200)}{37500\sigma^2} \wedge \varepsilon_0 < 50\mu - A_j(50\sigma)^2).$$

If we take a μ so that $150\mu - c(200) < 0$ ($\Leftrightarrow \mu < \frac{c(200)}{150}$). Then $P_{50}(\mu, \sigma) \to \Pr(\varepsilon_0 < 50\mu)$. Because we can move μ in the region $\mu < \min\{\frac{-c(200)+5c(100)}{100}, \frac{c(100)}{150}\}$ ($\Leftrightarrow \mu < \frac{-c(200)+5c(100)}{100}$), the distribution of ε_0 is identified for at all $t < \frac{-c(200)+5c(100)}{2}$. Combining these results, $F_{\varepsilon_0}(t)$ is identified for all $t \in \mathbb{R}$.

Identification of $P_q(\mu, \sigma)$ In this subsection, we discuss identification of F_N (the distribution of the number of lenders that visit the listing) and $P_q(\mu, \sigma)$ for all values of μ, σ and $q \in M$ under the assumption that lenders behave as if they are not pivotal. For the purpose of exposition, we start our discussion when $M = \{\$50\}$, i.e., when the lenders do not have any amount choice. Recall that we assumed that F_N has finite support, i.e., the support is $\{0, 1, \cdots, \bar{N}\}$ for some finite $\bar{N}$. First, the upper bound $\bar{N}$ is identified by the maximum requested amount by the borrower that has positive probability of being funded. If the borrower requests an amount that is larger than $\$50 \times \bar{N}$, then the loan is never funded. Conversely, for loans whose requested amount is less than $\$50 \times \bar{N}$, there is a positive probability of being funded. Hence $\bar{N}$ is identified by the maximum loan amount for which the probability of being funded is nonzero.[81]

Next we identify $\{P_0(\mu, \sigma), P_{50}(\mu, \sigma)\}$ and $f_N(0)$,..., $f_N(\bar{N})$, where $f_N(\cdot)$ is the pdf of F_N. In order to do so, consider listings which, if funded at an interest rate equal to the reserve interest rate, yields mean return μ and variance σ^2. Among such listings, consider listings with a requested amount just equal to $\$50 \times \bar{N}$. Then it follows that

$$\Pr(\text{fund} = 1 | x_{amt} = \$50 \times \bar{N}) = f_N(\bar{N}) \times P_{50}(\mu, \sigma)^{\bar{N}},$$

[80] This is the case when the intersection between U_{50} and U_{200} lies to the right of the intersection between U_{50} and U_{100}. ($\frac{100\mu - c(200) + c(100)}{30000\sigma^2} > \frac{c(50) - c(100) + 50\mu}{7500\sigma^2}$)

[81] In practice, Prosper has a cap regarding how much a borrower can request. The cap is at $\$25,000$. While the funding probability of listings that request $\$25,000$ is small, it is nonetheless strictly positive. Hence, assuming rich support for request amount is somewhat problematic.

where the left hand side is the observed funding probability, and the right hand side is the probability that $\bar{N}$ potential lenders visit the listing, multiplied by $P_{50}(\mu, \sigma)$. Recall that $P_{50}(\mu, \sigma)$ is just the probability that all lenders who visited the listing receive higher utility from bidding \$50 than from not lending (assuming that the listing is funded at an interest equal to the reserve interest rate). Given the strategy of the lenders that we described in section 4.2, a listing is funded if and only if there are enough lenders that are willing to fund it when the mean and the variance of the return from the listing is evaluated at the reserve interest rate. This implies that a listing with a loan amount equal to $\$50 \times \bar{N}$ is funded if and only if there are $\bar{N}$ bidders who visit the listing and all of them prefer to bid \$50 on the listing to not funding the listing (at the reserve interest rate).

Similarly, consider a listing with requested amount equal to $\$50 \times (\bar{N} - 1)$. Then,

$$\begin{aligned} & \Pr(\text{fund} = 1 | x_{amt} = \$50 \times (\bar{N} - 1)) \\ = \; & f_N(\bar{N}) \times C_1^{\bar{N}} \times P_0(\mu, \sigma) P_{50}(\mu, \sigma)^{\bar{N}-1} + f_N(\bar{N} - 1) \times P_{50}(\mu, \sigma)^{(\bar{N}-1)} + f_N(\bar{N}) \times P_{50}(\mu, \sigma)^{\bar{N}}. \end{aligned}$$

The right hand side is equal to the sum of three probabilities: the first term is the probability that $\bar{N}$ potential lenders visit the listing and $\bar{N} - 1$ of them decide to bid, the second term corresponds to the probability that $\bar{N} - 1$ potential lenders visit the listing, and all of them decide to bid, and the last term is the probability that there are $\bar{N}$ potential lenders, all of whom decide to bid. We repeat this process for all amounts $\{50, 100, \cdots, 50 \times \bar{N}\}$. This yields $\bar{N}$ equations (for each loan amount) and $\bar{N} + 1$ unknowns, $P_{50}(\mu, \sigma)$, $f_N(1)$,..., $f_N(\bar{N})$.[82] Now consider repeating the above exercise with a different μ and σ (say μ' and σ'). Then this yields $\bar{N}$ additional equations. Because we assume that F_N is invariant to (μ, σ), we have a total of $2 \times \bar{N}$ equations and $\bar{N} + 2$ unknowns ($P_{50}(\mu, \sigma)$, $P_{50}(\mu', \sigma')$, $f_N(0)$,..., $f_N(\bar{N})$). Assuming that F_N is invariant to (μ, σ), we can increase the number of equations at a faster rate than the number of observables. Hence $P_{50}(\mu, \sigma)$, and $f_N(0)$,..., $f_N(\bar{N})$ are identified for all μ and σ.

The preceding identification argument focused on the case when $M = \{\$50\}$, i.e., when there is no amount choice. We now briefly discuss identification when $M = \{\$50, \$100, \$200\}$. As before, we start with identification of $\bar{N}$: $\bar{N}$ is again identified by the maximum loan amount for which the probability of being funded is nonzero: $\$200 \times \bar{N}$ is the threshold loan amount, below which the

[82] Note that $P_0(\mu, \sigma) = 1 - P_{50}(\mu, \sigma)$ and $f_N(0) = 1 - \sum_{n=1}^{\bar{N}} f_N(n)$.

probability of being funded is positive, and above which the probability is zero.

In the presence of amount choice, the objects that we would like to identify are now $\{P_0(\mu,\sigma)$, $P_{50}(\mu,\sigma)$, $P_{100}(\mu,\sigma)$, $P_{200}(\mu,\sigma)\}$ and $f_N(0)$,..., $f_N(\bar{N})$. As before, consider listings which, if funded at an interest rate equal to the reserve interest rate, yields mean return μ and variance σ^2. Moreover, if we consider listings with amount equal to $\$200 \times \bar{N}$, we see that

$$\Pr(\text{fund} = 1 | x_{amt} = \$200 \times \bar{N}) = f_N(\bar{N}) \times P_{200}(\mu,\sigma)^{\bar{N}}.$$

For listings with amount equal to $\$200 \times (\bar{N}-1) + 100$,

$$\begin{aligned} & \Pr(\text{fund} = 1 | x_{amt} = \$200 \times (\bar{N}-1) + 100) \\ = \; & C_1^{\bar{N}} \times f_N(\bar{N}) \times P_{200}(\mu,\sigma)^{\bar{N}-1} P_{100}(\mu,\sigma) + f_N(\bar{N}) \times P_{200}(\mu,\sigma)^{\bar{N}}. \end{aligned}$$

Similarly, we can express the probability that the loan is funded for different loan amounts as a function of $P_q(\mu,\sigma)$ and f_N. The number of (independent) equations we end up with is $\bar{N}+2$, and the number of unobservables is $\bar{N}+3$. Assuming that F_N is invariant to (μ,σ), we can increase the number of equations at a faster rate than the number of observables. Hence $P_q(\mu,\sigma)$ and f_N are identified for all μ and σ.

We acknowledge that assuming F_N is invariant to (μ,σ) is a strong assumption: In fact it is stronger than we need. We only need F_N to be invariant to a small subset of the characteristic of the listings. Out of the many listing characteristics, it is natural to let F_N depend on some of them, such as the credit grade. This is possible, as long as there is some element x_k in the vector of listing characteristics x, to which F_N is invariant.

A.4 Estimation Procedure of $\Pr(s|x)$ and $f(r|s,x)$

We explain how to implement estimation of $\Pr(s|x)$ and $f(r|s,x)$ in this section. In order to estimate $f(r|s,x)$, we first divide the observations into 14 subsamples by the credit grade of the borrower (AA, A, B, C, D, E, and HR) and by home ownership. This is necessary because the estimation strategy by Gallant and Nychka (1987) requires continuous support for each covariate. Hence, for discrete conditioning variables we nonparametrically estimate $f(r|s,x)$ separately for

each value. For estimation of $\Pr(s|x)$, we divide the observations into 7 subsamples by the credit grade of the borrower and estimate a Probit model, i.e., $\Pr(s|x) = \Phi(P(x,s))$, where P is a second order polynomial of x and s. Let $\widehat{\Pr}(s|x)$ and $\widehat{f}(r|s,x)$ be the estimates of $\Pr(s|x)$ and $f(r|s,x)$. In addition to $f(r|s,x)$ and $\Pr(s|x)$, we need $\frac{\partial \Pr(s|x)}{\partial s}$ and $\frac{\partial f(r|s,x)}{\partial s}$, when we evaluate the first order condition of the borrower's problem. We compute $\frac{\partial \widehat{\Pr}(s|x)}{\partial s}$ and $\frac{\partial \widehat{f}(r|s)}{\partial s}$ by taking analytical derivatives of $\widehat{\Pr}(s|x)$ and $\widehat{f}(r|s,x)$, respectively.

A.5 Discussion of Partial Pooling

A.5.1 Additional Condition for Partially Pooling Equilibria at 36%

In order for there to exist an equilibrium with partial pooling among the low types, we need an extra condition in addition to the ones that we explained in the main text. The extra condition requires that the pooled types do not benefit from changing the reserve rate. Formally, let φ^m denote the marginal type, where borrowers with types below φ^m are pooled and borrowers above are not pooled. Moreover, let $s^m(\neq 0.36)$ be the largest reserve rate that the set of non-pooled types submit. Then the extra condition we need is

$$V_0(0.36, \varphi^m) = V_0(s^m, \varphi^m),$$

i.e., the marginal type is indifferent between being pooled and not pooled.

A.5.2 Identification of the Borrower's Model when There is Pooling

Our discussion in the main text focused on the case when there is no pooling. As long as we can identify $F_{\varepsilon|X}$ using the subset of the borrowers who are not pooled, we can identify $F_{\varphi|X}$ for the case of pooling as well. To see this, first note that we can identify $F_{\varphi|X}$ for borrowers who are not being pooled just as before. Now consider the terminal decision of the borrower who is being pooled:

$$\begin{cases} \text{repay: if } -(r \times x_{amt}) + \varepsilon_T \geq -F^{-1}_{\varphi|X}(\alpha^{pool}) \\ \text{default: otherwise} \end{cases},$$

where $F^{-1}_{\varphi|X}(\alpha^{pool})$ is a random variable with $\alpha^{pool} \sim U[0, m^{pool}]$ and m^{pool} is the fraction of borrowers who submit $s = 0.36$.[83] Note that $F^{-1}_{\varphi|X}(\alpha^{pool})$ is a random variable because we do not know the exact value of φ for pooled borrowers: We only know that φ is below $F^{-1}_{\varphi|X}(m^{pool})$. Given that the distribution of $\varepsilon_T + F^{-1}_{\varphi|X}(\alpha^{pool})$ can be identified and we have already identified the distribution of ε_T from markets with no pooling, it is immediate that we can identify the distribution of $F^{-1}_{\varphi|X}(\alpha^{pool})$ nonparametrically.

References

[1] Adams, William, Liran Einav, Jonathan Levin, "Liquidity Constraints and Imperfect Information in Subprime Lending," *American Economic Review*, Vol. 99(1), 2009, pp. 49–84.

[2] Aghion, Philippe and Patrick Bolton, "Contracts as a Barrier to Entry," *American Economic Review*, 77(3), 1987, pp.388–401.

[3] Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, Vol. 84(3), 1970, pp. 488-500.

[4] Austen-Smith, David and Roland Fryer, "An Economic Analysis of "Acting White"," *Quarterly Journal of Economics*, Vol. 120, 2005, pp. 551-83.

[5] Ausubel, Larry, "Adverse Selection in the Credit Card Market," mimeo, 1999.

[6] Berger, Allen N. and Gregory F. Udell, "Some Evidence on the Empirical Significance of Credit Rationing," *Journal of Political Economy*, Vol. 100 (5), 1992, pp. 1047-1077.

[7] Bhattacharya, S., "Imperfect Information, Dividend Policy and 'The Bird in the Hand' Fallacy," *Bell Journal of Economics*, Vol. 9(1), 1979, pp. 259-70.

[8] Cai, Hongbin, John Riley, and Lixin Ye, "Reserve price signaling," *Journal of Economic Theory*, Vol. 135, 2007, pp. 253 – 268.

[9] Cohen, Alma and Liran Einav, "Estimating Risk Preferences from Deductible Choice," *American Economic Review*, Vol. 97(3), 2007, pp. 745-788.

[83] Note that if α is the quantile of φ given X, i.e., $\alpha = F_{\varphi|X}(\varphi)$, then $\Pr(\alpha \leq t) = \Pr(F_{\varphi|X}(\varphi) \leq t) = \Pr(\varphi \leq F^{-1}_{\varphi|X}(t)) = t$. Hence α is a uniformly distributed random variable.

[10] Einav, Liran, Amy Finkelstein and Jonathan Levin, "Beyond Testing: Empirical Models of Insurance Markets," *Annual Review of Economics*, 2010, pp. 311-36.

[11] Einav, Liran, Mark Jenkins, and Jonathan Levin, "Contract Pricing in Consumer Credit Markets," *Econometrica,* Vol. 80 (4), 2012, pp. 1387-432.

[12] Fang, Hanming, "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices," *International Economic Review*, Vol. 47 (4), 2006, pp. 1151-1185.

[13] Freedman, Seth and Ginger Jin, "Learning by Doing with Asymmetric Information: Evidence from Prosper.com," mimeo, 2010.

[14] Gallant, A. Ronald, and Douglas W. Nychka, "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica*, Vol. 55(2), 1987, pp. 363-390.

[15] Gayle, George-Levi and Limor Golan, "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968-97," *Review of Economic Studies*, Vol. 79 (1), 2012, pp. 227-67.

[16] Hörner, Johannes and Nicolas Sahuguet, "A war of attrition with endogenous effort levels," *Economic Theory*, Vol. 47, 2011, pp.1-27.

[17] Hungerford, Thomas and Gary Solon, "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, Vol. 69 (1), 1987, pp. 175-177.

[18] Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue, "Inferring Asset Quality: Determining Borrower Creditworthiness in Peer-to-Peer Lending Markets," mimeo, 2010.

[19] Jullien, Bruno and Thomas Mariotti, "Auction and the Informed Seller Problem," *Games and Economic Behavior*, Vol. 56(2), 2006, pp. 225-58.

[20] Karlan, Dean and Jonathan Zinman, "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment," *Econometrica*, Vol. 77(6), 2009, pp. 1993-2008.

[21] Kihlstrom, Richard E. and Michael Riordan, "Advertising as a Signal," *Journal of Political Economy*, Vol. 92, 1984, pp. 427-450.

[22] Mailath, George J., "Incentive Compatibility in Signaling Games with a Continuum of Types," *Econometrica*, Vol. 55(6), 1987, pp. 1349-1365.

[23] Milgrom, Paul and John Roberts, "Limit Pricing and Entry under Incomplete Information: An Equilibrium Analysis," *Econometrica*, Vol. 50 (2), 1982, pp. 443-459.

[24] Milgrom, Paul and John Roberts, "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, Vol. 94 (4), 1986, pp. 796-821.

[25] Morris, Stephen and Hyun Song Shin, "Contagious Adverse Selection," *American Economic Journal: Macroeconomics*, Vol. 4, 2012, pp. 1-21.

[26] Paravisini, Daniel, Veronica Rappoport, and Enrichetta Ravina, "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios," mimeo, 2011.

[27] Powell, James, "Censored Regression Quantiles," *Journal of Econometrics*, Vol. 32, 1986, pp. 143-55.

[28] Prat, Andrea, "Campaign Advertising and Voter Welfare," *Review of Economic Studies*, Vol. 69, 2002, pp. 999-1017.

[29] Ravina, Enrichetta, "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets," mimeo, 2008.

[30] Rigbi, Oren (2011) "The Effects of Usury Laws: Evidence from the Online Loan Market," forthcoming in *Review of Economics and Statistics*.

[31] Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, Vol. 87(3), 1973, pp. 355-374.

[32] Stiglitz, Joseph E., and Andrew Weiss, "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, Vol. 71(3), 1981, pp. 393-410.

Table 1: Descriptive Statistics – Listings

Grade	Amount Requested		Reserve Rate		Debt/ Income		Home Owner		Bid Count		Fund Pr.	Obs.
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd		
AA	13,144.8	8,342.8	0.132	0.047	0.364	0.976	0.812	0.391	171.0	204.2	0.534	1,420
A	12,396.2	7,881.7	0.165	0.067	0.376	0.673	0.612	0.487	116.1	161.6	0.409	1,850
B	10,622.4	6,096.5	0.211	0.075	0.386	0.655	0.593	0.491	82.5	117.2	0.334	3,068
C	7,622.3	5,158.0	0.246	0.078	0.373	0.623	0.556	0.497	39.3	65.2	0.247	5,203
D	6,368.5	4,691.3	0.287	0.075	0.389	0.711	0.370	0.483	19.8	42.3	0.155	6,581
E	4,783.5	4,868.2	0.310	0.073	0.360	0.680	0.329	0.470	4.7	13.7	0.068	5,757
HR	4,350.7	4,599.4	0.315	0.069	0.308	0.641	0.221	0.415	2.2	7.3	0.030	11,362
All	6,603.9	5,937.8	0.274	0.089	0.354	0.679	0.393	0.488	31.1	84.1	0.158	35,241

This table presents summary statistics of listings posted on Prosper.com by credit grade. Debt/Income is the debt-to-income ratio of the borrower. Home Owner is a dummy variable that equals 1 if the potential borrower is a homeowner and 0, otherwise. Bid Count is the number of submitted bids by the lenders. Fund Pr. stands for the percentage of listings that are funded.

Table 2: Descriptive Statistics – Loans

Grade	Amount Requested		Reserve Rate		Contract Rate		Debt/ Income		Home Owner		Bid Count		Obs.
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	
AA	9,710	7,384	0.131	0.046	0.096	0.033	0.21	0.39	0.80	0.40	131.5	99.3	755
A	8,723	6,626	0.165	0.060	0.127	0.045	0.23	0.14	0.55	0.50	114.0	84.4	755
B	7,347	4,858	0.216	0.063	0.164	0.046	0.27	0.34	0.56	0.50	100.9	67.6	1,023
C	4,687	2,998	0.247	0.064	0.181	0.062	0.25	0.21	0.48	0.50	53.4	38.3	1,285
D	3,578	2,380	0.280	0.064	0.210	0.066	0.24	0.17	0.26	0.44	21.6	11.7	1,022
E	1,890	1,187	0.339	0.028	0.291	0.057	0.22	0.22	0.26	0.44	44.7	30.6	392
HR	1,690	1,288	0.339	0.036	0.300	0.057	0.20	0.44	0.17	0.38	17.6	10.4	339
All	5,821	5,285	0.233	0.086	0.179	0.079	0.24	0.28	0.47	0.50	80.0	76.7	5,571

This table reports the summary statistics of loans. Loans are the subset of listings that are funded. Contract Rate is the interest rate charged to the borrower. Debt/Income refers to the debt-to-income ratio of the borrower. Home Owner is a dummy variable that equals 1 if the potential borrower is a homeowner and 0, otherwise. Bid Count is the number of submitted bids by the lenders.

Table 3: Descriptive Statistics – Default Timing

	Default	Default Timing							
Grade	Prob.	mean	sd	10%	25%	50%	75%	90%	Obs
AA	0.149	18.7	8.9	9	12	17	25	35	113
A	0.211	18.7	9.2	8	11	16	26	33	159
B	0.297	17.3	8.0	8	11	16	23	28	303
C	0.309	17.2	8.3	7	10	16	23	30	397
D	0.321	17.8	8.9	7	10	16	25	31	328
E	0.372	18.2	9.0	7	10	17	26	32	146
HR	0.439	15.3	8.3	7	9	13	20	28	149
All	0.286	17.5	8.6	7	10	16	24	31	1,595

This table presents descriptive statistics of the repayment behavior of the borrowers, by credit grade. The first column reports the percentage of funded loans that end in default. The second and third columns report the average and standard error of the number of months until default for the subset of loans that end in default. Columns 4 through 8 report the quantiles of the number of months until default.

Table 4: Descriptive Statistics – Internal Rate of Return

Grade	mean	sd	10%	25%	50%	75%	90%	Obs
AA	-0.011	0.283	-0.449	0.061	0.082	0.110	0.132	755
A	-0.025	0.331	-0.767	0.072	0.094	0.135	0.181	755
B	-0.074	0.404	-0.871	-0.229	0.136	0.169	0.211	1,023
C	-0.060	0.413	-0.871	-0.211	0.135	0.196	0.256	1,285
D	-0.036	0.424	-0.865	-0.192	0.153	0.231	0.316	1,022
E	0.000	0.475	-0.861	-0.315	0.249	0.345	0.394	392
HR	-0.112	0.532	-0.886	-0.800	0.202	0.345	0.398	339
All	-0.046	0.402	-0.858	-0.100	0.121	0.187	0.281	5,571

This table reports the sample statistics of the internal rate of return (IRR) of the loans originated by Prosper. We present the average IRR, the standard error, and the quantiles of the distriubution. See footnote 21 on how we computed the IRR.

Table 5: Reduced Form Analysis - Funding Probability and Contract Interest Rate.

	Funded	Contract Rate
Reserve Rate	2.1368***	0.6834***
	(0.0263)	(0.0145)
Amount	−0.0001***	0.00001***
	(0.0000)	(0.0000)
Debt/Income	−0.7971***	0.0731***
	(0.0015)	(0.0037)
Home owner	−0.1513***	0.0137***
	(0.0004)	(0.0018)
Grade		
AA	3.6468***	−0.3013***
	(0.0044)	(0.0061)
A	3.0727***	−0.2670***
	(0.0033)	(0.0055)
B	2.5681***	−0.2347***
	(0.0022)	(0.0046)
C	1.8743***	−0.1862***
	(0.0014)	(0.0038)
D	1.2754***	−0.1329***
	(0.0011)	(0.0034)
E	0.5022***	−0.0499***
	(0.0014)	(0.0036)
Observation	35,241	35,241
R^2	0.2827	
Likelihood		−1,137

The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. In addition to the independent variables shown in the table, we also control for month dummies, day-of-the-week dummies, and hour-of-the-day dummies. Standard errors are robust-heteroskedasticity-consistent, and are presented in parentheses below the coefficients.

Table 6: Reduced Form Analysis - Repayment Behavior and Reserve Interest Rate.

	Default	Rate of Return	Default Time
Reserve rate	1.5365***	-0.5919***	3.6714***
	(0.4095)	(0.1313)	(0.8019)
Contract rate	2.1531***	0.0540	4.1446***
	(0.4091)	(0.1372)	(0.7463)
Amount	1.92E-05***	-4.51E-06***	3.30E-05***
	(4.38E-06)	(1.24E-06)	(7.64E-06)
Debt / income	0.0275	-0.0314	0.0726
	(0.0588)	(0.0197)	(0.1016)
Home owner	0.0633	-0.0471***	0.1399***
	(0.0366)	(0.0117)	(0.0714)
Grade			
AA	-0.1896	0.0595	-0.1714
	(0.1236)	(0.0402)	(0.2389)
A	-0.1543	0.0475	-0.1764
	(0.1083)	(0.0366)	(0.2100)
B	-0.0888	0.0224	-0.0999
	(0.0894)	(0.0320)	(0.1700)
C	-0.0780	0.0380	-0.0233
	(0.0777)	(0.0288)	(0.1664)
D	-0.1162	0.0636**	-0.0843
	(0.0773)	(0.0272)	(0.1406)
E	-0.2814***	0.1155***	-0.4632***
	(0.0878)	(0.0296)	(0.1484)
Observation	85,657	5,571	85,657
R^2		0.0224	
Likelihood	-4,686		-10,333

The first column reports estimated coefficients from the panel Probit model (expression (3)). The unit of observation is a loan - period. The dependent variable is an indicator variable that equals one if the loan ends in default at period t. The second column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. The third column shows estimated coefficients of the Cox's proportional hazard model (expression (5)). The dependent variable is the month from the start of the loan in which the borrower defaults. In addition to the independent variables shown in the table, we also control for month dummies, day-of-the-week dummies, and hour-of-the-day dummies. Standard errors are robust-heteroskedasticity-consistent, and are presented in parentheses below the coefficients.

Table 7: Parameter Estimates of the Borrower's and Lender's Model.

	Borrower		Lender Estimates by Grade			
Parameter	Estimates	Parameter	AA	A	B	C
σ_ε	6.0096	μ_N	4.5848	3.5513	4.3386	2.9504
	(0.2095)		(0.0635)	(0.1569)	(0.2761)	(0.0733)
λ_{AA}	3.0965	σ_N	0.7213	1.3940	1.1132	1.4640
	(0.0827)		(0.0442)	(0.0984)	(0.3151)	(0.0448)
λ_A	4.3398	μ_A	2.24×10^{-2}	1.91×10^{-2}	3.67×10^{-2}	3.46×10^{-2}
	(0.1384)		(1.64×10^{-3})	(1.72×10^{-3})	(8.22×10^{-3})	(1.79×10^{-3})
λ_B	9.4551	σ_A	2.22×10^{-2}	2.00×10^{-2}	1.75×10^{-2}	1.58×10^{-2}
	(0.3477)		(9.36×10^{-4})	(6.58×10^{-4})	(3.29×10^{-3})	(4.56×10^{-4})
λ_C	8.8111	μ_{ε_0}	−14.8775	−13.2908	−9.60346	−1.4651
	(0.3551)		(0.8411)	(0.7640)	(1.8669)	(0.0385)
		σ_{ε_0}	86.4102	62.7313	32.1786	88.4056
			(10.2508)	(4.0462)	(4.4980)	(2.6275)
		c_{100}	−1.6206	−1.7003	−0.5006	0.56680
			(0.3704)	(0.1413)	(0.0755)	(0.0238)
		c_{200}	−21.9644	−27.3998	−10.9928	−14.8229
			(3.1801)	(2.8967)	(2.7251)	(0.5049)
Obs	3,818		1,420	1,850	3,068	5,203

We report the parameter estimates of the borrower's model in the first column of this table and the estimation results of the lender's model in the rest. Time dummies are included in the estimation of the borrower's model, but we omit the estimates from the table. Standard errors are obtained by bootstrap and they are reported in parentheses.

Table 8: Quantile of the Borrower's Type Distribution.

Quantile	AA	A	B	C
25%	7.493	5.597	2.173	2.425
50%	9.367	6.766	2.789	2.957
75%	10.509	7.417	3.242	3.400

This table reports the quartile of the default cost of the borrower, φ, by credit grade.

Table 9: Parameter Estimates of the Regression of the Borrower's Type on the Characteristics.

Parameter	Estimates
Constant	16.01
	(0.86)
Log amount requested	−2.00
	(0.10)
Debt to Income Ratio	−1.25
	(0.11)
Home Ownership	0.26
	(0.15)
Grade	
AA	11.23
	(0.24)
A	8.49
	(0.22)
B	3.51
	(0.18)
Observation	3,818

The table presents the estimated coefficients of the Tobit model described in the last paragraph of Section 7. The dependent variable is the estimated borrower type, $\hat{\varphi}$. Standard errors are obtained by bootstrap and they are reported in parentheses.

Table 10: Expected Surplus for Different Market Designs by Credit Grade.

		Expected			Median		
		Borrower	Lender	Total	Borrower	Lender	Total
AA	Signaling	466.4	1642.6	2109.0	423.3	1645.2	2068.5
	Pooling	613.2	2531.4	3144.6	567.7	2587.1	3154.9
	Symmetric	675.7	2801.6	3477.2	633.0	2885.6	3518.6
A	Signaling	143.0	1470.4	1613.4	143.1	1488.6	1631.7
	Pooling	142.7	1491.7	1634.4	142.4	1497.5	1639.9
	Symmetric	143.3	1481.5	1624.8	143.1	1488.6	1631.7
B	Signaling	396.7	573.5	970.2	380.4	577.9	958.3
	Pooling	343.1	477.6	820.7	327.7	478.1	805.7
	Symmetric	398.4	579.8	978.2	382.2	581.7	963.9
C	Signaling	417.9	510.0	927.9	434.5	508.5	943.0
	Pooling	374.7	445.3	820.0	392.1	446.1	838.2
	Symmetric	425.5	518.7	944.2	446.7	517.3	964.0

The first three columns correspond to the expected surplus averaged across all borrower types. The last three columns correpond to the expected surplus of the median borrower type. The requested amount, debt-to-income ratio, and home ownership status are set to their median values. The units are in U.S. dollars.

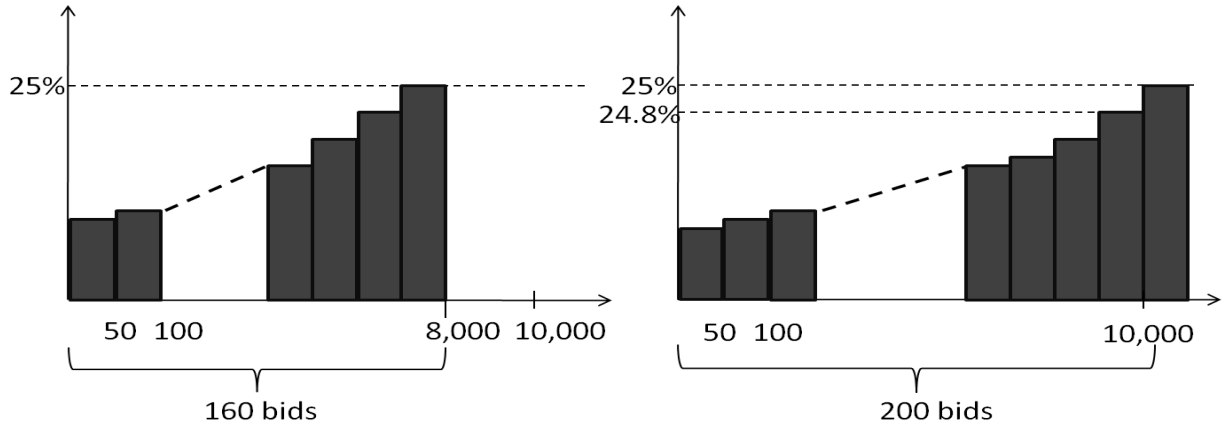


Figure 1: Funding Decision – The figure shows how a loan is funded for the simple case in which lenders only submit a bid with an amount of $50. The horizontal axis corresponds to amount and the vertical axis corresponds to the interest rate. The left panel illustrates a situation in which the requested amount is $10,000, and the listing has received 160 bids ($8,000). The right panel illustrates the situation in which the requested amount is $10,000, and it has attracted more than 200 bids.

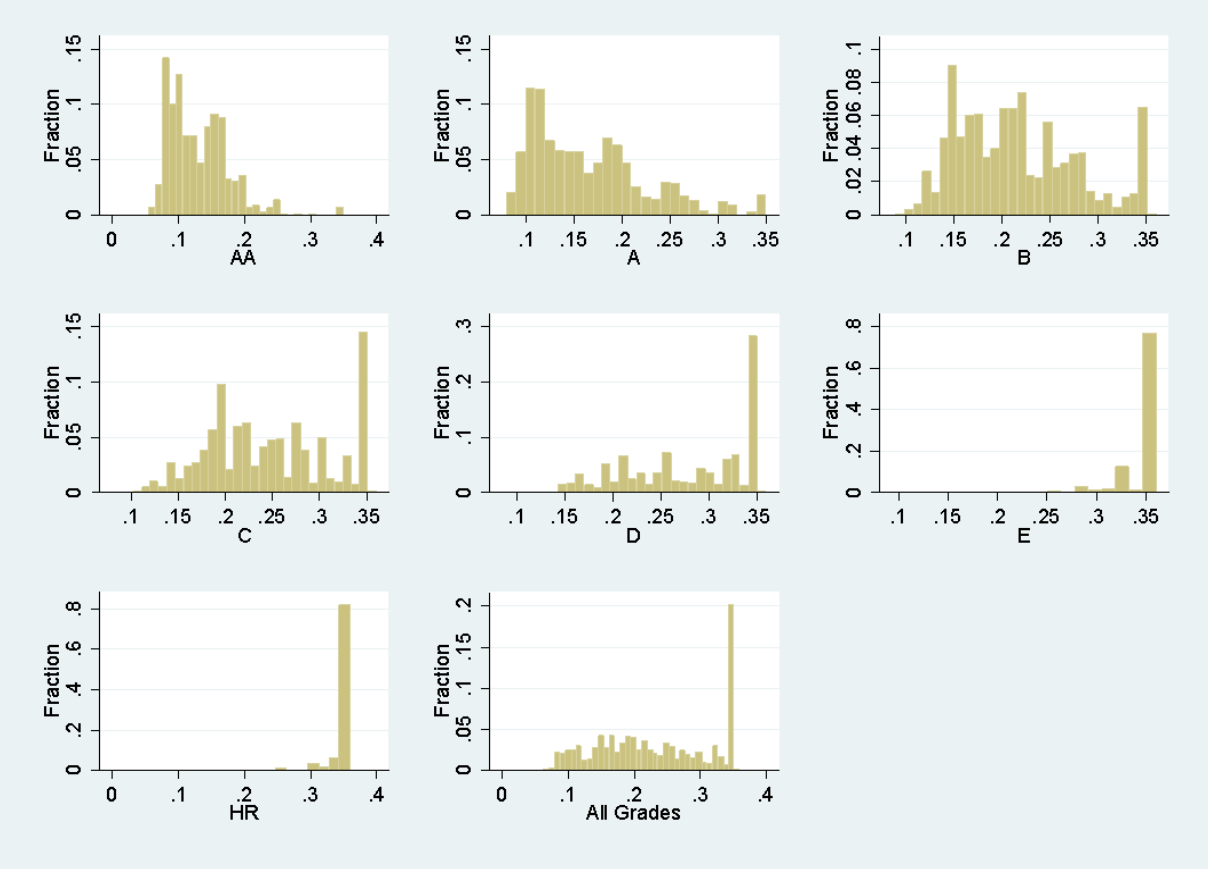


Figure 2: Distribution of Reserve Interest Rate by Credit Grade – We show the distribution of reserve interest rate by credit grade. The reserve interest rate is capped at 36% because of the usury law.

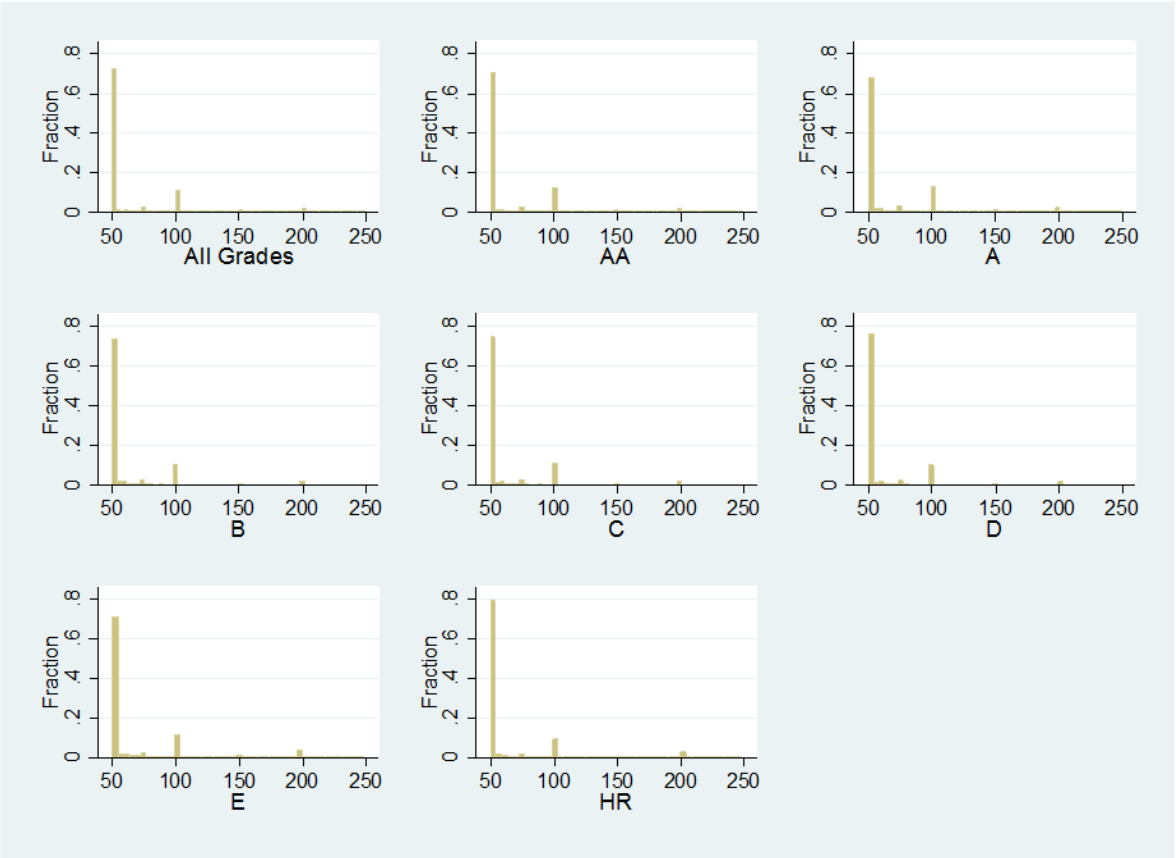


Figure 3: Distribution of Bid Amount – The figure shows the distribution of bid amount for each credit grade. Bids with amount exceeding $250 are not shown. The fraction of these bids is about 3.5%.

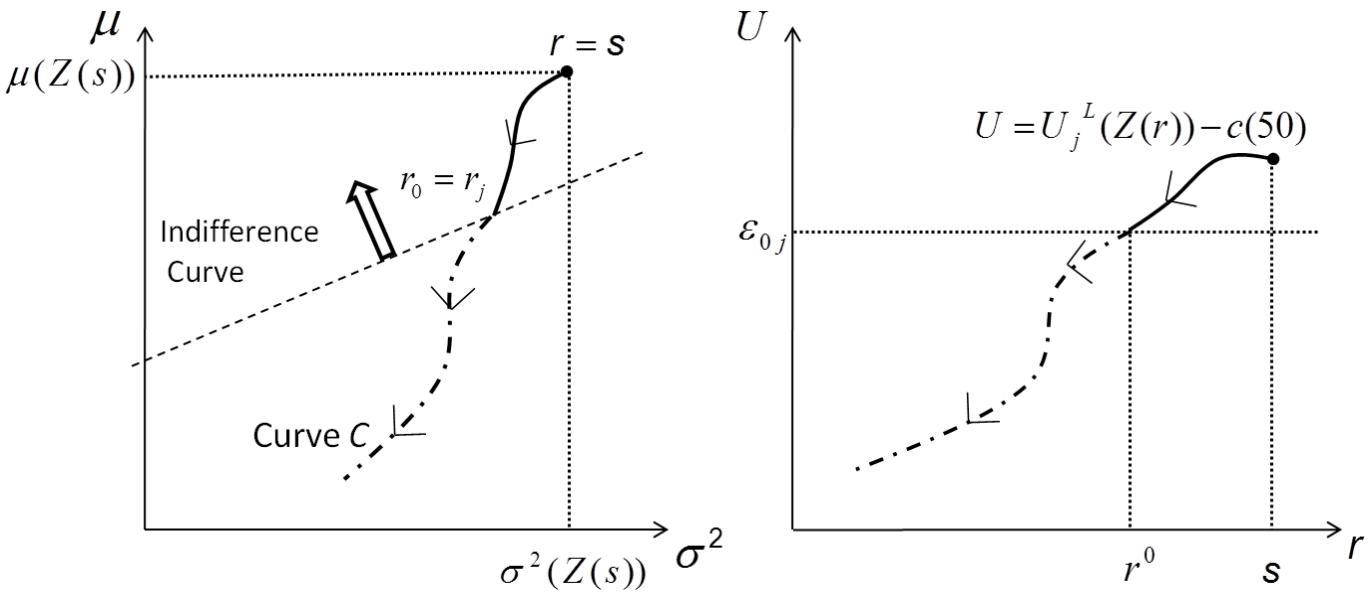


Figure 4: Graphical Representation of the Lender's Problem: Case of No Amount Choice – The figure illustrates how the lender should bid when there is no amount choice. In the left panel, the horizontal axis is σ^2 and the vertical axis is μ. For each listing and for each realization of the contract interest rate, we can assign a corresonding point on this $\mu - \sigma^2$ plane. Curve C corresponds to the mean and variance of a listing for different realizations of r. The dashed line is the lender's indifference curve. The right panel plots the lender's utility, $U_j^L(Z(r))$, against r. The dotted line at ε_{0j} corresponds to the lender's outside option.

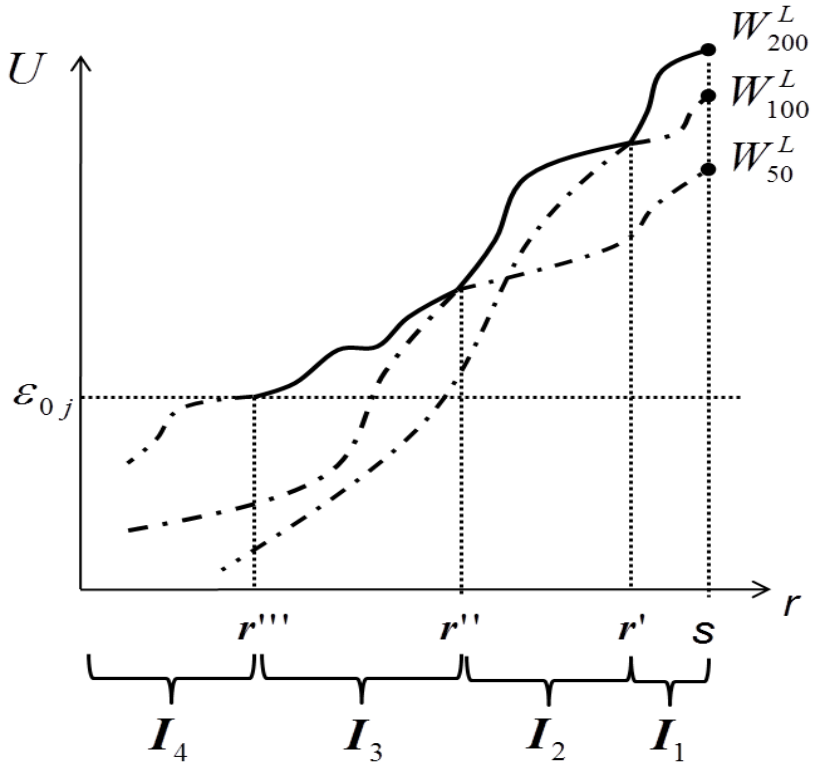


Figure 5: Graphical Representation of the Lender's Problem: Case of Amount Choice – The figure illustrates how the lender should bid when there is amount choice. The horizontal axis corresponds to the interest rate, and the vertical axis corresponds to the lender's utility. Each curve $W^L_q(r)$ illustrates the relationship between r and the lender's utility when the lender bids q. I_1 corresponds to the region of the active interest rate for which bidding \$200 is optimal. I_2, I_3, and I_4 correspond to the regions of the active interest rate for which bidding \$100, \$50, and \$0 is optimal, respectively.

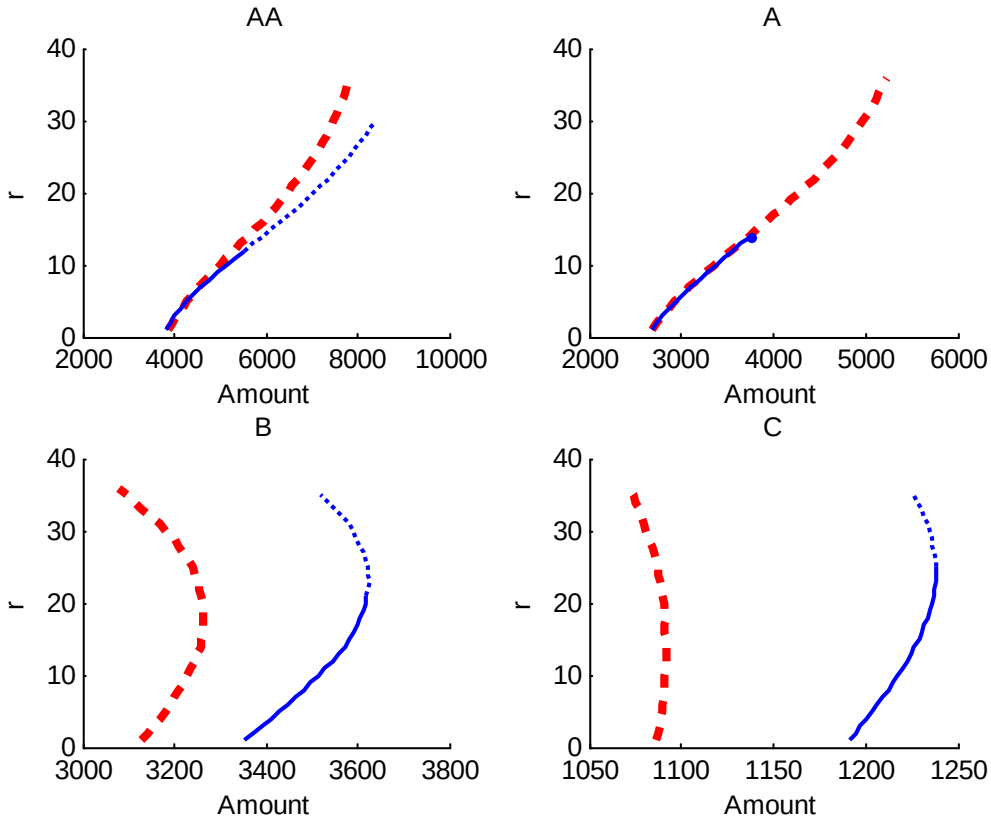


Figure 6: The Credit Supply Curve for the Borrower of the Median Type – The thick dotted curve corresponds to the credit supply curve under no signaling (i.e., pooling). The solid line corresponds to the credit supply curve under signaling, and the dotted line that lies on top of it corresponds to the credit supply curve under no asymmetric information. Borrower covariates are set to the median values.

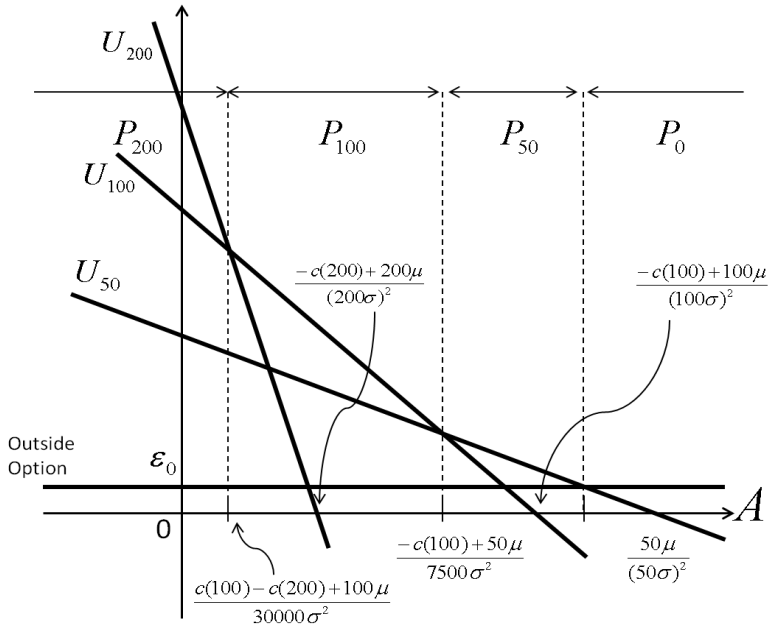


Figure 1: (Appendix) Lender utility as a function of A – The figure shows the regions of A that correspond to $P_0(\mu,\sigma)$, $P_{50}(\mu,\sigma)$, $P_{100}(\mu,\sigma)$, and $P_{200}(\mu,\sigma)$, for a given value of ε_0. This graph is drawn for the case in which the intersection between U_{200} and U_{100} is located to the left of the intersection between U_{100} and U_{50}. In the figure, three lines represent $U_{50} = 50\mu + 2500\sigma A - c(50)$, $U_{100} = 100\mu + 10000\sigma A - c(100)$, and $U_{200} = 200\mu + 40000\sigma A - c(200)$, respectively.