2015

Loan picking in p2p-lending. Is there wisdom in the crowd?

Mark Bärthel
University of Munich

_table

Loan picking in p2p-lending. Is there wisdom in the crowd?

Extended Abstract

for submission to the

First International Workshop on P2P Financial Systems (P2PFISY)

Abstract

With more than 8 billion dollars in loans issued by just the two largest crowdlending platforms in the U.S., as of October 2014, p2p-lending is now in the focus of regulators worldwide. At the moment, the p2p lending industry is undergoing a fast paced dynamic of policy changes by the platforms and adaptions to new regulation and legal framework. Using an extensive dataset from the largest p2p-lending platform in the U.S., we provide an empirical framework, contributing to several open questions regarding both platform design and market regulation. First, we identify the main drivers of risk in the p2p market for unsecured personal loans. These results also contribute to the question of borrower selection bias in the crowdlending market. Furthermore, we investigate the appropriateness of the risk-adjustment of interest rates by the platform. Thereby we address the effectiveness of the credit scoring system, implemented by the platform after switching from an auction mechanism to a posted-price regime. Finally, we provide evidence on how well the investor crowd evaluates potential risk driver, while funding loans. We find that, while the individual investor is inexperienced and most likely not able to derive risk implications from the loan characteristics, the crowd as a collective makes smart decisions and can process informations about the loans correctly. This evidence points towards a hidden process of learning by imitating, communicating and experience within the crowd.
Most regulatory approaches currently discussed, face a trade-off between eliminating risk associated with crowdlending as well as introducing certain quality standard and harming the business model of the platform by limiting their ability to service certain loans directly or by inducing costs. Our results introduce a fundamentally different approach on regulating the crowdlending market. We suggest that a democratization of the market characterized by liberal regulatory constraints could lead to a welfare optimum, both for lenders and borrowers.

JEL Classification: G11, G2, O31

Keywords: crowdfunding, p2p-lending, crowd intelligence, regulation, posted price, credit scoring

WORKING PAPER

15th of September 2014

1 Introduction - The crowdfunding market

Crowdfunding is a fast growing worldwide phenomenon with approximately \$2.7 billion funds raised in 2012 and a forecast of at least \$5.1 billion for 2013[1]. Thereby, crowdfunding can be divided into three major categories. The first stream are money-for-product platforms, where early-adopters enable the entrepreneur to start production by pre-ordering the product, motivated by price discrimination against later retail customers. A very popular example is Kickstarter with 73,598 successfully funded projects and a total of 1.397 billion dollars raised [2]. However, there are numerous examples that this first category also works for financing almost everything, from charitable causes to research projects.

The second stream are so called crowdvesting platforms, that allocate equity shares of mostly start-up companies based on specific valuations made by the platform. This form of crowdfunding is subject to the strongest regulatory hurdles and is therefore in a more developing stage compared to the previously mentioned. To date the Australian equity platform ASSOB is, with \$141 million funds raised, at the top of the ranking, compared to Seedmatch, the only real investing platform in the world, with €20.78 million invested capital [3]. For the further development of this form of crowdfunding, advances in regulation could be a determining factor Cumming [2013]. In the U.S. the JOBS (Jumpstart Our Business Startups) Act has addressed the needs of the equity crowdfunding community by allowing platforms to raise up to \$50 million from up to 2000 shareholder in a private placement under rule 505[4]. German platforms that raise funds above €100.000 have to choose special equity share types to avoid the costly prospectus liability and have switched from silent partnerships to "patriarchisches Darlehen"[5]. Due to the sensitivity of the information contained in the business plans of the innovative start-up companies, commonly funded by these platforms, the data availability for research on the matter of equity crowdfunding is very limited.

The third stream are peer-to-peer lending platforms that mostly fund private unsecured loans. The most prominent examples in the U.S. are Prosper.com and LendingClub.com (henceforth LC [6]). LC issued \$ 6.03 billion in loans to date and \$ 1.165 billion just during the third quarter of 2014. In 2013 public figures like John Mack (Morgan Stanley/KKR) and Lawrence H. Summers (Harvard/World Bank) are represented on the board of directors of LC [7]. In May, 2013 Google Inc. made a substantial investment in LC at a valuation of \$1.55 billion[8]. While the platform has undergone several improvements and policy changes, many due to improved legislation and clearance from the SEC, there are also various external factors that have catalyzed the enormous growth of peer-to-peer lending (henceforth p2p) .

[1] According to The Crowdfunding Industry Report '13 by Masssolution.com

[2] Statistics as of 14.11.14 from http://www.kickstarter.com/help/stats

[3] As of 14.11.14 from http://www.assob.com.au and www.seedmatch.de

[4] For more information see http://www.sec.gov/spotlight/jobs-act.shtml

[5] See Hornuf and Klöhn [2012] for a detailed examination of the legal framework in Germany.

[6] Lending Club = LC

[7] See https://www.lendingclub.com/public/board-of-directors.action

[8] See the WSJ for a good overview on investment in LC at:
http://online.wsj.com/news/articles/SB10001424127887323628004578458892382014094

Since 2011, private investors face historically low interest levels, while at the same time, banks and credit card companies cannot provide borrowers of unsecured personal loans with favorable conditions. In December 2013 the return on a 3 and 5 year constant maturities treasury bills was 0.81% and 1.74% respectively[9]. While the Federal Reserve's G.19 report on consumer credit released in December 2013 states that the average annual percentage rate on credit cards was 13.11% in August 2013. According to the Student Monitor Financial Services, 17% of the students were subject to an interest rate increase on their credit card in spring 2012. Moreover small business owners had to pay 15.6% on average on their business credit cards[10].
Another problem for many households are sever penalties for late payments. The research center www.demos.org reports that 24% of low and middle-income households reported increased interest rates due to late payments during 2008 and 2012.
Furthermore, according to FINAR[11] there is evidence for discrimination in credit card rates against colored and female customers.

P2p-lending platforms could therefore fill a gap in the market and facilitate personal loans between personal borrowers and small investors, while cutting out financial intermediaries and running a low cost/fees framework.
In fact, the business model of p2p-lending platforms today could be described as being based on exploiting a regulatory advantage. P2p-lending platforms have no exposure to the risks of the loans they facilitate as these are passed on to the investors on the platform. Therefore, they are not subject to the regulatory capital requirements many consumer banks and credit card companies face. A recent study by mastercard shows, how severely the Basel III framework will affect credit card companies in multiple ways [12].

Evidently this billion dollar industry has alarmed regulators all over the world. Their major concern is that investing in unsecured personal loans might be too risky for the inexperienced, small private investor. This concern is reasonable, as neither in the US and UK, nor in Germany, p2p borrowers have to provided any kind of confirmation about net worth, income or experience.
Some of the regulatory effort by the SEC in the US or the FCA in the UK is directed at providing small investors with the necessary information to asses risk and protect them from potentially dangerous investments. Under US regulation the loans originated on the LC platform are registered as securities and the platform has to periodically update a prospectus with the SEC.

This regulatory approach does not provoke a major trade-off problem as most platforms are happy to comply, as the compete with other platforms along these channels anyway.
Platforms like LC and Prosper.com undertake great efforts to make the past performance of loans as transparent as possible and provide inexperienced borrowers with all the information they need to make save and sound investment decisions. However, extensive documentation and disclosure standards and increased investor services could potentially be a threat to the low cost, low fees business model of the p2p lending platforms.

---

[9] Source: http://www.federalreserve.gov

[10] http://www.nsba.biz/wp-content/uploads/2012/07/Access-to-Capital-Survey.pdf

[11] http://www.finra.org

[12] See: www.mastercardadvisors.com/$_assets/pdf/basel_final.pdf$

Another regulatory approach aims at introducing certain minimum requirements for loan applications filed on crowdlending platforms. As illustrated in this chapter before, it is part of the business model of p2p-lending platforms to service loans that are no longer attractive for financial institutions that are subject to regulatory constraints. However, if a negative selection bias leads the p2p-lending platforms to become basins for loans that could not be funded anywhere else, the p2p lending market would not be sustainable. In practice, all the large p2p-lending platforms have self-imposed rather strict criteria for a loan to become listed. A great deal of loan applications is rejected, as we address later in this paper. Thereby the platforms face a tradeoff between maximizing the volume originated on the platform, which is directly linked to the fees and therefore the profit of the platform, and the default rates and risks that their investors face on the platform.
Interestingly, so far a "race-to-the-bottom" can not be observed amongst p2p-lending platforms. Still, any regulation aimed in this direction could potentially limit the ability of the platforms to service these loans and therefore would also limit the size of the total market for p2p loans.
This paper contributes to the prior literature on p2p lending and investor behavior on crowdfunding platforms in several ways. First, we provide insight on the performance of the posted-price credit rating system implemented by the platform after abolishing the auction mechanism in 2009. Second, we derive inferences that can be drawn from loan and borrower characteristics on credit risk. Finally, to the best of our knowledge, we are the first to study collective intelligence in the crowdfunding market, providing evidence that there is wisdom in the crowd.

The remainder of this paper is structured as follows. After briefly reviewing the literature on crowdfunding we develop the motivation and research questions for this paper with regard to the mechanisms on the LC platform. We then describe the research framework and present descriptive results from the data we have obtained. Subsequently, we present the aggregated results from our risk assessment and our analysis on the wisdom of the crowd.

2 Literature

The early literature on crowdfunding focused mainly on money-for-product platforms, as these platforms faced less regulatory constraints and emerged very early[1]. However, many of the main results are also applicable to crowdfunding via online platforms in general.
While the literature on relationship banking an venture capital suggests that geographical proximity is relevant for investment, this effect might be less prevalent for a platform design that facilitates transaction solely over the internet Manson [2007], Degryse and Ongena [2005]. Agrawal et al. [2010] find for musical projects that, with an average distance of 3,000 miles between investor an artist, spatial proximity might play a reduced role. The fact that local investors invest early and independent of the behavior of other investors can be explained by personal connections (family&friends). In their study about Kickstarter projects, Mollick [2013b] and Mollick [2013a] use a locational Gini coefficient and population data. They provide descriptive evidence that crowdfunding projects are geographically concentrated and the location plays a major role in the success of the projects.

During the early years of p2p lending many platforms relied on an auction process to allocate funds and even set rates. Therefore early work on the dynamics of the p2p funding process focused around the auction mechanism (Comment and Jarrell [1991],Chen et al. [2011], Ceyhan et al. [2011] and Zhang and Liu [2012] as well as Lee and Lee [2012]).

Herzenstein et al. [2011] reveal that a 1% increase in bids can increase the likelihood of an additional bid by 15% when the project has not received full funding yet. They also find a positive relation between strategic herding during the funding process and subsequent loan performance.
Puro et al. [2010] and Puro et al. [2011] provide evidence that lenders in p2p auctions engage in strategic bidding and that their behavior can be inhomogeneous.

Since both of the major p2p lending platforms have changed to a posted-price mechanism after 2009, some of this research is no longer applicable to the platforms design in 2014.
The dynamic of the funding process on the platforms shows repeating patterns of investor behavior. Agrawal et al. [2010] show that funding is highly skewed with less than 1% of project creators accounting for more than 73% of the funds on the platform, Agrawal et al. [2013].
Bayus and Kuppuswamy [2013] highlight the importance of past backer support and the project's funding cycle on the dynamic of receiving additional backer support.
Their results provide a somewhat differentiated view on herding as they suggest a decrease in herding and a so called "bystander" effect during the middle of the funding cycle of a project.

A main characteristic of online platforms is that they often develop a social dynamic in terms of friendship networks. Lin et al. [2011] show that under the old platform design of Prosper.com friendships between borrowers and other users of the platform influence their ex-ante outcomes. They advocate the view that the social dynamic on the platform can mitigate concerns about information asymmetries between borrowers and lenders, even in the absence of a financial intermediary that produces and preserves soft information.

[1] ArtistShare(2003), Sellaband(2006) and Kickstarter(2009) are prominent examples

An extant stream of literature suggests, that social networks in online platforms have a major effect on the funding success (Herrero-Lopez [2009],Freedman and Jin [2008], ?, Zhu and Iansiti [2009], Everett [2010] and Collier et al. [2010]).
There is evidence on the role of group leaders Klein [2010],Everett [2010] and high reputation investors Kim and Viswanathan [2013b], Kim and Viswanathan [2013a], as well as their incentives Hildebrand et al. [2011] in the networks. Due to policy changes some of the results might no longer be applicable to the 2014 platform design as for example group leader rewards have been eliminated on all major platforms.

A major stream of literature deals with the role of information asymmetries on crowdfunding and how they can be overcome. As most platforms have no external control systems, information asymmetries are very prevalent in crowdfunding. Therefore signaling through soft and non-verifiable information is found to be essential by several studies Garman et al. [2008], Iyer et al. [2009], Iyer et al. [2011], Zhu and Iansiti [2009], Krishna et al. [2009] and Zhu and Iansiti [2009].
There are indications that investors conduct search Garman et al. [2008] and aggregate sources of information Iyer et al. [2011] to infer infer creditworthiness.
Personal attributes like, gender, race and age that are conveyed by pictures or texts can have an effect on funding according to Pope and Sydnor [2008] and Ravina [2012] while Herzenstein and Andrews [2008] finds that financial information have a stronger effect.
Consequently Caldieraro et al. [2011],Yum et al. [2012] and Lambert and Schwienbacher [2010] find that borrowers who withhold non-verifiable information exhibit a significantly lower likelihood of delinquency, whereas Rainer and Stefanie [2010] finds only limited benefit from disclosing private information. Caldieraro et al. [2011] study the effect on non verifiable information, like the wordcount in the description of a loan, on the funding success of the loan. In accordance with the counter signaling theory they find that borrowers, providing no description for their loan, are more successful. Siegel and Young [2010] use photographs associated with loan listings on prosper to study the effect of trust on funding success and the terms of the loans. They find that trustworthy borrowers have a higher probability of funding and pay less interest on their loans. Ravina [2012] show that the effect of beauty, race age and other personal characteristics affect lenders decision while controlling for other formal borrower attributes like employment history and home ownership status.

However early evidence from Herzenstein and Andrews [2008] suggest that lenders behave rational and demographic attributes have a much smaller effect on their decision making than indicators like financial strength. Using early data from Prosper.com in 2008 and 2010, Freedman and Jin [2010] provide evidence that lenders learn over time both from responding to the performance of their own portfolios as well as from watching the market. While they also find great heterogeneity in the sophistication of individual lenders, they show that less sophisticated lenders improve over time and a later cohorts of new lenders have already learned from the past performance of the loans.

Askira Gelman [2013] address potential shortcomings in the LC loan assessment process as income verification is not mandatory and properly implemented in the rating. They argue income verification traditionally is a vital tool in the risk assessment of unsecured personal loans. Surprisingly, their purely descriptive results, using loan data until September 2012, indicate, that loans without income verification are preferred by investors and perform better in the default case. Using data from the auction era of Prosper.com Iyer et al. [2011], Iyer et al. [2009] find that lenders infer risk drivers from borrower attributes within credit categories.

3 Hypotheses

While banks employ very sophisticated techniques in screening and evaluating loan applications, as well as managing their risk actively by secularization and credit default assurance, for private investors the asset class of personal unsecured loans is new territory. There is an ongoing debate about the benefits and perils, accompanied by the request for stricter regulation and an applicable legal framework [1].

To a certain extent platforms like LC have self-imposed certain regulations and safety mechanisms to promote safe transactions on their platforms. In a first step, LC assesses if a loan application fits the strict requirements set by the platform. According to the prospectus filed by LC the initial criteria are a FICO score above 650, a deb-to-income ration (excluding mortgage) below 35%, a minimum credit history of 36 months, 6 or less inquiries in the last 6 months; and at least 2 revolving trade accounts [2]. LC is so strict in their assessment that about 90% of the loan applications are rejected [3].

After switching from an auction pricing mechanism to a posted-price mechanism in 2009, LC assigns interest rates to each borrower based on an internal system of credit classes. If the application passes the minimum listing requirements an interest rate is assigned to the loans according to a system of 35 risk grades A1-G5. The interest rate consists of a base rate of 5.05% and a risk&volatility adjustment uniquely determined by the credit grade for each loan. The risk evaluation and credit grade assignment process works as follows. Initially, LC assigns a model rank between 1 and 25 to each loan application based on the applicants FICO score and "certain other credit attributes[4]" that LC will not further disclose. In a second step, the loan sub-grade is adjusted regarding the requested loan amount and the maturity (36 or 60 months) of the application [5].
If a borrower misses a payment LC starts a collection process trying to contact the borrower and collecting the outstanding amount. The status of a loan reveals for how long the borrower has been late on his payments. The status "in grace period" indicates a late loan before moving to "late(16-30days", "late(31-120 days)" and finally "default(120+ days)" before the loan is "chargedoff". Furthermore, a borrower may prepay a loan at any time without a penalty or fee.

In theory, all the investor should care about with regard to risk is the credit grade of the loan as it is designed to distinctly assign the appropriate risk-adjusted interest rate to the loan. This raises the question of the effectiveness of the credit scoring system employed by LC. It can be assumed that there is still a fair amount of variation in terms of risk within each credit category. Therefore, the first part of this paper studies the sources of risk and develops several models to unveil the significant risk drivers accounting for the variation within the credit classes.

---

[1] See for example: Bradford [2011] or http://www.pepperlaw.com/pdfs/Peer-to-Peer_Meetup_Slides_121813.pdf as well as http://www.ft.com/cms/s/0/9c7a4896-aac5-11e3-9fd6-00144feab7de.htmlaxzz2xkZUORgu.

[2] Information was obtained from the latest prospectus available at: https://www.lendingclub.com/info/prospectus.action

[3] http://www.orchardplatform.com/rejected-loans-on-lendingclub/

[4] https://www.lendingclub.com/public/how-we-set-interest-rates.action

[5] Information on the posted-price mechanism was retrieved from: https://www.lendingclub.com/public/how-we-set-interest-rates.action

Kiefer and Larson [2006] point out an selection bias when credit scoring models are on data from extended loans and investigate the value of inference drawn from rejected applicant's characteristics. Using data from rejected applicants, Barakova et al. [2011] demonstrate that neglecting this information can affect the precision of credit scores. Ciampi et al. [2009] demonstrate that limited information on small business in Italy affect the predictive power of credit scoring models and advise the use of specialized models.
Glennon, Kiefer, Larson, and Choi [Glennon et al.] study the performance of credit scoring models and find, that while the model perform rather well in ranking the creditworthiness of individuals, they lack accuracy in predicting future delinquencies or defaults for groups of borrowers.

H1: Borrower and loan characteristics from the credit file significantly influence relevant risk metrics even within credit grades.

A key question associated both with platform design as well as the discussion about regulation is how well the crowd performs in assessing risk themselves and invest their money sensibly. It is straight forward that individual investors could also come up with the idea not to solely rely on the predictive powers of the credit score but to engage in loan picking behavior in order to improve returns. In fact, all the information in our dataset are also available to investors upon funding. Most investors use filters in order to select loans with certain characteristics, they suspect to be beneficial for the performance of the loan.

It remains an open question how well small, inexperienced investors can process this information and make correct inferences from the information provide for them. Screening popular blogs an forums associated with p2p lending reveals, that many investors discuss which characteristics are relevant and share their experience[6]. A stream of literature on collective intelligence argues that dynamics within a crowd can facilitate beneficial outcomes and outperform fixed structures Surowiecki [2005], Arazy et al. [2006], Brabham [2008], Malone et al. [2009], Yi et al. [2012], Martin [2012] and Prpic and Shukla [2012]
However, some investor obviously feel unable to process this information as several commercial services have emerged, offering assistance for picking loans and building portfolios that outperform their peers within the credit grade.

H2: The crowd as a collective, can correctly process all information available to them and infer the right conclusion regarding the risk assessment of a loans.

Consequently, the second part of this paper tries to provide empirical evidence on how well the crowd perceives risk factors in a loan application.

[6]See for example: http://blog.dmpatierno.com/post/3161338411/lending-club-genetic-algorithm or http://www.beatingbroke.com/lending-club-selecting-investments/

4 Methodology

In order to study the total risk associated with an investment in a loan, we split the total risk up into several components. The most obvious risk associated with an investment in a loan is the default of the loan. Obviously, the exposure of an individual investor to default risk also depends on his diversification strategy. Moreover, as all LC loans are repayed as an annuity in fixed monthly installments, the survivaltime until a default occurs is vital for the exposure of the investor at default and therefore the assessment of the loss-given-default of a loan. Furthermore, we study prepayment as a potential source of risk as it is rather prevalent on the platforms due to the lack of sanctions for the borrowers. Apart from the obvious reinvestment risk this could impose a potential threat to the diversification strategy of an investor if the prepayment is asynchronous across loan and borrower characteristics.

4.1 Probability of Default

In order to study the probability of default (henceforth PD) for a loan, we employ a logistic regression methodology. The full model is specified by equation 4.1:

$$
\begin{aligned}
P_i(ChargedOff) =& \Phi(\alpha + \beta_1 FundedAmount_i + \beta_2 AnnualIncome_i + \beta_3 DTI_i + \beta_4 FICO_i \\
&+ \beta_5 FractionalFunding_i + \beta_6 Term_i + \beta_7 IncomeVerified_i + \beta_8 TitleWordcount \\
&+ \beta_9 DescriptionWordcount + \gamma_1 \sum_{j=1}^{4} Ownership\_dummies_{i,j} \\
&+ \gamma_2 \sum_{k=1}^{13} Purpose\_dummies_{i,k}) + \gamma_3 \sum_{l=1}^{34} Subgrades\_dummies_{i,l}) \\
&+ \gamma_4 \sum_{m=1}^{3} Employment\_dummies_{i,m} + +\gamma_4 \sum_{n=1}^{5} YEAR\_dummies_{i,m}
\end{aligned}
\tag{4.1}
$$

Where $P_i(ChargedOff)$ refers to the probability that loan i is chargedoff. All variables are explained in table 8.1.

As we are dealing with heavily censored data, only 3.41% of the loans in the sample have finally matured, predicting default for the full sample imposes several challenges. In the literature several competing concepts are discussed in the p2p-lending background Iyer et al. [2011], Haltiwanger et al. [2014].

As the appropriateness of a duration model for estimating defaults based on personal loan data is still discussed controversially, we decide in this version of the paper to rely on the probit model. In this context however, we have to be careful with the interpretation of zeros in the data due to censoring. In order to mitigate this problem we estimate our model only for loans with a maturity of 36 months that were originated in the years 2010 or 2011 with standard errors clustered at the year and state level.

4.2 Time to Default

$$
\begin{aligned}
SurvivalMonths_i = & \beta_1 FundedAmount_i + \beta_2 AnnualIncome_i + \beta_3 DTI_i + \beta_4 FICO_i \\
& + \beta_5 FractionalFunding_i + \beta_6 Term_i + \beta_7 IncomeVerified_i + \beta_8 TitleWordcount \\
& + \beta_9 DescriptionWordcount + \gamma_1 \sum_{j=1}^{4} Ownership\_dummies_{i,j} \\
& + \gamma_2 \sum_{k=1}^{13} Purpose\_dummies_{i,k}) + \gamma_3 \sum_{l=1}^{34} Subgrades\_dummies_{i,l}) \\
& + \gamma_4 \sum_{m=1}^{3} Employment\_dummies_{i,m} + +\gamma_4 \sum_{n=1}^{5} YEAR\_dummies_{i,m}
\end{aligned}
\tag{4.2}
$$

The survivaltime in the model is the timespan from origination until the last payment of the defaulted loan in months. All other variables are, again, explained in table 8.1. The dependent variable is count data comprised of non-integer values between one an 44 months.
Therefore, we estimate the equation using a count data model. In order to test for overdispersion in the data we fit a negative binomial model both to the full sample and the matured sample, observing alpha values of 0.44 and 0.31. The likelihood-ratio test indicates that both alphas are different from zero at the 1% level. Therefore, the equidispersion property of the poisson distribution $E(y|x) = var(y|x) = \mu$ is violated and the negative binomial model with its less restrictive properties $E(y|x) = var(y|x) = \mu + \alpha\mu^2$ seems more applicable. Furthermore, as the chargedoff sample does not contain zeros in the survival months count variable, we adjust the probability function to fit the zero-truncated structure of our data as stated in equation 4.3, Rodriguez [2013].

$$
Pr(Y = y|\lambda, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{y!\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda} \right)^{\alpha^{-1}} \left(\frac{\lambda}{\alpha^{-1} + \lambda} \right)^{y} \tag{4.3}
$$

4.3 Loss given Default

For the finally chargedoff loans in the matured sample we calculate the loss-given-default according to 4.4.

$$
LGD_i = 1 - \left(\frac{ReceivedInterest_i + ReceivedLatefees_i + ReceivedPrincipal_i + Recoveries_i - CollectionFees_i}{FundedAmount_i} \right) \tag{4.4}
$$

In accordance with the risk evaluating framework of the world bank (Basel II/III), our LGD measure for loan i is defined as the percentage loss of a defaulted loan. In order to identify the determinants of LGD for defaulted loans, the model specified in equation 4.5 is estimated by OLS. Note that $SurvivaltimeMonths$ is now incorporated in the model.

$$
\begin{aligned}
LGD_i = & \alpha + \beta_1 FundedAmount_i + \beta_2 AnnualIncome_i + \beta_3 DTI_i + \beta_4 FICO_i \\
& + \beta_5 FractionalFunding_i + \beta_6 Term_i + \beta_7 IncomeVerified_i + \beta_8 TitleWordcount \\
& + \beta_9 DescriptionWordcount + \beta_{10} \textbf{SurvivaltimeMonths} + \gamma_1 \sum_{j=1}^{4} Ownership\_dummies_{i,j} \\
& + \gamma_2 \sum_{k=1}^{13} Purpose\_dummies_{i,k}) + \gamma_3 \sum_{l=1}^{34} Subgrades\_dummies_{i,l}) \\
& + \gamma_4 \sum_{m=1}^{3} Employment\_dummies_{i,m}
\end{aligned}
\tag{4.5}
$$

4.4 Prepayment risk

As borrowers on LC can prepay their loans at any time without any penalty, understanding the patterns of prepayment is vital for an comprehension of expected returns. $Prepay_i$ indicates if loan i was prepayed before its fixed maturity.

$$
\begin{aligned}
P_i(Prepay) = & \Phi(\alpha + \beta_1 FundedAmount_i + \beta_2 AnnualIncome_i + \beta_3 DTI_i + \beta_4 FICO_i \\
& + \beta_5 FractionalFunding_i + \beta_6 Term_i + \beta_7 IncomeVerified_i + \beta_8 TitleWordcount \\
& + \beta_9 DescriptionWordcount + \gamma_1 \sum_{j=1}^{4} Ownership\_dummies_{i,j} \\
& + \gamma_2 \sum_{k=1}^{13} Purpose\_dummies_{i,k}) + \gamma_3 \sum_{l=1}^{34} Subgrades\_dummies_{i,l}) \\
& + \gamma_4 \sum_{m=1}^{3} Employment\_dummies_{i,m}
\end{aligned}
\tag{4.6}
$$

4.5 Expected Returns

In a second step this papers tries to apply the findings from the previous section to the individual loans by predicting their expected returns. The differentiation between stated returns and expected returns is vital for our final analysis of intelligent investor behavior.

The approach builds on our findings from the previous section on default probabilities to predict expected returns for the full sample according. In a first step we need to predict the default probability for each individual loan. The predictions for default probabilities $Prob\widehat{(default)}_i$ are derived from the previously discussed robust Probit model.
In case the loan does not default, and therefore is either prepaid or fully paid at maturity, the return equals the interest rate of the loan under the reinvestment assumption.

For the case of default, an expected return needs to be predicted for the loan.
The default case is modeled employing the predicted time to default based on the zero-truncated-negative

binomial model trained over the chargedoff sample from the previous section. The predicted survival time in months $\widehat{SurvivaltimeMonths_i}$ multiplied by the monthly installments (principal+interest) of the loan, are modeled as the received cash-flows over the useful life of the loan. In order to come up with a compounding rate, comparable to the interest rate under the reinvestment assumption, we calculate the future value of the annuity. Equation 4.7 describes the full derivation of our expected return concept.

$$\begin{aligned}E(r_i) =&((1 - Prob(default)_i) * InterestRate_i) + Prob(default)_i * \\ &\left(\left(\frac{(\frac{(1+InterestRate_i)^{\widehat{SurvivaltimeMonths_i}}-1}{InterestRate_i}) * monthlyInstallment_i}{FundedAmount_i}\right)^{12/\widehat{SurvivaltimeMonths_i}} - 1\right)\end{aligned} \tag{4.7}$$

We are currently running a series of robustness checks and have estimated the expected return in various alternative ways. However the main results regarding the wisdom of the crowds analogy of this paper are robust and have not changed subject to the method we employed.

5 Data and descriptive results

The data was obtained from www.lendingclub.com in February 2014, and contains complete loan details for all loans issued on the platform until that time. Personal identifiable information however, have been removed by LC to protect member's privacy rights. Due to the transition from an auction to a posted price regime in 2009, and several other major policy changes on the platform, we restrict our analysis on loans issued from 2010 on. Apart from that we make no other outlier correction.
The sample contains 245,795 loans issued between the 5th of January 2010 and the 21th of February 2014.

It has been widely discuss how LC integrates in the loan market and which kind of borrowers choose p2p-lending as a borrowing alternative. By design, LC loans are unsecured personal loans as they are paid out to individuals that do not provide any form of collateral. Hence, the LC service naturally competes with financing alternatives for consumption and purchases not suitable for collateralization.
From the descriptive results we can see that the purpose of the loans is in 57,6 % of the cases debt consolidation and that debt consolidation combined with credit card debt make up nearly 80 % of all loans.
Therefore it is not only borrowers that have no other source of financing, but rather borrowers that use LC to refinance existing debt. However, he median overall FICO score of 694 in our sample is significantly lower than the median score of 711 reported by FICO.com[1] for the US. According to the US Census Bureau in 2010, 85.66 percent of all US citizens have a personal income of less then $70,000. In the full sample we have a mean annual income of $72,183 and a median of $62,000 which would indicate that LC borrowers are among the richest 18% of all US citizens.

Table 8.3 indicates that the nominal loan amounts vary by the purpose of the loan and that credit card and debt consolidation have higher mean amounts and lower FICO scores (untabulated both differences confirmed by a t-Test at the 1% level).
As expected table 8.4 demonstrates that DTI rations as well as duration rise and FICO scores decline, going from the best credit grades to the worst. Rather unexpected is the rise in annual income and income verification in the worst credit grades F and G. This might be in line with the major criticism abut credit scoring models in the US. Critics argue that even individuals with a healthy income, can get a bad scoring due to a lifestyle of frequently moving or simply short credit histories, that are regarded negative in the scoring model.

Considering the matured sample of finally charged-off or fully paid loans, table 8.7 shows the expected correlations for funded amount, annual income, dati ratio and the FICO score. Table 8.8 compares the charged-off sample with the fully paid sample. As expected the loan amounts and interest rates are higher while the annual income is lower in the charged-off sample. Interestingly the charged-off sample has a significantly higher percentage of income verification.

For the sub-sample of finally charged-off loans we can calculate the loss given defaults according to the previously discussed equation 4.4. Table **??** reveals that the credit grades assigned by LC actually reflect the LGDs in the default case with a slight exception for grade G (untabulated a t-test confirms the significance of the differences, at least at the 10% level).

[1]http://bankinganalyticsblog.fico.com/

In order to study wether the credit classes are also indicative of the survival time in month, we have to split the sample by the two possible durations, 36 and 60 months. Table 8.10 and table 8.11 reveal that the credit classes are not a sharp indicator of the survival time.

Table 8.5 illustrates that the credit grades assigned by LC correctly represent the level of expected return and the default probability that we have estimated for the full sample, excluding the 60 months term loans. As expected returns decline alongside the credit grades while the default probabilities are increasing.

6 Main results

Table 9.1 aggregates our results on the determinants of risk for the loans. The first column presents marginal effects from the robust probit model with default as the dependent variable, estimated for a sample of 36months term loans, originated in 2010 or 2011 as described earlier. The option to fund a loan as a whole was only introduced in 2012, therefore the corresponding variable is omitted in the model. As we are controlling for the sub-grade of the loan in the model, that should incorporate all the risk associated with the loan, we should not expect to find any significant coefficients.

The model however confirms, that loan amount, annual income, dti-ratio and FICO score have significant marginal effects on the default probability with the expected signs. Judging from the magnitude of the effects, it appears that the dti-ratio is not as well incorporated in the credit grades as the other factors.
Furthermore, compared with the reference category debt consolidation, an investment in a small business loan would ceteris paribus increase the risk of a default by 7.7%. We also find a significant increase in default risk for the purpose categories medical, moving, other and renewable energy. Compared to debt consolidation, loans with the purpose credit card have a 3.07% decreased risk of defaulting.
In the model estimated over the training sample we cannot observe any significant effects from the length of the title or the description, as could have been expected from the counter signaling theory. Also the home ownership status of the borrower and longterm employment have to significant effects. However the unemployment dummy indicates a significant marginal risk increase of 6.3% at the 1% level.

The second column shows the coefficients from an zero-truncated negative binomial model with survival time in months as the dependent variable estimated for the charged-off sample.
As the loan is repaid in fixed month installments, we are also concerned *when* the loan defaults as opposed to *if* he defaults with regard to our expected return measure. In terms of our continuous variables, the dti-ratio again has the biggest magnitude. A one-unit increase of the dti-ratio would, ceteris paribus, decrease the the number of months until default by $\exp(-0.3242) = 0.7231$ times. Surprisingly FICO score does not show the expected sign, but only to a very small magnitude as a one unit increase would reduce the number of months times $\exp(-0.0012) = 0.9988$.
Interestingly the purpose dummies do not all influence survival time to the same extend as they influence defaults. While Home improvement has not been significant before in the default model, the coefficients on renewable energy and credit card lost their significance in the model on survivaltime. The coefficients on medical, other and small business confirm that risk increasing effects for these categories from the default model. Curiously, unemployment has no significant effect on the survival time but short term employment does.

The next piece in our loan risk puzzle is the loss that occurs if default actually happens. The third columns presents the coefficients of an OLS model with LGD as the dependent variable, also estimated for the charged-off sample. In terms of the continuous variables the dti-ratio again has the greatest effect in terms of magnitude. For the first time the dummy controlling for the verification of income becomes significant, reducing the LGD in the default case. Note again, that we are controlling for funding time in the model which reduced the LGD significantly, as expected. While the purpose categories other and small business so far have always increased the risk of the loan, they now significantly reduce the LGD.

Figure 8.2 displays the means of the prepay dummy by subgrade, suggesting a greater prepay risk at both ends of the spectrum of credit grades. Consistent with our previous finding, the model in the last column, presenting the marginal effects on prepayment, suggest that most of our well known risk driver significantly decrease the probability of prepayment.

We can conclude that while controlling for each individual subgrade and year fixed effects there is still considerable variation regarding the risk metrics studied in this paper within the credit classes. This confirms our conjecture that investors can improve their expected return through loan picking.
A smart investor would therefore first select a credit class which is appropriate for his risk tolerance, as we have seen from the descriptive results that credit classes do represent overall risk correctly. Then he would set filters to narrow the loans available for investment down, based on the criteria he thinks is beneficial for him. Following this strategy, a well diversified investor could be able to increase his expected returns above the average appropriate for the credit class he is investing in.
Due to the privacy policy of LC, we do not have the data to investigate the behavior of an individual investor. However we are not interested in the lucky loan picking of any individual, but rather in the dynamic of the whole crowd.

If the crowd is able to infer the right conclusions from the past performance of the loans, one would expect that the crowd always allocates its funds to the loans first that they consider beneficial for them. This should become observable in the fundingtime of the loan, define as the time in days from the listing until the funding of the loan.
As described previously we use our findings about if and when a loans defaults, to estimate an expected return measure. Note that as we have trained our default model only for 36 months term loans, we can only make sensible predictions regarding default for loans of this maturity. 60 months loans are therefore excluded from the analysis of expected returns due to the lack of historic loan performance data. Untabulated we have run several robustness checks, introducing day of the week dummies etc. with consistent results.

Table 9.2 intents to compare the results from an OLS with fundingtime as the dependent variable in the first column with the results from another OLS model with our expected return measure as the dependent variable in the second column.
Borrower or loan characteristics, that increase the expected return, should decrease the time the loans needs to receive funding in order to indicate intelligent behavior on behalf of the crowd.
The coefficients on loan amount and funding in fractions can not be interpreted in this context as there is an underlying process that links these coefficients to fundingtime.
Analyzing the results we can conclude that the crowd correctly infers the effects of the dti-ratio and employment on expected return.
In accordance with the counter signaling theory the crowd mistrusts long descriptions for the loans and verification of income, both increase the fundingtime significantly at the 1% level. Nevertheless a meaningful title, containing more than 2 words, significantly reduces the fundtime. However both loans with longer descriptions and titles actually have significantly higher expected returns.
As the dummies on home ownership status are all insignificant in the model on fundingtime, it can be sug-

gested that the crowd does not use this filter to select loans. However, compared to the reference category other, the home ownership dummies actually have a highly significant impact on expected returns.

The most straight forward criteria to filter loans is probably the purpose they are intended to be used for. The crowd misinterprets the effect of the purpose categories credit card, house and wedding regarding their positive effect on expected returns. On the other hand the crowd correctly infers the effect of the categories small business, other, moving and home improvement.
We have included funding time as a dependent variable in the model in order to test its effect on expected returns directly. The coefficient on fundingtime is significantly negative, indicating a certain smartness of the crowd.

Therefore we conclude that the crowd understands the relation of certain borrower characteristics and the riskiness of the loans. Furthermore the crowd is actively trying to exploit the variance within a credit category to gain excess returns. As a collective they process the information from the most popular filters correctly and invest accordingly. We descriptor this process as crowd intelligence.

7 Conclusion

The mechanism how loan contracts are facilitated between individual borrowers and a crowd of mostly inexperienced lenders over the internet, is a fairly new topic to economic research. Both the market structures as well as the regulatory framework are still evolving. The population of borrowers applying for p2p-loans might not represent the average credit card or personal loan customer due to a selection bias. Therefore, certain borrower characteristics might have a different effect in a p2p lending context compared to traditional credit markets. The effect of geographic dispersion and the performance of the collection process by the platforms might also contribute to the uncertainty associated with a p2p investment.
As correct risk-adjustment by the platforms plays a key role for the functioning of the p2p lending market, credit scoring models have to adapt to this specialties of the market.

We disentangle the credit risk puzzle of p2p loans by demonstrating inferences that can be drawn from soft and hard information regarding the loan and the borrower. Thereby, we find certain shortcoming in the credit scoring models employed by the platforms during our sample period. Our results suggest that lenders can outperform a naively diversified portfolio by intelligently selection loans by the right characteristics.

Furthermore we find strong evidence for selection and loan filtering behavior by the investor crowd and point out the characteristics that investors focus on. Finally, we investigate if the investors manage to draw the right inferences from the credit information available to them. In this respect, some of our results are ambiguous regarding individual characteristics. We find significant evidence that there is wisdom in the crowd as a collective.

With regard to the development of a regulatory framework, our results advocate a liberal, deregulation approach that relies on the market dynamic. We believe that leaving the decision witch loans are funded and at which terms to the market, would most likely lead to a welfare optimum. The democratic process of the market can be trusted to the crowd intelligance phenomene. While the individual investor might be

8 Appendix I - Descriptive and Summary Statistics

8.1 Figures

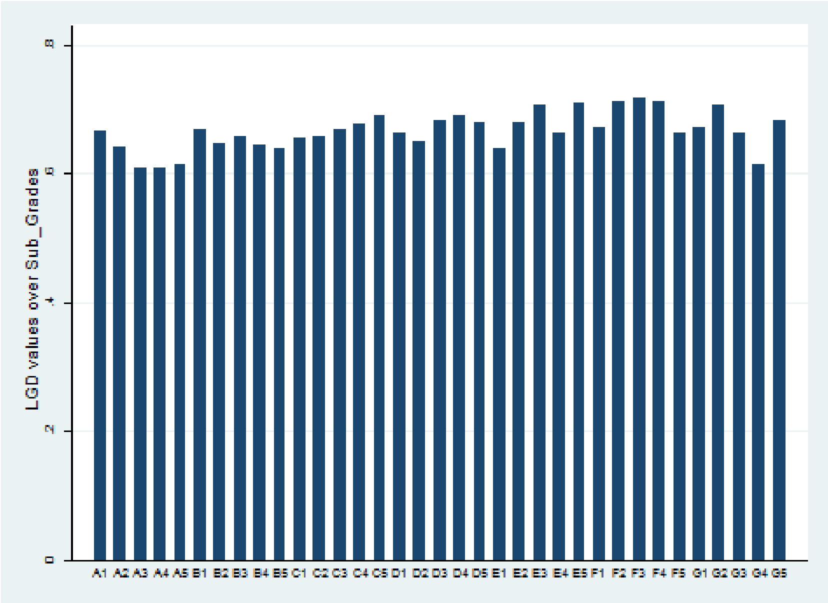

Figure 8.1: LGDs by Subgrade

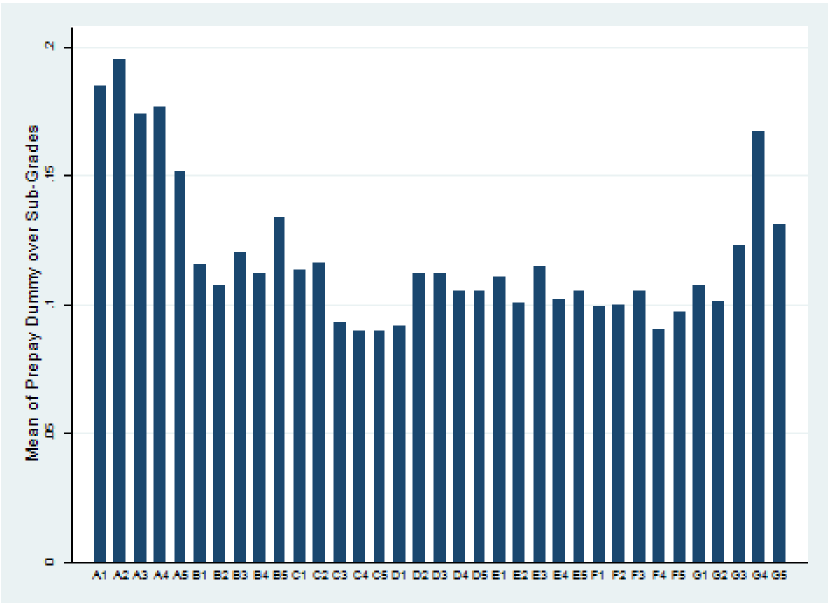

Figure 8.2: Prepay by Subgrade

Table 8.1: Variable Description

Variable Name	Description	Format
A. Continuous Variables		
Funded Amount	The total amount funded by investors for that loan at that point in time.	$ in thousand
Annual Income	The annual income provided by the borrower during registration.	$ in thousand
Debt-to-Income	The borrower's debt to income ratio, calculated using the monthly payments on the total debt obligations, excluding mortgage, divided by self-reported monthly income.	percent
Funding in fractions	The initial listing status of the loan. Possible values are "F" for fractional, "W" for whole-	dummy
FICO Score	The upper boundary of range the borrow's FICO belongs to at application.	Long
B. Dummies		
Income not verified	Indicates if income is verified by LC	Dummy
Emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.	Categorical
d_longempl	Indicates longterm employment: +10 years	Dummy
d_shortempl	Indicates shortterm employment: ¡= 3 years	Dummy
d_unempl	Indicates unemployment	Dummy
Term	The number of payments on the loan. Values are in months and can be either 36 or 60.	Dummy
Lenght of the title	Wordcount in Title $\geq 2 words$	Dummy
Lenght of the Description	Wordcount in Description $\geq 7 words$	Dummy
Home_ownership	Residential ownership status	Categorical
Purpose	Purpose of the loan	Categorical
C. Controls		
Average current balance of all accounts	Average current balance of all accounts	Long
Months since most recent account opened	Months since most recent account opened	Long
Months since most recent bankcard account opened	Months since most recent bankcard account opened	Long
Months since oldest installment account opened	Months since oldest installment account opened	Long
Num of accounts now delinquent	The number of accounts on which the borrower is now delinquent	Long
Num. of currently active bankcard accounts	Number of currently active bankcard accounts	Long
Num. of mortgage accounts	Number of mortgage accounts	Long
Num.of accounts opened (past 12 months)	Number of accounts opened in past 12 months	Long
Number of trades opened in past 24 months	Number of trades opened in past 24 months.	Long
Percent. of all bankcard accounts $\succ 75\% of limit$	Percentage of all bankcard accounts $\succ 75\% of limit.$	Percent
Ratio total current balance/high credit limit	Ratio of total current balance to high credit/credit limit for all bankcard accounts.	Percent
Total bankcard high credit/credit limit	Total bankcard high credit/credit limit	Percent
Total current balance of all accounts	Total current balance of all accounts	Long
Total open to buy on revolving bankcards	Total open to buy on revolving bankcards	Long

8.1.1 Full Sample

Table 8.2: Overview Status & Purpose

	Charged Off	Current	Default	Fully Paid	In Grace Period	Issued	Late (16-30 days)	Late (31-120 days)	Total
car	161	2238	5	1115	17	102	7	39	3684
credit_card	1042	43030	37	6115	197	3891	162	463	54937
debt_consolidation	4822	108470	165	19882	731	8154	478	1794	144496
educational	53	9	0	262	0	0	0	1	325
home_improvement	487	10306	14	2626	74	693	48	168	14416
house	93	960	1	403	9	49	7	24	1546
major_purchase	290	3771	8	1813	29	204	12	56	6183
medical	157	1540	7	570	16	164	10	30	2494
moving	125	1024	5	456	7	56	8	34	1715
other	895	8765	26	3279	75	537	52	198	13827
renewable_energy	29	138	0	60	3	9	1	4	244
small_business	621	2715	16	1197	28	168	24	111	4880
vacation	78	892	4	329	7	60	2	21	1393
wedding	144	1331	4	754	9	1	3	34	2280
Total	8997	185189	292	38861	1202	14088	814	2977	252420

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8.3: Variable Means by Purpose

	Funded Amount	Interest Rate	Annual Income	DTI	FICO	Income verified	Fractional funding	Term 36 months
car	7655.57	0.12	64571.31	12.57	723.30	0.25	0.90	0.75
credit_card	14128.88	0.13	71075.53	17.06	699.91	0.42	0.78	0.81
debt_consolidation	14889.52	0.14	71239.45	17.30	700.66	0.47	0.81	0.73
educational	6692.08	0.12	53471.37	11.21	718.31	0.19	1.00	0.95
home_improvement	13426.49	0.13	88862.17	13.56	715.79	0.44	0.83	0.73
house	15327.98	0.14	80245.43	12.46	719.70	0.47	0.87	0.72
major_purchase	9514.22	0.12	71944.07	12.77	719.98	0.32	0.87	0.80
medical	8686.88	0.15	71943.37	15.02	708.25	0.34	0.87	0.83
moving	7393.43	0.15	67218.96	14.17	706.30	0.30	0.89	0.88
other	9114.63	0.16	67212.28	14.94	706.78	0.36	0.86	0.79
renewable_energy	9648.16	0.14	75484.71	13.45	712.44	0.31	0.93	0.82
small_business	14920.85	0.15	82696.36	12.76	717.12	0.53	0.90	0.73
vacation	5877.19	0.15	63857.81	15.24	706.67	0.29	0.87	0.91
wedding	10468.11	0.14	69445.68	13.99	710.13	0.33	0.92	0.83
Total	13872.62	0.14	72090.15	16.51	703.20	0.44	0.81	0.76

Table 8.4: Variable Means by Credit Grade

	Funded Amount	Interest Rate	Annual Income	DTI	FICO	Income verified	Fractional funding	Term 36 months
A	12549.33	0.08	78853.34	14.21	742.22	0.32	0.80	0.97
B	12739.78	0.12	70039.19	16.32	705.55	0.38	0.80	0.89
C	14026.11	0.15	69138.40	17.27	692.82	0.48	0.81	0.69
D	14201.28	0.18	69343.12	17.42	685.81	0.48	0.84	0.66
E	18064.18	0.21	77266.85	17.65	684.57	0.62	0.83	0.33
F	19736.02	0.23	79615.29	17.87	680.70	0.69	0.81	0.15
G	22450.82	0.25	94228.25	17.55	677.32	0.71	0.83	0.06
Total	13872.62	0.14	72090.15	16.51	703.20	0.44	0.81	0.76

Table 8.5: Expected Returns by Credit Grade

	Interest Rate	expect. Return	Pr(default)	Fundingtime [in days]
A	0.08	0.04	0.04	8.35
B	0.12	0.05	0.08	7.96
C	0.15	0.04	0.12	7.67
D	0.18	0.03	0.16	7.92
E	0.21	0.02	0.21	7.80
F	0.23	-0.05	0.31	7.04
G	0.24	-0.14	0.43	9.14
Total	0.13	0.04	0.10	7.96

Table 8.6: Expected Returns by Purpose

	Interest Rate	expect. Return	Pr(default)	Fundingtime [in days]
car	0.11	0.06	0.07	7.56
credit_card	0.12	0.06	0.07	7.93
debt_consolidation	0.13	0.04	0.10	7.91
educational	0.12	0.06	0.14	7.55
home_improvement	0.12	0.04	0.09	8.23
house	0.13	0.05	0.09	8.89
major_purchase	0.12	0.06	0.08	7.74
medical	0.14	0.00	0.16	7.63
moving	0.15	-0.01	0.18	7.84
other	0.15	0.01	0.16	7.92
renewable_energy	0.14	-0.05	0.23	7.93
small_business	0.14	-0.04	0.22	9.96
vacation	0.15	0.01	0.16	7.54
wedding	0.14	0.05	0.09	8.47
Total	0.13	0.04	0.10	7.96

8.1.2 Matured Sub-Sample

Table 8.7: Correlation Table

	Chargedoff	Funded Amount	Annual Income	Debt-to-Income	FICO Score
Chargedoff	1				
Funded Amount	0.0491***	1			
Annual Income	-0.0655***	0.294***	1		
Debt-to-Income	0.0944***	0.0794***	-0.147***	1	
FICO Score	-0.150***	0.0435***	0.0659***	-0.237***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8.8: Comparison Fully Paid vs. Charged Off

	Total					**Chargedoff**		**Fully Paid**		**T-test**	**WMW-test**
Variable	Count	min	mean	max	sd	Count	mean	Count	mean	t-value	z-value
Funded Amount	41238	1000.00	12149.34	35000.00	7767.16	8102	12801.88	33136	11989.79	8.44***	7.32***
Interest Rate	41238	0.05	0.13	0.26	0.04	8102	0.15	33136	0.13	42.85***	42.40***
Annual Income	41238	4080.00	70172.14	2000000.00	50808.19	8102	61715.21	33136	72239.93	-16.77***	-22.95***
Income Verified	41238	0.00	0.37	1.00	0.48	8102	0.42	33136	0.36	10.03***	10.02***
Fractional Funding	41238	0.00	0.93	1.00	0.25	8102	0.95	33136	0.93	9.28***	9.27***
36 months term	41238	0.00	0.79	1.00	0.41	8102	0.66	33136	0.82	-32.44***	-32.04***

8.1.3 Chargedoff Sub-Sample

Table 8.9: LGD Table

	LGD (pos.)	Received Interest	Received Late Fee	Received Principal	Recoveries	Collection Fee
A	7502.26	656.83	0.14	4598.04	4.33	0.38
B	7943.25	1006.09	0.18	3844.43	5.11	0.29
C	9514.34	1298.43	0.24	3332.81	7.59	0.65
D	9166.73	1623.11	0.35	3486.97	11.93	0.75
E	11299.19	2716.21	0.45	4138.64	20.13	1.66
F	12424.51	3297.52	0.56	4074.82	30.71	2.48
G	13733.34	3924.83	1.15	4887.35	48.81	3.71
Total	8848.90	1314.97	0.25	3817.47	8.66	0.65

Table 8.10: Survivaltime in months - 36 months term loans

	count	min	p25	mean	p75	max
A	757	1	7	12.20343	16	37
B	1657	1	6	11.04406	14	42
C	1446	1	5	10.19018	13	44
D	1016	1	5	10.0374	14	40
E	229	1	5	10.37118	14	38
F	56	1	3.5	10.32143	17.5	31
G	15	1	7	9.8	13	22
Total	5176	1	5	10.73628	14	44

Table 8.11: Survivaltime in months - 60 months term loans

	count	min	p25	mean	p75	max
A	22	3	8	15.27273	22	36
B	393	1	7	14.35878	21	37
C	537	1	6	12.4581	18	36
D	619	1	6	12.01454	16	39
E	673	1	7	12.26003	17	41
F	395	1	5	10.93165	15	38
G	110	1	6	11.54545	16	30
Total	2749	1	6	12.34813	17	41

8.1.4 Fully Paid Sub-Sample

Table 8.12: Expected Returns by Credit Grade

	Interest Rate	expect. Return 1	Pr(default)	Fundingtime [in days]
A	0.07	0.04	0.05	8.58
B	0.11	0.05	0.08	8.07
C	0.15	0.05	0.12	7.49
D	0.17	0.04	0.15	7.76
E	0.19	0.04	0.19	7.95
F	0.21	0.00	0.25	8.14
G	0.22	-0.15	0.41	10.79
Total	0.12	0.05	0.09	8.06

Table 8.13: Expected Returns by Purpose

	Interest Rate	expect. Return 1	Pr(default)	Fundingtime [in days]
car	0.10	0.06	0.06	7.09
credit_card	0.12	0.06	0.07	8.15
debt_consolidation	0.12	0.05	0.09	8.21
educational	0.11	0.06	0.13	7.39
home_improvement	0.11	0.05	0.08	7.99
house	0.11	0.06	0.07	9.19
major_purchase	0.10	0.06	0.06	7.25
medical	0.11	0.02	0.12	7.80
moving	0.12	0.01	0.14	7.26
other	0.12	0.03	0.12	7.58
renewable_energy	0.11	-0.03	0.20	6.88
small_business	0.12	-0.02	0.18	9.53
vacation	0.12	0.03	0.12	6.79
wedding	0.12	0.05	0.07	7.96
Total	0.12	0.05	0.09	8.06

9 Appendix II - Main Results

Table 9.1: Sources of Risk

	Probit P(Default)		ZTNB on SurvTime		OLS on LGD		Probit P(Prepay)	
	b	se	b	se	b	se	b	se
main								
Funded Amount	-0.0007*	(0.00)	0.0041***	(0.00)	0.0002	(0.00)	-0.0014***	(0.00)
Annual Income	-0.0004***	(0.00)	-0.0002	(0.00)	-0.0001**	(0.00)	0.0001***	(0.00)
Debt-to-Income	0.0753**	(0.03)	-0.3242***	(0.09)	0.0965***	(0.01)	-0.1766***	(0.01)
FICO Score	-0.0002*	(0.00)	-0.0012***	(0.00)	0.0012***	(0.00)	0.0005***	(0.00)
o.Funding in fractions	0.0000	(.)	0.2434***	(0.04)	-0.0046	(0.01)	0.0016	(0.00)
Income not verified	0.0017	(0.00)	-0.0137	(0.01)	-0.0086***	(0.00)	0.0062***	(0.00)
Lenght of the title	-0.0039	(0.00)	-0.0009	(0.02)	-0.0013	(0.00)	0.0064***	(0.00)
Lenght of the Description	-0.0046	(0.00)	0.0093	(0.01)	0.0039*	(0.00)	0.0059***	(0.00)
Purpose Dummies:								
purpose==car	-0.0182	(0.01)	-0.0052	(0.05)	-0.0031	(0.01)	-0.0032	(0.00)
purpose==credit_card	-0.0307***	(0.01)	0.0237	(0.02)	0.0006	(0.00)	-0.0111***	(0.00)
purpose==educational	0.0330	(0.03)	-0.0631	(0.15)	-0.0501**	(0.03)	-0.0191	(0.02)
purpose==home_improvement	0.0055	(0.01)	-0.0642**	(0.03)	-0.0014	(0.00)	-0.0100***	(0.00)
purpose==house	-0.0065	(0.02)	0.0782	(0.06)	-0.0116	(0.01)	0.0348***	(0.01)
purpose==major_purchase	-0.0126	(0.01)	-0.0320	(0.04)	-0.0019	(0.01)	-0.0005	(0.00)
purpose==medical	0.0293**	(0.01)	-0.0876*	(0.05)	-0.0088	(0.01)	-0.0073	(0.01)
purpose==moving	0.0343***	(0.01)	-0.0647	(0.05)	-0.0107	(0.01)	-0.0034	(0.01)
purpose==other	0.0171***	(0.01)	-0.0731***	(0.02)	-0.0098***	(0.00)	-0.0090***	(0.00)
purpose==renewable_energy	0.0741***	(0.03)	-0.0880	(0.11)	-0.0253	(0.02)	-0.0419**	(0.02)
purpose==small_business	0.0770***	(0.01)	-0.0960***	(0.03)	-0.0095**	(0.00)	-0.0316***	(0.00)
purpose==vacation	0.0171	(0.02)	-0.0565	(0.07)	0.0078	(0.01)	0.0059	(0.01)
purpose==wedding	-0.0204	(0.01)	-0.0789	(0.05)	-0.0005	(0.01)	0.0041	(0.01)
Home Ownership Dummies:								
home_ownership==MORTGAGE	-0.9340	(22.23)	0.2235	(0.32)	0.0261	(0.05)	0.0731*	(0.04)
o.home_ownership==NONE	0.0000	(.)	0.0733	(0.52)	0.0074	(0.08)	0.0403	(0.06)
home_ownership==OWN	-0.9349	(22.23)	0.1881	(0.32)	0.0129	(0.05)	0.0686	(0.04)
home_ownership==RENT	-0.9318	(22.23)	0.1922	(0.32)	0.0192	(0.05)	0.0655	(0.04)
Employment Dummies:								
Longterm Employment	0.0085	(0.01)	0.0032	(0.02)	0.0028	(0.00)	-0.0079***	(0.00)
Shortterm Employment	0.0000	(0.00)	-0.0344**	(0.02)	-0.0027	(0.00)	0.0023	(0.00)
Unemployed	0.0630***	(0.01)	-0.0232	(0.03)	-0.0003	(0.00)	-0.0370***	(0.00)
o.esub_grade==G3	0.0000	(.)	0.0496	(0.20)	-0.0061	(0.03)	0.0073	(0.03)
o.esub_grade==G4	0.0000	(.)	0.1079	(0.23)	-0.0285	(0.04)	0.0011	(0.03)
2010b.year	0.0000	(.)	0.0000	(.)	0.0000	(.)	0.0000	(.)
2011.year	-0.0010	(0.00)	-0.2432***	(0.02)	0.0107***	(0.00)	-0.0894***	(0.01)
2012.year			-0.7084***	(0.02)	0.0092***	(0.00)	-0.2241***	(0.01)
2013.year			-1.5059***	(0.03)	0.0408***	(0.00)	-0.3497***	(0.00)
survivaltime_m					-0.0260***	(0.00)		
2014.year							-0.3990***	(0.00)
Constant			3.1127***	(0.43)	0.1520**	(0.07)		
lnalpha								
Constant			-1.5431***	(0.02)				
Sub grades:	Yes		Yes		Yes		Yes	
N	2.31e+04		7925.0000		7925.0000		2.46e+05	
r2_p	0.0633		0.0641				0.1802	
r2					0.8334			
chi2			3340.4858					
p	0.0000		0.0000		0.0000		0.0000	

Table 9.1 aggregates the results from our models studying sources of risk. The first column presents marginal effects from a robust probit model with default as the dependent variable, estimated for a sample of 36months term loans, originated in 2010 or 2011. The second column

shows the coefficients from an zero-truncated negative binomial model with survival time in months as the dependent variable estimated for the charged-off sample. The third columns presents the coefficients of an OLS model with LGD als the dependent variable, also estimated for the charged-off sample. The last column contains marginal effects from a probit model with prepay as the depended variable. All variables are explained in table 8.1.

Table 9.2: Wisdom of the Crowd

	OLS on Fundingtime		OLS on expect. Return	
	b	se	b	se
Funded Amount	0.0660***	(0.00)	0.0008***	(0.00)
Annual Income	-0.0016***	(0.00)	0.0001***	(0.00)
Debt-to-Income	-0.1465	(0.15)	-0.1665***	(0.00)
FICO Score	0.0037***	(0.00)	0.0002***	(0.00)
Funding in fractions	0.1472***	(0.03)	0.0301***	(0.00)
Income not verified	-1.3995***	(0.02)	-0.0013***	(0.00)
Lenght of the title	-0.0460*	(0.03)	0.0072***	(0.00)
Lenght of the Description	0.1792***	(0.02)	0.0074***	(0.00)
Purpose Dummies:				
purpose==car	-0.0122	(0.09)	0.0161***	(0.00)
purpose==credit_card	0.1164***	(0.03)	0.0234***	(0.00)
purpose==educational	0.3685	(0.49)	0.0111***	(0.00)
purpose==home_improvement	0.3512***	(0.05)	-0.0082***	(0.00)
purpose==house	0.5524***	(0.14)	0.0043***	(0.00)
purpose==major_purchase	0.0586	(0.07)	0.0130***	(0.00)
purpose==medical	0.1287	(0.10)	-0.0354***	(0.00)
purpose==moving	0.2363*	(0.12)	-0.0424***	(0.00)
purpose==other	0.3637***	(0.05)	-0.0205***	(0.00)
purpose==renewable_energy	0.2340	(0.33)	-0.0822***	(0.00)
purpose==small_business	1.4681***	(0.08)	-0.0954***	(0.00)
purpose==vacation	0.1698	(0.13)	-0.0188***	(0.00)
purpose==wedding	0.4055***	(0.11)	0.0065***	(0.00)
Home Ownership Dummies:				
home_ownership==MORTGAGE	0.5759	(0.69)	0.5199***	(0.00)
home_ownership==NONE	-0.0897	(1.03)	-0.3822***	(0.00)
home_ownership==OWN	0.9859	(0.69)	0.5145***	(0.00)
home_ownership==RENT	0.7721	(0.69)	0.5130***	(0.00)
Employment Dummies:				
Longterm Employment	0.1421***	(0.03)	-0.0068***	(0.00)
Shortterm Employment	0.1058***	(0.03)	-0.0011***	(0.00)
Unemployed	1.2605***	(0.05)	-0.0823***	(0.00)
2010b.year	0.0000	(.)	0.0000	(.)
2011.year	-0.6264***	(0.06)	0.0070***	(0.00)
2012.year	1.4023***	(0.05)	0.0145***	(0.00)
2013.year	0.0364	(0.05)	0.0128***	(0.00)
2014.year	-3.3502***	(0.06)	0.0105***	(0.00)
fundingtime			-0.0001***	(0.00)
Constant	5.0193***	(1.69)	-1.4883***	(0.01)
Sub grades:	Yes		Yes	
N	1.84e+05		1.84e+05	
r2_p				
r2	0.1126		0.8272	
chi2				
p	0.0000		0.0000	

Table 9.2 presents the results from our analysis on the smartness of the crowd. the first column present the coefficients from an OLS with fundingtime as the dependent variable. Fundingtime is defined as the time in days from the listing until the funding of the loan. The second column presents the results from an OLS model with our expected return measure as the dependent variables. All variables are explained in table 8.1.

Bibliography

Agrawal, A., C. Catalini, and A. Goldfarb (2010). The Geography of Crowdfunding. *Available at SSRN: http://ssrn.com/abstract=1692661*.

Agrawal, A., C. Catalini, and A. Goldfarb (2013). The Simple Economics of Crowdfunding. In *Available at http://www.nber.org/papers/w19133*, Volume 14. National Bureau of Economic Research.

Arazy, O., W. Morgan, and R. Patterson (2006, December). Wisdom of the Crowds: Decentralized Knowledge Construction in Wikipedia. *SSRN Electronic Journal*.

Askira Gelman, I. (2013). Show Us Your Pay Stub: Income Verification in P2P Lending. *SSRN Electronic Journal*.

Barakova, I., D. Glennon, and A. A. Palvia (2011, April). Adjusting for Sample Selection Bias in Acquisition Credit Scoring Models. *SSRN Electronic Journal*.

Bayus, B. L. and V. Kuppuswamy (2013). Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter. *Available at SSRN: http://ssrn.com/abstract=2234765*.

Brabham, D. C. (2008, February). Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Convergence: The International Journal of Research into New Media Technologies 14*(1), 75–90.

Bradford, S. (2011). Crowdfunding and the federal securities.

Caldieraro, F., M. C. Jr, J. D. Shulman, and J. Zhang (2011). Is Silence Golden? How Non-Verifiable Information Influences Funding Outcomes On Peer-to-Peer Lending Platforms. *Available at http://faculty.washington.edu/ jshulman/documents/Caldieraro%20Cunha%20Shulman%20Zhang%202011.pdf* (April).

Ceyhan, S., X. Shi, and J. Leskovec (2011). Dynamics of bidding in a P2P lending service. In *Proceedings of the 20th international conference on World wide web - WWW '11*, New York, New York, USA, pp. 547. ACM Press.

Chen, N., N. S. Lambert, and A. Gosh (2011). Auctions for Social Lending: A Theoretical Analysis. *Research Paper No.2078, Stanford Graduate School*.

Ciampi, F., C. Vallini, N. Gordini, and M. Benvenuti (2009). Are Credit Scoring Models Able to Predict Small Enterprise Default? Statistical Evidence from Italian Small Enterprises. *Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2214957*.

Collier, B., R. Hampshire, H. J. Heinz, and I. I. I. College (2010). Sending Mixed Signals: Multilevel Reputation Effects in Peer-to-Peer Lending Markets. *CSCW*, 197–206.

Comment, R. and G. Jarrell (1991). The Relative Signalling Power of Dutch-Auction and Fixed-Price Self-Tender Offers and Open-Market Share Repurchases. *The Journal of Finance 46*(4), 1243–1271.

Cumming, D. (2013). Demand Driven Securities Regulation: Evidence from Crowdfunding. *Available at SSRN: http://ssrn.com/abstract=2255229*.

Degryse, H. and S. Ongena (2005, February). Distance, Lending Relationships, and Competition. *The Journal of Finance 60*(1), 231–266.

Everett, C. R. (2010). Group membership , relationship banking and loan default risk: the case of online social lending. *Available at SSRN: http://ssrn.com/abstract=1114428* (765).

Freedman, S. and G. Z. Jin (2008). Do Social Networks Solve Information Problems for Peer-to-Peer Lending ? Evidence from Prosper.com. *Available at SSRN: http://ssrn.com/abstract=1304138*.

Freedman, S. and G. Z. Jin (2010). Learning by Doing with Asymmetric Information: Evidence from Prosper.com. *Available from http://www.indiana.edu/~spea/faculty/pdf/freedman_jin_1210.pdf*.

Garman, S., R. Hampshire, and R. Krishnan (2008). A Search Theoretic Model of Person-to-Person Lending. *Working Paper Available at http://www.heinz.cmu.edu/research/244full.pdf*, 1–32.

Glennon, D., N. M. Kiefer, C. E. Larson, and H.-s. Choi. Development and Validation of Credit Scoring Models. *Journal of Credit Risk, Forthcoming*.

Haltiwanger, J., J. Ham, R. Hampshire, A. Korinek, P. Leslie, R. Cooper, H. Cai, J. Brickley, E. Cantillon, and S. Borenstein (2014). The Signaling value of online social networks. *NBER Working paper. Available at: http://www.nber.org/papers/w19820*.

Herrero-Lopez, S. (2009). Social interactions in P2P lending. *Proceedings of the 3rd Workshop on Social Network Mining and Analysis 09*, 1–8.

Herzenstein, M. and R. L. Andrews (2008). The Democratization Of Personal Consumer Loans? Determinants Of Success In Online Peer-To-Peer Lending Communities. *Working Paper Available at http://www.prosper.com/downloads/research/democratization-consumer-loans.pdf* (302), 1–45.

Herzenstein, M., U. M. Dholakia, and R. L. Andrews (2011, February). Strategic Herding Behavior in Peer-to-Peer Loan Auctions. *Journal of Interactive Marketing 25*(1), 27–36.

Hildebrand, T., M. Puri, and J. Rocholl (2011). Skin in the Game : Incentives in Crowdfunding. *Working Paper Available at http://www.rhsmith.umd.edu/feaconference/docs/Session3PuriSkinintheGame.pdf* (October).

Hornuf, L. and L. Klöhn (2012). Crowdinvesting in Deutschland. *Journal of Banking Law and Banking 4*, 237–320.

Iyer, R., A. Ijaz, K. Erzo, and F. P. L. Kelly (2009). Screening in New Credit Markets Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending? *Available at SSRN: http://ssrn.com/abstract=1570115* (August).

Iyer, R., A. Ijaz, K. Erzo, and F. P. L. Kelly (2011). Inferring Asset Quality : Determining Borrower Creditworthiness in Peer-to-Peer Lending Markets. *Available at http://users.nber.org/~luttmer/peerlending*.

Kiefer, N. M. and C. E. Larson (2006, October). Specification and Informational Issues in Credit Scoring. *SSRN Electronic Journal*.

Kim, K. and S. Viswanathan (2013a). The Experts in the Crowd : The Role of Reputable Investors in a Crowdfunding Market. *Available at SSRN: http://ssrn.com/abstract=225824*.

Kim, K. and S. Viswanathan (2013b). The Signals in the Noise: The Role of Reputable Investors in a Crowdfunding Market. *Available at SSRN: http://ssrn.com/abstract=2258243*.

Klein, T. (2010). Microfinance 2 . 0 â Group Formation & Repayment Performance in Online Lending Platforms. *Available at: http://campus-for- finance.com/filebrowser/files/Papers/49a40613015_name.pdf*.

Krishna, K., L. H. Tan, and A. Fishman (2009). Financial intermediaries as facilitators of information exchange between lenders and reputation formation by borrowers. *International Review of Economics & Finance 18*(2), 301–305.

Lambert, T. and A. Schwienbacher (2010). An Empirical Analysis of Crowdfunding. *Available at http://www.crosnerlegal.com/images/47770544_An_Empirical_ Analysis_of_ Crowdfunding.pdf*.

Lee, E. and B. Lee (2012, September). Herding behavior in online P2P lending: An empirical investigation. *Electronic Commerce Research and Applications 11*(5), 495–503.

Lin, M., N. R. Prabhala, and S. Viswanathan (2011). Judging Borrowers By The Company They Keep : Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Available at SSRN: http://ssrn.com/abstract=1355679*.

Malone, T. W., R. Laubacher, and C. N. Dellarocas (2009, February). Harnessing Crowds: Mapping the Genome of Collective Intelligence. *SSRN Electronic Journal*.

Manson (2007). *Handbook of Research on Venture Capital (Google eBook)*. Edward Elgar Publishing.

Martin, T. A. (2012, April). The Jobs Act of 2012: Balancing Fundamental Securities Law Principals with the Demands of the Crowd. *Available at SSRN http://papers.ssrn.com/abstract=2040953*.

Mollick, E. (2013a). Swept Away by the Crowd? Crowdfunding, Venture Capital, and the Selection of Entrepreneurs. *Available at SSRN: http://ssrn.com/abstract=2239204*.

Mollick, E. (2013b). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*.

Pope, D. G. and J. R. Sydnor (2008). What is in a Picture? Evidence of Discrimination from Prosper.com. *Available at SSRN: http://ssrn.com/abstract=1220902*.

Prpic, J. and P. Shukla (2012). The Theory of Crowd Capital.

Puro, L., J. E. Teich, H. Wallenius, and J. Wallenius (2010, April). Borrower Decision Aid for people-to-people lending. *Decision Support Systems 49*(1), 52–60.

Puro, L., J. E. Teich, H. Wallenius, and J. Wallenius (2011). Bidding strategies for real-life small loan auctions. *Decision Support Systems 51*(1), 31–41.

Rainer, B. and P. Stefanie (2010). Privacy in Online Social Lending. *AAAI 2010 Spring Symposium on Intelligent Privacy Management Available from http://www.aaai.org/ocs/index.php/SSS/SSS10/paper/viewPDFInterstitial/1048/1472*, 23–28.

Ravina, E. (2012). Love & Loans : The Effect of Beauty and Personal Characteristics in Credit Markets. *Available at SSRN: http://ssrn.com/abstract=1101647*.

Rodriguez, G. (2013). Models for Count Data With Overdispersion. *Working Paper Available at data.princeton.edu/wws509/notes/c4a.pdf*.

Siegel, J. and L. A. Young (2010). Trust and Credit. *FA 2010 Atlanta Meetings Paper. Available at SSRN: http://ssrn.com/abstract=1343275*.

Surowiecki, J. (2005). *The Wisdom of Crowds*. Knopf Doubleday Publishing Group.

Yi, S. K. M., M. Steyvers, M. D. Lee, and M. J. Dry (2012, April). The wisdom of the crowd in combinatorial problems. *Cognitive science 36*(3), 452–70.

Yum, H., B. Lee, and M. Chae (2012, September). From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Electronic Commerce Research and Applications 11*(5), 469–483.

Zhang, J. and P. Liu (2012). Rational Herding in Microloan Markets. *Management Science 58*(5), 892–912.

Zhu, F. and M. Iansiti (2009). Entry into Platform-Based Markets. *Available at SSRN: http://ssrn.com/abstract=1350233*.