Signaling in Online Credit Markets[*]

Kei Kawai[†]   Ken Onishi[‡]

University of California Berkeley   Singapore Management University

Kosuke Uetake[§]

Yale School of Management

February 2018

Abstract

We study how signaling affects equilibrium outcomes and welfare in markets with adverse selection. Using data from an online credit market, we estimate a model of borrowers and lenders where low reserve interest rates can signal low default risk. Comparing a market with and without signaling relative to the benchmark case with no asymmetric information, we find that adverse selection destroys as much as 31% of total surplus, up to 86% of which can be restored with signaling. We also find the credit supply curves to be backward-bending for some markets, consistent with the prediction of Stiglitz and Weiss (1981).

JEL Code: D82, D83, G21, L15

KEYWORDS: Signaling, adverse selection, structural estimation

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.

Recently, there is a growing empirical literature in industrial organization that studies the effect of adverse selection on market outcomes and welfare.[1] In this paper, we ask the natural next question, how signaling affects equilibrium outcomes and welfare in markets with adverse selection. While the theory of signaling has been applied to a wide range of topics in industrial organization, there is very little empirical work that quantifies the extent to which signaling affects market outcomes and welfare relative to a market with no signaling (i.e., pooling). An empirical analysis of welfare seems especially important given that whether signaling improves or decreases total welfare relative to pooling is theoretically ambiguous.[2]

This paper studies these questions by building an estimable model of signaling in credit markets for unsecured loans using data from Prosper.com, an online peer-to-peer loan market. 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.[3] In this paper, we provide evidence that the borrower's reserve interest rate signals his creditworthiness and explore how signaling affects market outcomes and welfare.

Prosper.com is an online platform that matches potential borrowers with potential lenders with more than $280 million in funded loans (as of 2011). Established in 2006, it specializes in small-scale unsecured loans to individuals with a standardized loan repayment length of 36 months. The average funded loan is about $5,800 and debt consolidation is, by far, the most commonly stated purpose of the loan, accounting for about 46% of all listings.

While Prosper is a relatively young and small market, it is an ideal setting for investigating the effect of signaling on market outcomes and welfare. First, in this market we observe both the reserve interest rate choice of the borrower as well as the contract interest rate determined by the auction. The contract interest rate is the actual interest rate that the borrower faces in repayment and it is often strictly lower than the reserve interest rate. Because the reserve interest rate should not

---

[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]For a brief discussion of how signaling equlibriun can be pareto dominated by a pooling equilibrium, see Mas-Colell, Whinston and Green (1995), Chapter 13.C, p.454.

[3]We use borrower listing data from May through October of 2008 and the corresponding loan repayment data that goes until 2011. Our description of the institutional details pertain to this period.

affect the borrower's repayment behavior conditional on the contract interest rate, we can isolate the signaling value (as opposed to moral hazard) of the reserve rate by studying how the reserve rate correlates with the default probability conditional on the contract interest rate. Second, transaction in this market takes place online and basically all of the information that lenders observe about the borrowers are also available to the econometrician, unlike in traditional markets. This feature makes us somewhat less concerned about unobserved heterogeneity than in other settings.

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 the lenders may infer that this borrower faces difficulty borrowing from outside sources, which also raises concerns about the repayment ability of the borrower. This will lead lenders to charge high interest to compensate for the high risk. Of course, this intuition is not a complete explanation of signaling, because there needs to be a countervailing force that induces the borrower to post a higher reserve interest rate (otherwise, all borrowers would want to post a low reserve rate). In the market we study, the natural countervailing force is the probability of obtaining a loan. As long as the funding probability increases as a function of the reserve rate, this can counteract the incentive for the borrower to post a low reserve rate. 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. Given a trade-off between higher funding probability and higher interest rate, the heterogeneity in 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) value a decrease in the interest rate on the potential loan relatively more than an 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 relatively more than a decrease in the interest rate. As long as the 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 see whether the reserve interest rate functions as a signal in this market, we begin our analysis by providing results from a series of regressions. 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 it also leads to a more favorable contract interest rate on average even after

controlling for various observables and selection. 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.[4]

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 that those who post high reserve rates are more likely to default than those who post low reserve rates, even after *conditioning on the contract interest rate* (the actual interest rate that the borrower pays on the loans). Given that the reserve rate should not directly affect the borrower's repayment behavior conditional on the contract interest rate, this result suggests that there is informational value in the reserve rate. From the perspective of the lender, this implies that the reserve interest rate is informative about the creditworthiness of the borrower, i.e., the reserve rate is a signal of the borrower's unobserved type.

Given the results of our preliminary analysis, we devote the second part of our paper to developing and estimating a structural model of the online credit market that agrees with the basic findings of the preliminary analysis. 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 listing 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. This feature is very useful, because it allows us to condition on a particular quantile of the borrower's type distribution by simply conditioning on a quantile of the reserve rate distribution. Then the observed default probability at each quantile of the reserve rate distribution nonparametrically identifies the borrower's type. The distribution of the lenders' attitudes toward risk is also nonparametrically

---

[4]Note that it is probably safe to assume that many borrowers are aware of this trade-off: In a prominently displayed tutorial, Prosper informs the borrowers that setting a higher reserve rate increases the probability that the loan will be funded.

identified by relating the expected return of listings to their 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. We simulate the credit supply curve under pooling and no information asymmetry by re-computing the lenders' and borrowers' behavior using the estimates of our structural model. As pointed out by Stiglitz and Weiss (1981), the credit supply curve in loan markets may be backward bending, or non-monotonic in the interest rate, because of adverse selection.[5] The results of our counterfactual support their prediction.

With respect to welfare, we find that the cost of adverse selection can be as much as 31% of the total surplus created under no asymmetric information. We also find that while signaling restores up to 86% of the difference in the surplus between pooling and no asymmetric information in some markets, it destroys welfare in others.

The empirical findings of this paper directly apply only to the market of Prosper.com and our model is tailored to the setting in which agents signal through the reserve rate. However, our basic methodology can be extended to study other markets in which signaling plays an important role, for example, pricing of new issues in IPO markets (e.g., Allen and Faulhaber, 1989), convertible debt recalls (Harris and Raviv, 1985), etc. As long as both the signal and the ex-post performance are observable, our model and identification strategy can be used to quantify the effect of signaling on market outcomes and welfare.

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).[6] 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 researchers to answer questions regarding welfare and market design. Our paper goes in this direction.

The second strand of the literature to which our paper is related is the theoretical literature on

---

[5]Recently, Arnold and Riley (2012) shows that the credit supply curve cannot be globally backward-bending. They obtain this result under the assumption that the type of the borrowers affect the variance of the return but not the mean. In our setting, the type of the borrower affects the mean as well as the variance. Hence, the mean return from a loan can be globally concave in our setting.

[6]Freedman and Jin (2010) uses data from Prosper.com. Other papers that also use the data include Rigbi (2011), Ravina (2008), Iyer et al. (2010). In Iyer et al. (2010), the authors examine the lenders' ability to infer borrowers' creditworthiness. They find, among other things, that the reserve interest rate affects the contract interest rate, and note that signaling can be one interpretation of their finding.

signaling. Starting with the work of Spence (1973), signaling has been applied to a wide range of topics. Even confined to applications in industrial organization, signaling has been applied to advertising (e.g., Milgrom and Roberts, 1986), entry deterrence (e.g., Milgrom and Roberts, 1982), war of attrition (Hörner and Sahuguet, 2011), as well as credit markets (e.g., Bester 1986, Milde and Riley 1988). Bester (1986) shows that borrowers can signal their type through the amount of collateral and Milde and Riley (1988) show that borrowers can signal through the loan amount. 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).[7] The signaling mechanism that we consider in this paper is very similar to those studied in Cai, Riley and Ye (2007) and Jullien and Mariotti (2006).

In contrast to the large body of theoretical work, however, the empirical industrial organization literature on signaling is thin.[8] The few papers that study signaling in industrial organization include Gedge, Roberts and Sweeting (2012), Sahni and Nair (2016), and Backus, Blake, and Tadelis (2016). Gedge, Roberts and Sweeting (2012) study limit pricing in the airline industry in which an incumbent signals its costs through the price. Sahni and Nair (2016) study advertising as a signal of service quality. Backus, Blake and Tadelis (2016) study round prices as a signal of sellers' willingness to cut prices in the context of eBay.

Our paper is also related to the large empirical literature on screening. In particular, Adams, Einav 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 non-monotone in the loan size. A key feature of our paper that is different from theirs is that our paper examines signaling while their paper examines screening. Moreover, our model of credit supply has a large number of heterogenous lenders while their model has a single lender.

A Two-Type Example In order to illustrate how signaling can be sustained in our setting, we begin by presenting a model with just two types of borrowers, φ_L and φ_H ($0 < \varphi_L < \varphi_H$), where φ_L is interpreted as a "bad" type who has a low disutility of default and φ_H is interpreted as the "good" type who has a high disutility of default.[9] We identify the borrower's type with the disutility of default. The types are privately known to the borrowers.

Conditional on obtaining a loan at interest r, we let $u(r) + \varepsilon$ denote the borrower's utility from

---

[7]Relatedly, Roberts (2013) shows how the reserve rate can be used to overcome unobserved heterogeneity in auctions. He studies an environment in which there is informational asymmetry between the players and the econometrician, but there is no asymmetric information between the sellers and the buyers.

[8]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 have tended to focus on testing for the existence of signaling (a few exceptions are Gayle and Golan, 2012, and Fang, 2006).

[9]Note that there are a continuum of borrower types in the model that we take to the data.

repaying the loan and $-\varphi$ ($\varphi \in \{\varphi_L, \varphi_H\}$) denote the utility from default, where ε is a random shock that realizes after the borrower obtains the loan. The borrower's utility from obtaining a loan at interest r is then $E[\max\{u(r) + \varepsilon, -\varphi\}]$. We let $\lambda(\varphi)$ be the borrower's utility from not borrowing and assume that $\lambda(\varphi_H) > \lambda(\varphi_L)$. This assumption simply reflects the idea that "good" types who value their credit history, for example, have an easier time obtaining a loan from outside sources and hence have a higher outside option. The borrower chooses a reserve interest s at the time the borrower posts a listing.

As for the lender, we let the lender's utility from lending money be a function of the mean (μ) and variance (σ^2) of the return as $\mu - A_j \sigma^2$, where A_j is the lender specific random variable that determines her attitude toward risk. In this example, we assume that every borrower is randomly matched with two lenders each of whom decides whether or not to bid on the listing, and at what interest. The lender observes the reserve interest chosen by the borrower when making her decision. The contract interest is determined by a second price auction, i.e., the contract interest is equal to the second lowest interest if two bidders decide to bid on the loan, and it is equal to the borrower's reserve rate if only one lender bids on the loan. We assume, in this example, that whoever bids a lower interest rate becomes the sole lender. In this setting, the lender has a weakly dominant strategy: If we let r_j denote the interest rate at which lender j is indifferent between lending and not lending, the weakly dominant strategy for the lender is to bid r_j if $r_j \leq s$ and not bid otherwise.

In this simple example, a signaling equilibrium consists of two different reserve interests s_H and s_L such that φ_H prefers s_H to s_L and vice versa.[10] Table 1 and Figure 1 illustrate one signaling equilibrium for a particular parameterization of the primitives; $u(r) = -r$, $\lambda(\varphi) = 0.105\varphi$, A_j distributed uniform $[0, 1]$, ε distributed logit, and $\varphi \in \{1.7, 1.8\}$. In Table 1, we summarize the outcomes and utility levels associated with $s_H = 27.5\%$ and $s_L = 36\%$ for both types of borrowers when lenders believe that s_H signals high type and s_L signals low type. If a borrower posts s_H, the funding probability is relatively low (35.5%) but the average interest rate is relatively favorable (27.18%). On the other hand, if a borrower posts s_L, the funding probability is high (44.5%), but the average interest rate is relatively unfavorable (35.2%). For φ_H types, the favorable interest rate at s_H more than compensates for the low funding probability. Hence, the expected utility of choosing s_H (0.300) is higher than the expected utility of choosing s_L (0.299) for φ_H types. On the other hand, the expected utility of choosing s_L (0.302) is higher than the expected utility of choosing s_H (0.300) for φ_L types. Hence, reserve interests $s_H = 27.5\%$ and $s_L = 36\%$ can be supported as a separating equilibrium.

In Figure 1, we illustrate this equilibrium graphically. The horizontal axis of the figure corre-

---

[10]To be complete, one needs to specify the off-path beliefs of the lenders. One possibility is to assume that lenders belive that $\varphi = \varphi_L$ upon observing $s \neq s_H$: $\Pr(\varphi = \varphi_L | s \neq s_H) = 1$.

Reserve Rate	Lender Belief ($\hat{\varphi}$)	Funding Pr.	Interest Rate	Utility for φ_H	Utility for φ_L
$s_H = 27.5\%$	φ_H	35.5%	27.2%	0.300	0.300
$s_L = 36\%$	φ_L	44.5%	35.2%	0.299	0.302

Table 1: Outcomes and Utility Associated with Choosing s_H (27.5%) and s_L (36.0%). In this Table, we show the outcomes and utility levels that correspond to choosing s_H or s_L for each type of borrower.
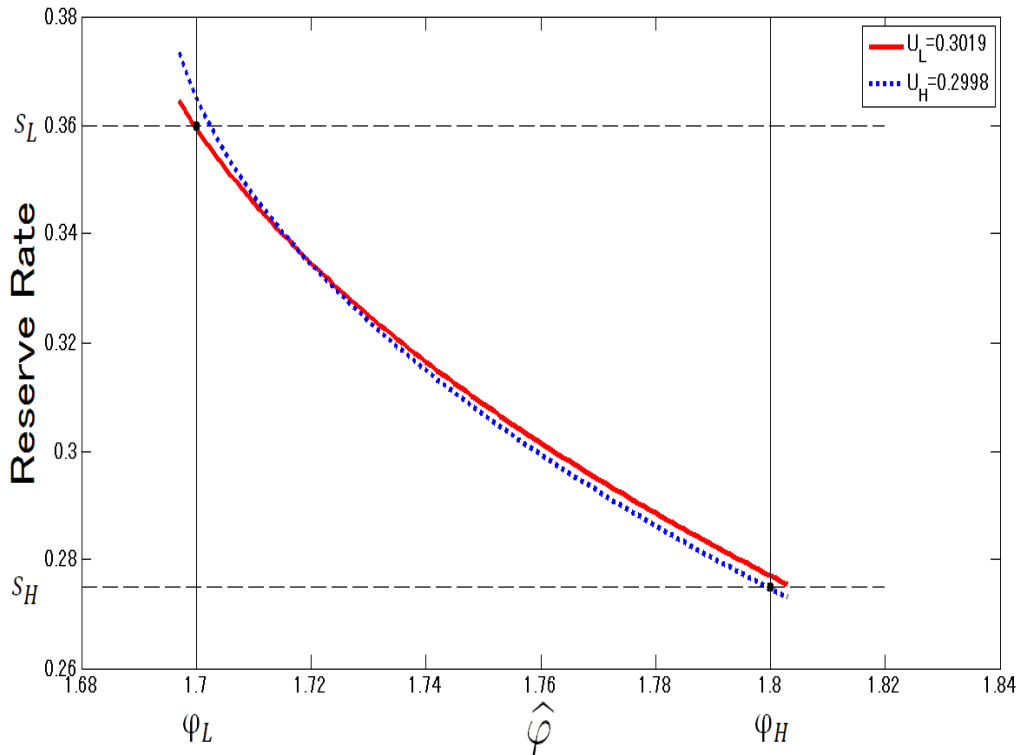

Figure 1: Illustration of a Separating Equilibrium: The dotted curve in the figure corresponds to the indifference curve of φ_H and the solid curve corresponds to the indifference curve of φ_L. The borrowers' utility level is higher to the northeast of the indifference curve. The horizontal axis corresponds to the beliefs of the lender ($\hat{\varphi}$) and the vertical axis corresponds to the reserve rate.

sponds to the beliefs of the lender ($\hat{\varphi}$) and the vertical axis corresponds to the reserve rate (s). The dotted curve in the figure corresponds to the indifference curve of φ_H at $(\hat{\varphi}_H, s_H)$ and the solid curve corresponds to the indifference curve of φ_L at $(\hat{\varphi}_L, s_L)$. The borrowers' utility level is higher to the northeast of the indifference curve.[11] The figure shows that a separating equilibrium can be supported in which type-φ_H chooses point s_H, while type-φ_L chooses s_L.

In the full model which we present in Section 4, the types of the borrowers are continuously distributed and the borrowers' decision to repay or default is modeled explicitly as a single-agent dynamic programming problem. Moreover, the lenders bid both an interest rate and an amount consistent with the actual way in which Prosper operates. The number of lenders are also random. However, our two type example captures the basic forces behind the signaling equilibrium. There is an inherent trade-off between the funding probability and the interest rate. Given that different types evaluate this trade-off differentially, it is possible to sustain a signaling equilibrium.

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 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 determine the interest rate. For details on other aspects of Prosper, see Section 1 of the online appendix and 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, and (4) 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 first creates an account with Prosper, who pulls the applicant's credit history from Experian, a third-party credit-scoring agency. As long as the credit score is above a certain threshold, the borrower can create a listing on Prosper's web site. Each listing contains information regarding the amount of loan requested, the reserve interest rate and the borrower's characteristics. The loan amount and the reserve interest are both variables that the borrower chooses, subject

---

[11]The utility level is higher to the north because raising s increases the probability of obtaining a loan. At the current parameter values, the marginal gain from increasing the funding probability outweighs the marginal loss from increasing the contract interest rate for both types.

to Prosper's conditions and usury laws.[12] During our sample period, the maximum loan amount allowed on Prosper was $25,000 and the usary law maximum was 36%.

The characteristics of the borrower that appear in the listing page include credit grade, home-ownership status, debt-to-income ratio, purpose of the loan, as well as other additional information such as 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 verified by Prosper.[13] Other information, such as debt-to-income ratio and purpose of the loan, is provided by the borrower without verification by Prosper. The most commonly stated purpose of the loan is debt consolidation, accounting for about 46% of all listings.

2. Lenders Bid   Prosper maintains a list of active listings on its Web site for potential lenders. If a potential 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 lender can bid on any active listing at any time.

For each active listing, Prosper displays the fraction of the loan funded and the active interest rate in addition to information regarding borrower characteristics, loan amount, and the reserve interest rate. The active interest rate corresponds to the standing marginal bid in multi-unit auctions. We will explain what the active interest is, in more detail below.

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 bidders become lenders. 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. 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 (e.g., 80% for the left panel in Figure 2 and 100% for the right panel). For listings that have yet to attract enough bids to reach the requested amount (i.e., fraction of loan funded is less than 100%), the lender does not observe the interest rate of any other bid. As for listings that have already received enough bids to cover the requested amount, (i.e., fraction of loan funded is equal to 100%) the lender observes the active interest rate, which is the interest rate of the marginal bid that brings the supply of money

---

[12]In a tutorial that walks borrowers through the listing process, Prosper advises borrowers to "Think of the interest your are paying on your next best alternative" when posting the reserve insterst rate.

[13]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.
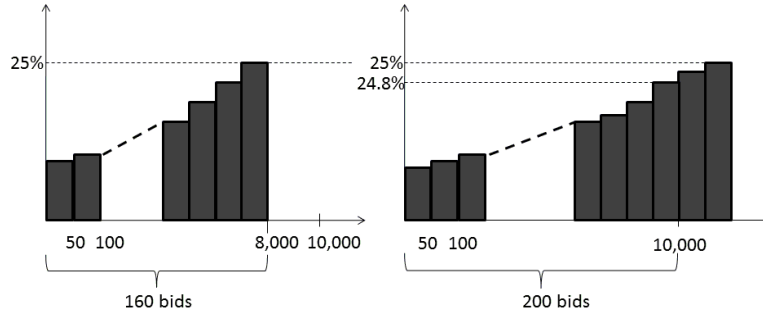
Figure 2: Funding Decision – The figure shows how a loan is funded for the simple case in which lenders only submit a bid with an amount of $50. The horizontal axis corresponds to amount and the vertical axis corresponds to the interest rate. The left panel illustrates a situation in which the requested 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.

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 that are still active, 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. *However, there are no partial loans for listings that have failed to attract enough bids to fund the total requested amount.* Hence the borrower would receive no loan in the situation depicted in the left panel of Figure 2.[14]

As for fully 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 2, 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.

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. 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.[15] 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.

---

[14]In our sample, about 20% of the borrowers whose loans are not funded relist on Prosper. Interestingly, borrowers that relist on Prosper do not adjust the reserve interest rate by much. The median borrower who relists does not change the reserve rate at all. The average change in the reserve rate is about 1.2%.

[15]The proceeds, which are about 1.05 cents on the dollar, on average, are distributed to the lenders. Our sample statistic and the regressions account for these proceeds.

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

Table 2: Descriptive Statistics – Listings: 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.

2.2   Data

The data for our analysis come directly from Prosper.com. The data set 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.[16] We also have monthly repayment data of the borrowers.

Our data consist of all listings that were created from May to October of 2008 (and the corresponding loan repayment data for funded listings which go until the end of 2011). Note that all loans in our sample have either matured or 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. We are left with a total of 35,241 listings, of which 5,571 were funded. Our online appendix contains a more detailed description of our data construction.

Table 1 reports sample statistics of the listings by credit grade. The mean requested amount is reported in the first column, and it ranges from a high of more than $13,000 for AA listings to a low of less than $5,000 for HR listings. 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.

In Figure 3, we present the distribution of the reserve rate across different credit grades. As expected, the reserve rate is higher for worse 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

---

[16]The only piece of information missing is the conversation that takes place between borrowers and potential lenders through the Prosper Web site.
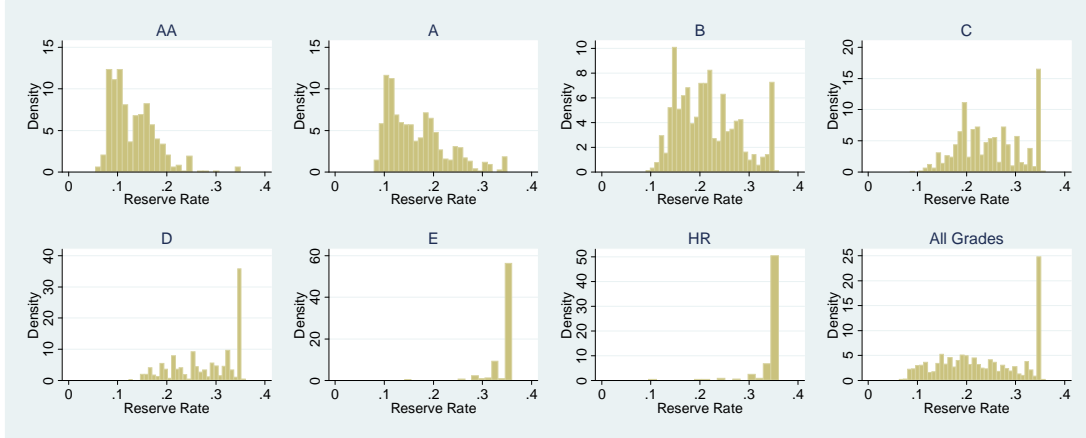
Figure 3: Distribution of Reserve Interest Rate by Credit Grade – We show the distribution of reserve interest rate by credit grade. The reserve interest rate is capped at 36% because of the usury law.

maximum for our sample. 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 among 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 (AA, A, B and C) in presenting some of our analysis below.

In Figure 4, we plot the cumulative distribution functions of the contract interest rate (r) conditional on the reserve interest rate (s), by credit grade. In the top left panel, we report two distributions for credit grade AA, one corresponding to $s = 10\%$ and the other corresponding to $s = 20\%$.[17] The other panels of Figure 4 plot similar distributions for credit grades A – HR and for $s = 10\%$, 20%, 30% and 36%. The figure shows that there is a stochastic dominance relationship between the distributions of contract interest rate given s' and s'' whenever $s' < s''$. This suggests that borrowers who post low reserve rates are more likely to obtain low interest. In the next section, we show that this relationship is true even after controlling for the fact that r is right-censored at the reserve rate.

In Figure 5, we report the distributions of the bid amount, again by credit grade. 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 the lenders bid either \$50 or \$100 for a given loan. 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 a discrete set of possible loan amounts rather than from a continuous set.

---

[17]Given that almost no borrowers in credit grade AA post a reserve rate of 30% or more (see Figure 3), we only report the distribution for s equal to 10% and 20% for credit grade AA.
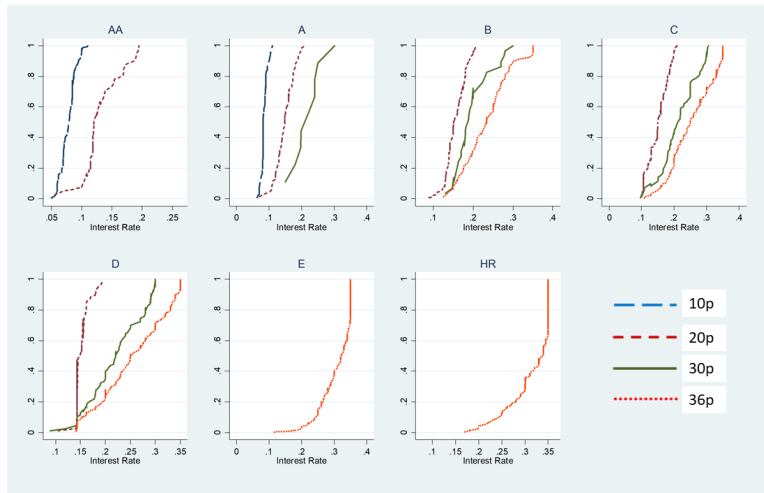
Figure 4: Distribution of Contract Interest Rate Given Reserve Interest Rate by Credit Grade: The figure plots the distribution of contract interest rate conditional on reserve interest rate $s = 10\%, 20\%, 30\%$, and 36%. The distribution for $s = 10\%, 20\%, 30\%$ are computed by pooling funded listings with $s \in [s-1, s+1]$. For $s = 36\%$, the distribution is computed by pooling funded listings with $s \in [34\%, 36\%]$.
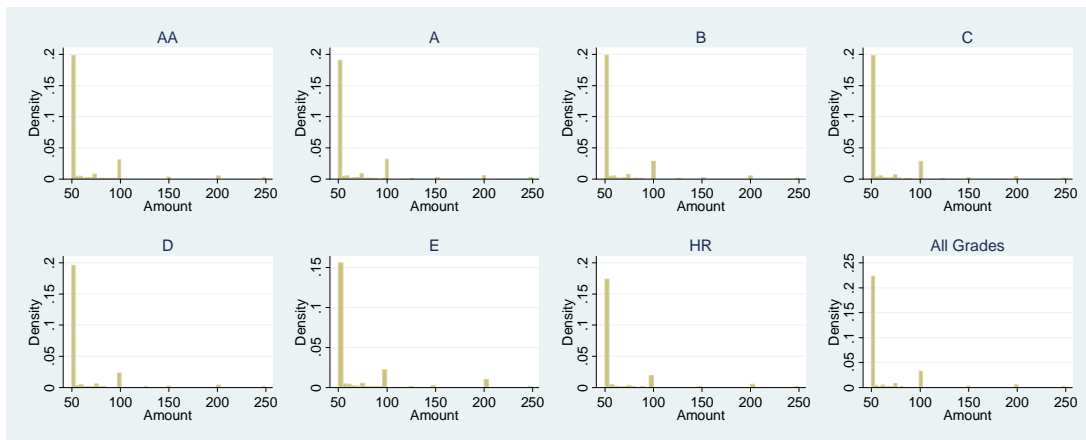

Figure 5: Distribution of Bid Amount – The figure shows the distribution of bid amount for each credit grade. Bids with amount exceeding \$250 are not shown. The fraction of these bids is about 3.5%.

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

Table 3: Descriptive Statistics – Loans: This table reports the summary statistics of loans. 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 reports sample statistics of listings that were *funded*, which is a subset of the set of listings. Note that the mean loan amount reported in Table 3 is smaller than the mean requested amount shown in Table 2, 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 3 is higher than in Table 2, for the obvious reason that listings need to attract sufficient number of bids to get funded: Recall that there is no partial funding for listings that fail to attract enough bids to cover the requested amount.

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. In the first column of Table 4, we report sample statistics regarding the default probability by credit grade. The average default probability is lowest for AA loans at 14.9%, while it is highest for HR loans at 43.9%. Table 4 also reports the mean and the quantiles of the internal rate of return (IRR) of the loans.[18] 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 our sample coincides with the period of economic downturn during the financial crisis. (Note that the return on the S&P was at -37% during 2008).[19] It may also reflect the fact that lenders were not fully aware of the creditworthiness of the pool of borrowers on Prosper.

Finally, Table 5 reports the summary statistics of the lenders. We find that lenders fund, on

---

[18]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. In Table 4, we report the annualized IRR.

[19]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 estimate of 1.1% is conservative because some lenders had not finished repaying by the day we retrieved our data.

Grade	Default Prob.	Mean IRR	sd	10%	25%	50%	75%	90%	Obs
AA	0.149	-0.011	0.283	-0.449	0.061	0.082	0.110	0.132	755
A	0.211	-0.025	0.331	-0.767	0.072	0.094	0.135	0.181	755
B	0.297	-0.074	0.404	-0.871	-0.229	0.136	0.169	0.211	1,023
C	0.309	-0.060	0.413	-0.871	-0.211	0.135	0.196	0.256	1,285
D	0.321	-0.036	0.424	-0.865	-0.192	0.153	0.231	0.316	1,022
E	0.372	0.000	0.475	-0.861	-0.315	0.249	0.345	0.394	392
HR	0.439	-0.112	0.532	-0.886	-0.800	0.202	0.345	0.398	339
All	0.286	-0.046	0.402	-0.858	-0.100	0.121	0.187	0.281	5,571

Table 4: Descriptive Statistics – Default Probability and Internal Rate of Return (IRR): This table reports the default probability and IRR of the loans originated by Prosper. We present the average IRR, the standard error, and the quantiles of the distriubution.

	Mean	sd	5%	25%	50%	75%	95%	Obs
Portfolio Size (# of Loans)	17.5	42.4	1	2	6	16	68	$29,176$
Portfolio Size ($)	$1,325$	$5,314$	50	126.6	350	1000	4950	$29,176$
Average IRR	-0.041	0.166	-0.332	-0.076	-0.005	0.039	0.144	$29,176$

Table 5: Descriptive Statistics – Lender Portfolio Characteristics: This table reports the summary statistics of the lender's portfolio. We present the number of loans that a lender owns in her portfolio, the total amount of loans lent by the lender, average IRR of loans in a portfolio, the standard deviation of IRR of the loans within a portfolio for the full sample, and the standard deviation of IRR of the loans within a portfolio for the lenders who own more than 10 loans.

average, 17.5 loans with a total portfolio size of about $1,325. The median lender funds 6 loans and a total of $350. This suggests that the lenders do not invest a lot of money on Prosper, in general. The mean IRR of a lender's portfolio is about -4.1% with a standard deviation of 16.6%.

3   Evidence of Signaling Through the Reserve Rate

In this section, we provide some reduced-form evidence that the borrower's reserve interest rate serves as a signaling device. In particular, we first show evidence that suggests that raising the reserve rate (1) increases the funding probability; (2) increases the contract interest rate; and (3) increases the default probability. We next argue that, taken together, these results suggest that the reserve rate serves as a signal.

While the baseline results that we present below are based on a relatively parsimonious specification of the reduced form, the online appendix contains results from richer specifications with interactions of covariates as well as specifications with additional covariates, such as text information and more detailed credit information of the borrowers. The online appendix also contains regression results for a subsample of listings in which we include a measure of creditworthiness as perceived by online survey respondents. The baseline results we report below are broadly consis-

tent with the results of these alternative specifications.

Funding Probability   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 and $\mathbf{x}_j$ is a vector of controls that include the requested amount, the debt–to–income ratio, a dummy variable for home ownership, the credit grade, calendar month, and hour of day the listing was created.

The first column of Table 6 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. In terms of the marginal effect, a 1 percentage point increase in s_j is associated with about a 0.32 percentage point increase in the funding probability.

Contract Interest Rate   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, s_j is the reserve rate, and $\mathbf{x}_j$ is the same vector of controls as before. The first equation relates the latent contract interest rate to the reserve rate and other listing characteristics. r_j^* is interpreted as the latent interest rate at which the loan is funded in the absence of any censoring. The second equation is the censoring equation, which accounts for the fact that the contract interest rate r_j is always less than the reserve rate, s_j. Note that if we were to run a simple 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.

We report the results from this regression in the second column of Table 6. As reported in the first row, we estimate β_s to be positive and significant, which seems to suggest that a lower reserve interest rate leads to a lower contract interest rate, consistent with our hypothesis. As we discuss next, borrowers who post high reserve rates are relatively less creditworthy. If we take this as given, the results of regression (2) seem to suggest that lenders charge higher interest to riskier borrowers.

	(1) Funded	(2) Contract Rate	(3) Default	(4) Rate of Return
Reserve rate	2.1368***	0.6834***	1.2324***	−0.5919***
	(0.0263)	(0.0145)	(0.3172)	(0.1313)
Contract rate			1.4085***	0.0540
			(0.3031)	(0.1372)
Amount (in $1000)	−0.1070***	0.0077***	0.0122***	−0.0045***
	(0.0024)	(0.0002)	(0.0031)	(0.0013)
Debt / income	−0.7971***	0.0731***	0.0090	−0.0314
	(0.0015)	(0.0037)	(0.0443)	(0.0197)
Home owner	−0.1513***	0.0137***	0.0529	−0.0471***
	(0.0004)	(0.0018)	(0.0284)	(0.0117)
Grade				
AA	3.6468***	−0.3013***	−0.0958	0.0595
	(0.0044)	(0.0061)	(0.0968)	(0.0402)
A	3.0727***	−0.2670***	−0.0740	0.0475
	(0.0033)	(0.0055)	(0.0846)	(0.0366)
B	2.5681***	−0.2347***	−0.0239	0.0224
	(0.0022)	(0.0046)	(0.0688)	(0.0320)
C	1.8743***	−0.1862***	−0.0309	0.0380
	(0.0014)	(0.0038)	(0.0592)	(0.0288)
D	1.2754***	−0.1329***	−0.0579	0.0636**
	(0.0011)	(0.0034)	(0.0550)	(0.0272)
E	0.5022***	−0.0499***	−0.1560*	0.1155***
	(0.0014)	(0.0036)	(0.0608)	(0.0296)
Observation	35,241	35,241	91,939	5,571
R^2	0.2827			0.0224
Likelihood			−1,137	−4,805

Table 6: Reduced Form Analysis - Funding Probability, Contract Interest Rate and Repayment Behavior: 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. The third 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. For regression (3), we report the results after 100 iterations. The last column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. 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 in all of the regressions. Standard errors are robust-heteroskedasticity-consistent and clustered at the loan level. They are presented in parentheses below the coefficients.

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 for the mean holds for different quantiles. The results of the quantile regressions are qualitatively similar. The results seem to imply that $F(r^*|s)$ first order stochastically dominates $F(r^*|s')$ for $s \geq s'$ (See also Figure 3).

The results of regressions (1) and (2) 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. Note that it is probably safe to assume that many borrowers are actually aware of this trade-off: In a prominently displayed tutorial, Prosper informs the borrowers that setting a higher reserve rate increases the probability that the loan will be funded. Given the dispersion in the reserve rate (See Figure 2), 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.

Repayment Behavior 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 Probit regressions of an indicator of default on observable characteristics of the loan as well as the reserve rate:

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

The first regression is a panel Probit in which the regressand, Default_{jt}, is a dummy variable that takes a value of 1 if borrower j defaults on the loan at period t.[20] We include period-t dummies, μ_t, and a borrower random-effect, α_j, in regression (3). 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 (r_j) in the regression, the effect captured by β_s is purely due to selection. In other words, β_s is not picking up the effect of moral hazard given that the reserve rate should not directly affect the behavior of the borrower once we condition on the contract interest rate.

The parameter estimates obtained from these regressions are shown in column (3) of Table 6. The coefficient associated with the reserve interest rate is positive and significant in column (3). In terms of the marginal effect, a 1 percentage point increase in s_j is associated with about a 1.23 percentage point increase in the monthly default probability in columns (3). This implies that

---

[20]In the online appendix, we also estimate the model in which the dependent variables is a dummy for ever default, which takes the value of one if borrower i ever defaults.

borrowers who post higher reserve interest rates tend to default more often, which is consistent with the notion that the reserve rate is informative about the type of the borrowers, i.e., the reserve interest rate can be used as a signal of the creditworthiness of the borrower. In the second and third rows, 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 default. The positive coefficient on the amount can be a result of either signaling or moral hazard. 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 signaling, and the latter would be consistent with moral hazard.[21]

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. As with our discussion of regression (3), the coefficient on s_j captures the pure selection effect given that we control for the contract interest rate in the regression.

The parameter estimates obtained from this regression are shown in the last column of Table 6. As expected, the coefficient on s_j is negative and significant ($\beta_s = -0.59$). This is consistent with the result of regression (3), where we examined the relationship between r_j and the default probability. The coefficient on r_j is positive, but not statistically significant. This may be due to moral hazard.[22]

Interpretation of the Results   Taken together, our regression results seem to indicate that (1) reserve rate choice involves a trade-off between a larger funding probability and a higher contract interest rate; (2) borrowers are heterogeneous with respect to how they evaluate this trade-off; (3) those who post high reserve rates tend to be less creditworthy; and (4) the lenders anticipate this

---

[21]One may wonder why the coefficient of home ownership is negative. One hypothesis is selection. It may be the case that most homeowners have easy access to credit outside of Prosper. Conditional on having to borrower on Prosper, homeownership may be a negative signal of the borrower's creditworthiness. Another possible explanation is that house prices depreciated substantially during the sample period. This may also explain why homeownership comes out as a bad signal. We note that there are other studies in similar contexts that also do not find homeownership to be a good signal of the borrower. For example, Adams, Einav and Levin (2009) finds that homeownership increases missed payments on subprime auto loans in one of their specifications.

[22]In our structural model, borrower repayment captures this ex-post moral hazard through the borrower's period utility function.

and charge higher interest to borrowers who post high reserve rates.

These results are informative about how signaling is sustained in equilibrium. Uncreditworthy borrowers are more willing to sacrifice a favorable interest rate for a bigger probability of being funded. Hence these borrowers prefer {high interest, high probability of receiving a loan} to {low interest, low probability of receiving a loan} while creditworthy borrowers prefer the latter to the former. This heterogeneity in preferences prevents either types from mimicking the other types and sustains a separation of types through signaling.

While the results that we present in this section correspond to relatively parsimonious specifications of the reduced form, the results are quite robust. In particular, the coefficient on the reserve rate remains very strong and significant across a range of alternative specifications. For example, we obtain similar results when we include more variables regarding the financial state of the borrower (bank card utilization, number of credit lines, etc.) or when we restrict the sample to the set of listings with a particular loan purpose (debt consolidation, paying for school etc.). Finally, we also conducted a small-scale survey to control for listing characteristics that are hard to include in regression form, such as text and pictures. The survey respondents were asked to rate listings based on pictures and text information provided by the borrowers. We find that the measure obtained in the survey is informative, but that the inclusion of the measure in the regressions does not weaken the explanatory power of the reserve rate. Hence, it is unlikely to affect our interpretation that the reserve rate is a signal of the borrower's creditworthiness. Section 3 of the online appendix contains more details about the experiment and the results.[23]

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 has three parts. The first part of the model pertains to the reserve interest rate choice of the borrowers, the second part to the lenders' bidding behavior and the third part to the borrowers' repayment behavior. We collect and discuss our modeling choices and assumptions at the end of this section.

---

[23]In addition, there are papers using additional graphical and textual data that report similar effect of the reserve rate on various outcome variables. For example, Ravina (2008) augments the Prosper data with the perceived attractiveness of the borrowers using the photos that borrowers post and Freedman and Jin (2010) includes variables such as social ties of the borrower, etc. Their findings are reassuring in the sense that inclusion of these additional variables does not change much the estimated coefficients of the reserve rate (see Table 5 of Freedman and Jin, 2010 and Table IV of Ravina, 2008). It seems very unlikely that unobserved heterogeneity explains away all of the correlation between the reserve rate and outcome variables.

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. We write the terminal decision of the borrower at period T as follows:

$$
\begin{cases}
\text{full repayment: if } u_T(r) + \varepsilon_T \geq D(\varphi) \\
\text{default: otherwise,}
\end{cases}
\tag{5}
$$

where $u_T(r) + \varepsilon_T$ denotes the period utility of the borrower if he repays the loan in full, r denotes the interest rate on the loan, and φ denotes the (unobserved) type of the borrower that determines the likelihood of repaying the loan. While there are many ways to interpret φ and $D(\cdot)$ – e.g., as unobserved liquid asset/wealth of the borrower – we adopt the interpretation of $D(\varphi)$ as the default cost of a borrower whose type is equal to φ.[24] That is, the borrower compares the utility of repaying the loan $(u_T(r) + \varepsilon_T)$ with the cost of default $(D(\varphi))$, choosing to repay the loan if and only if the former is greater than the latter.

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. Hence, borrowers with high φ are "good" types who value avoiding default and maintaining a good credit history.

We assume φ to be independent of ε_T, conditional on observables. The conditional independence of ε_T and φ may appear to be a very strong assumption, but mean independence is actually without loss of generality. To see this, if $E[\varepsilon_T|\varphi] \neq 0$, we can subtract $E[\varepsilon_T|\varphi]$ from both sides of equation (5) and by appropriately redefining $D(\cdot)$ and ε_T, we have an observationally equivalent model with $E[\varepsilon_T|\varphi] = 0$. This is possible because we allow $D(\cdot)$ (or equivalently, the distribution of φ) to be nonparametric.[25] While mean independence is not the same as independence, we think that this alleviates some of the concerns regarding our assumption. We come back to this point at the end of this section (Section 4.4).

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:

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

---

[24]φ is the unobserved type of the borrower that affects the propensity to make repayments. The exact interpretation of φ is not very important for our purposes. In the online appendix, we provide an observationally equivalent model in which φ is interpreted as wealth/asset of the borrower.

[25]Intuitively, one can think of this as loading all of the "systematic" component on φ. In other words, if $E[\varepsilon_T|\varphi] \neq 0$, we can load on φ the part of ε_T that is correlated with φ.

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 allow u_t to depend on t in order to capture any deterministic time dependence while we assume $\{\varepsilon_t\}$ to be i.i.d across t and mean zero.

We have presented the model up to now without making explicit the dependence of the primitives of the model on observable borrower/listing characteristics such as the credit grade. 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.

Borrower Reserve Rate Choice   Now 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 its effect on the probability that the loan is funded, and its effect on the contract interest rate, r. The borrower's problem is then to choose s, subject to the usury law limit of 36%, as follows:

$$\max_{s \leq 0.36} V_0(s, \varphi) = \max_{s \leq 0.36} \left[\Pr(s) \int V_1(r, \varphi) f(r|s) dr + (1 - \Pr(s))\lambda \right], \qquad (6)$$

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 λ 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. Borrowers with high values of λ have good outside option, e.g., borrowing money from relatives, friends, and local banks, etc. Borrowers with low values of λ have bad outside option. We suppress the dependence of $\Pr(s)$ and $f(r|s)$ on the characteristics of the borrower. Although $\Pr(s)$ and $f(r|s)$ are equilibrium objects, they are taken as exogenous and known by the borrower.

Note that an important choice variable for the borrower that we do not model is the loan amount. We treat the loan amount as part of the set of conditioning variables. Given that the borrowers' reserve rate choice has to solve equation (6) conditional on the optimal choice of the loan amount, treating the loan size as a covariate does not bias our estimates. To the extent that the loan size has a signaling effect, we will be able to pick this up directly when we estimate the distribution of types conditional on borrower covariates. We come back to this point at the end of this section.

The first term in the bracket in equation (6) 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 $t = 1$, is integrated against the distribution of the contract interest rate $f(r|s)$. The second term 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, λ.

In what follows, we assume that φ and λ are related as $\lambda = \lambda(\varphi)$, where $\lambda(\cdot)$ is an increasing

function of φ, where φ is the private type of the borrower we defined earlier. This assumption 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 $\lambda(\varphi)$.

The first–order condition associated with problem (6) is as follows,

$$\frac{\partial}{\partial s} \Pr(s) \left(\int V_1(r, \varphi) f(r|s) dr - \lambda(\varphi) \right) + \Pr(s) \int V_1(r, \varphi) \frac{\partial}{\partial s} f(r|s) dr = 0, \qquad (7)$$

for an interior solution. Equation (7) captures the trade-off 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 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 $s < 0.36$. 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 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.,*

$$\frac{\partial^2}{\partial s \partial \varphi} V_0(s, \varphi) < 0.$$

Proof. See Appendix. ∎

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. As we explained above, $\frac{\partial}{\partial s} V_0(s, \varphi)$ 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 and, conditional on default, the borrower does not have to bear the full cost of a high interest rate.

Before turning to the lenders' model, we briefly discuss the optimal reserve rate choice of the borrowers when the usury law limit is binding. Recall from our discussion of Figure 3 that there

is a non-negligible mass at exactly 36% for credit grades B and below, implying that the usury law maximum is a binding constraint for many borrowers in these credit grades. For credit grades B and C, the pattern in the data seem broadly consistent with partial pooling, i.e., separation of types below 36%, and pooling at 36%. For there to be partial pooling, 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 online appendix. For listings in credit grades B and C, we include them in the estimation of the model accounting for the fact that there is separation of types below 36%, and some pooling at 36%. For credit grades D and below, an even larger fraction of the borrowers submit a reserve interest rate at the usury law maximum, leaving little variation in the reservation interest rate. This means that data from these categories are not very informative about the signaling value of the reserve rate. Hence in our estimation, we only focus on 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. We let F_N denote its cumulative distribution function with support $\{0, 1, \cdots, \bar{N}\}$, where $\bar{N}$ 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 must decide 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 in addition to various characteristics of the listing, such as the reserve rate. 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, or $100. In the estimation, we assume that lenders face a discrete set of amount $M = \{\$50, \$75, \$100, \$200, \$500, \$1000\}$ to choose from.

Lender's Problem with No Amount Choice   We first describe the case when the lender can only bid $50, i.e., $M = \{\$50\}$, so that the lender's decision is whether to bid or not and what interest rate to bid. We later extend the model to the case with amount choice. Before the lender can decide what to bid, the lender must first form beliefs over the return she will make if she funds a part of the loan. Given that the average return from funding loans on Prosper was negative for this sample period, we do not want to impose rational expectations. We allow the lenders' beliefs to be different from the actual realized distribution of returns, albeit in a very simple way.

Following the standard specification used in the asset pricing literature, 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 lends to listing Z at contract interest rate r, as follows:

$$\begin{aligned} U &= \tilde{U}_j^L(Z(r)) - \tilde{\varepsilon}_{0j} \\ &\text{where } \tilde{U}_j^L(Z(r)) = \tilde{\mu}_j(Z(r)) - A_j \tilde{\sigma}_j^2(Z(r)) - c. \end{aligned}$$

$Z(r)$ is the random return from investing in Z at rate r, and $\tilde{\mu}_j(Z(r))$ and $\tilde{\sigma}_j^2(Z(r))$ are lender j's expectation of the return and variance. A_j is a lender specific random variable known only to lender j that determines her attitude toward risk and c and $\tilde{\varepsilon}_{0j}$ are deterministic and random opportunity costs of lending to listing Z. If we express $\tilde{\mu}_j$ and $\tilde{\sigma}_j^2$ as deviations from the mean and variance computed under rational expectations, we can rewrite the previous expression as follows:

$$\begin{aligned} U &= U_j^L(Z(r)) - \varepsilon_{0j} \\ &\text{where } U_j^L(Z(r)) = \mu(Z(r)) - A_j \sigma^2(Z(r)) - c \\ &\qquad \varepsilon_{0j} = \tilde{\varepsilon}_{0j} + U_j^L(Z(r)) - \tilde{U}_j^L(Z(r)), \end{aligned}$$

where ε_{0j} now includes lender forecasting error as well as $\tilde{\varepsilon}_{0j}$. $\mu(Z(r))$ and $\sigma^2(Z(r))$ are the expected return and variance computed under rational expectations.

For our baseline results, we make two important assumptions, which are (1) ε_{0j} does not depend on r; and (2) A_j and ε_{0j} are independent. The two assumptions are very convenient because the lender's model is then isomorphic to a model in which lenders have rational expectations. While these are strong assumptions, they still allow, for example, the lenders to be optimistic about the expected return. This would be the case if $\mu_j(Z(r)) = \mu(Z(r)) + \varepsilon_{\mu j}$ and $\varepsilon_{\mu j}$ has positive mean. However, the independence assumption implies that the lenders' beliefs over the variance coincide with the realized variance. We relax the independence assumption by allowing μ and σ^2 to depend on S&P 500 index and show the results in the online appendix (see Section 7 in the online appendix).

Note that the mean and variance depend on the characteristics of listing Z, such as the reserve interest rate, loan amount, credit grade, etc., in addition to r. The characteristics of the loan other than r are suppressed to simplify notation. Also, in order to account for the possibility that lenders' beliefs change over time, the mean of ε_{0j} is allowed to change each month. We also suppress this dependence.

In order to understand the lender's problem, it is useful to study it graphically. Figure 6 is a graphical representation of the lender's problem. In the left panel of this figure, we take σ^2 to be the horizontal axis and μ to be the vertical axis. Now, consider a listing Z. For each realization of the contract interest rate, consider the mean return, $\mu(Z(r))$, and the variance of the return, $\sigma^2(Z(r))$. Note that we can plot the points $(\mu(Z(r)), \sigma^2(Z(r)))$ on this $\mu - \sigma^2$ plane for each r.
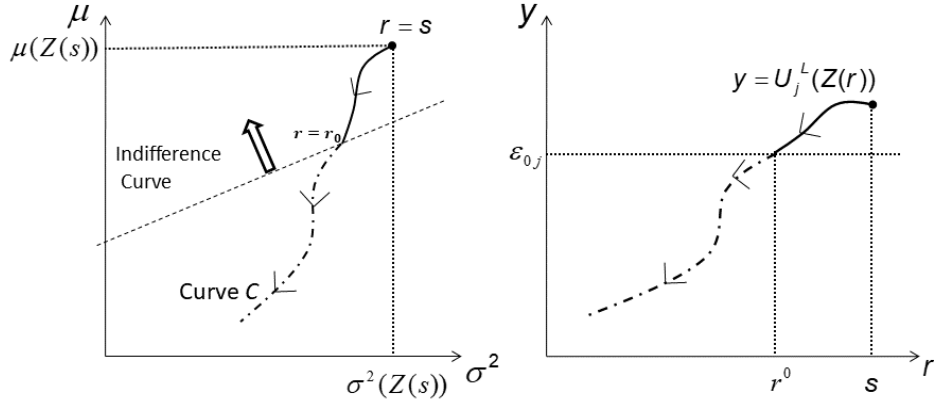
Figure 6: Graphical Representation of the Lender's Problem: Case of No Amount Choice – The figure illustrates how the lender should bid when there is no amount choice. In the left panel, the horizontal axis is σ^2 and the vertical axis is μ. For each listing and for each realization of the contract interest 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 $U_j^L(Z(r))$ against r.

Curve C in the left panel of Figure 6 illustrates the possible mean and variance for a given listing. The starting point of the curve corresponds to the return and variance associated with the case when the listing is funded at the reserve rate, so that $r = s$. As the contract rate is bid down from s, the corresponding point on the $\mu - \sigma^2$ plane changes, and this is shown as a movement along Curve C in the direction of the arrows. Note that we have also drawn a dashed line in the left panel of Figure 6. This is the lender's indifference curve, i.e., the set of points that makes the lender 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., $\mu - A_j\sigma^2 - c - \varepsilon_{0j} = 0$. Any point above this line gives the lender a strictly higher utility than the outside option, and vice versa. The lender is exactly indifferent between lending money and not lending money when the contract interest rate is r^0.

In the right panel of Figure 6, we plot the utility of the lender, $U_j^L(Z(r))$, as a function of r. 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. At $r = r^0$, the lender is indifferent between lending and not lending, which is reflected in the fact that $U_j^L(Z(r))$ crosses ε_{0j} at r^0. Note that as drawn in the figure, Curve C intersects with the lender's indifference curve only once, or equivalently, $U_j^L(Z(\cdot))$ crosses ε_{0j} just once. The analysis for the case in which Curve C intersects with the indifference curve multiple times is more or less the same so we assume it away to simplify exposition. Our Proposition 3, which covers the case with lender amount choice is general enough to allow for multiple intersections.

We now claim the following: Under the assumption that the lender behaves as if she is never pivotal (i.e., never marginal), and that $U_j^L(Z(\cdot))$ crosses ε_{0j} just once, bidding r^0 is a (weakly)

dominant strategy for the lender. That is, it is optimal for the lender to bid an interest rate that makes the lender indifferent between lending and not lending. We state this as a proposition below.

Proposition 2 *Suppose that $U_j^L(Z(\cdot))$ crosses ε_{0j} just once. Under the assumption that the lender behaves as if she is never marginal, it is a weakly dominant strategy for the lender to bid an interest rate that makes the lender indifferent between lending and not lending.*

Proof. See Appendix. ■

The reason for why this strategy is weakly-dominant is the same as why bidding one's value is weakly dominant in a second-price auction. That is, as long as the lender is infra-marginal (i.e., not pivotal), increasing the bid does not affect the contract interest rate. Hence, it is in the lender's best interest to bid her value. The proof of the proposition is in the Appendix.

While it is certainly restrictive, we think that assuming that lenders behave as if they will never be pivotal is a reasonable approximation of the lenders' behavior. 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 bidding strategically is also small – the difference between the lowest interest rate among the losing bids and the interest rate of the marginal bid is only about 0.03%, on average. For these reasons, we assume from now on that lenders behave as if they will not be pivotal. The online appendix contains a more detailed discussion of this assumption.

Lender's Problem with Amount Choice   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 a discrete set M or chooses not to bid. Note that if the lender bids amount q to listing Z at contract interest rate r, then $E[qZ(r)] = q\mu(Z(r))$ and $Var(qZ(r)) = q^2\sigma^2(Z(r))$. Hence, lender j's utility can be expressed as follows,

$$U = U_j^L(qZ(r)) - \varepsilon_{0j} = q\mu(Z(r)) - A_j(q\sigma(Z(r)))^2 - c_q - \varepsilon_{0j}, \qquad (8)$$

where the cost of lending now depends on q as c_q.[26]

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 7 for the case in which $M = \{\$50, \$100, \$200\}$.[27] The three curves in Figure 7 correspond to $U_j^L(50Z(\cdot))$, $U_j^L(100Z(\cdot))$, and $U_j^L(200Z(\cdot))$. Just as

---

[26]To the extent that borrowers have biased belifes regarding the return from funding loans, c_q will also capture this bias.

[27]Note that in the estimation, we set $M = \{\$50, \$75, \$100, \$200, \$500, \$1000\}$.

before, there is a (weakly) dominant strategy for the lender under the assumption that the bidder is not pivotal. For the case shown in Figure 7, a (weakly) dominant strategy can be described by the following bidding strategy:

bid amount \$200 at interest r' if active interest rate $\in [r', s]$

bid amount \$100 at interest r'' if active interest rate $\in [r'', r')$

bid amount \$50 at interest r''' if active interest rate $\in [r''', r'')$

do not bid if active interest rate $\in [0, r''')$,

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. Basically, the lender should bid an amount q that maximizes $U_j^L(qZ(r))$ when the active interest rate is r. The optimal interest rate associated with the amount is the minimum interest that makes $U_j^L(qZ(\cdot))$ higher than $U_j^L(q'Z(\cdot))$. We now state the previous analysis in the form of a proposition.

Proposition 3 *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 $q(k) \in \{\$0\} \cup M$, so that $U_j^L(q(k)Z) - \varepsilon_{0j} \geq U_j^L(q'Z) - \varepsilon_{0j}$ for all q' and $r \in I_k$.[28] Under the assumption that the lender behaves as if she is never pivotal, it is a dominant strategy to bid $q(k)$ and interest rate r_k when the active interest rate is in I_k.*

We conclude the lender's model by briefly discussing the relationship between the model and identification. Using the ex-post borrower repayment data, we can identify $\mu(Z(r))$ and $\sigma(Z(r))$ for each r.[29] In particular, we can identify $\mu(Z(s))$ and $\sigma(Z(s))$, where s is the reserve interest rate, i.e., we can identify the "starting 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 any amount $m \in M$ for listings that are not funded. 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.

---

[28]To be more precise, when $q(k) \neq 0$, $U_j^L(q(k)Z) - \varepsilon_{0j} \geq \max\{0, \max_{q \in M} U_j^L(q(k)Z) - \varepsilon_{0j}\}$ and when $q(k) = 0$, $0 \geq \max_{q \in M} U_j^L(q(k)Z) - \varepsilon_{0j}$.

[29]$\mu(Z(r))$ and $\sigma(Z(r))$ correspond to the mean and variance of $Z(r)$ computed assuming rational expectations.
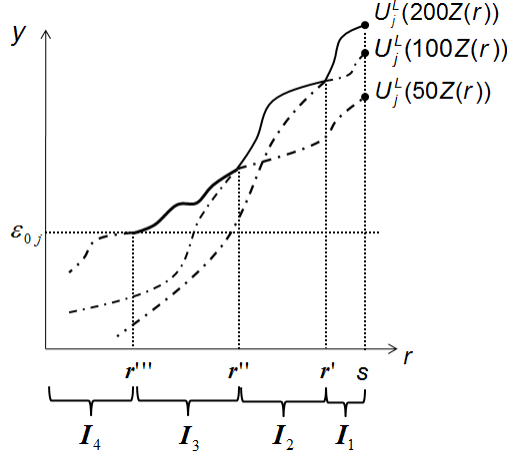
Figure 7: Graphical Representation of the Lender's Problem: Case of Amount Choice – The figure illustrates how the lender should bid when there is amount choice. Each curve $U_j^L(qZ(r))$ illustrates the relationship between r and the lender's utility net of ε_{0j} when the lender bids q. I_1 corresponds 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.

4.3 Equilibrium

We now discuss equilibrium existence and uniqueness.[30] There always exists an equilibrium of the model we described, but there may not exist a separating equilibrium.[31] Mailath (1987) (see also Mailath and von Thadden (2013)) provides sufficient conditions for the existence of a separating equilibrium. The main condition is a single-crossing condition on the borrower's utility function, V_0:

$$\frac{\partial}{ds}V_0(s,\varphi,\check{\varphi};X) \bigg/ \frac{\partial}{d\check{\varphi}}V_0(s,\varphi,\check{\varphi};X) \text{ is monotone increasing in } \varphi, \forall X, \tag{9}$$

where $V_0(s,\varphi,\check{\varphi};X)$ 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 $\check{\varphi}$.[32] X is a vector of conditioning variables such as borrower and listing characteristics. As long as lenders play the strategy described in Proposition 2, we can compute the funding probability and the distribution of r when lenders perceive the borrower to be of type $\check{\varphi}$. Hence, we can evaluate $\frac{\partial V_0}{d\check{\varphi}} \big/ \frac{\partial V_0}{ds}$ for any given parameter value.

---

[30]Since there always exists a pooling equilibrium, we focus on separating equilibria. As long as the lenders' beliefs off the equilibrium path are sufficiently pessimistic, all borrowers will find it optimal to post the same reserve rate.

[31]To be more precise, there exists an equilibrium in which all lenders have rational expectations. Note that our model is isomorphic to the model with rational expectations.

[32]More precisely,

$$V_0(s,\varphi,\check{\varphi}) = \Pr(s,\check{\varphi})\int V_1(r,\varphi)f(r|s,\check{\varphi})dr + (1 - \Pr(s,\check{\varphi}))\lambda(\varphi),$$

where we have suppressed the dependence on X. $\Pr(s,\check{\varphi})$ and $f(r|s,\check{\varphi})$ are the funding probability and the distribution of the contract interest rate, respectively, when the lender perceives the borrower to be of type $\check{\varphi}$.

While it is possible to check numerically whether the model satisfies (9) for a given parameter value, it is not easy to analytically characterize the set of parameters that satisfy these conditions. This is because V_0 is a fairly complicated object – it is an integral of a value function.

In what follows, we proceed by estimating the model without checking whether or not a separating equilibrium exists at any given parameter. Once we have estimated our parameters, we then check whether the sufficient conditions for separation are satisfied at the estimated values.[33] At the estimated parameter values, the conditions seem to generally hold.[34]

As for uniqueness, signaling models generally admit multiple equilibria because there are always pooling equilibria in which no information is transmitted. 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). In particular, D1 refinement is sufficient for ruling out equilibria because it implies the initial value condition of Mailath (1987).[35]

4.4 Model Discussion

In this section, we discuss some of our modeling choices and assumptions.

Independence of ε_t and φ   An important assumption we made in our borrower's repayment model is the independence of ε_t and φ. As we discussed above, mean independence of ε_t conditional on φ, i.e., $E[\varepsilon_t|\varphi] = 0$, is without loss of generality. This is because we can always redefine ε_t and φ – redefine ε_t as $(\varepsilon_t - E[\varepsilon_t|\varphi])$ and $D(\cdot)$ as $(D(\cdot) - E[\varepsilon_t|\cdot])$ – so that $E[\varepsilon_t|\varphi] = 0$. Given that we allow $D(\cdot)$ (or equivalently, the distribution of φ) to be nonparametric in our identification and estimation, this is without loss of generality. Although the assumption of independence between ε_t and φ is stronger than mean independence, it gives some credibility to the independence assumption.

Serial correlation in ε_t   Another assumption we made in our borrower's repayment model is 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

---

[33]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 very computationally intensive to include this condition in the estimation routine.

[34]For worse credit grades, condition (9) seems to be violated at the end of the reserve rate distribution. Note however, that when there is partial pooling at the maximum reserve rate of 36%, condition (9) need not hold close to 36%.

[35]We aknowledge that D1 is a relatively strong refinement criteria, and it would be desirable to obtain the same result with a weaker condition such as Divinity. However, because Divinity requires information about $F_\varphi(\cdot)$ – which is nonparametrically estimated – we found it hard to use Divinity for ruling out pooling equilibria.

31

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 $\{\varepsilon_t\}$.[36] Only the marginals of $\{\varepsilon_t\}$ are relevant for data generation. While this may appear to be a limitation, this means that as long as $\{\varepsilon_t\}$ is structural, our counterfactual policy is robust to serial correlation among $\{\varepsilon_t\}$.

Interpretation of φ Recall that the unobservable type of the borrower (φ) is interpreted as default cost in our model. However, we can write an alternative, observationally equivalent model where φ has the interpretation of unobserved income/assets of the borrower. We show this in the online appendix (see Section 3.2 in the online appendix). While there are several ways to model borrower heterogeneity – default cost, income, or some combination of the two – the implied default pattern may be very similar. For our purposes, the interpretation of types is not very important because it is structural to our counterfactual policy. This is not to say, however, that the distinction may be very important in other contexts.

Signaling through the loan amount In addition to the reserve interest rate, an important variable that the borrower needs to optimize over is the requested amount. While we do not explicitly model the amount choice of the borrower, abstracting away from the amount choice does not bias our results. First, even when the borrower optimizes over the requested amount, the borrower still chooses the reserve rate in accordance with equation (6). That is, conditional on the amount that the borrower requests, the borrower's reserve interest rate still solves equation (6). Second, we are allowing φ to be (arbitrarily) correlated with the requested amount in the identification allowing for the possibility that the requested amount can be informative about φ. To the extent that the requested amount has a signaling aspect, we will be able to capture it directly when estimating the distribution of φ as a function of covariates. In fact, we find some evidence that the loan amount is informative of φ, as we discuss in the results section.

Lender portfolio In our model, we abstracted from the portfolio decision of the lenders. In principle, however, lenders should care about the correlation between a given loan and existing

---

[36]Consider an extreme case when $\{\varepsilon_t\}$ takes on only two values, $\{+\infty, -\infty\}$. The following two cases are observationally equivalent: (1) $\{\varepsilon_t\}$ are perfectly correlated, so $\varepsilon_1 = \cdots = \varepsilon_T$, and $\Pr(\varepsilon_1 = +\infty) = p$ (2) $\{\varepsilon_t\}$ are independently (but not identically) distributed, with $\Pr(\varepsilon_1 = +\infty) = p$, and $\varepsilon_t = +\infty$ with probability 1 for all $t \geq 2$. For case (1), either $\varepsilon_1 = \cdots = \varepsilon_T = +\infty$ or $\varepsilon_1 = \cdots = \varepsilon_T = -\infty$ with probability p and $1 - p$. For case (2), $\varepsilon_1 = +\infty$ or $\varepsilon_1 = -\infty$ with probability p and $1 - p$, but $\varepsilon_t = +\infty$ with probability 1 for all $t \geq 2$. Note that (1) and (2) are different ($\{\varepsilon_t\}$ are correlated in (1) and independent in (2)), but we cannot identify between (1) and (2): In both cases, the borrower would default with probability p in the first period, and conditional on not defaulting in the first period, the borrower never defaults later. Our counterfactual results would be the same under either data generating process.

loans. Hence, the lender's utility should include a term that captures this correlation – which is currently missing.

As we discussed in Section 2.2, the average lender funds a total of 17.5 loans with a total portfolio size of about $1,300$. Given that this is a relatively small amount of money, the correlation in the returns among loans may not be of first-order importance to many lenders. However, to the extent that portfolio considerations are important, our estimate of the distribution of the lender's risk attitude (A_j) may pick up the correlation between the listing's return and the lender's other loan holdings as well as lender specific risk attitude.[37]

Cost of revising the bid   The optimal strategy described in Proposition 2 requires the lenders to submit new bids as the active interest rate changes. In the situation depicted in Figure 7, 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.[38] While this may be a strong assumption, it allows us to abstract from the dynamics of bidding and increases the tractability of the model. Also, in order to make sure that our estimates are not too sensitive to this assumption, we do not use the full implications of the dominant strategy (For example, we do not use the exact distribution of the realized contract interest rate). Our estimation only uses moments that are not too sensitive to this assumption as we describe below.

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, $F_{\varepsilon|X}$. 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 x_{amt} is the loan size and 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, $F_{\varepsilon|X}$, to depend on borrower/listing 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

---

[37]The lender may care about the variance of a given loan ($\sigma^2(Z)$) as well as the correlation between Z and other loans. The correlation term will get picked up by A_j in the estimation.

[38]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 7.

requested. We are allowing for this possibility. Second, note that we can normalize either $D(\cdot)$ or $F_{\varphi|X}$ (for some $X = X^*$) without loss of generality. For identification (and estimation) we normalize $D(\varphi) = -\varphi$.[39] It is also easy to see that we can normalize one of the constants in u_t without loss of generality: Hence we set $d_T = 0$.[40]

The intuition for our identification is quite simple. Recall that we model the borrowers' repayment decision as a sequence of binary decisions. Hence, if we knew the value of φ for each borrower, our model is a simple binary threshold crossing model and identification of the primitives easily follows from existing results (See, e.g., Mazkin (1992)).[41] The question then becomes how we identify φ. The identification of φ 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). Recall from Proposition 1 that the objective function of the borrower when choosing s (see equation (6)) satisfies the single crossing property. This guarantees that types with higher φ choose lower s conditional on X. In particular, if we take loans for which the reserve rate is equal to the α quantile of $F_{s|X}$, the borrowers all have φ equal to the $1 - \alpha$ quantile of $F_{\varphi|X}$. Then the default rate among borrowers with reserve rate equal to $F_{s|X}^{-1}(\alpha)$ identifies $F_{\varphi|X}(\alpha)$ (and hence the value of φ for each borrower). The following proposition states our results formally:

Proposition 4 $F_{\varphi|X}$, $F_{\varepsilon|X}$ and $\lambda(\cdot)$ are nonparametrically identified up to location normalizations. $\{d_t\}$ are also identified.

    Proof. See Appendix. ∎

Nonparametric identification of $F_{\varphi|X}$, $F_{\varepsilon|X}$ and $\lambda(\cdot)$ holds even when there is partial pooling at 36%. The proof of this Proposition when there is no pooling appears in the Appendix. The online appendix (see Section 4.1 of the online appendix) contains the proof of the proposition when there is partial pooling.

5.2 Identification of the Lender's Primitives

The primitives of the lender's 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 $\{1, .., \overline{N}\}$. Our proof of identification proceeds by first showing identification of F_A, F_{ε_0}, and c_q under the assumption

---

[39]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.

[40]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, $F_{\varepsilon|X}$.

[41]To be precise, the binary threshold crossing model identifies $F_{\varepsilon_t|X^*}$ for some X^*. In order to identify $F_{\varepsilon_t|X}$ for $X \neq X^*$, we also use the first-order condition (7). Once $F_{\varphi|X^*}$ and $F_{\varepsilon|X}$ are identified, $\lambda(\cdot)$ is identified from equation (7).

that $P_q(\mu, \sigma)$, which we will define below, is identified for all values of (μ, σ) and $q \in M \cup \{\$0\}$. We will then show that $P_q(\mu, \sigma)$ and F_N are identified. Given a listing with mean and variance of return equal to μ and σ^2, define $P_q(\mu, \sigma)$ to be the probability that funding q dollars gives higher utility to a lender than funding q' ($q' \neq q$) dollars. Formally, $P_q(\mu, \sigma)$ is expressed as follows:

$$
P_q(\mu, \sigma) = \begin{cases} \Pr\left(q\mu - A(q\sigma)^2 - c_q - \varepsilon_0 \geq \max\left\{0, \max_{q' \in M}\{q'\mu - A(q'\sigma)^2 - c_{q'} - \varepsilon_0\}\right\}\right) & \text{for } q \in M \\ \Pr(0 \geq \max_{q' \in M}\{q'\mu - A(q'\sigma)^2 - c_{q'} - \varepsilon_0\}) \text{ for } q = 0 \end{cases}.
$$

Note that $P_q(\mu, \sigma)$ corresponds to the probability that (A, ε_0) lie in the region defined by the inequalities in the expression above. By varying μ and σ, this region changes. Proposition 5 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}. In other words, $(F_A, F_{\varepsilon_0}, c_q)$ is identified if $P_q(\mu, \sigma)$ are identified.

Proposition 5 $(F_A, F_{\varepsilon_0}, c_q)$ *are identified if* $(P_q(\mu, \sigma), P_0(\mu, \sigma))$ *are identified.*

Proof. See online appendix. ∎

The next proposition claims that $P_q(\mu, \sigma)$ and F_N are both identified.

Proposition 6 $P_q(\mu, \sigma)$ *is identified for all* q *and* (μ, σ) *on the support of* (μ, σ). F_N *is also identified.*

Proof. See online appendix. ∎

The proof of Propositions 5 and 6 are contained in Section 4 of the online appendix. Here, we briefly discuss the intuition for why F_N and $P_q(\mu, \sigma)$ are identified. Consider a listing which has yet to be fully funded. Let x_{amt} denote the requested loan amount and (μ, σ^2) denote the mean return and variance of this listing *if funded at the reserve interest*, s. Under the strategy described in section 4.2, lender j bids an amount equal to q if and only if lender j's risk aversion parameter and the outside option, (A_j, ε_{0j}), are such that $U_j^L(qZ(s)) - \varepsilon_{0j} \geq \max\{\max_{q' \in M} U_j^L(q'Z(s)) - \varepsilon_{0j}, 0\}$. The probability of this event is $P_q(\mu, \sigma)$. Given that a listing is funded if and only if there is a sufficient number of potential bidders who are willing to fund it, we can express the probability that a listing is funded as a function of F_N and $\{P_q(\mu, \sigma)\}$. Since the probability that a listing is funded can be identified for all x_{amt}, μ, and σ^2, sufficient variation in x_{amt} and (μ, σ) identifies both F_N and $P_q(\mu, \sigma)$ if we assume that F_N is invariant to x_{amt} and (μ, σ).

Our identification relies on the fact that when a lender with (A_j, ε_{0j}) visits a listing that is still not fully funded, the lender submits a bid with amount q if and only if $U_j^L(qZ(s)) - \varepsilon_{0j} \geq \max\{\max_{q' \in M} U_j^L(q'Z(s)) - \varepsilon_{0j}, 0\}$, where $Z(\cdot)$ is evaluated at the return from funding the listing at s. Note that this lender behavior is consistent with the dominant strategy we described in section 4.2.

6 Estimation

We estimate our model in three steps. First, we estimate the conditional distribution of the contract interest rate given the reserve rate, $f(r|s,x)$, and the funding probability, $\Pr(s,x)$. We estimate these two functions nonparametrically as $f(r|s,x)$ and $\Pr(s,x)$ are both equilibrium objects. 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 $f(r|s,x)$ and $\Pr(s,x)$

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 credit grade, requested amount, debt-to-income ratio, and home ownership. We use a (second-order) Hermite series approximation to estimate $f(r|s,x)$, following Gallant and Nychka (1987). Our estimation of $\Pr(s,x)$ is based on a Probit model with flexible functional forms. The details regarding the estimation are contained in Section 4 of the online 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 Section 5.1).

In order to estimate θ_B, we maximize the likelihood of repayment and default for each borrower. Note that for any θ_B, our borrower's repayment model 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 x), where this relationship is implicitly defined by the borrower's first-order condition (equation (7)). 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 of repayment and default.

The actual computation of the likelihood proceeds as follows: First, recall that the borrower's choice of the reserve rate satisfies the first-order condition;

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

$$(10)$$

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 $\widehat{f}(r|s, x)$. We can also compute $V_1(r, \varphi, x; \theta_B)$ for each value of $\{r, \varphi, x\}$ given θ_B by recursively solving the borrower's dynamic problem. This allows us to back out the borrower's type, $\widehat{\varphi} \equiv \widehat{\varphi}(s, x; \theta_B)$, for each borrower. Note that Proposition 1 shows that the right-hand side of equation (10) is monotonic in φ, guaranteeing a unique solution given s and x (for unpooled types).[42]

The second step of our procedure is to compute the likelihood for a given sequence of repayment decisions for each borrower i, using $\widehat{\varphi}_i = \widehat{\varphi}(s_i, x_i; \theta_B)$. Borrower i's default probability at period t is

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

and the probability of paying back at period t is $1 - \Pr(\text{default at } t; \theta_B)$. Let ι_{it} be an indicator variable that is equal to 1 if borrower i repays at period t, and 0 otherwise.

Finally, the likelihood is written as

$$L(\theta_B) = \prod_{i=1}^{N_L} \left[\prod_{t=1}^{T_i} \Pr(\text{repay at } t; \theta_B)^{\iota_{it}} \times \Pr(\text{default at } t; \theta_B)^{(1-\iota_{it})} \right], \quad (12)$$

where N_L is the number of loans, $\{\iota_{it}\}$, is the sequence of repayment decisions, and $T_i \equiv \min\{1 + \sum_{\tau=1}^{T} \iota_{i\tau}, 36\}$, i.e., the number of periods until default or 36 periods, whichever is smaller. We obtain our parameter estimates by maximizing the likelihood function.[43]

---

[42]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. When we fail to solve for $\widehat{\varphi}$, we replace $\widehat{\varphi}$ with default values. We tried two different default values and the results seem to be pretty stable. The results from the different specifications are available on request.

[43]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 for the types that are not being pooled. 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.

6.3 Estimation of the Lender Model

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, $F_N(\cdot; \theta_L)$, the distribution of the lender's risk attitude, $F_A(\cdot; \theta_L)$, the distribution of the opportunity cost of lending, $F_{\varepsilon_0}(\cdot; \theta_L)$, and the lender's cost of bidding, c_q.

We use a (simulated) method of moments by matching the conditional funding probability and the number of bids. 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 listing characteristic x_i, reserve interest s_i, and parameter θ_L. As we will explain below, $fd(x_i, s_i; \theta_L)$ can be expressed as

$$fd(x_i, s_i; \theta_L) = 1\left\{\sum_{j=1}^{N} q_j^* \geq x_{i,amt}\right\} \text{ and} \tag{13}$$
$$q_j^* = \underset{q_j \in M \cup \{0\}}{\arg\max}\left\{1\{q_j \neq 0\}\left(U_j^L(q_j Z(s)) - \varepsilon_{0j}\right)\right\},$$

where N is the (random) number of potential lenders, $U_j^L(q_j Z(s))$ is the utility of lending q_j dollars at interest rate s (defined in expression (8)), and 1_E is an indicator function that equals one if event E is true. 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 (13). Suppose that there are N potential lenders and their risk attitude and outside option are $(A_j)_{j=1}^{N}$ and $(\varepsilon_{0j})_{j=1}^{N}$. When the loan is not fully funded yet, the optimal choice is given by the second equation in expression (13), where $U_j^L(q_j Z(\cdot))$ is evaluated at *the reserve interest rate s*. Now consider the right hand side of the first equation of (13). $x_{i,amt}$ 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 $\sum_{j=1}^{N} q_j^* \geq x_{i,amt}$.

In addition to the funding probability, we also match two moments. The first is the number of lenders, in particular, the number of lenders who bid an amount $q \in \{\$50, \$75, \$100, \$200, \$500, \$1000\}$ to unfunded listings. The second is the fraction of listings that receive no bids. These objects can be expressed as functions of the primitives as long as lenders play the strategy we described in section 4.2.

Note that the set of moments that we use in the estimation does not use the full implications of the strategy we defined in section 4.2. For example, we do not use information concerning the realization of the actual contract interest rate or the number of lenders who bid an amount equal to q conditional on the listing being funded. This is because these objects are quite sensitive to the particular dominant strategy described in section 4.2. If lenders are playing other strategies (say, because revising their bid is costly), the distribution over the contract interest rate could be quite different depending on how we specify the timing of lender arrival. Finally, we have suppressed the conditioning variables in our exposition, but we construct moment conditions for each conditioning variable.[44]

7 Results

The exact specification we use to estimate the model of the borrower is as follows: First, we set the period utility function as $u_t(r_j; \theta_B) = -r \times x_{amt} + \theta_t$, where $\{\theta_t\}$ are time dummies.[45] $D(\varphi)$ is normalized to $-\varphi$ and the outside option $\lambda(\varphi)$ is specified as a linear function with a credit grade specific slope as $\lambda_{x_{gr}}\varphi$. The type of each borrower is recovered nonparametrically for each borrower. Lastly, F_ε is specified to be a Type I extremum value distribution with standard error equal to σ_ε and the discount factor, β, is set to $0.95^{1/12}$.

As for the lenders' side, we estimate the distribution of potential lenders, F_N, the distribution of lender's risk attitude, F_{A_j}, the distribution of the outside option and forecasting error, $F_{\varepsilon_{0j}}$, and the costs of bidding for each amount choice, $\{c_m\}_{m \in M}$. 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 the risk attitude, F_{A_j}, and the outside option, F_{ε_0}, to be Normally distributed with $N(\mu_A, \sigma_A^2)$ and $N(\mu_{\varepsilon_0,\tau}, \sigma_{\varepsilon_0}^2)$. The mean of the distribution of $F_{\varepsilon_{0j}}$ is allowed to depend on the calendar month, τ, to capture changes in the lender's outside option or beliefs due to factors such as macro shocks. Given that one of c_m ($m \in M$) or $\mu_{\varepsilon_0,\tau}$ can be normalized to zero, we set $c_{50} = 0$.

We report the estimation results in Table 7 and Table 8. In Table 7, we report the distribution of the default cost of the borrower (φ), by credit grade and requested amount. Table 8 reports the estimates of the other parameters of the model.

Table 7 reports the quantiles of borrower type (φ) by credit grade and by requested amount in

---

[44]For our first two moments (funding probability and the number of lenders who bid an amount 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 the third moment (probability of receiving no bid), we just compute one moment for each credit grade. We then sum the moment conditions for each credit grade.

[45]Instead of including a separate dummy variable for each repayment period, we include dummies that correspond to a three-month period in order to reduce the number of parameters that we estimate. Hence, we have a single dummy for the first three months, another dummy for the next three months, etc. We normalize the first dummy, which leaves us with 11 dummies per credit grade for estimation.

Amount Quantile	Type Quantile	AA	A	B	C
25%	25%	10.683	7.437	2.966	3.196
	50%	10.872	7.635	3.365	3.479
	75%	10.966	8.735	3.867	4.012
50%	25%	9.999	7.085	2.944	3.041
	50%	10.302	7.229	3.207	3.424
	75%	10.595	8.013	3.726	4.031
75%	25%	7.984	5.685	2.755	3.099
	50%	8.196	5.947	2.923	3.463
	75%	8.427	6.406	3.454	3.977
All	25%	7.890	6.118	2.753	3.007
	50%	9.756	7.199	3.224	3.404
	75%	10.683	7.785	3.760	3.947

Table 7: Quantiles of the Borrower's Type Distribution: This table reports the estimated quartiles of the default cost of the borrower, φ, by credit grade and by requested amount. The unit is $1,000.

Parameter	Borrower Estimates	Parameter	Lender Estimates by Grade			
			AA	A	B	C
σ_ε	6.0096	μ_N	5.8044	6.3012	6.1769	5.6751
	(0.2095)		(0.2151)	(0.4492)	(1.0709)	(0.1370)
λ_{AA}	3.0965	σ_N	0.5510	0.5613	2.3968	1.0158
	(0.0827)		(0.0416)	(0.0967)	(1.0725)	(0.0832)
λ_A	4.3398	μ_A	5.71×10^{-2}	5.45×10^{-2}	3.49×10^{-2}	4.07×10^{-2}
	(0.1384)		(1.32×10^{-3})	(3.75×10^{-3})	(9.54×10^{-3})	(0.49×10^{-3})
λ_B	9.4551	σ_A	1.80×10^{-2}	1.49×10^{-2}	1.19×10^{-2}	1.17×10^{-2}
	(0.3477)		(4.50×10^{-4})	(1.44×10^{-3})	(3.02×10^{-3})	(1.20×10^{-4})
λ_C	8.8111	μ_{ε_0}	1.0843	0.8406	-1.6084	0.5439
	(0.3551)		(0.0342)	(0.3320)	(0.8168)	(0.0518)
		σ_{ε_0}	1.2959	0.1501	0.1734	0.1998
			(0.0935)	(0.0227)	(0.0491)	(0.0100)
		c_{75}	-15.0651	-24.2441	-9.1104	-21.7680
			(0.3386)	(1.6037)	(2.1378)	(0.4078)
		c_{100}	-28.1832	-38.1448	-29.3130	-27.5604
			(0.6601)	(4.0414)	(6.7876)	(0.9188)
		c_{200}	-36.5678	-43.4040	-25.5103	-25.3122
			(0.8889)	(3.0231)	(5.2462)	(0.6498)
		c_{500}	-0.6329	-0.5808	-0.7933	-0.7762
			(0.0148)	(0.0464)	(0.1897)	(0.0281)
		c_{1000}	14.5768	9.5608	6.5145	6.6795
			(1.1268)	(1.3798)	(1.4203)	(0.2811)
Obs	$3,818$		$1,420$	$1,850$	$3,068$	$5,203$

Table 8: Parameter Estimates of the Borrower's and Lender's Model. 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 both the borrower's and lender's model, but we omit the estimates from the table. Standard errors are obtained by bootstrap (150 times) and they are reported in parentheses. Borrower's model is scaled in $1,000, and the lender's model is scaled in $1.

units of $\$1,000$. Recall that one interpretation of φ is the borrower's default cost, with good types having high default cost (and hence less likely to default) and bad types having low default cost (and hence more likely to default). We find that the median default cost (across all requested loan amount; see second to last to row of Table 7) are estimated to be around $\$9,800$ for credit grade AA, $\$7,200$ for credit grade A, $\$3,200$ for credit grade B and $\$3,400$ for credit grade C.[46] We find that borrowers with credit grades AA and A have much higher default cost compared to borrowers with credit grades B or C. This seems natural given that we expect borrowers with good credit grades to have higher default cost. While we find it somewhat surprising that borrowers with credit grade C have slightly higher default cost than borrowers with credit grade B, this probably reflects the fact that borrowers of these two credit grades are not that different – as we report in Table 4, the average default probability of borrowers in these two credit grades were only about 1% apart.

Another general pattern that can be seen from Table 7 is that borrowers who request a large amount generally have lower values of φ. For example, the median AA borrower with a requested loan amount at the 75% quantile has a default cost equal to about $\$8,200$ – which is more than $\$2,000$ lower than the default cost of the median AA borrower with a requested amount at the 25% quantile. Note that conditioning on the requested amount tightens the distribution of φ significantly in credit grades AA and A, but less so for B and C. For example, the unconditional interquartile range of credit grade AA is about $\$2,800$ whereas the conditional interquartile range at the median requested amount is less than $\$600$. For credit grade C, the unconditional and the conditional interquartile range are about the same. These results imply that the requested amount has an important signaling value for credit grades AA and A, but not so much for credit grades B and C.

In the first two columns of 8 we report the parameter estimates of the borrowers' model. The rest of Table 8 presents the estimation results of the lenders' model. Recall that $\lambda_{x_{gr}}$ is a parameter that measures the relationship between the default cost of the borrower (φ) and the utility of the outside option ($\lambda_{x_{gr}}\varphi$). Our estimates of $\lambda_{x_{gr}}$ indicate that it 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 $\lambda_C = 8.8$). Our estimates imply that a $\$100$ increase in the default cost of the borrower translates to a $\$310$ increase in the outside option for credit grades AA, $\$430$ increase for credit grade A, and so on. The relatively large estimate of $\lambda_{x_{gr}}$ that we find for low credit grades compared to high credit grades may reflect the fact that the marginal increase in the default cost of the borrower for low credit grades leads to a disproportionate increase in the default probability. A given reduction in the default cost may translate to more credit at lower credit grades and hence to a higher utility from the outside option.

Columns 3 through 7 of Table 8 report the parameter estimates for the lenders. We estimated

---

[46]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.

	AA		A		B		C	
	Data	Model	Data	Model	Data	Model	Data	Model
Funding probability: Mean	53.4	47	40.9	36	33.4	26	24.7	22
Contract interest rate: Mean	9.6	7.7	12.7	11.1	16.4	12.7	18.1	17.4
Contract interest rate: Standard deviation	3.3	3.3	4.5	4	4.6	5	6.2	4.4

Table 9: Model Fit: The first row reports the mean funding probability in the data and the mean funding probability predicted from the model for each grade. The second row reports the mean contract interest rate in the data and the mean contract interest rate predicted from the model for each grade. The third row reports the standard deviation of the distribution of the contract interest rate in the data and the one predicted from the model for each grade.

a log Normal distribution with parameters (μ_N, σ_N) for the number of potential bidders. The estimated means of the lenders' risk aversion parameter (μ_A) range from 3.49×10^{-2} to 5.71×10^{-2}, which are comparable to the risk aversion estimates reported in Paravisini, Rappoport and Ravina (2013) who study a similar setting. Their estimate of the average risk aversion parameter among participants of Lending Club, another P2P lending web site, is about 3.68×10^{-2}, with a standard deviation of about 2.37×10^{-2}.

Finally, Table 8 reports our estimate of the mean of ε_{0j} (μ_{ε_0}), which range from -1.61 to 1.08. Note that we let μ_{ε_0} vary by month – the estimates in Table 8 correspond to the mean of ε_{0j} for October 2008, which is the last period of the data. The estimates of μ_{ε_0} for other months are not reported in the table but they are quite similar.

7.1 Model Fit

In this subsection, we report the fit of the model. To evaluate the fit, we solve the borrower's repayment problem and the lenders' model for each listing and simulate the outcome. In particular, we solve the borrower's problem 10,000 times to compute the mean and variance of the return for each interest rate. We then take 200 sets of potential lenders to simulate the distribution of outcomes. Table 9 reports the comparison between the data and the model prediction in terms of the funding probability, mean contract interest rate, and standard deviation of contract interest rates for each credit grade. Overall, we find that our structural model fits the data quite well.

Figure 8 reports the scatter plot of the predicted contract interest rates against the actual contract interest rate for funded listings. The solid line in each figure is a simple regression line and the dotted line is the 45 degree line. We find that the plots are generally clustered around the 45 degree line with the exception of Grade C, where the fit is somewhat worse than for other grades.

8 Counterfactual Experiment

In our counterfactual experiment, we compare the equilibrium market outcomes and welfare under three market designs – a market with signaling, a market without signaling (i.e., pooling) and a
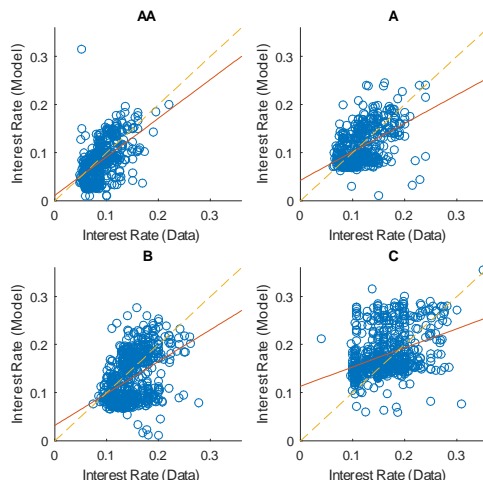
Figure 8: Model Fit: The figure shows the scatter plot of the actual contract interest rate and the predicted contract interest rate. The solid line is the fitted line and the dashed line is the 45 degree line.

market with no information asymmetry between borrowers and lenders. These counterfactuals are interesting because they allow us to empirically quantify the extent to which credit markets suffer from adverse selection and the extent to which signaling can affect market outcomes. In particular, the question of how adverse selection affects credit supply goes back to Stiglitz and Weiss (1981) but few empirical attempts have been made to study the effect.[47]

In Figure 9, we present the credit supply curves under no asymmetric information (solid curves) and that under pooling (dotted curves) for each of the four credit grades with observable listing characteristics set to the median of each credit grade. 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 supply varies from credit grade to credit grade.

The credit supply curve under no asymmetric information is computed by first fixing a borrower type, and computing the average credit supply for that type at each interest rate, assuming that lenders perfectly know the type of the borrower. We then average the credit supply over the distribution of borrower types to obtain the supply curves in the figure. In order to compute the credit supply curve under pooling, we assume that each borrower can post a secret reserve price as opposed to a public reserve price. That is, we let the borrower take out a loan only if the contract interest at the end of the bid closing period is less than the secret reserve price. Note that under this market design, it is a dominant strategy for each borrower to submit a secret reserve rate equal to

---

[47]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 also take x_{amt}, the borrower's requested amount, as exogenous. We acknowledge that these are potential limitations of our counterfactual experiment, and we report some sensitivity analysis in Section 7.3 of the online appendix.
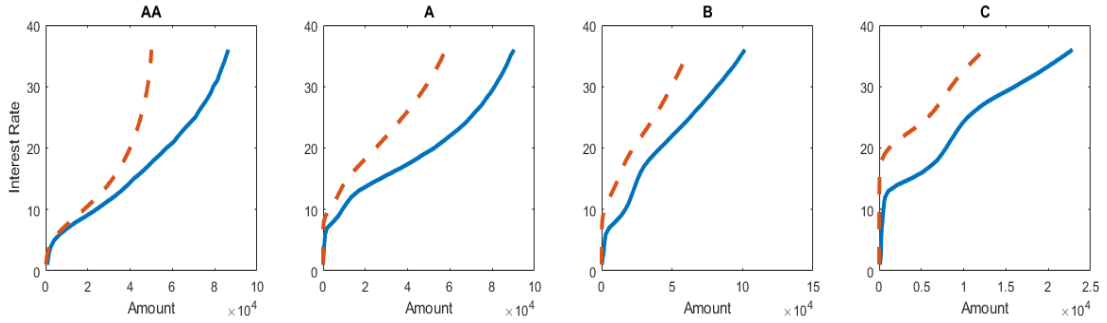
Figure 9: Credit Supply Curve: The thick dotted curve corresponds to the credit supply curve under no signaling (i.e., pooling). The solid line corresponds to the credit supply curve under no asymmetric information.

the interest rate at which the borrower is indifferent between borrowing and not borrowing.[48] This market design would induce pooling of types, i.e., at a given interest rate, there would be a mix of different borrowers who take out the loan, and the lenders have no way of differentiating among them. The supply curves under pooling trace out the average amount of credit that potential lenders are willing to supply to a pool of borrower types at each interest rate. The pool of borrowers consist of those whose secret reserve price is higher than the given interest rate.

Figure 9 makes clear the role of adverse selection in credit markets. At low interest rates, the pool of borrowers who are willing to take out a loan is not adversely selected. Hence, the average return from funding a loan under pooling is similar to that from funding a loan to a borrower who is randomly chosen from the set of all borrowers. This fact tends to make the difference between the credit supply under pooling and no information asymmetry relatively small at low interest rates. However, at higher interest rates, the set of borrowers who are willing to take out a loan consists of only low types. Hence, the average return from funding a loan under pooling will be much lower than that from funding a loan to a borrower that is randomly selected from all borrowers. This fact tends to make the differences between the two supply curves large at higher interest rates.

Depending on the characteristic of the listing, it is possible for the pool of borrowers to become so adversely selected that the mean return from lending becomes decreasing in the interest rate, as pointed out by Stiglitz and Weiss (1981). In this case, the supply curve becomes backward bending. Figure 10 shows two examples, one from credit grade AA and another from credit grade B. For the listing in grade AA, the requested amount is set to $21,000$, debt-to-income ratio to 21%, and home ownership to 1. For the listing for grade B, the requested amount is set to $10,000$, debt-to-income ratio to 28%, and home ownership to 0. For these listings, the worsening of borrower types at higher interest rates puts a downward pressue on the rate of return to the point that the downard pressure outweighs the increase in the return from higher interest rates.

Finally, we examine the welfare implications of signaling and information asymmetry. In Table

---

[48]For a more detailed computational procedure for obtaining the credit supply curves, see Section 7 of the online appendix.
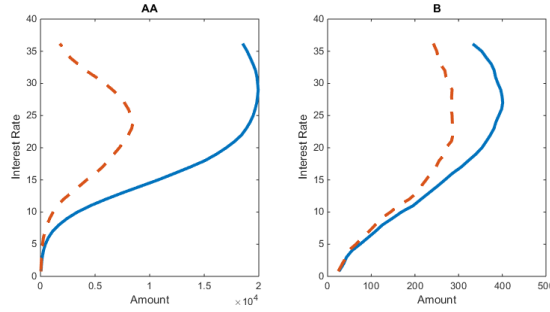
Figure 10: Examples of Backward Bending Credit Supply Curves: The figure shows two credit supply curves, one from credit grade AA and another from credit grade B. For the listing in grade AA, the requeted amount is set to $21,000$, debt-to-income ratio to 21%, and home ownership to 1. For the listing in grade B, the requeted amount is set to $10,000$, the debt-to-income ratio to 28%, and home ownership to 0.

| | | Welfare | | |
		Borrower	Lender	Total
	Signaling	260.6	52.6	313.2
		(18.2)	(2.2)	(19.8)
AA	Pooling	309.0	43.0	352.1
		(11.9)	(2.1)	(13.0)
	Symmetric	340.4	54.0	394.4
		(13.5)	(2.1)	(14.7)
	Signaling	361.4	28.0	389.5
		(15.8)	(2.7)	(17.4)
A	Pooling	329.8	13.5	343.3
		(12.4)	(1.8)	(12.7)
	Symmetric	399.5	29.5	429.0
		(15.8)	(2.7)	(17.3)
	Signaling	318.9	30.4	349.3
		(64.4)	(5.5)	(67.1)
B	Pooling	233.4	16.4	249.8
		(44.6)	(3.4)	(45.2)
	Symmetric	332.1	33.0	365.1
		(61.2)	(5.9)	(62.5)
	Signaling	434.4	21.2	455.6
		(16.7)	(0.6)	(17.1)
C	Pooling	321.9	10.2	332.1
		(12.0)	(0.4)	(12.3)
	Symmetric	452.2	24.3	476.5
		(13.4)	(0.7)	(13.0)

Table 10: Expected Surplus for Different Market Designs by Credit Grade. The standard errors are calculated by bootstrap and reported in parenthesis.

10, we report the surplus of the lenders and the borrowers per listing for each of the three different market designs we consider. The surplus of a borrower with type φ is the product of the probability of being funded ($\Pr(\text{funded})$) and the surplus conditional on borrowing ($E_r[V_1(r, \varphi) - \lambda(\varphi)]$), where the expectation is over the contract interest rate. The surplus of lender j is $E_r[U_j^L(r)]$, if a loan is funded and 0, otherwise. We compute the expected surplus by simulating the model using the estimates we obtained from our structural model. We then take the average over the borrower's type distribution. Observable listing characteristics such as the amount, debt-to-income ratio, and home ownership are set to the median values of each credit grade. Details of the computation are discussed in the online appendix.

First, consider the welfare of the borrowers 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 high credit grades while it is relatively severe in low credit grades. For example, the difference in borrower welfare is about \$30 for credit grade AA while it is over \$100 for credit grades B and C. Comparing these numbers to borrower welfare 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 lender welfare under pooling and under no asymmetric information, we find that welfare decreases under pooling in all credit grades. We also find that the expected surplus is higher under signaling relative to that under pooling in all credit grades for lenders.

The third column of Table 10 reports total welfare. Comparing the total surplus under pooling and no asymmetric information, we find that the cost of adverse selection can be quite large, with a 31.6% (\$115.3) decrease in total surplus for credit grade B and a 30.3% (\$144.4) 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 86.3% and 85.5% 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 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. In our counterfactual, we compare market outcomes 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 one of the first structural analyses of signaling in industrial organization, 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. 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.

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 $D'(\varphi) \leq \frac{\partial}{\partial \varphi} V_\tau(r, \varphi) < 0$ for some $\tau \leq T$, then the same conditions hold for $\tau - 1$. First, for $t = T$,

$$\frac{\partial}{\partial \varphi} V_T(r, \varphi) = \frac{\partial}{\partial \varphi} \int \max\{u_T(r) + \varepsilon_T, D(\varphi)\} dF_{\varepsilon_T}(\varepsilon_T) = D'(\varphi) \Pr\nolimits_T(r, \varphi),$$

where $\Pr_T(r, \varphi) = \Pr(u_T(r) + \varepsilon_T < D(\varphi))$. 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,

$$\frac{\partial}{\partial \varphi} V_t(r, \varphi) = \frac{\partial}{\partial \varphi} \int \max\{u_t(r) + \varepsilon_t + \beta V_{t+1}(r, \varphi), D(\varphi)\} dF_{\varepsilon_t}(\varepsilon_t)$$

$$= \frac{\partial}{\partial \varphi} \beta V_{t+1}(r, \varphi)(1 - \Pr_t(r, \varphi)) + D'(\varphi) \Pr_t(r, \varphi) \geq D'(\varphi),$$

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

$$\frac{\partial}{\partial r \partial \varphi} V_t(r, \varphi) = \frac{\partial}{\partial r} \left[\frac{\partial}{\partial \varphi} \beta V_{t+1}(r, \varphi)(1 - \Pr_t(r, \varphi)) + D'(\varphi) \Pr_t(r, \varphi) \right]$$

$$= \frac{\partial^2}{\partial r \partial \varphi} \beta V_{t+1}(r, \varphi)(1 - \Pr_t(r, \varphi)) + \frac{\partial}{\partial r} \Pr_t(r, \varphi) \times (D'(\varphi) - \frac{\partial}{\partial \varphi} \beta V_{t+1}(r, \varphi)) \leq 0.$$

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 Proof of Proposition 2

Suppose that the lender bids an interest rate, r_j, that is higher than r^0 (the interest rate at which the lender is indifferent between lending and not lending). 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.

A.3 Proof of Proposition 4

In this Appendix, we provide a proof of identification of $F_{\varphi|X}$, $F_{\varepsilon|X}$, and $\lambda(\varphi)$. We first note that we can normalize one of the constants in u_t without loss of generality: Hence we set $d_T = 0$.[49] Also, we can normalize the location of $F_{\varphi|X}$ at one point: Hence, we set $F_{\varphi|X^*}^{-1}(\alpha^*) = 0$ for some $\alpha^* \in (0,1)$ and X^*.[50] In what follows, we consider the case when there is no pooling. The Online Appendix contains the proof for the case when there is partial pooling.

Consider the repayment decision of the borrower with $F_{\varphi|X^*}^{-1}(\alpha^*)(= 0)$ at period $t = T$. The borrower's problem is as follows:

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

where x_{amt}^* is an element of X^*. This is simply a binary threshold-crossing model; hence using variation in r, we can nonparametrically identify the conditional distribution of ε_T given X^*, i.e., $F_{\varepsilon|X^*}$. Once $F_{\varepsilon|X^*}$ is identified, we can identify $F_{\varphi|X^*}^{-1}(\alpha)$ for all α given X^* by conditioning the sample on the α-quantile of s given X^* (i.e., samples with $s = F_{s|X^*}^{-1}(\alpha)$).[51] This is because $F_{\varphi|X^*}^{-1}(\alpha)$ is just a constant term in the binary threshold-crossing model where the distribution of ε_T given X^* has already been identified.

Now consider the $t = T - 1$ period problem with $X = X^*$:

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

where $V_T(r, F_{\varphi|X^*}^{-1}(\alpha))$ and $F_{\varphi|X^*}^{-1}(\alpha)$ have already identified. Similarly as before, we can nonparametrically identify the distribution of $(\varepsilon_{T-1} + d_{T-1})$ and the value of β using variation in r, given that this is a simple binary threshold crossing model. It should be clear that $\{\varepsilon_t + d_t\}_{t \leq T-2}$ can also be identified by looking at the borrower's period t problem and the associated default probability.

Now, we discuss how to identify $\lambda(\varphi)$. Rearranging the borrower's FOC in equation (7) evalu-

---

[49]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, $F_{\varepsilon|X}$.

[50]Given that $D(\varphi) = -\varphi$, if we set $\tilde{\varepsilon}_t = \varepsilon_t + \kappa$ $(\forall t)$, $\tilde{F}_{\varphi|X}(h) = F_{\varphi|X}(h + \kappa)$, $\tilde{d}_T = d_T$, $\tilde{d}_t = d_t - \beta\kappa$ $(t \in \{1, ..., T-1\})$ 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.

[51]Here, we are using the fact that $\varphi \perp \varepsilon|X^*$.

ated at $X = X^*$ and solving for $\lambda(\varphi)$, we obtain

$$\lambda(\varphi) = \int V_1(r, \varphi; X^*) f(r|s; X^*) dr + \frac{\Pr(s; X^*)}{\frac{\partial}{\partial s} \Pr(s; X^*)} \int V_1(r, \varphi; X^*) \frac{\partial}{\partial s} f(r|s; X^*) dr, \quad (2)$$

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

Lastly, we show that $F_{\varphi|X}$ and $F_{\varepsilon|X}$ are identified for any 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 same steps as above to identify $F_{\varphi|X}$ and $F_{\varepsilon|X}$. In order to see that $F_{\varphi|X}(0)$ is identified, consider a given profile, $(F_{\varepsilon|X}^*, F_{\varphi|X}^*, \lambda^*, d_t^*)$. Note that 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 $F_{\varphi|X}(0)$. This concludes the proof of identification when there is no pooling. In the online appendix we also discuss the case when there is pooling at $s = 36\%$.

References

[1] Ackerberg, Daniel, "Advertising, Learning, and Consumer Choice in Experience Good Markets: A Structural Empirical Examination", *International Economic Review*, Vol. 44, No. 3, August 2003, pp. 1007-1040

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

[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] Allen, Franklin and Gerald Faulhaber, "Signaling by Underpricing in the IPO Market," *Journal of Financial Economics*, Vol. 23, 1989, pp. 303-323.

[5] Arnold, Lutz and John Riley, "On the Possibility of Credit Rationing in the Stiglitz-Weiss Model," *American Economic Review,* Vol. 99, No. 5 (2009), pp. 2012-2021

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

[7] Backus, Matt, Tom Blake, and Steve Tadelis, "On the Empirical Content of Cheap-Talk Signaling: an Application to Bargaining," forthcoming in *Journal of Political Economy*.

[8] 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.

[9] Bester, Helmut, "Screening vs. Rationing in Credit Markets with Imperfect Information," *American Economic Review*, Vol. 75, No. 4, 1985, pp. 850-855

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

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

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

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

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

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

[16] 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.

[17] Gedge, Christopher, James Roberts, and Andrew Sweeting, "A Model of Dynamic Limit Pricing with an Application to the Airline Industry," mimeo, 2012.

[18] Harris, Milton and Artur Raviv, "A Sequential Signaling Model of Convertible Debt Call Policy," *Journal of Finance*, 1985, Vol.40, (5) pp. 1263-1281.

[19] Holt, Charles, and Susan Laury "Risk Aversion and Incentive Effects," *American Economic Review,* 92(5), 2002, pp. 1644-1655.

[20] Hörner, Johannes and Nicolas Sahuguet, "A War of Attrition with Endogenous Effort Levels," *Economic Theory*, Vol. 47, 2011, pp.1-27.

[21] Hungerford, Thomas and Gary Solon, "Sheepskin Effects in the Returns to Education," *Review of Economics and Statistics*, Vol. 69 (1), 1987, pp. 175-178.

[22] 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.

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

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

[25] Kim, Kyoo il, "Semiparametric Estimation of Signaling Games and Equilibrium Refinement," mimeo, 2012

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

[27] Mailath, George J. and Ernst-Ludwig von Thadden, "Incentive Compatibility and Differentiability: New Results and Classic Applications," *Journal of Economic Theory*, Vol. 148 (5): 1841–1861 (2013).

[28] Mas-Colell, Andreu, Michael Whinston and Jerry Green, *Microeconomic Theory*, Oxford University Press, 1995.

[29] Mazkin, Rosa "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, Vol. 60, 1992, 239-270.

[30] Milde, Hellmuth and John Riley, "Signaling in Credit Markets," *Quarterly Journal of Economics,* 103 (1), 1988, pp. 101-129.

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

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

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

[34] Paravisini, Daniel, Veronica Rappoport and Enrichetta Ravina "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios," mimeo, 2013.

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

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

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

[38] Roberts, James, "Unobserved Heterogeneity and Reserve Prices in Auctions," *RAND Journal of Economics*, forthcoming.

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

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