Not for Publication: Supplementary Material of "Signaling in Online Credit Markets"

Kei Kawai[*]
University of California Berkeley

Ken Onishi[†]
Federal Reserve Board

Kosuke Uetake[‡]
Yale School of Management

August 23, 2021

In this online appendix, we provide additional discussions, robustness checks, and proofs that are omitted from the main text. In Section 1, we demonstrate the workings of a simpler version of the model presented in the main text with two types of borrowers. The simplified model allows us to illustrate graphically the forces that make it possible to sustain a signaling equilibrium. In Section 2, we provide additional discussion of the institutional details and how our sample is constructed. Section 3 reports various robustness checks of our reduced-form analyses that we report in Section 3 of the main text. In particular, we explain the survey we ran to capture aspects of the listing that are hard to include in regression form, such as pictures and text. We present regression results that control for these measures. We also explain how we designed a field experiment to gain additional information about the signaling value of the reserve rate. In Section 4, we provide additional discussion of the modeling choices we make regarding the structural model. In particular, we discuss the choice of treating the loan amount as an exogenous variable. We show that even when the amount is endogenous, taking the loan amount as given does not lead to inconsistent estimates of model primitives. Section 5 discusses the details of the estimation. Section 6 reports additional parameter estimates of the structural model that we do not report in the main text. In Section 7, we report the computational details of our counterfactual simulations. Section 8 discusses sensitivity of the counterfactuals and estimates from an alternative model of lender beliefs. Finally, Section 9 collects all omitted proofs.

[*]University of California Berkeley, kei@berkeley.edu
[†]Federal Reserve Board: ken.t.onishi@frb.gov
[‡]Yale School of Management: kosuke.uetake@yale.edu

1 Simple Model of Signaling with Two Types of Borrowers

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

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

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

In this simple example, a signaling equilibrium consists of two different reserve interests s_H and s_L such that φ_H prefers s_H to s_L and vice versa.[1] Table 1 and Figure 1 illustrate one signaling equilibrium for a particular parameterization of the primitives; $u(r) = -r$, $\lambda(\varphi) = 0.105\varphi$, A_j distributed uniform $[0,1]$, ε distributed logit, and $\varphi \in \{1.7,\, 1.8\}$. In Table 1, we summarize the outcomes and utility levels associated with $s_H = 27.5\%$ and

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

Table 1: Outcomes and Utility Associated with Choosing s_H (27.5%) and s_L (36.0%).

Reserve Rate	Funding Pr.	Interest Rate	Utility for φ_H	Utility for φ_L
$s_H = 27.5\%$	35.5%	27.2%	0.300	0.300
$s_L = 36\%$	44.5%	35.2%	0.299	0.302

Note: The table shows the funding probability, average interest rate and borrower utility from choosing $s = 27.5\%$ and $s = 36\%$ when lenders believe that $s = 27.5\%$ signals H type and $s = 36\%$ signals L type.

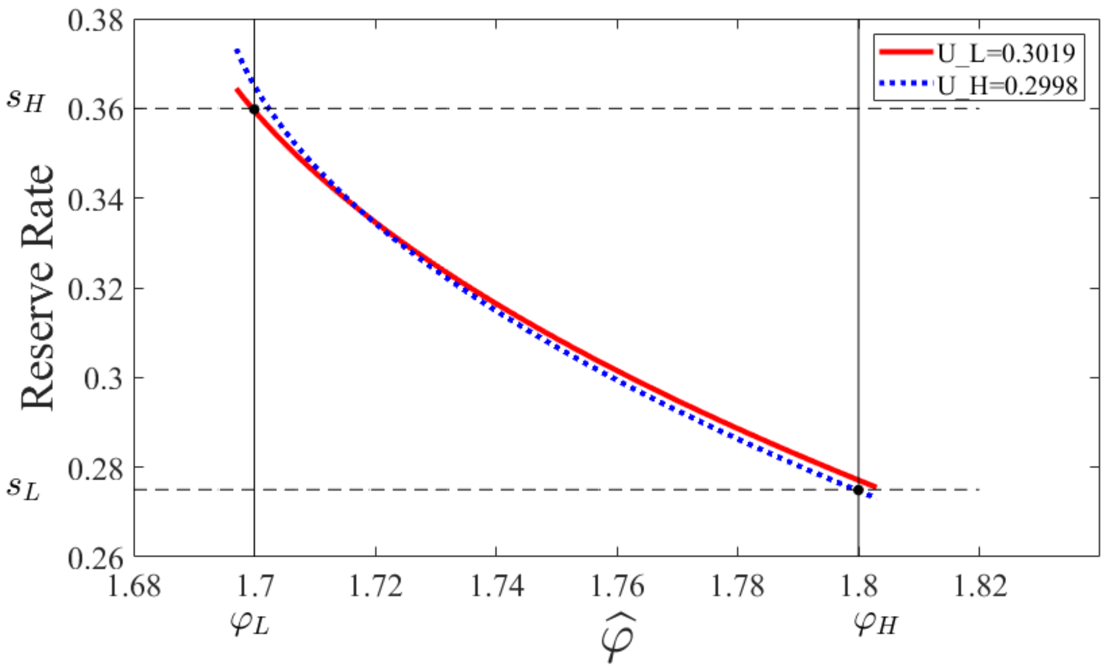
$s_L = 36\%$ for both types of borrowers when lenders believe that s_H signals high type and s_L signals low type. If a borrower posts s_H, the funding probability is relatively low (35.5%) but the average interest rate is relatively favorable (27.2%). On the other hand, if a borrower posts s_L, the funding probability is high (44.5%), but the average interest rate is relatively unfavorable (35.2%). For φ_H types, the favorable interest rate at s_H more than compensates for the low funding probability. Hence, the expected utility of choosing s_H (0.300) is higher than the expected utility of choosing s_L (0.299) for φ_H types. On the other hand, the expected utility of choosing s_L (0.302) is higher than the expected utility of choosing s_H (0.300) for φ_L types. Hence, reserve interests $s_H = 27.5\%$ and $s_L = 36\%$ can be supported as a separating equilibrium.

In Figure 1, we illustrate this equilibrium graphically. The horizontal axis of the figure corresponds to the beliefs of the lender ($\hat{\varphi}$) and the vertical axis corresponds to the reserve rate (s). The dotted curve in the figure corresponds to the indifference curve of φ_H at $(\hat{\varphi}_H, s_H)$ and the solid curve corresponds to the indifference curve of φ_L at $(\hat{\varphi}_L, s_L)$. The borrowers' utility level is higher to the northeast of the indifference curve.[2] The figure shows that a separating equilibrium can be supported in which type-φ_H chooses point s_H, while type-φ_L chooses s_L.

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

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

Figure 1: Illustration of a Separating Equilibrium



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

2 Institutional Details and Data Description

In this section, we provide additional description of the institutional details and the sample construction procedure. Note that Prosper has made major changes to its platform since the end of our sample period. For example, Prosper stopped using the auction format since December 31, 2010. The definitions of the credit grades have also changed since July of 2009. Our description of Prosper applies to the platform that was in use during our sample period (i.e., from May to October of 2008).

2.1 Institutional Details

Listing There are two types of listings on Prosper, open listings and closed listings. Closed listings become inactive after 7 days or after the listing attracts enough lenders to fully fund the loan, whichever occurs first. Open listings remain active for 7 days even after the loan is fully funded (additional bids may bring down the interest rate). Since less than one-fourth of the listings are closed listings, we work only with open listings in our sample.

A potential borrower may withdraw the listing at any time until the end of the bid submission period. About 27% of all listings are subsequently withdrawn. Most of the withdrawals occur immediately after the creation of the listing. Conditional on withdrawal, about 80% are withdrawn within one day. Most withdrawals are likely due to some mistake the borrower finds in the listing or due to experimentation by the borrower when learning how to use the system. We do not think that withdrawals occur as a response to borrowers seeing an unexpectedly high interest rate just before origination. We do not include withdrawn listings in our sample.

If a listing is not funded, a borrower may relist on Prosper. Only about 30.4% of the borrowers (excluding borrowers that withdraw) post a second listing when they are unsuccessful in obtaining a loan, however. The median change in the reserve rate between the original and the second listing is 0%. The average change in the reserve rate is about 1.2%. This suggests that dynamic considerations are not too large. Moreover, the vast majority (97.0%) of the sample consists of first-time borrowers.

Lenders Bid At the time the lender submits her bid, she observes the fraction of the loan funded, the bid amount of each of the submitted bids and the number of submitted bids in addition to the characteristics of the listing. For listings that have not been fully funded, these are all she observes. For listings that have attracted enough bids to fund the loan, she also observes the active interest rate as well as the losing bids in addition to what she observes for unfunded listings.

Repayment In about 4% of repayments, the borrower pays more than 50% in excess of the regular monthly loan repayments. There is no penalty for early payment: early repayments go directly into paying off the principal. We abstract from modeling early repayments because they unnecessarily complicate our model. The relevant information in the data that we use for estimating the model of the borrowers is the timing of default. We assume that the borrower has made regular repayments until default.

Being behind on loan payments by one month does not automatically imply that the borrower defaults on the loan. If a borrower's monthly payment is more than 15 days late, a late fee is charged. Usually, there is a three–month lag between the first missed payment until the loan is charged off by Prosper as default. For our estimation, we define the month of default as the first month of consecutive missed payments that subsequently result in default. While it is possible to consider an alternative model of the borrower that incorporates the number of months the borrower is behind schedule, this comes at a significant increase in computation (we need to track the number of months late as a state variable). Given that once the borrower misses a payment, the probability of ending up in default is considerable (more than 85%), we think that the results will not be affected very much by our simplifying assumption. Our definition of default is the same as that in Freedman & Jin (2010).

2.2 Sample Construction

We retrieved the data from the website of Prosper.com in January of 2012. Our initial data consist of all listings that were created between May to October of 2008 (and the corresponding loan repayment data for funded listings, which go until the end of 2011). We then drop closed listings, withdrawn listings, and listings that are registered in Texas. The reason for dropping listings registered in Texas is because a different interest rate cap is used there.

The rationale for restricting the sample to those that were listed between May and October of 2008 is as follows. First, there are substantial institutional differences across states before April 2008, such as interest rate caps. Second, the SEC issued a cease and desist order on Prosper in November of 2008 for Securities Act violations, which prevented Prosper from originating new loans for a substantial period of length. Hence, we have no observations from November 2008 to June 2009. Lastly, Prosper made changes to the minimum bid amount from $50 to $25 and also changed its definition of the credit grades after its relaunch in July 2009. Hence we focus on listings that are created between May to October of 2008.

2.3 Lender Portfolio Size and IRR

Table 3 of Section 2 in the main text reports summary statistics of the lenders' portfolio. Although there is substantial heterogeneity in portfolio size and portfolio performance across lenders, we do not find a strong correlation between the two. The first column of Table 2 below reports the results of a regression of portfolio-level IRR on the number of loans funded and the amount of total investment by a lender. The second column reports the results of the same regression but with logged independent variables. The unit of observation is a lender portfolio in each regression.

We find that the effect on IRR is quite small. The result implies that a one standard deviation increase in the number of loans leads to about a 0.2 percentage point increase in IRR, and a one standard deviation increase in the amount of investment leads to a 0.07 percentage point decrease in the IRR. The estimated coefficients are not statistically significant at the 5% significance level.

Table 2: Lender level IRR

	(1)	(2)
# Loans	4.64E-05	-
	(2.72E-05)	-
Amount Investment ($1000)	-1.32E-04	-
	(4.79E-04)	-
Log(# Loans)	-	1.50E-03
	-	(2.24E-03)
Log(Amt. Investment)	-	3.21E-03
	-	(2.26E-03)
Obs.	27,792	27,792

Note: The table reports the results of the OLS Regression of lender-level IRR. Standard errors are reported in parentheses.

3 Additional Reduced-form Results

In this section, we provide additional results for Section 3 of the paper. Section 3.1 reports coefficient estimates that are omitted from Table 4 of the main text. Section 3.2 reports the regression results when we use all credit grades for estimation. Section 3.3 discusses results from alternative specifications of the reduced-form model. In particular, we discuss the details of a survey that we conducted in order to obtain simple measures that summarize the richness of the information contained in pictures and texts. Section 3.4 reports the estimates of the funding probability as a nonparametric function of the reserve rate. Section 3.5 discusses why the funding probability is non-monotone in the reserve rate for credit grade D. Section 3.6 reports reduced-form results that support the assumption that the borrower's outside option (λ) is positively correlated with the borrower's default cost (φ). Section 3.7 discusses the details of the field experiment that we ran on Amazon's Mechanical Turk.

3.1 Coefficient Estimates for Covariates Omitted in Table 4

Table 4 in Section 3 of the main text does not report some of the coefficient estimates of the covariates included in the regression. Table 3 reports the coefficient estimates for all of the borrowers' financial characteristics and the word counts of the title and the description. We do not report the self-reported occupation and income dummies and other time dummies.

3.2 Including All Credit Grades

Table 4 in Section 3 of the main text reports the estimation results for the sample of listings in credit grades AA to C. We now report the regression results for all credit grades. Table

Table 3: Reduced Form Analysis: Funding Probability, Contract Rate and Repayment

	(1) Funded	(2) Contract rate	(3) Ever default	(4) IRR
Reserve rate	6.439***	0.461***	1.427**	-0.501***
	(0.267)	(0.015)	(0.603)	(0.153)
Contract rate			1.996***	0.188
			(0.702)	(0.185)
Amount ($1000)	-0.133***	0.007***	0.035***	-0.007***
	(0.003)	(1.84E-04)	(0.006)	(0.001)
Debt/income	-1.246***	0.057***	0.082	-0.004
	(0.058)	(0.003)	(0.086)	(0.023)
Home-ownership	0.004	9.54E-04	0.261***	-0.052***
	(0.034)	(0.002)	(0.058)	(0.014)
Grade				
AA	2.025***	-0.098***	-0.495***	0.045
	(0.070)	(0.004)	(0.118)	(0.028)
A	1.397***	-0.070***	-0.247***	0.028
	(0.055)	(0.003)	(0.091)	(0.022)
B	0.702***	-0.037***	-0.045	-0.010
	(0.040)	(0.002)	(0.066)	(0.017)
Amount delinquent	-0.015**	7.28E-04**	0.005	-0.004
	(0.005)	(2.72E-04)	(0.008)	(0.002)
Bank card utilization	-0.082	0.005	-0.030	0.021
	(0.052)	(0.003)	(0.088)	(0.022)
Current credit line	0.019*	-0.001*	-0.011	0.002
	(0.010)	(5.25E-04)	(0.016)	(0.004)
Current delinquencies	-0.376***	0.021***	-0.013	0.008
	(0.018)	(0.001)	(0.031)	(0.008)
Delinquencies last 7 years	-0.010***	5.62E-04***	0.002	1.02E-04
	(0.002)	(1.02E-04)	(0.003)	(8.28E-04)
Inquiries last 6 months	-0.118***	0.007***	0.059***	-0.019***
	(0.006)	(3.57E-04)	(0.011)	(0.003)
Length status	0.001*	-5.35E-04	-8.66E-04	1.36E-04
	(5.43E-04)	(2.89E-05)	(8.87E-04)	(2.08E-04)
Open credit lines	-0.026**	0.002**	0.008	-4.61E-04
	(0.010)	(5.47E-04)	(0.017)	(0.004)
Public records 10 years	-0.238***	0.012***	0.028	-0.017
	(0.025)	(0.001)	(0.042)	(0.011)
Public records 12 months	0.038	-0.005	0.072	0.036
	(0.086)	(0.005)	(0.168)	(0.041)
Total credit lines	-0.007***	4.15E-04***	0.001	-7.09E-04
	(0.002)	(8.66E-05)	(0.003)	(0.001)
Word count: description	2.02E-04***	-9.21E-06***	-1.00***	2.93E-05***
	(2.00E-05)	(1.06E-06)	(3.24E-05)	(7.68E-06)
Word count: title	0.001	-1.06E-04	-6.15E-04	6.00E-05
	(0.001)	(6.44E-05)	(0.002)	(4.81E-04)
Constant	-0.339**	0.073***	-1.714***	0.117**
	(0.130)	(0.008)	(0.236)	(0.056)
Occupation/income	X	X	X	X
Time dummies	X	X	X	X
Observations	11,539	11,539	3,737	3,737

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

4 reports the coefficient estimates for all of the borrowers' financial characteristics and the word counts of the title and the description.

3.3 Robustness of Regression Analyses: Alternative Specifications

We examine the robustness of the regression results that we report in Table 4 of the main text. In particular, we re-estimate the same regressions using subsets of the listings and include additional variables.

Restricting the Sample to Debt Consolidation We re-estimate the regressions in Section 3 of the main text using the subsample of listings and loans in which the stated purpose of the loan is debt consolidation. Debt consolidation is the most popular loan purpose, accounting for about 50% of all listings. Table 5 reports the regression results and shows that our main results remain similar.

Restricting the Sample to Listings with Reserve Rate *Less Than* the Cap Table 6 reports the estimates of the four regressions when we restrict the sample to those for which the reserve rates are strictly less than 36%. We find that the results are both qualitatively and quantitatively similar to what we report in Table 4 of the main text. We find that the coefficients on the reserve rate are larger in magnitude than what we report in the main text for regressions (1), (3), and (4). (The coefficients reported in the main text are 6.439, 0.461, 1.427, and -0.501, respectively).

Restricting the Sample to Listings with Reserve Rate *Equal to* the Cap We now estimate the regressions for the subset of listings for which the reserve rate is at the cap of 36%. Table 7 reports the results.

One interesting aspect of listings that are at the reserve price cap is that the reserve price has less signaling value. It seems possible that when this is the case, other aspects of the listing such as the amount requested may be more informative about the borrowers. Relative to the coefficients on the amount that we report in our main specification for the first two regressions (-0.13 and 0.0071) the coefficient on the amount reported in Table 7 are larger in magnitude. However, the estimated coefficients on the amount for the last two regressions are not precisely estimated. It is hard to determine from these regression results whether or not the requested amount has more signaling value when the reserve rates are at the cap.

Controlling for Urgency One aspect of listings that we do not account for in our reduced-form results is how quickly the borrower needs credit. In order to investigate heterogeneity

Table 4: Reduced Form Analysis: Funding Probability, Contract Rate and Repayment (All Credit Grades)

	(1) Funded	(2) Contract Rate	(3) Ever Default	(4) IRR
Reserve Rate	3.574***	0.556***	2.748***	-0.484***
	(0.185)	(0.015)	(0.789)	(0.131)
Contract Rate			2.789***	0.291**
			(0.797)	(0.139)
Amount ($1000)	-0.132***	0.009***	0.059***	-0.007***
	(0.003)	(2.10E-04)	(0.008)	(0.001)
Debt/income	-0.962***	0.076***	0.162	-0.033
	(0.046)	(0.004)	(0.113)	(0.021)
Home-ownership	-0.055**	0.004*	0.337***	-0.047***
	(0.025)	(0.002)	(0.079)	(0.013)
Grade				
AA	3.547***	-0.258***	-1.009***	0.085*
	(0.077)	(0.006)	(0.263)	(0.043)
A	2.917***	-0.226***	-0.624**	0.066
	(0.065)	(0.005)	(0.226)	(0.038)
B	2.294***	-0.191***	-0.311	0.034
	(0.053)	(0.005)	(0.189)	(0.033)
C	1.606***	-0.148***	-0.228	0.038
	(0.043)	(0.004)	(0.166)	(0.030)
D	1.006***	-0.100***	-0.139	0.051*
	(0.038)	(0.003)	(0.154)	(0.028)
E	0.368***	-0.033***	-0.270	0.096**
	(0.041)	(0.004)	(0.164)	(0.030)
Amount delinquent	-0.017***	0.009***	-0.004	9.07E-04
	(0.003)	(2.30E-04)	(0.0107)	(0.0022)
Bank card utilization	0.069**	-0.005	0.047	0.022
	(0.031)	(0.002)	(0.103)	(0.017)
Current credit lines	0.030***	-0.002***	-0.021	0.0024
	(0.007)	(0.005)	(0.022)	(0.004)
Current delinquencies	-0.124***	0.013***	0.029	-0.008**
	(0.007)	(0.001)	(0.022)	(0.004)
Delinquent last 7 years	-0.010***	0.001***	-0.004	0.001*
	(0.001)	(1.00E-04)	(0.004)	(6.40E-04)
Inquiries last 6 months	-0.071***	0.007***	0.075***	-0.0162***
	(0.004)	(3.00E-04)	(0.012)	(0.002)
Length status in months	5.65E-04	-3.00E-04	-0.001	1.91E-04
	(3.80E-04)	(3.00E-04)	(0.001)	(1.90E-04)
Open credit lines	-0.024***	0.002***	0.020	-0.001
	(0.007)	(0.001)	(0.023)	(0.004)
Public records 10 years	-0.138***	0.011***	0.036	-0.012
	(0.015)	(0.001)	(0.046)	(0.008)
Public records 12 months	0.080*	-0.009***	-0.065	0.060**
	(0.046)	(0.004)	(0.157)	(0.027)
Total credit lines	-0.009***	0.001***	0.003	-0.001
	(0.001)	(8.00E-04)	(0.004)	(0.001)
Word count: description	0.026***	-0.016***	-0.013***	0.093**
	(0.001)	(0.001)	(0.004)	(0.041)
Word count: title	0.257**	-0.200***	-0.353	-0.153
	(0.081)	(0.006)	(0.259)	(0.259)
Constant	-1.910***	0.214***	-2.080***	0.038
	(0.107)	(0.008)	(0.346)	(0.059)
Occupation/Income	X	X	X	X
Time dummies	X	X	X	X
Observations	35,236	35,236	5,530	5,530
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	

Table 5: Reduced Form Analysis: Debt Consolidation

	(1) Funded	(2) Contract rate	(3) Ever default	(4) IRR
Reserve rate	6.231***	0.499***	2.076**	-0.474*
	(0.394)	(0.0237)	(0.969)	(0.246)
Contract rate			2.330**	0.015
			(1.119)	(0.290)
Amount ($1,000)	-0.144***	0.008***	0.012	-3.75E-04
	(0.005)	(3.23E-04)	(0.010)	(0.002)
Debt/income	-1.865***	0.067***	0.181	0.001
	(0.092)	(0.006)	(0.203)	(0.036)
Home-ownership	0.033	-9.78E-04	0.391***	-0.091***
	(0.052)	(0.003)	(0.091)	(0.022)
Grade				
AA	1.991***	-0.100***	-0.537***	0.089***
	(0.110)	(0.006)	(0.190)	(0.045)
A	1.366***	-0.068***	-0.193	0.048
	(0.084)	(0.005)	(0.145)	(0.035)
B	0.672***	-0.036***	-0.111	0.026
	(0.059)	(0.003)	(0.102)	(0.026)
Amount delinquent	0.010	-8.11E-04	0.011	-0.008
	(0.008)	(4.55E-04)	(0.018)	(0.004)
Bank card utilization	-0.192**	0.013***	-0.345**	0.092**
	(0.086)	(0.005)	(0.149)	(0.037)
Current credit line	0.026	-0.001	0.003	7.87E-04
	(0.014)	(7.96E-04)	(0.025)	(0.006)
Current delinquencies	-0.538***	0.029***	-0.113	0.050***
	(0.036)	(0.002)	(0.071)	(0.017)
Delinquencies last 7 years	-0.017***	0.001***	-0.002	-2.97E-04
	(0.003)	(1.88E-04)	(0.006)	(0.002)
Inquiries last 6 months	-0.138***	0.008***	0.038**	-0.018***
	(0.010)	(6.02E-04)	(0.019)	(0.005)
Length status	4.09E-04	-2.97E-05	-2.36E-04	0.0001
	(8.35E-04)	(4.73E-05)	(0.002)	(3.48E-04)
Open credit lines	-0.036*	0.002**	-0.022	0.002
	(0.015)	(8.28E-04)	(0.026)	(0.006)
Public records 10 years	-0.301***	0.017***	0.105	-0.017
	(0.040)	(0.002)	(0.069)	(0.017)
Public records 12 months	-0.315	0.016	0.214	0.026
	(0.205)	(0.012)	(0.403)	(0.106)
Total credit lines	-0.006**	4.31E-04**	0.006	-0.001
	(0.002)	(1.37E-04)	(0.004)	(0.001)
Word count: description	2.07E-04***	-1.07E-05***	-6.89E-05	3.19E-05**
	(3.19E-05)	(1.75E-06)	(5.33E-05)	(1.26E-05)
Word count: title	0.002	-1.06E-04	0.005	-0.001
	(0.002)	(1.06E-04)	(0.003)	(7.94E-04)
Constant	0.053	0.043***	-1.612***	0.098
	(0.213)	(0.012)	(0.375)	(0.090)
Occupation/income	X	X	X	X
Time dummies	X	X	X	X
Observations	5,569	5,569	1,536	1,536

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

Table 6: Reduced Form Analysis: Less than the Cap

	(1) Funded	(2) Contract rate	(3) Ever default	(4) IRR
Reserve rate	8.878***	0.380***	2.364***	-0.698***
	(0.342)	(0.017)	(0.747)	(0.185)
Contract rate			1.986**	0.224
			(0.858)	(0.217)
Amount ($1000)	-0.135***	0.006***	0.032***	-0.006***
	(0.004)	(1.63E-04)	(0.006)	(0.001)
Debt/income	-1.235***	0.047***	0.087	-0.007
	(0.061)	(0.003)	(0.088)	(0.023)
Home-ownership	-0.019	0.002	0.271***	-0.055***
	(0.037)	(0.002)	(0.062)	(0.015)
Grade AA	2.115***	-0.088***	-0.456***	0.036
	(0.075)	(0.003)	(0.125)	(0.029)
Grade A	1.425***	-0.059***	-0.233**	0.026
	(0.058)	(0.003)	(0.096)	(0.023)
Grade B	0.685***	-0.028***	-0.078	-0.005
	(0.043)	(0.002)	(0.070)	(0.017)
Amount delinquent	-0.022***	0.001***	0.003	-0.003
	(0.006)	(2.77E-04)	(0.009)	(0.002)
Bank card utilization	-0.082	0.004	-0.031	0.017
	(0.057)	(0.003)	(0.094)	(0.023)
Current credit line	0.011	-6.06E-04	-0.006	5.07E-04
	(0.010)	(4.77E-04)	(0.017)	(0.004)
Current delinquencies	-0.430***	0.021***	-0.044	0.013
	(0.020)	(9.76E-04)	(0.036)	(0.009)
Delinquencies last 7 years	-0.011***	5.18E-04***	0.003	-5.96E-05
	(0.002)	(9.21E-05)	(0.003)	(8.55E-04)
Inquiries last 6 months	-0.123***	0.006***	0.059***	-0.018***
	(0.007)	(3.25E-04)	(0.012)	(0.003)
Length status	0.001**	-4.37E-05	-0.0004	6.52E-05
	(5.80E-04)	(2.54E-05)	(9.22E-04)	(2.10E-04)
Open credit lines	-0.022**	0.001**	0.006	8.49E-04
	(0.011)	(4.95E-04)	(0.018)	(0.004)
Public records 10 years	-0.245***	0.011***	0.013	-0.012
	(0.026)	(0.001)	(0.044)	(0.011)
Public records 12 months	0.037	-0.003	0.168	0.006
	(0.093)	(0.004)	(0.178)	(0.042)
Total credit lines	-0.007***	3.75E-04***	1.16E-04	-4.70E-04
	(0.002)	(7.80E-05)	(0.003)	(6.65E-04)
Word count: description	1.69E-04***	-5.24E-06***	-8.99E-05***	2.60E-05***
	(2.15E-05)	(9.52E-07)	(3.48E-05)	(7.90E-06)
Word count: title	0.001	-7.22E-05	-4.01E-04	4.71E-05
	(0.001)	(5.75E-05)	(0.002)	(4.90E-04)
Constant	-0.695***	0.0747***	-1.909***	0.155***
	(0.148)	(0.00644)	(0.253)	(0.0571)
Occupation/income	X	X	X	X
Time dummies	X	X	X	X
Observations	10,027	10,027	3,439	3,439

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

Table 7: Reduced Form Analysis: At the Cap

	(1) Funded	(2) Contract rate	(3) Ever default	(4) IRR
Contract rate			2.688	-0.046
			(1.885)	(0.550)
Amount ($1000)	-0.182***	0.020***	0.036	-0.006
	(0.014)	(0.001)	(0.029)	(0.008)
Debt/income	-1.257***	0.188***	0.554	-0.056
	(0.225)	(0.029)	(0.611)	(0.141)
Home-ownership	0.104	-0.013	0.604*	-0.093
	(0.107)	(0.012)	(0.261)	(0.074)
Grade				
AA	-	-0.151***	-	0.107
		(0.057)		(0.255)
A	2.236***	-0.244***	-0.029	-0.046
	(0.309)	(0.031)	(0.625)	(0.168)
B	0.873***	-0.098***	0.290	-0.058
	(0.128)	(0.014)	(0.271)	(0.080)
Amount delinquent	-0.012	-4.08E-04	0.016	-0.009
	(0.011)	(0.001)	(0.031)	(0.008)
Bank card utilization	-0.318**	0.036**	-0.257	0.086
	(0.151)	(0.017)	(0.335)	(0.102)
Current credit line	0.075***	-0.007**	-0.0570	0.016
	(0.028)	(0.003)	(0.057)	(0.017)
Current delinquencies	-0.179***	0.026***	0.004	0.014
	(0.037)	(0.005)	(0.099)	(0.031)
Delinquencies last 7 years	-0.013*	0.002**	-0.009	0.001
	(0.005)	(6.15E-04)	(0.014)	(0.004)
Inquiries last 6 months	-0.108***	0.013***	0.015	-0.011
	(0.018)	(0.002)	(0.040)	(0.012)
Length status	0.002	-1.84E-04	-0.006	0.001
	(0.002)	(2.07E-04)	(0.004)	(0.001)
Open credit lines	-0.091***	0.007**	-0.016	-0.005
	(0.029)	(0.003)	(0.061)	(0.019)
Public records 10 years	-0.362***	0.034***	0.139	-0.032
	(0.085)	(0.009)	(0.189)	(0.059)
Public records 12 months	0.336	-0.031	-1.134	0.331
	(0.252)	(0.028)	(0.778)	(0.185)
Total credit lines	-0.001	0.001	0.018	-0.004
	(0.005)	(5.31E-04)	(0.011)	(0.003)
Word count: description	3.58E-04***	-3.66E-05***	-3.13E-04**	5.53E-05
	(6.19E-05)	(6.63E-06)	(1.33E-05)	(3.65E-05)
Word count: title	0.004	-4.33E-04	-1.23E-04	-0.001
	(0.004)	(4.06E-04)	(0.008)	(0.002)
Constant	1.538***	-0.201***	-0.449	0.147
	(0.465)	(0.048)	(1.022)	(0.371)
Occupation/income	X	X	X	X
Time dummies	X	X	X	X
Observations	1,555	1,555	242	242

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

in urgency within our sample, we constructed an indicator variable *urgent* which takes the value one if the listing title or the listing description includes one of the following strings: *urgen-*, *quick-*, *fast*, *immediate-*, and *in a hurry*. 6.86% of the listings and 9.38% of the loans have $urgent = 1$. We now wish to show that even if we control for urgency, the relationship between s and the outcome variables are not fundamentally altered. Table 8 presents our reduced-form regression results that conditions on loan purpose (debt consolidation) and urgency. In particular, for each regression, we include a dummy for urgency as well as an interaction term between urgency and s. We condition on a particular loan purpose in order to control for as much heterogeneity as possible regarding the reason for attempting to borrow through Prosper.

We find that the coefficients on the urgency dummy are negative and statistically significant for regressions (1) and (3), and positive and statistically significant for regression (2). The results suggest that urgent listings are associated with lower probability of funding and default, and higher contract interest rates. We also estimate statistically significant coefficients for the interaction term in regressions (1), (2) and (3). The results suggest that the relationship between s and funding is amplified for urgent listings while the relationship between s and the contract interest rate is mollified for urgent listings. In all of the regressions, we find that the coefficient on the reserve rate remains robust and statistically significant.

Controlling for Selection Using Heckit For the default and the IRR regressions, we use the sample of funded listings. As long as the funding probability is a function of only observable variables, there is no need to control for sample selection. However, if there is selection on unobservables, we need to account for this in our estimates.

In order to control for potential selection on unobservables, we run a Heckman-style sample selection equation (Heckman 1976). Although the Heckit estimator is known to be quite sensitive to the specification of the error terms in the absence of instruments (See, e.g., LaLonde 1986) it allows for a simple way for controlling for selection without imposing too much structure. We report the Heckit estimates in Table 9 below:[3]

We find that the coefficient on the reserve rate is positive and statistically significant in the default regression. We also find that the coefficient on the reserve rate is negative and

[3]We estimate the following model for default:

$$\begin{aligned}\text{Default}_j &= \beta_s s_j + \beta_r r_j + \mathbf{x}_j' \beta_x + \epsilon_j,\\ \text{Funded}_j &= \mathbf{1}\{\gamma_s s_j + \gamma_r r_j + \mathbf{x}_j' \gamma_x + u_j \geq 0\},\end{aligned}$$

where the errors $\{\epsilon_j, u_j\}$ are distributed jointly Normal with covariance matrix Σ. We report the coefficient estimates for β. The estimates for IRR are analogous: we estimate a model in which the outcome variable in the first equation is replaced with IRR_j.

Table 8: Reduced Form Analysis: Urgent & Debt Consolidation

	(1) Funded	(2) Contract rate	(3) Ever default	(4) IRR
Reserve rate	6.040***	0.511***	1.728*	-0.454*
	(0.403)	(0.024)	(0.989)	(0.251)
Urgent	-0.387*	0.024*	-0.996**	0.105
	(0.229)	(0.013)	(0.444)	(0.091)
Reserve rate x Urgent	1.941*	-0.113**	4.002**	-0.264
	(1.015)	(0.055)	(1.874)	(0.419)
Contract rate			2.272**	0.010
			(1.124)	(0.292)
Amount ($1,000)	-0.144***	0.008***	0.010	5.78E-05
	(0.006)	(3.24E-04)	(0.010)	(0.002)
Debt/income	-1.861***	0.067***	0.173	0.002
	(0.092)	(0.006)	(0.200)	(0.036)
Home-ownership	0.025	-0.001	0.372***	-0.089***
	(0.053)	(0.003)	(0.091)	(0.022)
Grade				
AA	1.975***	-0.099***	-0.531***	0.080*
	(0.110)	(0.006)	(0.192)	(0.045)
A	1.366***	-0.068***	-0.170	0.041
	(0.084)	(0.005)	(0.146)	(0.035)
B	0.668***	-0.036***	-0.108	0.022
	(0.059)	(0.003)	(0.103)	(0.026)
Amount delinquent	0.004	-2.91E-04	-0.001	-0.005
	(0.009)	(0.001)	(0.021)	(0.005)
Bank card utilization	-0.188**	0.013***	-0.340**	0.090**
	(0.087)	(0.005)	(0.149)	(0.037)
Current credit lines	0.025*	-0.001	5.92E-04	0.001
	(0.014)	(7.98E-04)	(0.025)	(0.006)
Current delinquencies	-0.531***	0.028***	-0.105	0.049***
	(0.036)	(0.002)	(0.072)	(0.018)
Delinquent last 7 years	-0.016***	0.001***	-0.002	-2.83E-04
	(0.003)	(1.88E-04)	(0.006)	(0.002)
Inquiries last 6 months	-0.137***	0.008***	0.039**	-0.018***
	(0.010)	(6.03E-04)	(0.019)	(0.005)
Length status	4.25E-04	-3.18E-05	-0.002	1.32E-04
	(8.37E-04)	(4.74E-05)	(0.002)	(3.50E-04)
Open credit lines	-0.035**	0.002**	-0.020	0.002
	(0.015)	(8.31E-04)	(0.026)	(0.006)
Public records last 10 years	-0.305***	0.017***	0.098	-0.015
	(0.040)	(0.002)	(0.069)	(0.018)
Public records last 12 months	-0.295	0.015	0.183	0.029
	(0.206)	(0.012)	(0.407)	(0.107)
Total credit lines	-0.006***	4.23E-04**	0.006	-0.001
	(0.002)	(1.37E-04)	(0.004)	(0.001)
Word count: description	2.00E-04***	-1.03E-05***	-6.35E-05	2.69E-05**
	(3.27E-05)	(1.80E-06)	(5.49E-05)	(1.30E-05)
Word count: title	0.002	-1.13E-04	0.005	-0.001
	(0.002)	(1.07E-04)	(0.003)	(8.01E-04)
Constant	0.098	0.041***	-1.519***	0.099
	(0.215)	(0.012)	(0.379)	(0.092)
Occupation/income	X	X	X	X
Time dummies	X	X	X	X
Observations	5,517	5,517	1,521	1,521

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

Table 9: Heckit Regression Results for Ever Default and IRR

	(1) Ever default	(2) IRR
Reserve Rate	0.375**	-0.484***
	(0.174)	(0.155)
Contract Rate	0.689***	0.179
	(0.216)	(0.193)
Amount ($1000)	8.28E-03***	-6.61E-03***
	(1.53E-03)	(1.37E-03)
Debt/income	0.020	-3.29E-03
	(0.025)	(0.022)
Home-ownership	0.070***	-0.052***
	(0.016)	(0.014)
Grade		
AA	-0.095***	0.047*
	(0.031)	(0.027)
A	-0.052**	0.029
	(0.025)	(0.022)
B	-8.04E-03	-8.96E-03
	(0.019)	(0.017)
Amount delinquent	1.90E-03	-3.87E-03*
	(2.50E-03)	(2.24E-03)
Bank card utilization	-0.010	0.018
	(0.024)	(0.022)
Current credit line	-3.51E-03	2.28E-03
	(4.45E-03)	(3.98E-03)
Current delinquencies	5.05E-03	7.72E-03
	(9.00E-03)	(8.06E-03)
Delinquencies last 7 years	2.91E-04	9.98E-05
	(9.11E-04)	(8.16E-04)
Inquiries last 6 months	0.018***	-0.019***
	(3.14E-03)	(2.82E-03)
Length status	2.45E-04	1.41E-04
	(2.29E-04)	(2.05E-04))
Open credit lines	2.44E-03	-4.03E-04
	(4.59E-03)	(4.11E-03)
Public records 10 years	9.11E-03	-0.016
	(0.012)	(0.011)
Public records 12 months	9.16E-03	0.036
	(0.045)	(0.040)
Total credit lines	4.04E-04	-7.77E-04
	(7.14E-04)	(6.39E-04)
Word count: description	-2.62E-05***	2.97E-05***
	(8.46E-06)	(7.57E-06)
Word count: title	-2.09E-04	3.93E-05
	(5.30E-04)	(4.74E-04)
Constant	-0.079	0.177
	(0.165)	(0.148)
Occupation/income	X	X
Time dummies	X	X
Observations	10,683	10,683

Note: The first column reports the estimated coefficients of the linear probability model (expression (2)). The second column reports the estimated coefficients of the reserve rate regression model (expression (3)) *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels.

statistically significant in the IRR regression.

Controlling for measures of creditworthiness perceived from pictures and text In order to better control for aspects of the listing that are hard to capture in regression analysis, we ran an online survey on Mechanical Turk in which we ask the respondents to rate the "creditworthiness" of the borrower on a scale from 1 (least creditworthy) to 7 (most creditworthy) by showing the subjects pictures and text from listings drawn randomly from our sample. Note that the survey respondents were shown only the picture and the text of each listing, and not other aspects of the listing characteristics. The survey question is motivated by a study by Ravina (2019) who studies the effects of various personal attributes of the borrowers on Prosper (as perceived from the posted pictures) on outcome variables such as whether a loan is funded, contract interest rate, default, and internal rate of return. She finds that the survey question that asks the respondents about the perceived creditworthiness of the borrower is an important explanatory variable in her regressions.

In order to roll out the survey, we first created a random subset of listings ($N = 5,000$) that we pre-formatted. This corresponds to about 14% of the full sample (There are a total of 35,241 listings in the full sample). We posted the survey on Mechanical Turk during March 11th, 2016 and May 26th, 2016. Each survey respondent was then shown 25 or 50 listings randomly from the sample of $5,000$, in sequence.[4] Each respondent was paid $2 for participation. There are a total of 575 respondents, a small number of whom we exclude because they did not answer most of the questions. Hence, we end up with 562 respondents for our empirical analysis.

Table 10 presents the summary statistics of the sample of listings chosen for the survey for which we received at least one rating. We also report the summary statistics of the original sample so that it is straightforward to compare the two data sets. Note that the sample size is $4,438$, which is less than $5,000$. Given that we randomly assigned listings to respondents, there are listings that received 0 ratings. Table 10 suggests that the summary statistics of the listings that received ratings are generally similar to those of the full sample. In the third-to-last column of the table, we report the average rating. We first compute the mean rating for each listing (for listings with multiple ratings), and take the average across all listings by credit grade. Note that the survey respondents are not shown the credit grade of the borrower (or any other characteristics aside from the picture and text).

Panel (B) of Table 10 presents the summary statistics for the subset of loans with at least one rating. Again, we report the summary statistics side-by-side with those of the full

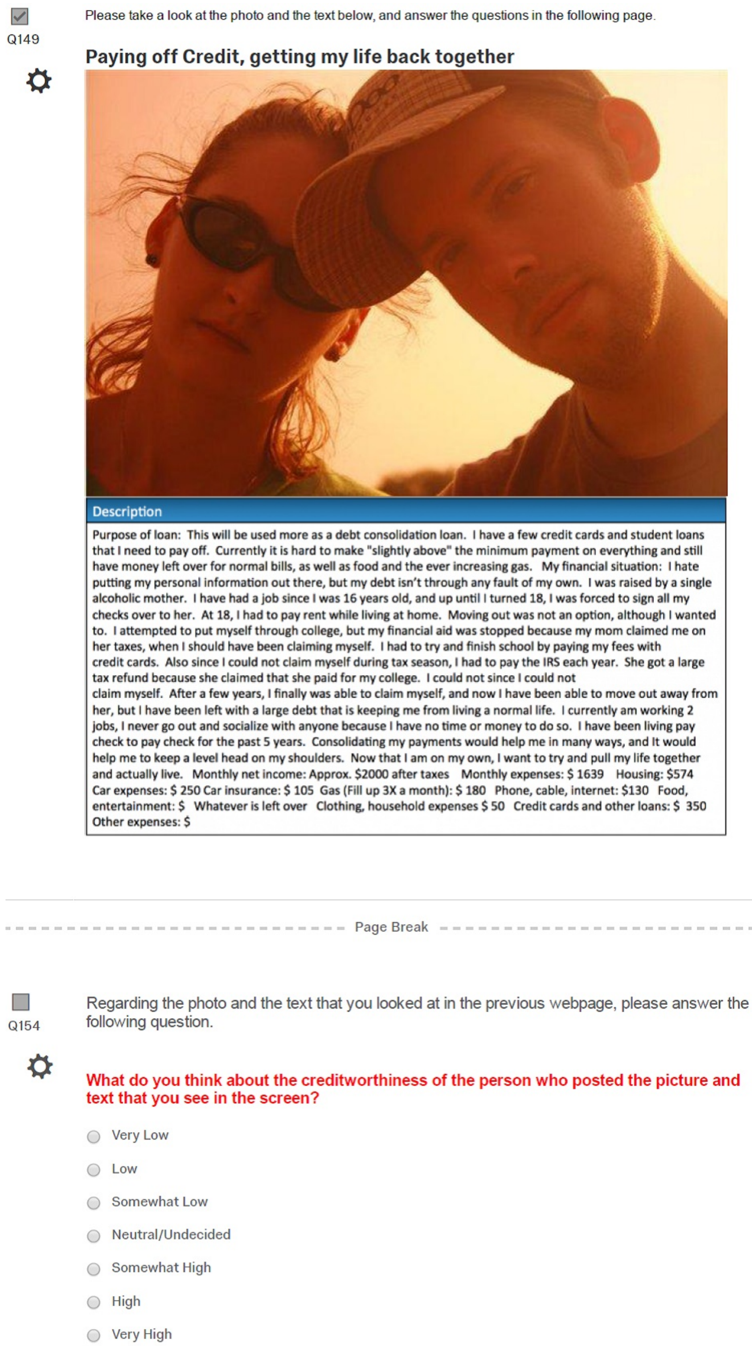
[4]For some respondents with slow connection, there were issues with loading the survey webpage in a timely manner. We first started out by showing 50 listings per respondent, but then changed to 25 listings per respondent.

Table 10: Descriptive Statistics– Survey vs Original

	Amount Requested		Reserve Rate		Debt/ Income		Home Owner		Bid Count		Fund Pr		Contract Rate		Rating	Obs	
Grade	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)		(1)	(2)
								Listing Characteristics									
AA	14444	13145	13.0	13.2	0.40	0.36	0.77	0.81	155	153.2	42.5	53.4			3.95	174	1,420
	(9453)	(8343)	(4.8)	(4.7)	(1.15)	(0.98)	(0.42)	(0.39)	(208.3)	(178.6)					(1.30)		
A	12540	12396	16.5	16.5	0.35	0.38	0.59	0.61	131.6	103.3	41.7	40.9			3.89	259	1,850
	(7870)	(7882)	(6.7)	(6.7)	(0.66)	(0.67)	(0.49)	(0.49)	(183.6)	(139.9)					(1.17)		
B	10544	10622	21.2	21.1	0.33	0.39	0.59	0.59	86.7	74.4	35.4	33.4			4.15	396	3,068
	(6237)	(6096)	(7.7)	(7.5)	(0.23)	(0.65)	(0.49)	(0.49)	(118.2)	(104.0)					(1.18)		
C	7484	7622	24.5	24.6	0.37	0.37	0.58	0.56	42.2	36.2	26.7	24.7			3.96	651	5,203
	(5147)	(5158)	(7.9)	(7.8)	(0.55)	(0.62)	(0.49)	(0.50)	(67.9)	(59.2)					(1.10)		
D	6740	6368	28.5	28.7	0.39	0.39	(0.38)	0.37	19.1	18.2	15.7	15.5			3.91	828	6,581
	(5244)	(4691)	(7.6)	(7.5)	(0.61)	(0.71)	(0.49)	(0.48)	(42.1)	(39.0)					(1.16)		
E	4665	4783	30.8	31.0	0.40	0.36	0.35	0.33	4.4	3.99	6.6	6.8			3.81	720	5,757
	(4546)	(4868)	(7.5)	(7.3)	(0.86)	(0.68)	(0.48)	(0.47)	(12.2)	(11.0)					(1.11)		
HR	4284	4351	31.5	31.5	0.33	0.31	0.22	0.22	2.3	1.9	2.6	3.0			3.78	1410	11,362
	(4476)	(4599)	(6.9)	(6.9)	(0.77)	(0.64)	(0.41)	(0.42)	(8.5)	(5.7)					(1.13)		
All	6713	6604	27.3	27.4	0.36	0.35	0.40	0.39	32.7	28.0	16.0	15.8			3.88	4438	35,241
	(6070)	(5938)	(9.0)	(8.9)	(0.71)	0.68	(0.49)	(0.49)	(88.6)	(74.4)					(1.15)		
								Loan Characteristics									
AA	10600	9710	13.6	13.1	0.17	0.21	0.77	0.80	323	258.1			9.8	9.6	4.23	74	755
	(7725)	(7,384)	(4.9)	(3.3)	(0.12)	(0.39)	(0.42)	(0.39)	(225.1)	(186.2)			(3.3)	(4.6)	(1.26)		
A	9727	8723	16.9	16.5	0.22	0.23	0.56	0.55	280.9	218.1			13.0	12.7	4.06	108	755
	(7224)	6626	(6.2)	6.0	(0.14)	(0.14)	(0.50)	(0.14)	(203.6)	(155.4)			(4.5)	(4.5)	(1.12)		
B	7234	7,347	21.8	21.6	0.26	0.27	0.60	0.56	211.4	181.2			16.2	16.4	4.24	140	1,023
	(4313)	(4858)	(6.2)	(6.3)	(0.15)	(0.34)	(0.49)	(34)	(110.0)	(113.9)			(4.1)	(4.6)	(1.10)		
C	4700	4687	25.0	24.7	0.23	0.25	(0.54)	0.48	123.8	112.6			18.5	18.1	3.93	174	1,023
	(3123)	(2998)	(6.5)	6.4	(0.14)	(0.21)	(0.50)	(0.21)	(82.8)	(72.2)			(5.9)	(6.2)	(1.08)		
D	3508	3578	28.7	28.0	0.24	0.24	0.25	0.26	93.8	86.7			21.3	21	3.88	130	1,022
	(2392)	(2380)	(5.7)	(6.4)	(0.13)	(0.17)	(0.44)	(0.17)	(65.7)	(60.8)			(6.4)	(6.6)	(0.99)		
E	1993	1890	32.7	33.9	0.31	0.22	0.31	0.26	39.9	36.4			28.3	29.1	3.82	48	392
	(1825)	(1187)	(4.9)	(2.8)	(0.53)	(0.22)	(0.47)	(0.22)	(27.8)	(20.3)			(6.4)	(5.7)	(1.07)		
HR	2073	1690	34.2	33.9	0.17	0.20	0.08	0.17	42.2	27.0			29.3	30	3.82	36	339
	(2341)	(1288)	(3.0)	(3.6)	(0.11)	(0.44)	(0.28)	(0.44)	(33.1)	(19.0)			(5.6)	5.7	(1.07)		
All	6045	5,821	23.6	23.3	0.23	0.24	0.49	0.47	170.4	144.0			18.0	17.9	4.02	710	5,571
	(5443)	(5285)	(8.3)	(8.6)	(0.19)	(0.28)	(0.50)	(0.28)	(158.2)	(132.4)			(7.4)	(7.9)	(1.10)		

Note: The table reports the descriptive statistics of the data used for the online survey and the full sample. Columns with (1) correspond to the data used for the survey and columns with (2) to the full sample.

Figure 2: Example of a Survey Question



Note: The image is an example of a picture that we show to survey-takers. A subject is shown a picture and a description of a listing first and then asked to answer questions regarding the borrower's creditworthiness on the next page.

sample. The sample of loans corresponds to the subset of the listings that were funded. The mean rating for the sample of loans is generally higher than the mean rating for the sample of listings. The average rating for the loans is 4.02 while that for the listings is 3.88. This suggests that the data that we collected from the survey is informative.[5] The summary statistics reported in Table 10 generally match that of the full sample.

Table 11 reports the estimates of the regressions that control for rating information. Because the sample size is much smaller than that of the full sample, we estimate a parsimonious specification in which we control for only variables that we use in our structural analysis.

We find that the coefficient on Rating is positive and significant in the first regression, suggesting that listings with higher Ratings are funded more, as expected. In the second column, we find that the coefficient on Rating is negative and significant, which implies that higher Ratings decrease the contract interest rate. This result is intuitive, because the interest rate on listings that are perceived to be creditworthy should be bid down.

The estimated coefficient on Ratings in the third column is negative, although it is not precisely estimated. In the fourth column the coefficient on Ratings is positive, but again, not statistically significant.

The second row of Table 11 reports the coefficient on the reserve rate. We find that the estimated coefficient on the reserve rate is statistically significant in all but the third regression despite the fact that the sample size is less than 10% of the original sample.

We take these results as suggestive that controlling for pictures and text information do not change our analysis much.

3.4 Monotonicity of the Funding Probability

In this section, we report estimates of the funding probability as a flexible function of the reserve rate, s. Specifically, we estimate the following semi-parametric specification of the funding probability:

$$\mathbb{E}[\mathbf{1}_{\text{Fund}}|s, X] = f(s) + X'\beta, \tag{1}$$

where $f(\cdot)$ is a nonparametric function that we are interested in and X is a vector of observable characteristics. We use an Epanechnikov kernel for the estimation of $f(\cdot)$ and include 4-th order polynomials of the variables that we use in our structural estimation (amount requested, debt-to-income ratio, home-ownership) in X. We estimate the regression separately by credit grade.

Figure 3 reports our estimates of $f(\cdot)$ and the histogram of s for each credit grade. We find monotonically increasing relationship between the funding probability and the reserve rate for all credit grades except for the region above 30% for credit grade A. Given that only

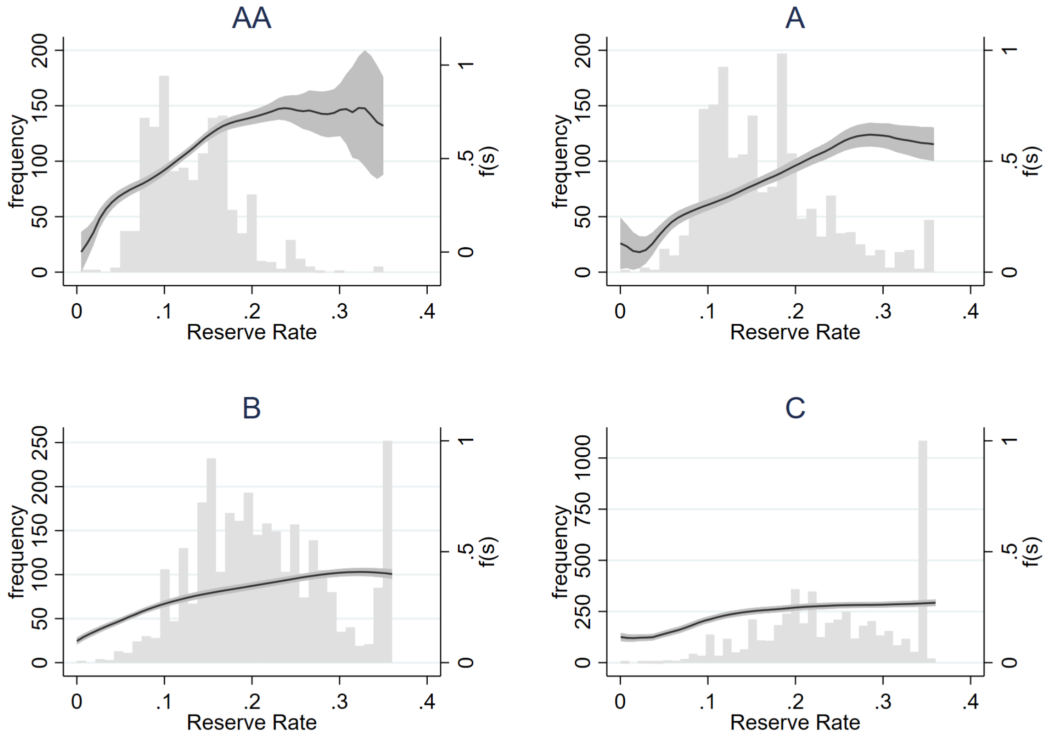
[5] Note that the survey respondents do not know whether or not a listing is funded.

Table 11: Reduced Form Analysis: Rating

	(1) Funded	(2) Contract rate	(3) Ever Default	(4) IRR
Rating	0.076**	-0.004*	0.030	-0.016
	(0.033)	(0.002)	(0.060)	(0.016)
Reserve rate	4.987***	0.508***	1.537	-0.872**
	(0.647)	(0.041)	(1.626)	(0.441)
Contract rate			3.715	0.008
			(2.047)	(0.564)
Amount ($1,000)	-0.092***	0.006***	0.011	-0.003
	(0.007)	(4.78e-4)	(0.012)	(0.003)
Debt / income	-2.202***	0.143***	-0.075	-0.005
	(0.244)	(0.016)	(0.518)	(0.137)
Home-ownership	-0.025	0.003	0.099	-0.036
	(0.079)	(0.005)	(0.140)	(0.036)
Grade				
AA	1.601***	-0.097***	-0.327	-0.002
	(0.170)	(0.010)	(0.310)	(0.076)
A	1.353***	-0.081***	-0.156	0.013
	(0.134)	(0.008)	(0.241)	(0.062)
B	0.728***	-0.047***	-0.016	-0.043
	(0.101)	(0.006)	(0.177)	(0.048)
Constant	-0.818***	0.098***	-1.482***	0.064
	(0.221)	(0.014)	(0.419)	(0.111)
Time dummies	X	X	X	X
Observations	1,480	1,480	495	495

Note: The first column reports the estimated coefficients of the Probit model (expression (1)). The unit of observation is a listing. The dependent variable is an indicator variable that equals one if the listing is funded and zero, otherwise. The second column reports the estimated coefficients of the Tobit model (expression (2)). The dependent variable is the contract interest rate charged to the borrower. The third column reports estimated coefficients from the Probit model (expression (3)). The unit of observation is a loan. The dependent variable is an indicator variable that equals one if the loan ends in default. The fourth column presents estimated coefficients of the OLS model (expression (4)). In this model, the unit of observation is a funded loan. *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels. Standard errors are presented in parentheses below the coefficients.

Figure 3: Probability of Funding



Note: The figures report the nonparametric estimates of the probability of funding as a function of the reserve interest rate for each credit grade.

about 5.8% of the listings in credit grade A has s greater than 30%, we do not think that this small region of non-monotonicity poses a substantial problem for our empirical analysis. Overall, Figure 3 suggests that the monotonicity of $\Pr(s)$ holds throughout the support of s.

We next formally test for the monotonicity of $f(\cdot)$ by modifying a test proposed by Gijbels et al. (2000) to allow for covariates. To explain the test, assume for the time being that β is known in equation (1). Consider T_{count} and T_{run} defined in Gijbels et al. (2000). We construct critical values for these test statistics using the bootstrap as follows. For each observation i, we generate a binary random variable y_i with $\Pr(y_i = 1) = \mu + \beta(X_i - \overline{X})$, where μ is the average funding probability in the sample, β is the parameter in equation (1), and $\overline{X}$ is the mean of X in the sample. Let $\mathbf{y} = \{y_i\}_{i=1}^N$ denote the collection of random variables. We repeat this procedure B times to obtain $(\mathbf{y}^b)_{b=1}^B$ where, for each b, $\mathbf{y}^b$ is a vector of N binary variables. For each b, we compute T^b_{count} and T^b_{run}, using $\{s_i, y^b_i\}_{i=1}^N$. The

Table 12: Test of Monotonicity

Grade	AA	A	B	C
	size 5%			
τ_{count}	0	0	0	0
τ_{run}	0	0	0	0
	size 10%			
τ_{count}	0	0	0	0
τ_{run}	0	0	0	0

critical values are the 95% quantiles of T^b_{count} and T^b_{run}. Note that if $\beta = 0$, this procedure is the same as the one in Gijbels et al. (2000). If we let c_{count} and c_{run} denote the critical values for T^b_{count} and T^b_{run}, our test can be expressed as $\tau_{\text{count}} \equiv \mathbf{1}_{\{T_{\text{count}} > c_{\text{count}}\}}$ and $\tau_{\text{run}} \equiv \mathbf{1}_{\{T_{\text{run}} > c_{\text{run}}\}}$.

We now discuss how we modify the test to allow for sampling error in β. Let $\Theta_\beta(\alpha)$ denote a $1-\alpha$-level confidence set for β. For each $\beta \in \Theta_\beta(\alpha)$, consider the associated test statistics $\tau_{\text{count}}(\beta)$ and $\tau_{\text{run}}(\beta)$. We then take the inf of $\tau_{\text{count}}(\beta)$ and $\tau_{\text{run}}(\beta)$ over $\beta \in \Theta_\beta(\alpha)$:

$$\overline{\tau}^{\alpha}_{\text{count}} = \inf_{\beta \in \Theta_\beta(\alpha)} \tau_{\text{count}}(\beta)$$

$$\overline{\tau}^{\alpha}_{\text{run}} = \inf_{\beta \in \Theta_\beta(\alpha)} \tau_{\text{run}}(\beta).$$

We now define our test as follows. Let α_1 and α_2 be positive numbers and $\alpha = \alpha_1 + \alpha_2$. In practice α is going to be 0.05 or 0.1. Take critical values c_{count} and c_{run} to be the $1-\alpha_1$ quantiles of T^s_{count} and T^s_{run}. Consider the test above for $1-\alpha_2$ confidence set for β. Given the construction of the test, our test if guaranteed to have size of the test less than α. In other words, we can construct a conservative test of monotonicity that accounts for the sampling error in β.

For computation of the test statistic, we use the Epanechnikov kernel for $f(\cdot)$ and select the bandwidth using cross-validation. We include in the vector X, the amount requested, homeownership, and debt-to-income ratio.[6] In order to compute the critical values of β, we used 200 bootstrap samples ($B = 200$). For the choice of k, a tuning variable for the test of Gijbels et al. (2000), we choose $k = 0.2$ as recommended by Gijbels et al. (2000).

The top and bottom panels of Table 12 report the results of the test for size 5% and 10%, respectively. For credit grades AA through C, we cannot reject the null of monotonicity at the 5% or at the 10% with either τ_{count} or τ_{run}.

[6] We chose not to include higher order polynomials of these variables in X for the monotonicity test because estimating a semi-parametric model with large number of covariates takes significantly more time.

3.5 Funding Probability and Reserve Rate for Credit Grade D

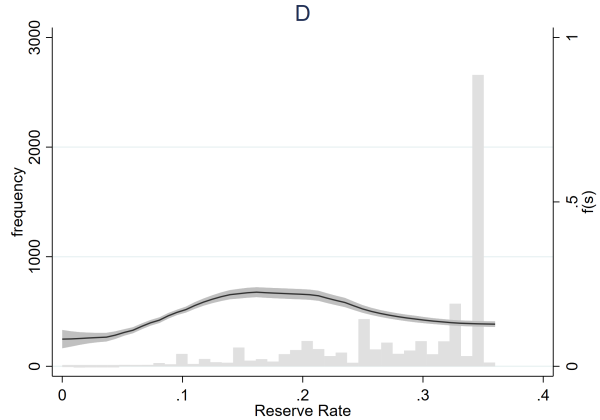
We do not include credit grades D, E, and HR in our analysis. We drop credit grades E and HR because there is little variation in the reserve rate of the borrowers in these credit grades. We also drop credit grade D because we find that the relationship between the reserve rate and the funding probability appears to be non-monotone. Figure 4 plots the estimate of $f(\cdot)$ when we estimate expression (1) for credit grade D. Unlike for credit grades AA-C, we find that the funding probability is first increasing in the reserve rate in the range $s \in [0, 0.20]$ and then decreasing in the range $s \in [0.20, 0.36]$. Indeed, we can reject monotonicity for credit grade D using the test we discussed in the previous section at the 5% level with τ_{count} and at the 10% with τ_{run}. The non-monotonicity suggests that the conditions for signaling appear to not be satisfied for $s \in [0.20, 0.36]$. In particular, given that borrowers can increase the funding probability and decrease the contract interest rate by reducing s in the range $s \in [0.20, 0.36]$, the borrowers' choice of s above 0.20 seems suboptimal.

While it is difficult to give a definitive answer for why some borrowers choose $s \in [0.20, 0.36]$, it seems likely that Prosper's tutorial had some effect on the borrowers' choices. As we mention several times in the paper, Prosper displays a tutorial in which borrowers are told that (1) higher reserve rates are more likely to be funded; and (2) to think about the interest the borrower would be paying on the next best alternative. Note that the same tutorial is given to all borrowers regardless of the credit grade. While the statement that higher reserve rates are more likely to be funded is correct, on average, we have already seen that the statement is incorrect once we condition on credit grade D and s in the range $[0.20, 0.36]$. It seems plausible that some borrowers believe the message of the tutorial despite not being applicable for their specific circumstance. The fact that the message of the tutorial is at odds with the actual funding probability for credit grade D makes it potentially difficult for the borrowers to form rational expectations about their listings' prospects, leading some to choose s that are higher than what is optimal. On the other hand, the fact that Prosper's advice coincides with the actual data patterns for credit grades AA through C is likely to have facilitated signaling in those credit grades.

3.6 Cost of Borrowing from Outside Sources and Creditworthiness

In our model, we make the assumption that the borrower's outside option, λ, is positively correlated with the borrower's default cost, φ. This assumption reflects the idea that a borrower who is unlikely to default (i.e., a high φ type) is likely to have an easier time obtaining a loan from outside sources, such as relatives, friends, and local banks, etc. (i.e. a high λ type). In this section, we provide evidence that supports this assumption.

Figure 4: Probability of Funding for Credit Grade D



In order to do so, we consider two regressions, one for default and another for the reserve rate. The residuals of the default regression reflect the borrowers' propensity to default, which is inversely related to the borrowers' default costs. We correlate these residuals with the residuals from the reserve rate regression. The residuals of the reserve rate regression are likely to be inversely related to the borrowers' outside option since higher reserve rates increase the funding probability (at the cost of higher interest rates). Positive correlation between the residuals of the two regressions would suggest that φ and λ are positively correlated.

We estimate the following linear probability model for default:[7]

$$\text{Default}_j = \mathbf{x}_j' \beta_x + \epsilon_j, \tag{2}$$

where $\mathbf{x}_j$ is the same vector of listing characteristics that we include in equation (3.3) of the main text. Note that s_j and r_j have been excluded from the right-hand side of (2). In the model, s_j is a sufficient statistic of φ, so we exclude s_j in order for the error term in (2) to better reflect the borrower's type φ. Similarly, we exclude r_j from the regression because r_j is determined after the lenders observe s_j and much of the effect of φ on default will load on r_j otherwise. The specification (2) includes only covariates that are observable to the lenders in the absence of reserve rate signaling. The first column of Table 13 reports the regression

[7]We opt to estimate the linear probability model instead of the Probit model (note that we estimate the Probit model in the main text) because the residuals in a Probit model are not uniquely determined even if the parameters β_x are known. To see this, consider the following Probit model

$$\text{Default}_j = \mathbf{1}\{\mathbf{x}_j' \beta_x + \epsilon_j \geq 0\}.$$

Given an outcome of $\text{Default}_j = 1$, any ϵ_j above $-\mathbf{x}_j'\beta_x$ rationalizes the outcome. Similarly, given the outcome $\text{Default}_j = 1$, any ϵ_j less than $-\mathbf{x}_j'\beta_x$ rationalizes the outcome.

results.

Once we have an estimate of β_x, we can obtain estimates of the residuals by subtracting $\mathbf{x}_j'\beta_x$ from the outcome variable, Default_j, as follows:

$$\hat{\epsilon}_j = \text{Default}_j - \mathbf{x}_j'\hat{\beta}_x,$$

where $\hat{\beta}_x$ is a vector of estimated parameters reported in column (1) of Table 13.

Next, we estimate a regression for the reserve rate, s_j:

$$s_j = \mathbf{x}_j'\beta_x + u_j, \tag{3}$$

where s_j is the reserve rate of listing j and $\mathbf{x}_j$ is the same vector of listing characteristics that we include in equation (2). Column (2) of Table 13 reports the results estimated on the sample of funded listings.

We then obtain the residuals by subtracting $\mathbf{x}_j'\beta_x$ from s_j as follows

$$\hat{u}_j = s_j - \mathbf{x}_j'\hat{\beta}_x,$$

where $\hat{\beta}_x$ is a vector of estimated parameters reported in column (2) of Table 13. Note that most of the coefficients in column (1) and (2) have the same sign, suggesting that observable characteristics of the borrower that tend to increase the default probability also increases the reserve rate. The fact that coefficients on observable characteristics have the same sign suggests that unobservable characteristics of the borrower that increase default are also likely to increase the reserve rate (See, e.g., Altonji et al. (2005)), consistent with our assumption that φ and λ are positively correlated.

Table 14 reports the correlation coefficient between $\hat{\epsilon}_j$ and $\hat{u}_j$ as well as the 95% confidence intervals. We also report the Spearman's ρ in the second row. We find that the correlation coefficient of $\hat{\epsilon}_j$ and $\hat{u}_j$ is 0.086, and statistically different from 0 at the 5% level. We also find that the Spearman's ρ is 0.072, and statistically different from zero at the 5% level. These results confirm that unobservable borrower characteristics that increase default are also likely to increase the reserve rate. Figure 5 plots the estimates of a nonparametric mean regression of $\hat{\epsilon}_j$ on $\hat{u}_j$. The figure shows that the mean of $\hat{\epsilon}_j$ is increasing in $\hat{u}_j$.

3.7 Field Experiment

As we discuss in the main text, we conduct a field experiment on Amazon's Mechanical Turk to obtain more direct evidence for the idea that the reserve rate serves as a signal. In the field experiment, we ask subjects to evaluate the creditworthiness of the borrowers on a

Table 13: Regression Results for Ever Default and Reserve Rate

	(1) Ever default	(2) Reserve Rate
Amount ($1000)	0.013***	4.90E-03***
	(1.37E-03)	(1.58E-04)
Debt/income	0.045*	0.026***
	(0.025)	(2.91E-03)
Home-ownership	0.074***	2.36E-03
	(0.016)	(1.89E-03)
Grade		
AA	-0.211***	-0.128***
	(0.025)	(2.88E-03)
A	-0.132***	-0.095***
	(0.021)	(2.48E-03)
B	-0.040**	-0.042***
	(0.018)	(2.11E-03)
Amount delinquent	2.40E-03	8.96E-04***
	(2.56E-03)	(2.97E-04)
Bank card utilization	-0.010	3.86E-03
	(0.025)	(2.90E-03)
Current credit line	-4.21E-03	-6.53E-04
	(4.59E-03)	(5.32E-04)
Current delinquencies	6.97E-03	0.013***
	(9.04E-03)	(1.05E-03)
Delinquencies last 7 years	5.88E-04	1.44E-04
	(9.47E-04)	(1.10E-04)
Inquiries last 6 months	0.024***	6.06E-03***
	(3.09E-03)	(3.59E-04)
Length status	2.38E-04	1.98E-05
	(2.36E-04)	(2.73E-05))
Open credit lines	3.88E-03	1.39E-03**
	(4.74E-03)	(5.49E-04)
Public records 10 years	0.011	8.58E-04
	(0.012)	(1.42E-03)
Public records 12 months	0.021	8.49E-03
	(0.051)	(5.87E-03)
Total credit lines	5.71E-04	1.91E-04**
	(7.36E-04)	(8.53E-05)
Word count: description	-2.16E-05**	6.33E-06***
	(8.71E-06)	(1.01E-06)
Word count: title	-1.16E-04	1.53E-04**
	(5.47E-04)	(6.34E-05)
Constant	0.097	0.148***
	(0.059)	(6.82E-03)
Occupation/income	X	X
Time dummies	X	X
Observations	3,737	3,737

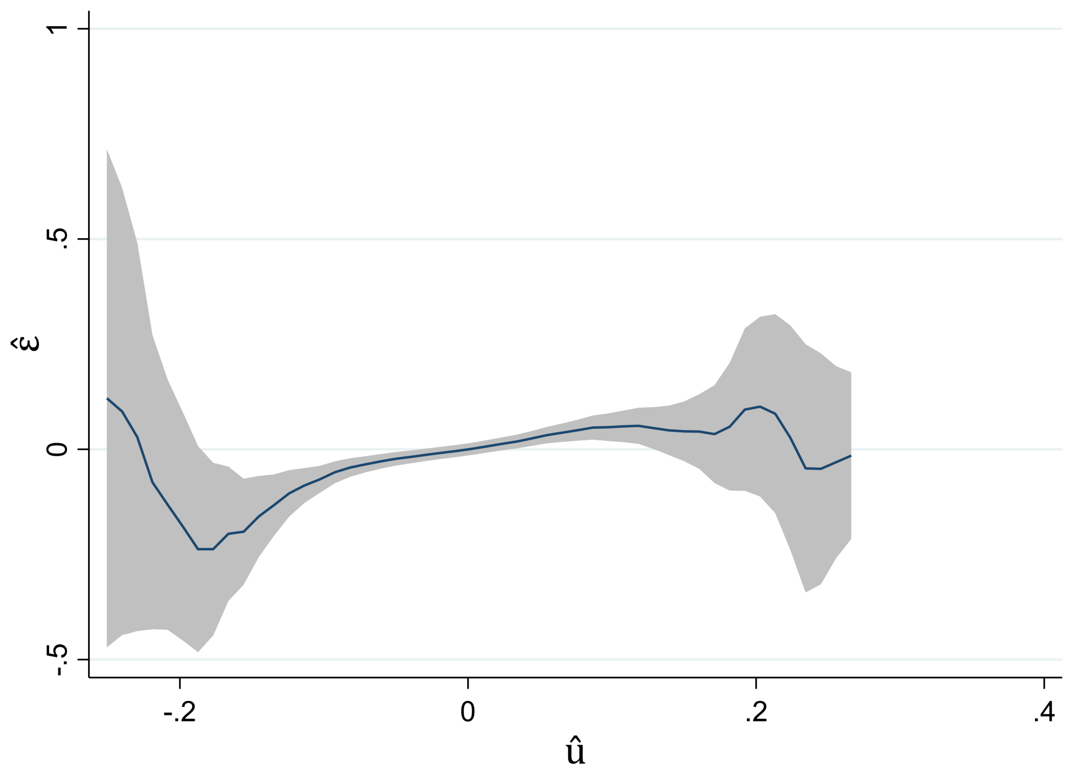
Note: The first column reports the estimated coefficients of the linear probability model (expression (2)). The second column reports the estimated coefficients of the reserve rate regression (expression (3)). *, ** and *** respectively denote significance at the 10%, 5%, and 1% levels.

Table 14: Correlation of $\hat{\epsilon}_j$ and $\hat{u}_j$.

		95% Conf. Interval
$Corr(\hat{\epsilon}_j, \hat{u}_j)$	0.086***	[0.054, 0.118]
Spearman's ρ	0.072***	[0.040, 0.104]

Note: *** denotes significance at the 1% level.

Figure 5: Regression of $\hat{\epsilon}$ on $\hat{u}$.



Note: The figure plots the nonparametric estimate of $\mathbb{E}[\hat{\epsilon}|\hat{u}]$.

scale of 1 (least creditworthy) to 7 (most creditworthy) by showing listings drawn randomly from our sample. 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 as we describe below, we randomize this aspect of the listing. In other words, we show, to some experimental subjects, a listing with a reserve rate that is higher than that in the actual listing while we show to others the same listing but with a lower reserve rate. This experimental design allows us to keep any listing characteristics (other than the reserve rate) fixed.

In this survey, we randomize the reserve rates of listing j that we show to respondent k as follows:

$$s_{jk} = s_j + d_{jk} \times std(s_j),$$

where s_{jk} is the reserve rate that we show to respondent k, s_j is the reserve rate of listing j observed in the data, $d_{jk} \in \{2, 1, 0, -1, -2\}$ is a random variable, and $std(s_j)$ is the standard deviation of the reserve rates among listing in a given credit grade. When d_{jk} is negative (positive), we show a reserve rate that is lower (higher) than the actual one.[8] When d_{jk} is 0, the respondents are shown the same reserve rate as the one in the data.

In order to run the field experiment, we first create a random subset of the listings ($N = 6,156$) that we pre-format. This corresponds to about 17% of the full sample (There are a total of 35,241 listings in the full sample).[9] The subjects of the experiment are shown listings that are drawn randomly from this subset. The instruction to the respondent, an example of a listing we show and the question we ask the respondent are shown on the next page.

Table 15 shows the summary statistics of the listings and the loans used in the survey. We report the sample statistics of the sample used for the experiment and that of the full sample side by side see for easy comparison. The summary statistics of the listings we use in the experiment are quite close to that of the full sample. The summary statistics of the evaluations are reported in Table 5 of the main text.

We posted the survey on Mechanical Turk during December 8th, 2020 and December 15th, 2020, and during July 26th and July 28th, 2021. Each survey respondent was then shown 25 listings randomly from the subsample of listings in sequence and was paid $2 for participation. There are a total of 900 respondents. Among the 900 respondents, 46.6% are female, more than 78% have at least a bachelors degree, and the average age is between 30-39.

[8] We drop the samples if $s_{jk} < 0.05$ as those interest rates seem unrealistically low.

[9] We do not use the full sample because we need to obtain multiple responses for a given listing. If we use the full sample, we risk the possibility that there are too few observations with multiple responses.

Survey Instruction

The purpose of this study is to understand what we infer about other people from text information people post online. In particular, we are interested in text information posted on an online peer-to-peer lending website. We want to know your impression of these people from the information they provided.

Example of a Survey Question

Listing Summary

Debt Consolidation
- Listing #179262

$10,000.00 @ **8.85%***

* The rate shown includes a servicing fee of 1.00% because this listing was created prior to October 15, 2008.

Bid Now | 0% funded | 0 bids | Ended

Listing expired

Borrower rate: **8.85% (10.24% APR)**

Term: **3 years**

Monthly payment: **$317.30**

Servicing fee: **1.00%**

Credit Grade	AA	A	B	C	D	E	HR
Score	760-900	720-759	680-719	640-679	600-639	560-599	520-559

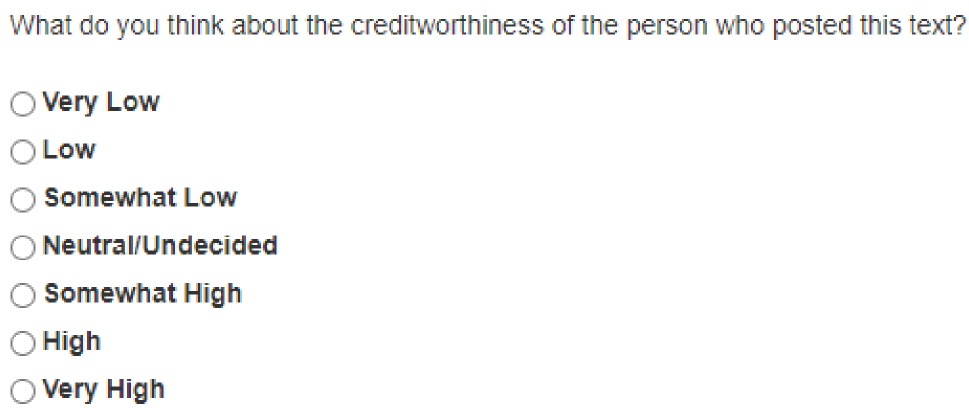
Description

I've come to prosper.com to look for a loan to consolidate my existing credit card debit into a single loan in order to avoid high interest charges that my cards carry. About 5 years ago while in college like many students, I accumulated some debt. Thanks for your consideration.

Here is an example of a listing created by a potential borrower on the lending website. Note that the listing contains information about how much the borrower is asking, and at what interest rate. In this example, the borrower is requesting $1,000 at 8.85 percent (or less). The borrower rate is the maximum interest rate that the borrower is willing to accept. The purpose of the loan is for debt consolidation. The table below shows the range of the credit score for each credit grade. This borrower's credit grade is C, so his/her credit score is between 640-679.

Note that the borrower posts a description about himself/herself. We will be showing you a series of credit information and text information that various borrowers posted on the listing. We want to know your impression of the borrowers' creditworthiness from the information they provided.

What do you think about the creditworthiness of the person who posted this text that you see on the screen?


What do you think about the creditworthiness of the person who posted this text?

- Very Low
- Low
- Somewhat Low
- Neutral/Undecided
- Somewhat High
- High
- Very High

[The end of the instruction.]

In Table 6 of the main text (Section 3), we report the coefficient estimates of the following regressions:

$$\begin{aligned} Eval_{jk} &= \beta_s s_{jk} + \alpha_j + \epsilon_{jk} \\ Eval_{jk} &= \beta_s s_{jk} + \alpha_j + \alpha_k + \epsilon_{jk}, \end{aligned}$$

where $Eval_{jk}$ is the evaluation of listing j by respondent k, s_{jk} is the reserve rate of listing j shown to respondent k and α_j and α_k are listing and respondent fixed effects. Respondent fixed effects control for heterogeneity in the respondent's average evaluation. We use the data from Grade AA to Grade C listings for estimation.

In Table 16, we report the estimated coefficients of the same regressions using the data from Grade AA to Grade HR listings. The estimated coefficients are similar to the results in the main text.

4 Additional Discussion of the Structural Model

In this section, we discuss in greater detail some of the modeling choices regarding our structural model. First, we discuss our decision to take the requested loan amount as given in Section 4.1. We then discuss additional modeling choices for the borrower in Section 4.2. Section 4.3 discusses the assumption that lenders do not take into account the probability of being pivotal. Finally, Section 4.4 discusses additional equilibrium conditions needed to support a partial pooling equilibrium in which a subset of the borrowers pool at $s = 0.36$.

4.1 Signaling Through the Loan Amount

In the main text, we focus on the borrower's choice of the reserve interest rate. Another important variable that the borrower needs to optimize over is the size of the loan. While we do not explicitly model the choice of the loan size, our identification and estimation allow for the possibility that the loan amount is chosen endogenously by the borrower, and that it may also be a signal of the borrower's type. We discuss how our identification and estimation allow for this possibility below.

Consider a slightly richer model than the one in the main text in which the borrower explicitly makes an amount choice a as well as a reserve price choice s as follows:

$$\max_{\substack{s\leq 0.36 \\ a\in[1\text{K},25\text{K}]}} V_0(s,a,\varphi) = \max_{\substack{s\leq 0.36 \\ a\in[1\text{K},25\text{K}]}} \left[\Pr(s,a)\int V_1(r,a,\varphi) f(r|s,a)dr + (1-\Pr(s,a))\lambda(\varphi)\right], \quad (4)$$

Table 15: Descriptive Statistics–Second Survey vs Original

	Amount Requested		Reserve Rate		Debt/ Income		Home Owner		Bid Count		Fund Pr		Contract Rate		Obs	
Grade	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)		(1)	(2)
							Listing Characteristics									
AA	13065	13,145	12.2	13.2	0.36	0.36	0.82	0.81	158.6	153.2	52.5	53.4			282	1420
	(8834)	(8,343)	(4.5)	(4.7)	(0.96)	(0.98)	(0.39)	(0.39)	(202.1)	(178.6)						
A	13170	12,396	16.3	16.5	0.32	0.38	0.66	0.61	107.9	103.3	38.1	40.9			336	1850
	(8415)	(7,882)	(6.5)	(6.7)	(0.25)	(0.67)	(0.47)	(0.49)	(155.2)	(139.9)						
B	10321	10,622	19.3	21.1	0.38	0.39	0.58	0.59	78.3	74.4	33.5	33.4			517	3068
	(6449)	(6,096)	(7.4)	(7.5)	(0.69)	(0.65)	(0.49)	(0.49)	(114.8)	(104.0)						
C	8449	7,622	22.5	24.6	0.36	0.37	0.52	0.56	30.5	36.2	19.4	24.7			833	5203
	(5708)	(5,158)	(8.7)	(7.8)	(0.51)	(0.62)	(0.5)	(0.50)	(55.4)	(59.2)						
D	6873	6,368	26.4	28.7	0.39	0.39	0.36	0.37	18.2	18.2	13.7	15.5			1074	6581
	(5117)	(4,691)	(8.9)	(7.5)	(0.72)	(0.71)	(0.48)	(0.48)	(40.9)	(39.0)						
E	5535	4,783	26.5	31.0	0.37	0.36	0.34	0.33	3.42	3.99	4.3	6.8			905	5757
	(5680)	(4,868)	(9.9)	(7.3)	(0.72)	(0.68)	(0.47)	(0.47)	(10.9)	(11.0)						
HR	4776	4,351	27.9	31.5	0.38	0.31	0.23	0.22	2.1	1.9	2.5	3.0			2209	11362
	(5170)	(4,599)	(9.8)	(6.9)	(0.92)	(0.64)	(0.42)	(0.42)	(8.4)	(5.7)						
All	7054.7	6,604	24.6	27.4	0.37	0.35	0.39	0.39	28.3	28.0	13.9	15.8			6156	35241
	(6424.0)	(5,938)	(10.0)	(8.9)	(0.77)	0.68	(0.49)	(0.49)	(82.0)	(74.4)						
							Loan Characteristics									
AA	9884	9,710	12.2	13.1	0.18	0.21	0.84	0.80	273.8	258.1			9.2	9.6	148	755
	(8073)	(7,384)	(4.2)	(3.3)	(0.14)	(0.39)	(0.37)	(0.39)	(221.3)	(186.2)			(3.3)	(4.6)		
A	8820	8,723	15.4	16.5	0.22	0.23	0.62	0.55	238.3	218.1			12.1	12.7	128	755
	(7036)	6,626	(5.1)	6.0	(9.9)	(0.14)	(0.49)	(0.14)	(185.7)	(155.4)			(4.3)	(4.5)		
B	6808	7,347	20.5	21.6	0.23	0.27	0.60	0.56	199.6	181.2			15.4	16.4	173	1023
	(4585)	(4,858)	(5.4)	(6.3)	(0.13)	(0.34)	(0.49)	(34)	(125.0)	(113.9)			(3.7)	(4.6)		
C	4682	4,687	22.7	24.7	0.25	0.25	0.49	0.48	118.0	112.6			16.8	18.1	162	1023
	(2722)	(2,998)	(6.5)	6.4	(0.12)	(0.21)	(0.50)	(0.21)	(71.3)	(72.2)			(5.7)	(6.2)		
D	3577	3,578	26.6	28.0	0.25	0.24	0.22	0.26	99.3	86.7			20.0	21	147	1022
	(2235)	(2,380)	(6.9)	(6.4)	(0.17)	(0.17)	(0.41)	(0.17)	(63.3)	(60.8)			(6.5)	(6.6)		
E	2029	1,890	32.4	33.9	0.24	0.22	0.21	0.26	46.7	36.4			28.1	29.1	39	392
	(1018)	(1,187)	(2.2)	(2.8)	(0.14)	(0.22)	(0.41)	(0.22)	(25.8)	(20.3)			(4.3)	(5.7)		
HR	1923	1,690	34.5	33.9	0.16	0.20	0.14	0.17	42.5	27.0			30.4	30	56	339
	(998)	(1,288)	(1.9)	(3.6)	(0.15)	(0.44)	(0.35)	(0.44)	(31.3)	(19.0)			(5.8)	5.7		
All	6144	5,821	21.2	23.3	0.22	0.24	0.51	0.47	168.2	144.0			16.5	17.9	853	5571
	(5685)	(5,285)	(8.4)	(8.6)	(0.14)	(0.28)	(0.5)	(0.28)	(156.1)	(132.4)			(7.6)	(7.9)		

Note: The table reports the descriptive statistics of the data used for the online survey and the full sample. Columns with (1) correspond to the data used for the survey and columns with (2) to the full sample.

Table 16: Evaluation of Borrower Creditworthiness and Reserve Rate

	(1) Evaluation	(2) Evaluation
Reserve Rate (%)	-0.645^{***}	-0.696^{***}
	(0.141)	(0.125)
Listing FE	X	X
Respondent FE		X
Observations	24,706	24,706

Note: The table reports the estimation results of the OLS regression, where the rating is regressed on the reserve rate and other characteristics. The star mark *** denotes significance at the 1% level. Standard errors are reported in parenthesis.

where $\Pr(s, a)$ is the probability that the borrower is able to obtain a loan, and $f(r|s, a)$ is the conditional distribution of the interest rate. By allowing $\Pr(s, a)$ and $f(r|s, a)$ to depend on s and a, we are explicitly allowing for the possibility that both the reserve rate and the loan amount to signal to the lenders the type of the borrower.

Taking the first order condition of (4), we obtain the following expressions,

$$\frac{\partial \Pr(s,a)}{\partial a}\left(\int V_1(r,a,\varphi) f(r|s,a)dr - \lambda(\varphi)\right) \\ + \Pr(s,a)\int V_1(r,a,\varphi)\frac{\partial f(r|s,a)}{\partial a}dr + \Pr(s,a)\int \frac{V_1(r,a,\varphi)}{\partial a} f(r|s,a)dr = 0, \text{ and} \tag{5}$$

$$\frac{\partial \Pr(s,a)}{\partial s}\left(\int V_1(r,a,\varphi) f(r|s,a)dr - \lambda(\varphi)\right) + \Pr(s,a)\int V_1(r,a,\varphi)\frac{\partial f(r|s,a)}{\partial s}dr = 0. \tag{6}$$

The first expression corresponds to the first-order condition with respect to a, and the second expression corresponds to the first-order condition with respect to s.[10] Note that the second expression is exactly the same as the expression we use in the main text to estimate the

[10]When endogenizing loan size, the specification of the model should be modified to include a term that captures the utility gain from meeting a desired level of loan size. Otherwise, (5) will almost always be negative. The following specification is one such possibility:

$$\max_{\substack{s_i \le 0.36 \\ a\in[1\text{K},25\text{K}]}} \left[\Pr(s,a)\int (V_1(r,a,\varphi) - \gamma \mathbf{1}_{\{a<a^*\}})) f(r|s,a)dr + (1-\Pr(s,a))\lambda(\varphi)\right]. \tag{7}$$

The only difference between this specification and that in the main text is the inclusion of the term $\gamma\mathbf{1}_{\{a<a^*\}}$, a penalty term that the borrower incurs if the loan amount is less than the desired amount, a^*. γ is a parameter that determines how large the penalty is and $\mathbf{1}_{\{a<a^*\}}$ is an indicator function for the event that the loan amount is less than the (borrower-specific) desired level a^*. Note that if the borrower chooses $a = a^*$, the

borrower's model.

There are two reasons why our estimates of the primitives – which are obtained only using (6) and *not* (5) – are consistent. The first reason is that estimating the model parameters using only a subset of the equilibrium restrictions (i.e., (6)) does not affect the consistency of the estimates (it may affect efficiency, however).[11] The second reason is that we make no functional form restrictions on how the borrower's type is related to a or other primitives of the model. Note that Section 5 in the main text shows nonparametric identification of $F_{\varphi|X}$ where X includes the loan amount. The purpose of showing that $F_{\varphi|X}$ is nonparametrically identified is precisely to allow a and other primitives to be informative about the type of the borrower in an arbitrary manner. To the extent that the requested amount has a signaling aspect, we will be able to capture it directly when estimating the distribution of φ as a function of a.

4.2 Alternative Model of Borrower's Type

In our specification, we model the private information of the borrower φ as default cost. A natural alternative would be to model the private information of the borrowers as income/asset of the borrower. In this section, we present a model in which φ is modeled as unobserved income/assets of the borrower. We then show that this model is isomorphic to the model in the paper in which φ has the interpretation of default cost.

Consider the following alternative model of the borrower:

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

borrower's problem (7) reduces to the following:

$$\max_{s\leq 0.36} \left[\Pr(s, a^*) \int (V_1(r, a^*, \varphi)) f(r|s, a^*) dr + (1 - \Pr(s, a^*))\lambda(\varphi)\right], \quad (8)$$

which is exactly equal to the model that we have in the main text (with a exogenously given). Hence model (7) is obervationally equivalent to the main specification while allowing for a channel that deters borrowers from reducing a. The corresponding first-order condition for optimality for this specification is for the left derivative (with respect to a) to be non-negative and the right derivative to be non-positive (as opposed to the derivative being equal to zero as in (5).

[11] We emphasize that when the borrowers are making the reserve rate choice, the borrowers are shown a tutorial informing them that a higher reserve rate increases the funding probability. Hence, it seems very likely that the borrowers are aware of the trade-off between high funding probability and high interest rates when choosing the reserve rate. We are less confident, however, that borrowers are fully aware of the effect of the amount on the contract interest rate and take this into consideration when choosing the amount. To the extent that some borrowers are not completely rational in this regard, using the FOC for the amount choice may lead to inconsistent estimates. Note that using only the FOC of the reserve rate choice would yield consistent parameter estimates even in this scenario. We feel that using the FOC for the reserve rate to recover default costs is more reliable than using the FOC of the amount choice in our setting.

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

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

Recall that the model in the paper is as follows:

$$\text{full repayment} \Leftrightarrow -(r \times x_{amt}) + \varepsilon_T \geq D(\varphi)$$

Redefining ε_T and D in the equation as $\varepsilon_T = -\tilde{u}_T^{-1}(-\tilde{\varepsilon}_T) + \omega_T$, and $D(\tilde{\varphi}) = -\tilde{\varphi}$, we find that the two specifications are observationally equivalent. Regardless of the source of unobserved heterogeneity that affects the propensity to default – whether it be default cost, income, or some combination of the two – the resulting specification will be similar and the difference will be only in the interpretation of φ.

4.3 Non-strategic Bidding of Lenders

The auction mechanism used in Prosper.com is not a Vickrey auction. Under the auction format used by Prosper, the equilibrium interest rate is determined by the bid of the marginal winner, and not, as is the case of a Vickrey auction, by the bid of the marginal loser. This means that a bidder has an incentive to increase her bid to the extent that the probability of being the marginal bidder is positive. However, if the probability of being marginal is very small, it is nearly optimal for the lender to bid the interest rate that makes the lender indifferent between funding the loan and not funding the loan. Moreover, the gain from taking into account the probability of being marginal is bounded by the difference between the marginal losing bid and the marginal winning bid.

Below, we provide evidence that the probability of being marginal is in fact very small and explain why the strategy described in Proposition 3 in the main text is approximately optimal.

In order to make our arguments concrete, consider the problem faced by bidder j. Let r^* and r^{**} ($r^* \leq r^{**}$) respectively denote the bid of the marginal winner and the bid of the marginal loser if j decides to put in a sufficiently low bid (e.g., interest of 0%). Then, the relationship between j's bid, r_j, and the outcome of the auction is as follows:

$$\begin{cases} j \text{ wins, contract interest rate is } r^* & \text{if } r_j \leq r^* \\ j \text{ wins, contract interest rate is } r_j & \text{if } r^* \leq r_j \leq r^{**} \\ j \text{ does not win} & \text{if } r_j \geq r^{**} \end{cases}$$

The problem of the bidder is as follows:

$$\begin{aligned} & \max_{r_j} \overbrace{\mathbf{E}[1_{\{r_j \leq r^*\}}(U_j^L(Z(r^*))]}^{j \text{ is infra-marginal}} + \overbrace{\mathbf{E}[1_{\{r_j \in [r^*, r^{**}]\}}(U_j^L(Z(r_j))]}^{j \text{ is marginal}} \\ = & \max_{r_j} \int_{r_j}^{\overline{s}} U_j^L(Z(r^*)) f_{r^*}(r^*) dr^* + \Pr(r^* < r_j \leq r^{**}) U_j^L(Z(r_j)), \end{aligned}$$

where $Z(r)$ is a random return from investing in the listing at rate r, $\overline{s}$ is the borrower's reserve rate, and $f_{r^*}(r^*)$ is the pdf of r^*. The first part of the objective function corresponds to the event in which r_j is infra-marginal and the second component corresponds to the event in which r_j is marginal. Differentiating the objective function with respect to r_j, we obtain the following expression:

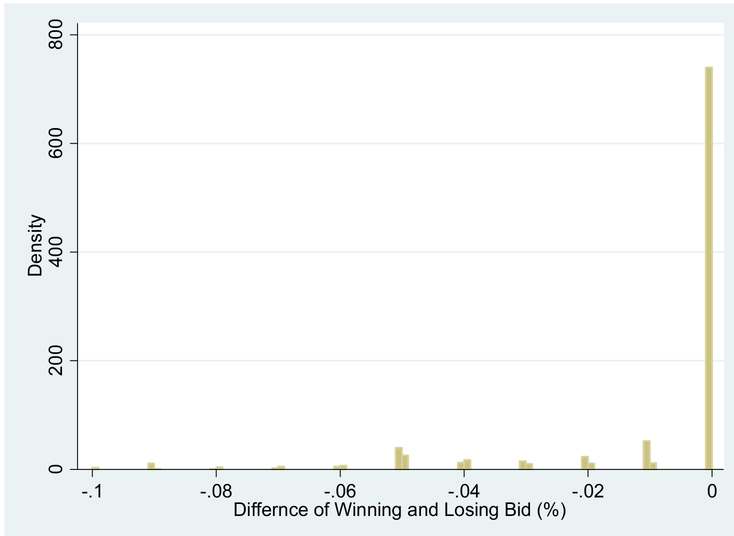
$$-U_j^L(Z(r_j)) f_{r^*}(r_j) + \frac{d}{dr_j} \Pr(r^* < r_j \leq r^{**}) U_j^L(Z(r_j)) + \Pr(r^* < r_j \leq r^{**}) \frac{d}{dr_j} U_j^L(Z(r_j)). \quad (9)$$

Consider first, the case in which the second and the third terms of expression (9) are nonexistent. This would be the case in a Vickrey auction. Note that the first term of expression (9) has the opposite sign as $U_j^L(Z(r_j))$ because $f_{r^*}(r_j) > 0$. The first term is equal to zero when $U_j^L(Z(r_j))$ is equal to the outside option (i.e., $U_j^L(Z(r_j)) = 0$), it is positive when $U_j^L(Z(r_j))$ is lower than the outside option, and it is negative when $U_j^L(Z(r_j))$ is higher than the outside option. This implies that, if we can ignore the second and third terms of expression (9), the objective function is maximized when r_j is set so as to equate the benefit of lending to the outside option.

We now argue that in the setting we study, the second and third terms of expression (9) are close to zero, and hence the root of (9) is approximately equal to the reserve interest. Intuitively, this is because the difference between the marginal losing bid and the marginal winning bid is typically very close together. Figure 6 plots the distribution of the difference between the two.[12] The figure shows that the difference is less than 0.05% for the vast majority of funded listings, and always less than 0.1%. When $\Delta = r^{**} - r^*$ is small, the

[12] We observe all of the winning bids. However, among the losing bids, we only observe the marginal losing bid and none of the other losing bids. The figure plots the difference between the marginal losing bid and the marginal winning bid. For lenders with a small bid amount, $r^{**} - r^*$ is exactly equal to what the figure plots. However, for a small set of lenders whose bid amount is large, $r^{**} - r^*$ will be larger than what the figure plots.

Figure 6: The Marginal Winning Bid and the Marginal Losing Bid



Note: The graph is the histogram of the difference in the winning bid and the losing bid (i.e., losing bid minus winning bid) for funded listings.

probability of being pivotal (i.e., $\{s_j \in [r^*, r^{**}]\}$) is small; hence bidder behavior can be approximated by behavior in a Vickrey auction.

We now show that the second and third terms of expression (9) are close to zero. First, focus on the second term of expression (9). Let F_{r^*} and f_{r^*} denote the distribution of r^*. Let $F_\Delta(\cdot|r^*)$ and $f_\Delta(\cdot|r^*)$ denote the conditional distribution of Δ $(= r^{**} - r^*)$ given r^*. Then,

$$\Pr(r^* < r_j \leq r^{**}) = \int_{r^*=0}^{r_j} [1 - F_\Delta(r_j - r^*|r^*)] f_{r^*}(r^*) dr^*.$$

Hence,

$$\begin{aligned} \frac{d}{dr_j} \Pr(r^* &< r_j \leq r^{**}) \qquad (10) \\ &= [1 - F_\Delta(0|r_j)] f_{r^*}(r_j) - \int_{r^*=0}^{r_j} f_\Delta(r_j - r^*|r^*) f_{r^*}(r^*) dr^* \\ &= f_{r^*}(s_j) - \int_{r^*=0}^{s_j} f_\Delta(s_j - r^*|r^*) f_{r^*}(r^*) dr^* \qquad (11) \end{aligned}$$

Now, note that $f_\Delta(\cdot|r^*)$ takes positive density only in a small interval $[0, \bar{\varepsilon}]$ as we showed in Figure 6. Hence, the second term of expression (10) can be approximated by $f_{r^*}(s_j)$ as follows:

$$
\begin{aligned}
& \int_{r^*=0}^{s_j} f_\Delta(s_j - r^*|r^*) f_{r^*}(r^*) dr^* \\
\approx \;& \int_{r^*=s_j-\bar{\varepsilon}}^{s_j} f_\Delta(s_j - r^*|r^*) f_{r^*}(r^*) dr^* && (\text{b/c } f_\Delta(\Delta|r^*) \approx 0 \text{ for } \Delta > \bar{\varepsilon}) \\
\approx \;& \int_{r^*=s_j-\bar{\varepsilon}}^{s_j} f_\Delta(s_j - r^*|s_j) f_{r^*}(s_j) dr^* && (\text{b/c } f_\Delta(s_j - r^*|r^*) \approx f_\Delta(s_j - r^*|s_j)) \\
\approx \;& f_{r^*}(s_j) && (\text{b/c } \int_{r^*=s_j-\bar{\varepsilon}}^{s_j} f_\Delta(s_j - r^*|s_j) dr^* = 1).
\end{aligned}
$$

In the second line of this expression, we have changed the region of integration from $[0, s_j]$ to $[s_j - \bar{\varepsilon}, s_j]$ because $f_\Delta(\Delta|r^*)$ is zero for $\Delta > \bar{\varepsilon}$. In the third line, we have used the fact that $f_{r^*}(r^*)$ can be approximated by $f_{r^*}(s_j)$ and that $f_\Delta(\cdot|r^*)$ can be approximated by $f_\Delta(\cdot|s_j)$ in $[s_j - \bar{\varepsilon}, s_j]$. The fourth line uses the fact that $\int_{r^*=s_j-\bar{\varepsilon}}^{s_j} f_\Delta(s_j - r^*|s_j) dr^*$ is approximately 1. Hence, we have shown that when $f_\Delta(\cdot|r^*)$ has positive support in a small neighborhood of 0, the second term of expression (9) is approximately zero.

The third term of expression (9) is also approximately zero if $\Pr(r^* < s_j \leq r^{**})$ is close to zero. This implies that the left-hand side of the first-order condition (9) is approximately $-U_j^L(Z(r_j)) f_{r^*}(r_j)$ which implies that the optimal interest is given by the interest rate r at which $U_j^L(Z(r)) = 0$.

4.4 Additional Condition for Partially Pooling Equilibria at 36%

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

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

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

5 Details of the Estimation

We denote by $\theta=\{\theta_b, \theta_\pi, \theta_r, \theta_p\}$ the parameters of the borrower's model. The borrower's period t utility function is specified as $u_t(r, X; \theta_b) = -r \times x_{amt} + d_t$, where r is the contract interest, x_{amt} is the loan size and d_t are time dummies that are allowed to depend on X. In particular, we estimate separate dummies by credit grade (AA, A, B, C) and latent type ($X^u = H, L$). The outside option is specified as a linear function of φ as $\lambda(\varphi, X; \theta_b) = \lambda\varphi$, where λ is a slope coefficient that is allowed to depend on credit grade and latent type. The distribution of ε, $F(\cdot|X; \theta_b)$, is specified as a type-I extreme value distribution.

We denote the distribution of the latent types as $\pi(s, X^o; \theta_\pi)$, the distribution of r as $f(r|s, X; \theta_r)$ and the funding probability as $\Pr(\text{fund}|s, X; \theta_p)$. We estimate these functions separately by credit grade, home-ownership and latent type. We specify $\pi(s, X^o; \theta_\pi)$ as a logistic function with second-order polynomials of s, requested amount and debt-to-income ratio. Similarly, we specify $f(r|s, X; \theta_r)$ as a second-order Hermite series in s, requested amount and debt-to-income ratio. We specify the funding probability as a Probit with second-order polynomials.

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 solve for the borrower's private type, $\varphi(s, X^o, L; \theta)$ and $\varphi(s, X^o, H; \theta)$ using the first order condition.

Consider the likelihood associated with each listing. If a listing is not funded, then the likelihood l_i is simply the probability of not being funded, i.e., $l_i(X^u, \theta) = 1 - \Pr(\text{fund}|s, X; \theta_p)$. If a listing is funded, we multiply the probability of being funded, the probability associated with the realization of r, and the likelihood associated with a given sequence of repayment decisions. Because the repayment decisions are modeled as outcomes of a sequence of binary threshold crossing models, it is easy to compute the likelihood. Note that the intercept of the binary threshold crossing model is either $\varphi(L)$ or $\varphi(H)$, the privately-known default costs that solve the first-order condition.

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 and θ^0:

$$\Pr(H|s, X^o, \text{outcome of } i; \theta^0) = \frac{\pi(s, X^o, \theta_\pi^0) l_i(H; \theta^0)}{\pi(s, X^o, \theta_\pi^0) l_i(H; \theta^0) + (1 - \pi(s, X^o; \theta_\pi^0)) l_i(L; \theta^0)}. \quad (12)$$

Now consider the likelihood assuming that the distribution of the latent type of the borrower is given by expression (12). 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 (12). We then take the maximizer, say θ^1, as the new initial parameter and repeat the procedure until convergence.

6 Additional Parameter Estimates of the Structural Model

6.1 Unreported Results on the Borrower Model

In the main text, we do not report the estimates of the time dummies d_t in the borrower's repayment model to save space. We report those estimates in this section.

Table 17 reports our estimates of the time dummies in units of $1,000. Recall that we have 12 time dummies, each of which corresponds to a 3-month period. We normalize the first one to zero. We find that the time dummies are generally negative in earlier periods and positive in later periods. Our estimates reflect the fact that, conditional on default, borrowers tend to default early in the repayment process (but not necessarily immediately after taking out the loan – note that the first time dummy is 0). Borrowers rarely default after having paid off most of their debt.

6.2 Unreported Results on the Lender Model

In the main text, we do not report the estimates of the time dummies in the lender's outside option to save space. Table 18 report the estimates. The unit is $1.

We find that, on average, the time dummies are close to zero. For credit grades B and C, some of the time dummies are larger, but even the time dummy with the largest magnitude is around -$25.

7 Computational Details of the Counterfactuals

7.1 Computation of the Credit Supply Curve

In this section, we describe the procedure for computing the credit supply curve, i.e., the total amount that lenders are willing to lend as a function of the interest rate. We discuss computation for the case of pooling and no asymmetric information. The supply curve under signaling can be obtained by truncating the one for no asymmetric information at the

Table 17: Parameter Estimates of the Borrower's Model–Time Dummies

	AA		A		B		C	
	High	Low	High	Low	High	Low	High	Low
d_1	-0.78	-0.76	-0.75	-0.74	-0.80	-0.79	-0.84	-0.95
	(0.14)	(0.14)	(0.15)	(0.15)	(0.15)	(0.14)	(0.17)	(0.20)
d_2	-0.61	-0.62	-0.56	-0.56	-0.64	-0.61	-0.67	-0.67
	(0.11)	(0.12)	(0.10)	(0.11)	(0.13)	(0.12)	(0.14)	(0.12)
d_3	-0.42	-0.42	-0.39	-0.37	-0.43	-0.43	-0.45	-0.48
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)
d_4	-0.24	-0.23	-0.23	-0.22	-0.26	-0.24	-0.29	-0.36
	(0.05)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.07)
d_5	-0.07	-0.06	-0.06	-0.07	-0.04	-0.05	-0.09	-0.06
	(0.01)	(0.01)	(0.01)	(0.11)	(0.01)	(0.02)	(0.02)	(0.01)
d_6	0.14	0.15	0.14	0.12	0.13	0.12	0.11	0.12
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
d_7	0.32	0.32	0.30	0.28	0.28	0.30	0.29	0.33
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
d_8	0.48	0.49	0.46	0.47	0.48	0.49	0.51	0.50
	(0.09)	(0.09)	(0.08)	(0.08)	(0.10)	(0.10)	(0.09)	(0.10)
d_9	0.70	0.68	0.64	0.63	0.65	0.66	0.71	0.70
	(0.13)	(0.14)	(0.12)	(0.13)	(0.13)	(0.12)	(0.14)	(0.14)
d_{10}	0.86	0.86	0.81	0.81	0.84	0.83	0.89	0.81
	(0.16)	(0.15)	(0.15)	(0.16)	(0.18)	(0.18)	(0.17)	(0.15)
d_{11}	1.05	1.05	0.99	0.99	1.02	1.02	1.07	0.91
	(0.22)	(0.23)	(0.20)	(0.19)	(0.20)	(0.20)	(0.22)	(0.19)

Note: Term of the loan is fixed at 36 months for all loans. Each dummy corresponds to a 3-month period. The dummy corresponding to the first 3 months is normalized to 0, leaving 11 dummies. The units are in \$1000. Standard errors are obtained by bootstrap (150 times) and they are reported in parentheses.

reserve rate. Note that the credit supply curves can be computed for each value of $x^o = (x_{amt}, x_{ho}, x_{dti})$ and x^u, where x^o is the vector of observable characteristics, x_{amt} is the requested amount, x_{ho} is the home-ownership status, x_{dti} is the debt-to-income ratio, and x^u is the latent type. The procedure we outline below is conditional on a particular value of x^o and x^u.

The procedure for computing the supply curve under pooling is as follows.

1. Fix $r \in [0, 0.36]$. Compute the set of borrower types that would prefer obtaining a loan at a contract interest rate r than not, i.e., $\Psi(r) = \{\varphi : V_1(r,\varphi) \geq \lambda(\varphi)\}$. Given that $\frac{\partial}{\partial\varphi}\lambda(\varphi) > 0$ and $\frac{\partial}{\partial\varphi}V_1 < 0$ (See Lemma 1 in Section 9 of the online appendix), we can characterize $\Psi(r)$ with a threshold value of the borrower's type $\overline{\varphi}(r)$ as $\Psi(r) = \{\varphi : \varphi \leq \overline{\varphi}(r)\}$. $\overline{\varphi}(r)$ can be obtained by solving $V_1(r,\varphi) = \lambda(\varphi)$ with respect to φ.

Table 18: Lender's Model–Time Dummies

	AA	A	B	C
d_1^L	-0.59	-0.35	2.25	1.35
	(0.12)	(0.07)	(0.55)	(0.27)
d_2^L	0.72	0.35	0.84	10.98
	(0.14)	(0.08)	(0.22)	(2.17)
d_3^L	-0.69	0.86	-26.27	5.11
	(0.44)	(0.17)	(9.01)	(1.02)
d_4^L	-0.76	-0.50	-3.97	12.97
	(0.15)	(0.10)	(3.27)	(2.6)
d_5^L	-2.04	-0.78	0.71	-25.05
	(0.41)	(0.16)	(0.22)	(14.20)

Note: The units are in \$1. Standard errors are reported in parentheses.

2. Take a draw of φ from $\Psi(r)$ according to the PDF $f_\varphi/F_\varphi(\overline{\varphi}(r))$. The PDF, $f_\varphi/F_\varphi(\overline{\varphi}(r))$, is just the conditional density of φ that is truncated above by $\overline{\varphi}(r)$. In practice, we fit a truncated Normal distribution on $f_\varphi/F_\varphi(\overline{\varphi}(r))$ when we take draws of φ from $f_\varphi/F_\varphi(\overline{\varphi}(r))$. We discuss this point in more detail later. Given a draw of φ and the estimated θ_b, simulate the borrower's repayment decision assuming that the borrower receives the loan with a contract interest rate of r. Repeat this step many times for different draws from $\Psi(r)$ to obtain the mean return and variance, $(\mu(r)\ , \sigma^2(r))$.[13]

3. Simulate the number of potential bidders, $\widetilde{N}$, drawn from F_N. Draw each bidder's risk attitude, $\{A_1, ..., A_{\widetilde{N}}\}$, and the value of the outside option $\{\epsilon_{01}, ..., \epsilon_{0\widetilde{N}}\}$, drawn from F_A and F_{ϵ_0}, respectively. For each potential bidder j, obtain her optimal amount choice $q_j^*(r)$ by solving the following problem:

$$\max\left\{\max_{q_j \in M}\left\{\left[q_j\mu(r)\ - A_j(q_j\sigma(r))^2\right] - c(q_j;\theta_L)\right\}, \epsilon_{0j}\right\},$$

where $(\mu(r)\ , \sigma^2(r))$ is obtained in step 2. The total credit supply, denoted by $Q_1(r)$, is obtained as $Q_1(r) = \sum_{j=1}^{\widetilde{N}} q_j^*(r)$.

4. Repeat step 3 many times and take the average of $Q_1(r)$. That gives a credit supply for r, which we denote as $\overline{Q}_1(r)$.[14]

5. Repeat steps 1 through 4 for each $r \in [0, 0.36]$ to obtain $\overline{Q}_1(r)$ for all $r \in [0, 0.36]$. $\overline{Q}_1(r)$ is the credit supply curve for the pooling case.

[13]In practice, we repeat this step 10,000 times.

[14]We repeat this step 500 times.

The procedure for computing the supply curve under no informational asymmetry is as follows:

1. Take a borrower type φ.

2. Given the type of the borrower we specify in step 1 (φ) and the estimated θ_b, simulate the borrower's repayment decision assuming that the borrower receives the loan with a contract interest rate of r. Repeat this step many times to obtain the mean return and variance, $(\mu(r,\varphi), \sigma^2(r,\varphi))$.[15]

3. Simulate the number of potential bidders, $\widetilde{N}$, drawn from F_N. Draw each bidder's risk attitude, $\{A_1, ..., A_{\widetilde{N}}\}$, and the value of the outside option $\{\epsilon_{01}, ..., \epsilon_{0\widetilde{N}}\}$, drawn from F_A and F_{ϵ_0}, respectively. For each potential bidder j, obtain her optimal amount choice $q_j^*(r,\varphi)$ by solving the following problem:

$$\max\left\{\max_{q_j \in M}\left\{\left[q_j\mu(r,\varphi)\ -A_j(q_j\sigma(r,\varphi))^2\right] - c(q_j;\theta_L)\right\}, \epsilon_{0j}\right\}$$

   where $(\mu(r,\varphi)\ , \sigma^2(r,\varphi))$ is obtained in step 2. The total credit supply, denoted by $Q_2(r,\varphi)$, is obtained as $Q_2(r,\varphi) = \sum_{j=1}^{\widetilde{N}} q_j^*(r,\varphi)$.

4. Repeat step 3 many times and take the average of $Q_2(r,\varphi)$ to obtain $\overline{Q}_2(r,\varphi)$.

5. Repeat steps 1 through 4 for each $r \in [0, indiff(\varphi)]$, where $indiff(\varphi)$ is the interest rate that makes the type φ borrower indifferent between borrowing and not borrowing. $\overline{Q}_2(r,\varphi)$ ($r \in [0, indiff(\varphi)]$) is the credit supply curve for no informational asymmetry. Note that we can solve for $indiff(\varphi)$ as a function of φ and the parameter of borrowers' utility, θ_b.

Finally, we discuss how we fit a truncated Normal distribution on $f(\varphi)/F(\overline{\varphi}(r))$ in step 2 of the pooling case when we take draws of φ. Note that $f(\varphi)/F(\overline{\varphi}(r))$ implicitly depends on the observable listing characteristics, x^o =(x_{amt}, x_{ho}, x_{dti}), and the latent type, x^u, and we need to take this into account when we fit a Normal distribution. The following steps describe the procedure given the latent type:

1. Take observations (s_i, x_i) such that x_i is close to x – In particular, take observations such that $x_{ho,i} = x_{ho}$, $|x_{amt,i} - x_{amt}| < 3000$ and $|x_{dti,i} - x_{dti}| < 0.1$.

[15] In practice, we repeat this step 10,000 times.

(a) For each (s_i, x_i) we find in step 1 with $s_i < 0.36$, calculate the corresponding borrower type, φ_i, by solving the borrower's first-order condition (FOC) in equation (4.3).

(b) For each (s_i, x_i) we find in step 1 with $s_i = 0.36$, we cannot directly calculate the types using equation (4.3) because the first-order condition is not satisfied. Instead, we use the fact that the borrower type corresponding to $s_i = 0.36$ must be worse than any type that submits a reserve rate less than 0.36. Denote the worst type among those that submit a reserve rate less than 0.36 by φ_{min}, $\varphi_{min} = \min_{s<0.36} \varphi_i$. We know that the type of the borrower, φ_i, that corresponds to (s_i, x_i) with $s_i = 0.36$ is less than φ_{min}.

2. We fit a Normal on $f(\varphi)/F(\overline{\varphi}(r))$ that maximizes the likelihood.

7.2 Computation of Welfare

In this section, we describe how we compute the expected welfare under three different market designs presented in Table 10 of the main text.

Note that once we can compute the welfare for any given listing, we can use the empirical distribution of x^o and estimated posterior probability of x^u in order to obtain the expected welfare by integrating the welfare over the empirical distribution. Below, we describe how we compute the welfare for a given listing with x^o, s, and x^u.

For the signaling equilibria, the procedure is as follows:

1. Take a listing. Given the reserve rate, s, we compute the type of the borrower (φ) using the borrower's FOCs. We also discuss this step in more detail at the end of this section.

2. Given the type of the borrower we obtained in step 1 and the estimated parameter θ_b, simulate the borrower's repayment decision, assuming that the borrower receives the loan for some contract interest rate $r \leq s$. Repeat this step many times to obtain the mean return and variance, $(\mu(r, \varphi), \sigma^2(r, \varphi))$.[16]

3. Repeat step 2 for each $r \in [0, s]$. This gives us the mean and variance of the return $(\mu(r, \varphi), \sigma^2(r, \varphi))$ for each realization of $r \leq s$.

4. Simulate the number of potential bidders, $\widetilde{N}$, drawn from F_N. Draw each bidder's risk attitude, $\{A_1, ...A_{\widetilde{N}}\}$, and the value of the outside option $\{\epsilon_{01}, ..., \epsilon_{0\widetilde{N}}\}$, drawn from F_A

[16]In practice, we repeat this step 10,000 times.

and F_{ϵ_0}, respectively. For each potential bidder j and r, obtain her optimal amount choice $q_j^*(r,\varphi)$ by solving the following problem:

$$\max\left\{\max_{q_j\in M}\left\{\left[q_j\mu(r,\varphi)\ -\ A_j(q_j\sigma(r,\varphi))^2\right]-c(q_j;\theta_L)\right\},\epsilon_{0j}\right\}$$

where $(\mu(r,\varphi),\ \sigma^2(r,\varphi))$ is obtained in step 4. The total credit supply is $Q_1(r,\varphi) = \sum_{j=1}^{\widetilde{N}} q_j^*(r,\varphi)$.

5. Take the minimum interest rate r^E for which the credit supply computed in step 4 exceeds the requested amount, i.e., $r^E = \min\{r \leq s|\ Q_1(r,\varphi) \geq x_{amt}\}$, where x_{amt} is the requested amount of the borrower.

6. Given the interest rate r^E we obtained in step 5, we compute the borrower and the lender welfare. If $Q_1(r,\varphi)$ was less than x_{amt} in step 6, the welfare of the lenders and the borrowers are set to 0.

7. Repeat step 4-6 many times and compute the average of the borrower and the lender welfare.[17]

We now discuss how we compute welfare under no signaling (i.e., pooling). The procedure is very similar to the one we discussed for signaling. The only differences are in step 2 and step 3.

2. For each r, we need to obtain the mean return and variance, $(\mu(r,\varphi)$,$\sigma^2(r,\varphi))$. Unlike the case for signaling, we compute the mean and variance by following steps 1 and 2 for computing the credit supply curve under pooling.

3. Repeat step 2 for each $r \in [0, indiff(\varphi)]$, where $indiff(\varphi)$ is the interest rate that makes the type φ borrower indifferent between borrowing and not borrowing. This gives us the mean and variance of the return $(\mu(r,\varphi)$,$\sigma^2(r,\varphi))$ for each realization of $r \leq indiff(\varphi)$.

The procedure for computing welfare under no asymmetric information is also very similar to the one we discussed for signaling. The only difference is in step 3. All other steps are exactly the same as the steps for signaling.

3. Repeat step 2 for each $r \in [0, indiff(\varphi)]$, where $indiff(\varphi)$ is the interest rate that makes the type φ borrower indifferent between borrowing and not borrowing. This

[17]We repeat this step 500 times.

gives us the mean and variance of the return ($\mu(r, \varphi)$,$\sigma^2(r, \varphi)$) for each realization of $r \leq indiff(\varphi)$.

In step 1 of all three cases, we need to compute the type of the borrower (φ) that corresponds to a given reserve rate, s. In principle, we can compute the type by solving for φ the borrower's FOCs as long as $s < 0.36$. However, at $s = 0.36$ we need to impute the borrower type as the FOCs may not hold. To impute the borrower's type, we first fit a truncated normal distribution to the listing with $s < 0.36$, and then draw φ from the fitted Normal distribution conditional on $\varphi < \varphi_{min}$, where φ_{min} is the borrower type at $s = 0.36$.

The procedure of how we fit a Normal distribution is as follows:

1. Take observations (s_i, x_i) such that x_i is close to x – In particular, take observations such that $x_{ho,i} = x_{ho}$, $|x_{amt,i} - x_{amt}| < 3000$ and $|x_{dti,i} - x_{dti}| < 0.1$.

2. For each observation, we compute the borrower type using the first-order condition, and compute the likelihood of a Normal distribution if $s_i < 0.36$. If $s = 0.36$, we treat it as being a censored observation.

3. We choose the parameter that maximizes the likelihood of the observation.

7.3 Additional Results on Counterfactual Supply Curves and Welfare

In Section 8 of the paper, we highlight the possibility that the credit supply curve under pooling can be backward bending and that welfare under signaling can be higher than that under pooling. In this section, we study how listing characteristics affect the shape of the credit supply curve and the welfare comparison between signaling and pooling. In order to do so, for each listing, we construct a dummy variable for whether or not expected welfare is higher under signaling than under pooling, $\mathbf{1}_{\{\text{Signaling}>\text{Pooling}\}}$, and another dummy variable for whether or not the credit supply curve is backward bending, $\mathbf{1}_{\{\text{Backward Bending}\}}$.[18] We then regress these dummy variables on the listing characteristics used in the structural model.

Table 19 reports the OLS estimation results. The first column reports the results for whether or not expected welfare is higher under signaling than that under pooling. The second column corresponds to whether or not the supply curve is backward bending.

In both regressions we find that the coefficients on the amount are estimated to be positive and statistically significant. As we report in Table 7 of the main text, the dispersion of

[18] We identify whether a supply curve is backward bending by comparing the supply amount at the highest interest rate (36%) and the maximum supply amount (in the interest rate range $[0\%, 36\%]$), and we classify a supply curve to be backward bending when the former is strictly smaller than the latter.

borrower type φ is increasing in the borrower's requested amount. As the variance of unobserved type increases, the value of signaling generally increases, which makes the expected social welfare under signaling more likely to be greater than that under pooling. Similarly, larger variance in the distribution of φ makes adverse selection more severe, resulting in the supply curve to be backward bending.

Table 19: Conditions for signaling to be welfare enhancing and the supply curves to be backward bending.

	(1)	(2)
	$\mathbf{1}_{\{\text{Signaling}>\text{Pooling}\}}$	$\mathbf{1}_{\{\text{Backward Bending}\}}$
Amount ($1000)	1.66E-02***	4.70E-02***
	(4.48E-04)	(3.98E-04)
Debt/income	-4.80E-04	3.07E-03
	(4.09E-03)	(3.62E-03)
Home-ownership	-8.45E-02***	-5.26E-03
	(5.89E-03)	(5.22E-02)
Grade		
AA	0.244***	-0.035***
	(0.009)	(0.008)
A	0.207***	0.080***
	(0.008)	(0.008)
B	0.137***	0.003***
	(0.007)	(0.006)
Constant	0.602***	-0.215***
	(0.006)	(0.006)
Observartions	11,541	11,541

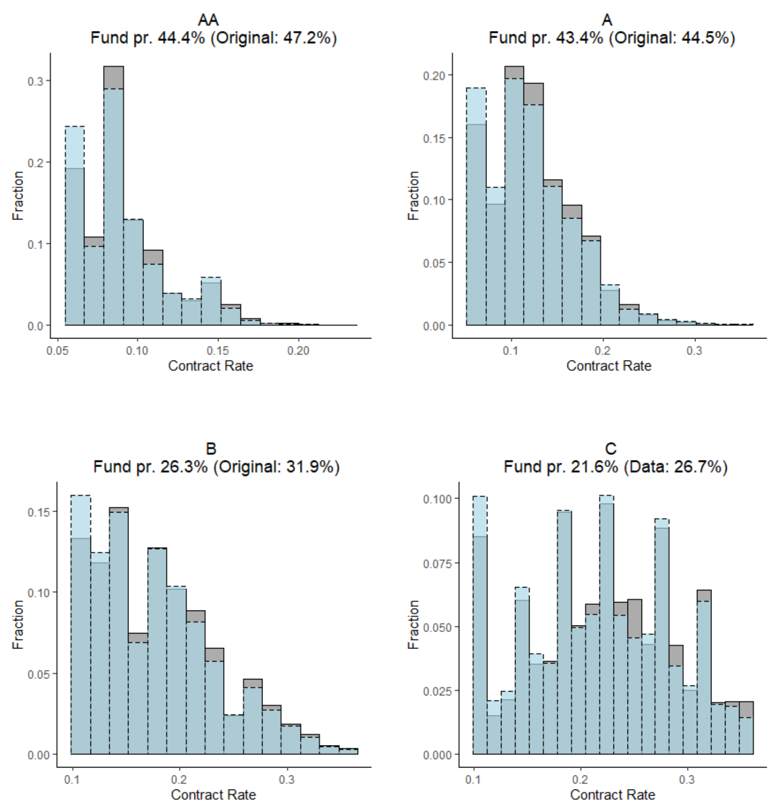
Note: This table reports the OLS estimation results. The first column reports the estimated coefficients with a dummy variable for whether or not expected welfare is higher under signaling than under pooling as the dependent variable. The second column reports the estimated coefficients with a dummy variable for whether or not the credit supply curve is backward bending as the dependent variable. The standard errors are reported in parentheses. *, **, and *** respectively denote significance at 10%, 5%, and 1% levels.

8 Sensitivity Analysis and Alternative Model of Lender Beliefs

8.1 Sensitivity of Counterfactual Simulations

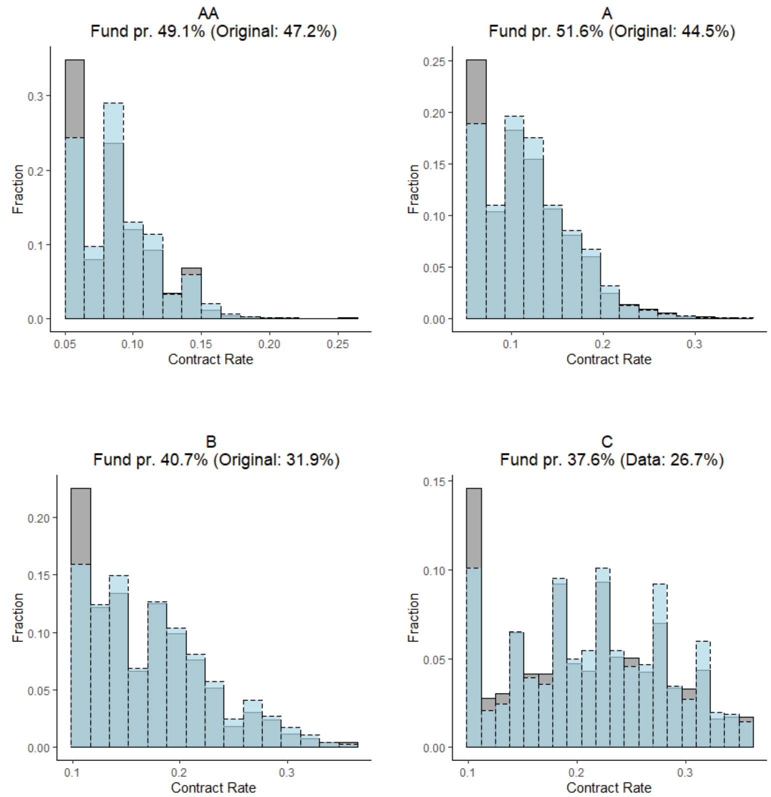
Our analysis of the counterfactual experiments in Section 8 of the main text keeps fixed the distribution of the number of potential bidders (F_N) and the requested amount. In order to gauge how sensitive the counterfactual results are to these assumptions, we decrease μ_N by 5% and decrease the requested loan amount by 5% and recompute the equilibrium and associated welfare under signaling in each case. We compare the equilibrium distribution of the contract interest rates and the funding probability under the perturbations with those of the baseline. We also compute welfare under signaling, pooling and symmetric information for each of the perturbations.

Figure 7: Sensitivity–Number of Potential Lenders



Note: The numbers below each credit grade are the funding probabilities for the sensitivity analysis and that of the baseline. The bars with solid lines correspond to the distribution of the contract interest rate for the sensitivity analysis and the bars with dashed lines correspond to those of the baseline.

Figure 8: Sensitivity–Amount



Note: The numbers below each credit grade are the funding probabilities for the sensitivity analysis and that of the baseline. The bars with solid lines correspond to the distribution of the contract interest rate for the sensitivity analysis and the bars with dashed lines correspond to those of the baseline.

Figures 7 and 8 compare the distribution of contract interest rate and the funding probability when we decrease the number of bidders and when we decrease the requested amount. When μ_N is decreased by 5%, we find that the funding probability decreases and contract interest rates increase, as expected. The change in the funding probability is about 1% to 5% depending on credit grade. When we decrease the requested loan amount, we find that the funding probability increases by about 2-11% and the interest rate decreases.

Table 20 reports the results of the sensitivity analysis in terms of welfare. The first panel corresponds to the sensitivity analysis with respect to the number of lenders and the second panel corresponds to the sensitivity analysis with respect to the requested amount. We report the welfare computed for the sensitivity analysis alongside those of the baseline results reported in the main text. While the welfare numbers change somewhat depending on the perturbation we introduce, the welfare ranking among signaling, pooling and symmetric information is mostly preserved. Moreover, we find that, for credit grades AA and A, signaling can restore much of the surplus difference between signaling and pooling as we find for our

Table 20: Welfare – Sensitivity

		AA		A		B		C	
		Sensi-tivity	Base-line	Sensi-tivity	Base-line	Sensi-tivity	Base-line	Sensi-tivity	Base-line
		Sensitivity: Number of Potential Lenders							
Borrower	Pooling	180.0	185.5	448.7	464	136.0	141.8	49.6	58
	Separating	270.9	253.6	470.9	487.5	89.2	150.7	44.6	39.1
	Symmetric	275.5	281.8	483.0	499.2	148.6	154.7	54.5	63.4
Lender	Pooling	50.13	102.2	154.9	265	115.12	217.8	110.62	262.9
	Separating	70.1	141.9	164.3	289.9	121.9	218.3	113.5	271.4
	Symmetric	74.9	144.8	171.9	295.6	176.1	231.7	138.8	277.1
Total	Pooling	230.1	287.7	603.6	729	251.2	359.6	160.2	320.9
	Separating	341.0	395.5	635.2	777.4	211.1	369.1	158.1	310.5
	Symmetric	350.4	426.6	654.9	794.8	324.7	386.4	193.3	340.5
% Recovered		92.2	77.6	61.6	93.7	-54.4	35.3	-6.3	-53.2
		Sensitivity: Amount							
Borrower	Pooling	140.5	185.5	625.3	464	139.4	141.8	59.3	58
	Separating	274.9	253.6	666.8	487.5	150.8	150.7	64.9	39.1
	Symmetric	284.9	281.8	667.5	499.2	151.8	154.7	65.0	63.4
Lender	Pooling	116.6	102.2	275.6	265	245.9	217.8	258.4	262.9
	Separating	142.4	141.9	294.5	289.9	245.7	218.35	266.2	271.4
	Symmetric	144.6	144.8	301.6	295.6	261.0	231.7	272.5	277.1
Total	Pooling	257.1	287.7	900.9	729	385.3	359.6	317.7	320.9
	Separating	417.3	395.5	961.3	777.4	396.5	369.05	331.1	310.5
	Symmetric	429.5	426.6	969.1	794.8	412.8	386.4	337.5	340.5
% Recovered		92.9	77.6	88.5	73.6	40.9	35.3	67.5	-53.2

Note: The table reports the expected surplus for different market designs by credit grade. The top panel shows the results when the parameter μ_N is decreased by 5%, and the bottom panel when x_{amt} is increased by 5%.

baseline results.

8.2 Alternative Model of Lender Beliefs

An important restriction of the model that we estimate in the main text is that the lenders' beliefs over the return from lending money are allowed to deviate from rational expectations in very limited ways. In particular, we assume that the lender's risk attitude (A_j) and ϵ_{0j} are independent. This assumption implies that the lenders' beliefs over the variance of the return to coincide with the realized variance.

In order to check the robustness of our results to this assumption, we estimate an alternative specification with more structure on the lenders' beliefs. In particular, we estimate a model in which the beliefs are given as follows,

$$\begin{aligned} \mu &= \mu_{RE} + \rho_1 \mu_{S\&P} \\ \sigma^2 &= \sigma^2_{RE} \times \exp(\nu_1 \sigma^2_{S\&P}), \end{aligned}$$

where μ_{RE} and σ^2_{RE} are the mean and variance of the realized return and $\mu_{S\&P}$ and $\sigma^2_{S\&P}$ are the monthly mean and the volatility index of the S&P 500.[19] ρ_1 and ν_1 are parameters to be estimated. An interpretation of this specification is that lenders might adjust their beliefs adaptively and form beliefs that deviate from rational expectations depending on macroeconomic conditions.

In Table 21, we report the estimates of the lender's parameters of this model. We find that our estimates of ρ_1 and υ_1 are quite small, although they are statistically significant.

Figure 9 compares the counterfactual credit supply curve under this alternative specification of the lender's belief with those of the baseline model. The first 8 figures in Panel (a) shows the the supply curves under the alternative specification and the figures in Panel (b) correspond to those under the baseline specification. As the figures show, the supply curves are qualitatively similar to the baseline results.

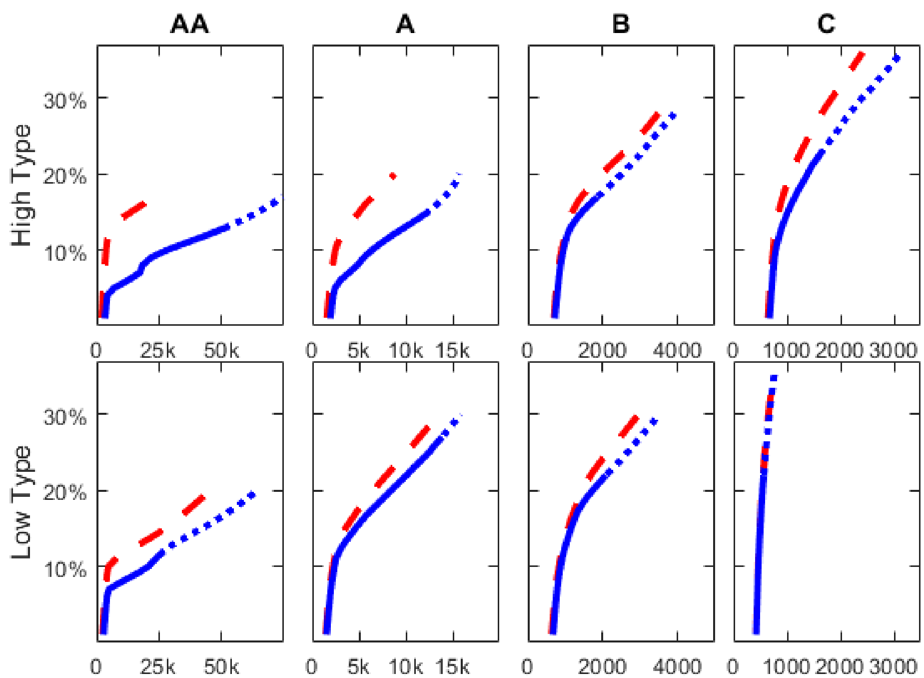
Lastly, we report welfare under pooling, signaling, and no asymmetric information using the alternative specification that we estimate. The results are reported in Table 22. The results show that our results are qualitatively similar to those in the main text.

9 Proofs

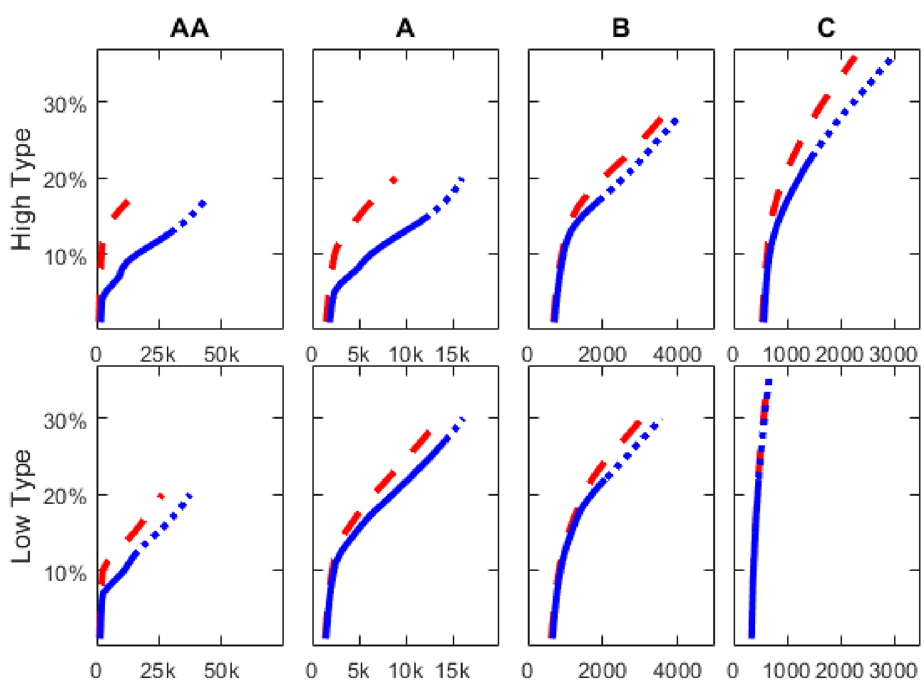
This section collects the proofs of the Propositions in the main text.

[19] The mean return of S&P 500 during the sample time period is about -10%, and the mean volatility is about 30.

Figure 9: Credit Supply Curve Under Alternative Specification of Lender Beliefs:



(a) Credit Supply under Alternative Beliefs



(b) Credit Supply under Baseline Model

Note: The thick dashed curve corresponds to the credit supply curve under no signaling (i.e., pooling) when μ and σ^2 depend on the S&P500 index. The solid line corresponds to the credit supply curve under signaling, and the dotted line under no asymmetric information.

Table 21: Parameter Estimates of the Lender's Model–Robustness Check

	AA	A	B	C
μ_N	6.684	5.137	4.454	5.140
	(2.291)	(1.700)	(1.502)	(1.673)
σ_N	0.521	0.899	0.677	0.641
	(0.176)	(0.308)	(0.235)	(0.222)
c_{75}	-0.302	0.813	0.854	1.302
	(0.102)	(0.275)	(0.289)	(0.441)
c_{100}	0.350	1.546	3.301	4.150
	(0.118)	(0.516)	(1.098)	(1.393)
c_{200}	0.454	1.197	0.905	4.702
	(0.153)	(0.403)	(0.306)	(1.591)
c_{500}	1.077	-3.930	-8.498	-8.098
	(0.361)	(1.334)	(2.869)	(2.734)
c_{1000}	-22.85	-244.1	-160.3	-441.2
	(7.672)	(82.472)	(54.415)	(149.268)
μ_A	4.01E-04	3.60E-04	-1.59E-04	1.04E-03
	(1.36E-04)	(1.22E-04)	(5.37E-05)	(3.57E-04)
σ_A	2.89E-04	7.84E-04	-5.90E-04	1.26E-03
	(9.76E-05)	(2.61E-04)	(2.01E-04)	(4.22E-04)
$\mu_{\varepsilon 0}$	13.59	13.05	18.16	77.31
	(4.509)	(4.380)	(5.989)	(26.089)
$\sigma_{\varepsilon 0}$	10.78	15.77	24.49	56.38
	(3.728)	(5.404)	(8.877)	(19.353)
ρ_1	-4.17E-03	2.64E-03	1.38E-03	1.27E-04
	(1.40E-03)	(8.92E-04)	(4.69E-04)	(4.29E-05)
ν_1	-7.12E-04	-7.05E-03	-1.97E-03	-1.41E-04
	(2.39E-04)	(2.39E-03)	(6.75E-03)	(4.79E-04)

Note: We report the parameter estimates of the lender's model. Standard errors are obtained by bootstrap (150 times) and they are reported in parentheses.

9.1 Proof of Proposition 1

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.,*

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

In order to prove Proposition 1, we first prove the following lemma.

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

Table 22: Welfare–Robustness Check with Alternative Lender Belief

		AA	A	B	C
Lender	Pooling	99.2	247.4	200.5	252.7
	Separating	137.8	270.8	204.0	255.0
	Symmetric	142.9	277.9	224.2	269.5
Borrower	Pooling	189.6	463.9	141.4	58.8
	Separating	282.2	497.4	147.7	58.5
	Symmetric	285	499.1	154.4	64.2
Total	Pooling	288.8	711.3	341.9	311.5
	Separating	420.0	768.2	351.7	313.5
	Symmetric	427.9	777.0	378.6	333.7
% Recovered		94.3%	86.6%	26.7%	9.0%

Note: The table reports the expected surplus for different market designs by credit grade under the specification for the robustness check.

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

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

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

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

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

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

$\frac{\partial}{\partial r \partial \varphi} V_t(r,\varphi) \leq 0$, note that

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

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

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

We are now ready to prove Proposition 1.

Proof. Note that

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

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

Consider the first term. Note that $\frac{\partial \Pr(s)}{\partial s}$ is positive by assumption, $\frac{\partial V_1}{\partial \varphi} < 0$ and $\frac{\partial \lambda}{\partial \varphi} > 0$. Hence the first term is negative. Now consider the second term. Note that for $s_0 < s_1$, $F(r|s_1)$ first-order stochastically dominates $F(r|s_0)$. Hence if $\frac{\partial}{\partial \varphi}V_1(r,\varphi)$ is non-increasing in r, then $\int \frac{\partial}{\partial \varphi}V_1(r,\varphi)dF(r|s_0) \geq \int \frac{\partial}{\partial \varphi}V_1(r,\varphi)dF(r|s_1)$ for any s_0 and s_1 s.t. $s_0 < s_1$. This implies that $\Pr(s)\frac{\partial}{\partial s \partial \varphi}\int V_1(r,\varphi)dF(r|s) \leq 0$. ■

9.2 Proof of Proposition 2

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

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

9.3 Proof of Proposition 4

To simplify exposition, we first prove Proposition 4 assuming that all listings receive at least 3 or more bids. We then provide a full proof of Proposition 4 that does not require this additional assumption.

Proposition 4 *Assume that for each s and X^o 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 each listing receives at least 3 bids and the irreducibility condition of Hall and Zhou (2003) is satisfied. Then,* $u_t(r)$, $F_{\varphi|X}$, $\lambda(\cdot)$ *are identified. Moreover, there exists an interval on which* $F_{\epsilon|X}$ *is identified.*

In order to prove Proposition 4, we first prove the following lemma.

Lemma 2 *For each r, there exists a subset of listings E_r that are funded at r for which* $\Pr(X^u|s, X^o, E_r)$ *is identified.*

Proof. Fix $K \in \mathbb{N}$ and $r \in \mathbb{R}$. Also fix s and X^o. Now consider the set of listings that attracts a total of $K + 3$ bids and the interest rate of the first K bids are all equal to r.[20] Denoting by $\mathbf{r}$ the $K + 3$ vector of interest rates, $\mathbf{r}$ takes the following form, $\mathbf{r} = \{r, \ldots, r, r_{K+1}, r_{K+2}, r_{K+3}\}$. Suppose that K is large enough so that the listing would have been funded with just the first K bids. We denote the set of such listings as E_r. We are interested in identifying $\Pr(X^u|s, X^o, E_r)$.

Note that all listings in E_r ends up with an interest rate of r. This is because the interest rate is determined by the marginal bid. If there are K bids with the exact same interest rate r, the last three bids will not move the marginal interest rate. Similarly, the marginal interest that was displayed to the bidders would have been constant at r.

Now consider the joint distribution of $(r_{K+1}, r_{K+2}, r_{K+3})$. Note that these bids are independent because the marginal interest is constant at r at the time each bid was submitted. Hence, the joint distribution can be expressed as follows:

$$F(\tau) = \Pr(L|s, X^o, E_r) \prod_{j=1}^{3} G_L(\tau_j|s, X^o, E_r) + \Pr(H|s, X^o, E_r) \prod_{j=1}^{3} G_H(\tau_j|s, X^o, E_r),$$

where $G_{X^u}(\tau_j|s, X^o, E_r)$ denotes the probability that the interest bid by the bidder is less than $\tau_j \in \mathbb{R}$ when borrower type is X^u. This is a special case of the finite mixture model considered in Hall and Zhou (2000). Hence, by Theorem 4.3 of Hall and Zhou (2003), the mixing probabilities $\Pr(X^u|s, X^o, E_r)$ and $G_{X^u}(\tau_j|s, X^o, E_r)$ are both identified under an irreducibility condition. ■

[20] As long as enough bids are at r, the proof goes through without any modifications.

Now we prove Proposition 4.

Proof. We first note that we can normalize $d_T^X = 0$ without loss of generality by shifting the distribution of ε appropriately. Also, we can normalize the location of $F_{\varphi|X^o,X^u}$ at some point $X^o = \overline{X^o}$: Hence we set $F^{-1}_{\varphi|\overline{X^o},H}(\alpha^*) = 0$ for some $\alpha^* \in (0,1)$ and $\overline{X^o}$. Note that α^* quantile of the reserve rate distribution $F^{-1}_{s|\overline{X^o},H}(\alpha^*)$ is identified because $\Pr(X^u|s,\overline{X^o})$ is identified whenever there are 3 or more bids per listing. Identification of $\Pr(X^u|s,\overline{X^o})$ when each listing receives at least 3 or more bids follows directly from Hall and Zhou (2003).

Consider the repayment decision of the borrower with $F^{-1}_{\varphi|\overline{X^o},H}(\alpha^*)(=0)$ at period $t=T$. These borrowers correspond to $X^u = H$ type borrowers who set reserve equal to the α^* quantile of the reserve rate distribution. The borrower's problem is as follows:

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

where $\overline{x}_{amt}$ is an element of $\overline{X^o}$. Now consider r in range $[r', r'']$ such that $\Pr(X^u = H|\overline{X^o}, r, s) = 1$. Conditioning on this event guarantees that the set of borrowers consist of only $X^u = H$ types. Using variation in $r \in [r', r'']$, we can nonparametrically identify the conditional distribution of ε_T given $\overline{X^o}$ and $X^u = H$, i.e., $F_{\varepsilon_T|\overline{X^o},H}$. Once $F_{\varepsilon_T|\overline{X^o},H}$ is identified, we can identify $F^{-1}_{\varphi|\overline{X^o},H}(\alpha)$ for all α by conditioning the sample on the α-quantile of s and r in the range such that the probability $X^u = H$ is one given $\overline{X^o}$ and $s = F^{-1}_{s|\overline{X^o}}(\alpha)$, i.e., $\Pr(H|\overline{X^o}, s) = 1$. Note that the repayment problem for the borrowers among this sample is simply a binary threshold crossing model with a known error distribution, $F_{\varepsilon_T|\overline{X^o},H}$. Hence, $F_{\varphi|\overline{X^o},H}$ are identified.

Now consider the $t = T-1$ period problem with $X^o = \overline{X^o}$, $s = F^{-1}_{s|\overline{X^o},H}(\alpha^*)$, and $r \in [r', r'']$ such that $\Pr(H|\overline{X^o}, s) = 1$:

$$\begin{cases} \text{repay: if } -(r \times \overline{x}_{amt}) + d_{T-1}^{\overline{X^o},H} + \beta V_T(r, F^{-1}_{\varphi|\overline{X^o},H}(\alpha^*)) + \varepsilon_{T-1} \geq -F^{-1}_{\varphi|X^*,H}(\alpha^*) = 0 \\ \text{default: otherwise,} \end{cases}$$

where $d_{T-1}^{\overline{X^o},H}$ denotes the $T-1$ dummy variable in the utility function for type H. Note that $V_T(r, F^{-1}_{\varphi|\overline{X^o},H}(\alpha^*))$ has already been identified. Hence, similar as before, we can nonparametrically identify the distribution of $(\varepsilon_{T-1} + d_{T-1}^{\overline{X^o},H}$ and the value of β using variation in r. It should be clear that the distribution of $\{\varepsilon_t + d_t^{\overline{X^o},H}\}_{t \leq T-2}$ can also be identified by looking at the borrower's period t problem and the associated default probability. A location normalization on $\{\varepsilon_t\}$ identifies $d_t^{\overline{X^o},H}$ and $F_{\varepsilon_t|\overline{X^o},H}$ separately.

We now show that the distribution of ε_t and φ are also identified for the low types, i.e., $F_{\varepsilon_t|\overline{X^o},L}$ and $F_{\varphi|\overline{X^o},L}$ are identified. Recall from Proposition 2 that for each r, there exists a subset of funded loans E_r such that the contract interest rate is r and we know $\Pr(X^u = L|s, \overline{X^o}, E_r)$. Because $F_{\varepsilon_t|\overline{X^o},H}$, $F_{\varphi|\overline{X^o},H}$ and time dummies $\{d_t^{\overline{X^o},H}\}$ are identified, we can identify the default probability for type H at each period t. Given that we know $\Pr(H|s, \overline{X^o}, E_r)$, we also know how much type-H borrowers contribute to the default probability at time t. This implies that we also identify the default probability of type-L borrowers at each period t. Then, by using variation in r and how the default probability changes for type L borrowers, we identify $F_{\varepsilon_T|\overline{X^o},L}$. Identification of $F_{\varepsilon_t|\overline{X^o},L}$ for $t < T$, time dummies $\{d_t^{\overline{X^o},L}\}$ and $F_{\varphi|\overline{X^o},L}$ is similar to that for the type H borrowers.

Because the utility parameters of the borrowers that determine repayment behavior are already identified when $X^o = \overline{X^o}$, it is easy to see that $\Pr(X^u|r, s, \overline{X^o})$ is identified. To see this, we observe the overall default probability given $r, s, \overline{X^o}$. We can also compute the probability of default conditional on type, i.e., $\Pr(\text{default}|r, s, \overline{X^o}, X^u)$. Given that the overall probability is a mixture of the conditional probabilities, the mixing probabilities $\Pr(X^u|r, s, \overline{X^o})$ must be identified (except in the degenerate case in which the default probabilities are exactly the same).[21] Similarly, the funding probability $\Pr(s, \overline{X^o}, X^u)$ is identified.

Now, we discuss how to identify $\lambda(\varphi)$. Rearranging the borrower's FOC in equation (6) evaluated at $X^o = \overline{X^o}$ and solving for $\lambda(\varphi)$, we obtain

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

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

Lastly, we show that $F_{\varphi|X}$ and $F_{\varepsilon|X}$ are identified for any X. To see that $F_{\varphi|X}$ and $F_{\varepsilon|X}$ are identified for any X, note that it is enough to identify $F_{\varphi|X}(0)$ – if $F_{\varphi|X}(0)$ is identified, we can follow the same steps as above to identify $F_{\varphi|X}$ and $F_{\varepsilon|X}$. In order to see that $F_{\varphi|X}(0)$ is identified, consider a given profile, $(F^*_{\varepsilon|X}, F^*_{\varphi|X}, \lambda^*, d^*_t)$. Note that the profile $(F^*_{\varepsilon|X}, F^*_{\varphi|X}, \lambda^*, d^*_t)$ is identified up to a single constant $F_{\varphi|X}(0)$. The set of profiles that are observationally equivalent to $(F^*_{\varepsilon|X}, F^*_{\varphi|X}, \lambda^*, d^*_t)$ are given by $\{(F_{\varepsilon|X}, F_{\varphi|X}, \lambda, d_t) : F_{\varepsilon|X}(h) = F^*_{\varepsilon|X}(h - \kappa),\ F_{\varphi|X}(h) = F^*_{\varphi|X}(h + \kappa),\ d_T = d^*_T,\ d_t = d^*_t - \beta\kappa\ (t < T),\ \lambda(\varphi) = \lambda^*(\varphi) + \beta\kappa\}$.

[21]To the extent that one type of borrower has strictly higher creditworthiness than the other, this will never occur.

Given that we have already identified $\lambda(\varphi)$, we can identify $F_{\varphi|X}(0)$. ■

9.4 Proof of Proposition 5

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.*

In this section, we prove identification of the primitives of the model of the lenders. Our proof of identification proceeds by first showing identification of F_A, F_{ε_0}, and c_q under the assumption that $P_q(\mu, \sigma)$, which we will define below, is identified for all values of (μ, σ) and $q \in M \cup \{\$0\}$. We will then show that $P_q(\mu, \sigma)$ and F_N are identified. Given a listing with mean and variance of return equal to μ and σ^2, define $P_q(\mu, \sigma)$ to be the probability that funding q dollars gives higher utility to a lender than funding q' $(q' \neq q)$ dollars. Formally, $P_q(\mu, \sigma)$ is expressed as follows:

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

Note that $P_q(\mu, \sigma)$ corresponds to the probability that (A, ε_0) lie in the region defined by the inequalities in the expression above. By varying μ and σ, this region changes. Lemma 3 below claims that with enough variation in μ and σ, we can recover the probability that (A, ε_0) is contained in an arbitrary set, i.e., identify F_A and F_{ε_0}.

Lemma 3 $(F_A, F_{\varepsilon_0}, c_q)$ *are identified if* $(P_q(\mu, \sigma), P_0(\mu, \sigma))$ *are identified.*

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

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

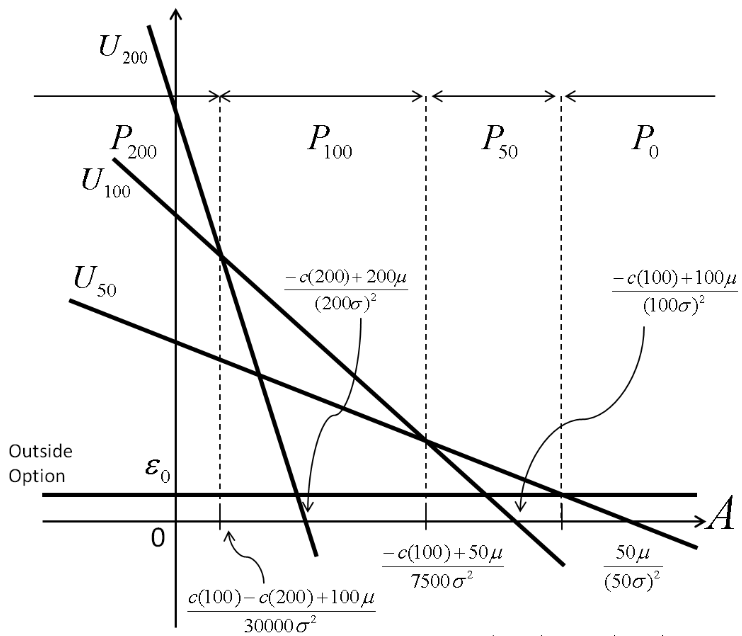
The two lemmas together prove Proposition 5.

First we provide a proof of the first lemma. We discuss the case in which the lender's amount choice is $M = \{\$50, \$100, \$200\}$ for exposition. The proof for the case in which $M = \{\$50, \$75, \$100, \$200, \$500, \$1000\}$ (the actual specification that we take to the data) is analogous.

Proof.

Recall that $P_q(\mu, \sigma)$ corresponds to the probability that (A, ϵ_0) falls into a region defined by inequalities. Fix a particular value of ϵ_{0j}, and $P_q(\mu, \sigma)$ can be considered as defining a

Figure 10: Lender Utility as a Function of A



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

region for A. The region of A that corresponds to $P_q(\mu,\sigma)$ is defined by the intersection of straight lines $U_q = U_q(A) \equiv q\mu - A(q\sigma)^2 - c(q)$ for $q = \$50$, $\$100$, and $\$200$ (U_{50}, U_{100}, U_{200}). Figure 10 illustrates this for the case of $\frac{100\mu - c_{200}+c_{100}}{30000\sigma^2} < \frac{c(50)-c_{100}+50\mu}{7500\sigma^2}$ ($\Leftrightarrow \mu > \frac{-c_{200}+5c_{100}-4c(50)}{100}$) (which ensures that the intersection between U_{200} and U_{100} is to the left of the intersection between U_{100} and U_{50}).

Note first that it is possible to assume $c_{50} = 0$ without loss of generality.[22] We also assume that $c_{200} > 3c_{100}$ for our proof below. This restriction is just for exposition: Identification for $c_{200} < 3c_{100}$ can be shown analogously. Now, consider $P_{200}(\mu,\sigma)$. Given μ and σ, bidding \$200 is optimal if the risk parameter A_j is sufficiently small and the outside option ϵ_0 is also sufficiently small. Hence, $P_{200}(\mu,\sigma)$ can be expressed as follows,

$$\begin{aligned} P_{200}(\mu,\sigma) &= \Pr(\{U_{200}(\mu,\sigma) > \max\{U_{50}(\mu,\sigma),\, U_{100}(\mu,\sigma)\}\} \cap \{\epsilon_0 < U_{200}(\mu,\sigma)\}) \\ &= \Pr(A_j < \overline{A}(\mu,\sigma) \wedge \epsilon_0 < 200\mu - A_j(200\sigma)^2 - c_{200}), \end{aligned}$$

where $\overline{A}(\mu,\sigma) = \frac{c_{100}-c_{200}+100\mu}{30000\sigma^2}$.[23]

[22] We can add a constant to c_{50}, c_{100}, c_{200}, and shift the distribution of ϵ_{0j} to the right without changing the distribution of outcomes.

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

Observe that if $\overline{A}(\mu,\sigma) < 0$, then as we let $\sigma \to 0$ (while keeping μ fixed), $P_{200}(\mu,\sigma)$ would tend to 0.[24] However, if $\overline{A}(\mu,\sigma) = 0$, then as $\sigma \to 0$, $P_{200}(\mu,\sigma)$ would converge to a positive number, i.e.,

$$\lim_{\sigma\to 0} P_{200}(\mu,\sigma) = \Pr(A_j < 0 \wedge \epsilon_0 < c_{200} - 2c_{100}),$$

where we have used the fact $\overline{A}(\mu,\sigma) = 0 \Longleftrightarrow c_{100} - c_{200} + 100\mu = 0$ and $A_j(200\sigma)^2 \to 0$.[25] Let us define μ^* as

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

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

Now we consider identification of F_A, given that $\{c_{100}, c_{200}\}$ has already been identified. Again note that

$$P_{200}(\mu,\sigma) = \Pr(A_j < \overline{A}(\mu,\sigma) \wedge \epsilon_0 < 200\mu - A_j(200\sigma)^2 - c_{200}).$$

Now take μ and σ so that $\overline{A}(\mu,\sigma) = \delta^+$, (or equivalently, $\mu = \frac{c_{200}-c_{100}+30000\sigma^2\delta^+}{100}$), where δ^+ is some positive number.[26] Then consider keeping $\overline{A}(\mu,\sigma)$ fixed at δ^+, but moving $200\mu - A_j(200\sigma)^2 - c_{200}$ by changing both μ and σ. In particular, as $\sigma \to 0$, we have

$$\begin{aligned} \mu &\to \frac{c_{200} - c_{100}}{100} \text{ and} \\ P_{200}(\mu,\sigma) &\to \Pr(A_j < \delta^+ \wedge \epsilon_0 < c_{200} - 2c_{100}) \\ &= \Pr(A_j < \delta^+)\Pr(\epsilon_0 < c_{200} - 2c_{100}), \end{aligned}$$

where we have used the independence assumption between A_j and ϵ_0 for going from the second line to the third line. By varying δ^+ (> 0), we can identify $\Pr(A_j < t)\Pr(\epsilon_0 < c_{200} - 2c_{100})$ for all $t > 0$. By taking $t \to +\infty$, we also identify $\Pr(\epsilon_0 < c_{200} - 2c_{100})$. Similarly, by taking μ and σ such that $\overline{A}(\mu,\sigma) = \delta^-$ for some negative constant, we can

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

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

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

identify $\Pr(A_j < t)\Pr(\epsilon_0 < c_{200} - 2c_{100})$ for all $t < 0$.[27][28] Combining these two results together, F_A is identified.

We now discuss identification of F_{ϵ_0} given that F_A and $\{c_{200}, c_{100}\}$ have been identified. Recall that $P_{200}(\mu, \sigma)$ can be expressed as follows,

$$P_{200}(\mu,\sigma) = \Pr(A_j < \overline{A}(\mu,\sigma) \wedge \epsilon_0 < 200\mu - A_j(200\sigma)^2 - c_{200}).^{29}$$

Suppose we take a μ so that $c_{100} - c_{200} + 100\mu > 0$ ($\Leftrightarrow \mu > \frac{c_{200}-c_{100}}{100}$). Now consider holding μ constant and taking the limit as $\sigma \to 0$. Then $P_{200}(\mu,\sigma) \to \Pr(\epsilon_0 < 200\mu - c_{200})$. Because we can move μ in the region $\mu > \frac{c_{200}-c_{100}}{100}$,[30] the distribution of ϵ_0 is identified for all $t > c_{200} - 2c_{100}$.

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

$$P_{100}(\mu,\sigma) = \Pr(\overline{A}(\mu,\sigma) < A_j < \frac{-c_{100}+50\mu}{7500\sigma^2} \wedge \epsilon_0 < 100\mu - A_j(100\sigma)^2 - c_{100}).^{31}$$

Again, take a μ so that $-c_{100} + 50\mu > 0$ and $c_{100} - c_{200} + 100\mu < 0$ ($\Leftrightarrow \frac{c_{100}}{50} < \mu < \frac{c_{200}-c_{100}}{100}$). As before, we take $\sigma \to 0$, while holding μ constant. Then $P_{100}(\mu,\sigma) \to \Pr(\epsilon_0 < 100\mu - c_{100})$. Because we can move μ in the region $\frac{c_{100}}{50}, < \mu < \frac{c_{200}-c_{100}}{100}$[32] the distribution of ϵ_0 is identified for all $t \in [c_{100}, c_{200} - 2c_{100}]$.

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

$$P_{50}(\mu,\sigma) = \Pr(A_j > \frac{-c_{100}+50\mu}{7500\sigma^2} \wedge \epsilon_0 < 50\mu - A_j(50\sigma)^2).^{33}$$

As before, take a μ so that $-c_{100} + 50\mu < 0$ ($\Leftrightarrow \mu < \frac{c_{100}}{50}$). Then $P_{50}(\mu,\sigma) \to \Pr(\epsilon_0 < 50\mu)$. Because we can move μ in the region $\frac{c_{100}}{50} > \mu$ ($> \frac{-c_{200}+5c_{100}}{100}$), the distribution of ϵ_0 is identified for at all $\frac{-c_{200}+5c_{100}}{2} < t < c_{100}$.

[27] We can apply the analogous argument here. We first fix $\overline{A}(\mu,\sigma)$ at some negative constant δ^-, but move $200\mu - A_j(200\sigma)^2 - c_{200}$ by changing both μ and σ. Then considering $\sigma \to 0$, we obtain $\mu \to \frac{c_{200}-c_{100}}{100}$, and $P_{200}(\mu,\sigma) \to \Pr(A_j < \delta^- \wedge \epsilon_0 < c_{200} - 2c_{100} = \Pr(A_j > \delta^-)\Pr(\epsilon_0 < c_{100} - 2c_{200})$. Hence, by moving δ^- appropriately, we identify $\Pr(A_j < t)\Pr(\epsilon_0 < c_{200} - 2c_{100})$ for all $t < 0$.

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

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

[30] Note that $c_{200} > 3c_{100}$ implies $\frac{c_{200}-c_{100}}{100} > \frac{-c_{200}+5c_{100}}{100}$.

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

[32] Note that $c_{200} > 3c_{100}$ implies $\frac{c_{200}-c_{100}}{100} > \frac{c_{100}}{50} > \frac{-c_{200}+5c_{100}}{100}$.

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

Lastly, consider $P_{50}(\mu, \alpha)$, when $\mu < \frac{-c_{200}+5c_{100}}{100}$:[34]

$$P_{50}(\mu, \sigma) = \Pr(A_j > \frac{150\mu - c_{200}}{37500\sigma^2} \wedge \epsilon_0 < 50\mu - A_j(50\sigma)^2).$$

If we take a μ so that $150\mu - c_{200} < 0$ ($\Leftrightarrow \mu < \frac{c_{200}}{150}$). Then $P_{50}(\mu, \sigma) \to \Pr(\epsilon_0 < 50\mu)$. Because we can move μ in the region $\mu < \min\{\frac{-c_{200}+5c_{100}}{100}, \frac{c_{100}}{150}\}$ ($\Leftrightarrow \mu < \frac{-c_{200}+5c_{100}}{100}$), the distribution of ϵ_0 is identified for at all $t < \frac{-c_{200}+5c_{100}}{2}$. Combining these results, $F_{\epsilon_0}(t)$ is identified for all $t \in \mathbb{R}$. ■

We now give a proof of Lemma 4.

Proof. First, it is clear that the ratio $P_q(\mu, \sigma)/P_{q'}(\mu, \sigma)$ for $q, q' \neq 0$ is identified from the ratio of lenders who bid q and q' for listings whose mean and variance of the return are mu and $sigma^2$ when funded at the marginal interest rate.[35] This implies that we only need to identify F_N and $P_0(\mu, \sigma)$.

Recall that we assumed that F_N has finite support, i.e., the support is $\{0, 1, \cdots, \overline{N}\}$ for some finite $\bar{N}$. First, the upper bound $\bar{N}$ is identified by the maximum number of observed bids.

Because the ratio $P_q(\mu, \sigma)/P_{q'}(\mu, \sigma)$ for $q, q' \neq 0$ is identified, we only need to show that $P_{50}((\mu, \sigma)$ is identified. Consider listings with a requested amount just equal to $\$50 \times \bar{N}$. We also fix other listings characteristics at X^o and s. Let $(\mu_{X^u}, \sigma_{X^u})$ denote the mean and standard error of the return if the listing in funded at the reserve rate (i.e., $r = s$) for type-X^u borrower. $(\mu_{X^u}, \sigma_{X^u})$ depend on X^o and s, but we suppress this dependence. The probability that the listing receives exactly k \$50 bids (and no bids with \$100, \$200, etc.) is as follows:

$$\begin{aligned}
&\Pr(k \text{ \$50 bids}|X^o, s) \\
&= \Pr(L|X^o, s) \sum_{j=k}^{\overline{N}} f_N(j) C_k^j P_{50}(\mu_L, \sigma_L)^k (1 - P_{50}(\mu_L, \sigma_L)^{j-k} \\
&+ \Pr(H|X^o, s) \sum_{j=k}^{\overline{N}} f_N(j) C_k^j P_{50}(\mu_H, \sigma_H)^k (1 - P_{50}(\mu_H, \sigma_H)^{j-k},
\end{aligned}$$

where $C_k^j = k!/j!(k-j)!$, $\Pr(L|X^o, s)$ and $\Pr(H|X^o, s)$ are the mixing probabilities. Note that the mixing probabilities are identified. The unknowns are $f_N(j)$, $P_{50}(\mu_L, \sigma_L)$ and

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

[35] By assumption there is a range (r', r'') such that $Pr(H|X^o, r, s) = 1$ for $r \in (r', r'')$. The ratio $P_q(\mu, \sigma)/P_{q'}(\mu, \sigma)$ is identified for values (μ, σ) as long as there exists (X^o, r, s) such that $Pr(H|X^o, r, s) = 1$ and the associated mean and variance of return are (μ, σ)

$P_{50}(\mu_H, \sigma_H)$. Note that there are $\overline{N}$ independent restrictions for each $k = 0 \cdots \overline{N} - 1$ and there are $\overline{N} + 2$ unknowns $(f_N(0) \cdots f_N(\overline{N} - 1), P_{50}(\mu_L, \sigma_L), P_{50}(\mu_H, \sigma_H)$. Now consider repeating the above exercise with a different X^o and s with associated mean and variance of return $(\mu'_{X^u}, \sigma'_{X^u})$. Then, this yields $\overline{N}$ additional equations. Because we assume that F_N is invariant to (μ, σ), we have a total of $2 \times \overline{N}$ equations and $\overline{N} + 4$ unknowns $(P_{50}(\mu_{X^u}, \sigma_{X^u})$, $P_{50}(\mu'_{X^u}, \sigma'_{X^u})$, $f_N(0), \cdots,$ $f_N(\bar{N}))$. Assuming that F_N is invariant to (μ, σ), we can increase the number of equations at a faster rate than the number of observables. Hence $P_{50}(\mu_{X^u}, \sigma_{X^u})$, and $f_N(0), \cdots,$ $f_N(\overline{N})$ are identified. ■

Proof of Proposition 4 with 2 or Less Bids Our proof of Proposition4 above made the assumption that there are at least 3 bids per listing. We now relax this assumption. Consider a listing in which the requested amount is \$50 $\times \overline{N}$, where $\overline{N}$ is the maximum number of bidders. $\overline{N}$ is immediately identified from the maximum number of actual bids. We fix listings characteristics at X^o and s. Let $(\mu_{X^u}, \sigma_{X^u})$ denote the mean and standard error of the return if the listing in funded at the reserve rate (i.e., $r = s$) for a borrower whose type is X^u. The mean and standard error $(\mu_{X^u}, \sigma_{X^u})$ are not identified yet.

Now consider the probability that a listing receives exactly k \$50 bids (but no bid with \$100, \$200, etc.):

$$\begin{aligned}
&\Pr(k \text{ \$50 bids}|s, X^o) \\
&= \Pr(L|s, X^o) \sum_{j=k}^{\overline{N}} f_N(j) C_k^j P_{50}(\mu_L, \sigma_L)^k (1 - P_{50}(\mu_L, \sigma_L))^{j-k} \\
&+ \Pr(H|s, X^o) \sum_{j=k}^{\overline{N}} f_N(j) C_k^j P_{50}(\mu_H, \sigma_H)^k (1 - P_{50}(\mu_H, \sigma_H))^{j-k}.
\end{aligned}$$

Unlike in our proof of Lemma 2, the mixing probabilities $\Pr(X^u|s, X^o)$ are not yet identified. The unknowns are $f_N(j)$, $\Pr(H|s, X^o)$, $P_{50}(\mu_L, \sigma_L)$ and $P_{50}(\mu_H, \sigma_H)$. Note that there are $\overline{N}$ independent restrictions for each $k \in \{0 \cdots \overline{N} - 1\}$ and there are $\overline{N} + 3$ unknowns $f_N(0) \cdots f_N(\overline{N}-1), \Pr(H|s, X^o), P_{50}(\mu_L, \sigma_L), P_{50}(\mu_H, \sigma_H)$. Now consider repeating the above exercise with a different s and X^o, say s' and $X^{o\prime}$ with associated mean and variance of return $(\mu'_{X^u}, \sigma'_{X^u})$. Then, this yields $\overline{N}$ additional equations. Because we assume that F_N is invariant to (μ, σ), we have a total of $2 \times \overline{N}$ equations and $\overline{N} + 4$ unknowns $f_N(0) \cdots f_N(\overline{N} - 1)$, $P_{50}(\mu_{X^u}, \sigma_{X^u})$, $P_{50}(\mu'_{X^u}, \sigma'_{X^u})$, $\Pr(H|s, X^o)$, $\Pr(H|s', X^{o\prime})$. Assuming that F_N is invariant to (μ, σ), we can increase the number of equations at a faster rate than the number of observables. Hence $\Pr(H|s, X^o)$, $P_{50}(\mu_{X^u}, \sigma_{X^u})$, and $f_N(0), \cdots,$ $f_N(\overline{N})$ are identified. The assumption that each listing receives at least 3 or more bids was used to identify $\Pr(H|s, X^o)$

in our proof of Proposition 4. Our discussion here shows that that assumption can be dispensed with.

Proof of Proposition 4 when There is Pooling Our identification argument in the main text focused on the case when there is no pooling. As long as we can identify $F_{\varepsilon|X}$ using the subset of the borrowers who are not pooled, we can identify $F_{\varphi|X}$ for the case of pooling as well. To see this, first note that we can identify $F_{\varphi|X}$ for borrowers who are not being pooled just as before. Now consider the terminal decision of the borrower who is being pooled:

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

where $F^{-1}_{\varphi|X}(\alpha^{pool})$ is a random variable with $\alpha^{pool} \sim U[0, m^{pool}]$ and m^{pool} is the fraction of borrowers who submit $s = 0.36$.[36] Note that $F^{-1}_{\varphi|X}(\alpha^{pool})$ is a random variable because we do not know the exact value of φ for pooled borrowers: We only know that φ is below $F^{-1}_{\varphi|X}(m^{pool})$. Given that the distribution of $\varepsilon_T + F^{-1}_{\varphi|X}(\alpha^{pool})$ can be identified and we have already identified the distribution of ε_T from markets with no pooling, it is immediate that we can identify the distribution of $F^{-1}_{\varphi|X}(\alpha^{pool})$ nonparametrically.

References

Altonji, J., Elder, T. & Taber, C. (2005), 'Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools', *Journal of Political Economy* **113**(1), 151–184.

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

Gijbels, I., Hall, P., Jones, M. C. & Koch, I. (2000), 'Tests for monotonicity of a regression mean with guaranteed level', *Biometrika* **87**(3), 663–673.

Heckman, J. (1976), 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models', *Annals of Economic and Social Measurement* **5**(4), 475–492.

LaLonde, R. J. (1986), 'Evaluating the econometric evaluations of training programs with experimental data', *The American Economic Review* **76**(4), 604–620.

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

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