Signaling in Online Credit Markets[*]

Kei Kawai[†]   Ken Onishi[‡]

University of California Berkeley   Federal Reserve Board

Kosuke Uetake[§]

Yale School of Management

December 2020

Abstract

We study how signaling affects equilibrium outcomes and welfare in an online credit market using detailed data on loan characteristics and borrower repayment. We build and estimate an equilibrium model in which a borrower may signal her default risk through the reserve interest rate. Comparing a market with and without signaling relative to the benchmark with no asymmetric information, we find that adverse selection destroys as much as 34% of total surplus, up to 78% of which can be restored with signaling. We also estimate backward-bending supply curves for some markets, consistent with the prediction of Stiglitz & 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 & Shin (2012)). An important source of inefficiency in these markets lies in the inability of agents who are of good types to distinguish themselves from the bad, 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 costs of signaling can be induced to take actions that reveal their true type in equilibrium. Hence signaling can prevent the market from unraveling, with possibly large welfare implications.

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). These issues seem especially salient today given that many of the fastest-growing marketplaces are online, and participants typically know very little about each other. While reputation mechanisms are known to improve efficiency for markets with frequent transactions, signaling may be an important mechanism for preventing markets from unraveling in settings with relatively few repeat transactions.

This paper studies these issues 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. A key feature of Prosper.com during the period that we study 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 website.[1] The reserve rate serves as an upper bound on the actual interest rate that the borrower pays, but the former is often strictly higher than the latter as the actual contract interest rate is determined by the outcome of a bidding process. We first provide reduced-form evidence that the reserve rate signals the borrower's creditworthiness. We do so by studying how default probability varies with the reserve price conditional on the contract interest rate. In the second part of the paper, we build and estimate an equilibrium model of borrowers and lenders that allows us to compare outcomes and welfare with and without signaling.

---

[1]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.

2

The idea that the reserve rate can signal the borrower's 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 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 serves as a signal in the market we study, 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 rate on the funding probability and on the actual interest rate conditional on the listing being funded. Our estimates show that a high reserve rate leads to a higher funding probability, but it also leads to a less favorable contract interest rate on average. 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.[2]

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* . 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 privately- known type.

Our third set of regressions further explore the information value of the reserve rate by combining our data with additional data that we collect from an online field experiment. In the experiment, we ask subjects of the experiment to examine the listings and rate the creditworthiness of the borrowers. In the experiment, we randomly vary the reserve rate of the listing so that we can analyze how the reserve rate affects the perception of the borrowers' creditworthiness. Our regression estimates with listing fixed-effects suggest that higher reserve rates lower the creditworthiness rating, on average. This result suggests that the reserve rate is informative of the borrower's type and that our reduced-form results on the relationship between the reserve rate and the default probability are robust to unobserved heterogeneity.

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 cost of borrowing, which are privately known to the borrowers (and not to the lenders). The borrowers choose which interest rate to post, where the choice reveals their private information in equilibrium. Once the loan is originated, the borrower faces monthly repayment decisions.

---

[2]Note that many borrowers are likely to be 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.

We model the repayment decisions as a finite period dynamic problem. We also allow for latent borrower types to capture unobserved heterogeneity.

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 bid, and how much to lend.

In terms of identification, the key primitives of the model that we wish to recover are the distribution of the borrowers' latent types, the distribution of the borrowers' private information and the distribution of the lenders' attitude toward risk. The borrowers' latent types are identified because potential lenders observe the latent types and bid accordingly. As long as multiple bids are observed per listing, our setting becomes analogous to the mixture model considered in Hall & Zhou (2003). For identifying the borrowers' privately-known default costs, 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 identified by relating the expected return of listings to their funding probability.

In our counterfactual experiment, we compare the equilibrium market outcomes and welfare under three alternative scenarios – 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 & Weiss (1981), the credit supply curve in loan markets may be backward bending, or non-monotonic in the interest rate when there is adverse selection.[3] The counterfactual results for the pooling equilibrium support their prediction.

With respect to welfare, we find that the cost of adverse selection can be as much as 34% of the total surplus created under no asymmetric information. We also find that

---

[3]Recently, Arnold & Riley (2009) show that the credit supply curve cannot be globally backward-bending. They obtain this result under the assumption that the type of borrowers affects the variance of the return but not the mean. In our setting, the type of borrowers affects the mean as well as the variance. Hence, the mean return from a loan can be globally concave in our setting.

while signaling restores up to 78% of the difference in the surplus between pooling and no asymmetric information in some markets, it destroys welfare in others.

While the empirical findings of this paper directly apply only to the market of Prosper.com, many other markets have features that potentially allow for participants to signal private information. Examples include list prices for real estate and used car markets, reserve prices in auctions and buy-it-now prices for certain online P2P markets. There are at least two features of our setting that make it particularly attractive for the study of signaling. First is that we observe the ex-post loan performance of the borrowers and the second is that we can correlate the loan performance with the reserve price conditional on the contract interest. The latter feature is especially important as it allows us to isolate the informational value of the signal from other aspects of the signal that may interact with moral hazard.

Related Literature　　Our paper is related to several strands of literature. First, our study is related to the literature on adverse selection in credit markets. Since the seminal work of Stiglitz & Weiss (1981), there have been many studies testing for adverse selection in credit markets. Examples include Berger & Udell (1992), Ausubel (1999), Karlan & Zinman (2009), and Freedman & Jin (2010).[4] 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. A closely related paper that studies welfare and market design in credit markets is Xin (2020). She focuses on the role of reputation while we focus on signaling.[5]

The second strand of the literature to which our paper is related is the literature on 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 & Roberts (1986)), entry deterrence

---

[4]Freedman & Jin (2010) use data from Prosper.com. Other papers that also use data from Prosper.com include Rigbi (2013), Ravina (2019), Iyer et al. (2016), Liu et al. (2020), and Xin (2020). In Iyer et al. (2016), 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.

[5]Xin (2020) also uses data from Propser, but her data are from a more recent period. She studies reputation by focusing on repeat borrowers. Because our data are from an earlier period, the vast majority (97.0%) of our sample consists of first-time borrowers.

(e.g., Milgrom & Roberts (1982)), war of attrition (Hörner & Sahuguet (2011)), as well as credit markets (e.g., Bester (1985), Milde & Riley (1988)). Bester (1985) shows that borrowers can signal their type through the amount of collateral and Milde & 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 et al. (2007), and Jullien & Mariotti (2006), for example).[6] The signaling mechanism that we consider in this paper is very similar to those studied in Cai et al. (2007) and Jullien & Mariotti (2006).

In contrast to the large body of theoretical work, however, the empirical industrial organization literature on signaling is thin.[7] The few papers that study signaling in industrial organization include Gedge et al. (2020), Sahni & Nair (2020), and Backus & Tadelis (2019). Gedge et al. (2020) study limit pricing in the airline industry in which an incumbent signals its costs through the price. Sahni & Nair (2020) study advertising as a signal of service quality. Backus & Tadelis (2019) study round prices as a signal of sellers' willingness to cut prices on eBay.

Our paper is also related to the large empirical literature on screening. In particular, Adams et al. (2009) and Einav et al. (2012) are two papers that are closely related to our paper. They consider how an auto insurer can screen borrowers using the down payment. 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 heterogeneous lenders while their model has a single lender.

2 Institutional Background and Data

2.1 Institutional Background

Prosper.com is an online peer-to-peer lending website that matches borrowers with lenders and provides loan administrative services for the lenders. While the market

---

[6]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.

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

was relatively small during the period of our analysis, more than $18 billion in personal loans have been originated through their platform by the end of 2020. In this section, we describe how Prosper operated during the period of our analysis (May to October 2008), with a particular emphasis on the auction mechanism used to determine the interest rate. The auction format was used throughout the sample period until 2010 when Prosper switched to a non-auction format.[8]

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 website. 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 to Prosper's conditions and usury laws.[9] During our sample period, the maximum loan amount allowed on Prosper was $25,000 and the usury law maximum was 36%.

The characteristics of the borrower that appear on the listing page include credit grade, home-ownership status, debt-to-income ratio, the 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.[10] 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 website for potential lenders. If a potential lender finds a listing to which she wishes to lend money, she

---

[8]For a description of how Prosper allocates funds after 2010, see Xin (2020).

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

[10]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.

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 interest rate is 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 1 and 100% for the right panel). For listings that have yet to attract enough bids to reach the requested amount (i.e., a 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., a 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 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*

Figure 1: Funding Decision
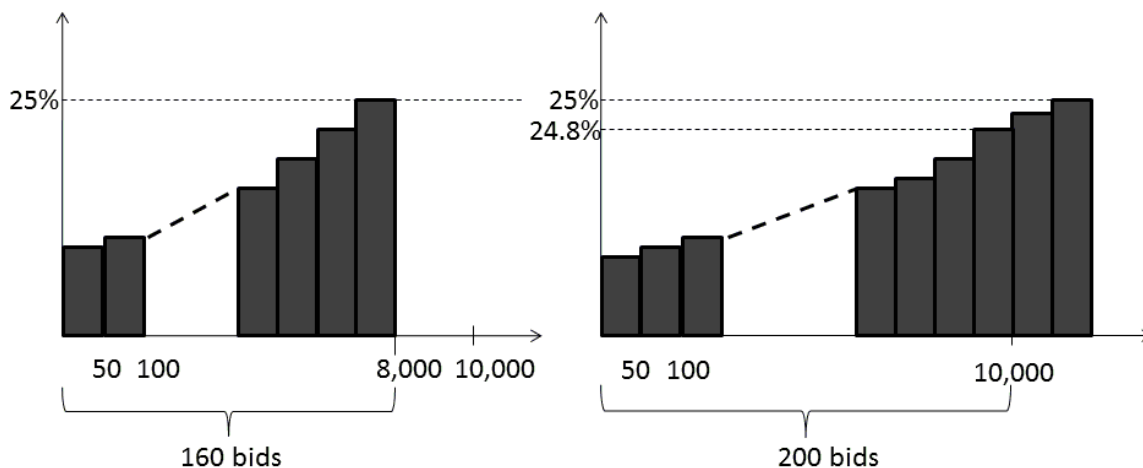
Note: The horizontal axis corresponds to amount and the vertical axis corresponds to the interest rate. The requested amount is $10,000 in both panels. The listing has received 160 bids ($8,000) in the left panel and 200 bids in the right.

amount. Hence the borrower would receive no loan in the situation depicted in the left panel of Figure 1.[11]

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 1, the listing is funded at 24.8% and the same rate applies to all lenders who submitted bids below 24.8%.

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.[12] 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.

---

[11]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%.

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

10

2.2 Data

The data for our analysis come directly from Prosper.com. The data contain virtually all the information on the listing that is available to the potential lenders as well as the ex-post performance of the loans. 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.[13] 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, canceled by Prosper, or missing parts of the data. We are left with a total of 35,241 listings, of which 5,571 were funded. Section 2.2 of the online appendix contains a more detailed description of our data construction.

Panel (A) of 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 more than $13,000 for AA listings to 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. 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.

Panel (B) of Table 1 reports the sample statistics of listings that were *funded*, which constitute a subset of the set of listings. Note that the mean loan amount reported in panel (B) is smaller than the corresponding mean requested amount shown in Panel (A), 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 Panel (B) is higher than in Panel (A), for the obvious reason that listings need to attract a 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.

In Figure 2, we present the distribution of the reserve rate across different credit grades. As expected, the reserve rate is higher for worse credit grades. One important point to note is that there is a mass at 36% for credit grades B and below. This is because 36% was 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

---

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

Table 1: Descriptive Statistics

Grade	(1) Amount Requested	(2) Reserve Rate (%)	(3) Debt/ Income	(4) Home Owner	(5) Bid Count	(6) Fund Pr.(%)	(7) Contract Rate (%)	Obs.
Panel (A): Listing Characteristics								
AA	13,145	13.2	0.36	0.81	153.2	53.4		1,420
	(8,343)	(4.7)	(0.98)	(0.39)	(178.6)			
A	12,396	16.5	0.38	0.61	103.3	40.9		1,850
	(7,882)	(6.7)	(0.67)	(0.49)	(139.9)			
B	10,622	21.1	0.39	0.59	74.4	33.4		3,068
	(6,096)	(7.5)	(0.65)	(0.49)	(104.0)			
C	7,622	24.6	0.37	0.56	36.2	24.7		5,203
	(5,158)	(7.8)	(0.62)	(0.50)	(59.2)			
D	6,368	28.7	0.39	0.37	18.2	15.5		6,581
	(4,691)	(7.5)	(0.71)	(0.48)	(39.0)			
E	4,783	31.0	0.36	0.33	3.99	6.8		5,757
	(4,868)	(7.3)	(0.68)	(0.47)	(11.0)			
HR	4,351	31.5	0.31	0.22	1.9	3.0		11,362
	(4,599)	(6.9)	(0.64)	(0.42)	(5.7)			
All	6,604	27.4	0.35	0.39	28.0	15.8		35,241
	(5,938)	(8.9)	0.68	(0.49)	(74.4)			
Panel (B): Loan Characteristics								
AA	9,710	13.1	0.21	0.80	258.1		9.6	755
	(7,384)	(3.3)	(0.39)	(0.39)	(186.2)		(4.6)	
A	8,723	16.5	0.23	0.55	218.1		12.7	755
	6,626	6.0	(0.14)	(0.14)	(155.4)		(4.5)	
B	7,347	21.6	0.27	0.56	181.2		16.4	1,023
	(4,858)	(6.3)	(0.34)	(34)	(113.9)		(4.6)	
C	4,687	24.7	0.25	0.48	112.6		18.1	1,023
	(2,998)	6.4	(0.21)	(0.21)	(72.2)		(6.2)	
D	3,578	28.0	0.24	0.26	86.7		21.0	1,022
	(2,380)	(6.4)	(0.17)	(0.17)	(60.8)		(6.6)	
E	1,890	33.9	0.22	0.26	36.4		29.1	392
	(1,187)	(2.8)	(0.22)	(0.22)	(20.3)		(5.7)	
HR	1,690	33.9	0.20	0.17	27.0		30.0	339
	(1,288)	(3.6)	(0.44)	(0.44)	(19.0)		5.7	
All	5,821	23.3	0.24	0.47	144.0		17.9	5,571
	(5,285)	(8.6)	(0.28)	(0.28)	(132.4)		(7.9)	

Note: Panel (A) reports summary statistics for listings and Panel (B) reports summary statistics for the subset of listings that are funded.

Figure 2: Distribution of Reserve Interest Rate by Credit Grade
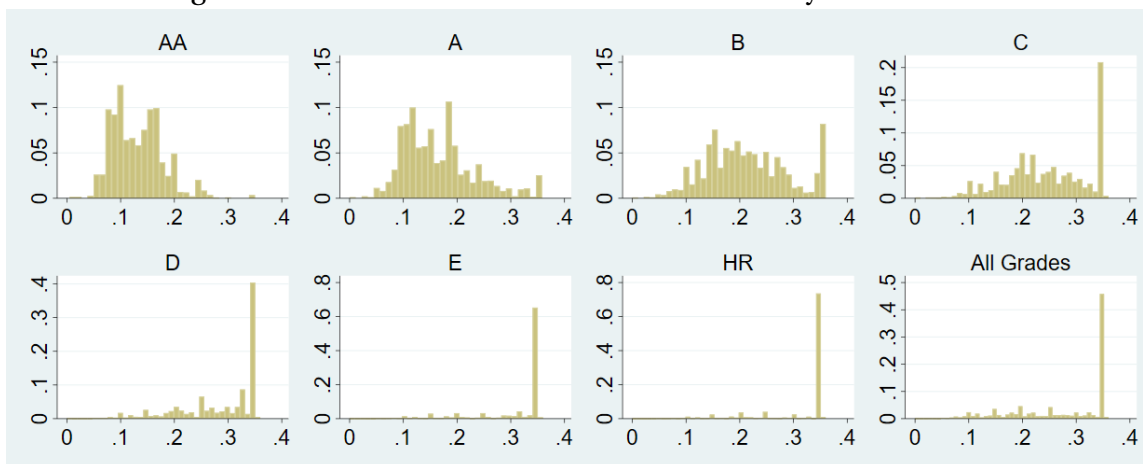
Note: The reserve interest rate is capped at 36% because of the usury laws.

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 E 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.[14]

In Figure 3, 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 $s = 20\%$.[15] The other panels of Figure 3 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 4, we report the distributions of the bid amount, again by credit grade. The

---

[14]As we document in Section 3.3 of the online appendix, the relationship between the reserve rate s and the funding probability is non-monotonic for a relatively large region of s for credit grade D. This suggests that conditions required for the reserve rate to serve as a signal are not met for credit grade D. Hence, we also drop credit category D from our structural analysis.

[15]Given that almost no borrowers in credit grade AA post a reserve rate of 30% or more (see Figure 2), we only report the distribution for s equal to 10% and 20% for credit grade AA.

Figure 3: Distribution of Contract Interest Rate Given Reserve Interest Rate
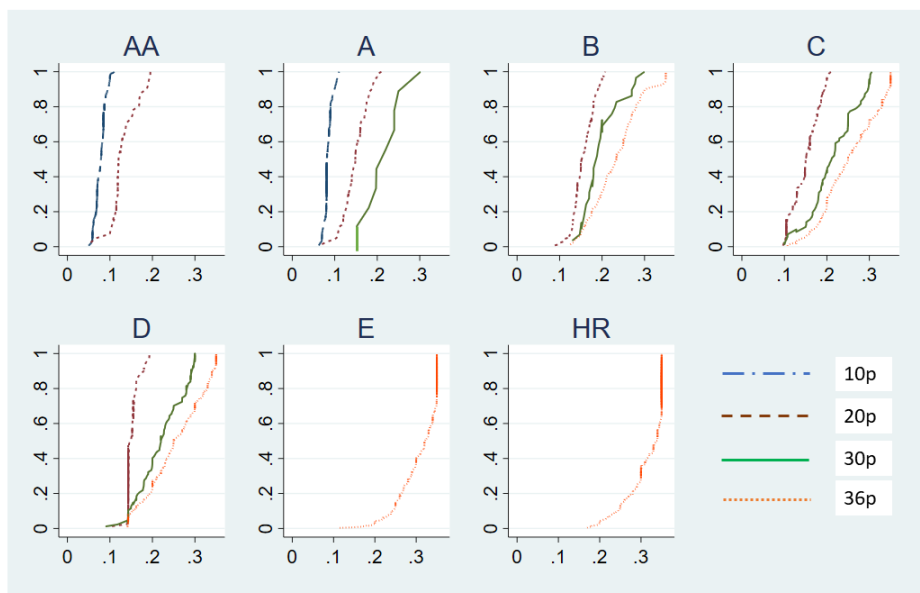
Note: The figure plots the distribution of contract interest rate conditional on reserve interest rate $s = 10\%, 20\%, 30\%$, and 36%. The distribution for $s = 10\%, 20\%, 30\%$ are computed by pooling funded listings with $s \in [s-1, s+1]$. For $s = 36\%$, the distribution is computed by pooling funded listings with $s \in [34\%, 36\%]$.

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.

For each loan originated by Prosper, we have monthly data regarding the repayment decisions of the borrower, i.e., we observe whether or not the borrower repaid the loan every month. In the first column of Table 2, 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 2 also reports the mean and the quantiles of the internal rate of return (IRR) of the loans. 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 are

Figure 4: Distribution of Bid Amount
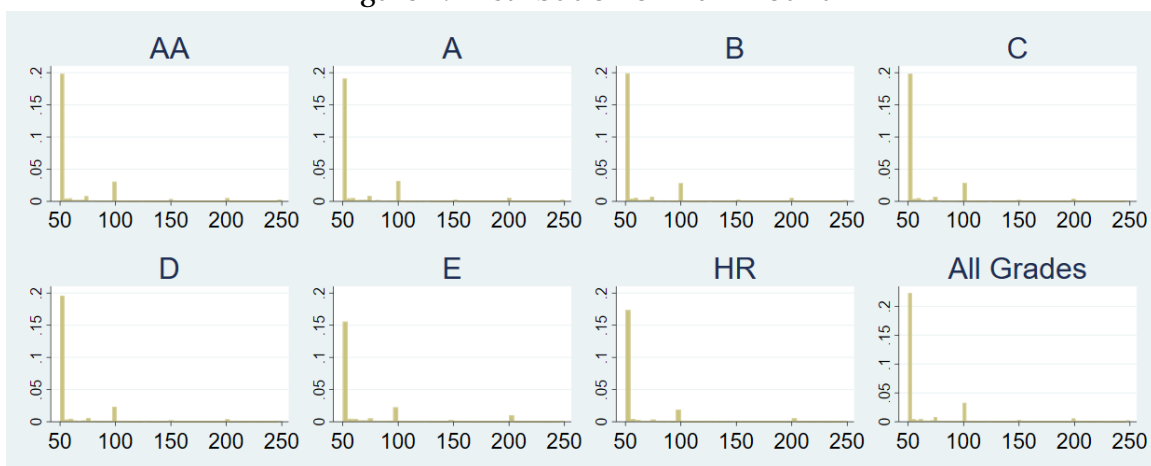
Table 2: Default Probability and Internal Rate of Return (IRR)

Grade	Default	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

likely to reflect the fact that our sample coincides with the period of economic downturn during the financial crisis. As a comparison, the average rate of return on loans originated by Prosper after July 1, 2009 has been consistently above 3%.

Finally, Table 3 reports the summary statistics of the lenders. We find that lenders fund, on 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 in Prosper in general. The mean IRR of a lender's portfolio is about −4.1%. As we report in the online appendix 2.3, we do not find a statistically significant relationship between lender portfolio size and IRR.

Table 3: Lender Portfolio Characteristics.

	Mean	sd	5%	25%	50%	75%	95%	Obs
# of Loans	17.5	42.4	1	2	6	16	68	29,176
Total Amount	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

3 Evidence of Signaling Through the Reserve Rate

In this section, we provide some reduced-form evidence that the borrower's reserve interest rate choice serves as a signal of the borrower's creditworthiness. 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 then discuss the results of a field experiment that we ran which complements our analysis. The design of the experiment allows us to test more directly whether or not the reserve rate is likely to contain information that affects lender beliefs. Taken together, we argue that the reduced-form results support the notion that the reserve rate serves as a signal.

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

$$\text{Funded}_j = 1\{\beta_s s_j + \mathbf{x}'_j \beta_x + \varepsilon_j \geq 0\}, \tag{3.1}$$

where Funded_j is a dummy variable for whether or not listing j is funded, s_j is the reserve rate and $\mathbf{x}_j$ is a vector of borrower characteristics that includes the requested amount, the debt–to–income ratio, home-ownership, as well as other borrower credit information such as current credit inquiries, delinquencies and bank card utilization, etc. We also include word counts of the listing description and title as well as dummy variables for credit grade, self-reported occupation and income brackets.

The first column of Table 4 reports the results of this regression.[16] 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 3.57 and it is statistically significant. In terms of the

---

[16]We do not report coefficient estimates for all variables because of space constraints. See Table 3 of Section 3.1 in the online appendix for the full results.

marginal effect, a 1 percentage point increase in s_j is associated with about a 0.32 percentage point increase in the funding probability. The pseudo R^2 decreases from 0.370 to 0.292 when we exclude s_j. In the online appendix, we estimate the funding probability as a nonparametric function of s by credit grade. Section 3.3 of the online appendix reports the results.

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

$$
\begin{aligned}
r_j^* &= \beta_s s_j + \mathbf{x}_j' \beta_x + \varepsilon_j, & (3.2) \\
r_j &= \begin{cases} r_j^* & \text{if } r_j^* \le s_j \\ \text{missing} & \text{otherwise} \end{cases} .
\end{aligned}
$$

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. The variable 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 estimates of the regression in the second column of Table 4. We find that our estimate of β_s is positive and significant, which suggests 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 (3.2) suggest that lenders charge higher interest to riskier borrowers. We note that the pseudo R^2 is above 1 which can occur when the outcome variable is a mixture of discrete and continuous distributions. The pseudo R^2 is 0.907 when we drop s_j from the first expression in (3.2).

The results of regressions (3.1) and (3.2) 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. It is very likely that borrowers are actually aware of this trade-off: In a prominently displayed tutorial,

17

Table 4: Reduced–Form Analysis: Funding Probability, Contract Rate and Repayment

	(1) Funded	(2) Contract Rate	(3) Ever Default	(4) IRR
Reserve Rate	3.57***	0.56***	2.75***	-0.48***
	(0.19)	(0.015)	(0.79)	(0.13)
Contract Rate			2.79***	0.29**
			(0.80)	(0.14)
Amount ($1000)	-0.13***	8.71E-03***	5.92E-02***	-6.90E-03***
	(2.80E-03)	(2.10E-04)	(8.37E-03)	(1.33E-03)
Debt/income	-0.96***	7.63E-02***	0.162	-0.0334
	(0.05)	(0.39E-02)	(0.113)	(0.0205)
Home owner	-0.05*	3.57E-03*	0.34***	-0.05***
	(0.0245)	(1.83E-03)	(0.08)	(0.01)
Grade				
AA	3.54***	-0.26***	-1.01***	0.08*
	(0.08)	(0.01)	(0.26)	(0.04)
A	2.91***	-0.23***	-0.62**	0.07
	(0.07)	(0.01)	(0.23)	(0.04)
B	2.29***	-0.19***	-0.31	0.03
	(0.05)	(0.01)	(0.19)	(0.03)
C	1.60***	-0.15***	-0.22	0.04
	(0.04)	(0.04)	(0.17)	(0.03)
D	1.01***	-0.10***	-0.14	0.05
	(0.04)	(0.01)	(0.15)	(0.03)
E	0.37***	-0.03***	-0.27	0.10**
	(0.04)	(0.01)	(0.16)	(0.03)
Other Credit Info	X	X	X	X
Word Count	X	X	X	X
Occupation/Income	X	X	X	X
Time dummies	X	X	X	X
Observations	35236	35236	5530	5530
R^2	0.370	1.380	0.118	0.071
R^2 w/o s_j	0.292	0.907	0.105	0.043
Log likelihood	-9646	330	-4717	

Note: We include several variables in the regression whose coefficients we do not report because of space constraints. These variables include bank card utilization, number of current and total credit lines, number of credit inquiries, amount delinquent, etc. We also include dummies for self-reported occupation and income levels as well as word counts of the title and description of the listing. See Table 3 of the Online Appendix for the full results. The row labeled R^2 reports the R^2 or pseudo R^2 of McFadden. The row labeled R^2 w/o s_j is the same as the row above without the reserve rate in the model. *, **, and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are reported in parenthesis.

Prosper informs the borrowers that choosing 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 borrower private information 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 loans. In order to do so, we run a Probit regression of an indicator of default on the the reserve rate, contract interest rate, and other controls:

$$\text{Default}_j = 1\{\beta_s s_j + \beta_r r_j + \mathbf{x}'_j \beta_x + \varepsilon_j \geq 0\}, \tag{3.3}$$

where the regressand, Default_j, is a dummy variable that takes a value of 1 if borrower j defaults on the loan. 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 of this regression are reported in column (3) of Table 4. Our estimate of β_s is positive and significant in column (3). In terms of the marginal effect, a 1 percentage point increase in s_j is associated with a 0.54 percentage point increase in the default probability. This means that 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 of the table, 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 bigger loans may be less creditworthy, or bigger loans may induce borrowers to default more often because of higher interest payments.[17] The former explanation would be consistent with signaling, and the latter would be consistent with moral hazard.[18]

We next 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:

$$IRR_j = \beta_s s_j + \beta_r r_j + \mathbf{x}_j' \beta_x + \varepsilon_j, \tag{3.4}$$

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 (3.3), 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 4. As expected, the coefficient on s_j ($\beta_s = -0.48$) is negative and significant. This is consistent with the result of regression (3.3) where we find that borrowers who post high reserve rates tend to default more often. The coefficient on r_j is also positive and statistically significant suggesting that an increase in the contract interest rate increases the rate of return on average.

We emphasize that our regressions control for a large number of covariates related to the borrower's credit information such as bank card utilization, current and past delinquencies on payments, etc. We also control for self-reported income levels and occupation as well as the length of the title and the description. The fact that we find coefficient on s_j to be significant even after controlling for a large number of variables related to the borrower's credit information suggests that the reserve rate is informative about the borrower's type. We also note that the difference in R^2 with and without s_j is

---

[17]Some of the results in Iyer et al. (2016) suggest that ex-post moral hazard is relatively modest in this setting.

[18]One may wonder why the coefficient on home-ownership is positive. One hypothesis is selection. It may be the case that most homeowners have easy access to credit outside of Prosper. Conditional on having to a borrower on Prosper, home-ownership 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 the coefficient on home-ownership comes out positive. We note that there are other studies in similar contexts that also find a negative relationship between home-ownership and creditworthiness. For example, Adams, Einav & Levin (2009) find that home-ownership increases missed payments on subprime auto loans in one of their specifications.

especially large for the IRR regression (0.071 and 0.043).

Unobserved Heterogeneity   Although we include a rich set of covariates in the regressions, there may be a correlation between the reserve rate s and listing characteristics that are omitted from the regressions biasing the coefficient on s away from zero. In order to address this concern, we show in Section 3.2 of the online appendix that our results are robust to alternative ways of controlling for covariates and restricting the estimating sample. We also conduct a survey in which we ask multiple choice questions regarding the respondent's impressions of a listing by showing pictures and text information from the listing. The purpose of the survey is to construct a simple measure that summarizes the richness of the information contained in texts and pictures. We find that inclusion of these measures as well as other specifications yield similar results. In particular, we find that the sign and significance of the coefficient on the reserve rate remain largely unchanged.

In order to further address concerns regarding unobserved heterogeneity, we conduct a field experiment on Mechanical Turk in which we ask the subjects of the experiment to evaluate the creditworthiness of the borrowers on a scale of 1 (least creditworthy) to 7 (most creditworthy) by showing listings drawn randomly from our sample.[19] We show the subjects listing characteristics such as the amount requested, credit grade, debt-to-income ratio, home-ownership, as well as the description of the listing. We also show the subjects the borrower reserve rate, but we randomize this aspect of the listing in order to study how perceived creditworthiness varies as we vary the reserve rate. The experimental design implies that the reserve rate that is shown to the respondents can be higher or lower than the actual reserve rate. Section 3.4 of the online appendix explains the details regarding the implementation of the field experiment.

Table 5 reports the summary statistics of the evaluations reported separately by credit grade. The first two rows correspond to listings and the next two rows correspond to those that were funded. The table shows that the mean evaluations for listings are perfectly ordered by credit grade with average evaluations at 5.77 for credit grade AA and lower for worse credit grades. We also find that funded loans generally have higher evaluations than unfunded loans within a given grade. The alignment of the evaluations with the credit grades and the outcome of the listing suggests that the

---

[19]Our design is motivated by Ravina (2019) who studies the effects of personal attributes on loan outcomes on Prosper.

evaluations are informative about the borrowers' creditworthiness.

Table 5: Descriptive Statistics–Evaluations

	AA	A	B	C	D	E	HR	All
All Listings	5.767	5.323	4.892	4.320	3.726	3.412	3.267	3.863
	(1.470)	(1.433)	(1.433)	(1.435)	(1.561)	(1.694)	(1.831)	(1.800)
# Obs.	810	923	1,553	2,508	3,232	2,801	6,632	18,459
Funded Listings	5.853	5.401	5.020	4.473	3.841	3.296	3.411	4.711
	(1.383)	(1.361)	(1.301)	(1.320)	(1.529)	(1.574)	(1.620)	(1.621)
# Obs.	415	403	511	514	415	135	180	2,573

We now report estimates from regressions in which we take the evaluations as the dependent variable and the reserve rate shown to the respondents as the independent variable. Column (1) of Table 6 reports the estimates of a regression with listing fixed effects and column (2) reports the estimates of a regression with listing and respondent fixed effects. Respondent fixed effects control for heterogeneity in the respondent's average evaluations. We find that the coefficients on the reserve rate are negative and statistically significant at the 5% level and 1% level, respectively. A standard deviation increase in the reserve rate decreases the respondent's evaluation by about 0.035 points and 0.042 points, respectively (the mean evaluation is about 3.86). These results suggest that the reserve rate has a direct impact on the lenders' beliefs.

Table 6: Evaluation of Borrower Creditworthiness and Reserve Rate

	(1) Evaluation	(2) Evaluation
Reserve Rate (%)	−0.389**	−0.469***
	(0.166)	(0.144)
Listing FE	X	X
Respondent FE		X
Observations	18,459	18,459

Note: ** and *** respectively denote significance at the 5%, and 1% levels. Standard errors are reported in parenthesis.

Interpretation of the Results   The results of Table 4 and 6 suggest that (1) the 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 and charge higher interest to borrowers who post high reserve rates.

Taken together, these results are informative about how signaling is sustained in equilibrium. Less creditworthy borrowers are more willing to sacrifice a favorable interest rate for a higher 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 type from mimicking the other and sustains a separation of types through signaling.

4 Model

In this section, we develop a model of borrowers and lenders on Prosper, which we then take to the data. Our model has three parts: (1) reserve rate choice of the borrowers; (2) lenders' bidding behavior; and (3) borrowers' repayment behavior. Features of the model include asymmetric information between borrowers and lenders and latent borrower types that capture unobserved borrower heterogeneity. We collect and discuss our modeling choices and assumptions at the end of this section. The online appendix contains a numerical example with two types of borrowers that illustrates graphically the forces that make it possible to sustain a signaling equilibrium.

4.1 Borrowers

Borrower Repayment We first describe the repayment decisions of the borrowers and work our way backwards. We model the monthly repayment decisions of the borrowers as a sequential decision of 36 periods (36 months) which is the length of the loans that Prosper originates. We write the decision of the borrower at the terminal period as follows:

$$
\begin{cases}
\text{full repayment: if } u_T(r) + \varepsilon_T \geq D(\varphi) \\
\text{default: otherwise,}
\end{cases}
\tag{4.1}
$$

where $u_T(r) + \varepsilon_T$ denotes the period utility of the borrower if she repays the loan in full, r denotes the interest rate on the loan, and φ denotes the (privately-known) type of the borrower that determines the likelihood of repaying the loan. While there are many ways to interpret φ and $D(\cdot)$, we adopt the interpretation of $D(\varphi)$ as the default

cost of a borrower whose type is equal to φ.[20] 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 listing characteristics. 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 $\mathbb{E}[\varepsilon_T | \varphi] \neq 0$, we can subtract $\mathbb{E}[\varepsilon_T | \varphi]$ from both sides of equation (4.1) and by appropriately redefining $D(\cdot)$ and ε_T, we have an observationally equivalent model with $\mathbb{E}[\varepsilon_T | \varphi] = 0$. This is possible because we allow $D(\cdot)$ (or equivalently, the distribution of φ) to be nonparametric.[21] While mean independence is not the same as independence, we think that this alleviates some of the concerns regarding our assumption.

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

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

where $u_t(r) + \varepsilon_t$ is the period t utility of repaying the loan, β is the discount factor, and $V_{t+1}(r, \varphi)$ is the continuation utility, which can be defined recursively. We 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 borrower/listing characteristics. This is purely for expositional purposes. In our identification and estimation, we let u_t, F_{ε_t}, and F_φ depend on observable characteristics as well as on latent types.

---

[20]Note that φ is the privately-known 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.

[21]Intuitively, if $\mathbb{E}[\varepsilon_T | \varphi] \neq 0$, one can load on φ the part of ε_T that is correlated with φ.

Borrower Reserve Rate Choice   Now we describe our model of the borrower's reserve rate choice. When the borrower determines the reserve rate, s, she 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 to maximize the ex-ante value function V_0 subject to the usury law limit of 36%, as follows:

$$\max_{s \leq 0.36} V_0(s, \varphi) = \max_{s \leq 0.36} \left[\Pr(s) \int V_1(r, \varphi) f(r|s) d r + (1 - \Pr(s)) \lambda \right], \qquad (4.2)$$

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 a good outside option, e.g., borrowing money from relatives, friends, and local banks, etc. Borrowers with low values of λ have a 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.

The first term in the bracket in equation (4.2) captures the borrower's expected utility in the event of obtaining a loan through Prosper: $V_1(r, \varphi)$, 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. 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 λ. 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 λ.

The first–order condition associated with problem (4.2) is as follows,

$$\frac{\partial}{\partial s} \Pr(s) \left(\int V_1(r, \varphi) f(r|s) d r - \lambda(\varphi) \right) + \Pr(s) \int V_1(r, \varphi) \frac{\partial}{\partial s} f(r|s) d r = 0, \qquad (4.3)$$

for an interior solution. Equation (4.3) captures the trade-off that the borrower faces in determining the reserve rate. 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.,*

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

The formal proof of this proposition as well as those of all other propositions are in Section 9 of the online appendix. To see the intuition for why SCP holds, consider the marginal utility from increasing s 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 2 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 Section 4.4 of 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 a 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.[22] 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 $\{0, 1, \cdots, \overline{N}\}$, where $\overline{N}$ is the maximum number of potential lenders. The potential lenders are heterogeneous with regard to their attitude toward risk and with regard to their opportunity cost of lending.

Each potential lender must decide whether or not to submit a bid 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 2.2, the vast majority of the bid amounts are either $50, or $100. In the estimation, we assume that lenders face a discrete set of amounts $M = \{\$50, \$75, \$100, \$200, \$500, \$1000\}$ 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 or not to bid 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

---

[22]While the mass at 36% for credit grade D is not as large as that for credit grades E and HR, the relationship between s and the funding probability is non-monotonic for a relatively large region of s for credit grade D (See Section 3.3 of the online appendix). Hence, we also drop credit category D from our analysis.

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:

$$U = \tilde{U}_j^L(Z(r)) - \tilde{\epsilon}_{0j}, \text{ where } \tilde{U}_j^L(Z(r)) = \tilde{\mu}_j(Z(r)) - A_j \tilde{\sigma}_j^2(Z(r)) - c.$$

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

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

The error term in the utility ϵ_{0j} now includes lender forecasting error as well as $\tilde{\epsilon}_{0j}$. $\mu(Z(r))$ and $\sigma^2(Z(r))$ are the mean and variance of the return computed under rational expectations.[23]

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 and ϵ_{0j} subsumes any lender specific forecasting error. 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 $\mu_j(Z(r)) = \mu(Z(r)) + \epsilon_{\mu j}$ and $\epsilon_{\mu j}$ has positive mean. However, the independence assumption implies that

---

[23]Note that the mean and variance depend on the characteristics of listing Z. 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.

Figure 5: Graphical Representation of the Lender's Problem: Case of No Amount Choice
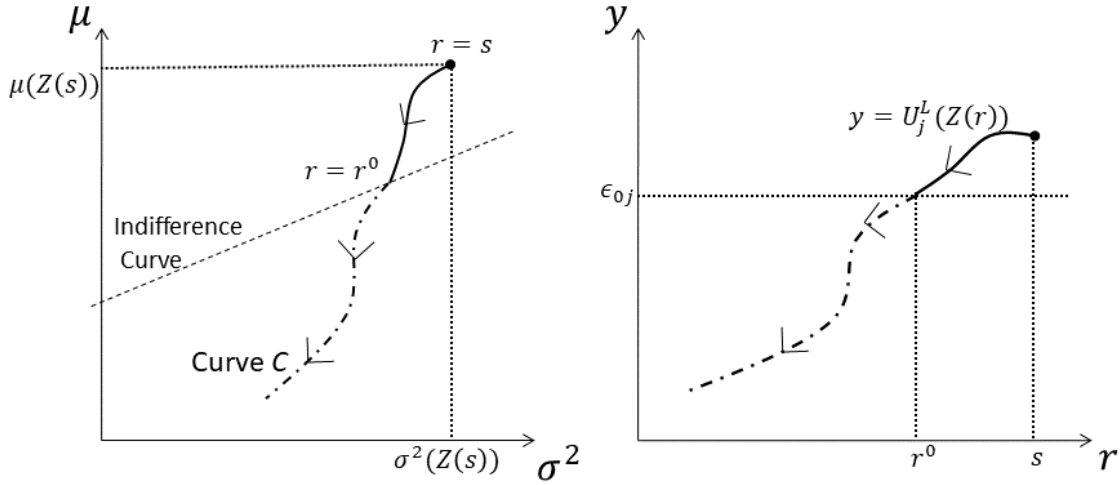
Note: In the left panel, the horizontal axis is σ^2 and the vertical axis is μ. Curve C corresponds to the mean and variance of a listing for different realizations of r. The dashed line is the lender's indifference curve. The right panel plots $U_j^L(Z(r))$ against r.

the lenders' beliefs over the variance coincide with the realized variance. We relax the independence assumption by allowing the lenders' beliefs to depend on the mean and volatility index of the S&P 500. Section 8.2 in the online appendix reports the results.

In order to understand the lender's problem, it is useful to study it graphically. Figure 5 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 r, consider the mean and the variance of the return. We can plot the points $(\mu(Z(r)), \sigma^2(Z(r)))$ on this $\mu - \sigma^2$ plane for each r.

Curve C in the left panel of Figure 5 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 5. 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., $\mu - A_j \sigma^2 - c - \epsilon_{0j} = 0$. 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 5, 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 the exposition. Our Proposition 3, which covers the case with the lender's 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.*

The formal proof is in Section 9 of the online appendix. The reason 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.

While it is certainly restrictive, we think that assuming that the lender behaves as if she 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 144 winning bids; see Table 1. 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 the lender behaves as if she will not be pivotal. Section 4.3 of 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, in which the borrower chooses q from a discrete set M or chooses not to bid. If the lender bids amount q to listing Z at contract interest rate r, then $\mathbb{E}[qZ(r)] = q\mu(Z(r))$ and $Var(qZ(r)) = q^2\sigma^2(Z(r))$. Hence, lender j's utility can be expressed as follows,

$$U = U_j^L(qZ(r)) - \epsilon_{0j} = q\mu(Z(r)) - A_j q^2 \sigma^2(Z(r)) - c_q - \epsilon_{0j}, \tag{4.4}$$

where the cost of lending now depends on q as c_q.[24]

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 6 for the case in which $M = \{\$50, \$100, \$200\}$.[25] The three curves in Figure 6 correspond to $U_j^L(50Z(\cdot))$, $U_j^L(100Z(\cdot))$, and $U_j^L(200Z(\cdot))$. Just as before, there is a (weakly) dominant strategy for the lender under the assumption that the bidder is not pivotal. For the case shown in Figure 6, 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 $U_j^L(q'Z(\cdot))$ for $q' \neq q$. We now state the above analysis in the form of a proposition.

---

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

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

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 $U_j^L(q(k)Z) - \epsilon_{0j} \geq U_j^L(q'Z) - \epsilon_{0j}$ for all q' and $r \in I_k$.[26] 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.*

Figure 6: Graphical Representation of the Lender's Problem: Case with Amount Choice
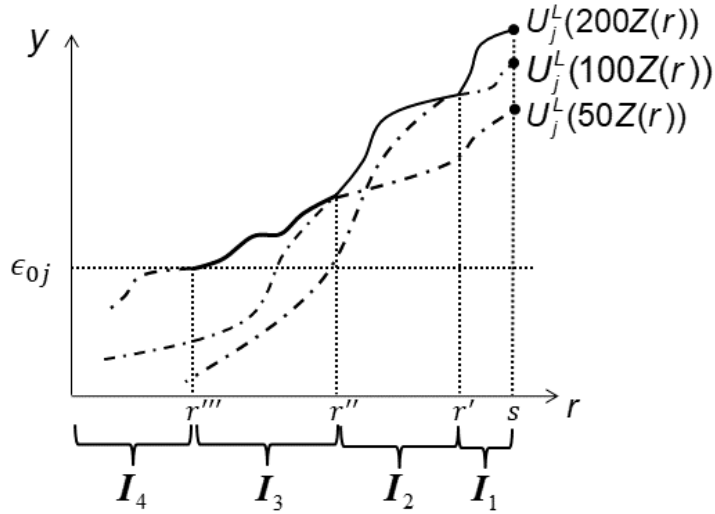
Note: Each curve $U_j^L(qZ(r))$ illustrates the relationship between r and the lender's utility net of ϵ_{0j} when the lender bids q. I_1 corresponds to the region of the active interest rate for which bidding \$200 is optimal. I_2, I_3, and I_4 correspond to the regions of the active interest rate for which bidding \$100, \$50, and \$0 is optimal, respectively.

4.3 Equilibrium

We now discuss equilibrium existence and uniqueness. There always exists an equilibrium of the model we described, but there may not exist a separating equilibrium. Mailath (1987) (see also Mailath & 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:

$$\frac{\partial}{\partial s} V_0(s, \varphi, \tilde{\varphi}; X) \bigg/ \frac{\partial}{\partial \tilde{\varphi}} V_0(s, \varphi, \tilde{\varphi}; X) \text{ is monotone in } \varphi, \forall X, \quad (4.5)$$

---

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

where $V_0(s, \varphi, \tilde{\varphi}; X)$ is the borrower's expected utility from posting a reserve interest rate s, when the borrower is of type φ, and the lenders perceive her to be of type $\tilde{\varphi}$. X is a vector of borrower and listing characteristics observed to the players.

Because V_0 is a value function of a dynamic decision problem, analytically characterizing the set of parameters that satisfy condition (4.5) is difficult. However, condition (4.5) is very intuitive. The numerator corresponds to the gain in borrower's utility from an increase in the reserve rate s, holding fixed lender beliefs. This utility gain occurs through the borrower's ability to obtain a loan at interest rates above s. Increases in the funding probability at high interest rates are more valuable for low types than for high types. The denominator in condition (4.5) corresponds to the utility gain when lenders perceive the borrower to be a better type holding fixed s. When lender beliefs improve, contract rate r tends to become lower, although the upper bound of the distribution of r stays constant at s. This change tends to be more valuable for high types. These two forces suggest that $\frac{\partial V_0}{\partial s} / \frac{\partial V_0}{\partial s}$ to be monotone decreasing in φ.

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.[27] At the estimated parameter values, the conditions seem to generally hold.[28]

As for the uniqueness of equilibrium, signaling models generally admit multiple equilibria because there are always pooling equilibria in which no information is transmitted. In our setting, D1 refinement is sufficient for guaranteeing a unique separating equilibrium because it implies the initial value condition of Mailath (1987).[29]

4.4 Model Discussion

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

---

[27]The reason why we do not include this condition in our estimation routine is that 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.

[28]For worse credit grades, condition (4.5) 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 (4.5) need not hold close to 36%.

[29]We acknowledge that D1 is a relatively strong refinement criterion, 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.

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., $\mathbb{E}[\varepsilon_t|\varphi]=0$, is without loss of generality. This is because we can always redefine ε_t as $\varepsilon_t - \mathbb{E}[\varepsilon_t|\varphi]$ and $D(\cdot)$ as $D(\cdot) - \mathbb{E}[\varepsilon_t|\cdot]$ so that $\mathbb{E}[\varepsilon_t|\varphi]=0$. 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 is a sequence of binary decisions (repay or default) for each borrower, in which default is an absorbing state: If a borrower defaults, we do not observe any repayment decisions from that point on. Unlike in a situation where there are distinct decisions for each of the T periods (i.e., no absorbing state), only the marginals of $\{\varepsilon_t\}$ are relevant for data generation precluding us from identifying correlation across $\{\varepsilon_t\}$. While this may appear to be a limitation, this means that as long as $\{\varepsilon_t\}$ is structural to our counterfactual policy, the counterfactual policy outcomes are robust to the independence assumption.

Interpretation of φ   Recall that the unobservable type of the borrower (φ) is interpreted as the default cost in our model. A natural alternative would be to model the private information of the borrowers as the income/asset of the borrower. In Section 4.2 in the online appendix, we discuss an alternative model in which φ is modeled as unobserved income/assets and show that the alternative model is isomorphic to our model. 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   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 (4.2) conditional on the optimal choice of the loan amount, treating the loan size as a covariate does not bias our estimates. We discuss this point in detail in Section 4.1 of the online appendix. 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.

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

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.[30]

Cost of revising the bid   The optimal strategy described in Proposition 3 requires the lenders to submit new bids as the active interest rate changes. In the scenario depicted in Figure 6, 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 a low cost of revising their bids.[31] 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 track the active interest rate or the exact timing of the bids. Our estimation only uses moments that are not too sensitive to this assumption as we describe below.

---

[30]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.

[31]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 6.

5 Identification

We now discuss identification of the model. Until now, it has not been important to distinguish between covariates that are observed only to the players and those that are observed to both the econometrician and the players. For identification, this distinction is important. We partition the covariates X in two as $X = (X^o, X^u)$ where X^o denotes observed characteristics and X^u denotes latent characteristics. In particular, we focus on the case in which X^u is a binary variable ($X^u \in \{L, H\}$) that captures the latent type of the borrower. Our identification argument does not hinge on the assumption that there are only two latent types, but the assumptions needed for identification are more transparent for this case.[32] We estimate the model with two types in our application.

We first consider identification of $\Pr(X^u | s, X^o)$, the distribution of the latent variable conditional on s, X^o. The intuition for identification is very simple. From the perspective of the lender, the borrowers are either $X^u = L$ or $X^u = H$. Now consider the interest rates of the first k (≥ 3) bids $\mathbf{r} = (r_1, \cdots r_k)$ submitted *before the listing becomes fully funded.* If bidders bid according to Proposition 3, the bids are statistically independent ($r_j \perp r_{j'}$). Note that this is not the case for bids that are submitted after the listing becomes funded, as the marginal bid, which is observed to the lender, may change over time.

The joint distribution of $\mathbf{r}$ evaluated at τ can be expressed as follows:

$$F_{\mathbf{r}|s,X^o}(\tau) = \Pr(L|s, X^o) \prod_{j=1}^{k} G_L(\tau_j | s, X^o) + \Pr(H|s, X^o) \prod_{j=1}^{k} G_H(\tau_j | s, X^o),$$

where $G_{X^u}(\tau_j | s, X^o)$ denotes the conditional probability that the interest bid by the bidder is less than $\tau_j \in \mathbb{R}$ given reserve rate s and type X^u. This is a special case of the finite mixture model considered in Hall & Zhou (2003). Hence, applying Theorem 4.3 of Hall & Zhou (2003), the mixing probabilities and $G_{X^u}(\tau_j | s, X^o)$ are both identified under an irreducibility condition and $k \geq 3$.[33] The irreducibility condition is a fairly weak condition that is satisfied generically.

---

[32]With two latent types, the results of Hall & Zhou (2003) apply. With more than two types, Kruskal's theorem (Kruskal 1977) or eigen decomposition can be used to show identification. See, e.g., Hu (2020).

[33]Requiring that there are always 3 or more bids per listing simplifies the exposition, but it is not necessary. See the subsection titled "Proof of Proposition 4 with 2 or Less Bids" at the end of Section 9.4. in the online appendix.

Now consider the other primitives of the borrower's model. These 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 time dummies as $u_t(r) = -(r \times x_{amt}) + d_t^X$, where x_{amt} is the loan size and d_t^X are period specific constant terms that are allowed to depend on X.

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, $X = (X^o, X^u)$. In particular, the distribution of φ can depend on the amount requested. To the extent that there is some signaling value in the requested amount, the conditional distribution of φ will depend on the amount requested. We are allowing for this possibility.

Second, note that we can normalize $D(\cdot)$ as $D(\varphi) = -\varphi$ so that we interpret the type of the borrower as her default cost. It is also easy to see that we can normalize one of the constants in u_t without loss of generality: Hence we set $d_T^X = 0$.

The identification of the borrowers' model relies on the fact that conditional on X there is a one-to-one relationship between the borrower's type φ and s in a signaling equilibrium. This implies that conditioning on s is equivalent to holding fixed φ. We can then use variation in r to identify φ and the distribution of ε. The following proposition states the results in the form of a proposition.

Proposition 4. *Assume that for each X^o and s there exists a nonempty range (r', r'') such that $\Pr(X^u = H|X^o, r, s) = 1$ for $r \in (r', r'')$. Assume also that the irreducibility condition of Hall & Zhou (2003) is satisfied. Then, $u_t(r)$, $F_{\varphi|X}$, $\lambda(\cdot)$ are identified. Moreover, there exists an interval on which $F_{\varepsilon|X}$ is identified.*

The assumption in Proposition 4 requires that there is a range of interest rates such that lenders would be willing to bid in that range if and only if the type of the borrower is H. If type-H borrowers are more creditworthy, the assumption seems quite natural.[34] Note that $F_{\varepsilon|X}$ may not be identified on the entire support because the range of realized interest rate is limited to at most $[0, 0.36]$.

Lastly, we briefly discuss identification of the lender's model. The primitives of the lender's model that we need to identify are the distribution of the coefficient of risk, F_A,

---

[34]A sufficient (but by no means necessary) condition on the primitives is that the highest φ among type H borrowers is higher than those among type L borrowers, $d_t^H \geq d_t^L (\forall t)$ and F_A is bounded.

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 $\{1, \ldots, \overline{N}\}$. These primitives are identified from the bids of the lenders before the listing is funded as long as lenders bid as if their bid is not pivotal.

Proposition 5. *Assume lenders bid according to the strategy described in Proposition 3. Then, F_A, F_{ϵ_0}, F_N and c_q are identified.*

6 Estimation

We estimate our model in two steps. First, we use the EM algorithm to estimate the borrower's utility function, the latent type distribution, the distribution of r, $f(r|s, X)$ and the funding probability. We specify the distribution of r and the funding probability as flexible functions of s and X as they are both equilibrium objects. We then estimate the primitives of the lender's model in the second step. While our discussion of identification in the previous section focused on nonparametric identification, we place parametric restrictions on some of the model primitives in our estimation, as we will describe below.

6.1 Estimation of the Borrower Model

We denote by $\theta = \{\theta_b, \theta_\pi, \theta_r, \theta_p\}$ the parameters of the borrower's model. In particular, the borrower's period t utility function is denoted as $u_t(r, X|\theta_b)$, the outside option as $\lambda(\varphi, X; \theta_b)$, and the distribution of ε as $F(\cdot|X; \theta_b)$, where $X = (X^o, X^u)$. The vector of observable characteristics X^o includes the credit grade, requested amount, debt-to-income ratio, and home-ownership. The default cost $D(\varphi)$ is normalized as $D(\varphi) = -\varphi$ (see Section 5). We parameterize the probability that a borrower is type H as $\pi(s, X^o; \theta_\pi)$. We also specify the distribution of r, $f(r|s, X; \theta_r)$, using a (second-order) Hermite series where θ_r is a vector of parameters. We specify $\Pr(\text{fund}|s, X; \theta_p)$ as a Probit with flexible functional forms. We leave the distribution of φ to be fully nonparametric.

In order to estimate θ, we first compute for each borrower two value functions, $V_1(r, \varphi, X^o, L; \theta_b)$ and $V_1(r, \varphi, X^o, H; \theta_b)$. We then compute the first-order condition

that determines the borrower's choice of the reserve rate,

$$\frac{\partial \Pr(s,X;\theta_p)}{\partial s}\left(\int V_1(r,\varphi,X;\theta_b)f(r|s,X;\theta_r)dr - \lambda(\varphi;\theta_b)\right)$$

$$+ \Pr(s,X;\theta_p)\int V_1(r,\varphi,X;\theta_b)\frac{\partial f(r|s,X;\theta_r)}{\partial s}dr = 0, \qquad (6.1)$$

where $V_1(r,\varphi,X;\theta_b)$ is either $V_1(r,\varphi,X^o,L;\theta_b)$ or $V_1(r,\varphi,X^o,H;\theta_b)$. 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 (and conditional on X^u), the type of borrower φ who found it optimal to choose s. This allows us to back out the borrower's types $\varphi(s,X^o,L;\theta)$ and $\varphi(s,X^o,H;\theta)$.[35]

Once we have the borrower's types, it is possible to compute the (pseudo-)likelihood associated with each listing conditional on X^u.[36] For example, the likelihood associated with a particular repayment behavior can be computed as the outcome from a sequence of binary probit models with intercept φ (see, e.g., expression (4.1)), where φ solves the borrower' FOC. We denote the likelihood associated with listing i by $l_i(X^u;\theta)$, where we emphasize the dependence on X^u.

In order to maximize the likelihood, we use the EM algorithm. In particular, consider an initial parameter θ^0. Now consider the probability that listing i is type H given the outcome of the listing such as whether or not the listing was funded, the realization of r and the repayment decisions:

$$\Pr(H|s,X^o,\text{outcome of }i;\theta^0) = \frac{\pi(s,X^o,\theta_\pi^0)l_i(H;\theta^0)}{\pi(s,X^o,\theta_\pi^0)l_i(H;\theta^0) + (1-\pi(s,X^o;\theta_\pi^0))l_i(L;\theta^0)}. \quad (6.2)$$

We maximize the following expression with respect to θ

$$L(\theta;\theta^0) = \sum_i \Pr(H|\theta^0)l_i(H;\theta) + (1-\Pr(H|\theta^0))l_i(L;\theta),$$

where $\Pr(H|\theta^0)$ is defined in expression (6.2). We then take the maximizer, say θ^1, as

---

[35]There are a small fraction of borrowers (less than 3.5% of the sample) for whom $\varphi(s,X^o,X^u)$ cannot be solved for both X^u even when $s < 36\%$. This is because for a small fraction of borrowers, $\Pr(s,X)$ is not increasing at (s,x).

[36]Note that we do not parameterize the distribution of φ. If we were to specify the likelihood for φ, the contribution of the likelihood from φ will be multiplicatively separable (additively separable for the log-likelihood). Hence, the value of θ that maximizes the expectation of the pseudo-likelihood will also maximize the expectation of the full likelihood.

the new initial parameter and repeat the procedure until convergence. The online appendix discusses in greater detail how we implement the EM algorithm.

6.2   Estimation of the Lender Model

We now discuss the estimation of the lender's model. The primitives that we estimate are 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, $F_{\epsilon_0}(\cdot; \theta_l)$, and the lender's cost of bidding, c_q.

We estimate the parameters using simulated method of moments by matching the conditional funding probability, the realized interest rate (for funded listings) and the number of bids (for unfunded listings). First, let $f d_i$ be a dummy variable which equals 1 if listing i is funded, and 0 otherwise. Then $\frac{1}{I} \sum_{i=1}^{I} f d_i$ gives the (empirical) probability that a listing is funded, where I is the number of observations. Likewise, let $f d(s, X; \theta_l)$ denote a random dummy variable which equals 1 if a listing is funded and 0 otherwise, given listing characteristic X, reserve interest s, and parameter θ_l. As we explain below, $f d(s, X; \theta_l)$ can be expressed as follows:

$$
f d(s, X; \theta_l) = 1 \left\{ \sum_{j=1}^{N} q_j^* \geq x_{i,amt} \right\} \text{ and} \tag{6.3}
$$
$$
q_j^* = \underset{q_j \in M \cup \{0\}}{\arg \max} \left\{ 1\{q_j \neq 0\} \left(U_j^L(q_j Z(s, X)) - \epsilon_{0j} \right) \right\},
$$

where N is the (random) number of potential lenders, $U_j^L(q_j Z(s, X))$ is the utility of lending q_j dollars at interest rate s (defined in expression (4.4)), 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 moment and the model expectation:

$$
\frac{1}{I} \sum_{i=1}^{I} f d_i - \mathbb{E}[f d_i(\theta_l)],
$$

where the expectation is with respect to N, $(A_j)_{j=1}^{N}$, $(\epsilon_{0j})_{j=1}^{N}$, and X^u. Note that we use $\Pr(H|\widehat{\theta})$ to take expectations with respect to X^u where $\widehat{\theta}$ is the estimated parameter from the borrower's model. The mean and the variance of $Z(s, X)$ is computed using the estimates of the borrower model.

We now explain why $f d_i(\theta_l)$ can be expressed as (6.3). Suppose that there are N po-

tential lenders and their risk attitude and outside option are $(A_j)_{j=1}^N$ and $(\epsilon_{0j})_{j=1}^N$. When the loan is not yet fully funded, the optimal choice of the lender is given by the second equation in expression (6.3), 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 (6.3). $x_{i,amt}$ is the loan amount requested by borrower i, 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 realized interest rate for funded listings. The model prediction of the interest rate can be computed by considering an expression analogous to expression (6.3) for different values of r, $fd(r, X, \theta_l)$, and taking the lowest r for which $fd(r, X, \theta_l) = 1$.[37] The second set of moments is the number of lenders, in particular, the number of lenders who bid an amount $q \in M$ to unfunded listings. 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 do not use the full implications of the strategy we defined in Section 4.2. For example, we do not use the interest rate of each bid or the number of lenders who bid $q \in M$ for funded listings. This is because these objects are quite sensitive to the particular dominant strategy described in 4.2. If lenders are playing other strategies (say, because revising their bid is costly), the interest rate of each bid, the number of bids for funded listings, etc., could be quite different depending on how we specify the timing of lender arrival.[38] Finally, we have suppressed the conditioning variables in our exposition, but we construct moment conditions for each credit grade.

---

[37]We divide the space of r into 7 bins ([5%, 10%], [10%, 15%], [15%, 20%], [20%, 25%], [25%, 30%], [30%, 34%], [34%, 36%]) and match the proportion of listings that fall in each bin.

[38]For example, suppose that a particular bidder is interested in funding a given listing at or above 20%. If the lender views the listing before the active interest rate decreases to below 20%, the lender bids on the listing. However, if the lender views the listing after the active interest rate decreases to below 20%, then the lender will not bid. In this sense, the number of bids for funded listings will depend on assumptions regarding when lenders view the listing.

7 Results

7.1 Specification and Model Estimates

The exact specification that we use to estimate the model of the borrower is as follows. First, we set the period utility function as $u_t(r|\theta_b) = -r \times x_{amt} + d_t$, where $\{d_t\}$ are time dummies.[39] We allow the dummies to differ by X^u and credit grade. $D(\varphi)$ is normalized to $-\varphi$ and the outside option $\lambda(\varphi)$ is specified as a linear function with grade and latent-type specific slope as $\lambda\varphi$. The type of each borrower is recovered nonparametrically for each borrower. Lastly, F_ε is specified to be a Type I extreme value distribution with a 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 the number of potential lenders F_N, the distribution of lender's risk attitude, F_A, the distribution of the lender shock, F_{ϵ_0}, and the costs of bidding associated with each amount choice, $\{c_q\}_{q \in M}$. We specify N to be distributed log normal with parameters μ_N and σ_N^2, and we specify the risk attitude, A_j, and the outside option, ϵ_{0j}, to be distributed Normal as $N(\mu_A, \sigma_A^2)$ and $N(\mu_{\epsilon_0}, \sigma_{\epsilon_0}^2)$. The mean of ϵ_{0j} is allowed to depend on the calendar month to capture changes in the lender's outside option or lenders' beliefs due to factors such as macro shocks. We normalize $c_{50} = 0$. The parameters of the lender's model are allowed to differ by credit grade.

We report the estimation results in Table 7 and Table 8. Table 7 reports the quantiles of the borrowers' privately-known type distribution (φ) by credit grade, latent type, and requested amount.[40] The table also reports the fraction of high and low types by credit grade and requested amount in the shaded rows. Recall that we interpret φ as the borrower's default cost, with good types having high values of φ (and hence less likely to default) and bad types having low values of φ (and hence more likely to default). We find that the median default cost (across all requested loan amounts; see second to the last row of Table 7) are estimated to be around \$22,000 for the high types of credit grade AA and \$16,000 for the low types of credit grade AA. For credit grade A, the median

---

[39]Instead of including a separate dummy variable for each repayment period, we include dummies that correspond to each 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, and so on. We normalize the first dummy, which leaves us with 11 dummies per credit grade and latent type.

[40]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.

default cost for the high types is around \$15,000 and it is around \$7,350 for the low types. We find that the default cost is highest for credit grade AA and lowest for credit grade C, which is intuitive. The fraction of $X^u = H$ types is 50.9% (grade AA), 90.1% (grade A), 25.5% (grade B) and 88.5% (grade C), on average.

Table 7: Borrower's Type Distribution

Amount Quartile	Type Quantile	AA		A		B		C	
		High	Low	High	Low	High	Low	High	Low
1st	25%	26.671	18.603	18.051	8.807	4.471	4.179	3.268	2.240
	50%	27.939	19.074	18.527	9.024	4.615	4.308	3.339	2.288
	75%	28.725	19.763	19.078	9.244	4.760	4.442	3.445	2.357
	-	50.3%	49.7%	90.3%	9.7%	16.9%	83.1%	82.4%	17.6%
2nd	25%	23.003	16.628	15.679	7.866	3.851	3.617	2.988	2.043
	50%	23.934	17.301	16.381	8.245	4.038	3.790	3.054	2.106
	75%	24.938	17.840	17.228	8.507	4.270	4.007	3.112	2.138
	-	48.4%	51.6%	92.4%	7.6%	25.1%	74.9%	90.3%	9.7%
3rd	25%	17.503	14.572	12.986	6.314	3.221	3.086	2.777	1.860
	50%	19.388	15.520	14.150	6.898	3.501	3.293	2.834	1.944
	75%	19.782	15.875	15.024	7.352	3.584	3.398	2.922	1.958
	-	46.1%	53.9%	92.9%	7.1%	29.2%	70.8%	89.9%	10.1%
4th	25%	13.799	10.194	9.516	4.600	2.426	2.239	2.181	1.458
	50%	13.867	11.173	9.807	4.749	2.763	2.617	2.611	1.522
	75%	14.369	13.108	11.106	5.348	2.932	2.758	2.740	1.706
	-	57.0%	43.0%	84.5%	15.5%	29.1%	70.9%	90.8%	9.2%
All	25%	16.332	13.227	11.827	5.556	2.933	2.752	2.682	1.793
	50%	21.986	15.968	15.021	7.352	3.603	3.398	2.892	1.977
	75%	25.940	18.184	17.104	8.483	4.297	4.025	3.179	2.143
	-	50.9%	49.1%	90.1%	9.9%	25.5%	74.5%	88.5%	11.5%

Note: The table reports the estimated quartiles of the default cost of the borrower, φ, by credit grade, latent type, and requested amount. The units are in \$1,000. The columns labeled High correspond to the high types ($X^u = H$) and those labeled Low correspond to low types ($X^u = L$). The table also reports the fraction of latent types in the shaded rows.

Another general pattern that can be observed from Table 7 is that borrowers who request a large amount generally have lower values of φ. For example, the high type borrower in credit grade AA with a requested loan amount in the highest quartile has an estimated median default cost of \$13,867 – which is much lower than the corresponding value among borrowers with requested loan amount in the lowest quartile (\$27,939). Note that conditioning on the requested amount tightens the distribution of φ significantly in many cases. For example, the interquartile range among all high type borrowers in credit grade AA is about \$9,600 (\$25,940-\$16,332) whereas the conditional interquartile range among those who are in the first quartile of the requested amount is about \$2,000 (\$28,725-\$26,671). There are a few market segments such as the 4th

quartile for credit grade C where the interquartile range does not shrink by conditioning on the requested amount. The interquartile range among all high type borrowers in credit grade C is about $500 (=$3,179-$2,682) while the interquartile range for those in the fourth quartile of the requested amount is $559 (=$2,740-$2,181). The general pattern, however, seems to be that the requested amount is informative of the borrower's type.

Table 8: Parameter Estimates of the Borrower's and Lender's Model

Borrower Parameter	Latent Type High	Low	Parameter	Lender Estimates by Grade AA	A	B	C
σ	6.66		μ_N	6.10	5.15	4.45	5.14
	(1.28)			(1.32)	(0.98)	(0.91)	(1.00)
λ_{AA}	1.41	2.10	σ_N	0.51	0.90	0.67	0.63
	(0.28)	(0.40)		(0.10)	(0.17)	(0.19)	(0.13)
λ_A	2.02	3.99	c_{75}	-0.30	0.78	0.87	1.28
	(0.42)	(0.68)		(0.06)	(0.15)	(0.25)	(0.26)
λ_B	6.07	6.91	c_{100}	0.36	1.56	3.28	4.13
	(1.11)	(1.28)		(0.07)	(0.30)	(0.87)	(0.81)
λ_C	8.00	8.97	c_{200}	0.44	1.16	0.90	4.71
	(1.50)	(1.53)		(0.09)	(0.24)	(0.27)	(0.94)
			c_{500}	1.11	-3.91	-8.47	-8.09
				(0.22)	(0.74)	(2.02)	(1.57)
			c_{1000}	-22.69	-239.90	-160.20	-441.10
				(4.43)	(46.58)	(57.82)	(85.32)
			μ_A	3.99E-04	3.51E-04	-1.55E-04	1.03E-03
				(8.00E-05)	(7.20E-05)	(5.19E-05)	(2.10E-04)
			σ_A	2.94E-04	7.74E-04	-5.93E-04	1.26E-03
				(6.00E-05)	(1.44E-04)	(1.46E-04)	(2.46E-04)
			$\mu_{\epsilon_{0j}}$	13.93	13.32	18.15	77.23
				(2.70	(2.73)	(5.75)	(15.10)
			$\sigma_{\epsilon_{0j}}$	10.47	15.46	23.80	52.88
				(2.13)	(3.00)	(6.55)	(11.01)
Obs.	3,818			1,420	1,850	3,068	5,203

Note: We report the parameter estimates of the borrower's model in the first two columns and those of the lender's model in the last four columns. 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 (100 times) and they are reported in parentheses.

Table 8 reports the estimates of the other parameters of the model. The first two columns of Table 8 report the parameter estimates of the borrowers' model. The last four columns report the parameter estimates of the lenders' model. Recall that λ is

a parameter that captures the relationship between the default cost of the borrower (φ) and the utility of the outside option ($\lambda \varphi$). Our estimates of λ translate to mean (averaged across latent types) outside option of \$32,390 for credit grade AA, \$30,296 for grade A, \$23,083 for grade B, and \$22,523 for grade C.[41]

The last four columns of Table 8 report the parameter estimates for the lenders. We estimate 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 -1.55e-04 to 1.03e-03. Our estimates suggest that the lenders are close to being risk neutral. Our estimates of risk aversion imply that the average lender is indifferent between \$0 and a lottery which gives \$10 with probability 50% and $-\$x$ with probability 50%, where x is \$9.96 for credit grades AA and A, \$10.01 for B, and \$9.89 for C.

Table 8 also reports our estimates of the mean of ϵ_{0j} (μ_{ϵ_0}) and the cost of funding different amounts (c_q). Because we normalize the coefficient on $q\mu(Z(r))$ in the lender's utility to 1, the estimates are measured in units of dollars. We find that the mean of the outside option is lower for better credit grades (\$13.9 and \$13.3 for credit grades AA and A) while it is higher for low credit grades (\$18.2 and \$77.2 for credit grades B and C).[42]

7.2 Model Fit

We simulate the outcome by solving the borrower's repayment problem and the lenders' model for each listing to evaluate the model fit. In particular, we solve the borrower's problem 10,000 times to compute the mean return and the average default probability. We take 100 sets of potential lenders to simulate the funding probability and the contract interest rate.

Table 9 reports the fit of the borrower's model. The table compares the borrower's default probability and the IRR by credit grade. The table shows that the prediction errors are reasonably small overall, except for credit grade B where the model overpre-

---

[41]One reason why the value of the outside option is relatively similar across credit grades is because the relationship between the funding probability and the reserve rate is steeper for higher credit grades and flatter for lower credit grades (See Section 3.3 of the online appendix). This translates to smaller values of λ for high credit grades and larger values of λ for low credit grades since λ is primarily identified by expression (4.3).

[42]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 very similar. Section 6.2 of the online appendix reports the estimates.

dicts default and underpredicts the IRR relative to the data.[43]

Table 9: Model Fit–Borrower Default Probability and IRR

| | Default | | IRR | |
Grade	Model	Data	Model	Data
AA	0.167	0.149	-0.014	-0.010
A	0.216	0.211	-0.015	-0.025
B	0.352	0.297	-0.097	-0.074
C	0.273	0.309	-0.048	-0.060

Figure 7: Model Fit–Funding Probability & Distribution of Interest Rate
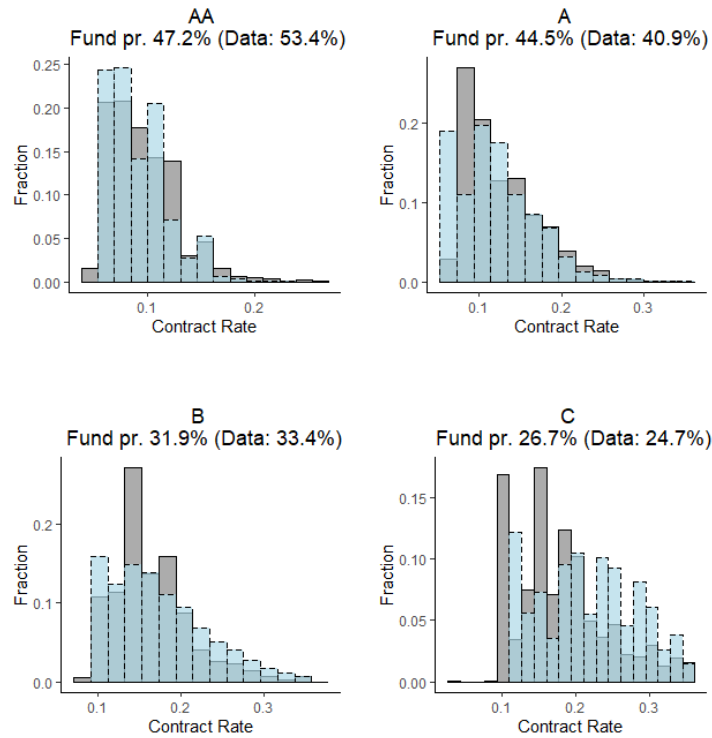
Note: The numbers below each credit grade are the funding probabilities predicted by the model, and those in the data. The bars with dashed lines correspond to the model prediction of the contract interest rate and the bars with solid lines correspond to the actual distribution.

Figure 7 plots the actual (solid bars) and the simulated (dashed bars) distributions of the realized interest rate by credit grade. The model does relatively well-fitting the

---

[43]One reason why the fit is not ideal is because the borrowers' private default costs are inverted out using the first-order condition with respect to s. Because borrower default costs are determined by a choice that is not directly informative of default, the model has difficulty fitting the repayment decisions of some borrowers.

overall shape of the actual distribution, although the model does not replicate the mass at round numbers observed in the data (e.g., 10%, 15%). The figure also includes the actual and the predicted funding probabilities. The model underpredicts the funding probability for credit grade AA somewhat, but does better for other credit grades.

8   Counterfactual Experiment

In our counterfactual experiment, we compare market outcomes and welfare under three possible scenarios – a market with signaling, a market without signaling (i.e., pooling) and a market with no information asymmetry between borrowers and lenders. Given that we cannot implement the benchmark first-best outcome of no asymmetric information, we do not interpret our counterfactual results as predictions of what would happen in response to changes in the design of the platform. Instead, the counterfactuals illustrate the extent to which an online market suffers from adverse selection and the extent to which signaling restores or reduces efficiency holding fixed the set of market participants.[44] We think that studying how signaling restores efficiency in markets with adverse selection is a salient issue for many online marketplaces in which market participants typically know very little about each other. Moreover, our counterfactuals offer an illustration of how adverse selection can lead to a backward bending credit supply curve, a very influential idea that goes back to Stiglitz & Weiss (1981) but one without much empirical evidence based on microdata.

In Figure 8, we present the credit supply curves for each credit grade and latent type. 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 panels reflecting the fact that the amount of credit supply varies considerably depending on credit grade and latent type. The curves in each panel correspond to the credit supply curve under signaling, symmetric information, and pooling. Below, we explain each in turn.

---

[44]We treat F_N and x_{amt} as exogenous. In order to see that our conclusions are *qualitatively* robust to small changes in these variables, we report sensitivity analysis in Section 8.1 of the online appendix. In particular, we decrease μ_N by 5% and decrease the requested amount by 5%, recomputing welfare in each of these cases. We find that the welfare ranking among signaling, pooling and symmetric information remains mostly the same. Moreover, we find that in each of the sensitivity analyses, signaling can restore much of the surplus difference between signaling and pooling in credit grades AA and A as we report below.

- **Credit supply curve under signaling**: The solid curves in Figure 8 correspond to the credit supply curves under signaling. These supply curves correspond to the average amount of credit that potential lenders are willing to supply under the actual mechanism used by Prosper. The supply curves are drawn for the borrower with the median φ in each credit grade and latent type.[45] The credit supply curves trace the average amount of credit that potential lenders are willing to supply to the median borrower at different interest rates. Note that the supply curves are truncated at the reserve interest rates. The borrower does not have access to credit above the reserve rate in the signaling equilibrium.

- **Credit supply curve under symmetric information:** The dotted curves that lie on top of the signaling credit supply curves correspond to the credit supply curves under symmetric information. The supply curves are again drawn for the median borrower type within each latent type and credit grade. The credit supply under symmetric information is computed assuming that the lenders can directly observe φ (as opposed to learning φ through the reserve rate). Note that under both signaling and symmetric information, the lenders have perfect knowledge of the borrower types in equilibrium: The lenders know that they are lending to a borrower of a particular type. Thus, the credit supply curves under these two scenarios partly coincide.

  The difference between the two is that the borrowers do not need to signal their type by the reserve price under symmetric information. Hence, they can borrow at rates that are higher than the reserve rate that they would choose under signaling. This means that the credit supply curve for symmetric information extends beyond the reserve rate all the way until the point at which the borrower is indifferent between borrowing and not borrowing. The truncated supply curve under the signaling equilibrium can be viewed as capturing the cost that borrowers must pay (i.e., the surplus that has to be burned) in order to differentiate herself from lower types in the signaling equilibrium.

- **Credit supply curve under no signaling (i.e., pooling):** The thick dashed curve in each of the panels correspond to the credit supply curve under asymmetric

---

[45]We set the other characteristics of the borrower to their median value within each credit grade. In particular, the requested amount, debt-to-income ratio and home-ownership are set to ($10,000, 21\%$, 1), ($10,000, 27\%$, 1), ($10,000, 28\%$, 1), and ($7,000, 28\%$, 1) for credit grades AA through C.

information with no signaling (i.e., pooling). This curve is computed under the assumption that each borrower posts a *secret* reserve rate. 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 rate. Note that under this scenario, it is a dominant strategy for each borrower to submit a secret reserve rate equal to the interest rate at which the borrower is indifferent between borrowing and not borrowing.[46] Secret reserve rates 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 curve traces out the average amount of credit that potential lenders are willing to supply to a pool of borrower types at a given interest rate. The pool of borrowers corresponds to the set of borrowers whose secret reserve price is higher than the given interest rate.

Figure 8 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 of the median borrower. This fact tends to make the difference between the credit supply under pooling (dashed curve) and the credit supply curve for the median type relatively small at lower 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 lower than that from funding a loan to a the median borrower. This fact tends to make the differences between the two supply curves larger 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 & Weiss (1981). The supply curve is backward bending in this case. Figure 9 shows a credit supply curve from credit grade B with latent type set to low, requested amount set to $25,000$, debt-to-income ratio to 24%, and home-ownership to 0. The credit supply curve under pooling is backward bending for this case. As the interest rate goes up, the set of borrowers who are willing to take out the loan at those interest rates becomes progressively worse. The worsening of the borrowers puts a downward pressure on the rate of return to the point that

---

[46]Section 7.1 of the online appendix discusses details regarding the computation of the credit supply curve.
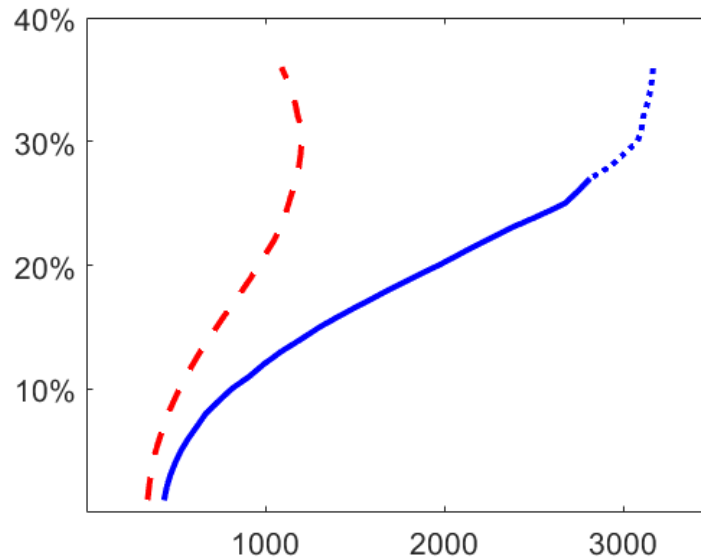
Figure 8: Credit Supply Curve
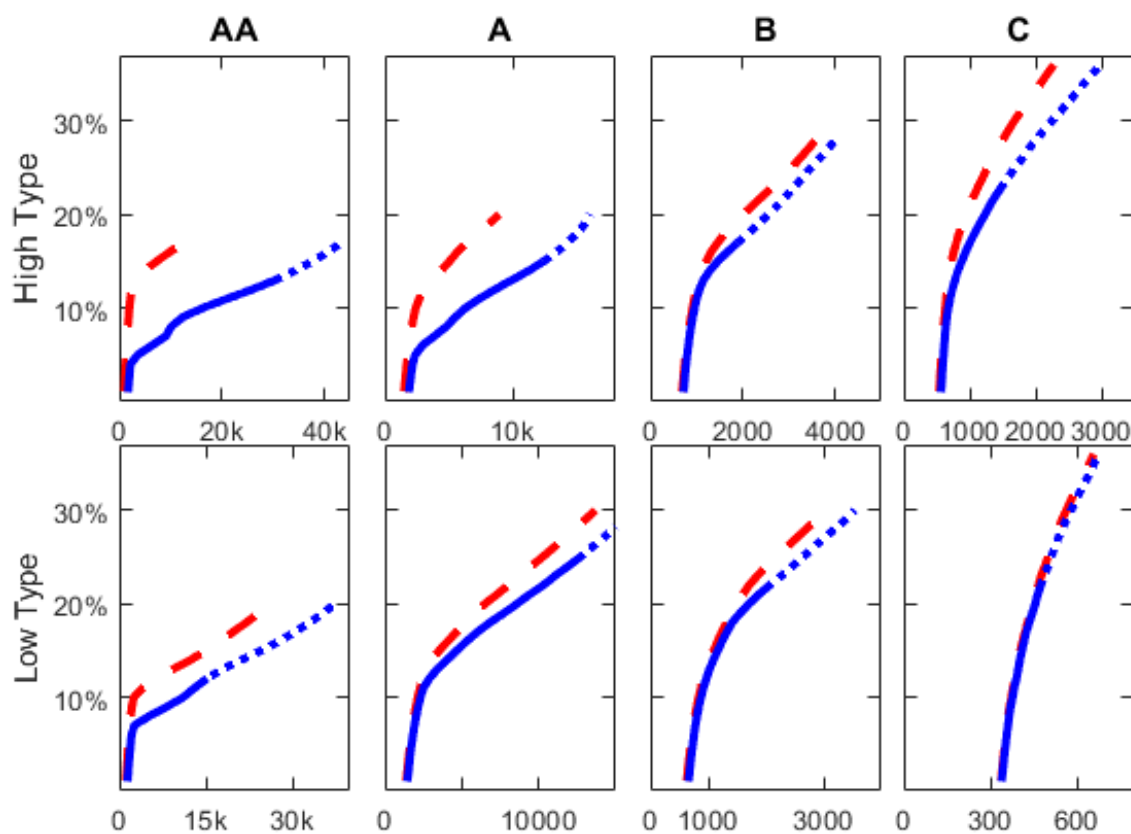
Note: The thick dashed curve corresponds to the credit supply curve under no signaling (i.e., pooling). The solid curve corresponds to the credit supply curve under signaling for the median type, and the dotted curve that lies on top of it corresponds to the credit supply curve under symmetric information.

the downward pressure outweighs the increase in the return from higher interest rates.

Finally, we examine the welfare implications of signaling and information asymmetry. In Table 10, we report the average surplus of the lenders and the borrowers per listing for each of the three different scenarios we consider. The surplus of a borrower with type φ is the product of the probability of being funded Pr(funded) and the surplus conditional on borrowing $\mathbb{E}_r[V_1(r, \varphi) - \lambda(\varphi)]$, where the expectation is over the contract interest rate. The surplus of lender j is $\mathbb{E}_r[U_j^L(r)]$, 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 for every listing and latent type X^u. We then take the expectation with respect to the distribution of X^u. Details of the computation are

50

Figure 9: Example of a Backward Bending Credit Supply Curve

Note: The figure shows credit supply curves from credit grade B with requested amount set to $25,000, debt-to-income ratio to 24%, home-ownership to 0, and the latent type to low. The thick dashed curve corresponds to the credit supply curve under pooling and the solid (dotted) curve corresponds to the credit supply curve for the median type under signaling (symmetric information).

discussed in Section 7.2 of the online appendix.

First, consider the welfare of the borrowers in the first panel of Table 10. Comparing the borrower welfare under pooling and under no asymmetric information, we find that the welfare loss from information asymmetry is relatively modest in low credit grades while it is relatively large in high credit grades. For example, the difference in borrower welfare is about $5 for credit grade C while it is about $95 for credit grade AA. Comparing these numbers to borrower welfare under signaling, we find that welfare improves relative to pooling in all credit grades except for credit grade C. For credit grade C, the borrower welfare under pooling is higher than that under signaling. When the borrower's type distribution is relatively concentrated, the costs of signaling can outweigh the benefits. 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 panel of Table 10. We find that lender welfare is highest under symmetric information, second-highest under signaling and lowest under pooling across all credit grades. We find that the welfare gains from signaling are considerable for credit grades AA and A.

Table 10: Welfare

		AA	A	B	C
Borrower	Pooling	185.5	464	141.8	58
		(42.6)	(19.9)	(13.8)	(9.5)
	Separating	253.6	487.5	150.7	39.1
		(29.0)	(21.0)	(14.0)	(10.0)
	Symmetric	281.8	499.2	154.7	63.4
		(29.1)	(20.6)	(14.0)	(9.9)
Lender	Pooling	102.15	265	217.8	262.9
		(46.6)	(14.4)	(26.9)	(18.9)
	Separating	141.85	289.9	218.35	271.4
		(49.8)	(14.3)	(18.9)	(6.6)
	Symmetric	144.75	295.6	231.7	277.1
		(49.0)	(15.8)	(28.0)	(19.4)
Total	Pooling	287.65	729	359.6	320.9
		(29.7)	(29.4)	(58.2)	(30.6)
	Separating	395.5	777.4	369.1	310.5
		(32.5)	(31.0)	(34.3)	(13.4)
	Symmetric	426.6	794.8	386.4	340.5
		(26.2)	(31.1)	(58.6)	(32.1)
Percent Recovered		77.6	73.6	35.3	-53.2

Note: The table reports the expected surplus for different market designs by credit grade. The standard errors are calculated by bootstrap and reported in parenthesis.

The third panel 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 32.6% ($139) decrease in total surplus for credit grade AA. We also find that in some markets, signaling can restore a large fraction of the potential welfare loss from adverse selection. For credit grades AA and A, about 77.6% and 73.6% of the welfare loss are recovered through signaling.

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 evidence suggesting that the reserve interest rate posted by potential borrowers serve as a signaling device. 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.

While the findings of the paper directly apply to the market of Prosper.com, our analysis may be salient to many online P2P markets that have seen rapid growth in recent years.

References

Adams, W., Einav, L. & Levin, J. (2009), 'Liquidity constraints and imperfect information in subprime lending', *American Economic Review* **99**(1), 49–84.

Akerlof, G. A. (1970), 'The market for lemons: Quality uncertainty and the market mechanism', *Quarterly Journal of Economics* **84**(3), 488–500.

Arnold, L. & Riley, J. (2009), 'On the possibility of credit rationing in the stiglitz-weiss model', *American Economic Review* **99**(5), 2012–2021.

Ausubel, L. (1999), 'Adverse selection in the credit card market', *mimeo.* .

Backus, Matt, B.-T. & Tadelis, S. (2019), 'On the empirical content of cheap-talk signaling: an application to bargaining', *Journal of Political Economy* **127**(4), 1599–1628.

Berger, A. & Udell, G. (1992), 'Some evidence on the empirical significance of credit rationing', *Journal of Political Economy* **100**(5), 1047–1077.

Bester, H. (1985), 'Screening vs. rationing in credit markets with imperfect information', *American Economic Review* **75**(4), 850–855.

Cai, H., Riley, J. & Ye, L. (2007), 'Reserve price signaling', *Journal of Economic Theory* **135**, 253–268.

Einav, L., Jenkins, M. & Levin, J. (2012), 'Contract pricing in consumer credit markets', *Econometrica* **80**(4), 1387–1432.

Fang, H. (2006), 'Disentangling the college wage premium: Estimating a model with endogenous education choices', *International Economic Review* **47**(4), 1151–1185.

Freedman, S. & Jin, G. (2010), 'Learning by doing with asymmetric information: Evidence from prosper.com', *mimeo.* .

Gayle, G.-L. & Golan, L. (2012), 'Estimating a dynamic adverse-selection model: Labour-force experience and the changing gender earnings gap 1968-97', *Review of Economic Studies* **79**(1), 227–267.

Gedge, C., Roberts, J. & Sweeting, A. (2020), 'A model of dynamic limit pricing with an application to the airline industry', *Journal of Political Economy* **128**(3), 1148–1193.

Hall, P. & Zhou, X.-H. (2003), 'Nonparametric estimation of component distributions in a multivariate mixture', *Annals of Statistics* **31**(1), 201–224.

Hörner, J. & Sahuguet, N. (2011), 'War of attrition with endogenous effort levels', *Economic Theory* **47**(1), 1–27.

Hu, Y. (2020), 'The econometrics of unobservables', *mimeo* .

Hungerford, T. & Solon, G. (1987), 'Sheepskin effects in the returns to education', *Review of Economics and Statistics* **69**(1), 175–187.

Iyer, R., Khwaja, A., Luttmer, E. & Shue, K. (2016), 'Screening peers softly: Inferring the quality of small borrowers', *Management Science* **62**(6), 1554–1577.

Jullien, B. & Mariotti, T. (2006), 'Auction and the informed seller problem', *Games and Economic Behavior* **56**(2), 225–258.

Karlan, D. & Zinman, J. (2009), 'Observing unobservables: Identifying information asymmetries with a consumer credit field experiment', *Econometrica* **77**(6), 1993–2008.

Kruskal, J. (1977), 'Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics', *Linear Algebra and Applications* **18**(2), 95–138.

Liu, X., Wei, Z. & Xiao, M. (2020), 'Platform mispricing and lender learning in peer-to-peer lending', *Review of Industrial Organization* **56**(2), 281–314.

Mailath, G. (1987), 'Incentive compatibility in signaling games with a continuum of types', *Econometrica* **55**(6), 1349–1365.

Mailath, G. & von Thadden, E.-L. (2013), 'Incentive compatibility and differentiability: New results and classic applications', *Journal of Economic Theory* **148**(5), 1841–1861.

Milde, H. & Riley, J. (1988), 'Signaling in credit markets', *Quarterly Journal of Economics* **103**(1), 101–129.

Milgrom, P. & Roberts, J. (1982), 'Limit pricing and entry under incomplete information: An equilibrium analysis', *Econometrica* **50**(2), 443–459.

Milgrom, P. & Roberts, J. (1986), 'Price and advertising signals of product quality', *Journal of Political Economy* **94**(4), 796–821.

Morris, S. & Shin, H. S. (2012), 'Contagious adverse selection', *American Economic Journal: Macroeconomics* **4**(1), 1–21.

Ravina, E. (2019), 'Love & loans: The effect of beauty and personal characteristics in credit markets', *mimeo.* .

Rigbi, O. (2013), 'The effects of usury laws: Evidence from the online loan market', *Review of Economics and Statistics* **95**(4), 1238–1248.

Roberts, J. (2013), 'Unobserved heterogeneity and reserve prices in auctions', *RAND Journal of Economics* **44**(4), 712–732.

Sahni, N. & Nair, H. (2020), 'Does advertising serve as a signal? evidence from a field experiment in mobile search', *Review of Economic Studies* **87**(3), 1529–1564.

Spence, M. (1973), 'Job market signaling', *Quarterly Journal of Economics* **87**(3), 355–374.

Stiglitz, J. E. & Weiss, A. (1981), 'Credit rationing in markets with imperfect information', *American Economic Review* **71**(3), 393–410.

Xin, Y. (2020), 'Asymmetric information, reputation, and welfare in online credit markets', *mimeo.* .