Marketplace Lending: A New Banking Paradigm?*

Boris Vallée  
Harvard Business School  

Yao Zeng  
University of Washington  

June, 2018

Abstract

Marketplace lending relies on screening and information production by investors, a major deviation from the traditional banking paradigm. Theoretically, the participation of sophisticated investors improves screening outcomes and also creates adverse selection among investors. In maximizing loan volume, the platform trades off these two forces. As the platform develops, it optimally increases platform prescreening intensity but decreases information provision to investors. Using novel investor-level data, we find that sophisticated investors systematically outperform, and this outperformance shrinks when the platform reduces information provision to investors. Our findings shed light on the optimal distribution of information production in this new lending model. (JEL G21, G23, D82)

---

*We are grateful to Itay Goldstein, Wei Jiang, Andrew Karolyi (the Editors), Shimon Kogan, Gregor Matvos, Adair Morse (discussants), Chiara Farronato, Simon Gervais, David Scharfstein, Andrei Shleifer, Jeremy Stein, and Xiaoyan Zhang and three anonymous referees for helpful comments. We also thank seminar and conference participants at the RFS FinTech Workshops at Columbia University and Cornell Tech, NY Fed/NYU Stern Conference on Financial Intermediation, NYU Stern FinTech Conference, Texas Finance Festival, Bentley University, Chinese University of Hong Kong Shenzhen, London School of Economics, Peking University, Tsinghua PBC School of Finance, the University of Oregon, and the University of Washington. The authors have received in-kind support from LendingRobot for this project in the form of a proprietary data set of LendingRobot transactions. These data were provided without any requirements. We thank Yating Han, Anna Kruglova, Claire Lin, Kelly Lu, Lew Thorson, and, in particular, Botir Kobilov for excellent research assistance. All remaining errors are our own.
Lending marketplaces, also commonly referred to as peer-to-peer lending platforms, such as Lending Club, Prosper, and Funding Circle, have been rapidly gaining market share in consumer and small business lending over the last decade. This rapid development has important implications on the consumer lending market, and more broadly on banking.

Designed as a two-sided platform, a lending marketplace bears no balance sheet exposure to the issued loans but brings innovations to traditional banking on both the borrower and investor side. The innovation on the borrower side relies on low-cost information technology to collect standardized information from dispersed individual borrowers on a large scale. With such information, lending platforms prescreen loans in the sense that they use scalable algorithm to gauge the riskiness of the underlying loan applications, list some of them on the platform, and allocate the listed applications into risk buckets.

Innovation on the investor side—that is, the focus of this paper—is equally important and complementary. Instead of pooling loans and issuing safe, information-insensitive claims, platforms provide detailed information on each prescreened loan application to investors, relying on investors to further screen individual borrowers and to directly invest in individual loans. These investors include informationally sophisticated investors, both among retail and institutional investors, as well as passive and unsophisticated investors. These diverse investors fully bear loan risks and perform large-scale borrower screening, and the resultant joint information production between the platform and investors challenges the traditional role of banks as being the exclusive information producer on behalf of investors (Gorton and Pennacchi 1990; Dang et al. 2017).

The joint information production between the platform and investors of varying sophistication poses several research questions, which we address in this paper. First, are more sophisticated investors on lending platforms screening borrowers more intensively and thereby systematically outperforming less sophisticated investors? If so, how does sophisticated investor

---

1Loans issued by these platforms represented one-third of unsecured consumer loans volume in the United States in 2016. Their revenues are predicted to grow at a 20% yearly rate over the next 5 years (see IBIS World 2017).

2Currently, more than half of retail borrowers on lending platforms use marketplace loans to refinance existing loans or pay off credit balances. Although the refinancing market represents an attractive entry point for platforms to achieve scale rapidly, it does not represent their whole addressable market, which includes any form of consumer or small business loan.

3The platforms are typically segmented between retail and institutional investors and randomly allocate loans between the two pool of investors. Platforms in general offer a passive investing feature as well.
outperformance relate to the platform’s prescreening and information provision? Finally, given the heterogeneity of investors, what is the optimal platform design in terms of loan prescreening and information provision to investors to maximize volumes? Answering these questions is essential for understanding how platform and investor information production interact with each other. This understanding could further speak to the promises and pitfalls of marketplace lending.

Our study is also motivated by a puzzling event during the development of the marketplace lending industry. On November 7, 2014, Lending Club, the largest lending platform, removed half of the 100 variables on borrower characteristics that it previously provided to investors. This change was unanticipated and surprised many market participants, because it was the only investor-unfriendly move in Lending Club’s history.\footnote{We provide more relevant institutional details, including about this specific event in Section 1.2.} Given that information transparency is crucial for investor screening, what is the economic rationale behind this reduction of the information set provided to investors?

In addressing the previously mentioned questions, we develop a model for marketplace lending and test its predictions using a novel data set that includes borrower and investor data.

To start, we theoretically argue that informationally sophisticated investors actively use information provided by the platform to screen listed loans (beyond platform prescreening) and identify good loans to invest. In contrast, unsophisticated investors do not screen; they invest in a listed loan passively if the platform prescreens intensely enough so that they can break even on average, or they do not invest at all. Hence, sophisticated investors outperform unsophisticated ones. Because sophisticated investors can identify and finance good loans before unsophisticated investors invest, sophisticated investor participation helps boost the volume of loans financed on the platform when unsophisticated investors are reluctant to invest.

However, the heterogeneity in investor sophistication creates an endogenous adverse selection problem among investors, which can hurt volume. Because sophisticated investors can identify and finance good loans before unsophisticated investors invest, sophisticated investor participation lowers the average quality of loans eventually facing unsophisticated investors. Being aware of this adverse selection problem, unsophisticated investors require a higher interest rate, or
equivalently a lower loan price to break even, resulting in a higher prevailing interest rate on the platform. This higher interest rate (i.e., lower loan price) reduces the amount of loan applications on the platform, hurting volume. If adverse selection becomes too severe, unsophisticated investors may not break even and may exit the market as a whole, leading to even lower volume.

Hence, to maximize volume, the platform optimally trades off these positive and negative effects of sophisticated investor participation. When platform prescreening cost is initially high, the platform optimally chooses a low prescreening intensity but distributes more information to investors. This environment encourages sophisticated investor participation, boosting volume even if unsophisticated investors do not participate. When platform prescreening cost becomes low as the platform develops, it optimally reverses the policies by choosing a high prescreening intensity such that unsophisticated investors are willing to invest, but at the same time distributes less information to mitigate the adverse selection caused by sophisticated investors.

Testing the model predictions crucially relies on data of investors of heterogenous sophistication. Although borrower-level data are made public by the platforms, data on investor characteristics and their loan portfolios is not publicly available. Fortunately, we obtain a rich data set provided by LendingRobot, an algorithmic third party, which includes portfolio composition for a large set of retail investors on two largest lending platforms, Lending Club and Prosper. Therefore, we can study sophisticated investor screening and outperformance within the same investor segment and across platforms. Importantly, our sample includes a significant source of heterogeneity in terms of sophistication: some investors invest by themselves, whereas others rely on the various screening and order-placing technologies offered by LendingRobot.

Our empirical analysis progresses in several steps. First, we show that more sophisticated investors rely on different loan characteristics to screen the loans they finance, which points to their information advantage. Being selected by sophisticated investors predicts a significantly lower probability of default for a given loan, meaning that sophisticated investors systematically outperform unsophisticated investors over time and across all risk buckets. We find that loans

5Traditionally, the breakdown between retail and institutional investors represents a natural source of heterogeneity in terms of sophistication, and platform public data allow to identify which loans are sold to retail (fractional loans) or institutional investors (whole loans). However, this distinction is not informative in marketplace lending, as the allocation between the retail and institutional investor segments is randomized by platforms, and each segment itself also has a large heterogeneity of investor sophistication. Therefore, studying the impact of investor sophistication needs to be conducted within these segments.
selected by sophisticated investors have a default rate on average 3% lower than the average loan, or loans picked by unsophisticated investors, which corresponds to a reduction of more than 20% of the average default risk. Indeed, sophisticated investors produce information.

Using the 2014 Lending Club event described above, we then implement a difference-in-differences methodology to establish causal evidence of the impact of a large reduction in platform information provision on sophisticated investors’ performance. We find that sophisticated investor outperformance drops by more than half at the time of the reduction. We rationalize this unanticipated event, corresponding to the platform “evening the playing field,” by referring to the theoretical argument that platforms actively manage adverse selection. Under our rationalization, this event suggests that lending platforms value unsophisticated investors and therefore act as if they “protect” them, even in the absence of specific regulation.

Finally, we find that platforms’ prescreening intensity also has been improving in the sense that platform risk buckets are increasingly precise at predicting default. In addition to attracting unsophisticated investors directly, this increased precision is also likely to mitigate adverse selection by reducing the heterogeneity of loans within a given risk bucket. Consistent with platforms managing adverse selection, we also find robust time-series evidence that sophisticated investor outperformance has become lower in recent years.

Although our empirical tests mainly rely on the heterogeneity within the retail investor segment, our findings have external validity for the institutional investor segment of marketplace lending, as the heterogeneity in sophistication is comparable across investor segments. Many institutional investors, such as pension funds, only apply a little screening (for instance, only relying on a grade threshold) as retail investors do; while other institutional investors, such as hedge funds, develop highly sophisticated investment strategies that are comparable to what LendingRobot offers to investors.

We focus on the robust features in the development of marketplace lending so far while leave a number of interesting questions for future research. These include, for example, the overall welfare implications of marketplace lending, and whether marketplace lending poses any financial stability concerns due to the increasing participation of institutional investors.

Our paper contributes to the burgeoning literature on marketplace lending by directly exam-
ining the sharing of information production between platforms and investors, one key factor that makes marketplace lending special as a new banking model. So far, the literature of marketplace lending has mainly focused on how borrowers’ soft information improves lending outcomes (e.g., Duarte, Siegel, and Young 2012; Iyer et al. 2015; see Morse 2015 for a review). To the best of our knowledge, we are the first to study how investors’ characteristics affect loan screening outcome and how the participation of sophisticated investors interacts with the optimal platform design. Paravisini, Rappoport, and Ravina (2016) also use a sample of investor portfolio data from Lending Club in the 2007–2008 period, but they mainly test a classical asset pricing relationship between risk aversion and wealth rather than focusing on marketplace lending itself. On optimal platform design, our work complements a few recent papers that study the motivation behind Propser’s switch from an auction-based pricing mechanism to platform direct pricing in its early stage (Franks, Serrano-Velarde, and Sussman 2016), and its effect on borrowers learning about their cost of credit (Liskovich and Shaton 2017). Our work also complements that of Hildebrand, Puri, and Rocholl (2016), who use Prosper’s earlier policy change of removing origination rewards to examine how changes to investors’ incentives affect loan interest rates.

In the context of marketplace lending competing with traditional banking, de Roure, Pelizzon, and Thakor (2018) and Tang (2018) study whether these two types of lenders complement or substitute each other when providing credit to different borrower segments. Balyuk (2017) examines whether households improve their credit access from traditional banks when using marketplace lending. These papers support traditional banking as a benchmark to marketplace lending. Our work complements theirs by focusing on the investor side of marketplace lending and exploring how the role of these investors deviates from that of traditional bank depositors.

In addition, our paper also relates to the growing literature on shadow banking. The notion of shadow banking has many different aspects and interpretations, but in a recent review, Adrian and Ashcraft (2016) define shadow banks as nonbank financial institutions that conduct credit, maturity, and liquidity transformation by pooling and tranching, issuing safe claims, and tapping wholesale funding. In the shadow banking context, investors cannot outperform each other within a single tranche and cannot produce information on an individual borrower. Information

\footnote{For example, prior to the 2008 financial crisis, mortgage lenders created and sold debt-like agency mortgage-backed securities to money market mutual funds.}
production at the loan level is thus low, especially for investors in the safest tranches. In contrast, in the absence of pooling and tranching, investors on marketplaces have to fully bear the risk of the individual loans they select based on the loan-level information they can access. Consequently, any investor in marketplace lending, regardless of their risk preferences, has the potential to produce information at the borrower level and outperform other investors on a risk-adjusted basis. Consistent with this difference, Keys et al. (2010) find that securitization leads to loosened screening, while in marketplace lending this risk may be mitigated by the sharing of information production between the platforms and investors.

Marketplace lending belongs to the broader and burgeoning field of FinTech lending, which can be loosely defined as technology-enabled lending practices without direct human interactions. FinTech lending includes other online lenders, notably in the residential lending markets, which bear some balance sheet risk and rely on wholesale funding models closer to the precrisis originate-to-distribute model. Recently, Buchak et al. (2017) find that online lenders account for almost a third of nonbank mortgage originations by 2015, and Fuster et al. (2018) find that they process mortgage applications faster than traditional banks and better adapt to variations in the volume of mortgage applications.

Finally, our paper is related to the theoretical literature on information acquisition by banks and investors (Hauswald and Marquez 2003, 2006), in particular studies on the resultant endogenous adverse selection (Glode, Green, and Lowery 2012; Biais, Foucault, and Moinas 2015, for example). Closer to us is the work by Fishman and Parker (2015), Bolton, Santos, and Scheinkman (2016), and Yang and Zeng (2017), who consider a related trade-off: investor information acquisition helps guide efficient production but also introduces adverse selection that may lower gains from trade. The contribution of our paper is to embed this trade-off in an outer-level optimal platform design problem, where the project supply is endogenous. Although our study is conducted for marketplace lending, it sheds light on other financial market design contexts in which adverse selection may be a concern, as suggested in Rochet and Tirole (2006).
1 Institutional Details of Marketplace Lending

In this section, we describe several key aspects of marketplace lending that are relevant to our study: platform information collection, platform prescreening, the funding model (including both investor screening and platform information distribution), and changes in investor composition. We refer interested readers to the review of Morse (2015) for general institutional details of lending marketplaces.\textsuperscript{7}

1.1 Platform information collection and prescreening

1.1.1 Information collection

By design, lending platforms only collect information on borrowers via online self-reporting and through credit pulls. Thus, information collection itself is standardized and mostly costless.\textsuperscript{8} A fraction of the self-reported information is verified by the platform by requiring supporting documents. Under the current practice, borrowers and platform employees do not personally interact.

1.1.2 Prescreening

Armed with the information they collect, platforms perform loan prescreening on two margins. On the extensive margin, they decide to accept or reject the application; an accepted application is subsequently listed on the platform and made available to investors. On the intensive margin, they allocate the listed loan to a risk bucket, called a “grade” or a “subgrade.” Currently, Prosper classifies its listed loans into 7 grades, whereas Lending Club uses a scoring system of 35 subgrades. These risk buckets map into interest rates, that is, loan prices. For a given loan maturity and at any given point in time, a risk bucket is associated with a single interest rate

\textsuperscript{\textsuperscript{7}}Marketplace Lending received important coverage in 2016, following some governance issues at the main platform, Lending Club (see Rudegeair and Andriotis 2016). Although these events revealed serious governance issue at Lending Club, they do not speak to the marketplace lending economic model.

\textsuperscript{\textsuperscript{8}}At the inception of marketplace lending, platforms frequently collected soft information, that is, nonstandardized answers to questions, such as a description of the project to be financed, or even pictures, which were an important part of the “peer-to-peer” aspect that was initially supported. Since the early years, platforms have progressively stopped collecting soft information to standardize the information set and to streamline the application process. Collecting more information, especially if not standardized or raising privacy issues, would increase the drop-off rate during the online application, which would effectively make information collection costly for the platform.
Platform prescreening is scalable but costly. It is not equivalent to simply picking a FICO score threshold and listing loans above that threshold. Rather, it may potentially use many variables. The development of such a prescreening model requires sophisticated data analysis, which involves a fixed cost. The more precise the allocation of loans into risk buckets is, including those that are not listed, the costlier the screening model is to develop.

The platform’s screening model also evolves over time as the platform learns from the growing pool of loan applications that are listed and loans that are financed. The increasing data available to platforms suggests that the cost associated with prescreening decreases over time.

1.2 Marketplace funding model and information distribution

1.2.1 Funding model

The funding model of marketplace lending heavily relies on investors screening and investing in loans individually, as is the case for loans issued on Lending Club or Prosper.\textsuperscript{9} This represents a deviation from the traditional banking model in which depositors hold a demandable debt contract issued by the bank without having any knowledge of the underlying loans (Gorton and Pennacchi 1990; Dang et al. 2017).\textsuperscript{10,11}

1.2.2 Information distribution

Platforms provide investors with a set of standardized information for each listed loan. This information is costless to investors but typically a subset of the information that the platform collects. Platforms distribute this information both on their Web sites and through their Application Programming Interfaces (APIs), which are customized programming protocols that allow more sophisticated investors to develop their own automated algorithms to place orders.

\textsuperscript{9}More precisely, individual investors purchase notes that are backed by loans. See Morse (2015) for a detailed description of this process.

\textsuperscript{10}A few other FinTech firms follow a different model, with some players keeping the loans on their balance sheet, while obtaining wholesale funding from another institution, as OnDeck does, or by implementing tranched securitization on a large pool of loans. These platforms, while doing online lending and collecting information as previously described, are not creating marketplaces per se and are closer to the shadow banking sector. Some online lenders follow a hybrid model, such as Avant (both marketplace and balance sheet funding) or SoFi (both balance sheet funding and securitization). Lending Club and Prosper jointly captured around half of the lending activity by FinTech firms in 2016, whereas the third largest player, SoFi, only captured less than 7% of the market.

\textsuperscript{11}Lending Club also experimented with securitization, but it remains marginal in its funding model.
This provision of information has two main purposes. First, it allows investors to check the quality of the loan they purchase. Second, information provision makes it possible for investors to further screen loans if they choose to do so. These two characteristics are also different from traditional banking or modern securitization as studied in the literature, where no information or only aggregate information is provided to end investors. The choice of the information set provided to investors has a direct impact on their screening cost, as investors have to exert more effort to identify good loans when the information they access is more restricted.

1.2.3 Lending Club change in investor information set

On November 7, 2014, Lending Club removed half of the 100 variables on borrowers’ characteristics that it shared with investors previously. This removal affected new loan listings available on the Web site, listed loan information available through the API, as well as historical data.

This change was unanticipated. Anecdotal evidence suggests that it was motivated by Lending Club’s desire to “even the playing field between investors.” While Lending Club and other platforms regularly adjust the number of available variables, this negative change in information set is unique by its magnitude.

1.2.4 Role of investor screening in marketplace lending

Investor screening plays an important role on lending platforms as it directly impacts whether a loan eventually will be funded. In turn, this funding model allows the platforms to better calibrate their prescreening intensity and loan pricing by monitoring investors’ funding activities. Platforms also prescreen and price listed loans in expectation of investors’ ultimate investing decisions. Investor screening therefore impacts the way the platform allocates loans to risk buckets, as well as the interest rates that the platform attributes to each risk bucket to clear the market. These interest rates have evolved over time, as Figure 2 illustrates.

13 If total commitments to a loan by the expiration date are less than the requested amount but are above a certain threshold, the borrower can accept that lesser amount or withdraw the loan request.
14 This current setting is an evolution from the earlier practice of platforms to run auctions for setting the interest rate, which created liquidity issues. See Franks, Serrano-Velarde, and Sussman (2016) and Liskovich and Shaton (2017) for a discussion on the auction model.
15 See the discussion on this point in Lending Club Investor Day 2017 Presentation, which is available at https://ir.lendingclub.com/Cache/1001230258.pdf.
1.3 Investor composition

Lending platforms initially targeted retail investors only but increasingly opened up to institutional investors. Today, most lending platforms, including Lending Club, Prosper, and Funding Circle, target both retail and institutional investors. Typically, institutional investors purchase whole listed loans, while retail investors invest in a fractional loan, meaning that each loan is divided into small USD 25 notes that bear the credit risk of the loan. This process allows retail investors to diversify interest risks with a small portfolio size, encouraging their participation.

The past a few years have witnessed a significant upward trend of institutional and generally informationally sophisticated investor participation. Morse (2015) suggests that the share of institutional investors on lending platforms has increased from less than 10% to more than 80% since 2012. The Lending Club 2017 Investor Day Presentation also indicates the following breakdown of investors: banks and other institutional investors (55%), managed accounts (29%), and self-directed retail investors (13%).

Within each segment of investors, the heterogeneity of investor sophistication has been increasing as well. For example, sophisticated third parties, such as LendingRobot, NSR Invest (serving retail investors), and Orchard (serving institutional investors) are increasingly providing investors with algorithmic tools to screen loan and automatically execute orders. A number of hedge funds have also publicly announced that they are investing on lending platforms (see Johnson 2014).

Within each segment, investors therefore can be further classified into two types: unsophisticated investors that buy loans passively or using self-directed rules, and sophisticated investors who actively produce information, screen loans using a proprietary screening model, and execute orders at high speed. Consistent with this trend of more informationally sophisticated investor participating, the speed for funding loans has increased dramatically, with the most popular loans being funded in seconds (see Figure D.1 in the appendix).

16Managed accounts are passive investment vehicles that are distributed by conventional third-party marketers to high net worth individuals.
2 Theoretical Framework

2.1 The model

2.1.1 Environment

Four types of risk-neutral agents are present in our model: (1) a platform, (2) a continuum of \( x_0 \) of loan applicants, each of which has a project and can submit a loan application, (3) a mass \( \Omega \) of sophisticated investors, and (4) a competitive fringe of unsophisticated investors.\(^{17}\) There are three dates, \( t = 0, 1, 2 \), with no time discount. The lending platform prescreens loan applications at \( t = 0 \). The applications that are screened in are listed on the platform at \( t = 1 \), and investors decide whether to further screen and subsequently buy these listed loans. Loan cash flows are realized at \( t = 2 \). We specify agents’ objectives and strategies below.

2.1.2 Loan applications

Each penniless loan applicant has one project that requires an initial investment \( I \). The project, if funded, pays \( R = R_H \) with probability \( \pi_0 \) (“good project”) and \( R = R_L \) with probability \( 1 - \pi_0 \) (“bad project”) at \( t = 2 \), where \( \pi_0 \in (0, 1) \) and \( R_L < I < R_H \). Following the banking literature that focuses on bank information production, assuming that applicants do not know their type ex ante allows us to focus on adverse selection on the investor side.\(^{18}\)

The number of loan applications \( x_0 \) is endogenous, and depends on the equilibrium price \( p \geq I \) on the platform. We assume that the supply curve of applications \( x_0(p) \) is an increasing function of \( p \), meaning that a higher price (i.e., a lower interest rate) attracts more applications. Because borrowers do not know their types ex ante, it is also natural to assume that \( \pi_0 \) does not depend on \( p \) as a benchmark.\(^{19}\) We provide a microfoundation for this reduced-form supply

---

\(^{17}\)In reality, investor-level information sophistication is distributed across a wide spectrum. Our model represents a useful benchmark that captures the heterogeneity in information sophistication.

\(^{18}\)Prominent examples are Hauswald and Marquez (2003, 2006), who are among the first to model bank information acquisition explicitly in a bank competition framework. In their models, borrowers can be good or bad, but they do not know their types ex ante. More recent examples in the theoretical literature include Glode, Green, and Lowery (2012) and Fishman and Parker (2015), in both a seller’s asset can be good or bad, a buyer can acquire information about it, whereas the seller itself does not know its type ex ante. Empirically, Liskovich and Shaton (2017) also find that many borrowers in the marketplace lending context do not know their cost of credit ex ante. Moreover, since our model features one risk bucket, we only need to assume that borrowers do not know their quality relative to the other loans of the same risk bucket.

\(^{19}\)We are aware of the classic borrower adverse selection issues in the banking context, but whether a lower interest rate attracts relatively better or worse applicants is still an open question in the marketplace lending
curve in Appendix B, which follows standard specifications in the literature.

### 2.1.3 Platform prescreening, pricing, and information distribution

The platform’s strategy is a triple \( \{ \pi_p, p, \mu \} \), which we specify in order below.

First, the platform prescreens the pool of loan applications and lists some of them at \( t = 0 \). Specifically, the platform chooses the interim probability \( \pi_p \) of a listed loan being good at \( t = 1 \) before any further investor screening, where \( \pi_p \in [\pi_0, 1] \). A higher \( \pi_p \) implies that the platform produces more information, and the average listed loan is more likely to be good and is indicative of a higher platform prescreening intensity. We assume that the platform can always screen in a good project but may fail to screen out a bad project.\(^{20}\) Accordingly, the number of loan applications being screened in and listed on the platform is

\[
x_p = \frac{\pi_0}{\pi_p} x_0 \leq x_0.
\]

Although we model only one risk bucket on the platform for simplicity, a recursive interpretation of our model covers platform prescreening and subsequent allocation of listed loans into multiple risk buckets. A higher \( \pi_p \) implies that loan applications unfit for the safest risk bucket will be more accurately screened out. These screened-out applications are prescreened again to see if they are fit for the second safest bucket, and again unfit loans will be more accurately screened out, and so forth.\(^{21}\) Therefore, a higher \( \pi_p \) also implies that, empirically, the classification of loan applications into risk buckets is more predictive of loan default risks.

Platform prescreening is costly. Since platform’s prescreening is implemented through scalable algorithms, we assume that prescreening cost increases in \( \pi_p \), but not on the number of applications processed. We take a parametric form \( C(\pi_p) = \frac{1}{2} \kappa (\pi_p - \pi_0)^2 \) where \( \kappa \geq 0 \) to facilitate closed-form solutions. A higher \( \kappa \) implies a higher prescreening cost.

\(^{20}\)This modeling choice is a reduced-form approach to parsimoniously capture the outcome, rather than the detailed process, of platform prescreening. It makes the model analytically tractable while is still general enough to capture all the possible prescreening outcomes in reality: \( \pi_p = \pi_0 \) means that the platform simply lists all the loan applications without any prescreening, and therefore \( x_p = x_0 \), whereas \( \pi_p = 1 \) means that the platform perfectly screens out all the bad projects and all the \( \pi_0 x_0 \) good projects are screened in.

\(^{21}\)Formally modeling this recursive process requires at least three types of loan applications. Because it does not affect the qualitative predictions of our model, we abstract away from this aspect for theoretical clarify.
Second, after prescreening, the platform assigns an interest rate to listed loans, which is modeled by a price $p$. Intuitively, a higher loan price $p$ means a lower interest rate in reality. In equilibrium, the platform holds rational expectations and thus $p$ will be determined by the marginal investor’s offer price, which we detail below.\(^{22}\)

Third, the platform distributes variables of listed loans to investors, which effectively determines sophisticated investors’ information cost $\mu \geq 0$ in further screening the listed loans. Since these variables are already provided to the platform by the loan applicants, we assume that it is costless for the platform to change $\mu$. As in reality, platforms do not collect any direct revenues from distributing information to investors regardless of $\mu$.

Finally, the platform’s objective is to maximize the expected volume of eventually financed loans by investors on the platform, regardless of their type, minus its prescreening cost. This objective function is motivated by the compensation scheme of platforms, which are typically a percentage of volume. Notably, the volume may be different from the amount of listed loans $x_p$, because some listed loans may not be financed by investors in equilibrium.

\subsection*{2.1.4 Sophisticated investors}

Sophisticated investors maximize expected profits. Each sophisticated investor may screen and buy at most one listed loan on the platform at $t = 1$. Her strategy is to choose whether to become informed and to offer a (latent) loan price $p_i$. An sophisticated investor has two potential advantages over an unsophisticated investor: information and speed, consistent with the discussion in Section 1 that sophisticated investors have an edge in both aspects. In Appendix C, we consider alternative specifications to illustrate that both advantages play roles in driving our results.

First, a sophisticated investor can purchase an information technology at cost $\mu$, which is set by the platform. This technology, capturing a screening algorithm in reality, allows her to become perfectly informed of one listed loan at $t = 1$.\(^{23}\) We denote by $\omega$ the population of

\(^{22}\)This modeling choice of having both the platform price and the investor price captures both the current practice that the platforms set the interest rate in expectation of investors’ financing decisions and the earlier practice of platforms running auctions among investors for setting the interest rate.

\(^{23}\)For tractability, we assume that an informed sophisticated investor is able to know the true type of one listed loan. This assumption can be easily motivated by the various capital constraints that the investors face. Our qualitative predictions would not change if we instead assumed that the informed sophisticated investor can screen all listed loans but only receives a noisy signal about each loan.
sophisticated investors who become informed, where $0 \leq \omega \leq \Omega$. When becoming informed, a sophisticated investor passes on a bad listed loan but offers price $p_i$ to buy a good one.

Second, sophisticated investors, when becoming informed, can screen and buy loans before uninformed investors do. This is consistent with informed investors use faster technology to place orders through the API. The screening outcome obtained by an informed investor is nonverifiable and nontransferable.

If a sophisticated investor chooses not to pay $\mu$, she remains uninformed and is essentially identical to an unsophisticated investor in equilibrium.

### 2.1.5 Unsophisticated investors

Unsophisticated investors also maximize expected profits. Each unsophisticated investor may also buy at most one loan on the platform at $t = 1$. An unsophisticated investor cannot become informed and may buy a loan only after informed investors move. Thus, his strategy is to offer a (latent) loan price $p_u$ given his updated belief, which determines his financing decision. Especially, $p_u = 0$ suggests that unsophisticated investors cannot break even and thus do not participate on the platform. Because unsophisticated investors are competitive, they are the marginal investors on the platform if they participate. Otherwise, informed sophisticated investors are the marginal investors.

### 2.1.6 Equilibrium definition

We formally define a sequential equilibrium as follows:

**Definition 1.** Given $R_H$, $R_L$, $I$ and $\pi_0$, the sequential equilibrium is defined as a collection of $\{\pi_p, \mu, p, p_i, p_u, \omega\}$ such that:\footnote{More formally, each element of the collection is defined on its corresponding information set; we omit the detailed specification of the information sets for simplicity.}

1. Given $\pi_p$, $\mu$ and $\omega$, the price $p_u$ gives uninformed investors expected zero profit.
2. Given $\pi_p$, $\mu$ and $\omega$, the price $p_i$ maximizes the expected profit of informed investors.
3. Given $\pi_p$ and $\mu$, the population $\omega$ of sophisticated investors find it optimal to acquire the information technology and become informed.
4. The platform’s choices of \( \pi_p \) and \( \mu \) maximize the expected volume of financed loans minus its prescreening cost.

5. The platform’s choice of \( p \) satisfies rational expectations in the sense that it is equal to the marginal investor’s offer price:

\[
p = \begin{cases} 
  p_i, & \text{if } p_u = 0, \\
  p_u, & \text{if } p_u > 0.
\end{cases}
\]

6. Agents use the Bayes’ rule to update their beliefs and follow sequential rationality.

### 2.2 Equilibrium Analysis

#### 2.2.1 Uninformed investors

We derive the equilibrium by backward induction. We first consider uninformed investors’ financing decision on the platform, under any given generic \( x_p, \pi_p, \mu \), as well as the population of informed sophisticated investors \( 0 \leq \omega \leq \Omega \).

Consider the pool of listed loans facing uninformed investors. Uninformed investors move after the \( \omega \) informed investors screen \( \omega \) listed loans and potentially finance \( \pi_p \omega \) good loans. Hence, uninformed investors’ updated posterior belief \( \pi'_p \) of a listed loan being good is

\[
\pi'_p(\omega) = \frac{\pi_p(x_p - \omega) + \frac{(1 - \pi_p)\omega + (x_p - \omega)}{x_p}}{\omega, } \leq \pi_p,
\]  
(2.1)

where \( (\cdot)_+ = \max\{0, \cdot\} \). As \( \pi'_p(\omega) \) is decreasing in \( \omega \), uninformed investors’ posterior expected value of a listed loan in this pool is as follows and also decreases in \( \omega \):

\[
V'(\omega) = \frac{\pi_p(x_p - \omega) + R_H + (1 - \pi_p)(\omega + (x_p - \omega))}{1 + \pi_p\omega + (x_p - \omega)} \cdot R_L.
\]  
(2.2)

Because uninformed investors are competitive, they receive zero profit in equilibrium. Therefore, they participate if and only if they can meet the investment requirement, while still breaking even:

\[
V'(\omega) \geq I,
\]  
(2.3)
and the price $p_u$ they offer is

$$p_u(\omega) = \begin{cases} V'(\omega), & \text{if } V'(\omega) \geq I, \\ 0, & \text{if } V'(\omega) < I. \end{cases} \quad (2.4)$$

The fact that both $\pi_p'(\omega)$ and $V'(\omega)$ are decreasing in $\omega$ indicates an important endogenous adverse selection problem, introduced by more sophisticated investors becoming informed. As more informed investors pick up more good loans from the platform, the pool of listed loans facing uninformed investors becomes less valuable. Thus, uninformed investors may leave the market, hurting volume. As will be shown later, this adverse selection problem also leads to fewer loan applications, hurting volume even if uninformed investors still participate.

### 2.2.2 Informed investors

Sophisticated investors, if becoming informed, buy good loans before uninformed investors can invest. Thus, informed investors’ optimal offer price to an identified good loan is the loan applicant’s outside option, depending on uninformed investors’ participation decision:

$$p_i(\omega) = \begin{cases} p_u(\omega) = V'(\omega), & \text{if } V'(\omega) \geq I, \\ I, & \text{if } V'(\omega) < I, \end{cases} \quad (2.5)$$

where $V'(\omega)$ is determined in (2.2). Because the value of a good loan $R_H$ is always higher than sophisticated investors’ offer price $p_i(\omega)$, sophisticated investors enjoy a positive information rent. This implies that informed sophisticated investors outperform unsophisticated ones.

As long as there are good listed loans available, a sophisticated investor becomes informed if and only if the unconditional information rent exceeds the information cost set by the platform:

$$\pi_p(R_H - p_i(\omega)) \geq \mu, \quad (2.6)$$

which is more (less) likely to be satisfied when the information cost $\mu$ is lower (higher).
2.2.3 Platform

Under Definition 1, the platform price is $p = 0$ when no investor participates and is pinned down by the marginal investor’s (latent) price when some investors participate:

$$p(\omega) = \begin{cases} 
  p_u(\omega) = V'(\omega), & \text{if } V'(\omega) \geq I, \\
  p_1(\omega) = I, & \text{if } V'(\omega) < I,
\end{cases}$$

(2.7)

where $V'(\omega)$ is determined in (2.2), and, thus, $p(\omega)$ is a decreasing function of $\omega$. Intuitively, unsophisticated investors are the marginal investor when they participate; otherwise, informed sophisticated investors are the marginal investor.

Having characterized investors’ participation and pricing decisions in the subgame, we solve for the platform’s optimal prescreening and information distribution policies $\{\pi_p, \mu\}$.

**Theorem 1.** There exists two thresholds of platform prescreening cost $0 < \kappa \leq \overline{\kappa}$ such that:

1. If $\kappa \geq \overline{\kappa}$, the platform optimally chooses a low $\pi_p$ and a low $\mu$. In this case, sophisticated investors become informed and they invest, while unsophisticated investors do not participate. The volume is $\min\{\pi_0 x_0(I), \pi_p \Omega\}$.

2. If $\kappa \leq \underline{\kappa}$, the platform optimally chooses a high $\pi_p$ and a high $\mu$, where $\pi_{p*} > \pi_p$ and $\mu > \underline{\mu}$. In this case, sophisticated investors do not become informed, while all uninformed investors participate. The volume is $\frac{\pi_0 x_0(p(0))}{p_p}$.

3. If $\underline{\kappa} < \kappa < \overline{\kappa}$, the platform optimally choose either $\{\pi_p, \mu\}$ or $\{\pi_{p*}, \overline{\mu}\}$, depending on which case gives the platform a higher expected payoff.

The mathematical expressions of the thresholds $\underline{\kappa}$, $\overline{\kappa}$, and the optimal policies $\pi_{p*}$, $\mu$, $\pi_p$, and $\overline{\mu}$ are given in Appendix A.

The following proposition is a direct time-series implication of Theorem 1:

**Proposition 1.** As prescreening cost $\kappa$ goes from high to low over time, the platform optimally:

a. increases prescreening intensity $\pi_p$, but $\pi_p < 1$ even if $\kappa = 0$;

b. increases investor information cost $\mu$, that is, distributes less information.

Theorem 1 and, in particular, Proposition 1 speak to the optimal platform design as the
platform’s prescreening cost becomes lower over its life cycle.\textsuperscript{25} Providing less information to
investors may seem surprising at first given that marketplace lending relies on investor screening.

To better understand the results and the underlying economic trade-offs, we discuss the first
two generic equilibrium cases of Theorem 1 in detail by considering what happens if the platform
deviate from its optimal strategies prescribed by Theorem 1. To do this, we formally show in
the proof of Theorem 1 (in Appendix A) that under any generic platform policy \( \{\pi_p, \mu\} \), the
economy must be in one of the following four types of subgame equilibrium. The respective
volume in each type of subgame equilibrium is given as follows:

<table>
<thead>
<tr>
<th>Equilibrium investor participation and platform volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ( \mu )</td>
</tr>
<tr>
<td>Low ( \pi_p )</td>
</tr>
<tr>
<td>High ( \pi_p )</td>
</tr>
</tbody>
</table>

where we call the top-left equilibrium type 1 equilibrium, top-right type 2 equilibrium, bottom-right type 3 equilibrium, and bottom-left type 4 equilibrium. The equilibrium volumes satisfy

\[
0 < \min\{\pi_0x_0(I), \pi_p\Omega\} < \frac{\pi_0x_0(p(\Omega))}{\pi_p} < \frac{\pi_0x_0(p(0))}{\pi_p}.
\] (2.8)

According to Theorem 1, a type 2 (type 4) equilibrium, where both the optimal \( \pi_p \) and \( \mu \) are
low (high), happens when platform prescreening cost is high (low).

First, we consider the trade-off underlying case (a) in Theorem 1 when \( \kappa \) is high. This implies
that the platform optimally chooses a low prescreening intensity \( \pi_p \) to save its cost. Because
the resultant quality of an average listed loan is low, unsophisticated investors cannot break
even and thus will not participate according to condition (2.3). If the platform chose a high
\( \mu \), that is, provided little information to sophisticated investors, sophisticated investors would
find it too difficult to become informed and thus stay out of the market as well. Hence, the
economy would be in a type 1 equilibrium, where no investor participates and the volume is

\textsuperscript{25} Platform prescreening cost decreases over time because of improvements in information technologies
and screening algorithms. See the Lending Club Investor Day 2017 Presentation available at ir.Lending
accordingly 0. To improve screening efficiency, the platform optimally chooses a low \( \mu \) to attract sophisticated investors to become informed, according to their participation condition (2.6). In this case, informed sophisticated investors become the marginal investor and the platform price is \( p = I \). They identify and buy good loans up to their capacity, helping the platform boost the volume to either \( \pi_0 x_0(I) \) or \( \pi_p \Omega \), whichever is larger.\(^{26}\) The economy is in a type 2 equilibrium, where the screening efficiency concern dominates.

Next, we consider case (b) in Theorem 1 when \( \kappa \) is low. By condition (2.3), the platform optimally chooses a high prescreening intensity \( \pi_p \) to attract uninformed, unsophisticated investors. Accordingly, uninformed investors become the marginal investor. If the platform chose a low \( \mu \), that is, provided much information to sophisticated investors, they would become informed and thus buy good loans earlier than uninformed investors. Notice that this strategy would introduce adverse selection: the quality of an average loan facing uninformed (and marginal) investors would become lower, leading to a lower platform price \( p(\Omega) \). This would in turn lead to a lower amount of loan applications \( x_0(p(\Omega)) \), driving the economy into a type 3 equilibrium and hurting volume. If adverse selection were severe enough, it would deter uninformed investors from participating at all. Hence, to eliminate adverse selection, the platform optimally chooses a high \( \mu \) so that sophisticated investors do not become informed, according to condition (2.6). In this case, all the investors are uninformed, leading to a higher platform price \( p(0) \). This implies a higher amount of loan applications \( x_0(p(0)) \). Ultimately, all the listed loans \( \frac{\pi_0 x_0(p(0))}{\pi_p} \) will be financed on the platform, yielding the highest possible volume for the platform. The economy is in a type 4 equilibrium, where the adverse selection concern dominates.

We note that, the type 3 equilibrium, where sophisticated investors become informed and participate while uninformed unsophisticated investors also participate, reflects a transition phase from a type 2 to a type 4 equilibrium. In this transition phase, the platform still provides some information to sophisticated investors, which is inherited from a type 2 equilibrium. All the listed loans are financed, but adverse selection is active. This transition phase may capture the current developing stage of some platforms in which platforms start to discourage sophisticated

\(^{26}\) Although our baseline model takes a reduced-form supply curve of loan applications regardless of the type, this result is robust in a possible extension, in which only good loan applicants apply when the marginal investor is informed (because bad loan applications know that their applications will be identified and rejected for sure). In this case, the platform does not prescreen, and the platform price is still \( p = I \), only \( \min\{\pi_0 x_0(I), \pi_p \Omega\} \) good loan applicants will apply and then be listed, while the volume eventually financed is still \( \min\{\pi_0 x_0(I), \pi_p \Omega\} \).
2.3 Empirical predictions

Our model generates a set of empirical predictions, which we subsequently bring to the data.

**Prediction 1.** Sophisticated (and informed) investors outperform unsophisticated (and uninformed) investors at any loan price.

**Prediction 2.** When their information cost becomes higher, sophisticated investors are less likely to become informed and thus their outperformance may shrink, at any loan price.

Predictions 1 and 2 focus on investors and derive from the subgame equilibrium, where the platform’s strategy is considered to be a parameter from the investors’ perspective. They speak to our first two research questions on sophisticated investor information production and its response to platform policies. Specifically, the two predictions come from conditions (2.5) and (2.6) of the model. Although they are theoretically straightforward, it is not empirically obvious whether there is any room for sophisticated investors to screen and outperform, and if so, to what extent the outperformance responds to platform policy changes. Moreover, in the traditional banking model, depositors, however sophisticated they are, are unlikely to outperform others within the deposit market only. Thus, empirically observing sophisticated investor outperformance serves as a smoking gun of marketplace lending being a different banking model.

**Prediction 3.** The platform may increase the information cost of sophisticated investors as it develops by distributing fewer variables to investors.

**Prediction 4.** The platform may increase its prescreening intensity as it develops, but it will eventually keep an intermediate prescreening intensity.

Predictions 3 and 4 focus on optimal platform design and derive from the full equilibrium. They are derived from Proposition 1, a direct time-series implication of Theorem 1, and speak to our third research question regarding the optimally joint information production. Beyond 27

---

27Theoretically, this transition phase also can be sustained as a full equilibrium in a potential model extension, where the platform actively learn from sophisticated investors’ screening criteria to further reduce its own prescreening cost, that is, where $\kappa(\omega)$ is decreasing in $\omega$. 

---

20
the key intuition of managing investor adverse selection, which we have articulated above, an optimally intermediate prescreening intensity is justified by the fact that as long as no investor becomes informed and all uninformed investors participate, further increasing $\pi_p$ only hurts volume.

3 Data and Investor Sophistication

3.1 Data

Our investor-level data are provided by LendingRobot, a leading robo-advisor for retail investors on lending marketplaces.\textsuperscript{28} The data covers all transactions executed by LendingRobot users between January 2014 and February 2017, which represents more than $120$ million invested on the two major lending platforms, Lending Club and Prosper, and all historic transactions from portfolios monitored by LendingRobot. We combine these data with the publicly available loan-level data from Lending Club and Prosper.

LendingRobot provides an automated investment tool for its clients that relies on a sophisticated screening model calibrated on historical data from the platforms.\textsuperscript{29} This tool also allows to execute orders at high speed through an API. Independently, LendingRobot also offers a free monitoring tool that helps investors monitor their portfolios on Lending Club and Prosper without automated investing. Thus, we observe both portfolios invested under the help of LendingRobot’s investment tool, as well as portfolios built by investors themselves.\textsuperscript{30}

The LendingRobot data are organized at the investor level, as shown in Figure 1. We access a set of variables at each level of this data structure.

- **User**: Each user represents a distinct physical investor.

- **Account**: A user can have one or several accounts. An account represents a portfolio of notes that an investor holds on a single lending platform: Lending Club or Prosper.

There are three types of accounts in our data. A monitor-only account is an account in

\textsuperscript{28}LendingRobot was acquired by NSR Invest, its main competitor, in August 2017.

\textsuperscript{29}For more details on LendingRobot screening model, refer to http://blog.lendingrobot.com/research/predicting-the-number-of-payments-in-peer-lending/.

\textsuperscript{30}Investors who use the monitoring tool may only rely on Lending Club’s and Prosper’s passive investment feature or on their own discretion.
which LendingRobot only monitors the portfolio but does not screen or buy notes through its technology. In a robot account, notes screening and purchase are executed by the LendingRobot investment tool. In an advanced account, investors can further implement their own screening criteria, combined or not with LendingRobot’s screening model, and rely on LendingRobot to automatically execute the orders when relevant loans appear on the platform.

- **Note**: Retail investors invest in notes, each of which is backed by a single loan. The information available at the note level is its nominal value, which is 25 USD or 50 USD, and the underlying loan identifier.

- **Loan**: Each loan is associated with a large set of financial characteristics of the borrower at the loan issuance, made public by lending platforms. These variables include loan amount, FICO score, debt-to-income ratio, employment length, three-digit ZIP code, and many others. For brevity, we refer interested readers to Morse (2015) for the description of loan-level data, which is standard in the literature. We observe these variables for all the loans issued by Lending Club and Prosper.
3.2 Investor sophistication

Our data provides us with portfolios of retail investors with heterogeneous levels of sophistication, which is key to our study.

Within our data, monitor-only accounts are the less sophisticated ones as they do not implement LendingRobot’s investment tool with information and speed advantages. We also consider the average retail investor on a given platform, which we can observe by collapsing the whole platform borrowers’ data, as unsophisticated.

Robot accounts represent sophisticated investors, as they invest through the LendingRobot investment tool which provides both screening and automated execution at high speed.

Advanced accounts also represent sophisticated investors. While some advanced accounts might be potentially even more sophisticated than robot accounts, advanced accounts are however at risk of making mistakes compared to the robot benchmark, as they rely at least partly on their own screening criteria, which are automatically executed by LendingRobot. For the purpose of our analysis, we consider advanced accounts as a different type of sophisticated investors without specifying whether they are more or less sophisticated than the robot accounts.

While our data focuses on the segment of retail investors, the heterogeneity in sophistication within this segment is arguably comparable to the heterogeneity of sophistication within the institutional investor segment. As discussed in the institutional background, traditional mutual and pension funds usually buy loans on the platforms passively while hedge funds screen loans actively before investing. In addition, LendingRobot indeed performs hedge-fund-like order execution for its robot and advanced accounts.

3.3 Summary statistics

Table 1 provides the summary statistics. Our sample is representative: the difference between Lending Club and Prosper accounts on LendingRobot, in terms of both the number of accounts

---

31 Monitored-only investors are selected on having registered to LendingRobot, a fact that might suggest a higher level of sophistication compared with fully naive investors. However, this potential selection effect can only bias against finding differences between portfolios of monitored-only and robot investors in our empirical analysis. Such a bias would only weaken, not strengthen, our outperformance results.

32 Per a discussion with LendingRobot, a few of the investors with an advanced account are institutional investors, such as family offices, but the majority are individual investors. LendingRobot displays a warning to investors: that they must proceed at their own risk when selecting an advanced account.

---
and the amount invested, is comparable to the difference between the two platforms overall.

On portfolio size, while robot accounts are the most representative type of accounts, advanced accounts are, on average, larger. The distribution of portfolio size is skewed, with a few investors having invested more than one million dollars, driving the average amount invested significantly above the median. Portfolio size should not be interpreted as a proxy for sophistication, however, as any robot account invests automatically along the same criteria, independently of its size.

On portfolio risk, accounts are on average modestly tilted towards riskier loans compared to the overall platform average, as exhibited by a higher average interest rate of portfolios.

[Insert Table 1]

4 Empirical Analysis

4.1 Investor screening

Our theoretical analysis suggests that informationally sophisticated investors may actively screen listed loans on the platforms. Thus, the first step of our empirical analysis is to study whether more sophisticated investors indeed screen differently, which is implied by Prediction 1.

For this purpose, we focus on loans invested by different LendingRobot account types and conduct the following empirical test at the loan level, separately for Lending Club and Prosper loans over our sample period from 2014 to 2016. We restrict our analysis to fractional loans on these platforms, that is, loans that are offered to retail investors only.

We define three loan-level indicator variables, RobotAccounti, AdvancedAccounti, and MonitoredOnlyAccounti, which are equal to 1 if at least two robot, advanced, or advanced accounts invested in a given loan i. We use these variables as left-hand-side variables, and run linear probability cross-sectional regressions on borrower characteristics.33 We include interest rate fixed effects in these cross-sectional regressions and, therefore, effectively identify screening that occurs among loans with the same price. Such screening is aimed at picking loans with lower default risk, keeping the interest rate constant.34 Concretely, we run the following

---

33 The results are similar when conducting logit regressions.
34 We abstract from the question of whether loans with a given interest rate are on average fairly priced at a given time.
cross-sectional regression at the loan level:

$$\text{Prob}(\text{TypeAccount}_i = 1) = \beta \times \text{BorrowerCharacteristics}_i + \sum_{j=1}^{n} IR_j + \sum_{t=1}^{T} m_t + \epsilon_i; \quad (4.1)$$

where BorrowerCharacteristics$_i$ are characteristics of borrower $i$ provided to the investors by the lending platform, $IR_j$ are fixed effects for the $n$ distinct levels of interest rate paid to investors (i.e., the loan price at issuance, which is a function of the loan grade given by the platform, the loan maturity, and the issuance period) observed in our data, $m_t$ are month-of-issuance fixed effects, and $\epsilon_i$ is the error term. Note that loan maturity (either 3 or 5 years) is implicitly controlled for by interest rate fixed effects, as within the same risk bucket interest rate varies with maturity only. Table 2 displays the results.

[Insert Table 2]

There are two key takeaways from this analysis. First, a large number of specific borrower characteristics strongly predict an investment by robot accounts, suggesting that robot accounts indeed actively screen loans within a risk bucket.\(^{35}\) This also suggests that LendingRobot’s screening model considers the risk associated with these characteristics to be mis-estimated or not fully incorporated by the platform when listing the loans. For variables positively (negatively) correlated with risk, a positive coefficient in Column 1 suggests that LendingRobot’s screening model considers that the platform overpenalizes (underpenalizes) this borrower characteristic. Conversely, a negative coefficient points out to characteristics that are underpenalized (overpenalized) when the platform lists the loan. For instance, the positive coefficient on loan amount or revolving utilization in Column 1 suggests that LendingRobot’s screening model considers that Lending Club excessively penalizes borrowers that take a large loan or have large revolving balance.\(^{36}\)

Second, screening behavior varies by the type of investors. First, although advanced accounts screen differently from robot accounts, their screening criteria are largely consistent with each

\(^{35}\)We note that this is different from the usual perception that robot advisors are a form of passive investors.

\(^{36}\)When comparing Column 1 to Column 4, we observe that the regression coefficients are different for the two platforms. For some characteristics, they have opposite signs. This result stresses that investor screening only can be interpreted relative to platform prescreening at the intensive margin and reveals that LendingRobot believes that the prescreening models significantly differs between Lending Club and Prosper.
other: the coefficients in Column 2 are largely comparable to the ones in Column 1. This again suggests that both types of accounts represent sophisticated investors. Consistently with our view that monitored-only accounts are less sophisticated, they appear to screen significantly differently from robot accounts as well as from advanced accounts do.

We note that the observed correlation between investment decisions and borrower characteristics is not necessarily causal. It might result from the loans with some characteristics having been first picked up and financed by more sophisticated investors, and then less sophisticated investors invest in loans with the opposite characteristics. For instance, if more sophisticated investors massively invest in loans with high FICO scores within a risk bucket, less sophisticated investors will be mechanically more likely to invest in loans with a relatively lower FICO score. This interpretation is consistent with the view that monitored-only accounts represent the less sophisticated segment of investors who eventually pick up the loans being passed on by more sophisticated investors.

In Table D.1 in the appendix, we run a similar analysis as a robustness check using the log of the number of accounts, for each type of accounts, as the dependent variable. We find consistent regression coefficients.

4.2 Investor performance

Having documented that more sophisticated investors actively screen loans and do so differently from less sophisticated investors, we test whether active screening allows more sophisticated investors to systematically outperform average and unsophisticated investors on the platform. Specifically, we test Prediction 1 by investigating whether, controlling for interest rates, having sophisticated investors participate in a loan predicts a lower likelihood of default.

Following the literature, we measure performance at the loan level, by using an indicator variable for the loan being in default or charged off. A loan enters default status when it is more than 121 days past due and enters charge-off status after 150 days.37

We first plot the share of defaulted loans by subgrades as of December 2017 for every fractional loan on the Lending Club platform, as well as for the subsets of loans in which at least two

37We rely on loan status data from Lending Club as of December 2017. Although some of the loans have not yet matured, the majority of loan defaults happen in their first year, as documented on the platform Web sites and in the literature (Morse 2015). We also control for monthly vintage in our regressions.
robot, advanced and monitor-only accounts invested. The results in Figure 3 suggest systematic outperformance achieved by robot and advanced accounts over monitor-only accounts and the average Lending Club loans, across the whole spectrum of loan risk.

[Insert Figure 3]

We conduct regressions to more precisely explore investor outperformance. We regress the performance indicator on indicator variables for participation by different types of investors. We use ordinary least squares (OLS) specifications because of the many fixed effects and to help interpret the economic magnitude of the coefficients.\(^{38}\)

Specifically, we run the following cross-sectional regression at the loan level:

\[
\text{Prob}(\text{ChargedOff}_i = 1) = \beta_1 \times \mathbb{I}_{\text{TypeAccount}} + \sum_{j=1}^{n} IR_j + \sum_{t=1}^{T} m_t + \epsilon_i, \quad (4.2)
\]

where ChargedOff\(_i\) is an indicator variable for the loan \(i\) to be in default or charged off, \(\mathbb{I}_{\text{TypeAccount}}\) is an indicator variable equal to 1 if at least two accounts of a given type are invested in the loan, and \(IR_j\) and \(m_t\) are the same interest rate and month-of-issuance fixed effects as those in specification (4.1). Table 3 shows the results. Robot accounts are shown in Column 1, advanced accounts in Column 2, and monitored-only accounts in Column 3. The following Columns interact \(\mathbb{I}_{\text{TypeAccount}}\) with year fixed effects and with loan grade fixed effects.

[Insert Table 3]

In Table 3, Column 1 documents that the default rate of loans in which robot accounts invest is significantly lower than that for the whole Lending Club loan population over the 2014–2016 period, controlling for interest rates. This outperformance is measured within each subgrade, and therefore cannot result from any compositional effect across the subgrades. The economic magnitudes are particularly large: the OLS specification suggests a reduction by more than 3% of the default rate, comparing to a 14% average default rate for the Lending Club loan population. This outperformance therefore translates into a reduction in average default rate by more than 20%. Column 2 shows that over the full sample period, advanced accounts’ outperformance is

\(^{38}\)The results are robust under a logit specification.
slightly larger than that of robot accounts. By contrast, Column 3 shows that participation by monitored-only accounts only weakly predicts a lower default rate with an economic magnitude four times smaller, consistent with monitored-only accounts being less sophisticated.

Moreover, Columns 7 and 8 illustrate that the reduction in default rate obtained by robot and advanced accounts is broadly increasing in loan risk, but this larger outperformance is partly driven by riskier loans having a higher average default rate.

This analysis empirically establishes that informationally sophisticated investors systematically outperform less sophisticated ones. It suggests that sophisticated investors produce additional information about the underlying loans. It also suggests that platform prescreening leaves significant room for further investor screening, representing a deviation from the traditional banking model where banks screen loans exclusively on behalf of depositors.

4.3 Screening cost and investor performance

Given the outperformance of sophisticated investors, we follow Predictions 2 and 3 to investigate the relationship between platform information provision and investor screening performance.

For this purpose, we exploit the Lending Club 2014 event, that is, the unexpected shock to investor information set described in Section 1.2 with a difference-in-differences setting. While the shock affects all investors, it affects investors differentially, because unsophisticated investors are less likely to use the removed variables than are sophisticated investors.

We run the following cross-sectional regression at the loan level:

\[
\text{Prob}(\text{ChargedOff}_i = 1) = \beta_1 \times \mathbb{1}_{\text{robot}_i} + \beta_2 \times \mathbb{1}_{\text{robot}_i} \times \text{Post}_i + \beta_3 \times \mathbb{1}_{\text{advanced}_i} + \beta_4 \times \mathbb{1}_{\text{advanced}_i} \times \text{Post}_i + \beta_5 \times \mathbb{1}_{\text{monitor}_i} + \beta_6 \times \mathbb{1}_{\text{monitor}_i} \times \text{Post}_i + \sum_{j=1}^{n} IR_j + \sum_{t=1}^{T} m_t + \epsilon_i, \quad (4.3)
\]

where \(\mathbb{1}_{\text{robot}_i}\) is an indicator variable equal to 1 if at least two robot accounts are invested in the loan \(i\), \(\mathbb{1}_{\text{advanced}_i}\) is an indicator variable equal to 1 if at least two advanced accounts are invested in the loan \(i\), \(\mathbb{1}_{\text{monitor}_i}\) is an indicator variable equal to 1 if at least two monitor-only accounts are invested in the loan \(i\), \(\text{Post}_i\) is an indicator variable for the loan \(i\) to be issued in
the period after the shock to the information set, and $IR_j$ and $m_t$ are the same interest rate and month fixed effects as those in specification (4.1).

We find empirical evidence consistent with Prediction 2 that when investors’ screening cost increases, sophisticated investor outperformance is reduced. The results are displayed in Table 4. Column 1 implements this specification for all Lending Club loans issued in the period spanning 3 months before the event month to 3 months after. Column 2 restricts the loan sample to loans with grades C to G. Column 3 restricts the period to 2 months before the event month to 2 months after. Column 4 restricts the sample to loans that have either robot accounts or monitor-only accounts invested in, thereby implementing a difference-in-differences between the two groups.

This analysis reveals that the increase in investor screening cost significantly impacts the screening performance of robot accounts, as well as that of advanced accounts to a smaller extent. On the other hand, monitor-only accounts are not affected at all. This result is robust to all specifications, and is more pronounced for lower grade loans. The magnitude of the effect is large: the outperformance of robot accounts in the period immediately preceding the shock drops by more than a half, compared to that at the time of the shock.\footnote{This drop in performance could not be immediately observed at the time of the shock, as it takes time for loan performance to be revealed to investors.} When comparing the effect between robot and advanced accounts, we observe that advanced accounts are less affected by the shock. But note that this smaller effect for advanced accounts is partly driven by advanced accounts’ relatively lower outperformance before the time of the shock.

To further pin down the causal impact of the information shock on sophisticated investor performance, we implement an event-study type analysis to investigate the exact timing of the change in performance. This zooming-in is important to rule out that an underlying trend on the platform, for instance a gradual improvement in platform prescreening intensity, may drive our result.

For this purpose, we implement two regressions with two different samples: all fractional Lending Club loans, the control group, and all the Lending Club loans in which at least two
robot accounts invested, the treatment group. For both regressions, the dependent variable is an indicator variable equal to 1 if the loan is charged off as of December 2017, and the explanatory variables are month fixed effects. We still control for interest rate fixed effects to alleviate concerns over potential compositional changes during the sample period. The reference period for the month fixed effects is months 5 and 6 after the event month.

Figure 4 shows the result, which illustrates that the overall fractional loan performance on the Lending Club platform is unaffected by the shock, while the performance of the loans invested by robot accounts sharply deteriorates at the time of the shock.\footnote{Because the constant of each regression is absorbed by the interest rate fixed effects, each line only speaks to the relative evolution of default rate in each sample over that period, and not to the initial level of charged off in each loan population before the shock.} The sharpness in the change of performance, as well as its synchronicity with the time of the shock, are supportive of a causal interpretation.

\[\text{[Insert Figure 4]}\]

We ensure that our results do not come from a possibly sharp change in the composition of listed loans that may not be captured by the interest rate fixed effects. For this purpose, Table D.2 in the appendix illustrates that the pool of listed loans indeed remains unchanged after the shock in terms of major observable characteristics of the borrowers.

Finally, consistent with Prediction 3, this difference-in-differences analysis provides an economic rationale for Lending Club’s reduce information provision to investors. It is consistent with the platform “evening the playing field” to actively manage the potential adverse selection problem introduced by sophisticated investors.\footnote{An alternative explanation for the reduction in the information set would be that it reduces noise, making it easier for certain investors, typically the unsophisticated ones, to pick good loans. However, this hypothesis would predict that these investors are better off after the change, but we do not observe this finding in our data.}

### 4.4 Trends in platform prescreening intensity and investor performance

According to Prediction 4, we study the time-series evolution of platform prescreening and investor relative performance as a final step of our empirical analysis.
4.4.1 Platform prescreening

Based on the recursive interpretation of platform prescreening in our model, a higher prescreening intensity $\pi_p$ means good (bad) loans are more precisely screened in (out) in any risk bucket, that is, a higher accuracy of the classification of loan applications into different risk buckets. We thus explore how this accuracy, the empirical counterpart of $\pi_p$, evolves over time.

For this purpose, we follow the literature (e.g., Iyer et al. 2015) to build receiver operating characteristic (ROC) curves—which graph the true positive rate against the false positive rate—obtained when using Lending Club subgrades as a predictor of loans being charged off as of December 2017. The larger the area below a ROC curve is, the more precise the predictor is. The test is computed separately for loans issued and graded in 2014, 2015, and 2016 to understand how platform prescreening evolves over time.

The results are displayed in Figure 5, which documents how Lending Club subgrades and Prosper grades have been improving explanatory power for default over time. This evolution suggests an improvement in the accuracy of platform prescreening. Lending Club subgrades also appear to better predict default than Prosper grades, as the areas below the ROC curves are larger for Lending Club.42

[Insert Figure 5]

As a complementary test, we also explore whether the platforms have been changing the risk thresholds for a loan application to be listed (before allocating them into a risk bucket). We plot the evolution of the share of borrowers on Lending Club whose FICO score is below 670 and 660, and whose debt-to-income ratio is above 30% and 35%. Results are displayed in Figure 6, showing an increasing share of ex ante high-risk loans being listed. This figure is consistent with the overall increase in loan prices on Lending Club from Figure 2. Listing an increasing share of ex ante high-risk or short-credit-history applications, while still improving the accuracy of their risk classification, further supports the hypothesis that platforms are increasing their prescreening intensity.

[Insert Figure 6]

42 Using grades instead of subgrades for Lending Club still yields a better predictor than Prosper grades.
Overall, these two figures illustrate that platforms improve over time the accuracy of the classification within listed loans, which corresponds to a higher but still intermediate level of platform information production. These trends are consistent with Prediction 4.

4.4.2 Evolution of investor screening performance from 2014 to 2016

We finally explore the evolution of investor screening performance over our sample period. For each year, we plot the share of defaulted loans by subgrades (as of December 2017) for the entire Lending Club platform, as well as the subsets of loans in which at least two robot, advanced and monitor-only accounts invested. Figure 7 shows that the outperformance of robot and advanced investors appear to decrease over time.\footnote{However, the 2016 chart should be interpreted with a grain of salt as a large share of the loans from that year have not matured as of December 2017.}

Using regressions, Columns 4 to 6 of Table 3 also explore the time series of outperformance by investor type, by interacting the indicator of having a given type of accounts participating in a loan with year fixed effects. For both robot and advanced accounts, outperformance is significantly stronger in 2014 than in 2015, and even more so than in 2016. Between the two more sophisticated types, robot accounts outperform advanced accounts in 2014, but the pattern reversed in the 2 years that follow. Consistently with monitor-only accounts being less sophisticated, we find no time trend for them.

We view these patterns consistent with the platforms actively managing the potential adverse selection problem introduced by more sophisticated investors. We also note that, while the Lending Club information shock we investigate above plays an important role in the evolution between 2014 and 2015, it cannot account for the whole trend, which likely results as well from the improvement in prescreening by the platforms we previously document.

5 Conclusion

Different from the conventional banking paradigm, one prominent feature of the burgeoning marketplace lending is the joint information production by the platform and investors. In
maximizing loan volume, the platform trades off the improving screening outcome and an adverse selection problem, both resulted from the participation of informationally sophisticated investors. Thus, intermediate levels of platform prescreening intensity and information provision to investors are optimal. Using novel investor-level data, we empirically show that, compared to less sophisticated investors, more sophisticated investors screen loans differently and significantly outperform. However, the outperformance shrinks when the platforms reduce the information set available to investors. These empirical facts are consistent with platforms dynamically managing adverse selection, and shed light on the optimal distribution of information production between the platform and investors.
References


Figure 2: Evolution of interest rates on Lending Club by subgrades

This figure plots the evolution of interest rates for the different risk buckets (subgrades) for Lending Club.
This figure displays the share of fractional loans issued on Lending Club in the 2014 to 2016 that were in default or charged off as of December 2017. These shares are plotted over Lending Club subgrades, which map into a given interest rate at a given time. This share is calculated for the whole Lending Club platform, as well as for the restricted samples of loans which have at least two robot accounts invested in this loan, at least two advanced accounts invested in this loan, and at least two monitored-only investors invested in the loan. A loan enters default when it is 121+ days past due and enters charge-off status after 150 days.
Figure 4: Change to investor screening costs: Difference-in-differences analysis

This figure plots regression coefficients from a difference-in-differences analysis. The left-hand-side variable of the regression is an indicator variable equal to 1 if the loan is charged off as of December 2017. The explanatory variables are month fixed effects, and the regression includes interest rate fixed effects. The reference period is months 5 and 6 after the change in information set. Each line results from a distinct regression, where the sample is all fractional Lending Club loans for the blue line (control), and all Lending Club, where are at least two robot accounts are invested in (treatment). Because the constant of each regression is absorbed by the interest rate fixed effects, each line only speaks about the evolution, and not the absolute level of charged-off loans in each loan population. Segments represent confidence intervals at 10%, where standard errors are clustered at the interest rate level. The bottom panel restricts the sample to loans with grades C to G.
Figure 5: Platform prescreening: ROC curve of Lending Club subgrades and Prosper grades on charged-off loans

This figure plots the ROC curve, which graphs the true positive rate against the false positive rate, obtained when using Lending Club subgrades and Prosper grades as a predictor of charged-off loans. The larger the area below a ROC curve, the more precise the predictor is. The test is separately computed for loans issued and graded by each platform in 2014, 2015, and 2016.
Figure 6: Platform prescreening: Evolution of the share of borrowers above FICO and debt-to-income thresholds

This figure plots the evolution of the share of borrowers on the Lending Club platform whose FICO score is below 670 (680) and 660 and whose debt-to-income ratio is above 30% and 35%, for both Lending Club and Prosper, respectively.
These figures display the share of fractional loans issued on Lending Club in 2014, 2015, and 2016 that are in default or charged off as of December 2017. These shares are plotted over Lending Club subgrades, which map into a given interest rate at a given time. This share is calculated for the whole Lending Club platform, as well as for the restricted samples of loans which have at least two robot accounts invested in this loan, at least two advanced accounts invested in this loan, and at least two monitored-only investors invested in the loan. A loan enters default when it is 121+ days past due and enters charge-off status after 150 days.
<table>
<thead>
<tr>
<th>Platform</th>
<th>Number</th>
<th>Total amount invested (1)</th>
<th>Median amount invested (2)</th>
<th>Mean amount invested (3)</th>
<th>Max amount invested (4)</th>
<th>Avg. int. rate (6)</th>
<th>Platform Avg. int. rate (7)</th>
<th>Risk tolerance (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lending Club</td>
<td>7,368</td>
<td>138,633,952</td>
<td>3,050</td>
<td>18,815.7</td>
<td>3,712,900</td>
<td>18.98%</td>
<td>-</td>
<td>15.76%</td>
</tr>
<tr>
<td>Robot</td>
<td>4,435</td>
<td>56,692,279</td>
<td>1,600</td>
<td>12,783.6</td>
<td>2,102,925</td>
<td>19.34%</td>
<td>7.96%</td>
<td>-</td>
</tr>
<tr>
<td>Advanced</td>
<td>2,933</td>
<td>81,703,628</td>
<td>5,925</td>
<td>27,936.8</td>
<td>3,712,900</td>
<td>18.83%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Monitor-only</td>
<td>636</td>
<td>13,309,525</td>
<td>4,650</td>
<td>20,926.9</td>
<td>722,750</td>
<td>19.20%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prosper</td>
<td>1,616</td>
<td>21,039,794</td>
<td>2,425</td>
<td>13,019.7</td>
<td>658,639</td>
<td>19.84%</td>
<td>-</td>
<td>16.32%</td>
</tr>
<tr>
<td>Robot</td>
<td>1,095</td>
<td>13,421,524</td>
<td>1,900</td>
<td>12,257.1</td>
<td>630,937</td>
<td>19.86%</td>
<td>8.01%</td>
<td>-</td>
</tr>
<tr>
<td>Advanced</td>
<td>521</td>
<td>7,618,145</td>
<td>3525</td>
<td>14,622.4</td>
<td>658,639</td>
<td>19.80%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Monitor-only</td>
<td>126</td>
<td>1,699,350</td>
<td>1,925</td>
<td>13,486.9</td>
<td>155,575</td>
<td>16.54%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

This table provides summary statistics on the proprietary data set used for the empirical analysis. The data covers all transactions executed by LendingRobot users between January 2014 and February 2017, which represents more than $120 million invested on the two major lending platforms, Lending Club and Prosper, as well as all historic transactions from portfolios monitored by the company. Robot accounts are automatically invested based on LendingRobot credit model and investors risk tolerance. Trades are executed through the lending platforms’ API. Advanced accounts combine LendingRobot screening tool with tailored screening criteria and are executed through the platform APIs. Monitor-only accounts are managed by the investor themselves, and LendingRobot only tracks the portfolio performance, without performing any trades. Risk tolerance corresponds to the target return the investor select between 6.3% and 8.5% and associated maximum losses between 7% and 12%.
Table 2: Investor screening, 2014-2016

<table>
<thead>
<tr>
<th></th>
<th>Lending Club</th>
<th></th>
<th>Prosper</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Robot (1)</td>
<td>Advanced (2)</td>
<td>Monitored (3)</td>
<td>Robot (4)</td>
</tr>
<tr>
<td>Loan amount</td>
<td>0.005***</td>
<td>0.008***</td>
<td>0.015***</td>
<td>0.015***</td>
</tr>
<tr>
<td></td>
<td>(18.89)</td>
<td>(27.97)</td>
<td>(36.16)</td>
<td>(25.83)</td>
</tr>
<tr>
<td>FICO score</td>
<td>0.000</td>
<td>0.001***</td>
<td>-0.001***</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(1.44)</td>
<td>(11.62)</td>
<td>(-10.56)</td>
<td>(-0.01)</td>
</tr>
<tr>
<td>Annual income</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.000***</td>
<td>-0.000***</td>
</tr>
<tr>
<td></td>
<td>(7.18)</td>
<td>(13.42)</td>
<td>(9.83)</td>
<td>(-2.90)</td>
</tr>
<tr>
<td>Employment length</td>
<td>0.002***</td>
<td>0.007***</td>
<td>0.001***</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(8.96)</td>
<td>(19.42)</td>
<td>(5.47)</td>
<td>(1.43)</td>
</tr>
<tr>
<td>Debt to income</td>
<td>-0.001***</td>
<td>-0.002***</td>
<td>0.001***</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(-4.67)</td>
<td>(-10.36)</td>
<td>(8.71)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>Home ownership</td>
<td>0.033***</td>
<td>0.054***</td>
<td>0.006**</td>
<td>-0.017**</td>
</tr>
<tr>
<td></td>
<td>(9.96)</td>
<td>(14.33)</td>
<td>(2.53)</td>
<td>(-2.71)</td>
</tr>
<tr>
<td>Open accounts</td>
<td>0.002***</td>
<td>0.001***</td>
<td>0.000</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>(7.04)</td>
<td>(5.73)</td>
<td>(0.89)</td>
<td>(3.30)</td>
</tr>
<tr>
<td>First credit line</td>
<td>-0.000</td>
<td>-0.001***</td>
<td>-0.001***</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-1.56)</td>
<td>(-2.50)</td>
<td>(-19.17)</td>
<td>(-0.62)</td>
</tr>
<tr>
<td>Delinquency</td>
<td>-0.005***</td>
<td>-0.019***</td>
<td>-0.006***</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-6.70)</td>
<td>(-18.68)</td>
<td>(-6.73)</td>
<td>(-0.24)</td>
</tr>
<tr>
<td>Term</td>
<td>-0.012</td>
<td>-0.066***</td>
<td>0.045**</td>
<td>-0.000</td>
</tr>
<tr>
<td></td>
<td>(-1.59)</td>
<td>(-7.65)</td>
<td>(6.89)</td>
<td>(-0.42)</td>
</tr>
<tr>
<td>Inquiries, last 6 months</td>
<td>-0.038***</td>
<td>-0.088***</td>
<td>-0.003**</td>
<td>-0.008***</td>
</tr>
<tr>
<td></td>
<td>(-14.47)</td>
<td>(-28.10)</td>
<td>(-2.00)</td>
<td>(-3.59)</td>
</tr>
<tr>
<td>Revolving utilization rate</td>
<td>0.000***</td>
<td>0.000***</td>
<td>0.000***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.39)</td>
<td>(5.36)</td>
<td>(2.93)</td>
<td></td>
</tr>
<tr>
<td>Derogatory public records</td>
<td>-0.042***</td>
<td>-0.044***</td>
<td>0.003*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-5.07)</td>
<td>(-6.65)</td>
<td>(1.92)</td>
<td></td>
</tr>
<tr>
<td>Mortgage</td>
<td>0.045***</td>
<td>0.075***</td>
<td>0.017***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.20)</td>
<td>(15.44)</td>
<td>(8.09)</td>
<td></td>
</tr>
<tr>
<td>Loan purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Car</td>
<td>0.051***</td>
<td>0.021*</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.33)</td>
<td>(1.89)</td>
<td>(1.08)</td>
<td></td>
</tr>
<tr>
<td>Credit card</td>
<td>0.063***</td>
<td>0.077***</td>
<td>-0.010***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8.37)</td>
<td>(13.71)</td>
<td>(-4.53)</td>
<td></td>
</tr>
<tr>
<td>Home improvement</td>
<td>-0.005</td>
<td>-0.027***</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-1.52)</td>
<td>(-6.34)</td>
<td>(-1.11)</td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>-0.103***</td>
<td>-0.189***</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-7.23)</td>
<td>(-16.90)</td>
<td>(0.95)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interest rate FEs</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month FEs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster</td>
<td>Interest rate</td>
<td>Interest rate</td>
<td>Interest rate</td>
<td>Interest rate</td>
<td>Interest rate</td>
<td>Interest rate</td>
</tr>
<tr>
<td>Observations</td>
<td>365,685</td>
<td>365,685</td>
<td>365,685</td>
<td>38,047</td>
<td>38,047</td>
<td>38,047</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.284</td>
<td>0.222</td>
<td>0.222</td>
<td>0.115</td>
<td>0.173</td>
<td>0.215</td>
</tr>
</tbody>
</table>

This table displays coefficients from OLS regressions, where the dependent variable is an indicator variable equal to 1 if at least two robot accounts invested in this loan (Columns 1 and 4), at least two advanced accounts invested in this loan (Columns 2 and 5), and at least two monitored-only investor invested in the loan (Columns 3 and 6). Robot accounts are invested automatically based on LendingRobot credit model and investors risk tolerance. Trades are executed through the lending platforms’ API. Advanced accounts combine LendingRobot screening tool with tailored screening criteria and are executed through the platform APIs. Monitor-only accounts are managed by the investor themselves, and LendingRobot only tracks the portfolio performance, without performing any trades. The sample includes all Lending Club fractional loans issued between 2014 and 2016 for Columns 1 to 3, and all Prosper fractional loans issued between 2014 and 2016 for Columns 4 to 6. Standard errors of the coefficients are clustered by platform interest rate, and t-statistics are reported below the coefficients. *, **, and *** represent statistical significance at the 10%, 5%, and 1% confidence levels, respectively.
Table 3: Screening performance

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.031***</td>
<td>-0.044***</td>
<td>-0.008***</td>
<td>-0.084***</td>
<td>-0.070***</td>
<td>-0.005</td>
<td>0.012*</td>
<td>-0.015***</td>
<td>0.007***</td>
<td></td>
</tr>
<tr>
<td>(-10.84)</td>
<td>(-18.04)</td>
<td>(-4.68)</td>
<td>(-20.56)</td>
<td>(-19.86)</td>
<td>(-1.27)</td>
<td>(1.66)</td>
<td>(-3.64)</td>
<td>(2.21)</td>
<td></td>
</tr>
<tr>
<td>Account type x 2015</td>
<td>0.051***</td>
<td>0.029***</td>
<td>-0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10.38)</td>
<td>(7.11)</td>
<td>(-1.27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x 2016</td>
<td>0.075***</td>
<td>0.050***</td>
<td>-0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13.66)</td>
<td>(12.42)</td>
<td>(-0.45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade B</td>
<td>-0.019***</td>
<td>-0.015***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-3.36)</td>
<td>(-3.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade C</td>
<td>-0.003***</td>
<td>-0.027***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-6.06)</td>
<td>(-4.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade D</td>
<td>-0.037***</td>
<td>-0.019***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-6.06)</td>
<td>(-4.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade E</td>
<td>-0.047***</td>
<td>-0.019***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-6.06)</td>
<td>(-4.58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade F</td>
<td>-0.005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-3.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account type x Grade G</td>
<td>-0.081***</td>
<td>-0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-3.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This table contains the OLS regression coefficients for fractional loans originated on the Lending Club platform for the period 2014–2016. The dependent variable is an indicator variable for the loan being charged off or in default status as of December 2017. Explanatory variables are indicator variables equal to 1 if at least two robot, advanced, and monitor-only accounts are invested in this loan. Robot accounts are invested automatically based on LendingRobot credit model and investors risk tolerance. Trades are executed through the lending platforms’ API. Advanced accounts combine LendingRobot screening tool with tailored screening criteria and are executed through the platform APIs. Monitor-only accounts are managed by the investor themselves, and LendingRobot only tracks the portfolio performance, without performing any trades. Grades A to G are as per Lending Club typology. Standard errors of the coefficients are clustered by interest rate, and t-statistics are reported below the coefficients. *, **, and *** represent statistical significance at the 10%, 5%, and 1% confidence levels, respectively.
## Table 4: Difference-in-differences analysis

<table>
<thead>
<tr>
<th></th>
<th>-3/+3 months window (1)</th>
<th>Grade below C (2)</th>
<th>-2/+2 months window (3)</th>
<th>Control group:  window below (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robot account</td>
<td>-0.072***</td>
<td>-0.076***</td>
<td>-0.074***</td>
<td>-0.098***</td>
</tr>
<tr>
<td></td>
<td>(-7.00)</td>
<td>(-5.34)</td>
<td>(-6.98)</td>
<td>(-10.85)</td>
</tr>
<tr>
<td>Robot account x Post</td>
<td>0.040***</td>
<td>0.049***</td>
<td>0.037**</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(3.20)</td>
<td>(3.01)</td>
<td>(2.68)</td>
<td>(3.65)</td>
</tr>
<tr>
<td>Advanced account</td>
<td>-0.057***</td>
<td>-0.064***</td>
<td>-0.053***</td>
<td>0.043***</td>
</tr>
<tr>
<td></td>
<td>(-8.03)</td>
<td>(-6.20)</td>
<td>(-6.14)</td>
<td></td>
</tr>
<tr>
<td>Advanced account x Post</td>
<td>0.013*</td>
<td>0.008</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
<td>(0.71)</td>
<td>(1.42)</td>
<td></td>
</tr>
<tr>
<td>Monitor-only account</td>
<td>0.013*</td>
<td>0.020**</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(2.15)</td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>Monitor-only account x Post</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.09)</td>
<td>(-0.19)</td>
<td>(1.71)</td>
<td></td>
</tr>
</tbody>
</table>

### Month FE

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
</table>

### Interest rate FE

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
</table>

### Cluster

<table>
<thead>
<tr>
<th></th>
<th>Int. rate</th>
<th>Int. rate</th>
<th>Int. rate</th>
<th>Int. rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>65,859</td>
<td>35,880</td>
<td>37,615</td>
<td>11,283</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.059</td>
<td>0.030</td>
<td>0.060</td>
<td>0.071</td>
</tr>
</tbody>
</table>

This table displays the regression coefficients from linear probability regressions. The dependent variable is an indicator variable for a given loan to be charged off as of December 2017. $\mathbb{1}_{\text{Robot}}$ is an indicator variable equal to 1 if at least two robot accounts are invested in the loan, $\mathbb{1}_{\text{Advanced}}$ is an indicator variable equal to 1 if at least two advance accounts are invested in the loan, $\mathbb{1}_{\text{Monitor}}$ is an indicator variable equal to 1 if at least two monitor-only account are invested in the loan, $Post$ is an indicator variable for being in the period after the shock to the information set. All regressions include interest rate fixed effects and month fixed effects. Standard errors of the coefficients are clustered by interest rate, and $t$-statistics are reported below the coefficients. *, **, and *** represent statistical significance at the 10%, 5%, and 1% confidence levels, respectively.
Appendix

A Proof of Theorem 1

The proof proceeds in three steps. First, we characterize different types of subgame equilibria as described in the main text. Second, we show that either type 1 or type 3 subgame equilibrium cannot sustain a full equilibrium; in other words, the full equilibrium only admits either a type 2 or a type 4 subgame equilibrium. Third, we show that a type 2 (type 4) subgame equilibrium happens in a full equilibrium when the platform’s information cost $\kappa$ is larger (smaller) than a threshold.

Step 1. This step repeatedly uses the two participation conditions (2.3) and (2.6). To recap, condition (2.3) specifies that uninformed investors invest on the platform if and only if the expected value of a listed loan, after potential screening by informed sophisticated investors, is higher than the investment requirement. Condition (2.6) specifies that sophisticated investors become informed and screen list loans if and only if their expected profit from screening is higher than the information cost.

By construction, the determination of the four types of subgame equilibria is governed by whether neither, at least one, or both of the two participations are satisfied. Denote type $j$ subgame equilibrium profile by $SE_j$, $j \in \{1, 2, 3, 4\}$, which is a subset of the $(\pi_p, \mu)$-space denoted by $S = \{ (\pi_p, \mu, \cdot) | \pi_0 \leq \pi_p \leq 1, \mu \geq 0 \}$.

In a type 1 subgame equilibrium, none of the investors participates. This implies that $\omega = 0$ and condition (2.5) further implies that $p_i(0) = I$. Thus, conditions (2.3) and (2.6) give that

$$SE_1 = \{ (\pi_p, \mu, \cdot) | V'(0) < I, \pi_p(R_H - I) < \mu \} \, . \quad (A.1)$$

In a type 2 subgame equilibrium, sophisticated investors become informed but uninformed investors do not participate. This implies that $\omega = \min \{\Omega, x_p\}$ and condition (2.5) further implies that $p_i(\min \{\Omega, x_p\}) = I$. Thus, conditions (2.3) and (2.6) give that

$$SE_2 = \{ (\pi_p, \mu, \cdot) | V'(\Omega) < I, \pi_p(R_H - I) \geq \mu \} \, . \quad (A.2)$$
where $V'(\Omega) = V'(\min\{\Omega, x_p\})$ by construction.

In a type 3 subgame equilibrium, sophisticated investors become informed and uninformed investors also participate. This implies that $\omega = \min\{\Omega, x_p\}$, and condition (2.5) further implies that $p_i(\min\{\Omega, x_p\}) = V'(\min\{\Omega, x_p\})$. Thus, conditions (2.3) and (2.6) give that

$$SE_3 = \{(\pi_p, \mu, \cdot)|V'(\Omega) \geq I, \pi_p(R_H - V'(\Omega)) \geq \mu\}, \quad (A.3)$$

where $V'(\Omega) = V'(\min\{\Omega, x_p\})$ by construction.

In a type 4 subgame equilibrium, sophisticated investors do not become informed but all the uninformed investors participate. This implies that $\omega = 0$, and condition (2.5) further implies that $p_i(0) = V'(0)$. Thus, conditions (2.3) and (2.6) give that

$$SE_4 = \{(\pi_p, \mu, \cdot)|V'(0) \geq I, \pi_p(R_H - V'(0)) < \mu\}. \quad (A.4)$$

Because $V'(\omega)$ is decreasing in $\omega$, it immediately follows that

$$\bigcup_{j \in \{1,2,3,4\}} SE = S,$$

implying that one of the four types of subgame equilibrium must happen. Direct calculation of the equilibrium volume in each subgame equilibrium concludes this step.

**Step 2.** Suppose $(\pi_p, \mu, \cdot) \in SE_1$, that is, the full equilibrium admits a type 1 subgame equilibrium as characterized in (A.1). The equilibrium volume is 0.

Consider an alternative equilibrium profile $(\pi_p, \mu', \cdot)$, $\mu' < \mu$ such that $\pi_p(R_H - I) \geq \mu'$. Notice that $\mu'$ must exist because $0 < \pi_0 \leq \pi_p$. Because $V'(0) \geq V'(\Omega)$, we have that $(\pi_p, \mu', \cdot) \in SE_2$ as characterized in (A.2). Because changing $\mu$ is costless for the platform, this implies that the platform will then find it profitable to deviate to a type 2 subgame equilibrium by decreasing $\mu$ to $\mu'$ without changing $\pi_p$, enjoying a higher volume $\min\{\pi_0x_0(I), \pi_p\Omega\} > 0$. Thus, a full equilibrium can only admit a type 2 subgame equilibrium, but not a type 1 subgame equilibrium.

Similarly, suppose $(\pi_p, \mu, \cdot) \in SE_3$, that is, the full equilibrium admits a type 3 subgame equilibrium as characterized in (A.3). The equilibrium volume is $\frac{\pi_0x_0(\mu(\Omega))}{\pi_p}$. 

47
Consider an alternative equilibrium profile \((p, \mu', \cdot)\), \(\mu' > \mu\) such that \(\pi_p(R_H - V'(0)) < \mu'\). Notice that \(\mu'\) must exist because \(0 < \pi_0 \leq \pi_p\) and \(I \leq V'(0) < R_H\). Again because \(V'(0) \geq V'(\Omega)\), we have that \((\pi_p, \mu', \cdot) \in SE_4\) as characterized in (A.4). Because changing \(\mu\) is costless for the platform and \(x_0(p(0)) > x_0(p(\Omega))\), this implies that the platform will then find it profitable to deviate to a type 4 subgame equilibrium by increasing \(\mu\) to \(\mu'\) without changing \(\pi_p\), enjoying a higher volume \(\frac{\pi_0 x_0(p(0))}{\pi_p} > \frac{\pi_0 x_0(p(\Omega))}{\pi_p}\). Hence, a full equilibrium can only admit a type 4 subgame equilibrium, but not a type 3 subgame equilibrium.

**Step 3.** Consider the hyperplane \(V'(0) = I\) that decomposes the \((\pi_p, \mu, \cdot)\)-space into two half-spaces:

\[
S_L = \{ (\pi_p, \mu, \cdot) | V'(0) < I \},
\]

and

\[
S_H = \{ (\pi_p, \mu, \cdot) | V'(0) \geq I \}.
\]

Notice that the hyperplane \(V'(0) = I\) gives a unique platform prescreening intensity

\[
\widehat{\pi_p} = \frac{I - R_L}{R_H - R_L} \in (0, 1),
\]

and by construction, \(SE_1 \subset S_L\) and \(SE_4 \subset S_H\) according to conditions (A.1) and (A.4). Thus, \(SE_1 \cap SE_4 = \emptyset\). In particular, (A.1) and (A.4) imply that there must exist a \(\widehat{\mu} > 0\) such that

\[
(SE_1 \cup SE_4) \cap \{ (\pi_p, \mu, \cdot) | \mu > \widehat{\mu} \} = S \cap \{ (\pi_p, \mu, \cdot) | \mu > \widehat{\mu} \}.
\]

Hence, if

\[
C(\widehat{\pi_p}) = \frac{1}{2} \kappa((\pi_p - \pi_0)^+) = \min\{\pi_0 x_0(I), \pi_p \Omega\},
\]

by the argument in Step 2, a type 2 subgame equilibrium is achievable by choosing

\[
\pi_p = \frac{\Omega}{\kappa} + \pi_0\quad \text{and}\quad \mu = 0,
\]

in which case we define

\[
\kappa = \frac{2 \min\{\pi_0 x_0(I), \pi_p \Omega\}}{(\pi_p - \pi_0)^+} > 0.
\]

48
Otherwise if
\[ C(\pi_p) = \frac{1}{2} \kappa((\pi_p - \pi_0)_+)^2 \leq \frac{\pi_0 x_0(p(\Omega))}{\pi_p}, \]
a type 4 subgame equilibrium is achievable by choosing \( \pi_p \) such that it solves
\[ \frac{\partial}{\partial \pi_p} \left( \frac{\pi_0 x_0(p(\Omega))}{\pi_p} - C(\pi_p) \right) = 0, \]
and choosing \( \pi > \hat{\mu} \), in which case we define
\[ \tilde{\pi} = \frac{2\pi_0 x_0(p(\Omega))}{\pi_p ((\pi_p - \pi_0)_+)^2} > 0. \]
Because \( \pi \geq \kappa \), this concludes the proof.

B The Supply Curve of Loan Applications

This appendix provides one simple microfoundation for the upward-sloping supply curve of loan applications \( x_0(p) \).

Keeping the initial setting in the baseline model, we further assume that a loan applicant, when successfully establishes a loan from a lending marketplace, may incur an adoption cost.\(^{44,45}\) This adoption cost reflects the fact that marketplace lending is still relatively new, and thus a loan applicant may incur physical or cognitive costs to understand and follow its rules and trust them. Different applicants may have different adoption costs, and it is natural that some applicants, who are already familiar with marketplace lending, may not incur any adoption cost.

Specifically, we assume that the population of potential loan applicants who do not have any adoption costs is \( q(0) > 0 \), and the population of potential loan applicants who have a positive adoption cost \( z \) is \( q(z) \geq 0 \), where \( z \in (0, +\infty) \) and follows an arbitrary distribution \( F(z) \). Consistent with our baseline model, the loan applicants do not know their type ex ante.

Following the literature (Hauswald and Marquez 2006; de Roure, Pelizzon, and Thakor 2018, for

\(^{44}\)If a loan applicant goes to the platform, applies, and is not funded, there will be no adoption cost.

\(^{45}\)This notion of adoption cost follows the classic Hotelling model and has been commonly used in the theoretical literature of bank competition. For example, Hauswald and Marquez (2006) use it to model that different banks may specialize in different borrower characteristics. Using a marketplace lending context similar to ours, de Roure, Pelizzon, and Thakor (2018) introduce a notion of “acquisition cost” to reflect how costly it is for a borrower to leave a bank (with which the borrower has a relationship) to join a lending platform.
example), we also assume that the adoption cost is independent to the type of loan applicants.

Under this setting, any platform price \( p \) that attracts a potential loan applicant to apply has to be at least the investment requirement \( I \) plus the adoption cost of that applicant. Thus, the population of actual loan applications at any platform price \( p \) is given by

\[
x_0(p) = \begin{cases} 
q(0) + \int_{0}^{p-I} q(z) dF(z), & \text{if } p > I, \\
q(0), & \text{if } p = I,
\end{cases}
\]

which is increasing in \( p \). Note that, because loan applicants do not know their type ex ante, the above expression does not depend on the distribution of their types. This microfoundation provides us with the reduced-form supply curve of loan applications we use in the baseline model.

C Alternative Specifications of Investor Sophistication

Without fully solving the model, this appendix discusses three alternative specifications of investor sophistication. The aim is to qualitatively illustrate that in our baseline model, both information and speed advantages are important in driving sophisticated investor outperformance and adverse selection among investors. We keep other model setting and assumptions unchanged.

**Alternative specification 1.** Both types of investors are always uninformed, but sophisticated investors can acquire a technology to become faster in the sense that it allows them to fund a listed loan before other investors do so. Because faster investors are uninformed, they fund a listed loan if and only if \( \pi_p R_H + (1 - \pi_p) R_L \geq I \). In addition, because faster investors do not screen loans, slower investors’ interim posterior belief of a listed loan being good is still \( \pi_p \). It follows that slower investors use the same funding criterion as faster investors. Therefore, faster investors cannot outperform slower ones and sophisticated investors never choose to become faster. Subsequently, there is no adverse selection among investors.

**Alternative specification 2.** Both types of investors are always informed about a listed loan facing them, but sophisticated investors can acquire a technology to become faster. Because all investors are informed, they use the same criterion to screen listed loans, always offer the same price \( I \) to an identified good loan and pass on an identified bad loan, regardless of their
speed. Thus, bad loans are never funded, and conditional on an investment on the platform, faster investors cannot outperform slower ones.\footnote{Note that these results may change if we instead assume at least three types of projects, in particular, more than one type of positive NPV project. Then faster investors may take higher-value projects, leaving lower-value projects for slower investors. But such an assumption would require a change in the model setting and more assumptions, making the predictions not directly comparable to the ones used in our baseline model.}

**Alternative specification 3.** Both types of investors are equally fast, but sophisticated investors can acquire a technology to become informed about a listed loan. An informed investor offers price $I$ to an identified good loan and passes on an identified bad loan. However, because uninformed investors are equally fast, an identified bad loan (passed on by an informed investor) will never reach them. Rather, they fund a listed loan if and only if $\pi_p R_H + (1 - \pi_p) R_L \geq I$ under any given prescreening intensity $\pi_p$. Thus, there is no adverse selection. Moreover, the bad loans that are passed on by informed, sophisticated investors are never funded.

**D Additional Tables and Figures**
<table>
<thead>
<tr>
<th></th>
<th>Lending Club</th>
<th></th>
<th>Prosper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Robot (1)</td>
<td>Advanced (2)</td>
<td>Monitored (3)</td>
</tr>
<tr>
<td>Loan amount</td>
<td>0.011***</td>
<td>0.018***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(16.82)</td>
<td>(26.95)</td>
<td>(36.81)</td>
</tr>
<tr>
<td>FICO score</td>
<td>0.000***</td>
<td>0.002***</td>
<td>-0.001***</td>
</tr>
<tr>
<td></td>
<td>(2.71)</td>
<td>(12.32)</td>
<td>(-10.21)</td>
</tr>
<tr>
<td>Annual income</td>
<td>0.002***</td>
<td>0.003***</td>
<td>0.001***</td>
</tr>
<tr>
<td></td>
<td>(7.63)</td>
<td>(12.20)</td>
<td>(9.20)</td>
</tr>
<tr>
<td>Employment length</td>
<td>0.006***</td>
<td>0.015***</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(9.30)</td>
<td>(20.99)</td>
<td>(7.31)</td>
</tr>
<tr>
<td>Debt to income</td>
<td>-0.001***</td>
<td>-0.004***</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(-4.00)</td>
<td>(-14.26)</td>
<td>(9.81)</td>
</tr>
<tr>
<td>Home ownership</td>
<td>0.086***</td>
<td>0.137***</td>
<td>0.015***</td>
</tr>
<tr>
<td></td>
<td>(8.98)</td>
<td>(14.31)</td>
<td>(4.03)</td>
</tr>
<tr>
<td>Open accounts</td>
<td>0.005***</td>
<td>0.004***</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(7.51)</td>
<td>(6.99)</td>
<td>(1.06)</td>
</tr>
<tr>
<td>First credit line</td>
<td>-0.002**</td>
<td>-0.003***</td>
<td>-0.002***</td>
</tr>
<tr>
<td></td>
<td>(-2.51)</td>
<td>(-4.63)</td>
<td>(-9.86)</td>
</tr>
<tr>
<td>Delinquency</td>
<td>-0.012***</td>
<td>-0.043***</td>
<td>-0.011***</td>
</tr>
<tr>
<td></td>
<td>(-7.13)</td>
<td>(-22.24)</td>
<td>(-10.14)</td>
</tr>
<tr>
<td>Term</td>
<td>-0.054***</td>
<td>-0.170***</td>
<td>0.057***</td>
</tr>
<tr>
<td></td>
<td>(-2.94)</td>
<td>(-11.68)</td>
<td>(6.13)</td>
</tr>
<tr>
<td>Inquiries, last 6 months</td>
<td>-0.096***</td>
<td>-0.162***</td>
<td>-0.005**</td>
</tr>
<tr>
<td></td>
<td>(-14.95)</td>
<td>(-22.42)</td>
<td>(-2.02)</td>
</tr>
<tr>
<td>Revolving utilization rate</td>
<td>0.001***</td>
<td>0.001***</td>
<td>0.000***</td>
</tr>
<tr>
<td>Derogatory public records</td>
<td>-0.113***</td>
<td>-0.117***</td>
<td>0.007***</td>
</tr>
<tr>
<td>Mortgage</td>
<td>0.118***</td>
<td>0.193***</td>
<td>0.028***</td>
</tr>
<tr>
<td>Car</td>
<td>0.130***</td>
<td>0.073***</td>
<td>0.015</td>
</tr>
<tr>
<td>Credit card</td>
<td>0.171***</td>
<td>0.199***</td>
<td>-0.019***</td>
</tr>
<tr>
<td>Home improvement</td>
<td>-0.010</td>
<td>-0.078***</td>
<td>-0.010**</td>
</tr>
<tr>
<td>Medical</td>
<td>-0.287***</td>
<td>-0.407***</td>
<td>0.007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interest rate FEs</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month FEs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cluster</td>
<td>Int.rate</td>
<td>Int.rate</td>
<td>Int.rate</td>
<td>Int.rate</td>
<td>Int.rate</td>
<td>Int.rate</td>
</tr>
<tr>
<td>Observations</td>
<td>365,685</td>
<td>365,685</td>
<td>365,685</td>
<td>38,047</td>
<td>38,047</td>
<td>38,047</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.355</td>
<td>0.362</td>
<td>0.290</td>
<td>0.282</td>
<td>0.386</td>
<td>0.280</td>
</tr>
</tbody>
</table>

This table displays coefficients from OLS regressions, where the dependent variable is the log of 1 plus the number of robot accounts invested in this loan (Columns 1 and 4), advanced accounts invested in this loan (Columns 2 and 5), and monitored-only investor invested in the loan (Columns 3 and 6). Robot accounts are automatically invested based on LendingRobot credit model and investors risk tolerance. Trades are executed through the lending platform’s API. Advanced accounts combine LendingRobot screening tool with tailored screening criteria and are executed through the platform APIs. Monitor-only accounts are managed by the investor themselves, and LendingRobot only tracks the portfolio performance, without performing any trades. The sample includes all Lending Club fractional loans issued between 2014 and 2016 for Columns 1 to 3 and all Prosper fractional loans issued between 2014 and 2016 for Columns 4 to 6. Standard errors of the coefficients are clustered by platform interest rate, and t-statistics are reported below the coefficients. *, **, and *** represent statistical significance at the 10%, 5%, and 1% confidence levels, respectively.
Table D.2: Borrower characteristics before and after the change in the information set

<table>
<thead>
<tr>
<th></th>
<th>2 months before</th>
<th>2 months after</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loan amount (in USDk)</td>
<td>14.85</td>
<td>15.29</td>
</tr>
<tr>
<td>FICO score</td>
<td>696</td>
<td>697</td>
</tr>
<tr>
<td>Annual income (in USDk)</td>
<td>72.94</td>
<td>74.38</td>
</tr>
<tr>
<td>Employment length</td>
<td>5.6</td>
<td>5.5</td>
</tr>
<tr>
<td>Debt to income</td>
<td>18.7</td>
<td>18.6</td>
</tr>
<tr>
<td>Home ownership</td>
<td>11.5%</td>
<td>11.8%</td>
</tr>
<tr>
<td>Open accounts</td>
<td>11.8</td>
<td>11.6</td>
</tr>
<tr>
<td>First credit line (in years)</td>
<td>18.0</td>
<td>17.9</td>
</tr>
<tr>
<td>Delinquency</td>
<td>0.345</td>
<td>0.333</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>58,566</td>
<td>56,335</td>
</tr>
</tbody>
</table>

This table displays the averages for key borrowers’ characteristics for the 2-month period before and after the change in the information set provided to investors by Lending Club.
Panel A plots the median time for loans to sell out. Panel B plots the share of loans that sell out faster than a 10-, 5-, and 1-minute threshold. The sample only covers Lending Club loans in which LendingRobot-assisted investors participated.

Figure D.1: Time for loans to sell out