2017

The Impact of Social Disclosures Within Fixed Rate Peer-to-Peer Lending Markets

Robert A. Jordan
Nova Southeastern University, robert.anthony.jordan@gmail.com

THE IMPACT OF SOCIAL DISCLOSURES WITHIN FIXED-RATE PEER-TO-PEER LENDING MARKETS

By
Robert A. Jordan

A DISSERTATION

Submitted to
H. Wayne Huizenga School of Business and Entrepreneurship Nova Southeastern University

in partial fulfillment of the requirements for the degree of

DOCTOR OF BUSINESS ADMINISTRATION

2017

A Dissertation
Entitled

THE IMPACT OF SOCIAL DISCLOSURES WITHIN FIXED-RATE PEER-TO-PEER LENDING MARKETS

By

ROBERT A. JORDAN

We hereby certify that this Dissertation submitted by Robert A. Jordan conforms to acceptable standards, and as such is fully adequate in scope and quality. It is therefore approved as the fulfillment of the Dissertation requirements for the Degree of Doctor of Business Administration.

Approved:

	11-13-17
Hyungkee Young Baek, Ph.D. Chairperson	Date
	11/13/2017
David D. Cho, Ph.D. Committee Member	Date
	11/15/2017
Darshana Palkar, Ph.D. Committee Member	Date
	11/15/17
Suri Weisfeld-Spolter, Ph.D. Chair, Doctoral Program	Date
	27 NOV 2017
J. Preston Jones, D.B.A. Dean, H. Wayne Huizenga College of Business and Entrepreneurship	Date

Nova Southeastern University
2017

CERTIFICATION STATEMENT

I hereby certify that this paper constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

Signed____________________

Robert A. Jordan

CERTIFICATION STATEMENT

I hereby certify that this paper constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

Signed_____________________

Robert A. Jordan

ABSTRACT

THE IMPACT OF SOCIAL DISCLOSURES WITHIN FIXED-RATE PEER-TO-PEER LENDING MARKETS

By

Robert A. Jordan

Financial journals have just begun to examine the implications of unsecured fixed-rate loans between lenders and borrowers administered over the internet. This study observes 31,550 loans issued between June 2007 and April 2013 with a 36-month term, that are fully paid or charged off, based on a data set from the largest P2P lending website. Initial findings within peer-to-peer (P2P) lending markets have identified that social disclosures may influence these markets. The result of this analysis unambiguously confirms social disclosures influence lenders and the factors significant for funding a loan are inconsistent with the factors significant to repayment of the loan. Prescriptive filters based on social disclosures can improve the likelihood of selecting a creditworthy borrower and increase the models explanatory power. The study finds that distinct forms of social disclosure and specific content within social disclosures predict the amount of funding received and probability of loan repayment.

ACKNOWLEDGEMENTS

I must first thank God and prayer, for blessing me with the strength and ability to complete the dissertation. Second, I must thank my wife, Britney Jordan, for her constant support, encouragement, and sacrifice required for me to obtain my educational goals. I would also like to thank my mother and father who have helped as editors and research assistants, but most importantly as cheerleaders throughout the doctoral process.

The contributions of Dr. Baek and Dr. Cho cannot be overstated, their input cannot be measured, and their guidance throughout the years has been transformative on multiple levels. It is a blessing to have been able to work with individuals both high in character and intelligence.

Finally, Octo Consulting Group and the Huizenga School of Business and Entrepreneurship, are both greatly appreciated for affording me the ability to pursue a doctoral education and professional career in parallel.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Introduction

Importance of the Problem

Peer-to-Peer (P2P) lending platforms democratize finance by enabling a more efficient flow of funds between capital-seekers and capital-providers. Historically, the financial systems of an economy consists of three main components: 1) financial markets, 2) financial institutions, and 3) financial regulators (Merton, 1974; Schwienbacher, 2010). This architecture has limited capital-providers in debt markets to only the wealthy and those capable of meeting the regulatory requirements imposed on financial institutions. Technology has disrupted this framework and opened access to previously closed financial markets in ways that were not possible merely ten years ago. The potential benefits to society through this financial revolution are limitless, however due to the infancy of these markets thoughtful research is required to ensure the perceived benefits are fully realized. In short, P2P lending has the potential to become a financial innovation that is parallel to the Savings Bank Movement of 1810[1], or the Micro Finance Movement[2].

The benefits of crowdfunding are epitomized within the financial innovations of P2P lending. P2P lending was first introduced by the United Kingdom based Zopa Company ("Zopa") established in 2005. Since the launch of Zopa, an estimated 52 active

---

[1] [1] The impact of the Savings Bank Movement of 1810 enabled the economically disenfranchised of Dumfries, Ireland to gain financial independence through bank accounts earning 4% interest

[2] The Micro Finance movement started by the Grameen Bank in the 1970's, provided underserved populations in Bangladesh access to capital for entrepreneurial endeavors.

P2P lending platforms can be found online[3]. These platforms act as the intermediary performing the matching function enabling capital-seekers and capital-providers to efficiently exchange information about security prices and offerings in order to overcome information asymmetries and minimize transaction costs (Bakos, 1991). The more a P2P lending platform is able to acquire and match borrowers to lenders the bigger the networking effect and overall success of the market (Caillaud & Jullien, 2003; Damiano & Li, 2008). From a market theory perspective, the economic rationale for the existence of financial institutions and instruments is related to transaction costs; thus, the surviving institutions and instruments are those that have the lowest transaction costs relative to potential benefits (Schwienbacher, 2010). The transaction costs are a key value proposition of P2P lending. Online lending platforms benefit from a lower cost of capital and transaction cost by having underwriting handled by a pool of lenders as opposed to a single bank. The borrowers benefit from lower interest rates for unsecured debt, and investors have the opportunity to earn above prime interest rates on their investments. Not to mention, P2P lending platforms create a simple non-threating online user experience for both amateur and expert investors alike. The loans appeal to borrowers seeking to consolidate or pay off credit card debt, repay high interest rate loans, or borrow funds for other general purposes. P2P lending requires only weeks for borrowers to receive capital versus a longer more iterative process required in retail banking. Both P2P lending and traditional banks qualify borrowers in similar manners, however, in contrast to the traditional banking processes for unsecured loans, the P2P lending application process is completely online lowering overhead cost. The human interaction

[3] http://www.p2p-banking.com/countries/germany-international-p2p-lending-statistics-february-2017/#more-5573

required in brick and mortar traditional banking has been exchanged for social disclosures. Social disclosures are defined as voluntary information provided by the borrower within the loan descriptions or facilitated through online borrower and lender interactions on the P2P platform. Social disclosures are a financial innovation and a paradigm shift that is ushering in a new area of debt market research.

Leading financial journals suggests that a close relationship between banks and borrowers reduces information asymmetries and improves borrowers' access to credit which leads to an overall improvement in their performance (Castelli, Dwyer, & Hasan, 2006). This notion is supported by Diamond's (1991) demonstration that a successful bank relationship lowers the equilibrium probability of default. Further, Rajan (1992) finds informational asymmetries are reduced for small businesses based on the length of time of their relationship with their lender and the number of creditors the firm uses. This concept is also consistent with Von Thadden's (1995) view that the efficiency of an investment is improved by a debt contract with periodic monitoring. This line of argument suggests that a closer bank relationship will be associated with better firm performance, and that a borrower's optimal strategy is to establish a long-term relationship and to borrow from one, or a limited number of banks. In stark contrast, the P2P environment is a market where borrowers are relatively numerous and participate in the market sufficiently infrequently to not acquire relationships with the lenders. The new relationship forged between borrowers and lenders is based solely on their online interactions through social disclosures which forms the basis for the research problem.

Foundational Theory

P2P lending is consistent with the fundamental principles of Two-sided Market Theory, Financial Intermediation Theory, and Agency Theory. Two-sided markets are characterized by multiple sets of groups interacting through an intermediary (Allen & Santomero, 1997; Caillaud & Jullien, 2003; Damiano & Li, 2008). Value is created in two-sided markets by enabling direct interactions between two distinct groups. The interactions create positive and negative same-side network effects. Members of one group exhibit a preference regarding the number of users in the other group. Borrowers and Lenders represent the two groups in the P2P lending market. Specifically, the Lending Club platform is structured as a per transaction charge two-sided market, where investors pay Lending Club a service fee equal to one percent (1%) of the amount of any borrower payments received within 15 days of the payment due date. Positive same-side benefits are produced by increasing the number for lenders enabling more loans to be fully funded. Additionally, borrowers exhibit a preference for more lenders in order to increase the likelihood of having loans fully funded.

Financial Intermediation Theory details the exchange relationships and functionalities of capital-providers providing funds to an intermediary institution, and the financial institution providing funds to capital-seekers (Allen & Santomero, 1997). Intermediaries overcome asymmetric information problems by acting as delegated monitors that perform the function of converting risky investments into lower risk investments through diversification and matching small deposits with large loans and large deposits with small loans. Accordingly, the purpose of financial intermediaries is to resolve market imperfections. In a perfect market, savers and investors have perfect information to identify each other directly and the financial intermediary's role becomes less important. P2P lending platforms further disintermediation by providing direct borrower contact to a pool of lenders. Lending Club still performs delegated monitoring, diversification, and matching functions, but bypasses traditional banking by facilitating the channeling of funds directly between lenders and borrowers.

Agency Theory is also present in P2P lending from information asymmetries when the risk-taking party knows more about its intentions than the party paying the consequences (Akerlof, 1970). In other words, the motivation to act in a self-interested manner results in moral hazards where the actions of one party change to the detriment of another after the financial transaction. Consistent with Jensen and Meckling (1976), agency costs are incurred via monitoring expenditures by the principal, the bonding expenditures by the agent, and the residual loss. In P2P lending, borrowers are motivated to act in their own best interests (Jensen, 1976). The true intent of the borrower's loan purpose is unknown to lenders at the time of investment and creates a moral hazard for borrowers and adverse selection problem within the Lending Club P2P environment. As a result, when borrowers miss payments and loans become late Lending Club charges monitoring fees and residuals losses occur in the form of collections and default. Borrower default is a consequence of moral hazard and a key component of the research problem.

The Research Problem

In contrast to previous literature predicated on face-to-face borrower and lender relationships, advances in technology enable P2P platforms to conduct social intermediation from geographically dispersed locations. Despite these innovations, the benefits of this new market are not completely understood due to new forms of information asymmetries.

It is currently unclear whether social disclosures are inadvertently or advertently being used by borrowers with poorer credit ratings to manipulate lenders (Berkovich, 2011; Herzenstein, Sonenshein, & Dholakia, 2011; Iyer, Khwaja, Luttmer, & Shue, 2009). The national debt-to-income ratio for consumer is 11.66% compared to 13.81% for Lending Club borrowers (Emekter, Tu, Jirasakuldech, & Lu, 2015). Based on this figure alone, it is reasonable to assume poorer quality borrowers, “lemons”, are attracted to P2P lending and will have an incentive to produce flattering social disclosures. Conversely, by ignoring social disclosures on P2P platforms the positive benefits to lenders and potential financial innovations could be lost. This research analyzes whether P2P lenders can reduce information asymmetries through social disclosure information provided by borrowers on the leading fixed-rate P2P lending platform. Ambiguity exists in understanding the determinants of creditworthiness from social disclosures and the subsequent impact on fixed-rate peer-to-peer lending platforms. This study is designed to address this problem by testing the social disclosures lenders are able to observe and their impact on funding time, investment, loan repayment and default.

Contributions of the Study

The literature has yet to fully understand and test the effects of information produced by social disclosures within P2P markets. It would be intellectually dishonest to overlook the adverse selection, selection biases, and moral hazards inherent in the P2P market where the lender has less knowledge of the borrowers' creditworthiness than the borrower and anonymity and geography between lenders and borrowers is high (Akerlof, 1970). In the absence of relationship banking described in the seminal work of Rajan (1992), the significance of social disclosures as a substitute for human interaction is of increased importance. Moreover, both good and bad borrowers are cognizant of the asymmetric information risk and may have a competitive incentive to obfuscate creditworthiness to lenders through social disclosures. Financial economists have a vested interest and responsibility for understanding these dynamics of P2P lending especially within an unsecured debt market where no collateral is backing the borrower's loan. The ability for lenders to separate good and bad borrowers might be possible through analysis of the social disclosures provided on the lending platform. This study combines multiple social disclosure analysis approaches found in the current stream of literature to answer this question.

This is the first study of its kind that observes social disclosures within a fixed-rate P2P platform. The majority of previous studies focus on reverse auction lending formats. Literature on reverse auction P2P lending platforms is fundamentally different due to negative same-side market effects from lender competition, whereas additional lenders on P2P fixed-rate platforms are a positive same-side market effect. Moreover, the reverse auction format for P2P lending has been abandoned by the leading P2P lenders in favor of fixed-rate models and the change has created a void in the current literature. Furthermore, the data and variables used in this analysis is no longer provided by the Lending Club platform in any format for the sample period of June 2007 through April 2013. This study also incorporates the methodologies of related social disclosure peer-reviewed articles from Herzenstein et al. (2011), Lewis (2011), and Michels (2012). The culmination of multiple approaches into one study enables this research to observe the interaction between a wide-array of social disclosure explanatory variables found to be significant in previous studies. The results from this analysis are designed to fill the void by unambiguously identifying the social disclosures that influence lenders to invest in a loan and the social disclosures significant for borrower repayment of the loan. Thus, studying social disclosures on a fixed interest rate lending platform is a primary area of inquiry in this research and a principal contribution to the field of finance.

CHAPTER II

Review of Literature

Debt-based P2P Lending Literature Review

Overall, debt-based crowdfunding research related to social disclosures on fixed-rate P2P lending platforms is limited. To substantiate this claim a literature review beginning in chronological order was performed. The first paper reviewed was Bachmann et al. (2011) "Online Peer-to-Peer Lending – A Literature Review". In this article, Bachman discusses the main results of forty-three scientific articles related to peer-to-peer lending. Feller, Gleasure, and Treacy (2013) provide a seminar review based on the different forms of crowdfunding and include discussion feedback from the group of crowdfunding researchers in attendance (Feller, Gleasure, & Treacy, 2013b). Six studies provided a comprehensive overview of the crowdfunding literature focusing on capital-seeking, capital-providing, and the role of intermediary parties. Two P2P lending empirical studies, Mach , Carter, & Slattery (2014) and Emekter et al. (2015), have been produced related to pricing Notes and evaluating risk on fixed-rate P2P lending sites, but were limited to hard credit information on the Lending Club platform. In Table 1 a list of the major works and their contributions to the field are provided.

Table 1 Literature Review Key Findings

Article	Platform	Key Findings
Pope et al. (2011)	Prosper	Capital-providers have been shown to discriminate against capital-seekers based on profile photos, race, obesity, and appearance which challenges the value of social disclosures on Prosper.
Herzenstein et al. (2011)	Prosper	Finds that herding behavior, defined as a greater likelihood of bidding in auctions with more existing bids, on P2P loan auctions on Prosper.com. The results of an empirical study provide evidence of strategic herding behavior by lenders such that they have a greater likelihood of bidding on an auction with more bids (a 1% increase in the number of bids increases the likelihood of an additional bid by 15%), but only to the point at which it has received full funding. The study also finds a positive association between herding in the loan auction and its subsequent performance, that is, whether borrowers pay the money back on time.
Herzenstein et al. (2011)	Prosper	Provides evidence of higher default rates from each additional borrower identity claim related to trustworthiness, personal success, economic hardship, work ethic, morality, and religion within a social disclosure.
Lewis (2011)	eBay Motors	The study test whether bidder behavior is casually influenced by information on the auction web page. The study observes the information voluntarily disclosed along with hard characteristics such as model, year, mileage, transmission, and the accessories equipped on the vehicle.
Michels (2012)	Prosper	Reports disclosures for high-risk borrowers increase bidding activity by 18.21% and each disclosure provided reduces default probability by 5.37%.
Chen et al. (2012)	Prosper	Identifies a fundamental difference between a reverse auction model and fixed-rate model and that the auction model implies that the interest rate for a loan is a function of the

Article	Platform	Key Findings
		number of bids from interested capital-providers.
Lin et al. (2013)	Prosper	Performed an analysis on 4,139 social groups and verified friends using Prosper's reverse auction platform and found borrower that had verified friends, defined as a friend that accepted a friend request from a borrower with a validated account, signaled positively at the 1% significance level to lenders (Lin et al., 2013).
Chen et al. (2014)	Prosper	Investigated whether using an auction model in crowdfunding markets leads to an optimal result for market participants and found the reverse auction method was more complicated and less transparent than a fixed-rate model for capital providers.
Mollick (2014)	Kickstarter	Finds factors such as word count, misspellings, updates, comments, duration, number of investors, and number of Facebook Friends signal higher quality using a dataset of over 48,500 Kickstarter platform reward-based projects with combined funding over $237 million.
Mach et al. (2014)	Lending Club	Calculated loan performance for small business loans and proved business loans were 250 times more likely for default compared to other loan categories.
Emekter et al. (2015)	Lending Club	Uses Lending Club data to confirm that high credit grade, low debt-to-income ratio, high FICO score and low revolving line utilization are the most significant hard information factors associated with lower default risk.
Dorfleitner et al. (2016)	Auxmoney Smarva	Analyzes two P2P lending markets: one with social disclosures and one that primarily uses hard information. The study finds social disclosure factors are important for lenders when hard financial information is not available. The study also finds that social disclosure factors do not have much predictive power with respect to default probability.

A recurring factor within the literature is that social disclosures heavily impact lender decision-making. Social disclosures allow borrowers to voluntarily provide any information they believe is important for lenders to make an investment decision. Common examples reviewed in the literature include personal information, loan purpose details, explanations for the borrower's creditworthiness, and identity claims. Previous findings in the literature demonstrated that this form of information has a positive effect on establishing trust and influencing the likelihood of financing, lowering interest rates and decreasing the probability of loan default (Allison, Davis, Short, & Webb, 2015; Duarte, Siegel, & Young, 2012; Michels, 2012; Mollick, 2014; Pope & Sydnor, 2011). However, P2P lending studies demonstrate that the impacts of social disclosures on loan performance, in terms of default, are inconsistent.

For example, Dorfleitner et al. (2016) finds interest rate and hard facts are the main drivers of the default probability. The study also finds that social disclosure factors are important to loan funding, but social disclosures do not significantly predict default probability (Dorfleitner et al., 2016). This differs from Michels (2012) that reports explanatory disclosures increase bidding activity and each disclosure provided reduces default probability for higher risk borrowers. Conversely, Herzenstein et al. (2011) provides evidence of higher default rates when borrowers provide identity claims related to trustworthiness, personal success, economic hardship, work ethic, morality, and religion within a loan listing. While Lewis (2011) finds that if borrowers include qualifying words to describe an asset the likelihood of funding and default changes significantly. The findings from these works are not mutually exclusive. Further, they do not account for a social disclosure that expresses creditworthiness using credit explanations, identity claims, and qualifying phrases within the same disclosure. Thus, a comprehensive side-by-side analysis of each approach is necessary and performed in this study.

Text Analysis

On a surface level, text analysis on social disclosure have been shown to influence perceptions of quality across crowdfunding markets in general (Mollick, 2014; Pitschner & Pitschner-Finn, 2014). Mollick (2014) and Pitschner et al. (2014) provide a framework for crowdfunding research by focusing on measuring the probability of an entrepreneur reaching a desired funding goal by analyzing the total number of funding providers, Project goal, Funding level, Backers, Category, Updates, Comments, Duration, Word Count, Misspellings, and the total dollar amount provided on crowdfunding platforms (Meer, 2014; E. R. Mollick, 2012; Pitschner & Pitschner-Finn, 2014). A subset of these factors such as Word Count, Misspellings, Updates, and Comments are also applicable to social disclosures on P2P lending platforms. The forms of social disclosures proven to be significant within fundraising markets, has yet to be fully tested within P2P fixed-rate markets to my knowledge. On crowdfunding platforms such as Kickstarter, social disclosures are provided in the form of "Comments" and "Updates" in which investors can express enthusiasm or displeasure about the loan, product or project. Updates represent efforts by entrepreneurs to reach out to current and potential investors in order to reduce information asymmetries (Mollick, 2014). Comments represent questions that current and potential investors may have about the opportunity. This information is publicly available to all investors for decision making purposes (Mollick, 2014).

Specifically, Mollick (2014) descriptive statistics found the chance of success for projects with spelling errors is 13% less than those without errors and not providing timely Updates to investors reduces the chance of funding success by just over 13%. In addition to Mollick (2014), Pitschner et al. (2014) finds Word Count significant across all models in their study, however, the research only observed non-profit organizations. Interestingly, the aforementioned studies have elected to not analyze the overall quality of the social disclosures in terms of readability. This study includes such a measure through the Flesch Index reading score that calculates text readability and the grade level of a loan description narrative.

Identity Claims

Previous literature has proven that social disclosures play an important role in mitigating information asymmetries between borrowers and lenders (Berger & Udell, 1995; Petersen & Rajan, 1994). However, the first content analysis incorporated into this study, Herzenstein et al. (2011), finds identity claims written in borrower loan narratives can also adversely influence lender decisions. Identity claims are defined as personal character qualifying words used by the borrower in their loan descriptions. The six identity claims used in the study are categorized as trustworthy, economic hardship, hardworking, successful, moral, and religious. Each claim is coded as a dichotomous variable that receives a value of zero or one and the number of identity claims within each narrative was found to influence both loan funding and performance. Herzenstein et al. (2011) finds that unverifiable information affects lending decisions above and beyond the influence of objective verifiable information. The Herzenstein et al. (2011) article uses data from the peer-to-peer lending website Prosper.com and a data set is comprised of 2006 and 2007 loans which operated under the reverse auction model. Under this model lenders competed on loans by bidding down the interest rate until the loan auction ended. Herzenstein et al. (2011) conclusions suggest that identity claims can be used to mask the borrower's true creditworthiness. The number of identities that borrowers claim in their narratives were positively correlated with the probability of default, while the funded amount and number of identity claims were positively correlated. In other words, as the number of identity claims increased the funding increased, but default rate also increased. This finding suggests the existence of a moral hazard that encourages bad actors to increase the number of identity claims in order to spur investment that ultimately results in higher default rates for lenders. At a more granular level of identity claims, a trustworthiness identity claim that reads "I am very reliable and trustworthy and always repay my debts," was more likely to result in both funding and loan repayment (Herzenstein, Sonenshein, et al., 2011). These findings give reason to further study the specific content within social disclosures and suggest identity claims are significant for lenders investment decisions.

Keywords and Qualifiers

The word choice and word sequence has been found to play a role in online markets. The Lewis (2011) article test whether social disclosures provide sufficient detail to address information asymmetries in markets where anonymity and geography between buyers and sellers is high. In this market, buyers must solely depend upon the information provided on the car auction web page to evaluate quality and purchasing decisions. Lewis (2011) finds that use of negation, minimizing, and maximizing phrases causally influence investor decision making on auction listings. The study observes the information voluntarily disclosed along with hard characteristics such as model, year, mileage, transmission, and the accessories equipped on the vehicle. Special attention is placed on the coding of key text phases and qualifier phrases. Lewis (2011) performs an analysis of the keywords “rust”, “scratch”, and “dent” used on the eBay Motors auction site and develops a corpus or words in order to code dummy variables for “no x,” meaning any negation; “small x,” meaning any favorable qualifier; “big x” implying an unfavorable qualifier, and “x,” meaning the phrase is used without qualification. Using the keyword “rust” as an example, Lewis (2011) created 4 dummies for rust: (1) "No Rust" (2) "Small Rust" (3) "Rust" and (4) "Big Rust". Lewis (2011) finds that a loan description that read “my car has no rust” has a positive impact on prices and the other three rust qualifier variables have negative effect. Lewis (2011) then performs hedonic regressions to deconstruct the price of an automobile sale into the cars component parts with a focus on photos and text. The analysis suggest that keywords and phrases provided through social disclosures are important for investors in online markets.

Borrower Explanations

Michels (2012) delved into the specific context of loan descriptions explanations to extract the importance of a borrower justifying their circumstances and ability to repay the loan. Michels (2012) finds that lenders are influenced by the unverifiable disclosures made by borrowers and receive lower interest rates as a result. The Michel (2012) research supports the auction theory concept that disclosures deemed as credible help the lender gauge the value of the loan, therefore, increasing the number of bids. The study’s results demonstrate that for each additional social disclosure there is a 1.27 percentage point reduction in interest rate and an 8 percent increase in bidding activity. These findings were derived from 500 manually coded loan listings from the Prosper P2P lending platform containing the presence of specific voluntarily provided unverified information. Specifically, Michels (2012) scored the purpose of the loan, income amount, income source, education, amount of other debt, interest rate on other debt, explanation for poor credit grade, listing of monthly expenses, and a picture of a person (presumably the borrower). For example, social disclosures related to education, clarifying poor credit, itemizing monthly expenses, lowering debt rate, and disclosing other amounts of debt would be scored if the borrower indicated the successful completion of an education program, explanation of life circumstances that led to poor credit, listed the dollar value of monthly expenses, stated the numerical interest rate on their other debts, or the borrower provided the dollar value of existing amounts of other debt owed. Michels (2012) reports that explanations increase bidding activity by 18.21% and each disclosure provided reduces default probability by 5.37% for higher risk borrowers. The results of Michel (2012) indicate that disclosures influence lenders by increasing the number of bids and decreasing the interest rate charged on a loan. Furthermore, the analysis proves disclosures are more important for borrowers with poorer credit.

Hard Credit Information

Social disclosures within P2P lending represent a new innovation produced by web 2.0 technologies, however, verifying borrower information dates back to the first credit agency, The Retail Credit Company (now Equifax, Inc), was first founded in 1899. The Retail Credit Company began the credit reporting industry by collecting and selling information on creditworthy customers based a person's home, furnishings and character, among other factors (Myers & Forgy, 1963). P2P lenders collect similar information today to produce a borrower's risk profile and loan grade. Previous work by Emekter (2015) uses Lending Club data to confirm that high credit grade, low debt-to-income ratio, high FICO score and low revolving line utilization are the most significant factors associated with lowering default risk. The significant credit information variables identified by Emekter (2015) successfully separates the good borrowers from the bad borrowers and decreases the probability of default to 5.36% for the highest grade A Lending Club loans (FICO 780+). However, the Emekter (2015) study does not account for verified information and the different implications that verified information produces. Lending Club has three states of verification: Income Verified, Income Source Verified, and Not Verified. Income Verified is regarding the actual income that the borrower indicated to be confirmed and the Income Source Verified is confirmation of where the income is originating from, such as retirement, self-employed, business, disability, or regular W-2 employment. In some instances, Lending Club will verify both the source and the actual income. Figure 1 Percentage of Loans with Income Verification, the Lending Club platform does not verify 100% of the income information for the issued loans which may explain the discrepancies in the Emekter et al. (2015) study between the highest risk grade G loans (FICO 640-659) having a 30.34% Charge Off rate compared to lower risk F graded loans (FICO 660-678) at a 33.08% Charge Off rate. Counter intuitively, if a borrower is selected by the Lending Club proprietary algorithms for Income Verification or Income Source Verification this might indicate that the borrower was detected as risky based on their loan application information. As seen in Figure 2

Percentage of Loans Charged Off by Income Verification, Not Verified loans outperform both Income Verified and Income Source Verified Loans, which raises questions that have not been answered in previous related studies[4].

The remainder of this study consists of multiple interrelated social disclosure concepts to determine if social disclosures can separate good borrowers from bad borrowers in fixed-rate P2P lending platforms. Beginning with Chapter 3, the different approaches for analyzing social disclosures are reviewed and hypotheses are formed for the unanswered questions. Following the review of social disclosure articles, Chapter 4 provides the Data and Variable Descriptions, and Methodology that will be used to test the hypotheses, and the subsequent results. Chapter 5 summarizes the aforementioned contributions. Finally, the next areas of potential research are examined in the Conclusion.

[4] https://www.lendingclub.com/public/income-verification.action

Figure 1 Percentage of Loans with Income Verification

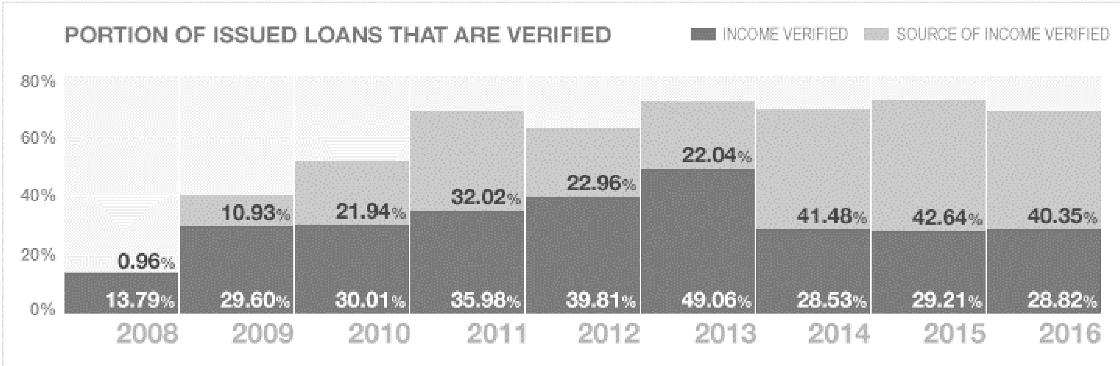
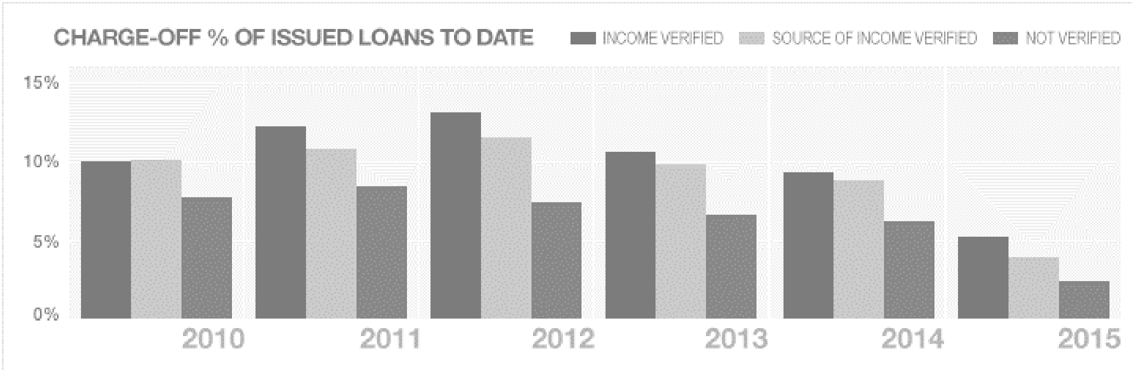


Figure 2 Percentage of Loans Charged Off by Income Verification



CHAPTER III

Hypotheses Development

As demonstrated in Table 1, the majority of the literature on P2P lending has revolved around the Prosper.com ("Prosper") reverse auction lending platform (D. Chen & Han, 2012; Duarte et al., 2012; Herzenstein, Sonenshein, et al., 2011; Lin, Prabhala, & Viswanathan, 2013; Michels, 2012; Pope & Sydnor, 2011). From 2005 to 2014, Prosper.com was the largest U.S. based firm. As of December 1, 2014, Prosper boasted over one million members and $2 billion in funded loans. Naturally, research was written within a reverse auction context where lenders competed against each other to offer borrowers the lowest interest rate. Multiple reverse auction studies support the ability of social disclosures to convey quality and influence lender behavior that can help minimize information asymmetries while reducing funding time and increasing the probability of funding (Agrawal, Catalini, & Goldfarb, 2013; Michels, 2012; Moritz & Block, 2016; Pope & Sydnor, 2011). However, studies have also found a fundamental difference between a reverse auction model and the fixed-rate model. The reverse auction model implies the interest rate for a loan is a function of the number of bids from interested capital-providers (D. Chen & Han, 2012). This format fosters an environment for buyer's remorse where lenders are incentivized to submit bids that were not aligned to the borrower's actual credit worthiness (Kawai, Onishi, & Uetake, 2014). Substantiating this finding, Chen et al. (2014) investigated and analyzed the results of the auction model used on Prosper and demonstrated the reverse auction method was more complicated and less transparent than a fixed-rate model for capital-providers (N. Chen, Ghosh, &

Lambert, 2014). For these reasons, Prosper filed with the United States Securities and Exchange Commission (SEC) to discontinue the reverse auction loan structure effective December 19, 2010 in favor of platform-established fixed-rates.

The adoption of fixed-rate P2P lending platforms and documented issues with reverse auction lending markets has created unanswered questions in the literature concerning social disclosures. Platform-established fixed-rates now dominate the peer-to-peer lending market, but only represent a fraction of the peer-reviewed literature. Accordingly, this study uses data from the Lending Club fixed-rate P2P platform, which eclipsed Prosper with over $15 billion in loans issued, as of March 2016. The key articles related to this study, Michels (2012) and Herzenstein et al. (2011), use data from the Prosper 1.0 reverse auction-lending platform that is no longer in operation, and Lewis (2011) uses the eBay Motors auction data that is limited to vehicle sales. Dorfleitner et al. (2016) study fixed-rate lending platforms but uses partial data from a third-party platform Wise Clerk to determine default. The data used from the Wise Clerk site is voluntarily provided and could be subject to selection bias. As a result, previous studies could not observe *ex post* loan default, were not collectively exhaustive in their analysis, or have been limited to auction formats.

The properties of auction models are based on the assumptions that all of the bidders are risk-neutral, each bidder has a private valuation for the item independently drawn from some probability distribution, the seller possesses symmetric information about their own valuation of the item, and the payment is represented as a function of only the bids (McAfee & McMillan, 1987). This model for auctions is fundamentally different from operations under a fixed-rate format. Addressing each auction property in order, in the market leading Lending Club P2P fixed-rate format lenders are risk neutral and select their risk via the loan grade. Second, the valuation of the loan is conveyed through the loan grade and interest rate. Third, borrowers possess symmetric information about their own creditworthiness, however the significance of this information is less important given 99.26% of loans issued receive their requested funding amount[5]. In other words, borrowers that are not rejected via the Lending Club screening process will have their loans almost fully funded. This is due to both retail and institutional lender investment as well as Lending Club subsidiaries investment in loans. Lastly, the payment and interest rates are not a function of the bids and there is a ceiling on the maximum amount of investment that can be received. In P2P fixed-rate lending borrowers no longer have the ability to obtain a lower interest rate or gain additional funding through persuasive social disclosures. In this environment, the incentives for borrower to use social disclosures are minimized and borrowers should not receive substantial economic gains from the social disclosures. Counterintuitively, 92.6% of borrowers in the sample provided some form of social disclosure. I posit that social disclosures will actually remain significant indicators and strong predictors of loan funding and repayment success in fixed-rate formats. I believe that the human element within borrowers causes them to provide disclosures even when it is unadvisable or negatively impacts their loans. I also believe lenders are not completely rational and are susceptible to compassion filled loan descriptions completely unrelated to the borrower's ability to repay the debt obligation. Lastly, I also trust social disclosures minimize asymmetric information between borrowers and lenders enabling cognizant lenders to discern and separate good and bad

[5] 92.21% of investment is by ordinary and institutional investors and 7.05% is from Lending Club subsidiaries.

borrowers. Thus, the first four research questions determine which forms of social disclosures are significant under fixed-rate platform parameters.

<u>Social Disclosure Forms Hypothesis</u>

The online relationship between capital-seekers and capital-providers is significant within P2P fundraising markets and the same is expected within P2P lending markets. I test this assumption through dependent variables that measure the Duration of time required to fund a loan, the amount of funding received by the borrower, the amount of principal recovered by the lender, and whether the loan was fully repaid. The Duration of time required to fully fund a loan is observable by subtracting the loan submission date from the loan issue date. The Total Invested variable provides the percentage of the borrower's requested loan amount that was funded, while the Percentage Invested only includes funding provided by peers (excludes Lending Club subsidiaries). Lastly, Total Recovered Principal measures the percentage of principal returned to lenders and Loan Status equal to Fully Paid is tested to determine the *ex post* influence of social disclosures.

The social disclosures used in this analysis are categorized as either form or content disclosures. Beginning with the forms, there are multiple formats information about the borrower is expressed to lenders. For example, Lin (2013) and Mollick (2014) identify quality signals in the loan descriptions and find that loan descriptions that contain typographical errors are less likely to be fully funded by project backers. I posit that loan descriptions that contain typographical errors, calculated through the Misspellings variable, should also indicate poorer creditworthiness to lenders and will be negatively related to Total Invested, Percent Invested, Total Recovered Principal, and Loan Status. P2P lenders all have access to the same information provided by the borrower; therefore, accurate high-quality loan descriptions are hypothesized to be more attractive to lenders and have a negative relationship with Duration times. Applying the same logic, I hypothesize that Word Count and Flesch Index indicate high quality loan descriptions that are detailed and well written. These forms of social disclosure will have a negative relationship with Duration and positive relationship to Total Invested and Percent Invested. I believe the overall presentation of the loan description is meaningful to lenders and may influence their investment decision leading to the following hypotheses.

H_1: Controlling for objective verifiable information, increasing the different forms of social disclosures decreases funding Duration.

H_2: Controlling for objective verifiable information, increasing the different forms of social disclosures increases investment from lenders.

There is a clear distinction between receiving funds and repayment of funds with any debt or credit obligation. The same forms of social disclosure positively associated with increasing investment are used to evaluate the return of investment. On crowdfunding platforms such as Kickstarter borrower and lender direct interaction takes the form of "Updates" and "Comments" in which investors can express enthusiasm or displeasure about the loan, product or project. Updates represent efforts by entrepreneurs to reach out to current and potential funders in order to inform interested investors about developments in a project (Mollick, 2014). Comments from current and potential investors that are answered by entrepreneurs on crowdfunding sites were found to be positively associated with achieving or exceeding funding goals (Mollick, 2014). Mollick (2014) also finds that meeting funding goals improves the ability of projects to fulfill their obligations to funders on time. Consistent with Mollick (2014), the Updates and Questions Answered variables in this study is expected to be an indicator of loan funding success as well as repayment success. Borrowers that respond to lender questions are expected to be more responsible and creditworthy individuals. These forms of disclosures are expected to proxy the relationship banking described in Rajan (1992) and to reduce information asymmetries and the likelihood of default. This analysis expects to show creditworthy borrowers are engaged and responsive, provide lengthy loan descriptions, update their loan listing, respond to lender questions, minimize misspelling errors, and post well-written descriptions. These forms of disclosure are expected to positively result higher Total Recovered Principal and probability of a Fully Paid loan leading to the following hypotheses:

H_3: Controlling for objective verifiable information, increasing the forms of social disclosures increases lender Total Recovered Principal.

H_4: Controlling for objective verifiable information, increasing the forms of social disclosures increases the probability of loans being Fully Paid.

<u>Content Analysis Hypothesis</u>

On a deeper level, understanding the specific content within social disclosures from borrowers will also substantially benefit the finance community. Previous research drawn from psychology and behavioral economics all demonstrate that voluntary unverifiable disclosures influence investing decisions (DellaVigna & Gentzkow, 2009;

Michels, 2012; Nisbett, Zukier, & Lemley, 1981). These studies also find that investors tend to incorporate information that is false or irrelevant into their decision making and also overlook conflicts of interest (Cain, Loewenstein, & Moore, 2005; Malmendier & Shanthikumar, 2007). Correctly parsing the meaningful content associated with creditworthy borrowers is difficult. I hypothesize that discerning investors can use specific content provided within social disclosure to predict creditworthiness. Three studies, Herzenstein et al. (2011), Lewis (2011), and Michels (2012), establish a methodology for delving deeper into the content of social disclosures being provided in online markets. Using the same dependent variables described in H_1-H_4 the relationship between specific content and loan performance is evaluated.

This analysis combines the three separate studies in order to confirm or reject the ability of specific content within social disclosures to influence both obtaining funding and repaying debt. Beginning with Herzenstein et al. (2011), the article finds that the identity claims that increase loan funding are less predictive of loan performance relative to other identities. Consistent with Herzenstein (2011), I expect to see identity claims for Trustworthy, Successful, Hardworking, Moral, and Religion negatively associated with Duration and positively associated with Total Invested and Percent Invested. Building upon Lewis (2011), analysis of keyword and qualifier phrases are shown to impact investors' decision making. Keywords such as "Rust" and "Dent" are the most relevant factors to automobile buyers and providing qualifying context around these keywords increases investment. I hypothesize that keywords and qualifiers may indicate the level of attention, focus, and understanding that a borrower possesses regarding debt instruments. On the other hand, using keywords may also provide lenders a sense of comfort and security with borrowers that use a common vocabulary and terminology. I expect combinations of "credit", "loan", and "debt" keywords and their qualifier phrases to have an impact on lender investment decisions. Accordingly, the "credit", "loan", and "debt" keywords and qualifiers are expected to result in a negative relationship with Duration and a positive relationship with Total Invested and Percent Invested. Lastly, Michels (2012) examines social disclosure explanations and their ability to mitigate information asymmetries. Monthly expense and interest rate explanations are two forms of explanations that indicate the borrower has set forth a measured plan that typically includes how the loan will be repaid. I posit that these types of explanations will also have a negative relationship with Duration and positive relationship with Total Invested and Percent Invested. In sum, the content analysis variables produced across the three studies lead to the following Duration and Investment hypotheses:

H_5: Controlling for objective verifiable information, increasing the number of identity claims, keywords and qualifiers, and borrower explanations within social disclosures decreases funding Duration.

H_6: Controlling for objective verifiable information, increasing the number of identity claims, keywords and qualifiers, and borrower explanations within social disclosures increases funding percentage in terms of Total Invested and Percent Invested by lenders.

Interestingly, *ex post* observations of the content analysis variables may produce sign changes based on the riskiness of the loan grade. For example, lenders may perceive a borrower identity claim for a low risk "A" grade loan to be different from the same identity claim on a high-risk "G" grade loan. Identity claims used for grade "A" loans may have less significance to lenders since the hard information, such as credit score, is higher. Specifically, Herzenstein et al. (2011) found higher-risk borrowers use more identity claims and the same results are expected in this study. As a result, I expect to see Successful and Hardworking identity claims to be positively related to Total Recovered Principal and Loan Status since these claims indicate an ability to repay the loan. While the Economic Hardship identity claim is expected to carry a negative relationship with repayment variables because the borrower is indicating a history of financial circumstances that could prevent loan repayment. In terms of keywords and qualifiers, I suspect that lower-risk and higher risk borrowers will use keywords in equal proportions. Both good and bad borrowers will feel compelled to use the same keywords in their loan descriptions to attract investment. Therefore, I posit that keywords and qualifiers will not be significant. Based on the Michels (2012) study, I expect educational explanations to indicate the borrower's potential earning potential and ability to repay debt obligations. I also expect the monthly expense and interest rate explanations to indicate the borrower's plan for lowering household expenses required to service the loan. For these reasons, I expect a positive relationship between the repayment variables and education, monthly expense, and interest rate explanations. Conversely, poor credit and other debt amount explanations fail to indicate how the borrower will be reducing expenses to increase income and service the debt obligation. As a result, these explanations are expected to have a negative relationship with Total Recovered Principal and the Loan Status of Fully Paid. Accordingly, I posit that specific content analysis variables will improve loan repayment leading to the following hypotheses:

H_7: Controlling for objective verifiable information, increasing the number of Successful

and Hardworking identity claims, Educational, Monthly Expense, and Interest Rate explanations within social disclosures increases the Total Recovered Principal.

H_8: Controlling for objective verifiable information, increasing the number of Successful and Hardworking identity claims, Educational, Monthly Expense, and Interest Rate explanations within social disclosures increases the probability of loans being Fully Paid.

The answers to each of these hypotheses, H_1-H_8, will indicate the loan characteristics that influence lenders to invest and the social disclosures that separate good borrowers from bad borrowers. The methodology and test results are provided for each hypothesis in the following section.

CHAPTER IV

Methodology and Results

Sample Data

Lending Club is the world's largest P2P lending company with loan originations exceeding $2 billion dollars. The San Francisco, California based firm established in 2007 host an online lending platform that enables borrowers to obtain a loan, and investors to purchase unsecured Notes (fractions of a loan) from borrowers. In 2011, $261 million of loans were originated on the Lending Club platform, and increased 2.75 times to $718 million in 2012 and to $1.9 billion in 2013. In 2014, the year Lending Club became a public company, the firm originated $3.5 billion in loans (Puls, 2015). In order to qualify for a loan, a number of factors are considered including but not limited to, the information provided on the loan application, information provided about the borrower by credit bureaus, borrower credit score, debt-to-income ratio, length of credit history, the number of other accounts that the applicant has open, payment history with open accounts, and recent credit inquiries. From these inputs, the Lending Club platform either rejects or accepts the loan application and subsequently assigns a credit grade, interest rate, and creates the loan listing. The loans have a maximum value of $35,000 and average loan size of $10,775.29 in my sample.

The sample was primarily formed from Lending Club publicly available data that can be downloaded via a comma separated values (CSV) spreadsheet. However, the publicly available data downloads do not contain all of the same information available to lenders during the loan funding window. The data, however, contain a website link to the original loan listing. The information lenders are able to review during the sample period on the platform is provided in Figure 3 Lending Club Listing. To obtain the exact same information available to lenders each of the loan listing's uniform resource identifier (URL) is used to collect the additional information for inclusion with the publicly available data set from Lending Club. The resulting data set captures additional social disclosure information, Lending Club platform information, and hard credit information. The combined URL and CSV data was then filtered to the sample period beginning June 2007 and ending April 2013. Additional filters were applied to limit the Loan Status to only "Fully Paid" or "Charged Off" and the loan term to 36-months. To correct for inflation from measurement error, I Winsorize the outliers in the Flesch Index score. The Winsorizing process involved setting the negative Flesch Index scores to a value of zero (Dixon, 1960). After applying all data filters and parameters the final sample consist of 31,550 loans with a 36-month terms and a terminal loan status.

Figure 3 Lending Club Listing

Credit card refinancing

Glossary

Borrower Member Loan 7381957 | Lending Club Prospectus
Next »

Invest in **Member_9043865**

Amount Requested	**$8,500**	Review Status	**Under Review**
Loan Purpose	**Credit card refinancing**	Funding Received	**$7,725 (90.88% funded)**
Loan Grade	B1	Investors	**135 people funded this loan**
Interest Rate	**9.99%**	Listing Expires in	**13d 9h (10/7/13 3:29 PM)**
Loan Length	**3 years (36 payments)**	Loan Status	**In Funding**
Monthly Payment	**$274.24 / month**	Loan Submitted on	**9/23/13 3:29 PM**

Member_9043865's Profile (all information not verified unless noted with an "*")

Home Ownership	**OWN**	Gross Income	**$3,817 / month**
Current Employer	**n/a**	Debt-to-Income (DTI)	**15.62%**
Length of Employment	**n/a**	Location	**HIGHLAND, CA**

Member_9043865's Credit History (as reported by credit bureau on 9/23/13)

Credit Score Range:	**685-689**	Accounts Now Delinquent	**0**
Earliest Credit Line	**01/2001**	Delinquent Amount	**$0.00**
Open Credit Lines	**8**	Delinquencies (Last 2 yrs)	**0**
Total Credit Lines	**16**	Months Since Last Delinquency	**n/a**
Revolving Credit Balance	**$12,038.00**	Public Records On File	**1**
Revolving Line Utilization	**50.40%**	Months Since Last Record	**85**
Inquiries in the Last 6 Months	**0**	Months Since Last Major Derogatory	**n/a**

Loan Description

Questions & Answers

No questions yet.

- What is your intended use for the loan proceeds?
- What are your current monthly expenses (rent, transportation, utilities, phone, insurance, food, etc)?
- What are your monthly expenses related to housing (rent, mortgage(s), home equity loan and / or line of credit, utilities, insurance, taxes, etc)?
- If you have a delinquency in the last 2 years, please explain the reason.
- Please explain the reason why you carry a large revolving credit balance.
- If you have a public record on file, please briefly explain the event and its resolution.
- If you are paying a mortgage, please break down all monthly housing related expenses (mortgage payment, insurance, taxes, etc).
- If using your loan for multiple purposes, what are the purposes and how are you allocating the money across them?
- Please itemize for each of the credit cards you plan to pay off the card name (Visa, MasterCard, etc - Please not include bank issuer of card), outstanding balance, current interest rate, and current minimum monthly payment.

https://www.lendingclub.com/browse/loanDetail.action?loan_id=7381957&previous=browse 1/2

Dependent Variables

The chief concern in this study is whether social disclosures influence investment and indicate credit worthiness. A lender will only benefit if the social disclosures can be reliably used to separate good borrowers from bad borrowers. The methodology is designed to test the social disclosures Lenders are able to observe and then produce results that indicate if the Funding Amount Invested, Charge Off rate, and Percentage of Recovered Principal increases or decreases based on the form of social disclosures and content within the social disclosure. That said, the first dependent variable analyzes the Duration of time required to fund loans with social disclosures. The Duration variable measures the attractiveness of the loan listing based on time required to fund the loan. Duration is calculated as the difference between the Loan Submitted Date and Loan Issue Date. Duration is calculated in hours to account for the lending platform being open for investment twenty-four hours each day.

$$Duration \ = (Loan\ Submitted\ Date - Loan\ Issue\ Date) * 24$$

The second dependent variable examines the percentage of investment by retail and institutional lenders as well as Lending Club subsidiary investment. The dependent variable provides an *ex ante* value that indicates the types of social disclosures that attract or deter lenders to invest in a particular loan and is calculated as:

$$Total\ Invested\ Percentage \ = \frac{fund\_amnt}{loan\_amnt}$$

The third dependent variable examines the percentage of investment by only retail and institutional lenders in P2P lending platforms between the loan submission date and the loan issue date. The dependent variable provides an *ex ante* value that indicates the types of social disclosures that attract or deter lenders to invest in a particular loan and is calculated as:

$$Percent\ Invested\ Percentage\ = \frac{fund\_amnt\_inv}{fund\_amnt}$$

For the fourth dependent variable, I observe the overall Total Recovered Principal returned to the lender. Total Recovered Principal measures the percentage of principal paid by the borrower and is calculated as:

$$Total\ Recovered\ Principal = \left(\frac{Total\ Recovered\ Principal}{Funded\ Amount}\right)$$

The last area of inquiry is the Loan Status. A Loan Status equal to 1 represents a Fully Paid loan and Loan Status equal to 0 represents a Charged Off loan. Loan Status measures the social disclosures' predictive ability in P2P lending as a binary variable. In a binary logistic regression, a dependent variable is the probability of the event to occur, in this case it is a Loan Status of Full Paid Off f_i. To convert this number into a number between zero and one, the following transformation is used:

$$p_i = \frac{1}{1 + e^{-f_i}}$$

Using a sample of 31,639 Lending Club loans, I propose to test each dependent variable in hopes of producing a significant contribution in the field of finance. A summary of the aforementioned variables are provided in Table 2 Dependent Variables.

Table 2 Dependent Variables

P2P Dependent Variables	Abbreviation	Variable Description
Duration	duration	Calculation of the Note Submitted on Date minus the Loan Issued on Date to derive duration measured in hours.
Total Invested	tot_inv (%)	The total amount invested by both retail and institutional investors as well as Lending Club subsidiaries
Percent Invested	pct_inv (%)	The total amount committed by only retail and institutional investors for the loan expressed as a percentage of the funded amount.
Total Recovered Principal	pct_rec_prncp (%)	The total amount funded by lenders divided by the recovered principal amount. This is a continuous variable from 0% to 100%.
Loan Status	loan_status	Current status of the loan. This is a dummy variable that is assigned Fully paid = 1, Charged off = 0

Social Disclosure Forms Independent Variables

Analysis of the different types of social disclosures provided in this study is used to understand if the presentation of the social disclosure information is a significant predictor in P2P lending markets. The first variable measured is the loan description Word Count determined by using a formula that calculates the number of words in the description based on the count spaces and length of the text within the loan description. This number is then subtracted from the length of the text with spaces to calculate the number of words in the description result. A count of Questions Answered (QA) and Updates are performed as a proxy for social disclosures. The QA captured from each loan listing webpage are denoted with a “Q:” for question or an “A:” for answer. A count of records with “A:” responses are counted for each loan to determine the number of questions the borrower answered related to their loan request. In a similar fashion, Updates to the loan description provided by the borrower are denoted by “borrower added on:” with an appended date. Each borrower Update is totaled for hypothesis testing and analysis. Lastly, I include two quality measures by incorporating Misspellings and the Flesch Index variables that capture whether the social disclosure conforms to the standard English grammar and punctuation rules. The Misspelling variable is calculated using Andrew Golding and Dan Roth's "Winnow-based spelling correction algorithm," published in 1999, which is able to recognize about 96% of context-sensitive spelling errors, in addition to ordinary non-word spelling errors. The Flesch Index reading score indicates how difficult a passage in English is to understand. Higher scores indicate material that is easier to read, while lower scores are more difficult to read. Use of this scale is included in word processing programs and services such as Microsoft Office

Word, WordPerfect, and WordPro. The next group of variables relate to the qualitative content analysis.

Content Analysis Independent Variables

This study recreates the scoring approach used in Herzenstein et al. (2011), Lewis (2011), and Michels (2012), for content analysis that captures personal identity claims, keyword qualifiers, and credit explanations to assess social disclosures. Beginning with Herzenstein et al. (2011), I measure the number of personal identity claims related to trustworthiness, personal success, economic hardship, work ethic, morality, and religion within a social disclosure. Identity claims are scored each time borrower references keywords related to trustworthiness, personal success, economic hardship, work ethic, morality, or religion. To develop this list of keywords related to each identity claim type the Merriam-Webster Thesaurus of synonyms and related words are included for each identity claim. A calculation is then used to sum the number of identity claims found in the identity claim results column. Herzenstein et al. (2011) with the assistance of 10 research assistants also coded demographic information on the Prosper Lending platform using both narratives and the borrower's picture. Note, based on my review of loan descriptions the majority of the demographic information related to race and gender was most likely based on the picture provided by the borrower. However, profile pictures are not allowed on Lending Club. Due to this difference in platforms, I have excluded the race demographic information in the data set and in the regression results.

Table 3 Identity Claim Examples

Identity Claim Criteria	Example
Trustworthy – Lenders can trust the borrower to pay back the loan on time.	Loan ID 3153865 added on 01/16/13: "I've accumulated this debt after 3 years or rough times and need to pay it off the right way. I'm a very reliable and responsible person. I have done my calculations and Lending Club would help me save on the interests that are charged on my credit cards by far! My goal is to start fresh w/ no debt."
Successful – The borrower is someone with a successful business, job, or career.	Loan ID 607833 added on 11/01/10: "The recession hit me hard as I graduated from grad school with a lot of student loans and a difficult job market. Since then, I've found a great, stable job where I am flourishing. I've been diligently paying off my school loan payments and chiseling away at my credit card debt, but my credit card company has raised rates to a ridiculous level. I would be a great borrower--I'm responsible, I have a great credit score and pay all of my bills on time. This debt is weighing heavily on me, and I'm looking for a little help now to get some peace of mind, and I hope to pay it forward to others in the future."
Economic Hardship – The borrower is someone in need because of hardship, as a result of difficult circumstances, bad luck, or other misfortunes that were or were not, under the borrower's control.	Loan ID 2091941 added on 11/18/12: "Hurricane Sandy property repairs Borrower added on 11/18/12: Pay for significant Hurricane Sandy damages incurred by my home and vehicle. This tragedy hit my family very hard. I have a great track record of paying back my loans in full, on time. I have a very stable job as a software engineer at a prominent financial services company with a strong salary."
Hardworking - The borrower will work very hard to pay back the loan back.	Loan ID 4308544 added on 04/13/13: "This loan will be used pay pay off credit card debt. I am responsbile and my employment is very stable. I have worked in the same industry for 30 years and with my current employeer for 6 years. I'm hard working, reliable and look forward to being debit free in less than 3 years! Thank you."
Moral – the borrower is an honest or moral person.	Loan ID 828449 added on 07/24/11: "I request this loan to pay off credit cards, currently charging me over 24%APR. My current job is stable and very rewarding. This loan will provide the relief I so much need and help me deal with my debt situation, leaving me in a

Identity Claim Criteria	Example
	better position to quickly pay off my loan. I've been dealing with this debt burden for too long and I pride my credit worthiness and moral credibility.”
Religious – the borrower is a religious person.	Loan ID 668663 added on 02/06/11: “Loan to go to Italy on vacation and for a Spiritual Break - I am an Episcopal Priest who works at a parish and helps to run a homeless shelter...do not get to take much time away as my mother lives with me and I have to hire a nurse when I do get to travel. Blessings!”

Note: Multiple identity claims can exist in a single loan description and each is scored.

Next, the data is filtered to nouns most frequently used (for example “credit” had a frequency of 1833, compared to 33 for “market”). The top 3 keywords “Credit”, “Loan”, and “Debt” were selected from this list to be the variables that interact with qualifier words. The Top 100 nouns used are provided in the table below. Note, “Will”, “Still”, “Back”, “Few”, “Most”, “May”, “Part”, “Lot”, “Put”, and “Use” are within the top 100 most frequently used nouns but were removed because they were reasonably assumed to be used as verbs within the context of P2P lending.

Third, I analyze concordance for each word chosen, by observing how it was used in context, by examining a list of qualifying phrases surrounding each word. A list of keywords words and qualifiers were derived from the corpus and Merriam-Webster thesaurus of related words. For example, for negations, I used “no”, “not”, “never”, “nothing”, “free” (as in “debt free”), “zero”; for adjectives, I used “small”, “minor”, etc.

Fourth, I search for each of the three nouns (“Credit”, “Loan”, and “Debt”) and where they were found, I perform a secondary search for any of the qualifiers within 50 characters of the noun. I perform multiple calculations to pair keywords and qualifiers and then score them appropriately as dummy variables Credit Combination, Loan Combination, and Debt Combination. I then score each row of data to determine the count of Credit references with No Credit, Less Credit, and Credit if no qualifiers are present. I repeat this process for Loan and Debt keywords in order to produce three dummy variables to capture Credit Combination, Loan Combination, and Debt Combinations of keywords and qualifiers.

Table 4 Loan Description Noun Frequency

Top 100 Nouns Used in Loan Descriptions Frequency Table (*n*=2,000)									
credit	1833	income	216	bank	104	wedding	67	process	40
loan	1715	stable	216	save	104	term	63	saving	40
debt	1045	company	190	life	103	secure	62	employer	39
card	790	business	181	family	99	living	61	care	38
interest	688	balance	180	rent	94	cost	59	couple	38
payment	636	current	177	wife	92	future	57	date	38
one	449	two	173	total	90	close	56	salary	38
time	447	thanks	157	way	90	end	56	single	37
job	429	well	147	start	88	opportunity	55	state	37
rate	402	club	141	cash	83	vehicle	53	capital	36
money	368	great	140	insurance	81	investment	52	finance	36
month	358	purchase	136	order	79	property	52	people	36
help	341	house	133	account	75	right	52	industry	35
high	331	full	123	cover	74	buy	51	project	35
need	319	amount	121	bill	72	purpose	51	support	35
home	304	score	121	goal	71	hope	48	request	34
good	264	school	118	move	71	employment	47	second	34
work	234	budget	115	low	70	love	42	gas	33
year	230	college	107	position	68	day	41	market	33
car	226	history	107	student	68	place	40	phone	33

Notes: The TextStat tool analysis of word frequency was cross-referenced against a dictionary of English nouns to develop a corpus of words based on a random sample of 2,000 loan descriptions.

Table 5 Keyword and Qualifier Examples

Qualifier Criteria	Example
X – no qualifier phrases used in the 50 characters preceding keywords Credit, Loan, and Debt	Loan ID 4277144 added on 04/11/13: “Refinancing Credit Cards”
No X – negation phrases used in the 50 characters preceding keywords Credit, Loan, and Debt	Loan ID 3290065 added on 01/31/13: “Never been late a single day and even without the loan …”
Less X – minimization phrases used in the 50 characters preceding keywords Credit, Loan, and Debt	Loan ID 4300149 added on 04/15/13: “This loan is for debt consolidation. I have a few small balance loans/ credit cards …”

Note: Multiple keywords and qualifiers can exist in a single loan description. Each type of qualifier criteria is only scored once per loan description.

The last content analysis approach adopted is from Michels (2012). Michels (2012) measures purpose of the loan, income amount, income source, education, amount of other debt, interest rate on other debt, explanation for poor credit grade, listing of monthly expenses, and a picture of a person (presumably the borrower) using the Prosper Lending platform. Lending Club provides the purpose of the loan, income amount, income source, and does not allow borrowers to provide a picture. The remaining variables applicable to by Lending Club are explanations related to education, poor credit, borrower's monthly expenses, other debt rate, and the amount of other expenses. This subset of explanation variables are scored in this analysis. A key difference between Michels (2012) and my study are the number of loans being reviewed. I have 31,639 records compared to the sample of 500 records used in Michels (2012). In order to identify the records that have Education, Poor Credit, Monthly Expenses, Other Debt Rate, and Amount of Debt explanation. I use an algorithm to search each loan explanation type for the keywords associated with Education, Poor Credit, Monthly Expenses, Other Debt Rate, and Amount of Debt explanations. The algorithm identified 10,661 records that were reviewed and manually coded. Results were coded in the same manner as Michels (2012) where one (1) point is assigned for each explanation disclosed. The following criteria and corpus were used for manually coding the data.

Table 6 Credit Explanation Scoring Examples

Explanation Criteria	Example
Education - borrower indicates they successfully completed an Education Program.	Loan ID 2378969 added on 12/08/12: "I am seeking this loan to get myself out of credit card debt's high interest rates. I am a college graduate and I work very hard to get ahead in life. I volunteer in my community in my spare time as a way of giving back. I hope I can help your investments while you help me wipe out debt!"
Poor Credit - borrower explains why his or her credit grade is low.	Loan ID 2381074 added on 12/10/12: "After graduating college I unfortunately became a victim to a debt consolidation scam. They were supposed to pay my bills while I paid them. They did not make those payments. I have worked the last 5 years to get my credit back. This situation has caused so many credit problems for me. Thank you."
Monthly Expenses - borrower lists their monthly expenses.	Loan ID 3153832 added on 01/16/13: "Budget-Rent-$1,825.00, Utilities and Internet-$368.00, Food-$500.00. No transportation cost. Have been with company for six years, in current position for two."
Interest Rate on Other Debt - the listing reports the interest rate on at least one of the borrower's other debts.	Loan ID 3642613 added on 03/06/13: "A credit card account is charging me 23.99% - i'm just trying to reduce my interest rate to make faster progress in paying off my debt."
Amount of Other Debt - the borrower reports outstanding balances of other debt.	Loan ID 3373085 added on 03/12/13: "want to pay off two credit cards. one has had about a $6000 balance for a few years now. just seem like i am not making a dent. the other card is under $1800, that was mostly from buying home heating oil this winter. just want to pay them both off and keep the one for emergencies. thank you.."

Notes: Multiple explanations can be used in the same loan description. Each type of explanation is only scored once per loan description.

The independent variables used in this study were produced from a variety of sources to create a unique data set. Algorithms for Word Count, Updates, Questions Answered, and Flesch Index were designed as well as manual scoring of each loan description. The manual scoring required for the content analysis was performed by three research assistants and then revised by one of the lead research assistant to ensure consistency in the methodology. The complete list of independent variables represented in the study is provided in Table 7 Independent Variables below:

Table 7 Independent Variables

Variable	Abbreviation	Variable Description
Word Count	word_count	Word Count is based on the number of words in the Loan Description and calculated using a word parsing formula.
Updates	updates	Updates to the loan descriptions during the funding window occurs when the borrower appends additional information to the loan description originally provided.
Questions Answered	qa_total	The count of the lenders questions answered during the funding window.
Misspellings	misspellings	Misspellings are based on the number of Misspellings in the Loan Description and calculated using Andrew Golding and Dan Roth's Winnow-based spelling correction algorithm.
Flesch Index	flesch_index	The Flesch Index readability score uses the sentence length (number of words per sentence) and the number of syllables per word in an equation to calculate the reading ease. Texts with a very high Flesch Index reading ease score (about 100) are very easy to read.
Trustworthy	trustworthy	Borrower indicated they are Trustworthy then the dummy variable takes the value of 1 and 0 otherwise.
Successful	successful	Borrower indicated they are Successful then the dummy variable takes the value of 1 and 0 otherwise.
Economic Hardship	economic_ hardship	Borrower indicated they are Economic Hardship then the dummy variable takes the value of 1 and 0 otherwise.
Hardworking	hardworking	Borrower indicated they are Hardworking then the dummy variable takes the value of 1 and 0 otherwise.
Moral	moral	Borrower indicated they are Moral then the dummy variable takes the value of 1 and 0 otherwise.
Religious	religious	Borrower indicated they are Religious then the dummy variable takes the value of 1 and 0 otherwise.
Credit Combination	credit_combo	The keyword "Credit" was provided in the loan description without qualifiers preceding the keyword, or the keyword "Credit" was provided in the loan description with negation qualifiers preceding the keyword, or the keyword "Credit"

Variable	Abbreviation	Variable Description
		was provided in the loan description with minimizing qualifiers preceding the keyword.
Loan Combination	loan_combo	The keyword "Loan" was provided in the loan description without qualifiers preceding the keyword, or the keyword "Loan" was provided in the loan description with negation qualifiers preceding the keyword, or the keyword "Loan" was provided in the loan description with minimizing qualifiers preceding the keyword.
Debt Combination	debt_combo	The keyword "Debt" was provided in the loan description without qualifiers preceding the keyword, or the keyword "Debt" was provided in the loan description with negation qualifiers preceding the keyword, or the keyword "Debt" was provided in the loan description with minimizing qualifiers preceding the keyword.
Education Explanation	education	A 1 is coded when the borrower indicates they successfully completed an education program, 0 otherwise.
Explanation of Poor Credit	poor credit	A 1 is coded when the borrower explains the life circumstances that the led to the Poor Credit.
Monthly Expenses Explanation	monthly expenses	A point is awarded when the borrower provides the dollar value of at least one Monthly Expense, 0 otherwise.
Other Debt Rate Explanation	other debt rate	A point is awarded when the borrower states the numerical Interest rate on their Other Debts
Amount of Other Debt	amount of other debt	A 1 is coded when the borrower provides the dollar value of an existing Amount of Other Debt owed, 0 otherwise.

Control Variables

The control variables include both endogenous and exogenous factors to control for platform determined, loan application, and market variables. Within the platform-determined variables, I control for the Verification Status, which contains three states: Income Verified, Income Source Verified and Not Verified. The Lending Club platform can request the borrower verify income, the income source, or both income and income source, but do not indicate in the publicly available data when both income and income source are verified. For these reasons, I use a higher level of abstraction and convert Verification Status to a binary variable for either Verified or Not Verified coded as 0 or 1 respectively. Within the loan application variables, Loan Purpose comprises 14 variables that are standardized for borrowers to select from when completing the loan application. The 14 available Loan Purposes include car, credit card, debt consolidation, educational, home improvement, house, major purchase, medical, moving, other, renewable energy, small business, vacation, and wedding loans. Note the "Other" loan purpose category is made the reference variable in this analysis.

The first of the loan application variables is the Funded Amount Requested. The Funded Amount Requested is total amount of money requested by the borrower and has been converted to a log value with base 10 in the analysis. This list of Loan Purposes has been converted into dummy variables taking the value of 0 or 1. The next set of loan application variables are verifiable through third-party service providers and Lending Club verification mechanisms for the borrower's FICO Score, Debt-to-Income (DTI) ratio, Home Ownership, Revolving Credit Line Utilization, and Monthly Income. All of these variables were found to be significant in Emekter et al. (2015). Consistent with

Emekter et al. (2015) the annual income was converted to Monthly Income by dividing the annual income by 12 months. I then take the log of the Monthly Income for consistency across the study.

The last set of control variables represents market factors that are present during the study. The Loan Volume variable controls for the daily number of loans issued on the platform within a 24-hour period. The Credit Spread variable is the platform assigned interest rate minus the 2-year Treasury rate. The Credit Spread variable enables the study to control for the market interest rate changes over time. I also control for economic sentiment changes throughout the year that can be measured monthly using the Michigan Consumer Sentiment Report. The loan issue date and Michigan Consumer Sentiment Report for the corresponding month control for periods with better economic sentiment that historically result in higher rates of default. The Dorfleitner (2016) article introduces the January effect (Turn Year) concept for P2P credit markets and suggest that periods with better economic sentiment predict a higher probability of default. This observation is limited to loans in the January time window and are extended across the entire calendar year in this study. The Log of Per Capita Wages is also included as a control variable and is based on the three-digit zip code prefix provided in the Lending Club data. The three-digit zip code prefix s used to determine the borrower's city in order to calculate the city's population and divide the population by the total aggregated wages for that city based on 2012 Census data. In Table 8 Control Variables a summarized list of variables is provided.

Table 8 Control Variables

Control Variables	Abbreviation	Variable Description
Verification Status	verified_status	Indicates if income was verified by Lending Club, not verified, or if the income source was verified. The verification status was converted to a binary variable where 0=Verified and 1 = Not verified
Car	car	Borrower indicated the purpose of the loan is for a car. The dummy variable takes the value of 1 and 0 otherwise.
Credit Card	credit_card	Borrower indicated the purpose of the loan is for a credit card. The dummy variable takes the value of 1 and 0 otherwise.
Debt Consolidation	debt_consolidation	Borrower indicated the purpose of the loan is for a debt consolidation. The dummy variable takes the value of 1 and 0 otherwise.
Educational	educational	Borrower indicated the purpose of the loan is for an educational. The dummy variable takes the value of 1 and 0 otherwise.
Home Improvement	home_improvement	Borrower indicated the purpose of the loan is for a home improvement. The dummy variable takes the value of 1 and 0 otherwise.
House	house	Borrower indicated the purpose of the loan is for a house. The dummy variable takes the value of 1 and 0 otherwise.
Major Purchase	major_purchase	Borrower indicated the purpose of the loan is for a major purchase. The dummy variable takes the value of 1 and 0 otherwise.
Medical	medical	Borrower indicated the purpose of the loan is for a medical. The dummy variable takes the value of 1 and 0 otherwise.
Moving	moving	Borrower indicated the purpose of the loan is for a moving. The

Control Variables	Abbreviation	Variable Description
		dummy variable takes the value of 1 and 0 otherwise.
Other (reference)	other (reference)	Borrower indicated the purpose of the loan is for "other" (reference). The variable is excluded as a reference variable.
Renewable Energy	renewable_energy	Borrower indicated the purpose of the loan is for a renewable energy. The dummy variable takes the value of 1 and 0 otherwise.
Small Business	small_business	Borrower indicated the purpose of the loan is for a small business. The dummy variable takes the value of 1 and 0 otherwise.
Vacation	vacation	Borrower indicated the purpose of the loan is for a vacation. The dummy variable takes the value of 1 and 0 otherwise.
Wedding	wedding	Borrower indicated the purpose of the loan is for a wedding. The dummy variable takes the value of 1 and 0 otherwise.
Log Funded Amount Requested	log_amount_requested	The total amount requested by the borrower converted to a log with base 10.
FICO Score	FICO_avg	Borrower's credit score is calculated as the average between the high and low FICO score.
Debt-to-Income (DTI)	dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested Lending Club loan, divided by the borrower's self-reported monthly income.
Home Ownership	home_own	The home ownership status provided by the borrower during registration and has been converted to a binary variable for 0=Ownership and 1=Rent.
Married	married	Borrower indicated they are Married then the dummy variable takes the value of 1 and 0 otherwise.

Control Variables	Abbreviation	Variable Description
Divorced	divorced	Borrower indicated they are Divorced then the dummy variable takes the value of 1 and 0 otherwise.
Single	single	Borrower indicated they are Single then the dummy variable takes the value of 1 and 0 otherwise.
Engaged	engaged	Borrower indicated they are Engaged then the dummy variable takes the value of 1 and 0 otherwise.
Children	children	Borrower indicated they are Children then the dummy variable takes the value of 1 and 0 otherwise.
Revolving Line Utilization	revol_util	Revolving line utilization rate is the amount of credit the borrower is using relative to all available revolving credit.
Log Monthly Income	mthly_inc_log($)	The annual income provided by the borrower during registration has been converted to monthly income (gross income/12 months) with log base 10.
Log Loan Volume	loan_vol_log	The daily volume of new loans on the platform determined by the issue date. Converted to log base 10.
Credit Spread	credit_spread (%)	Provides the loan interest rate minus the Federal Reserve 2 year Treasury Bonds interest rate.
Consumer Sentiment	consumer_sentiment	Provides Michigan Consumer Sentiment Report data corresponding to the issue date month of the loan.
Log Per Capita Wages	log_per_capita_wage	Based on the three-digit zip code prefix provided in the Lending Club data the borrower city population is divided by the total aggregated wages for that city based on 2012 Census data. Converted to a log base 10.

The descriptive statistics from the sample is provided in Tables 10, 11, and 12. The descriptive statistics have also been constructed to provide both the high–level analysis of the data, but also to demonstrate the difference between loans listed with social disclosures and without social disclosures present. Beginning with the total loan amount funded equals $340,919,425 between the periods of June 1, 2007 to April 1, 2013. These loans all have a loan status of either Charged Off or Fully Paid. From this total $51,570,350 of the loans were Charged Off and $289,349,075 of the loans were Fully Paid. At the highest level of aggregation this equals 84.9% of loans issued within the sample period were Fully Paid and 15.1% were Charged Off.

Table 9 Dependent Variable Descriptive Statistics

	Mean	Median	Std. Deviation	Minimum	Maximum
duration (hrs)	198.63	177.10	106.01	6.75	774.92
tot_inv(%)	99.26%	100.00%	5.47%	10.25%	100.00%
pct_inv (%)	92.21%	100.00%	21.37%	0.00%	100.00%
pct_rec_prncp(%)	89.83%	100.00%	24.70%	0.00%	100.00%
loan_status	0.85	1.00	0.355	0	1
N	31,550	31,550	31,550	31,550	31,550

Notes: Table 9 describes the dependent variables used in the analysis for sample period of June 2007 to April 2013. Duration is a continuous variable based on the number of hours between the loan submission and loan issue date and time. The total_inv (%), pct_inv (%), and pct_rec_prncp (%) are continuous variables expressed as percentages. The loan status is a binary variable equaling 1 for Fully Paid and 0 for Charged Off loans. The lack of variation in the Median and Maximum statistics is a result of loans receiving an average of 99.26% of the requested loan amount and Lending Club policies that prevent loans receiving funding in excess of 100% of the requested amount.

Table 10 Independent Variables Descriptive Statistics

	Mean	Median	Std. Deviation	Minimum	Maximum
word_count	61.10	39.00	71.812	0	1056
updates	0.97	1.00	0.863	0	28
qa_total	1.35	1.00	1.784	0	15
misspellings	0.56	0.00	1.540	0	37
flesch_index	64.9	65.71	17.99	0	160.48
trustworthy	0.22	0.00	0.414	0	1
successful	0.03	0.00	0.158	0	1
economic_hardship	0.05	0.00	0.220	0	1
hardworking	0.32	0.00	0.465	0	1
moral	0.17	0.00	0.372	0	1
religious	0.00	0.00	0.050	0	1
credit_combo	0.54	1.00	0.498	0	1
loan_combo	0.51	1.00	0.500	0	1
debt_combo	0.38	0.00	0.485	0	1
edu_exp	0.03	0.00	0.157	0	1
poor_credit_exp	0.01	0.00	0.087	0	1
mthly_expense_exp	0.03	0.00	0.180	0	1
oth_rate_exp	0.07	0.00	0.253	0	1
amnt_oth_debt_exp	0.02	0.00	0.153	0	1
N	31,550	31,550	31,550	31,550	31,550

Notes: Table 10 describes the independent variables used in the analysis for the sample period June 2007 to April 2013. The word_count is the total number of words used in the loan description. Updates is the count of revisions to the loan description. The qa_total is the count of borrower answers to lender questions. Misspellings are the number of grammatical errors identified. The flesch_index variable is a calculated readability score. A reference to trustworthy, successful, economic_hardship hardworking, moral, and religious is scored 1 if present and 0 otherwise. Credit, loan, and debt keywords and phrases are scored using credit_combo, loan_combo, or debt_combo variables with a 1 if present and 0 otherwise. Borrower explanations for education experience, poor credit, monthly expenses, other interest rates, and amount of other debt, are scored using edu_exp, poor_credit_exp, mthly_expense_exp, oth_rate_exp, and amnt_oth_debt_exp with a 1 if present and 0 otherwise.

Table 11 Control Variable Descriptive Statistics

	Mean	Median	Std. Deviation	Minimum	Maximum
verified_status	0.51	1.00	0.500	0	1
car	0.02	0.00	0.156	0	1
credit_card	0.17	0.00	0.372	0	1
debt_consolidation	0.50	1.00	0.500	0	1
educational	0.01	0.00	0.108	0	1
home_improvement	0.06	0.00	0.240	0	1
house	0.01	0.00	0.096	0	1
major_purchase	0.04	0.00	0.204	0	1
medical	0.01	0.00	0.120	0	1
moving	0.01	0.00	0.109	0	1
other (reference)	0.09	0.00	0.284	0	1
renewable_energy	0.00	0.00	0.045	0	1
small_business	0.04	0.00	0.191	0	1
vacation	0.01	0.00	0.087	0	1
wedding	0.02	0.00	0.137	0	1
loan_amount_req_log	3.94	4.00	0.30	2.70	4.54
FICO_avg	709.49	702.00	35.67	612.00	847.50
dti	14.30	14.19	7.29	0.00	34.96
home_own	0.48	0.00	0.500	0	1
married	0.06	0.00	0.234	0	1
divorced	0.01	0.00	0.080	0	1
single	0.02	0.00	0.140	0	1
engaged	0.00	0.00	0.042	0	1
children	0.16	0.00	0.366	0	1
revol_util(%)	51.13%	52.90%	27.38%	0.00%	119.00%
mthly_inc_log	3.69	3.70	0.25	0.00	5.77
loan_vol_log	1.58	1.54	0.39	0.00	2.58
credit_spread	11.20%	11.09%	3.54%	2.15%	24.64%
consumer_sentiment	71.25	72.90	6.93	55.30	90.40
per_capita_wage_log	4.39	4.38	0.11	4.00	5.06
N	31,550	31,550	31,550	31,550	31,550

Notes: Table 11 describes the control variables used in the analysis for the sample period June 2007 to April 2013. Dummy variables are used for the verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, other (reference), renewable_energy, small_business, vacation, wedding loan characteristics. The loan_amount_req_log is a log (Base 10) of the funding amount requested. FICO_avg is the average of a borrower high and low FICO score. Debt-to-income (DTI) is the ratio provided in the Lending Club dataset. Borrower demographic information is captured with dummy variables for home_own, married, divorced, single, engaged, and children. The revol_util variables captures the amount of revolving credit being used across all borrower accounts. The mthly_inc_log captures the log (Base 10) value of the borrower's monthly salary. The loan_vol_log captures the log (Base 10) volume of loans submitted on the Lending Club platform daily. The credit spread is the Lending Club provided interest rate minus the 2-yr treasury bond interest rate. The per_capita_wage_log captures the log (Base 10) per capita wages based on the 3-digit zip code of the borrower's home address.

Panel A of Table 12 demonstrates that $316,001,850 (92.7%) contained social disclosures and the remaining $24,917,575 (7.3%) did not contain social disclosures. The percentage of loans Charged Off with social disclosures is 14.9%, which is moderately below the sample average of 15.1%. However, for non-disclosure loans the Charge Off rate is 17.9%, which is 2.75% percentage points higher than the overall sample. the non-disclosure loans Charge Off rate slightly improves to 16.9%, however the overall percentage of loans issued that are ultimately Charged Off is lower at 14.8%. When observing the same break-down as a count in Table 14 Count of Disclosures and Funded Amount by Grade, the non-disclosure loans Charge Off rate slightly improves to 16.9%, however the overall percentage of loans issued that are ultimately Charged Off is lower at 14.8%. At first glance, the descriptive statistics indicate that there is a difference between loans that provide social disclosures and Charged Off loans have a higher dollar value on average. Table 12 and Table 13 also demonstrate the majority of loans issued on the platform are of the A, B, and C grade variety. Within each of the grades the Fully Paid versus Charged Off percentages do not deviate significantly between Charged Off and Fully Paid loans with the exception of E and F grades of non-disclosure loans which outperform social disclosure loans of the same grade. Within the E grade of stratification 19.5% of non-disclosure loans are Charged Off compared to E grade social disclosure loans 25.5% Charge Off rate. Similar results exist for F grade non-disclosure loans Charge Off rate of 25.6% while social disclosure F grade loans are Charged Off at 32.8%.

Table 12 Value of Disclosures and Funded Amount by Grade

Grade	Charged Off ($)	%	Fully Paid ($)	%	Total ($)	%
Panel A: SD	$47,105,325	14.9%	$268,896,525	85.1%	$316,001,850	92.7%
A	$4,876,900	6.3%	$72,505,800	93.7%	$77,382,700	22.7%
B	$14,311,950	12.9%	$96,248,975	87.1%	$110,560,925	32.4%
C	$12,647,200	18.6%	$55,391,875	81.4%	$68,039,075	20.0%
D	$9,581,675	23.8%	$30,664,750	76.2%	$40,246,425	11.8%
E	$3,409,150	25.8%	$9,808,025	74.2%	$13,217,175	3.9%
F	$1,480,850	34.7%	$2,784,900	65.3%	$4,265,750	1.3%
G	$797,600	34.8%	$1,492,200	65.2%	$2,289,800	0.7%
Panel B: ND	$4,465,025	17.9%	$20,452,550	82.1%	$24,917,575	7.3%
A	$251,975	5.1%	$4,737,050	94.9%	$4,989,025	1.5%
B	$1,467,275	16.9%	$7,196,750	83.1%	$8,664,025	2.5%[6]
C	$1,319,575	22.7%	$4,495,950	77.3%	$5,815,525	1.7%
D	$985,250	27.4%	$2,605,975	72.6%	$3,591,225	1.1%
E	$224,225	17.8%	$1,036,725	82.2%	$1,260,950	0.4%
F	$104,900	29.4%	$251,700	70.6%	$356,600	0.1%
G	$111,825	46.6%	$128,400	53.4%	$240,225	0.1%
Total	$51,570,350	15.1%	$289,349,075	84.9%	$340,919,425	100.0%

Notes: Table 12 Panel A describes loans categorized with Social Disclosure (SD) and Panel B Non-disclosure (ND) represents loans with fewer than 10 words in the loan description. The number of words was set to 10 to offset date and time text automatically appended to the beginning of each loan description by the Lending Club platform. The total dollar value and percentage of loans Charged Off and Fully Paid is provided at both the panel level and expressed as an overall total.

Table 13 Count of Disclosures and Funded Amount by Grade

Grade	Charged Off	%	Fully Paid	%	Total	%
Panel B: SD	4273	14.6%	24955	85.4%	29228	92.4%
A	495	6.3%	7325	93.7%	7820	24.7%
B	1277	13.0%	8531	87.0%	9808	31.0%
C	1177	18.5%	5184	81.5%	6361	20.1%
D	865	24.0%	2733	76.0%	3598	11.4%
E	286	25.5%	835	74.5%	1121	3.5%
F	111	32.8%	227	67.2%	338	1.1%
G	62	34.1%	120	65.9%	182	0.6%
Panel A: ND	408	16.9%	2003	83.1%	2411	7.6%
A	32	6.2%	483	93.8%	515	1.6%
B	135	16.7%	673	83.3%	808	2.6%
C	112	19.9%	450	80.1%	562	1.8%
D	86	25.2%	255	74.8%	341	1.1%
E	24	19.5%	99	80.5%	123	0.4%
F	10	25.6%	29	74.4%	39	0.1%
G	9	39.1%	14	60.9%	23	0.1%
Total	4681	14.8%	26958	85.2%	31639	100.0%

Notes: Table 13 Panel A describes loans categorized with Social Disclosure (SD) and Panel B Non-disclosure (ND) represents loans with fewer than 10 words in the loan description. The number of words was set to 10 to offset date and time text automatically appended to the beginning of each loan description by the Lending Club platform. The total count and percentage of loans Charged Off and Fully Paid is provided at both the panel level and expressed as an overall total.

Moving on to Table 15 and Table 16, that data demonstrates that credit card and debt consolidation are overwhelmingly the most listed purpose for using the Lending Club platform representing 66.7% of all loans issued in the sample. Again, non-disclosure loans have higher rates of being Charged Off in every category except Educational loans where 21.7% of social disclosure loans are Charged Off relative to 15.4% of non-disclosure loans. For both social disclosure and non-disclosure Small Business loans are the riskiest loan issued with over 27% of the loans being Charged Off. Furthermore, Small Business loans represent 4.4% of the total loan value funded but represent 8.4% of the Charged Off loan value.

Table 14 Count of Disclosures and Funded Amount by Loan Purpose

Loan Purpose	Charged Off	%	Fully Paid	%	Total	%
Panel A: SD	4,273	14.6%	24,955	85.4%	29,228	92.4%
Car	70	9.6%	662	90.4%	732	2.3%
Credit card	590	12.1%	4,276	87.9%	4,866	15.4%
Debt Consolidation	2,190	14.9%	12,490	85.1%	14,680	46.4%
Home Improvement	237	13.3%	1,550	86.7%	1,787	5.6%
House	39	14.4%	232	85.6%	271	0.9%
Major Purchase	118	9.2%	1,168	90.8%	1,286	4.1%
Medical	82	19.2%	344	80.8%	426	1.3%
Moving	52	14.5%	306	85.5%	358	1.1%
Other	416	16.4%	2,125	83.6%	2,541	8.0%
Renewable Energy	10	17.2%	48	82.8%	58	0.2%
Small Business	298	27.0%	804	73.0%	1,102	3.5%
Vacation	34	14.8%	195	85.2%	229	0.7%
Wedding	65	11.6%	495	88.4%	560	1.8%
Educational	72	21.7%	260	78.3%	332	1.0%
Panel B: ND	408	16.9%	2,003	83.1%	2,411	7.6%
Car	6	10.7%	50	89.3%	56	0.2%
Credit card	60	15.4%	330	84.6%	390	1.2%
Debt Consolidation	198	16.8%	979	83.2%	1,177	3.7%
Home Improvement	25	15.3%	138	84.7%	163	0.5%
House	2	9.5%	19	90.5%	21	0.1%
Major Purchase	14	14.6%	82	85.4%	96	0.3%
Medical	6	16.7%	30	83.3%	36	0.1%
Moving	5	25.0%	15	75.0%	20	0.1%
Other	50	20.1%	199	79.9%	249	0.8%
Renewable Energy	2	40.0%	3	60.0%	5	0.0%
Small Business	28	27.2%	75	72.8%	103	0.3%
Vacation	1	7.7%	12	92.3%	13	0.0%
Wedding	5	11.6%	38	88.4%	43	0.1%
Educational	6	15.4%	33	84.6%	39	0.1%
Total	4,681	14.8%	26,958	85.2%	31,639	100.0%

Notes: Table 14 Panel A and B describe the count of loans categorized by Loan Purpose and by Social Disclosure (SD) and Non-disclosure (ND). ND represents loans with fewer than 10 words in the loan description. The number of words was set to 10 to offset date and time text automatically appended to the beginning of each loan description by the Lending Club platform. The total count and percentage of loans Charged Off and Fully Paid is provided at both the panel level and expressed as an overall total.

Table 15 Social Disclosure versus Non-disclosure Paid and Charged Off by Purpose

Loan Purpose	Charged Off ($)	%	Fully Paid ($)	%	Total ($)	%
Panel A: SD	$47,105,325	14.9%	$268,896,525	85.1%	$316,001,850	92.7%
Car	$494,625	10.0%	$4,470,775	90.0%	$4,965,400	1.5%
Credit card	$7,396,675	13.1%	$49,149,250	86.9%	$56,545,925	16.6%
Debt Consolidation	$25,931,725	14.6%	$151,104,250	85.4%	$177,035,975	51.9%
Home Improvement	$2,479,275	13.5%	$15,866,350	86.5%	$18,345,625	5.4%
House	$404,400	12.8%	$2,756,675	87.2%	$3,161,075	0.9%
Major Purchase	$1,025,850	10.6%	$8,622,900	89.4%	$9,648,750	2.8%
Medical	$544,225	18.8%	$2,352,575	81.2%	$2,896,800	0.8%
Moving	$295,175	13.7%	$1,857,750	86.3%	$2,152,925	0.6%
Other	$3,117,425	16.7%	$15,530,050	83.3%	$18,647,475	5.5%
Renewable Energy	$102,825	21.2%	$382,150	78.8%	$484,975	0.1%
Small Business	$3,997,550	29.6%	$9,497,350	70.4%	$13,494,900	4.0%
Vacation	$185,425	14.7%	$1,078,475	85.3%	$1,263,900	0.4%
Wedding	$612,925	11.9%	$4,552,575	88.1%	$5,165,500	1.5%
Educational	$517,225	23.6%	$1,675,400	76.4%	$2,192,625	0.6%
Panel B: ND	$4,465,025	17.9%	$20,452,550	82.1%	$24,917,575	7.3%
Car	$59,400	15.9%	$313,075	84.1%	$372,475	0.1%
Credit card	$607,575	14.4%	$3,624,525	85.6%	$4,232,100	1.2%
Debt Consolidation	$2,444,400	17.9%	$11,227,425	82.1%	$13,671,825	4.0%
Home Improvement	$245,100	13.7%	$1,542,400	86.3%	$1,787,500	0.5%
House	$39,175	18.3%	$174,500	81.7%	$213,675	0.1%
Major Purchase	$155,300	21.7%	$559,125	78.3%	$714,425	0.2%
Medical	$26,700	12.3%	$190,425	87.7%	$217,125	0.1%
Moving	$24,000	17.5%	$113,225	82.5%	$137,225	0.0%
Other	$360,075	22.6%	$1,233,775	77.4%	$1,593,850	0.5%
Renewable Energy	$31,000	59.9%	$20,750	40.1%	$51,750	0.0%
Small Business	$362,125	28.3%	$918,450	71.7%	$1,280,575	0.4%
Vacation	$12,000	24.4%	$37,175	75.6%	$49,175	0.0%
Wedding	$51,925	15.1%	$293,000	84.9%	$344,925	0.1%
Educational	$46,250	18.4%	$204,700	81.6%	$250,950	0.1%
Total	$51,570,350	15.1%	$289,349,075	84.9%	$340,919,425	100.0%

Notes: Table 15 Panel A and B describe the dollar value of loans categorized by Loan Purpose and by Social Disclosure (SD) and Non-disclosure (ND). ND represents loans with fewer than 10 words in the loan description. The number of words was set to 10 to offset date and time text automatically appended to the beginning of each loan description by the Lending Club platform. The total count and percentage of loans Charged Off and Fully Paid is provided at both the panel level and expressed as an overall total.

Next, we observe the descriptive statistics for the dependent variables used in Table 16 Dependent Variable Descriptive Statistics. Clear differences between the Percent Invested and the Duration times exist between Fully Paid loans and Charged Off loans. Investors funded 93% of the Fully Paid loans, whereas less investment was made into loans that were eventually Charged Off. The Duration times for loans that are eventually Fully Paid are also lower at every loan grade compared to the Charged Off loans. The Percentage of Recovered Principal is obviously higher for Fully paid loans, and as expected the A grade Charged Off loans have a 5% higher recovered principal percentage relative to the next highest loan grade and 13% higher than the worst loan grade.

Table 16 Dependent Variable Descriptive Statistics

Grade	Tot_Inv (%)	Pct. Inv. (%)	Duration (hrs.)	Tot. Rec. Prncp (%)
Panel A: Charged Off	99%	88%	215.09	35%
A	99%	96%	202.59	42%
B	99%	92%	210.36	37%
C	99%	88%	206.99	35%
D	99%	87%	215.43	33%
E	99%	72%	255.74	33%
F	98%	62%	272.13	29%
G	98%	61%	269.88	34%
Panel B: Fully Paid	99%	93%	195.77	99%
A	99%	96%	191.95	99%
B	99%	95%	189.97	99%
C	99%	91%	195.70	99%
D	100%	89%	204.65	100%
E	99%	79%	233.28	99%
F	99%	72%	252.38	99%
G	98%	69%	252.86	98%
Total	99%	92%	198.63	90%

Notes: Table 16 Panel A describes the dependent variables for Charged Off loans by Loan Grade and Panel B: Describes the Fully Paid loans by Loan Grade. Percent Invested (Pct. Inv.) is the percentage of funding received from ordinary and institutional investors. Duration is the average number of hours between the Note Submission and Note Issue date. The Percentage of Recovered Principal (Tot. Rec. Prncp) is the average percentage of Principal repaid by the borrower.

In Table, 18, 19, 20, and 21 the content analysis variables demonstrate differences between Fully Paid and Charged Off loans. The Word Count on average is higher for Fully Paid loans, while the Misspellings are lower, Updates are higher and Questions Answered are higher. The Flesch Index readability score is consistent across both Charged Off and Fully paid loans and most loan grades. Interestingly, it appears higher risk borrowers (E, F, and G grades) in provide fewer Updates to their loan descriptions, which may be correlated to their higher words count. Lower grades also average about one additional lender question compared to lower risk loan grades. There is parity between the majority of content analysis independent variables with the exception of the hardworking identity claim, educational explanation, and other debt rate explanation. These variables are further reviewed in the regression analysis section.

Table 17 Independent Variables Social Disclosure Forms Averages

Grade	Word Count	Updates	QA Total	Misspellings	Flesch Index
Panel A: Charged Off	58.25	0.94	1.28	0.70	64.80
A	55.56	1.12	0.87	0.68	65.29
B	50.77	0.99	1.07	0.56	65.36
C	57.78	0.96	1.27	0.66	65.02
D	62.63	0.96	1.48	0.83	65.12
E	67.74	0.65	1.79	0.97	60.88
F	84.74	0.39	1.99	1.07	63.81
G	90.28	0.20	2.21	1.04	60.77
Panel B: Fully Paid	63.95	0.98	1.36	0.54	64.63
A	61.10	1.07	1.09	0.45	64.65
B	62.05	1.04	1.27	0.52	64.86
C	63.76	0.91	1.42	0.57	64.75
D	69.26	0.88	1.85	0.66	64.98
E	80.97	0.60	2.07	0.83	61.64
F	90.59	0.47	2.38	0.77	59.92
G	80.30	0.35	2.88	0.63	63.84
Total Average	63.10	0.97	1.35	0.56	64.65

Notes: Table 17 Panel A describes the text analysis independent variables for Charged Off loans by Loan Grade and Panel B: Describes the Fully Paid loans by Loan Grade. Word count is the average number of words in the loan description, misspellings are the average number of errors found using the Winnow-based spelling correction algorithm. Updates, Questions and Answers (QA Total) and Flesch Index are average values.

Table 18 Independent Variables Herzenstein et al. (2011) Identity Claims Frequency

Grade	Trustworthy	Successful	Economic Hardship	Hardworking	Moral	Religious
Panel A: Charge Off	22%	3%	6%	28%	18%	0%
A	21%	3%	5%	28%	17%	0%
B	17%	2%	4%	23%	13%	0%
C	18%	3%	5%	26%	16%	0%
D	21%	3%	6%	27%	16%	0%
E	23%	1%	6%	32%	21%	1%
F	25%	3%	7%	36%	19%	1%
G	27%	3%	10%	27%	25%	0%
Panel B: Fully Paid	25%	3%	7%	34%	19%	0%
A	23%	3%	5%	34%	17%	0%
B	21%	2%	5%	31%	16%	0%
C	22%	2%	5%	31%	16%	0%
D	23%	3%	6%	33%	18%	1%
E	27%	4%	7%	38%	21%	0%
F	32%	3%	9%	36%	20%	0%
G	30%	6%	10%	34%	23%	0%
Total Average	22%	3%	5%	32%	17%	0%

Notes:

Table 18 describes the frequency of the content analysis independent variables for Identity Claims and average HSCORE across the loan grades. Panel A describes Charged Off loans by Loan Grade and Panel B: Describes the Fully Paid loans by Loan Grade. HSCORE is average number Identity Claims used per loan description in the Sample. Trustworthy, Successful, Economic Hardship, Hardworking, Moral, and Religious identity claims are scored a 1 if used in the loan description and 0 otherwise.

Table 19 Independent Variables Lewis (2011) Qualifier Phrases Counts

Grade	Credit	No Credit	Less Credit	Loan	No Loan	Less Loan	Debt	No Debt	Less Debt
Panel A: Charge Off	40%	2%	10%	41%	3%	3%	24%	2%	9%
A	42%	3%	9%	42%	3%	3%	23%	2%	10%
B	42%	2%	9%	45%	2%	2%	26%	2%	9%
C	41%	3%	11%	44%	3%	3%	26%	2%	10%
D	42%	2%	13%	46%	3%	3%	28%	2%	10%
E	41%	3%	9%	46%	4%	4%	23%	2%	10%
F	44%	2%	7%	36%	2%	2%	22%	1%	4%
G	25%	1%	11%	31%	7%	7%	21%	1%	7%
Panel B: Fully Paid	40%	3%	10%	45%	3%	3%	25%	3%	9%
A	42%	3%	10%	48%	3%	3%	26%	2%	9%
B	41%	3%	11%	46%	3%	3%	27%	2%	10%
C	41%	3%	10%	46%	3%	3%	26%	3%	10%
D	41%	2%	11%	45%	3%	3%	25%	3%	10%
E	39%	3%	11%	45%	2%	2%	24%	3%	10%
F	41%	2%	7%	45%	3%	3%	25%	3%	5%
G	37%	4%	11%	38%	4%	4%	19%	3%	7%
Total	41%	3%	10%	46%	3%	3%	26%	2%	10%

Notes: Table 19 describes the frequency of keywords and qualifying phrases used in a loan description. Panel A describes Charged Off loans by Loan Grade and Panel B: Describes the Fully Paid loans by Loan Grade. Credit, No Credit, Less Credit, Loan, No Loan, Less Loan, Debt, No Debt, and Less Debt are scored a 1 if used in the loan description and 0 otherwise. Credit, Loan, and Debt combination are scored a 1 if any of the components are equal to 1.

Table 20 Michels (2012) Borrower Explanations

Grade	Education	Poor Credit	Monthly Expenses	Other Debt Rate	Amount of Other Debt
Panel A: Charged Off	2.23%	0.68%	2.18%	3.88%	0.71%
A	0.57%	0.19%	3.04%	4.55%	1.90%
B	1.35%	0.78%	1.98%	3.75%	0.99%
C	1.55%	0.93%	1.86%	4.65%	1.16%
D	2.10%	0.74%	3.26%	3.26%	1.47%
E	2.26%	1.61%	2.90%	3.23%	0.65%
F	3.31%	0.00%	1.65%	4.13%	0.00%
G	2.82%	0.00%	1.41%	4.23%	0.00%
Panel B: Fully Paid	3.61%	0.62%	3.04%	7.54%	1.33%
A	2.24%	0.47%	3.30%	7.42%	1.63%
B	2.68%	0.78%	3.80%	7.19%	1.92%
C	2.59%	0.94%	3.18%	7.10%	1.37%
D	2.51%	0.87%	3.38%	7.23%	1.54%
E	4.07%	1.28%	3.75%	8.57%	0.96%
F	8.20%	0.00%	2.34%	8.59%	0.39%
G	2.99%	0.00%	1.49%	6.72%	1.49%
Total	2.47%	0.75%	3.29%	6.81%	1.56%

Notes: Table 20 describes the frequency of borrower Explanations that meet the scoring criteria. Panel A describes Charged Off loans by Loan Grade and Panel B: Describes the Fully Paid loans by Loan Grade. MSCORE is the average number of borrower explanations used per loan description in the sample. Education, Poor Credit, Monthly Expenses, Other Debt Rate, and Amount of Other Debt are scored a 1 if used in the loan description and 0 otherwise.

Regressions

Ordinary Least Squares regression, Tobit, and Binomial logistic regression are used in this analysis to answer each hypothesis. For testing Duration for H_1 and H_5, ordinary least squares linear estimation is used to explain the distribution of the dependent variable against different models of independent variables. The OLS is to finds the set of weights for (a and b) that provide the best unbiased estimate for the Duration variable and provides minimum-variance mean-unbiased estimation. The OLS takes the form of:

$$\widetilde{Y_t} = a + bX_t + \widetilde{\varepsilon_\iota}$$

Where
$\widetilde{Y_t}$= dependent variable (percent funding by investors)
X_t= each independent variable term
$\widetilde{\varepsilon_\iota}$= error term that captures the difference between actuals and the predicted model

The Tobit model is used in this analysis for H_2, H_3, H_6, and H_7 in order to estimate linear relationships between variables due to censoring in the dependent variables. For example, the dependent variables Total Investment, Percent Investment, and Total Recovered Principal have a maximum value of 100%. The upper bound limitation of 100% indicates observations are being censored from above. The Tobit Model is designed to handle censoring from above and cases when the values are the maximum threshold of 100% (Tobin, 1958). In terms of P2P lending the true value might be equal to a number higher than the upper threshold if the Lending Club platform allowed for funding to exceed the borrower's requested amount. In order to examine the determinants of Percent Invested and Percentage of Recovered Principal variables while accounting for zero values, the following latent regression model is used.

$$y_t^* = x'_t\beta + \mu_t$$

The model assumes the latent variable is Y_t^* and that the variable linearly depends on a set of exogenous variables x'_t and β a vector, which determines the relationship between exogenous variables and the latent variable. The threshold value is set to T ($T=100\%$) because there is censoring from above. The model below can be used for Tobit model specifications and maximum likelihood techniques to create estimates for the censored variables.

$$Y_t \begin{cases} \dfrac{Y_t^*}{100\%} & \dfrac{\text{If } y_t^* < 100\%}{0 \text{ otherwise}} \end{cases}$$

Binomial Logistic Regression will be used to test the effect of signals on a loan being Charged Off or Fully Paid, for H_4 and H_8. Loan Status is the dependent variable and the sample is limited to only loans that have reached maturity. The dependent variable is the probability of the event to occur, and Loan Status of Fully Paid f_i is converted into a number between zero and one, the following transformation is used:

$$p_i = \frac{1}{1 + e^{-f_1}}$$

This study assumes that f_i is an unobserved continuous number representing the likelihood of a default. Therefore, higher f_i value is indicative of higher probability of full payment. Where p_i is the probability that full payment will occur. It is further assumed that n independent variables in the binary logistic regressions are linearly related to f_i. As previously referenced, the Emekter et al. (2015) study also uses binomial logistic regression to test the likelihood of default. In contrast to previous work, this study observes loans between June 2007 and April 2013 and removes assumptions that every loan that is late will be charged off and every loan that is current will always be fully paid. Instead, we limit our observations to only *ex post* 36-month term Fully Paid and

Charged Off loans in order to unambiguously understand the determinants of creditworthiness on peer-to-peer platforms. The methodology is designed to test the social disclosures Lenders perceive as credible then produce results that indicate if the Fully Paid loans decrease or increase in the same direction as the hypothesized sign. The standard logistic regression is used that follows the format below:

$$f_i = f\left(B_{0,}B_1X_i, B_2X_i, B_3X_i, B_4X_i, \dots, B_nX_i\right) + \mathcal{E}_i$$

Lastly, a test of mean difference between two populations of loan descriptions is performed between with loan with and without descriptions. Additionally, examination of *ex post* returns to show the difference in performance between loans will also be performed. For P2P lending the *ex post* returns are calculated as the total recovered principal plus interest, minus the principal paid by the lender. The summary of the findings from the multiple forms of analysis are then provided in the conclusion section of the dissertation.

Duration Results

In Table 21 the Duration dependent variable measures the number of hours required between loan submission date and loan issue date. In the first model for Duration, each of the independent variables were found to be significant and carry the sign consistent with hypothesis H_1. The Updates variable is the most significant form of social disclosure (Model 1: β = -6.083, SE = 0.708) across each of the models. Each Update provided by the borrower results in roughly a 6-hour reduction in the Duration time required to fund the loan. In addition, noteworthy when the social disclosure forms are combined with the content variables, Model 3, there is a slight increase in the size of the beta coefficients for Updates (Model 3: β = -6.107***, SE = 0.710). It is also important to note a sign change for Word Count between Model 1 and Model 3. When the word count is combined with content variables each additional word increases the time required to fully fund the loan (Model 3: β = 0.034***, SE = 0.012). Questions Answered and Flesch Scores are significant at the 10% level and are found to help reduce the Duration time required in both Model 1 and Model 3. Consistent with expectations, Misspellings carry the hypothesized positive sign and increases Duration time in both models. The beta coefficient for Misspellings (Model 1: β =1.672***, SE = 0.366), however becomes smaller when combined with the content explanatory factors (Model 3: β =1.583***, SE = 0.367) suggesting that specific content within the loan description may mitigate Misspelling errors to lenders. Interestingly, the forms of social disclosure within Model 1 have a positive interaction with content variables.

The results suggest the forms of social disclosure influence content variables more than content variables influence the forms of social disclosure. Identity claim coefficients in Model 2 increase in Model 3 when combined with each form of social disclosure. The following increases in beta coefficients are observed for each identity claim variable: Trustworthy (Model 2: β = -0.017*, SE = 1.658, Model 3: β = -0.699*, SE = 1.675), Successful (Model 2: β = -0.688*, SE = 3.358, Model 2: β = -2.543*, SE = 3.401), Hardworking (Model 2 β = -6.043***, SE = 1.277, Model 3 = -6.409***, SE = 1.323). The same results are found for all of the key words and qualifiers for Credit Combo (Model 2 β = -1.756*, SE = 1.194, Model 3 β = -2.142*, SE = 1.203), Loan Combo (Model 2: β =-3.13***, SE = 1.084, Model 3 = -3.428***, SE = 1.103), and Debt Combo (Model 2: β = -3.826***, SE = 1.195, Model 3 β= -4.018***, SE = 1.207). The

Duration is also reduced when loan explanations are combined with the different forms of social disclosure. The findings from Michels (2012) explanation variables indicate Education explanation reduces funding duration by 1 hours (Model 2: $\beta = -0.757^{*}$, SE = 3.386, Model 3 = -1.747*, SE = 3.401). Monthly Expense explanations are funded over one and half hour's sooners (Model 2: $\beta = -6.181^{**}$, SE = 3.031, Model 3 = β -7.883**, SE = 3.085) and Other Debt Rate and Amount of Other Debt explanations are funded roughly over 30 minutes faster respectively (Model 2: $\beta = -1.949^{*}$, SE = 2.179, Model 3 = $\beta = -2.681^{*}$, SE = 2.210, Model 2: $\beta = -1.281^{*}$, SE = 3.569, Model 3: $\beta = -1.757^{*}$, SE = 3.577). Notably, content disclosures related to a borrower's economic hardship, morals, religion, and poor credit explanation were not significant factors for lender investment in Model 2 and 3. Furthermore, the findings support H_5 that the specific content within social disclosures reduces the funding time. This implies social disclosures reduce the time required for lenders to make investment decisions.

Total Invested Results

In Table 22 the Total Invested dependent variable measures both the retail and institutional investors as well as the Lending Club subsidiaries investment in a loan. The Total Invested model is the most important variable for borrowers seeking loans on the Lending Club platform. Each form of social disclosure proved to be significant across each model, but coefficients did not always carry the hypothesized sign. The Word Count (Model 1: $\beta = -0.001^{**}$, SE = 0.00), Updates (Model 1: $\beta = -0.014^{*}$, SE = 0.018), and Flesch Index (Model 1: $\beta = -0.001^{*}$, SE = 0.001) variables had an unexpected negative relationship with Total Invested that is inconsistent with H_2. The Misspellings, as expected, has a negative relationship (Model 1: $\beta = -0.12^*$, SE = 0.00). The Questions Answered (Model 1: $\beta = 0.029^{***}$, SE = 0.10) variable is the only social disclosure that increased Total Investment by 2.9% for each question answered by borrowers. These results may be attributable to Lending Club subsidiaries funding loans that were less attractive to retail and institutional investors and would not be issued unless the subsidiary invested. Similar findings for Total Investment are also seen within the content variables.

Identity Claim variables for Trustworthy, Successful, Hardworking, and Religious also have a negative relationship with the Total Invested variable. These identity claims each resulted in less Total Investment with Religious identity claims reducing Total Invested by 22.7% per claim. The only positive relationship amongst the identity claim variables is the Moral identity claim, which increase the Total Invested by 6.3% per claim. Additional findings demonstrate keywords and qualifiers have negative relationships with Credit and Loan variables. Contrary to H_6, the mention of credit and loan phrases reduce investment between 1.2% and 4.1% across the respective models. The only positive relationship found is the Debt keyword and qualifier (Model 2: $\beta = 0.03^*$, SE = 0.033, Model 3: $\beta = 0.035$, SE = 0.34). Lastly, the explanation variables for Other Debt Rate and Amount of Other Debt proved to have a negative relationship with the Total Invested variable for both Model 2 and Model 3. The findings for Total Investment are in stark contrast to the findings for Percent Invested.

<u>Percent Invested Results</u>

In Table 23 the Percent Invested dependent variable table demonstrates the amount of investment provided by only retail and institutional investors. The results suggest that different forms of social disclosures and the specific language used within the disclosure influence retail and institutional investment. Hypothesis H_2 proved true for Word Count, Updates, Questions Answered and Flesch Index. These variables were significant across Model 1 and Model 3. As expected, the only social disclosure not found to be significant or carry the hypothesized sign for Percent Invested was the Misspellings variable. The most notable variables include Updates and Questions Answered that have large beta coefficients that indicate that for each Update, there is an 11% increase in the Percent Invested and each Question Answered increases the Percent Invested between 0.5% and 0.6% between each respective model. For each unit increase in Word Count there is a .004% change in Percent Invested in Model 1. The results also indicate that a 1-unit increase in Flesch Index score equates to .1% increase in the funding amount invested by retail and institutional investors in Model 1 and Model 3. The findings also show that the retail and institutional lenders change the amount of investment based on the content provided in the loan description.

The results suggest that removing Lending Club Subsidiary funding and limiting investment to retail and institutional changes the sign and significance of content variables. The identity claim analysis indicates a 3.7% to 4.9% increase in retail and institutional lender investment based on the hardworking identity claim (Model 2: β = 0.049***, SE = 0.004, (Model 3: β = 0.037*, SE = 0.004). This is a significant change from the Total Invested models where Hardworking has a negative relationship with Total Investment. However, claims for Trustworthy, Successful, Economic Hardship, and Moral maintain a negative relationship with Percent Invested in Model 2. Interestingly, Model 3 findings support for H_6, while Model 2 is not strongly supported. When identity claims and the forms of social disclosures are combined in Model 3 Trustworthy and Successful identity claims change signs and demonstrate a positive relationship with the Percent Invested variable. Religious claims also increase substantially from .3% in Model 2 to 2.5% in Model 3. The keyword and qualifier analysis also have mixed results. Combinations of Credit keywords and qualifiers have a negative relationship with Percent Invested, while combinations of loan and debt keywords and qualifiers have a positive influence on the Percent Invested in both Model 2 and Model 3. The explanations provided by borrowers also had unexpected results inconsistent with H_6. The Poor Credit explanation variable is significant at the 1% level and had a positive relationship with the Percent Invested dependent variable (Model 2: β = 0.014***, SE = 0.023, Model 3: β = 0.069***, SE = 0.022). A Poor Credit explanation increased retail and intuitional investment by 1.4% and 6.9% respectively across each model. Counterintuitively, providing an Educational or Other Interest Rate explanation reduced retail and institutional investment. In Model 2 and Model 3 Education explanations reduced funding by 3.2% and .3% respectively (Model 2: β = -0.032***, SE = 0.011, Model 3: β = -0.003***, SE = 0.003). In addition, obtaining lower interest rate explanations reduced Percent Invested by 3.1% in Model 2 and 2.8% in Model 3. The results also indicate that providing a breakdown of monthly expenses and other debt obligations increase lender investment. Loans with Monthly Expense explanations received 3.2% and 3.4% more lender investment and the Other Amount of Debt Explanations received 8.7% and 7.5% more investment in Model 2 and 3 respectively. In the following analysis, the findings related to funding a loan in H_1, H_2, H_5, and H_6 are analyzed from a loan repayment lens to test Total Recovered Principal and Loan status for H_3, H_4, H_7, and $H_{8.}$

Total Recovered Principal Results

In Table 24 the Total Recovered Principal returned to lenders from borrowers is examined. The Total Recovered Principal is arguably the most important dependent variable to lenders on the Lending Club platform. Understanding the factors associated with creditworthiness makes a direct and tangible impact on the lenders portfolio performance. The results indicate H_3 is true for Word Count and that for every additional word used in the loan description there is a .04% increase in Model 1 and .02% increase in Model 3 for Total Recovered Principal. In other words, a loan description of 100 words is expected to increase the Total Recovered Principal by 4.0% and 2.0% respectively ceteris paribus. H_3 is also true for the Updates variable and is positively related to the Total Recovered Principal at the 1% significance level. The results indicate that for each additional Update there is a negative 5% increase in Total Recovered Principal. This finding might suggest higher quality borrowers are more likely to make subsequent updates to their original loan description. These results might also suggest that poorer borrowers update more often because they know their true creditworthiness and feel more justification is required. The Questions Answered variable also demonstrates a positive relationship with Total Recovered Principal as hypothesized in H_3 (Model 1: $\beta = 0.026^{***}$, SE = 0.004, Model 3: $\beta = 0.025^{***}$, SE = 0.004). Each time that a borrower answers a question raised by a lender there is a 2.6% increase in Model 1 and 2.5% increase in Model 3 for Total Recovered Principal. Misspellings are consistent with the hypothesized sign for H_3 (Model 1: β = -0.03***, SE = 0.004, Model 3: β = -0.027***, SE = 0.004). Each Misspelling in the loan description is associated with a 3.0% and 2.7% reduction in the Total Recovered Principal in Model 1 and 3 respectively. The Flesch Index score was also significant across both of the models suggesting a well-written loan description is an indicator of credit worthiness and loan repayment. Content within the loan descriptions also proved to be significant.

Consistent with Hypothesis H_7, the specific content within the loan disclosure is found to be a predictor of loan performance. Beginning with identity claims, we observe that the Hardworking and Trustworthy claims increases the Total Recovered Principal between in both Model 2 and Model 3. Borrowers identifying as Trustworthy repaid 2.1% more principal in Model 1 and 2.3% more in Model 3. The Hardworking variable made the most significant impact increasing the Total Recovered Principal by 9% and 8% (Model 2: β = 0.09***, SE = 0.016, Model 3: β = 0.088***, SE = 0.017). Equally important, several identity claims result in a reduction of the Total Recovered Principal. Successful, Economic Hardship, Moral and Religious are all negatively related to recovering the loan principal. Religious claims reduce the recovered principal the most of the identity claims (Model 2: β = -0.08*, SE = 0.126, Model 3: β = -0.085*, SE = 0.126) followed by claims of being Successful (Model 2: β = 0.066*, SE = 0.042, Model 3: β = -0.055***, SE = 0.042). All of the Lewis (2011) keyword and qualifier variables proved to be significant at the 1% level in both Model 2 and Model 3 with debt keywords and qualifiers increasing the recovered principal by 9%. The most significant findings across all of the content variables are the Michels (2012) explanation variables. The results show having demonstrated successful completion of education increase the Total Recovered

Principal 18.3% in Model 2 and 19.2% in Model 3. Additionally, borrowers that provided the other rates of interest repaid approximately 10% more principal. These findings demonstrate a clear difference between Total Invested, Percent Invested, and Total Recovered Principal. The content variables significant for a lender to fund a loan are inconsistent with borrower repayment of the loan. The explanations detrimental to lenders are Poor Credit explanations and borrowers listing their other debt obligations. Interestingly, Poor Credit explanation borrowers pay 7.3% less principal in Model 2 and 4.7% less principal in Model 3. This suggest that the interaction between Poor Credit explanations different forms of social disclosure result in 2.6% more principal being recovered. In the last regression analysis, the Total Recovered Principal is further examined in terms of total loan repayment and default.

Loan Status Results

The Lending Club platform has two terminal statuses for loans that have reached maturity, "Charged Off" and "Fully Paid". In Table 25 the binomial logistic regression results are provided that indicate the explanatory variables significant for "Fully Paid" loans. The Loan Status results indicate that the most significant social disclosure indicator is the Questions Answered variable. Consistent with H_8, for each question answered the odds of borrower fully repaying the loan increase 3.8% in Model 1 and 3.6% in Model 3. While for each Misspellings the odds of fully paying the loan decrease by 6.8% in Model 1 and 6.1% in Model 3. Noteworthy, for each unit increase in Word Count there is a .2% increase in likelihood of repayment in Model 1. This increase based on Word Count reduces to zero when interaction with content variables are added in Model 3. In terms of identity claims, the results indicated that the only claim significant for “Fully Paid” loans is Hardworking (Model 2: β = 0.211***, SE = 0.042, Model 3: β = 0.21***, SE = 0.044). While the only claim significant for “Charged Off” loans in Moral (Model 2: β = -0.165***, SE = 0.057, Model 3 = -0.155***, SE = 0.058). Each of these claims are significant at the 1% level. The Hardworking claim increases the likelihood of full loan repayment by approximately 21%, while the Moral claim reduced the likelihood of loan repayment between 16.5% and 15.5% in Model 2 and Model 3. All of the keyword and qualifier variables were significant at the 1% level and positively related to the Loan Status of “Fully Paid” across both models. The most significant findings were in the loan explanations that demonstrate Education and Interest Rate on Other Debt separate good and bad borrowers. Educational explanations increased the likelihood of a “Fully Paid” loan status by approximately 40% (Model 2: β = -0.418***, SE = 0.127, Model 3 = 0.4***, SE = 0.128). Similarly, the Interest Rate on Other Debt increased the likelihood of full loan repayment by roughly 38% (Model 2: 0.379***, SE = 0.083, Model 3 = 0.384***, SE = 0.084). These findings not only support H_8, but also indicate the ability of social disclosures to signal borrower creditworthiness.

Table 21 Duration Analysis Results

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
(Constant)	-66.34**	30.094	-69.146**	29.640	-64.906**	30.102
word_count	-0.008*	0.009	-	-	0.034***	0.012
updates	-6.083***	0.708	-	-	-6.107***	0.710
qa_total	-0.371*	0.313	-	-	-0.308*	0.314
misspellings	1.672***	0.366	-	-	1.583***	0.367
flesch_index	0.003*	0.029	-	-	0.005*	0.030
trustworthy	-	-	-0.017*	1.658	-0.699*	1.675
successful	-	-	-0.688*	3.358	-2.543*	3.401
economic_hardship	-	-	2.382	2.426	0.335	2.492
hardworking	-	-	-6.043***	1.277	-6.409***	1.323
moral	-	-	2.561	1.759	1.779	1.774
religious	-	-	10.022	10.440	8.017	10.439
credit_combo	-	-	-1.756*	1.194	-2.142*	1.203
loan_combo	-	-	-3.13***	1.084	-3.428***	1.103
debt_combo	-	-	-3.826***	1.195	-4.018***	1.207
edu_exp	-	-	-0.757*	3.386	-1.747*	3.401
poor_credit_exp	-	-	7.254	6.052	4.495	6.057
mthly_expense_exp	-	-	-6.181**	3.031	-7.883**	3.085
oth_rate_exp	-	-	-1.949*	2.179	-2.681*	2.210
amnt_oth_debt_exp	-	-	-1.281*	3.569	-1.757*	3.577
verified_status	-23.131***	1.115	-22.665***	1.116	-23.117***	1.115
car	-4.134*	3.753	-4.165*	3.756	-4.617*	3.753
credit_card	-4.525**	2.267	-2.371*	2.383	-2.757*	2.382
debt_consolidation	-8.473***	2.009	-6.226***	2.092	-6.466***	2.091
educational	9.27	5.169	11.66**	5.178	9.506	5.182
home_improvement	-2.428*	2.800	-3.192*	2.804	-3.635*	2.804
house	6.593	5.719	6.308	5.729	5.146	5.723
major_purchase	-4.487*	3.066	-5.109*	3.068	-5.097*	3.065
medical	4.069	4.662	3.57	4.664	3.355	4.662
moving	0.258	5.088	0.333	5.094	0.083*	5.089
renewable_energy	-1.501*	11.772	-3.689*	11.780	-2.627*	11.767
small_business	23.7***	3.282	23.976***	3.264	21.603***	3.305
vacation	1.734	6.244	0.028*	6.249	0.606	6.242
wedding	-2.912*	4.191	-2.827*	4.196	-3.604*	4.192
loan_amount_req_log	117.782***	2.226	118.152***	2.221	118.547***	2.237
FICO_avg	-0.078***	0.025	-0.07***	0.025	-0.078***	0.025
dti	0.19**	0.079	0.179**	0.079	0.178**	0.079
home_own	2.026	1.144	2.622**	1.147	2.59**	1.147
married	-0.686*	2.347	1.064	2.317	-0.282*	2.348

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
divorced	16.955***	6.533	17.425***	6.534	16.056**	6.534
single	-2.409*	3.761	-1.159*	3.752	-2.069*	3.760
engaged	-1.157*	12.505	1.262	12.499	-1.123*	12.498
children	2.226	1.529	3.549**	1.496	2.331	1.532
revol_util(%)	-0.003*	0.024	-0.002*	0.024	0.001*	0.024
mthly_inc_log	-11.518***	2.618	-11.925***	2.613	-11.765***	2.622
loan_vol_log	33.922***	1.862	33.05***	1.856	33.527***	1.864
credit_spread	-1.606***	0.246	-1.499***	0.239	-1.718***	0.247
consumer_sentiment	-0.967***	0.085	-1.096***	0.084	-0.998***	0.085
per_capita_wage_log	-22.662***	5.038	-22.181***	5.040	-21.735***	5.036
R Square	0.239		0.238		0.240	
Adjusted R Square	0.238		0.237		0.239	
F statistic	223.94		280.48		203.51	
Number of Observations	31549		31550		31549	

Notes: Table 21 provides the results of the ordinary least squares regression for dependent variable Duration using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

Table 22 Total Invested Analysis Results

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
(Intercept):1	-0.116*	0.838	-0.58*	0.795	-0.083*	0.837
(Intercept):2	-0.211***	0.011	-0.239***	0.011	-0.218***	0.012
word_count	-0.0005**	0.0002	-	-	0.0004*	0.0003
updates	-0.014*	0.018	-	-	-0.01*	0.018
qa_total	0.029***	0.010	-	-	0.03***	0.010
misspellings	-0.012*	0.009	-	-	-0.011*	0.009
flesch_index	-0.0007*	0.0008	-	-	-0.0007*	0.0008
trustworthy	-	-	-0.072*	0.041	-0.065*	0.043
successful	-	-	-0.046*	0.076	-0.025*	0.080
economic_hardship	-	-	-0.04*	0.059	-0.009*	0.063
hardworking	-	-	-0.063*	0.033	-0.046*	0.036
moral	-	-	0.063*	0.044	0.077*	0.046
religious	-	-	-0.227*	0.217	-0.238*	0.222
credit_combo	-	-	-0.041*	0.034	-0.034*	0.035
loan_combo	-	-	-0.023*	0.030	-0.012*	0.031
debt_combo	-	-	0.03*	0.033	0.035*	0.034
edu_exp	-	-	0.122	0.093	0.136	0.097
poor_credit_exp	-	-	0.504	0.348	0.522	0.355
mthly_expense_exp	-	-	0.116	0.083	0.157	0.088
oth_rate_exp	-	-	-0.037*	0.052	-0.019*	0.054
amnt_oth_debt_exp	-	-	-0.104*	0.089	-0.089*	0.091
verified_status	0.033*	0.032	0.026*	0.031	0.031*	0.032
car	0.068*	0.106	0.071*	0.103	0.073*	0.105
credit_card	0.113	0.063	0.128**	0.065	0.125	0.067
debt_consolidation	0.09*	0.055	0.1*	0.056	0.095*	0.058
educational	0.059*	0.135	0.061*	0.131	0.059*	0.135
home_improvement	0.032*	0.075	0.029*	0.073	0.035*	0.075
house	0.25	0.174	0.233	0.169	0.236	0.173
major_purchase	-0.019*	0.085	-0.02*	0.082	-0.016*	0.085
medical	-0.061*	0.132	-0.039*	0.128	-0.044*	0.132
moving	0.008*	0.158	0.023*	0.154	0.021*	0.158
renewable_energy	0.177	0.382	0.147	0.366	0.163	0.380
small_business	0.083*	0.086	0.071*	0.083	0.095*	0.086
vacation	0.093*	0.224	0.081*	0.217	0.096*	0.224
wedding	0.058*	0.112	0.069*	0.109	0.068*	0.112
loan_amount_req_log	-1.519***	0.084	-1.488***	0.082	-1.52***	0.085
FICO_avg	0.006***	0.001	0.006***	0.001	0.006***	0.001
dti	-0.002*	0.002	-0.002*	0.002	-0.002*	0.002
home_own	-0.014*	0.033	-0.012*	0.032	-0.013*	0.033

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
married	-0.13**	0.053	-0.14***	0.051	-0.13**	0.053
divorced	-0.213*	0.128	-0.248**	0.123	-0.232*	0.127
single	0.062*	0.100	0.066*	0.097	0.079*	0.101
engaged	0.174	0.363	0.165	0.349	0.174	0.359
children	-0.055*	0.040	-0.068*	0.038	-0.054*	0.040
revol_util	-0.107*	0.072	-0.101*	0.070	-0.1*	0.072
mthly_inc_log	0.089*	0.076	0.11	0.074	0.09*	0.076
loan_vol_log	0.854***	0.046	0.832***	0.043	0.85***	0.046
credit_spread	15.326***	0.912	15.617***	0.849	15.18***	0.912
consumer_sentiment	0.022***	0.002	0.021***	0.002	0.021***	0.002
per_capita_wage_log	0.0719	0.141	0.061	0.138	0.064	0.141
R Square	0.031		0.03		0.031	
Adjusted R Square	0.03		0.029		0.03	
Number of Observations	31549		31550		31549	

Notes: Table 22 provides the results of the Tobit regression for dependent variable Total Invested using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

Table 23 Percent Invested Analysis Results

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
(Intercept):1	-1.181***	0.099	-1.378***	0.102	-1.172***	0.098
(Intercept):2	-1.389***	0.006	-1.331***	0.006	-1.394***	0.006
word_count	0.0002***	0.000003	-	-	0.0004***	0.000003
updates	0.11***	0.003	-	-	0.11***	0.003
qa_total	0.006***	0.001	-	-	0.005***	0.001
misspellings	-0.007***	0.001	-	-	-0.007***	0.001
flesch_index	0.0006***	0.00001	-	-	0.0007***	0.00001
trustworthy	-	-	-0.002*	0.006	0.006*	0.005
successful	-	-	-0.015*	0.011	0.012*	0.011
economic_hardship	-	-	-0.019**	0.008	-0.001*	0.008
hardworking	-	-	0.049***	0.004	0.037***	0.004
moral	-	-	-0.005*	0.006	-0.001*	0.006
religious	-	-	0.003*	0.035	0.025*	0.034
credit_combo	-	-	-0.006*	0.004	-0.003*	0.004
loan_combo	-	-	0.006*	0.004	0.004*	0.004
debt_combo	-	-	0.001*	0.004	0.006*	0.004
edu_exp	-	-	-0.031***	0.011	-0.003*	0.011
poor_credit_exp	-	-	0.014*	0.023	0.069***	0.022
mthly_expense_exp	-	-	0.032***	0.011	0.034***	0.010
oth_rate_exp	-	-	-0.031***	0.007	-0.028***	0.007
amnt_oth_debt_exp	-	-	0.087***	0.013	0.075***	0.013
verified_status	-0.07***	0.004	-0.092***	0.004	-0.069***	0.004
car	0.08***	0.013	0.084***	0.013	0.08***	0.013
credit_card	0.06***	0.007	0.065***	0.008	0.065***	0.008
debt_consolidation	0.06***	0.007	0.064***	0.007	0.06***	0.007
educational	-0.028*	0.015	-0.079***	0.016	-0.029*	0.015
home_improvement	0.043***	0.009	0.044***	0.010	0.045***	0.009
house	0.043**	0.019	0.036*	0.020	0.046**	0.019
major_purchase	0.04***	0.010	0.058***	0.011	0.041***	0.010
medical	0.025*	0.016	0.018*	0.016	0.025*	0.016
moving	0.037**	0.017	0.031*	0.018	0.033*	0.017
renewable_energy	0.052*	0.042	0.099**	0.045	0.056*	0.042
small_business	-0.034***	0.010	-0.067***	0.011	-0.029***	0.010
vacation	0.046**	0.023	0.072***	0.024	0.049**	0.023
wedding	0.035***	0.013	0.032**	0.014	0.037***	0.013
loan_amount_req_log	-0.41***	0.008	-0.415***	0.008	-0.409***	0.008
FICO_avg	0.003***	0.000	0.003***	0.000	0.003***	0.000
dti	0*	0.000	0.001**	0.000	0*	0.000
home_own	-0.001*	0.004	-0.005*	0.004	-0.002*	0.004

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
married	-0.017**	0.007	-0.036***	0.008	-0.019***	0.007
divorced	-0.025*	0.020	-0.039*	0.021	-0.022*	0.020
single	0.013*	0.012	0.006*	0.013	0.013*	0.012
engaged	-0.094**	0.037	-0.141***	0.039	-0.093**	0.037
children	-0.009*	0.005	-0.025***	0.005	-0.01*	0.005
revol_util	0.028***	0.008	0.038***	0.009	0.026***	0.008
mthly_inc_log	0.052***	0.009	0.051***	0.009	0.054***	0.009
loan_vol_log	0.382***	0.006	0.458***	0.006	0.379***	0.006
credit_spread	3.649***	0.089	3.628***	0.089	3.654***	0.088
consumer_sentiment	0.005***	0.000	0.007***	0.000	0.005***	0.000
per_capita_wage_log	-0.013	0.017	-0.0232	0.018	-0.015	0.017
R Square	0.34		0.292		0.349	
Adjusted R Square	0.339		0.291		0.344	
Number of Observations	31549		31550		31549	

Notes: Table 23 provides the results of the Tobit regression for dependent variable Percent Invested using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

Table 24 Total Recovered Principle Analysis Results

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
(Intercept):1	0.048*	0.376	-0.421*	0.370	0.1	0.375
(Intercept):2	-0.182***	0.006	-0.183***	0.006	-0.187***	0.006
word_count	0.0004***	0.0001	-	-	0.0002*	0.0001
updates	0.05***	0.009	-	-	0.05***	0.009
qa_total	0.026***	0.004	-	-	0.025***	0.004
misspellings	-0.03***	0.004	-	-	-0.027***	0.004
flesch_index	0.0004*	0.0004	-	-	0.0004*	0.0004
trustworthy	-	-	0.021*	0.021	0.023*	0.022
successful	-	-	-0.066*	0.042	-0.055*	0.042
economic_hardship	-	-	-0.032*	0.031	-0.012*	0.032
hardworking	-	-	0.09***	0.016	0.088***	0.017
moral	-	-	-0.053**	0.022	-0.045**	0.023
religious	-	-	-0.08*	0.126	-0.085*	0.126
credit_combo	-	-	0.048***	0.015	0.052***	0.015
loan_combo	-	-	0.048***	0.014	0.049***	0.014
debt_combo	-	-	0.091***	0.015	0.09***	0.015
edu_exp	-	-	0.183***	0.047	0.192***	0.047
poor_credit_exp	-	-	-0.073*	0.076	-0.047*	0.076
mthly_expense_exp	-	-	0.061*	0.041	0.08*	0.042
oth_rate_exp	-	-	0.101***	0.029	0.109***	0.030
amnt_oth_debt_exp	-	-	-0.009*	0.048	-0.014*	0.048
verified_status	-0.012*	0.014	-0.023*	0.014	-0.011*	0.014
car	0.19***	0.052	0.192***	0.052	0.197***	0.052
credit_card	0.204***	0.029	0.151***	0.030	0.153***	0.030
debt_consolidation	0.134***	0.025	0.085***	0.026	0.082***	0.026
educational	-0.032*	0.062	-0.062*	0.062	-0.047*	0.062
home_improvement	0.025*	0.036	0.045*	0.036	0.047*	0.036
house	0.085*	0.074	0.1	0.073	0.11	0.073
major_purchase	0.188***	0.042	0.203***	0.042	0.201***	0.042
medical	-0.088*	0.057	-0.068*	0.057	-0.069*	0.057
moving	0.018*	0.064	0.022*	0.064	0.027*	0.064
renewable_energy	-0.14*	0.141	-0.095*	0.141	-0.119*	0.141
small_business	-0.258***	0.038	-0.233***	0.038	-0.211***	0.038
vacation	-0.01*	0.080	0.003*	0.080	0.004*	0.080
wedding	0.095*	0.054	0.102	0.054	0.108**	0.054
loan_amount_req_log	-0.448***	0.029	-0.453***	0.029	-0.466***	0.029
FICO_avg	0.002***	0.000	0.002***	0.000	0.002***	0.000
dti	-0.005***	0.001	-0.006***	0.001	-0.005***	0.001
home_own	-0.017*	0.014	-0.029**	0.014	-0.029**	0.015

	Model 1		Model 2		Model 3	
	Beta	**Std. Error**	**Beta**	**Std. Error**	**Beta**	**Std. Error**
married	-0.002*	0.030	-0.015*	0.029	-0.008*	0.030
divorced	-0.122*	0.079	-0.118*	0.079	-0.109*	0.079
single	0.067*	0.049	0.055*	0.049	0.06*	0.049
engaged	0.073*	0.169	0.065*	0.169	0.079*	0.169
children	-0.06***	0.019	-0.071***	0.019	-0.062***	0.019
revol_util	-0.029*	0.031	-0.031*	0.031	-0.039*	0.031
mthly_inc_log	0.406***	0.034	0.431***	0.034	0.41***	0.034
loan_vol_log	0.236***	0.020	0.259***	0.020	0.239***	0.020
credit_spread	-3.858***	0.295	-3.338***	0.288	-3.717***	0.295
consumer_sentiment	-0.002*	0.001	-0.001*	0.001	-0.001*	0.001
per_capita_wage_log	0.163	0.064	0.163	0.064	0.145	0.064
R Square	0.048		0.05		0.052	
Adjusted R Square	0.047		0.049		0.05	
Number of Observations	31549		31550		31549	

Notes: Table 24 provides the results of the Tobit regression for dependent variable Total Recovered Principal using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

Table 25 Loan Status Analysis Results

	Model 1		Model 2		Model 3	
	Exp(B)	**Std. Error**	**Exp(B)**	**Std. Error**	**Exp(B)**	**Std. Error**
Constant	-5.04***	0.947	-5.743***	0.937	-4.925***	0.950
word_count	0.002***	0.000	-	-	0	0.000
updates	0.035	0.022	-	-	0.036	0.022
qa_total	0.038***	0.010	-	-	0.036***	0.010
misspellings	-0.068***	0.010	-	-	-0.061***	0.010
flesch_index	-0.001	0.001	-	-	-0.001	0.001
trustworthy	-	-	0.074	0.055	0.072	0.055
successful	-	-	-0.092	0.107	-0.111	0.109
economic_hardship	-	-	-0.112	0.076	-0.093	0.079
hardworking	-	-	0.211***	0.042	0.21***	0.044
moral	-	-	-0.165***	0.057	-0.155***	0.058
religious	-	-	-0.219	0.306	-0.231	0.307
credit_combo	-	-	0.152***	0.037	0.154***	0.038
loan_combo	-	-	0.144***	0.034	0.143***	0.035
debt_combo	-	-	0.245***	0.038	0.233***	0.039
edu_exp	-	-	0.418***	0.127	0.4***	0.128
poor_credit_exp	-	-	-0.136	0.187	-0.105	0.189
mthly_expense_exp	-	-	0.103	0.108	0.13	0.111
oth_rate_exp	-	-	0.379***	0.083	0.384***	0.084
amnt_oth_debt_exp	-	-	-0.051	0.126	-0.036	0.127
verified_status	-0.034	0.035	-0.044	0.034	-0.035	0.035
car	0.413***	0.135	0.432***	0.135	0.439***	0.136
credit_card	0.526***	0.071	0.383***	0.075	0.381***	0.075
debt_consolidation	0.346***	0.061	0.212***	0.063	0.206***	0.064
educational	-0.272*	0.143	-0.283**	0.144	-0.292**	0.144
home_improvement	0.042	0.089	0.111	0.089	0.106	0.089
house	0.12	0.184	0.181	0.184	0.191	0.184
major_purchase	0.426***	0.108	0.463***	0.108	0.468***	0.109
medical	-0.199	0.135	-0.133	0.135	-0.147	0.135
moving	0.124	0.159	0.15	0.159	0.155	0.160
renewable_energy	-0.205	0.344	-0.123	0.342	-0.15	0.343
small_business	-0.72***	0.090	-0.607***	0.089	-0.601***	0.091
vacation	0.029	0.197	0.069	0.197	0.078	0.198
wedding	0.333**	0.141	0.367***	0.142	0.371***	0.142
loan_amount_req_log	-0.592***	0.072	-0.615***	0.071	-0.635***	0.072
FICO_avg	0.007***	0.001	0.008***	0.001	0.007***	0.001
dti	-0.015***	0.002	-0.015***	0.002	-0.015***	0.002
home_own	-0.009	0.036	-0.035	0.036	-0.036	0.036

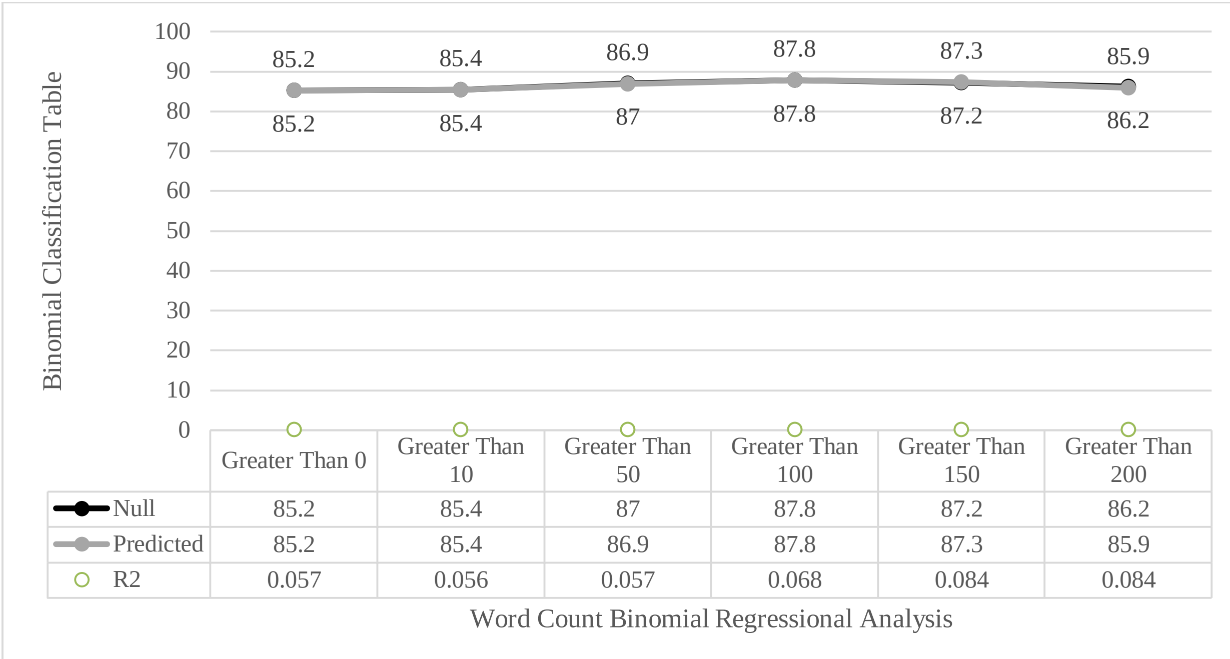
	Model 1		Model 2		Model 3	
	Exp(B)	**Std. Error**	**Exp(B)**	**Std. Error**	**Exp(B)**	**Std. Error**
married	0.057	0.079	0.057	0.078	0.049	0.079
divorced	-0.152	0.208	-0.12	0.208	-0.123	0.209
single	0.11	0.125	0.094	0.125	0.095	0.126
engaged	0.25	0.481	0.23	0.477	0.208	0.478
children	-0.094*	0.048	-0.102**	0.047	-0.095**	0.049
revol_util(%)	-0.001	0.001	-0.001	0.001	-0.001	0.001
mthly_inc_log	1.027***	0.085	1.081***	0.085	1.031***	0.086
loan_vol_log	0.183***	0.050	0.197***	0.049	0.195***	0.050
consumer_sentiment	-0.012***	0.003	-0.011***	0.003	-0.011***	0.003
per_capita_wage_log	0.453***	0.161	0.428***	0.161	0.408**	0.161
Cox & Snell R Square	0.059		0.059		0.061	
Nagelkerke R Square	0.104		0.104		0.107	
Number of Observations	31550		31549		31549	

Notes: Table 25 provides the results of the binomial logistic regression for dependent variable Total Recovered Principal using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

Robustness Testing

Additional fine-tuning of the model is performed by observing inflection points in the data for each form of social disclosure using binomial logistic regression. Beginning with the Word Count, I set the Word Count at different intervals and observe no difference in the classification model's ability to predict loan default. In Figure 5 Update Analysis , the fourth and fifth Update made by the borrower appear to be the inflection point where the predicted model begins to outperform the null classification model. The R2 increases from .054 for a single Update to .299 at the fifth Update demonstrating a significant improvement in the model's explanatory power. In Figure 6, the number of Questions Answered R2 slightly increases across each interval. Whereas, there is a noticeable improvement in the predicted model for Misspellings at the fifth Misspelling in Figure 7. In loan descriptions with five or more Misspellings, the predicted model demonstrates 5.5% improvement in accuracy over the null classification model. The explanatory power of the model also improves with each subsequent Misspellings beginning with a R2 of .063 for the first Misspelling and ending with a value of .242 for the fifth Misspelling. Lastly, the Flesch Index score has a significant impact on the model's predictive power. Flesch Index scores greater than 125 correctly identified 100% of the observations (n=54) with a R2 value of .636. From these findings, there appear to be separating conditions between good and bad borrowers observable through different forms of social disclosure.

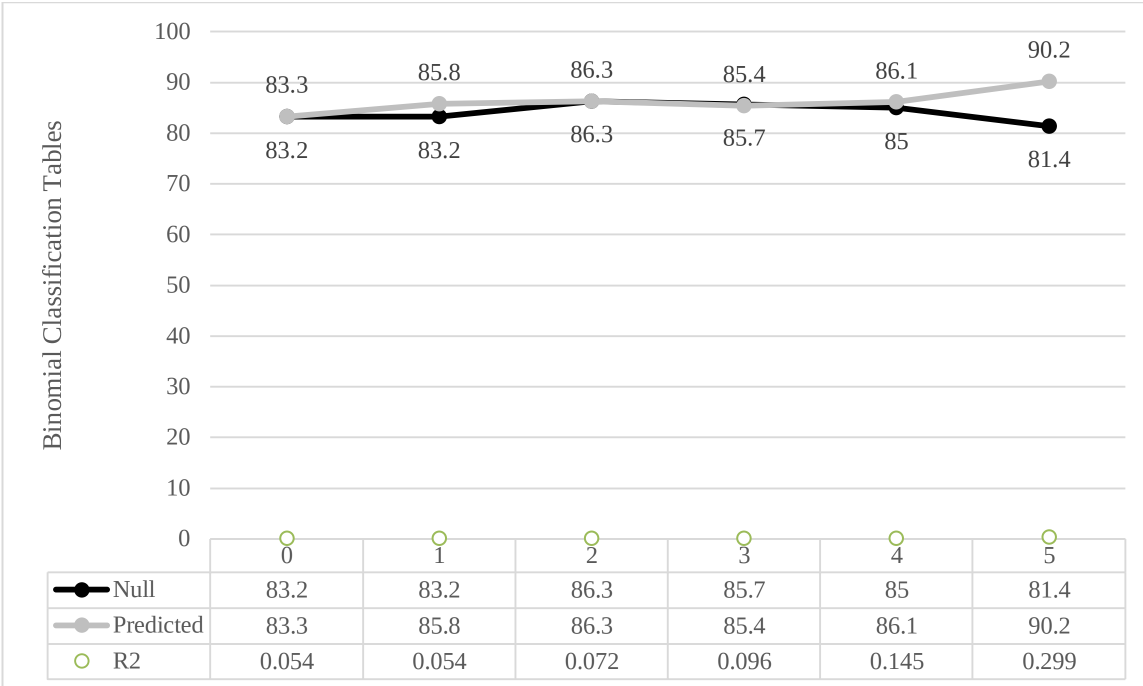
Figure 4 Word Count Analysis



Word Count	**N**	**Null**	**Predicted**	**R^2**
Greater Than 0	31550	85.2	85.2	0.057
Greater Than 10	29149	85.4	85.4	0.056
Greater Than 50	13538	87	86.9	0.057
Greater Than 100	5568	87.8	87.8	0.068
Greater Than 150	2756	87.2	87.3	0.084
Greater Than 200	1524	86.2	85.9	0.084

Notes: This table provides the results of the binary logistic regression for dependent variable Loan Status using Lending Club data. The model measures word_count predictive ability and controls for: verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, renewable_energy, small_business, vacation, wedding, loan_amount_req_log, FICO_avg, dti, home_own, married, divorced, single, engaged, children, revol_util(%), mthly_inc_log, loan_vol_log, credit_spread, consumer_sentiment, per_capita_wage_log.

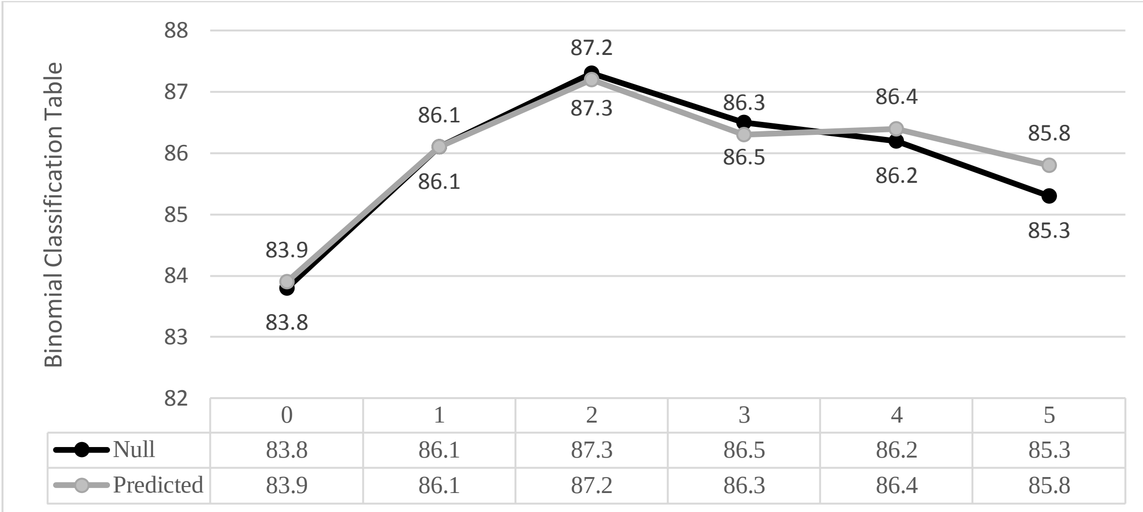
Figure 5 Update Analysis



Updates	**N**	**Null**	**Predicted**	**R^2**
0	8040	83.2	83.3	0.054
1	18416	83.2	85.8	0.054
2	3750	86.3	86.3	0.072
3	881	85.7	85.4	0.096
4	274	85	86.1	0.145
5+	102	81.4	90.2	0.299

Notes: This table provides the results of the binary logistic regression for dependent variable Loan Status using Lending Club data. The model measures Updates predictive ability and controls for: verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, renewable_energy, small_business, vacation, wedding, loan_amount_req_log, FICO_avg, dti, home_own, married, divorced, single, engaged, children, revol_util(%), mthly_inc_log, loan_vol_log, credit_spread, consumer_sentiment, per_capita_wage_log.

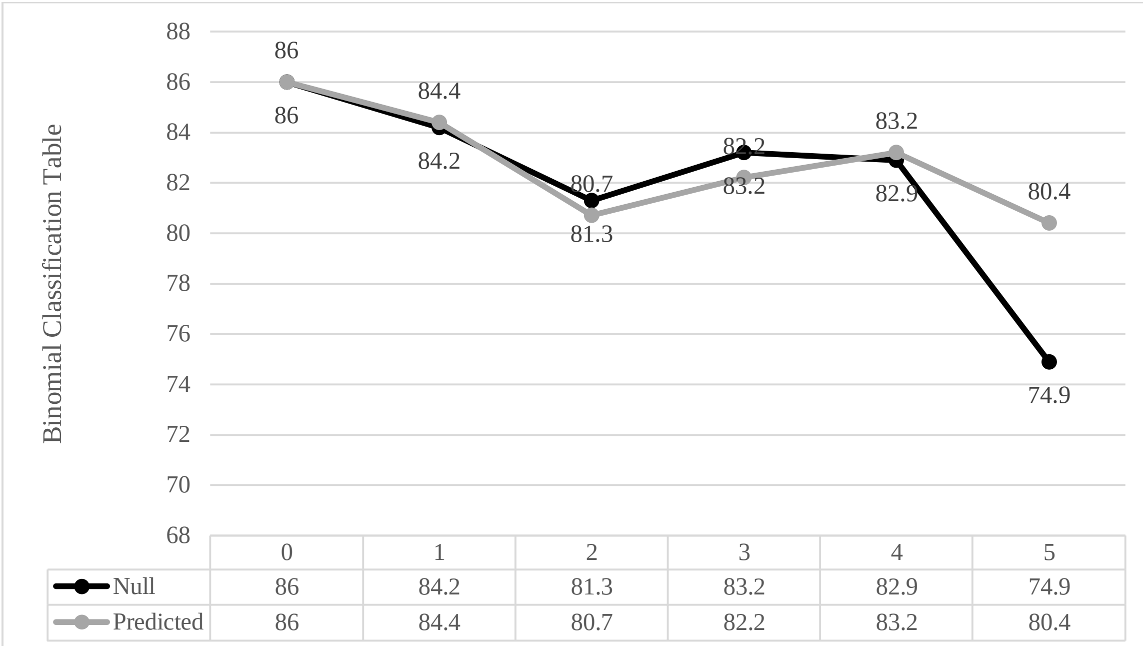
Figure 6 Questions Answered Analysis



QA Total	N	Null	Predicted	R^2
0	14205	83.8	83.9	0.058
1	6548	86.1	86.1	0.064
2	4397	87.3	87.2	0.055
3	2761	86.5	86.3	0.067
4	1679	86.2	86.4	0.065
5+	883	85.3	85.8	0.095

Notes: This table provides the results of the binary logistic regression for dependent variable Loan Status using Lending Club data. The model measures Updates predictive ability and controls for: verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, renewable_energy, small_business, vacation, wedding, loan_amount_req_log, FICO_avg, dti, home_own, married, divorced, single, engaged, children, revol_util(%), mthly_inc_log, loan_vol_log, credit_spread, consumer_sentiment, per_capita_wage_log.

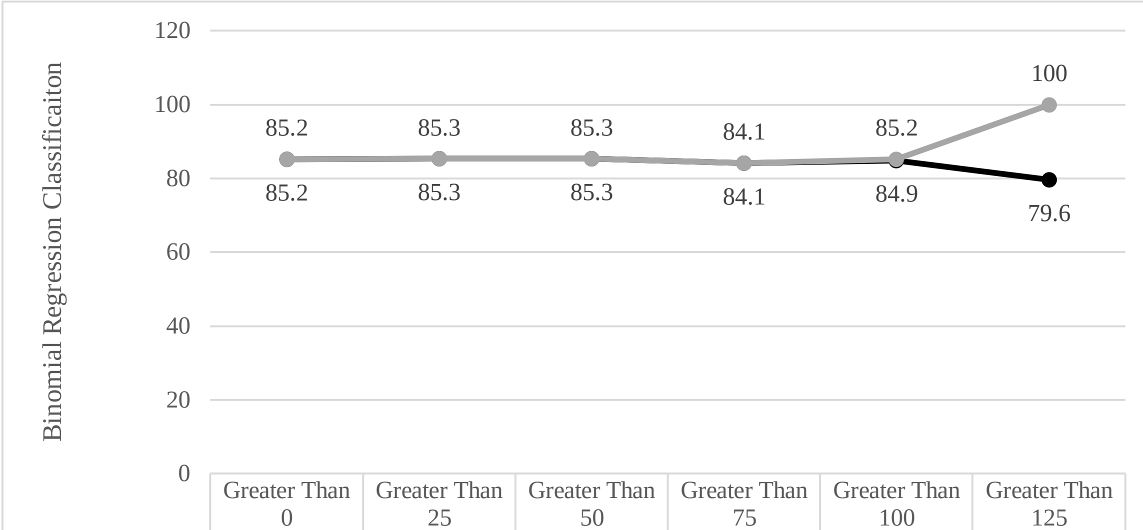
Figure 7 Misspellings Analysis



Misspellings	N	Null	Predicted	R^2
0	23048	86	86	.054
1	4988	84.2	84.4	0.063
2	1672	81.3	80.7	0.073
3	760	83.2	82.2	0.126
4	392	82.9	83.2	0.112
5+	199	74.9	80.4	0.242

Notes: This table provides the results of the binary logistic regression for dependent variable Loan Status using Lending Club data. The model measures Misspellings predictive ability and controls for: verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, renewable_energy, small_business, vacation, wedding, loan_amount_req_log, FICO_avg, dti, home_own, married, divorced, single, engaged, children, revol_util(%), mthly_inc_log, loan_vol_log, credit_spread, consumer_sentiment, per_capita_wage_log.

Figure 8 Flesch Index Analysis



	Greater Than 0	Greater Than 25	Greater Than 50	Greater Than 75	Greater Than 100	Greater Than 125
Null	85.2	85.3	85.3	84.1	84.9	79.6
Predicted	85.2	85.3	85.3	84.1	85.2	100

Flesch Index	**N**	**Null**	**Predicted**	**R^2**
Greater Than 0	31340	85.2	85.2	0.056625
Greater Than 25	30806	85.3	85.3	0.056
Greater Than 50	26178	85.3	85.3	0.056
Greater Than 75	8645	84.1	84.1	0.058
Greater Than 100	675	84.9	85.2	0.108
Greater Than 125	54	79.6	100	.636

Notes: This table provides the results of the binary logistic regression for dependent variable Loan Status using Lending Club data. The model measures Flesch Index predictive ability and controls for: verified_status, car, credit_card, debt_consolidation, educational, home_improvement, house, major_purchase, medical, moving, renewable_energy, small_business, vacation, wedding, loan_amount_req_log, FICO_avg, dti, home_own, married, divorced, single, engaged, children, revol_util(%), mthly_inc_log, loan_vol_log, credit_spread, consumer_sentiment, per_capita_wage_log.

The Loan Status results from the regression analysis are further tested and analyzed for robustness in the following section. I first test for the difference between high-grade and low-grade loans. The high-grade loans are represented by A, B, and C grade loans and low-grade loans reflect D, E, F, and G grade loans. Separation between each of these groups enables observation of borrowers with different levels of creditworthiness based on hard credit information. In Table 26, the test of two means test is performed. Loan Status is used as the test variable in the analysis to determine whether the mean variance between Loan Status of Fully Paid is significantly different between high-grade and low-grade loans. The results indicate the null hypothesis should be rejected due to a probability of less than 1% (p<.001). In Table 27, an additional Two-Means test is performed between loans with social disclosures and without social disclosures in the loan description. Word Count variable is used to separate the sample populations of loans with greater than one word in the loan description from the loan descriptions left blank by the borrower. The test results in Table 27 indicate a less than 5% (p<.038) probability that non-disclosure loans are equal to loans with social disclosures. The loan grades and non-disclosure loans are used as selection criteria filters in the final models within this analysis.

Table 26 Two-Means Test for High Grade and Low Grade Loans

Test Variable	**Grade Group**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Loan Status	High Grade = 1	25874	.88	.33	.002
	Low Grade=0	5765	.75	.434	.006

Independent Samples Test									
Loan Status	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. 2-tailed	Mean Diff.	Std. Error Diff.	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	2019.754	0.000	24.85	31637	0.000	.127	.05	.117	.137
Equal variances not assumed			20.95	7320	0.000	.127	.06	.115	.139

Table 27 Two-Means Test for Social Disclosures versus Non-Disclosure

Test Variable	**Word Count**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Loan Status	Word Count > 1	31525	.85	.355	.002
	Word Count < 1	114	.82	.389	.036

Independent Samples Test

Loan Status	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. 2-tailed	Mean Diff.	Std. Error Diff.	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	4.296	.038	1.092	31637	.275	.036	.033	-.029	.102
Equal variances not assumed			.996	113.680	.321	.036	.037	-.036	.109

A final model is constructed based on only the variables found significant for fully repaid loans. In Model 1, each significant variable is included without any selection filters. In Model 2, the loan Grade Group selection criteria are used to limit observations to only low-grade loans. In Model 3, only loans that contain social disclosures are selected. The results indicate that loan volume becomes insignificant across each model and the children attribute also loses significance in Model 1 and 2. Interestingly, the most significant social disclosure, Educational Explanation, becomes insignificant when observing only low-grade loans. Furthermore, the findings may indicate educational success is only a positive influence on the loan performance if the borrower is low-risk. In terms of performance, Model 2 modestly outperforms the null classification model percentage correct of 74.8% with a predicted model of 74.9%. Model 2 was also able to successfully predict 2.1% of the loans that would be Charged Off and was 99.4% correct in predicting Fully Paid loans. These findings suggest that social disclosures are more important for predicting the higher-risk borrowers. Compared to Model 1 and 3, these models failed to outperform the null classification model of 85.2%. Model 1 and Model 2 each respectively predicted .9% of the loan Charge Offs and 99.9% of the Fully Paid loans. In other words, in the sample Model 1 was able to correctly predict 41 loan defaults out of 4,679 and Model 3 was able to predict 42 loan defaults out of 4,645 Charged Off loans. The results of each model are provided in Table 28.

Table 28 Significant Variables Model

	Model 1		Model 2		Model 3	
	Exp(B)	Std. Error	Exp(B)	Std. Error	Exp(B)	Std. Error
Constant	-3.441***	0.947	-3.33*	1.770	-10.855***	0.830
qa_total	0.037***	0.010	0.064***	0.016	0.012	0.010
misspellings	-0.055***	0.009	-0.058***	0.015	-0.057***	0.009
hardworking	0.213***	0.041	0.173**	0.075	0.243***	0.041
moral	-0.1**	0.049	-0.097	0.089	-0.103**	0.048
credit_combo	0.16***	0.037	0.259***	0.071	0.164***	0.037
loan_combo	0.146***	0.034	0.141**	0.064	0.143***	0.034
debt_combo	0.238***	0.038	0.265***	0.073	0.242***	0.038
edu_exp	0.423***	0.127	0.179	0.204	0.39***	0.127
oth_rate_exp	0.409***	0.082	0.562***	0.159	0.429***	0.081
car	0.344***	0.129	0.694**	0.326	0.47***	0.131
credit_card	0.278***	0.064	0.096	0.126	0.377***	0.064
debt_consolidation	0.114**	0.051	0.038	0.094	0.205***	0.051
educational	-0.242*	0.139	-0.455**	0.232	-0.276**	0.140
major_purchase	0.398***	0.101	0.401**	0.201	0.44***	0.101
small_business	-0.593***	0.079	-0.493***	0.130	-0.741***	0.078
wedding	0.33**	0.136	0.395	0.240	0.321**	0.135
loan_amount_req_log	-0.537***	0.070	-0.602***	0.129	-0.75***	0.070
FICO_avg	0.006***	0.001	0.005***	0.002	0.015***	0.001
dti	-0.017***	0.002	-0.009**	0.004	-0.018***	0.002
children	-0.075	0.047	-0.069	0.082	-0.1**	0.047
mthly_inc_log	0.929***	0.080	0.673***	0.150	1.07***	0.079
loan_vol_log	-0.096	0.058	0.07	0.104	0.051	0.046
consumer_sentiment	-0.006**	0.003	-0.016***	0.005	-0.011***	0.003
per_capita_wage_log	0.456***	0.157	0.478	0.297	0.397**	0.157
Cox & Snell R Square	0.062		0.041		0.055	
Nagelkerke R Square	0.110		0.061		0.097	
Number of Observations	31,623		5,760		31545	

Notes: Table 28 provides the results of the binomial logistic regression for dependent variable Total Recovered Principal using Lending Club data. Model 1 includes the forms social disclosure variables; Model 2 includes the content variables. Model 3 is the integration model with the forms of social disclosures and content variables. P-value of .10 is represented with a *, p-value of .05 of less is represented with two **, and p-value of .01 or less is represented with three ***. Cells with "-" indicate the areas that did not have adequate data to perform the analysis or variables not included in the model.

CHAPTER V

Conclusions

This study has examined the influence of social disclosures within fixed rate P2P lending from both a funding and repayment perspective. The results from the analysis of social disclosures can be summarized in two primary ways. First, the factors significant for funding a loan are inconsistent with the factors significant to repayment of the loan. Second, prescriptive filters based on social disclosures can improve the likelihood of selecting a creditworthy borrower.

The social disclosures advantageous to borrowers are not always in the best interest of lenders. In terms of funding, H_1, H_2, H_5 and H_6 proved different forms of social disclosures and the specific content within the disclosure influence Duration times, Total Investment, and the Percent Investment. However, factors such as Education and Interest Rate on Other Debt were negatively associated with funding a loan in H_5 and H_6. These same two variables proved the most important to borrower repayment of the loan. These findings create a clear conflict between borrowers and lenders within P2P lending. For example, Poor Credit and Amount of Other Debt explanations are rewarded with higher levels of lender investment. Meanwhile educational success and obtaining a more beneficial interest rate are punished with lower levels of lender investment. Educational success demonstrates an ability to earn higher incomes, and a reduction in borrower interest rates increases the monthly cash flow necessary to service the debt. Both of these factors would reasonably improve a borrower's ability to repay their debt obligations. In practice, it appears lenders base investment decisions on more compassionate or altruistic social disclosures. For example, the Religious identity claim was the second most important claim in terms Percent Invested for retail and institutional lenders. Ironically, Religious claims also have the strongest negative relationship with the Total Recovered Principal. These findings may suggest that religious borrowers are less creditworthy than borrowers without religion. Alternatively, the findings may also indicate uncreditworthy borrowers unscrupulously make religious claims to appeal to religious lenders regardless of their ideological background. It also important to discuss that a subset of the variables significant for receiving funds are also predictors of loan performance.

The study provides strong evidence that claiming to be Hardworking resonates with lenders and is a good indicator of a borrower's true creditworthiness. I posit that there are intrinsic qualities associated with being Hardworking that share a strong correlation with creditworthiness. Being a hard worker may indicate the person is willing to work and sacrifice more in order to repay their debts. In addition to a borrower claiming Hardworking characteristics, exhibiting these same characteristics on the platform indicates credit worthiness.

This study produces findings that reveal prescriptive ways borrowers indicate creditworthiness through different forms of social disclosure on the loan listing. The results suggest it is advantageous for borrowers to invest time into well-written loan descriptions and remain engaged with potential lenders. Variables such as Updates, Questions Answered, and Flesch Index improve a borrowers funding amount from retail and institutional investors while lowering the time required to receive the loan. Deeper analysis of each of these variables demonstrate borrowers that Update their listing five times or have a readability score over 125 are more likely to fully pay their loan. The work required to Update loan listings or develop clear and concise loan descriptions maybe akin to the qualities found in Hardworking individuals. For example, each Question Answered increases the Total Recovered Principal by roughly 2.5%. Answering questions requires effort and borrowers that work hard to obtain funding appear to also be the same individuals working hard to repay their debts. Quite the opposite, Lenders should also observe the effort the borrower places into the loan description in terms of Misspellings. Careless mistakes may be indicative of a lack of hard work and consistently indicated higher risk and poorer credit borrowers. In the analysis of Misspellings, the model outperforms the null classification model when there are five or more Misspellings in a loan description. Furthermore, Misspellings reduce the Total Recovered Principal by up to 3% per error. Borrowers that take the time to proof read their writing and respond to lender questions are statistically more likely to repay their debt obligations. The importance of these results and the methodology used in this research is of increased importance given the changing direction in P2P lending.

Based on this study, social disclosures enable retail and institutional investors to separate good and bad borrowers. However, as of December 10, 2015 the loan description as well as question and answer fields were removed from the Lending Club platform for security and privacy issues. Lending Club Lending Club stated the change was necessary to prevent borrowers from publishing personally identifiable information. As a result, social disclosures are an endangered financial innovation on fixed-rate P2P platforms. Social disclosures contributed to the “peer” qualities of the platform and these changes may now reduce some lenders’ ability to determine borrower creditworthiness. Social disclosures represent a financial innovation facilitating the democratization of finance to more investors and bridging the human interaction found in traditional banking relationships. Dissemination of these findings are of increased importance in light of the Lending Club platform policy change, and to advocate the advantages of keeping “peer” interaction within P2P lending.

REFERENCES

Agrawal, A. K., Catalini, C., & Goldfarb, A. (2013). *Some simple economics of crowdfunding*. Retrieved from

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

Allen, F., & Santomero, A. M. (1997). The theory of financial intermediation. *Journal of Banking & Finance, 21*(11), 1461-1485.

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice, 39*(1), 53-73.

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., . . . Funk, B. (2011). Online Peer-to-Peer Lending–A Literature. *Journal of Internet Banking and Commerce, 16*(2), 1-18.

Bakos, J. Y. (1991). A strategic analysis of electronic marketplaces. *MIS quarterly*, 295-310.

Berger, A. N., & Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *Journal of business,* 351-381.

Berkovich, E. (2011). Search and herding effects in peer-to-peer lending: evidence from prosper. com. *Annals of Finance, 7*(3), 389-405.

Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. *RAND Journal of Economics*, 309-328.

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies, 34*(1), 1-25.

Castelli, A., Dwyer, G. P., & Hasan, I. (2006). *Bank Relationships and Small Firms' Financial Performance*. Retrieved from

Chen, D., & Han, C. (2012). A Comparative Study of online P2P Lending in the USA and China. *Journal of Internet Banking and Commerce, 17*(2), 1.

Chen, N., Ghosh, A., & Lambert, N. S. (2014). Auctions for social lending: A theoretical analysis. *Games and Economic Behavior, 86*, 367-391.

Damiano, E., & Li, H. (2008). Competing matchmaking. *Journal of the European Economic Association*, 789-818.

DellaVigna, S., & Gentzkow, M. (2009). *Persuasion: empirical evidence*. Retrieved from

Diamond, D. W. (1991). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of political Economy*, 689-721.

Dixon, W. J. (1960). Simplified estimation from censored normal samples. *The Annals of Mathematical Statistics, 31*(2), 385-391.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., & Kammler, J. (2016). Description-text related soft information in peer-to-peer lending–Evidence from two leading European platforms. *Journal of Banking & Finance, 64*, 169-187.

Duarte, J., Siegel, S., & Young, L. (2012). Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies, 25*(8), 2455-2484.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending. *Applied Economics, 47*(1), 54-70.

Feller, J., Gleasure, R., & Treacy, S. (2013a). Crowdfunding: Past Research, Future Directions.

Feller, J., Gleasure, R., & Treacy, S. (2013b). From the wisdom to the wealth of crowds: A metatriangulation of crowdfunding research. *TOTO Research Project*, V2.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011). Strategic herding behavior in peer-to-peer loan auctions. *Journal of Interactive Marketing, 25*(1), 27-36.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell me a good story and I may lend you money: the role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research, 48*(SPL), S138-S149.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2009). *Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending?* Paper presented at the AFA 2011 Denver Meetings Paper.

Jensen, M. C. M., William H. (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics, 3*(4), 305–360. doi:10.2139/ssrn.94043

Kawai, K., Onishi, K., & Uetake, K. (2014). Signaling in online credit markets. *Available at SSRN 2188693*.

Lewis, G. (2011). Asymmetric information, adverse selection and online disclosure: The case of eBay motors. *The American economic review, 101*(4), 1535-1546.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science, 59*(1), 17-35.

Mach, T. L., Carter, C. M., & Slattery, C. R. (2014). Peer-to-Peer lending to small business. *Federal Reserve Board Finance and Economics Discussion Series*(2014-10).

Malmendier, U., & Shanthikumar, D. (2007). Are small investors naive about incentives? *Journal of Financial Economics, 85*(2), 457-489.

McAfee, R. P., & McMillan, J. (1987). Auctions and bidding. *Journal of economic literature, 25*(2), 699-738.

Merton, R. C. (1974). On The Pricing of Corporate Debt: The Risk Structure of Interest Rates* *The Journal of Finance, 29*(2), 449-470. doi:10.1111/j.1540-6261.1974.tb03058.x

Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review, 87*(4), 1385-1413.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing, 29*(1), 1-16.

Moritz, A., & Block, J. H. (2016). Crowdfunding: A literature review and research directions *Crowdfunding in Europe* (pp. 25-53): Springer.

Myers, J. H., & Forgy, E. W. (1963). The Development of Numerical Credit Evaluation Systems. *Journal of the American Statistical Association, 58*(303), 799-806. doi:10.1080/01621459.1963.10500889

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 13*(2), 248-277.

Petersen, M. A., & Rajan, R. G. (1994). The Benefits of Lending Relationships: Evidence from Small Business Data. *The Journal of Finance, 49*(1), 3-37. doi:10.1111/j.1540-6261.1994.tb04418.x

Pitschner, S., & Pitschner-Finn, S. (2014). Non-profit differentials in crowd-based financing: Evidence from 50,000 campaigns. *Economics Letters, 123*(3), 391-394.

Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture?: Evidence of Discrimination from Prosper.com. *Journal of Human Resources, 46*(1), 53-92. doi:10.3368/jhr.46.1.53

Puls, T. (2015). Market Gives LendingClub Too Much Credit. *Stock Strategist*. Retrieved from http://news.morningstar.com/articlenet/article.aspx?id=714839&part=3

Rajan, R. G. (1992). Insiders and outsiders: The choice between informed and arm's-length debt. *The Journal of Finance, 47*(4), 1367-1400.

Schwienbacher, A. L., Benjamin (2010). *Crowdfunding of Small Entrepreneurial Ventures*: Oxford University Press, Forthcoming.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, 24-36.

Von Thadden, E.-L. (1995). Long-term contracts, short-term investment and monitoring. *The Review of Economic Studies, 62*(4), 557-575.