Voluntary Involuntary Disclosure

By

Randall J. Hucks

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2015

Doctoral Committee:

Professor Raffi J. Indjejikian, Chair
Associate Professor Venkatesh K. Nagar
Professor Amiyatosh K. Purnanandram
Associate Professor Catherine Shakespeare
Professor J. Frank Yates

Acknowledgements

I thank the members of my dissertation committee: Raffi Indjejikian (Chair), Venky Nagar, Cathy Shakespeare, Amiyatosh Purnanandam and Frank Yates. Thanks as well to Chris Williams, Greg Miller, Ryan Ball, Rick Cazier, Jed Neilson, Robin Queen, Julie Boland and Jason Chen for their helpful comments and suggestions, and to Kathleen Hucks for her research assistance. I also gratefully acknowledge the Paton Accounting Fellowship and the Thomas W. Leabo Memorial Scholarship for their financial support. All remaining errors are my own.

Table of Contents

List of Figures

List of Tables

Abstract

Prior examination of financial disclosures associates increasing linguistic complexity with poor firm performance, but cannot differentiate between competing signaling and managerial obfuscation hypotheses. In order to assess whether the lexical properties of written language signal discloser quality, I analyze borrower submitted free-form descriptions of anonymous peer-to-peer loans and find that borrower misspellings predict lower funding rates, longer time to fund, higher default rates, and lower yields to lenders. My findings suggest that the lexical properties of writing are a signal of the writer's quality.

Chapter I: Introduction

Recent literature regarding large-sample textual analysis of corporate disclosures has examined the effects of linguistic complexity on the ability of financial statement users to interpret the information contained within those disclosures (Lehavy, Li & Merkley, 2011; Lee, 2012). This work establishes that the lexical properties of written language affect the behavior of financial statement users, and extends the textual analysis literature which suggests that firms intentionally obfuscate the information in their annual reports via linguistic complexity (Li, 2008). However, because this literature studies the linguistic complexity of financial statements of firms with varying levels of operational complexity, and with varying strategic needs, much of the textual analysis literature suffers from endogeneity concerns (Li, 2010). Several of these studies test a set of joint hypotheses: that textual analysis is able to reveal something about the disclosing firm; and that managerial language is a strategic choice. Furthermore, in Brochet et al., (2012), which suggests that linguistic complexity in conference calls is a function of language barriers and is therefore not strategic, the managers of foreign firms are in a controlled environment and are likely coached in terms of communication style. In order to clearly establish that the lexical properties of written language are an attribute of disclosure that impacts how the disclosure is used, it is important to distinguish strategic choices in writing from the existence of certain lexical properties.

In this study, I attempt to answer whether the lexical properties of written language in financial disclosure reveal information about the discloser. Specifically, I examine whether spelling ability affects lender decisions and predicts borrower quality in an online peer-to-peer lending setting where voluntary disclosures by borrowers are common, and the general information environment is far sparser than in traditional capital markets. [1][2] The key aspect of misspelling as my measure in this context is that it is unlikely that someone would deliberately misspell. Therefore, misspellings are unlikely to represent a strategic choice, and I can eliminate communication strategy as a confounding issue.[3]

Prior literature suggests that spelling ability (or inability) affects perceptions of intelligence (Figueredo & Varnhagen, 2005; Scott et al., 2014; Vignovic & Thompson, 2010), credibility (Everard & Galleta, 2006; Stiff, 2012, Vignovic & Thompson, 2010), conscientiousness (Figueredo & Varnhagen, 2005; Everard & Galleta, 2006; Vignovic & Thompson, 2010, Boland & Queen, 2015) and employability (Scott et al., 2014), while propensity to misspell is negatively related to education (Hargittai, 2006) and intelligence (Kreiner et al., 2002). Furthermore, prior research shows that misspelling is strongly correlated with reading disability (Russell, 1955; Hinshaw, 1992). Reading disabilities have been linked to behavioral problems (Hinshaw, 1992; Beitchman et al., 1996; Cohen et al., 1993), and tend to be persistent, depending on other environmental factors (Beitchman et al., 1996; Rutter & Yule, 1975; Spreen, 1988). Consistent with these findings, individuals with reading disabilities from all socioeconomic backgrounds have poorer educational and occupational outcomes than the non-disabled, which are related to insolvency and bankruptcy (Spreen, 1988; Pape et al., 2011, Devaney & Lytton, 1995). Because hard credit information is backward looking, and explains

[1] By this I mean that there are far fewer signals of quality, such as analyst reports, media reports, networks and relationships.
[2] The relative lack of information allows me to more easily extract the effects of the disclosures and their misspellings while also allowing me to control for almost anything the lender may have access to. At the same time, such an environment may limit the generalizability of my study.
[3] While misspelling and linguistic complexity are different constructs, both are lexical properties of written language and can influence a person's perceptions of the communication and potentially its use (Rennekamp, 2012; Stiff, 2012).

only a small portion of future default, misspelling could contain soft information about borrower type, and could therefore be predictive of future loan outcome (Petersen, 2004.)

Taking these findings together, I hypothesize that misspellings directly and negatively influence the capital allocation decisions of lenders, which I measure as the dollar amount of funding provided per day, and as the time it takes to fully fund the loan.[4] I also predict that misspellings are predictive of adverse loan outcomes, such as incidence of default and decreased returns. Consistent with this perspective, I find that misspellings in borrowers' voluntary disclosures influence loan decisions, and are predictive of loan outcomes. In particular, I find that misspellings predict lower funding rates for loans, longer time to fully fund the loans, higher default rates, and lower yields to lenders. For instance, I find that each misspelling in a loan description predicts a marginal decrease in funding of $30 per day, a marginal increase in the time to fully fund of 2.6%, a marginal increase in the default rate of 0.8% and a marginal decrease in yield of 0.6%.

Peer-to-peer lending is a means for borrowers to "crowdfund" projects such as debt consolidations or home improvements from online lenders. Lending companies serve as intermediaries by attracting borrowers with better loan terms than those offered by traditional lenders as well as a quick and convenient loan application process, and by providing underwriting and collections services for lenders.[5] The data for my study are obtained from Lending Club, the largest peer-to-peer lending company in the United States. The information lending companies such as Lending Club provide to lenders includes loan characteristics,

[4] As the price (interest rate) of the loan is set before a loan description is completed, misspellings cannot affect price; therefore, misspelling's effect on loan demand can be measured purely with respect to the supply of funds. Because almost all of the loans in the sample are fully funded, I measure demand using funding rate, measured in dollars per day, as well as time to fully fund, measured in days. 98.6% of the loans in the sample are fully funded.
[5] A Lending Club survey of 20,913 Lending Club borrowers indicates that borrowers borrow at rates 31% lower than the rates on their other unsecured debt.

borrower characteristics, borrower credit history and a written description of the loan composed entirely by the borrower and submitted at his option after the underwriting process is completed.[6] The fact that the loan description is submitted after the loan is underwritten allows me to distinguish the demand effects of the loan description while holding the interest rate constant.[7] Figure 1 provides an example of a loan listing by Lending Club as presented to the lenders.

In contrast to traditional lending, peer-to-peer lending is almost always anonymous. Anonymity implies that a lender is unable to acquire information about a borrower beyond what is included in the loan application. Consequently, a borrower's description of the loan and related representations potentially play a more important role in a peer-to-peer lending decision than in a traditional loan setting. However, because a borrower's written description of the loan is optional, its contents and presentation are entirely discretionary. For example, these descriptions often include representations of the borrower's personality traits, pleas for assistance, as well as unverifiable claims about the borrower's financial security. Importantly, although a borrower's decision to provide these disclosures is deliberate and potentially strategic, his propensity for spelling and grammar errors is likely not.[8] Hence, I suggest that misspellings in voluntary disclosures represent a clear signal of poor borrower quality and are used by lenders to make their loan decisions.

Prior peer-to-peer lending studies have examined various features of written loan descriptions, other than misspellings, and have found that lenders often rely on these descriptions despite the "cheap-talk" nature of borrowers' claims. For example, Michels (2012) and

[6] As of March 2014, loan descriptions are no longer provided to lenders due to privacy concerns and increase regulatory requirements imposed on peer-to-peer lenders.
[7] Because the loan descriptions are written and submitted after the terms of the loan (interest rate, amount and term) are set, the terms of the loan are exogenous to the contents of the description.
[8] Because this is in the context of for-profit microlending, it is unlikely that any empathy received for presenting oneself as pitiable will outweigh the negative effects of misspelling on lenders' perceptions. To the extent that any borrowers are purposely misspelling, this will bias against finding a result.

Herzenstein et al., (2011) find that borrowers that provide these types of voluntary disclosures often attract more lenders and borrow at lower interest rates than others.[9] In a more recent study, Gao & Lin (2013) find that more readable, positive, and less deceptive disclosures are positively related to higher lender demand and higher quality loans. In contrast to the focus of prior studies, the distinguishing feature of misspellings is that they are not premeditated or intentional. Whereas borrowers may deliberately provide more readable and less deceptive disclosures or strategically engage in cheap-talk, they do not deliberately misspell. Hence, I do not expect measures of readability and deception employed in prior peer-to-peer lending studies to fully capture the effect of misspellings. Nevertheless, I control for measures employed in prior studies (e.g., Flesch-Kincaid ease of reading, Gunning-Fog index) in my tests and find that my findings are qualitatively unaffected by their inclusion.

My paper contributes directly to the textual analysis in capital markets disclosure literature by establishing that the characteristics of the language used within the disclosure provides a meaningful signal of participant quality. Recent work in this area has suggested that this could potentially be the case, leaving it an open question (Brochet et al., 2012; Bloomfield, 2008). My findings imply that firms which exhibit certain lexical properties in their disclosures, whether strategically or nonstrategically, may be incurring a hidden cost by doing so. With regard to the literature surrounding peer-to-peer lending, my paper extends the work in "understanding what individuals can do to signal creditworthiness through narratives" (Morse, 2015). Given the nature of the peer-to-peer lending market, my paper also answers the call from John Campbell's 2006 Presidential Address to the American Finance Association (Campbell, 2006), which sought attention to issues of household investing mistakes and their attendant

[9] These studies use Prosper.com data from the period before it was shut down by the SEC. In this period, loans were not priced by Prosper.com, but rather were posted with all borrower-submitted information and rates were set via auction. As a result, information contained within the disclosure could affect interest rate.

welfare costs. In addition, while many studies have shown relations between misspelling and perceptions of the misspellers (e.g., Scott et al., 2014), misspellings and traits of the misspellers (e.g., Hargittai, 2006), and even misspellings and their effects on the behavior of counterparties (e.g., Everard & Galleta, 2006), to my knowledge, my paper is the first to find that misspellings are predictive of real misspeller performance. My paper also establishes that in the context of a debt market, misspellings are a potentially important source of soft information when information is sparse and information asymmetries are potentially severe. Finally, although my study is set in the distinctive peer-to-peer lending setting, my findings suggest that misspellings in disclosures are consequential in any setting where information asymmetries loom large and reliable information is limited, such as other online markets or in contracting settings.

The rest of my paper proceeds as follows: Section II provides information about peer-to-peer lending and discusses the development of my hypotheses; Section III includes information about the data used in the analyses; Section IV explains the research design and discusses the results; Section V contains additional analysis on the effects of misspelling; and Section VI concludes.

Chapter II: Related Literature and Hypothesis Development

Peer-to-peer lending is a way for borrowers to "crowdsource" capital over the internet from multiple lenders in the form of unsecured loans. Borrowers apply for loans via websites such as LendingClub.com or Prosper.com, and these companies then act as underwriters for the loans. Following the underwriting process, a borrower can reject or accept the terms offered by the website, and if the borrower accepts, the loan is made available to the website's pool of lenders. This partial disintermediation allows for lower interest rates by avoiding costs associated with traditional brick-and-mortar banks. However, it also implies that information asymmetries are potentially more severe. Without the soft information which traditional lenders obtain from their interactions with borrowers in the loan application process, anonymous lenders over the internet are at a relative disadvantage. In light of this disadvantage, evidence suggests that anonymous lenders are willing to consider information that is unverifiable cheap-talk (Michels, 2012), or that more traditional lenders are legally barred from using for ethical reasons (Pope & Sydnor, 2009, Ravina, 2012).

In this paper, I examine data provided freely to the public by LendingClub.com.[10] Lending Club is the market leader in peer-to-peer lending activity in the United States, having facilitated over $9 billion in loans since its inception in 2007, including over $1.6 billion in the first quarter of 2015. 72% of the loans originated through Lending Club are for debt consolidation or credit card refinancing according to the borrowers, and borrowers claim to be

[10] This data can be downloaded at https://www.lendingclub.com/info/download-data.action

from every state in the country with the exception of Iowa, Idaho, Maine, Mississippi, Nebraska, North Dakota and Vermont. The borrowers must be long-term residents of the United States, and must be at least 18 years old and have a verified bank account.

To obtain a loan from Lending Club, the potential borrower submits to a credit check at LendingClub.com. Lending Club then evaluates the information and provides a menu of loan terms (i.e., interest rate, amount, loan term) to the borrower. Upon acceptance of a loan, the borrower is given the opportunity to complete an optional, free-form disclosure, which is the subject of this study. This takes place after the interest rate, amount and loan term are set, and therefore the disclosure cannot affect the terms of the loan. There are no formal requirements for this section, which is simply labeled "Loan Description." Borrowers also have the option to answer scripted questions that lenders submit in this section.[11]

When the application is complete, and the terms of the loan are set, the loan is released to the Lending Club lender base.[12] [13] Lenders at Lending Club must reside in of one of 27 states which permit lending in these markets.[14] At no point is any personally identifiable information about the borrower released to the lenders; the borrower remains anonymous. This anonymity forces lenders to consider only the information which is submitted through the loan listing when making their decisions. Once the loans are posted, lenders can select the loans to invest in, and

---

[11] As of October, 2013, demand for these loans has increased to the point where 100% of the loans which pass through Lending Club's screening process are fully funded. Because of this, and because of the regulatory costs associated with filing each loan with the SEC, Lending Club has recently stopped allowing borrowers to submit these loan descriptions.

[12] Lending Club's main rival, Prosper, recently indicated that 80% of new loans originated through their site went to institutional lenders (Cortese, 2014), and it seems that institutions are similarly active at Lending Club.

[13] Aside from information obtained from media articles describing the investor base, and rules set by Lending Club, I have no information about the Lending Club lender base (e.g. Lender type, size of note purchased, number of lenders to a given borrower.)

[14] The 27 states are: California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Kentucky, Louisiana, Maine, Minnesota, Mississippi, Montana, Nevada, New Hampshire, New York, Rhode Island, South Dakota, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin and Wyoming. Residents of all other states except Kansas, Maryland, Ohio, Oregon and Washington, DC can buy and sell these notes in a secondary market.

the amount (in increments of $25) to invest in each loan. These decisions are driven by the risk preferences of the lenders, as well as any strategies that the lenders believe will allow them to increase returns or satisfy biases.[15] When the loan is fully funded, the proceeds (less origination fees) are released to the borrower's verified bank account. Figure 2 contains a graphic depicting the lending process. Payments are made by automatic monthly electronic funds transfers from the borrower's bank account. In the event of a failed payment, Lending Club is responsible for bringing delinquent borrowers back to current status, or for turning the loan over to an external collections agency. Loans which are 121 days past due are in default, with charge off occurring no later than 30 days after default.

Prior studies of peer-to-peer lending have focused almost exclusively on Prosper Marketplace (hereafter Prosper), Lending Club's primary competitor in the peer-to-peer lending market. Prior to 2009, Prosper used an auction pricing mechanism in setting its interest rates. In that setting, researchers proxied for lender demand using interest rate and number of bids. In the Lending Club setting, funding rate and time to fund are much clearer indications of lender demand, as the price of the loan is fixed and demand is reflected in how quickly the loan is funded. An additional key difference between the settings is that on Prosper.com, borrowers were allowed to post pictures and join social network "groups."

There are three main branches of the literature on how lenders deal with the greater adverse selection problems of online unsecured lending: picture based analysis; social network based analysis; and borrower submitted text based analysis, as in this study. For an excellent, thorough review of the nascent peer-to-peer lending literature, please see Morse (2015).

[15] Lin & Visnawathan (2014) find that there is home bias in peer-to-peer lending. They conjecture this may be a pro-social rather than a returns boosting bias.

To perform appearance based analysis, authors test whether subjective judgments of borrower submitted pictures affect lender demand for those borrowers' loans. Duarte et al. (2012) find that subjective judgment of borrower trustworthiness is correlated with lower interest rates, and predicts lower default and higher net yield. Pope & Sydnor (2009) and Ravina (2012) find that lender demand decreases when the borrower appears to be black, but their findings related to loan outcomes are mixed. Pope & Sydnor (2009) also finds that lender demand increases when the picture indicates the borrower is female, but this has no predictive power over loan outcome. Ravina (2012) examines the effect of beauty in this setting, showing that loans with pictures of more beautiful people are correlated with greater lender demand, but also worse loan outcomes. Pictures are no longer allowed as part of the loan listings at Prosper, and have never been permitted at Lending Club.

In the original Prosper setting, one mechanism which was available to borrowers to signal their quality was the ability to join groups, which are frequently but not necessarily affinity based (e.g., employer, alumni, religion). Some of these groups provided attestation services to the borrowers who joined for an upfront fee. Maier (2014) studies these groups, showing that group membership increases the probability that a loan is funded, but that these verification services must be in place to affect the borrower's interest rate, and to predict loan outcome. Hildebrand et al. (2013) also investigates these groups, and finds that when a borrower pays a fee to the group leader in exchange for the leader's bids, this is viewed by lenders as a signal of quality, while it actually is predictive of worse loan outcomes. These groups are no longer a part of the Prosper lending process, and were never present in the Lending Club market.

With respect to the user submitted textual submissions, several papers have investigated whether and how lenders use these loan descriptions to mitigate their concerns about information asymmetry. Michels (2012) and Herzenstein et al., (2011) show that borrowers frequently use loan descriptions to make unverifiable claims about themselves, or about their current financial standing, finding that these unverifiable claims increase lender demand, on average. With respect to loan outcomes, Michels (2012) finds that the lender response to the disclosures about financial standing is negatively related to default. On the other hand, the results in Herzenstein et al., (2011) show that identity claims, while positively related to lender demand, are also positively related to default. This suggests that lenders were making suboptimal investment decisions with respect to identity claims. Through textual analysis, Gao & Lin (2013) assess the effect of various writing characteristics on lender demand and their relation to loan default. They use measures of readability, optimism, objective language, and deception and find that more readable and positive and less deceptive disclosures are related to higher lender demand and higher quality loans.

Additionally, prior literature shows that peer-to-peer lenders are subject to home bias (Lin & Viswanathan, 2014) and that business loans attract more lender demand, but default at a higher rate (Mach et al., 2014.)[16] I control for these findings in my tests by controlling for the borrower's home state, as well as by controlling for the stated purpose of the loan.

The peer-to-peer lending setting described above suggests that the relationship between borrowers and lenders is fraught with significant information asymmetries and yet borrower-specific information is limited. When information asymmetries are severe, prior research has found that soft information about borrowers can be useful in lending decisions (Berger & Udell, 1995, Petersen, 2004). Soft information includes decision useful information which is less quantitative and less easily transmitted between decision makers, such as reputation or

[16] Mach et al., (2014) performed their analyses over data from Lending Club prior to December 31, 2012

personality traits like intelligence and credibility. In the context of financial markets, this soft information can take the form of the lexical properties of financial disclosures.

In this paper, I choose to study the informational effects of the presence of one lexical property in particular: misspelling. I choose to focus on misspelling rather than other types of prescriptive language such as grammar, because prior research has shown that misspelling, particularly in an electronic communication context, is associated with a number of constructs which could affect creditworthiness. Experimental research has established that misspelling adversely affects perceptions of intelligence (Figueredo & Varnhagen, 2005, Scott et al., 2014, Vignovic & Thompson, 2010), credibility (Everard & Galleta, 2006, Stiff, 2012, Vignovic & Thompson, 2010), conscientiousness (Figueredo & Varnhagen, 2005, Everard & Galleta, 2006, Vignovic & Thompson, 2010) and employability (Scott et al., 2014). In contrast, grammar errors may be less salient to the reader, and may even appear to be more formal than the correct form (i.e., hypercorrected forms) (Queen & Boland, 2015).[17]

Lender demand for a given loan is a reflection of the lender's perception of the borrower's quality. This perception is formed in large part by the hard data available in each borrower's credit report, such as the FICO score, the borrower's monthly income, or even the interest rate that Lending Club assigns to the loan. Lenders in aggregate form beliefs over the likelihood of repayment in conjunction with the interest rate on the loan to determine whether they would like to invest in the borrower, and if so, in what amount. Given this risk-reward tradeoff, some borrowers will naturally appear more attractive to some lenders than others given lenders' varying risk preferences as well as differing models with which individual lenders are assessing borrower quality. Because misspellings in loan descriptions may also affect lenders'

---

[17] Hypercorrected forms are grammatical errors which violate syntactic constraints but may actually indicate increased formality (e.g., "*He gave the book to Paul and I.*", "*Whom is speaking to me?*")

perceptions of borrower quality in terms of some credit risk factor not captured by a credit report (e.g., education, conscientiousness, employability), I posit that lender demand is decreasing in the number of borrower misspellings. I formalize this hypothesis as follows:

> **H1:** Lender demand is negatively related to the incidence of misspelling in the loan description.

Lender demand is represented both by a positive choice to invest as well as the decision over how much to invest, with larger amounts corresponding to greater demand. In aggregate, lender demand can be measured by funding rate, or the number of dollars invested per day, assuming that lenders make their decisions somewhat uniformly over time.[18] If, on average, lender perceptions of borrower quality (holding hard credit report information constant) are deteriorating in the number of misspellings in the loan descriptions, then I expect a negative relation between funding rate and the number of misspellings in the loan description, controlling for other information available to the lenders. Conversely, I also measure lender demand using the number of days from the posting of the loan it takes to fully fund the loan. A loan that is in higher demand will fund more quickly than a loan that is in lower demand, all else equal. Therefore, if lender perceptions of borrower quality (holding hard credit report information constant) are deteriorating in the number of misspellings in the loan descriptions, then I expect a positive relation between the number of days to fund the loan and the number of misspellings in the loan description, controlling for other information available to the lenders.

If lenders are making decisions based on the borrowers' misspellings, then it is reasonable to examine the relation between borrower misspellings and *ex post* loan outcomes. In

---

[18] In my tests, I control for month-to-month heterogeneity in overall lender demand by using Year-Month fixed effects. Therefore, my assumption that lenders make their decisions uniformly over time needs only to hold within any given month.

assessing borrower type, lenders will review the borrower's hard credit information, and make judgements of the borrower's creditworthiness based on the information contained therein. However, hard credit information such as FICO score, income level and employment history is backward looking, and only explains a small proportion of the variation in default probability.[19] Because of this, misspelling may contain soft information that is useful in assessing borrower type above and beyond what is contained in the borrower's hard credit report (Petersen, 2004.) Therefore, misspelling could be a predictor of future default, even after taking into account the borrower's hard credit information.

Misspelling is related to characteristics which may be related to borrower creditworthiness. Prior studies have shown negative relations between misspelling and education (Hargittai, 2006) as well as misspelling and intelligence (Kreiner et al., 2002). Furthermore, prior research shows that misspelling is strongly correlated with reading disability (Russell, 1955; Hinshaw, 1992). Reading disabilities have been linked to both externalizing (e.g., aggression and hyperactivity) and internalizing (e.g. anxiety and neuroticism) behavioral problems (Hinshaw, 1992; Beitchman et al., 1996; Cohen et al., 1993), and tend to be persistent, depending on other environmental factors (Beitchman et al., 1996; Rutter & Yule, 1975; Spreen, 1988). Consistent with these findings, individuals with reading disabilities from all socioeconomic backgrounds have poorer educational and occupational outcomes than the non-disabled (Spreen, 1988; Pape et al., 2011). Occupational outcomes and the attendant income disparities have been shown to predict delinquency, default, insolvency and bankruptcy (Devaney & Lytton, 1995). That these attributes are associated with spelling ability suggests that misspellings in loan descriptions could be predictive of loan outcomes.

[19] For instance, in my seasoned sample, a model of default regressed on interest rate, FICO and debt-to-income ratio explains only 3% of future defaults.

Alternatively, any decreased lender demand for misspellers' loans could be related to a "reverse halo effect", whereby lenders allow the undesirable misspelling to cloud their judgment of overall borrower quality, which may be unrelated (Nisbett & Wilson, 1977.) Indeed, the linguistics literature typically seeks to deny that typos and grammar mistakes reveal writer characteristics (Queen and Boland, 2015). If misspelling is not related to overall borrower quality, then I would expect that *ex post* loan outcomes would be unrelated to the number of misspellings in the loan descriptions. Further, even if misspellings are a signal of borrower intelligence and education, actual creditworthiness may not be related to education or intelligence (Dunn & Kim, 1999, Boyes et al., 1989), or the predictive power of misspellings could be captured by the information in the hard credit report. However, if misspellings reveal some additional dimension of borrower quality which constitutes an omitted credit risk factor, then it may be an optimal reaction for lenders to fund misspellers' loans less rapidly. I formalize my hypothesis in the alternative form below:

H2: *Ex post* loan outcomes are negatively related to the incidence of misspelling in the loan description.

I operationalize *ex post* loan outcomes using two proxies that are of interest to lenders: default and yield. Lenders can only fail to profit on a loan if the borrower defaults, and therefore it is important to the lender that the borrower avoids default. At the same time, default does not fully capture the profitability of a loan to the lender. The timing of the default, as well as the loss given default, also determines how profitable or unprofitable the loan is. Therefore, I view positive *ex post* loan outcomes (from the perspective of the lender) as lower incidence of default as well as higher yield to the lenders. Consequently, if there is a positive relation between the number of borrower misspellings and the probability of default, and a negative relation between the number of borrower misspellings and the yield to lenders, then misspellings are a valid signal of borrower quality.

Chapter III: Sample and Data

In this study I analyze data from the dataset which is freely available from Lending Club's website.[20] The sample includes all loans posted for funding by Lending Club from their inception in June of 2007 through November of 2013. Figure 1 contains an example of the set of information made available to a lender when considering a particular loan. This includes loan characteristics, borrower characteristics, borrower credit history, and the focus of the study, the loan description.[21] I track misspellings which occur in this section of the loan profile. Because I am interested in the value of misspellings as a signal of borrower quality, I am interested only in the effects of misspelling only on those borrowers who choose to complete loan descriptions.[22] Therefore, I perform my demand tests only over the 103,344 loans with descriptions. To test the effects of misspelling on loan outcomes, I further restrict the sample to all loans issued in November, 2011 or earlier, in order to allow the loans to become "seasoned." I do this to ensure that I am only testing loans which have had a reasonable amount of time from issuance to default.[23] Table 1 contains summary statistics for the loans split into all loans, and loans which have been "seasoned."

---

[20] This data can be downloaded at https://www.lendingclub.com/info/download-data.action. The data used in this study was downloaded on January 23, 2014.
[21] This however, does not include any information about the lenders.
[22] I exclude loans without loan descriptions because I am interested in the effects of misspelling on those who misspell, i.e., the average treatment effect on the treated. While the choice to disclose or not to disclose is interesting, it is outside of the scope of this paper.
[23] Lending Club defines a seasoned loan as a loan which has matured at least 10 months. Because default occurs a minimum of five months after dispersal of loan funds, I choose a more restrictive definition of a seasoned loan as a loan which has matured at least 24 months.

Dependent Variables

The dependent variables for testing my hypotheses are dollars invested per day by lenders (*FUND_PER_DAY*); days between listing and issuance of the loan (*FUND_DAYS*); default (*DEFAULT*); and net loan yield (*RETURNS*). Dollars invested per day is calculated as the total amount of funding supplied by lenders for a particular loan divided by the number of days between loan listing and loan funding. The mean (median) funding rate for my sample is $1,890 ($1,429) dollars per day, with a standard deviation of $1,680. The mean (median) number of days between loan listing and loan funding is 9.08 days (8 days) with a standard deviation of 4.65 days. Default is an indicator variable which is set to one when a loan is either listed as charged off or in default as of the download date. 11.8% of the loans in my seasoned sample went into default. Net loan yield is calculated as the internal rate of return of the loan assuming that all payments were received by the lenders uniformly over the length of the loan. The mean (median) yield in the sample is 4.2% (11.7%), and the standard deviation of lender yield is 27.5%.

Loan Characteristics

The loans that Lending Club facilitates have three dimensions: amount (*LOAN_AMT*), duration (*60_MONTH*) and interest rate (*INT_RATE)*. Mean (median) loan size in the sample is $13,750 ($12,000), and the standard deviation of loan size is $7,964.[24] The variable *60_MONTH* is an indicator set to one if the duration of the loan is five years, or zero if the duration is three years. 77% of sample loans have three year terms, and the remaining 23% have five year terms. Mean (median) interest rate (*INT_RATE*) in the sample is 13.4% (13.1%), with a standard deviation of 4.3%. Due to the rapid growth of this market, the loans in the sample are weighted

[24] For parsimony, I report only the descriptive statistics of the full sample here. Descriptive statistics for the seasoned sample are available in Table 1.

towards the end of the sample period. The number of months between the beginning of the sample period and the loan application date (*LIST_MONTH*) indicates that the mean (median) loan in the sample was listed in June, 2012 (October, 2012), while the 25th percentile is in December of 2011.

Borrower Characteristics

Lending Club provides to the lenders a few personal characteristics of their borrowers which are potentially related to their creditworthiness, including length of employment (*EMP_LENGTH*), monthly income (*INCOME*), whether Lending Club has verified that income *(VERIFIED*) and the ratio of monthly debt payments to monthly income (*DTI*). The mean (median) employment length in the sample is 6.5 (6) years, with a standard deviation of 3.8 years. Mean (median) monthly income for borrowers in the sample is $5,962 ($5,083), with a standard deviation of $4,513, and 63.2% of borrowers had their income verified by Lending Club. The mean (median) debt-to-income ratio in the sample is 16.1% (15.9%) with a standard deviation of 21.4%.

Borrower Credit History

Most of the data which Lending Club procures from the borrower and provides to the lender concerns the borrower's credit history. A borrower's loan description includes information about accounts currently delinquent (*DELINQUENT*), how much is currently delinquent (*DELINQ_AMT*), the number of delinquencies in the last two years (*DELINQ_2YRS*) and the time since the borrower's most recent delinquency (*TIME_SINCE_DELINQ*). Current delinquencies are virtually nonexistent, with a mean amount currently delinquent across the sample of $5. The average borrower in the sample has had 0.2 delinquencies in the previous two years, and the mean (median) borrower has had a delinquency in the past 10 years (never).

In addition to information on delinquency, the loan listings provide the borrower's FICO score (*FICO*) which is a one-size-fits-all summary statistic for the borrower's credit history. The mean (median) FICO score for borrowers in the sample is 703 (695), which is reflective of the sample containing prime and near-prime borrowers. Also listed are the borrower's months of credit history (*CREDIT_HIST*), the number of credit inquiries in the last six months (*INQUIRIES*), the number of public records such as bankruptcies and liens (*PUB_RECORD*) as well as the time since the borrower's most recent record (*TIME_SINCE_RECORD*). The average borrower in the sample has 15 years of credit history. Inquiries are somewhat common, with a mean number of inquiries of 0.83. Public records are far less common, with a mean number of records of 0.08.

The listing also includes the total balance of the borrower's revolving credit accounts (e.g. credit cards) (*REV_BAL*) and the percentage of the credit limits that are being used (*REV_UTIL*). Lastly, the loan listing provides the number of credit accounts that the borrower currently has open (*OPEN_ACC)* as well as the total number of credit accounts the borrower has ever had (*TOTAL_ACC*). For the sample, the mean (median) borrower has a revolving balance of $15,816 ($11,788), which constitutes 55.7% (58.2%) of his credit limit, and has had 24 (22) lines of credit, of which 10.6 (10) are open at the time of the loan application.

Loan Description Characteristics

The variable of interest in this study is the number of misspellings (*MSPL*) contained in the loan descriptions. In order to count the number of misspellings, I remove all text in the loan description that is not provided by the borrower (e.g., text appended by Lending Club.) I then use the Lingua::EN::Fathom module in Perl to count the number of words within each loan description which do not match the internal dictionary of the module. This variable is denoted as

MSPL. Note that because this procedure for detecting misspellings works in much the same way as a word processor's spell checker, it does not recognize homophonic spelling errors (i.e., errors in the form of words which are also in the dictionary as in *their* instead of *there*). To the extent that these homophonic spelling errors are present in the loan descriptions, and they affect lenders' perceptions of borrower quality, my tests likely understate the impact of misspellings.[25] The mean number of misspellings in a loan description is 0.28. 8.9% of all loan descriptions feature at least one misspelling, and 2.7% of all loan descriptions contain more than one misspelling.

In addition to the number of misspellings in each loan description, I also consider two alternative specifications for misspellings. First, I consider the square root of the number of misspellings, *SQRT_MSPL*, to account for the possibility that successive misspellings are decreasingly consequential. Second, I consider the percentage of words misspelled, *PCT_MSPL*, to account for the possibility that misspellings relative to the size of the loan description (in total number of words) is what influences lenders' perceptions of borrower quality. On average, 1.0% of all words in a loan description are misspelled. In untabulated tests, I find that my results are qualitatively unaffected by the choice of misspelling specification.

I use the same module in Perl to obtain an overall word count for each loan description (*WORDS*) and to compute the Gunning-Fog Index (*FOG*) and Flesch-Kincaid grade level score (*KINCAID*), two measures that will be used as control variables in my tests. The mean (median) number of words in a loan description for the sample is 38 (26). The mean (median) Fog Index in the sample is 10.3 (9.8), indicating that the average loan description requires approximately 10 years of education to read. The mean (median) Flesch-Kincaid index for loan descriptions in the

[25] In an experimental study, Figueredo & Varnhagen (2005) find that homophonic spelling errors are judged less harshly than non-homophonic errors. If this also holds in peer-to-peer lending settings, then the effects of homophonic spelling errors are likely to be less consequential.

sample is 7.8 (8.3) indicating that an 8th grader could read the average loan description (Li, 2008).

Table 2 provides the univariate correlations for selected variables, including each dependent variable, each loan description characteristic, loan amount, interest rate, list month, employment length, monthly income and FICO score. In the univariate, I find that misspelling is strongly ($p<.001$) negatively correlated with funding rate, strongly positively correlated with default, and strongly negatively correlated with net yield to lenders. The size of the correlation between misspelling and funding rate is smaller than those of interest rate, employment length or income, suggesting that misspelling is not as much of a driver of lender demand as these other variables. On the other hand, the correlation between misspelling and default is among the largest univariate correlations between any variable and default, and it is the same as the correlation between interest rate and default. This suggests that misspelling is as accurate of a predictor of default as the result of Lending Club's algorithmic prediction of loan risk, the assigned interest rate.

I provide the results of a construct validity test of associations with my *MSPL* variable in Table 3. Borrowers are more likely to misspell if they write a longer description, or if the grade level at which they are writing is higher. I also find that misspellings are more likely for borrowers with higher interest rates and lower FICO scores. This provides some comfort that misspellings are measuring some dimension of low borrower quality. Along these lines, I find that misspellings are also negatively related to borrower income, yet they are also positively related to employment length. This could suggest that the type of person who is prone to misspell finds career advancement to be difficult.

Chapter IV: Research Design and Results

I use OLS regressions to test the effect of misspellings on the level of funding per day (my first proxy for lender demand) in a posted price setting. The regression includes each control variable included in Appendix A, as well as indicator variables for loan purpose (e.g., Credit Card Refinancing, Home Improvement), indicator variables for home ownership type (e.g., Own, Mortgage) and indicator variables for each combination of state, year and month in which a borrower applied for a loan (e.g., Alabama, 2009, March.) I use State-Year-Month fixed effects in order to control for potential heterogeneity among investor bases from month to month or from state to state. Standard errors are clustered by State-Year-Month[26]. In functional form:

$$FUND\_PER\_DAY = \beta_0 + \beta_1 MSPL + \sum \beta_i Controls + STATE\_YEAR\_MONTH\ FE + \varepsilon_{it}$$

The coefficient of interest in this model is β_1.

Table 4 includes the results of this regression without the loan description characteristics (1), with loan description characteristics other than misspelling (2), and the full model including misspellings (3). I find that misspellings are strongly negatively related to funding rate, with the marginal effect of each misspelling decreasing funding rate by $30 dollars per day. This represents a decrease of 2% relative to the median funding rate of $1,429 per day. I find that, as in Gao & Lin (2013), the Gunning-Fog index is negatively associated with lender demand, while the Flesch-Kincaid ease-of-read score is positively related to lender demand. The effect of misspelling on lender demand is similar in magnitude to the effects of a one-standard deviation

[26] This results in 2,927 clusters, allowing me to make valid inferences while simultaneously allowing for errors to be correlated across time and location.

increase in Fog Index ($24/day) or a decrease in the Flesch-Kincaid ease of read ($40/day), and indicates that each of the three affect lender demand in different ways. I also find that each additional word in a loan description increases funding rate by about $0.35 per day. This finding is in keeping with the results of Michels (2012) which showed that additional disclosures within the borrower's loan description were positively associated with lender demand.

With respect to loan characteristics, I find that interest rate is positively related to funding rate, which suggests that lenders in this market are more risk tolerant than the Lending Club pricing algorithm. The duration of the loans issued to the borrower is significantly positively related to funding rate, suggesting that lenders in this area prefer longer cash flow streams and are tolerant of duration risk. I find that loan amount is highly positively related to funding rate. Given the proxy of dollars funded per day, I believe this relation is mechanical.

As I would expect, among borrower characteristics, employment length and income are positively related to funding rate, while debt-to-income is negatively related to funding rate. However, a borrower having had his income verified by Lending Club is strongly statistically and economically negatively related to funding rate, with a decrease in funding rate of $226 per day. This could be because lenders are aware that Lending Club chooses those borrowers which are verified, most likely for reasons which indicate that those borrowers are less creditworthy.

Most of the borrower credit history variables show significant associations with funding rate in ways that I would expect based on common understanding of the effects of those variables on creditworthiness. Delinquencies in the prior two years, number of credit inquiries in the prior six months, number of public records and revolving credit balance each are considered detrimental to a borrower's creditworthiness, and each is negatively related to funding rate. Similarly, time since delinquency and the number of open accounts are considered positive indicators of creditworthiness, and each is positively related to funding rate. However, I find that credit history and FICO score are each negatively related to funding rate, while the percentage of available credit being utilized is positively related to funding rate. I believe these findings could be explained by lenders giving those borrowers with short credit history the benefit of the doubt relative to other creditors, lenders viewing FICO as inadequate in this setting given its “one-size-fits-all” approach, and lenders viewing high credit utilization as verification that the borrower will use the proceeds of the loan to pay off debt, as most borrowers claim.

In addition to the OLS regression over funding per day, I also use a set of Cox proportional hazard models over the number of days it takes to fund the loan in order to assess misspelling’s effects on lender demand. The models include each of the independent variables from the loan description characteristics, the loan characteristics, the borrower characteristics and the borrower’s credit history. I also control for the purpose of the loan and the borrower’s residential status. I use Year-Month fixed effects to control for potential heterogeneity in the lender base from month to month, and my standard errors are clustered by State-Year-Month.

The results of the Cox proportional hazard models are presented in Table 5 without the loan description characteristics (1), with loan description characteristics other than misspelling (2), and the full model including misspellings (3). Model coefficients are presented in such a way that negative numbers indicate a decrease in the rate of “failure” (in this case, full funding of the loan.) I find that each misspelling is associated with a 2.6% increase in the amount of time that it takes to fund a loan on average. I find that a one standard deviation decrease in Flesch-Kincaid ease of read is associated with a 1.9% increase in time to fund, but that Fog Index is not significantly associated with time to fund, in contrast with Gao & Lin (2013). I also find that a one standard deviation increase in the number of words written in the description is associated with a 1.7% decrease in the time it takes to fund the loan, consistent with the findings in Michels (2012).

With respect to loan characteristics, I find that interest rate is negatively related to time to fund, which further indicates that lenders in this market are risk tolerant. Loan duration is significantly positively related to time to fund, contradicting the results in the funding rate tests, and suggesting that lenders in this area are not tolerant of duration risk. I find that loan amount is highly positively related to time to fund. Because it would necessarily require more lender money to fill a larger loan, I believe this relation is largely mechanical.

Among borrower characteristics, employment length and income are negatively related to time to fund, as lenders prefer borrowers who exhibit strong, stable earning power. As in the funding rate test, a borrower having had his income verified by Lending Club is strongly statistically and economically positively related to time to fund, with an increase in time to fund of 25.7%. This suggests that lenders are aware that borrowers are chosen for verification based on their creditworthiness.

Most of the borrower credit history variables show significant relations with time to fund in ways that I would expect. Delinquencies in the prior two years, number of credit inquiries in the prior six months, number of public records and revolving credit balance each are considered detrimental to a borrower's creditworthiness, and each is positively related to time to fund. Similarly, time since delinquency, total number of accounts and the number of open accounts are considered positive indicators of creditworthiness, and each is negatively related to time to fund. Again, I find that credit history is positively related to time to fund, while the percentage of available credit being utilized is negatively related to time to fund. As suggested in the discussion of Table 4, I believe these findings could be explained by lenders giving those borrowers with short credit history the benefit of the doubt relative to other creditors, and lenders viewing high credit utilization as verification that the borrower will use the proceeds of the loan to pay off debt, as most borrowers claim.

To examine the relation between misspellings and default, I use OLS regressions with State-Year-Month clustering and fixed effects. I chose to use a linear probability model rather than a probit model in order to make the results more interpretable and to allow for the use of fixed effects.[27] The model can be represented as:

$$DEFAULT = \beta_0 + \beta_1 MSPL + \sum \beta_i Controls + STATE\_YEAR\_MONTH\ FE + \varepsilon_{it}$$

The coefficient of interest in this model is β_1.

Table 6 includes the results of this regression without the loan description characteristics (1), with loan description characteristics other than misspelling (2), and the full model including misspellings (3). I find that misspellings strongly predict future default. Each misspelling increases the marginal probability that a loan will default by 0.8%. The regressions show that, as in Gao & Lin (2013), Flesch-Kincaid ease-of-read is negatively related to default, while there is a weakly positive relation between Gunning-Fog index and default. The coefficient on misspellings in Model 3 represents an increase of 7% in the default rate per misspelling, relative to the average default rate of 11.8% across my sample. This effect is on the order of the effect of a one-standard deviation decrease in the Flesch-Kincaid ease of read (1.3%) or a one-standard deviation increase in Gunning-Fog index (1.0%). I find that the number of words in the loan description is statistically significantly negatively related to the incidence of default, which is consistent with the Michels (2012) finding that the number of textual disclosures is negatively related to default.

[27] Untabulated results from estimating default tests using probit without fixed effects are similar in economic and statistical significance.

Unsurprisingly, I find that interest rate and longer loan term are positively related to default. Each of these variables represents different aspects of loan risk, and is expected to be positively related to default. Consistent with the magnitudes of the univariate correlations in Table 2, I find that the economic effect of one misspelling on default rate (.008) is almost a quarter as large as that of a one standard deviation increase in interest rate (.036). Similarly unsurprisingly, among borrower characteristics, I find that income is negatively related to default.

Regarding borrower credit history, FICO is negatively associated with default while recent credit inquiries and revolving account utilization are positively associated with default. Each of these results is expected based on common perceptions of these variables' effects on creditworthiness. However, time since the most recent delinquency is strongly positively related to default.[28] This result is counterintuitive, and may be because recent delinquencies are impounded too strongly in variables such as FICO and interest rate in terms of predicting credit risk in this particular market.

In order to assess the effects of misspelling on both default and loss given default, I use OLS regressions to test the effect of misspellings on the net yield to lenders. The regression includes each control variable, indicator variables for loan purpose, indicator variables for home ownership type and State-Year-Month fixed effects. Standard errors are clustered by State-Year-Month. In functional form:

$$RETURNS = \beta_0 + \beta_1 MSPL + \sum \beta_i Controls + STATE\_YEAR\_MONTH\ FE + \varepsilon_{it}$$

The coefficient of interest in this model is β_1.

[28] As *TIME_SINCE_DELINQ* is an ordered categorical variable, I denoted those loans with no delinquencies reported as having the highest value possible, therefore this result is not an artifact of those without delinquencies being removed.

Table 7 includes the results of this regression without the loan description characteristics (1), with loan description characteristics other than misspelling (2), and the full model including misspellings (3). I find that misspellings are strongly predictive of lower returns. On average, the marginal effect of each misspelling on lender yield is a decrease of 0.64%. This represents a relative decrease of 15% relative to the median return of 4.2% across my sample. I find that, consistent with Gao & Lin (2013), the Gunning-Fog index is negatively associated with lender yield, while the Flesch-Kincaid ease-of-read score is positively related to lender yield. The magnitude of the effect of one misspelling on lender yield is approximately half of the effect of a one-standard deviation increase in Fog Index (1.3%) or a one-standard deviation decrease in the Flesch-Kincaid ease of read (1.5%). Consistent with Michels (2012) findings that additional textual disclosures are associated with improved loan outcomes, I find that each additional word written in a loan description predicts an additional 1.3 basis points in lender yield.

With respect to loan characteristics, I find that loan duration is negatively related to lender yield. This suggests that Lending Club's pricing algorithm over the sample time period underweighted duration risk. I also find that interest rate is strongly related to lender yield, which suggests an appropriate relation between risk and return. As one might expect, borrower monthly income is strongly positively associated with lender yield, and debt-to-income and income verification are strongly negatively associated with lender yield.

Several of the borrower's credit history variables show coefficients that one might expect based on the common perceptions of their effects on creditworthiness. For instance, the number of recent inquiries and the number of public records are strongly negatively related to investor yield.

Chapter V: Additional Analysis

In order to investigate which loan listing characteristics affect the relations between misspelling and loan demand, and misspelling and loan outcomes, I perform a number of cross-sectional partitions on the Lending Club data. I partition the data into top- and bottom-terciles for each of the following variables: interest rate, FICO, length of credit history, total number of credit accounts ever held by the borrower, employment length and income. Over each of these partitions, I perform the four empirical tests noted in Section IV of this paper, namely OLS regressions with State-Year-Month clustering and fixed effects for *FUND_PER_DAY*, *DEFAULT* and *RETURNS*, and a Cox Proportional Hazard Model with State-Year-Month clustering and Year-Month fixed effects for *FUND_DAYS*. The results of these tests partitioned by employment length, income and interest rate are in Table 8, Table 9 and Table 10, respectively, while the remaining tests are untabulated for parsimony.

With regard to lender demand, the cross-sectional tests reveal that misspellings negatively impact lender demand regardless of the situation, as evidence in each of Table 8, Table 9 and Table 10 suggests. This is consistent with the view that, from an *ex ante* lender's perspective, the misspellings are perceived with a "reverse halo effect" (Nisbett & Wilson, 1977.) On average, the lenders do not know specifically *why* they view the misspeller's loan listing as less desirable, just that in all circumstances, they *do* view the misspeller's loan listing as less desirable. Alternatively, the consistency with which misspellings decrease lender demand could be because each lender perceives the misspeller to be flawed from a credit risk standpoint in the lender's own way.

In contrast to the findings specific to lender demand, I find several key factors which attenuate or exacerbate the predictive power of misspellings in terms of loan outcomes. The relations between misspelling and default and between misspelling and lender yield are attenuated by the ability of the borrower to keep a job and generate income. Table 8 shows that for borrowers with a long employment history at their current job, misspelling is not predictive of default or decreased lender yield. Similarly, Table 9 suggests that high levels of borrower income mitigate the incremental contribution of misspelling to actual credit risk, as there is no relation between misspelling and default or misspelling and lender yields in the high income group. These findings suggest that misspelling reveals something about the borrower's future ability to pay off his debt (e.g., employability, intelligence, education and conscientiousness) more so than the borrower's willingness to do so (e.g., credibility).

In a somewhat contrasting finding, Table 10 shows that the effect of misspelling on loan outcome variables is modulated by the interest rate. These tests show that misspellings predict loan outcomes only when the borrower's interest rate is relatively low. That misspellings only predict creditworthiness for those loans with lower interest rates may suggest that Lending Club's underwriting process may be partially selecting on the dimension of creditworthiness that misspelling captures, by offering riskier borrowers less risky loans. Alternatively, these findings could indicate that the underwriting process is capturing the dimension of creditworthiness revealed through misspelling only for some borrowers, and therefore inappropriately providing low rates for other borrowers who misspell.

Chapter VI: Conclusion

In this study, I find that misspelling in the loan description section of a borrower's loan listing is associated with a decrease in funding rate from lenders, an increase in the time it takes to fully fund a loan, an increased incidence of default, and lower net yield to lenders. These findings are consistent with my hypotheses that misspellings lead to decreased lender demand, and that misspellings are a signal of low borrower quality, distinct from the hard information provided in a credit profile. These results are also consistent with prior work in social psychology which suggests that misspellings can affect perceptions of conscientiousness, competence and employability, or even that misspellings exhibit a "reverse halo effect." The results of additional analyses also suggest that the relation between misspellings and loan outcomes is driven by characteristics which affect ability to pay (e.g., employability, intelligence, education and conscientiousness), rather than willingness to pay (e.g., credibility).

To my knowledge my paper is the first to demonstrate that the lexical properties of writing in financial disclosures signal the communicator's quality, rather than merely impeding communication. My paper is also the first study to establish that misspellings predict the future performance of the misspelling party. Given that prior research has established that misspellings affect perceptions of misspellers, this finding fills an important gap in the spelling literature. This paper contributes to the literature on peer-to-peer lending by demonstrating that the information transmitted by misspelling is useful; misspellings help alleviate information asymmetry. With regard to the household finance literature, my work shows that proper spelling is important in retail lending contexts, for both borrowers and lenders. Finally, although my study is set in the distinctive peer-to-peer lending setting, my findings suggest that misspellings in disclosures are consequential in settings where information asymmetries are important and reliable information is limited.

Figures

Figure 1: Example Loan Listing

Other

Help | Lending Club Account

Borrower Member Loan 1044298 | Lending Club Prospectus

Amount Requested	$16,000	Funding Received	$16,000 (100.00% funded)
Loan Purpose	Other	Investors	149 people funded this loan
Loan Grade	A4	Note Status	Current
Interest Rate	7.90%	Loan Submitted on	11/27/11 2:59 PM
Loan Length	3 years (36 payments)		
Monthly Payment	$500.65 / month		

Member_1274662's Profile (all information not verified unless noted with an "*")

Home Ownership	RENT	Gross Income	$2,500 / month *
Current Employer	MORE MAZDA	Debt-to-Income (DTI)	17.88%
Length of Employment	5 years	Location	PHOENIX, AZ

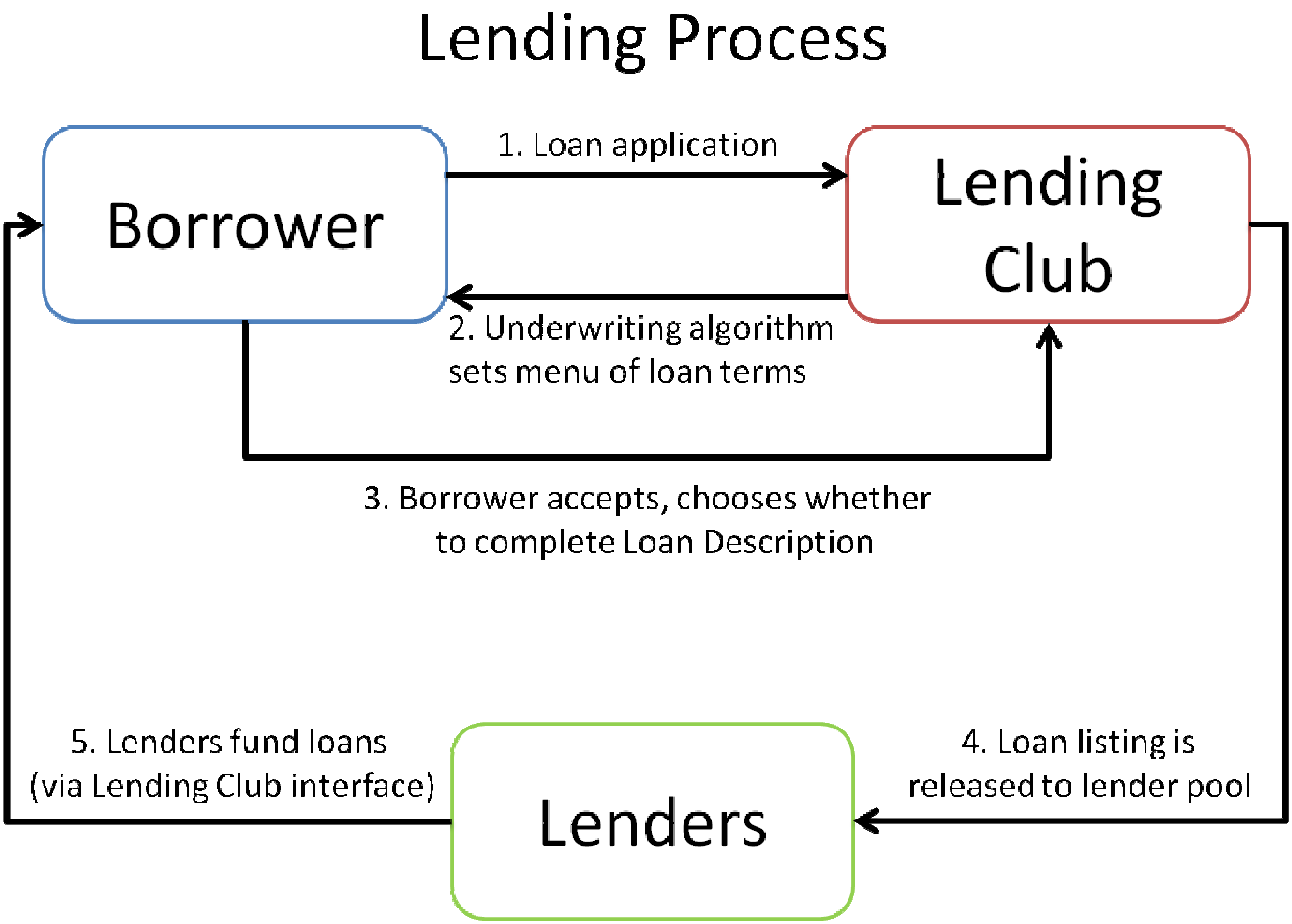
Member_1274662's Credit History (as reported by credit bureau on 11/27/11)

Credit Score Range:	720-724	Delinquent Amount	$0.00
Earliest Credit Line	12/2000	Delinquencies (Last 2 yrs)	0
Open Credit Lines	7	Months Since Last Delinquency	58
Total Credit Lines	12	Public Records On File	0
Revolving Credit Balance	$13,932.00	Months Since Last Record	n/a
Revolving Line Utilization	53.20%	Months Since Last Major Derogatory	n/a
Inquiries in the Last 6 Months	0	Collections Excluding Medical	0
Accounts Now Delinquent	0		

Loan Description

Borrower added on 11/27/11 > Pay off credit cards, make one monthly payment and decrease maximum amt payed overall. Have payed all previous loans on time.

Figure 2: Lending Club Loan Process Diagram



Tables

Table 1
Summary Statistics

Variable	All Sample Loans N	Mean	P25	Median	P75	Std Dev	Seasoned Loans Only N	Mean	P25	Median	P75	Std Dev
Dependent Variables												
FUND_PER_DAY	103,344	$1,890	$850	$1,429	$2,400	$1,680	25,700	$1,326	$686	$1,120	$1,678	$1,032
FUND_DAYS	103,344	9.08	6	8	12	4.65	25,700	9.17	6	9	13	4.30
DEFAULT	103,344	5.2%	0	0	0	22.3%	25,700	11.8%	0	0	0	32.2%
RETURNS	103,344	9.6%	9.6%	13.4%	17.2%	21.6%	25,700	4.2%	7.1%	11.7%	15.6%	27.5%
Loan Characteristics												
LOAN_AMT	103,344	13.750	7.750	12.000	19.000	7.964	25,700	11.500	6.000	10.000	15.000	7.375
60_MONTH	103,344	22.8%	0	0	0	42.0%	25,700	25.9%	0	0	1	43.8%
INT_RATE	103,344	13.4%	10.2%	13.1%	16.2%	4.3%	25,700	12.0%	9.3%	11.9%	14.3%	3.6%
LIST_MONTH	103,344	61.0	55	65	72	14.8	25,700	39.3	32	42	49	12.0
Borrower Characteristics												
EMP_LENGTH	103,344	6.5	3	6	11	3.8	25,700	5.7	3	5	9	3.6
INCOME	103,344	5.962	3.750	5.083	7.083	4.513	25,700	5.781	3.417	5.000	6.917	4.968
VERIFIED	103,344	63.2%	0	1	1	48.2%	25,700	52.7%	0	1	1	49.9%
DTI	103,344	16.1%	10.5%	15.9%	21.4%	7.5%	25,700	13.3%	8.1%	13.4%	18.6%	6.7%
Borrower Credit History												
DELINQUENT	103,344	0%	0	0	0	4%	25,700	0%	0	0	0	0%
DELINQ_AMT	103,344	0.005	0	0	0	0.431	25,700	0.000	0	0	0	0.000
DELINQ_2YRS	103,344	0.20	0	0	0	0.62	25,700	0.14	0	0	0	0.48
TIME_SINCE_DELINQ	103,344	9.9	4	14	14	5.3	25,700	10.4	5	14	14	5.1
CREDIT_HIST	103,344	180	124	164	221	84	25,700	164	108	149	203	81
FICO	103,344	703	680	695	720	33	25,700	716	685	710	740	36
INQUIRIES	103,344	0.83	0	0	1	1.05	25,700	0.89	0	1	1	1.09
TIME_SINCE_RECORD	103,344	11.7	12	12	12	1.2	25,700	11.8	12	12	12	1.0
PUB_RECORD	103,344	0.08	0	0	0	0.31	25,700	0.06	0	0	0	0.24
REV_BAL	103,344	15.816	6.353	11.788	20.065	19.100	25,700	13.590	3.771	8.961	17.286	16.041
REV_UTIL	103,344	55.7%	37.9%	58.2%	75.7%	25.0%	25,700	48.1%	24.8%	48.2%	71.4%	28.2%
OPEN_ACC	103,344	10.6	7	10	13	4.6	25,700	9.4	6	9	12	4.4
TOTAL_ACC	103,344	24.0	16	22	31	11.2	25,700	22.2	14	21	29	11.4
Loan Description Characteristics												
MSPL	103,344	0.28	0	0	0	0.78	25,700	0.50	0	0	1	1.13
SQRT_MSPL	103,344	0.22	0	0	0	0.49	25,700	0.35	0	0	1	0.61
PCT_MSPL	103,344	1.0%	0	0	0	4.0%	25,700	1.0%	0	0	0.7%	3.6%
WORDS	103,344	38.0	11	26	50	44.2	25,700	65.1	18	44	86.5	70.7
FOG	103,344	10.3	7.9	9.8	12.7	5.3	25,700	10.2	8	10	12.6	5.0
KINCAID	103,344	8.3	5.6	7.8	10.2	4.8	25,700	8.2	5.9	8	10.2	4.3

FUND_PER_DAY is the amount, in dollars per day, invested in the loan by investors through LendingClub's website. *FUND_DAYS* is the number of days that it takes the loan to be fully funded. *DEFAULT* is an indicator variable for whether the loan ended in default. *RETURNS* is net yield to the investors (before fees) in percentage terms. *LOAN_AMT* is the size of the loan in thousands of dollars. *60_MONTH* is an indicator variable for whether the loan is for 60 months. 0 indicates a 36 month loan. *INT_RATE* is the stated interest rate of the loan, in percentage terms. *LIST_MONTH* is months since beginning of data set at time of application. *EMP_LENGTH* is an ordered categorical variable for length of employment at current employer at time of application. *INCOME* is borrower submitted monthly income in thousands of dollars per month. *VERIFIED* is an indicator of whether borrower employment information is confirmed by LendingClub. *DTI* is debt-to-income ratio at time of application. *DELINQUENT* is an indicator of whether any of borrowers credit accounts are delinquent at time of application. *DELINQ_AMT* is the amount of debt on borrower's credit report that is delinquent at time of application. *DELINQ_2YRS* is the number of delinquencies reported on borrower's credit report within the last two years at the time of application. *TIME_SINCE_DELINQ* is an ordered categorical variable for time since last delinquency on borrower's credit report at time of application, where lower numbers indicate more recent delinquency. *CREDIT_HIST* is the length of borrower credit history, in months at time of application. *FICO* is borrower's FICO score at time of application. *INQUIRIES* is the number of hard credit inquiries on borrower's credit report within the last six months at time of application. *PUB_RECORD* is the number of public records on borrower's credit report at time of application. *TIME_SINCE_RECORD* is an ordered categorical variable for time since last public record on borrower's credit report at time of application, where lower numbers indicate more recent record. *REV_BAL* is the amount of revolving debt in thousands of dollars on borrower's credit report at time of application. *REV_UTIL* is the percentage of credit limit on revolving accounts being used by borrower at time of application. *OPEN_ACC* is the number of open credit accounts held by the borrower at time of application. *TOTAL_ACC* is the number of credit accounts ever held by the borrower at time of application. *MSPL* is the number of misspellings within the text of the borrower submitted loan description. *SQRT_MSPL* is the square root of the number of misspellings within the text of the borrower submitted loan description. *PCT_MSPL* is the number of misspellings within the text of the borrower submitted loan description divided by the total number of words in the description. *WORDS* is the total number of words in the borrower submitted loan description. *FOG* is Fog Index of the borrower submitted loan description. *KINCAID* is the Flesch-Kincaid ease of reading index, which is increasing in sentence length and syllables per word.

Table 2
Univariate Correlations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) *FUND_PER_DAY*	1.000	**-0.525**	**-0.077**	**0.178**	**-0.043**	**-0.021**	**0.047**	**0.051**	**0.680**	**0.186**	**0.307**	**0.166**	**0.352**	**-0.018**
(2) *FUND_DAYS*	**-0.473**	1.000	**0.055**	**-0.021**	**0.037**	**0.029**	0.009	0.004	**0.195**	-0.008	**-0.153**	**-0.029**	**0.045**	**0.087**
(3) *DEFAULT*	**-0.066**	**0.048**	1.000	**-0.376**	**0.053**	**0.051**	0.000	-0.003	**-0.025**	**0.064**	**-0.204**	**-0.032**	**-0.062**	**-0.032**
(4) *RETURNS*	**0.096**	**-0.043**	**-0.904**	1.000	**-0.018**	**-0.049**	0.000	0.004	**0.190**	**0.811**	**0.166**	**0.042**	**0.022**	**-0.534**
(5) *MSPL*	**-0.042**	**0.027**	**0.058**	**-0.046**	1.000	**0.281**	**0.038**	**0.045**	-0.010	-0.006	**-0.156**	0.001	**-0.023**	**0.023**
(6) *WORDS*	**-0.062**	**0.028**	**0.066**	**-0.054**	**0.361**	1.000	**0.115**	**0.143**	**0.012**	**-0.061**	**-0.297**	**-0.101**	**-0.021**	**0.072**
(7) *FOG*	**0.033**	0.010	-0.002	0.003	**0.041**	**0.080**	1.000	**0.918**	**0.066**	-0.003	**-0.012**	**-0.039**	**0.087**	**0.026**
(8) *KINCAID*	**0.043**	0.004	-0.006	0.008	**0.043**	**0.083**	**0.922**	1.000	**0.066**	-0.003	**-0.021**	**-0.050**	**0.078**	**0.026**
(9) *LOAN_AMT*	**0.547**	**0.178**	**-0.022**	**0.075**	-0.010	**-0.013**	**0.056**	**0.062**	1.000	**0.214**	**0.190**	**0.167**	**0.472**	**0.072**
(10) *INT_RATE*	**0.191**	0.006	**0.058**	**0.161**	-0.008	**-0.057**	**0.011**	**0.012**	**0.268**	1.000	**0.171**	**0.048**	-0.003	**-0.669**
(11) *LIST_MONTH*	**0.278**	**-0.100**	**-0.195**	**0.169**	**-0.150**	**-0.347**	**0.018**	**0.018**	**0.194**	**0.190**	1.000	**0.138**	**0.090**	**-0.211**
(12) *EMP_LENGTH*	**0.131**	**-0.025**	**-0.033**	**0.035**	-0.004	**-0.097**	**-0.020**	**-0.024**	**0.160**	**0.051**	**0.148**	1.000	**0.186**	-0.009
(13) *INCOME*	**0.187**	**0.046**	**-0.039**	**0.039**	**-0.016**	**-0.016**	**0.052**	**0.046**	**0.340**	0.007	**0.040**	**0.106**	1.000	**0.094**
(14) *FICO*	**-0.031**	**0.060**	**-0.032**	**-0.107**	**0.024**	**0.076**	**0.018**	**0.010**	**0.057**	**-0.636**	**-0.208**	-0.009	**0.079**	1.000

Bolded correlation coefficients are significant at the p<.001 level.

FUND_PER_DAY is the amount, in dollars per day, invested in the loan by investors through LendingClub's website. *FUND_DAYS* is the number of days that it takes the loan to be fully funded. *DEFAULT* is an indicator variable for whether the loan ended in default. *RETURNS* is net yield to the investors (before fees) in percentage terms. *MSPL* is the number of misspellings within the text of the borrower submitted loan description. *WORDS* is the total number of words in the borrower submitted loan description. *FOG* is Fog Index of the borrower submitted loan description. *KINCAID* is the Flesch-Kincaid ease of reading index, which is increasing in sentence length and syllables per word. *LOAN_AMT* is the size of the loan in thousands of dollars. *INT_RATE* is the stated interest rate of the loan, in percentage terms. *LIST_MONTH* is months since beginning of data set at time of application. *EMP_LENGTH* is an ordered categorical variable for length of employment at current employer at time of application. *INCOME* is borrower submitted monthly income in thousands of dollars per month. *FICO* is borrower's FICO score at time of application.

Table 3
Determinants of Misspelling

VARIABLES	*Dependent Variable: MSPL*
Loan Description Characteristics	
WORDS	0.006***
	(37.96)
FOG	-0.000
	(-0.48)
KINCAID	0.003***
	(3.01)
Loan Characteristics	
LOAN_AMT	-0.000
	(-1.00)
60_MONTH	0.016**
	(2.08)
INT_RATE	0.004***
	(4.38)
Borrower Characteristics	
EMP_LENGTH	0.007***
	(11.85)
INCOME	-0.002***
	(-3.12)
VERIFIED	-0.001
	(-0.21)
DTI	-0.000
	(-0.31)
Borrower Credit History	
DELINQUENT	-0.058**
	(-2.46)
DELINQ_AMT	-0.005***
	(-3.42)
DELINQ_2YRS	-0.005
	(-1.25)
TIME_SINCE_DELINQ	-0.000
	(-0.75)
CREDIT_HIST	0.000***
	(4.48)
FICO	-0.000***
	(-4.04)
INQUIRIES	0.001
	(0.52)
PUB_RECORD	0.020
	(1.41)
TIME_SINCE_RECORD	0.003
	(0.85)
REV_BAL	-0.000
	(-0.61)
REV_UTIL	-0.103***
	(-7.30)
OPEN_ACC	-0.004***
	(-5.98)
TOTAL_ACC	0.001*
	(1.87)
State-Year-Month Fixed Effects	Yes
Observations	103,344
Pseudo R-squared	0.106

Year-Month clustered z-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 4
Investor Funding per Day and Misspellings

VARIABLES	*Dependent Variable: FUND_PER_DAY* (1)	(2)	(3)
Loan Description Characteristics			
MSPL			-29.563***
			(-6.10)
WORDS		0.172	0.354***
		(1.58)	(2.96)
FOG		-4.518**	-4.573**
		(-2.38)	(-2.41)
KINCAID		8.239***	8.367***
		(3.82)	(3.88)
Loan Characteristics			
LOAN_AMT	110.618***	110.457***	110.461***
	(66.81)	(66.99)	(67.01)
60_MONTH	37.371*	36.170*	36.615*
	(1.70)	(1.65)	(1.67)
INT_RATE	7.980***	8.112***	8.194***
	(3.20)	(3.25)	(3.29)
Borrower Characteristics			
EMP_LENGTH	10.744***	10.986***	11.185***
	(8.76)	(8.97)	(9.13)
INCOME	2.082*	2.027*	1.960
	(1.74)	(1.70)	(1.64)
VERIFIED	-226.023***	-225.905***	-226.089***
	(-17.98)	(-17.96)	(-17.98)
DTI	-2.833***	-2.789***	-2.775***
	(-4.49)	(-4.42)	(-4.40)
Borrower Credit History			
DELINQUENT	-89.681	-90.508	-92.243
	(-0.92)	(-0.93)	(-0.95)
DELINQ_AMT	-0.608	-0.356	-0.512
	(-0.11)	(-0.06)	(-0.09)
DELINQ_2YRS	-26.709***	-27.152***	-27.319***
	(-3.44)	(-3.50)	(-3.51)
TIME_SINCE_DELINQ	2.411**	2.403**	2.403**
	(2.47)	(2.46)	(2.46)
CREDIT_HIST	-0.433***	-0.428***	-0.425***
	(-7.96)	(-7.89)	(-7.83)
FICO	-0.588**	-0.584**	-0.605**
	(-2.35)	(-2.33)	(-2.41)
INQUIRIES	-16.870***	-16.916***	-16.940***
	(-3.74)	(-3.75)	(-3.76)
PUB_RECORD	-75.705***	-75.791***	-75.261***
	(-2.60)	(-2.60)	(-2.58)
TIME_SINCE_RECORD	4.820	4.715	4.798
	(0.64)	(0.63)	(0.64)
REV_BAL	-1.506***	-1.510***	-1.515***
	(-6.02)	(-6.04)	(-6.07)
REV_UTIL	138.836***	138.415***	135.769***
	(6.16)	(6.14)	(6.02)
OPEN_ACC	10.164***	10.087***	9.975***
	(7.27)	(7.22)	(7.15)
TOTAL_ACC	0.439	0.457	0.468
	(0.78)	(0.81)	(0.83)
Purpose Indicators	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes
State-Year-Month Fixed Effects	Yes	Yes	Yes
Observations	103,344	103,344	103,344
R-squared	0.304	0.304	0.304

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 5
Hazard Model: Days to Fully Fund and Misspellings

	Dependent Variable: FUND_DAYS		
VARIABLES	**(1)**	**(2)**	**(3)**
Loan Description Characteristics			
MSPL			-0.026***
			(-6.89)
WORDS		0.000***	0.000***
		(3.85)	(5.80)
FOG		-0.002	-0.002
		(-1.45)	(-1.48)
KINCAID		0.004**	0.004**
		(2.00)	(2.07)
Loan Characteristics			
LOAN_AMT	-0.020***	-0.021***	-0.021***
	(-25.95)	(-26.23)	(-26.23)
60_MONTH	-0.041**	-0.042**	-0.041**
	(-2.51)	(-2.55)	(-2.55)
INT_RATE	0.011***	0.011***	0.011***
	(6.74)	(6.81)	(6.86)
Borrower Characteristics			
EMP_LENGTH	0.009***	0.009***	0.009***
	(9.71)	(9.93)	(10.12)
INCOME	0.003***	0.003***	0.003***
	(4.60)	(4.61)	(4.55)
VERIFIED	-0.297***	-0.297***	-0.297***
	(-27.97)	(-27.95)	(-27.98)
DTI	-0.001	-0.001	-0.001
	(-1.55)	(-1.54)	(-1.52)
Borrower Credit History			
DELINQUENT	-0.023	-0.022	-0.024
	(-0.32)	(-0.30)	(-0.32)
DELINQ_AMT	0.001	0.001	0.001
	(0.22)	(0.21)	(0.18)
DELINQ_2YRS	-0.025***	-0.026***	-0.026***
	(-3.93)	(-4.00)	(-4.00)
TIME_SINCE_DELINQ	0.002**	0.002**	0.002**
	(2.57)	(2.56)	(2.57)
CREDIT_HIST	-0.000***	-0.000***	-0.000***
	(-8.81)	(-8.64)	(-8.59)
FICO	0.000	0.000	-0.000
	(0.05)	(0.06)	(-0.05)
INQUIRIES	-0.006*	-0.006*	-0.006*
	(-1.90)	(-1.87)	(-1.89)
PUB_RECORD	-0.073***	-0.074***	-0.073***
	(-3.27)	(-3.28)	(-3.25)
TIME_SINCE_RECORD	-0.012**	-0.013**	-0.012**
	(-2.20)	(-2.22)	(-2.19)
REV_BAL	-0.001***	-0.001***	-0.001***
	(-3.94)	(-3.95)	(-3.95)
REV_UTIL	0.076***	0.075***	0.072***
	(4.31)	(4.27)	(4.11)
OPEN_ACC	0.004***	0.004***	0.004***
	(4.18)	(4.11)	(3.99)
TOTAL_ACC	0.002***	0.002***	0.002***
	(4.83)	(4.84)	(4.86)
Purpose Indicators	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes
Observations	103,344	103,344	103,344
Pseudo R-squared	0.007	0.007	0.007

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 6
Default and Misspellings

VARIABLES	*Dependent Variable: DEFAULT* (1)	(2)	(3)
Loan Description Characteristics			
MSPL			0.008***
			(3.62)
WORDS		-0.000***	-0.000***
		(-4.12)	(-5.29)
FOG		0.002*	0.002*
		(1.72)	(1.76)
KINCAID		-0.003**	-0.003**
		(-2.27)	(-2.37)
Loan Characteristics			
LOAN_AMT	0.000	0.000	0.000
	(0.62)	(0.97)	(0.98)
60_MONTH	0.027***	0.028***	0.027***
	(3.88)	(3.94)	(3.89)
INT_RATE	0.010***	0.010***	0.010***
	(7.68)	(7.65)	(7.60)
Borrower Characteristics			
EMP_LENGTH	0.000	0.000	0.000
	(0.71)	(0.40)	(0.24)
INCOME	-0.004***	-0.004***	-0.004***
	(-5.55)	(-5.58)	(-5.55)
VERIFIED	0.006	0.007	0.007
	(1.28)	(1.40)	(1.48)
DTI	0.001	0.001	0.001
	(1.46)	(1.41)	(1.38)
Borrower Credit History			
DELINQ_2YRS	0.001	0.001	0.002
	(0.27)	(0.28)	(0.30)
TIME_SINCE_DELINQ	0.001**	0.001**	0.001**
	(2.51)	(2.52)	(2.54)
CREDIT_HIST	0.000*	0.000	0.000
	(1.75)	(1.55)	(1.50)
FICO	-0.000*	-0.000*	-0.000*
	(-1.67)	(-1.70)	(-1.66)
INQUIRIES	0.019***	0.018***	0.018***
	(8.51)	(8.42)	(8.40)
PUB_RECORD	0.020	0.020	0.020
	(0.80)	(0.79)	(0.80)
TIME_SINCE_RECORD	-0.004	-0.004	-0.004
	(-0.68)	(-0.68)	(-0.67)
REV_BAL	-0.000	-0.000	-0.000
	(-1.26)	(-1.26)	(-1.27)
REV_UTIL	0.027**	0.027**	0.028**
	(2.34)	(2.38)	(2.47)
OPEN_ACC	0.001	0.001	0.001
	(0.80)	(0.90)	(1.01)
TOTAL_ACC	-0.000	-0.000	-0.000
	(-1.24)	(-1.36)	(-1.41)
Purpose Indicators	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes
State-Year-Month Fixed Effects	Yes	Yes	Yes
Observations	25,700	25,700	25,700
R-squared	0.044	0.045	0.046

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 7
Returns and Misspellings

	Dependent Variable: RETURNS		
VARIABLES	**(1)**	**(2)**	**(3)**
Loan Description Characteristics			
MSPL			-0.639***
			(-3.16)
WORDS		0.009***	0.013***
		(3.48)	(4.54)
FOG		-0.256**	-0.259**
		(-2.44)	(-2.48)
KINCAID		0.341***	0.351***
		(2.93)	(3.02)
Loan Characteristics			
LOAN_AMT	0.011	0.000	0.000
	(0.31)	(0.01)	(0.00)
60_MONTH	-2.899***	-2.952***	-2.926***
	(-4.41)	(-4.48)	(-4.45)
INT_RATE	0.223*	0.228*	0.234**
	(1.88)	(1.92)	(1.97)
Borrower Characteristics			
EMP_LENGTH	0.006	0.021	0.029
	(0.11)	(0.38)	(0.52)
INCOME	0.295***	0.296***	0.293***
	(5.03)	(5.04)	(5.02)
VERIFIED	-0.639	-0.683*	-0.713*
	(-1.54)	(-1.65)	(-1.72)
DTI	-0.054*	-0.052*	-0.052*
	(-1.74)	(-1.68)	(-1.67)
Borrower Credit History			
DELINQ_2YRS	-0.114	-0.114	-0.122
	(-0.25)	(-0.25)	(-0.27)
TIME_SINCE_DELINQ	-0.149***	-0.150***	-0.151***
	(-3.04)	(-3.06)	(-3.08)
CREDIT_HIST	-0.005*	-0.004*	-0.004*
	(-1.93)	(-1.73)	(-1.69)
FICO	0.005	0.006	0.005
	(0.44)	(0.47)	(0.44)
INQUIRIES	-1.503***	-1.490***	-1.490***
	(-8.02)	(-7.95)	(-7.93)
PUB_RECORD	-4.903**	-4.854**	-4.873**
	(-2.29)	(-2.26)	(-2.27)
TIME_SINCE_RECORD	-0.658	-0.654	-0.659
	(-1.43)	(-1.41)	(-1.43)
REV_BAL	0.010	0.010	0.010
	(0.75)	(0.74)	(0.74)
REV_UTIL	-2.236**	-2.262**	-2.339**
	(-2.09)	(-2.12)	(-2.19)
OPEN_ACC	-0.046	-0.052	-0.057
	(-0.72)	(-0.81)	(-0.89)
TOTAL_ACC	0.020	0.023	0.024
	(0.82)	(0.93)	(0.98)
Purpose Indicators	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes
State-Year-Month Fixed Effects	Yes	Yes	Yes
Observations	25,700	25,700	25,700
R-squared	0.017	0.018	0.018

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 8

Effects of Misspelling - Partitioned by Employment Length

Dependent Variable:	***FUND_PER_DAY***		***FUND_DAYS***		***DEFAULT***		***RETURNS***	
Partitioning Variable:	***EMP_LENGTH***		***EMP_LENGTH***		***EMP_LENGTH***		***EMP_LENGTH***	
VARIABLES	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**	**Long**	**Short**
Loan Description Characteristics								
MSPL	-23.290**	-17.295**	-0.021***	-0.018***	0.006	0.010***	-0.277	-0.645**
	(-2.42)	(-2.37)	(-3.39)	(-2.92)	(1.09)	(2.83)	(-0.58)	(-2.12)
WORDS	-0.034	0.167	0.000*	0.000***	-0.000***	-0.000***	0.012*	0.016***
	(-0.15)	(1.42)	(1.81)	(3.15)	(-2.91)	(-4.12)	(1.66)	(3.76)
FOG	-6.021*	-2.745	-0.002	-0.002	0.001	0.000	-0.273	-0.102
	(-1.67)	(-0.89)	(-0.87)	(-0.81)	(0.58)	(0.20)	(-1.31)	(-0.61)
KINCAID	8.409**	6.554*	0.003	0.004	-0.002	-0.001	0.301	0.145
	(2.00)	(1.92)	(0.82)	(1.27)	(-0.57)	(-0.60)	(1.22)	(0.78)
Loan Characteristics								
LOAN_AMT	113.394***	109.121***	-0.019***	-0.021***	0.001	0.000	-0.060	0.027
	(51.82)	(56.56)	(-19.00)	(-17.13)	(1.45)	(0.40)	(-0.77)	(0.44)
60_MONTH	59.024**	-3.331	0.012	-0.097***	0.042***	0.011	-3.872***	-1.969*
	(2.02)	(-0.11)	(0.61)	(-4.18)	(2.96)	(0.97)	(-3.17)	(-1.91)
INT_RATE	12.500***	4.953	0.013***	0.009***	0.007**	0.012***	0.293	0.101
	(3.43)	(1.46)	(5.65)	(3.47)	(2.39)	(5.86)	(1.18)	(0.56)
Borrower Characteristics								
EMP_LENGTH	26.622	70.735***	-0.017	0.058***	0.033**	-0.006**	-2.413**	0.590***
	(1.05)	(14.68)	(-0.99)	(13.23)	(2.58)	(-2.30)	(-2.16)	(2.63)
INCOME	0.693	4.105	0.002**	0.004**	-0.004***	-0.004***	0.331***	0.351***
	(0.49)	(1.31)	(2.04)	(2.35)	(-4.00)	(-3.36)	(3.74)	(2.81)
VERIFIED	-230.034***	-208.347***	-0.278***	-0.306***	-0.000	0.008	0.014	-0.633
	(-11.98)	(-12.95)	(-18.70)	(-20.44)	(-0.04)	(1.05)	(0.01)	(-0.96)
DTI	-1.570	-3.266***	0.000	-0.001	0.000	0.001	-0.019	-0.079
	(-1.33)	(-3.45)	(0.16)	(-1.47)	(0.16)	(1.49)	(-0.28)	(-1.52)
Borrower Credit History								
DELINQUENT	-124.045	23.001	-0.038	-0.029				
	(-0.77)	(0.11)	(-0.37)	(-0.19)				
DELINQ_AMT	8.987	-12.367*	0.001	-0.001				
	(0.77)	(-1.95)	(0.09)	(-0.14)				
DELINQ_2YRS	-45.768***	-6.460	-0.032***	-0.006	0.004	0.008	-0.276	-0.832
	(-3.61)	(-0.46)	(-2.94)	(-0.61)	(0.40)	(1.02)	(-0.29)	(-1.10)
TIME_SINCE_DELINQ	4.102**	3.102**	0.000	0.003***	0.001	0.001*	-0.144	-0.155**
	(2.02)	(2.02)	(0.37)	(2.61)	(1.20)	(1.74)	(-1.30)	(-1.99)
CREDIT_HIST	-0.432***	-0.280***	-0.000***	-0.000***	-0.000**	0.000**	0.011*	-0.007*
	(-3.92)	(-3.35)	(-4.59)	(-4.04)	(-2.28)	(2.04)	(1.86)	(-1.76)
FICO	-0.206	-0.810**	0.001**	-0.000	-0.000	-0.000	0.004	0.011
	(-0.48)	(-2.12)	(2.44)	(-0.47)	(-1.37)	(-1.48)	(0.17)	(0.57)
INQUIRIES	-25.959***	-9.647	-0.012**	-0.005	0.023***	0.019***	-1.629***	-1.661***
	(-3.05)	(-1.46)	(-2.17)	(-0.88)	(5.32)	(5.59)	(-4.13)	(-5.60)
PUB_RECORD	-95.253**	-36.053	-0.061*	-0.064	0.033	0.066	-5.204	-8.686**
	(-2.06)	(-0.68)	(-1.77)	(-1.44)	(0.73)	(1.48)	(-1.42)	(-2.49)
TIME_SINCE_RECORD	10.225	11.372	-0.008	-0.014	-0.000	0.003	-0.672	-1.282*
	(0.80)	(0.89)	(-0.91)	(-1.35)	(-0.03)	(0.28)	(-0.71)	(-1.89)
REV_BAL	-1.630***	-1.506***	-0.001***	-0.000	-0.000	-0.000	0.017	-0.004
	(-5.38)	(-2.87)	(-3.51)	(-1.15)	(-1.43)	(-0.50)	(0.75)	(-0.15)
REV_UTIL	163.069***	141.768***	0.127***	0.030	0.009	0.013	-0.878	-1.824
	(3.64)	(4.30)	(4.05)	(1.06)	(0.39)	(0.77)	(-0.40)	(-1.16)
OPEN_ACC	9.153***	5.640***	0.004**	0.001	0.000	0.001	0.033	-0.048
	(3.50)	(2.70)	(2.42)	(0.65)	(0.10)	(0.64)	(0.25)	(-0.47)
TOTAL_ACC	2.899***	0.432	0.003***	0.003***	-0.001*	-0.000	0.020	0.033
	(2.74)	(0.50)	(4.05)	(3.43)	(-1.89)	(-0.86)	(0.43)	(0.78)
Purpose Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	No	No	Yes	Yes	No	No	No	No
State-Year-Month Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	34,045	37,562	34,045	37,562	6,275	11,747	6,275	11,747
(Pseudo) R-squared	0.296	0.330	0.007	0.008	0.054	0.054	0.022	0.024

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Variable definitions can be found in Appendix A

Table 9
Effects of Misspelling - Partitioned by Borrower Income

Dependent Variable:	*FUND_PER_DAY*		*FUND_DAYS*		*DEFAULT*		*RETURNS*	
Partitioning Variable:	*INCOME*		*INCOME*		*INCOME*		*INCOME*	
VARIABLES	**High**	**Low**	**High**	**Low**	**High**	**Low**	**High**	**Low**
Loan Description Characteristics								
MSPL	-28.676**	-33.357***	-0.018***	-0.032***	0.006	0.011***	-0.552	-0.872**
	(-2.42)	(-6.07)	(-2.61)	(-4.79)	(1.51)	(2.77)	(-1.45)	(-2.42)
WORDS	0.475	0.541***	0.000***	0.000***	-0.000	-0.000***	0.006	0.023***
	(1.59)	(4.14)	(3.08)	(3.86)	(-1.28)	(-4.82)	(0.98)	(4.42)
FOG	-6.541	-2.407	-0.003	-0.002	0.002	0.004*	-0.298*	-0.456**
	(-1.59)	(-0.99)	(-1.03)	(-0.66)	(1.13)	(1.80)	(-1.71)	(-2.34)
KINCAID	10.779**	5.424**	0.003	0.004	-0.004	-0.004*	0.399*	0.481**
	(2.43)	(2.10)	(1.16)	(1.43)	(-1.57)	(-1.82)	(1.92)	(2.27)
Loan Characteristics								
LOAN_AMT	103.075***	108.508***	-0.022***	-0.036***	0.001	0.003**	-0.048	-0.137
	(53.96)	(53.73)	(-23.01)	(-21.57)	(1.61)	(2.49)	(-0.84)	(-1.55)
60_MONTH	33.298	-57.196**	-0.045**	-0.075***	0.005	0.024*	-0.722	-3.179***
	(0.90)	(-2.35)	(-2.05)	(-3.23)	(0.38)	(1.93)	(-0.64)	(-2.88)
INT_RATE	24.473***	2.809	0.017***	0.009***	0.011***	0.009***	0.187	0.246
	(5.51)	(1.08)	(6.36)	(3.42)	(5.48)	(3.68)	(0.99)	(1.18)
Borrower Characteristics								
EMP_LENGTH	9.433***	8.514***	0.004***	0.009***	0.001	0.001	-0.065	0.055
	(3.80)	(6.04)	(2.79)	(5.90)	(0.73)	(0.70)	(-0.75)	(0.53)
INCOME	-1.764	60.825***	-0.001	0.083***	-0.001**	-0.031***	0.091**	2.664***
	(-1.29)	(9.74)	(-1.26)	(9.99)	(-2.10)	(-6.17)	(2.09)	(6.06)
VERIFIED	-248.508***	-193.024***	-0.281***	-0.308***	0.010	0.012	-0.795	-1.492**
	(-10.68)	(-15.35)	(-17.28)	(-21.96)	(1.19)	(1.25)	(-1.14)	(-1.96)
DTI	0.492	0.367	0.002*	0.004***	0.000	-0.001	-0.038	0.044
	(0.32)	(0.50)	(1.77)	(4.83)	(0.27)	(-0.97)	(-0.61)	(0.85)
Borrower Credit History								
DELINQUENT	-150.740	105.801	-0.220**	0.106				
	(-0.90)	(0.57)	(-2.00)	(0.50)				
DELINQ_AMT	6.818	-22.546	0.009	0.035				
	(0.60)	(-0.30)	(1.13)	(0.48)				
DELINQ_2YRS	-49.909***	-32.612***	-0.021**	-0.051***	0.000	0.015	0.176	-1.599*
	(-3.44)	(-3.10)	(-2.24)	(-3.62)	(0.01)	(1.50)	(0.22)	(-1.70)
TIME_SINCE_DELINQ	2.358	4.374***	0.003**	0.004***	0.002**	0.001	-0.193**	-0.130
	(1.18)	(3.44)	(1.99)	(2.75)	(2.32)	(1.41)	(-2.32)	(-1.42)
CREDIT_HIST	-0.571***	-0.253***	-0.000***	-0.000***	-0.000	0.000	0.003	-0.004
	(-4.80)	(-3.70)	(-5.12)	(-4.41)	(-0.29)	(0.74)	(0.62)	(-0.94)
FICO	0.371	-1.438***	0.000	-0.001**	0.000	-0.000*	-0.024	0.016
	(0.76)	(-5.24)	(1.04)	(-2.55)	(0.60)	(-1.74)	(-1.21)	(0.74)
INQUIRIES	-48.142***	-2.459	-0.021***	0.003	0.016***	0.019***	-1.390***	-1.507***
	(-5.39)	(-0.45)	(-3.82)	(0.50)	(4.13)	(4.85)	(-4.26)	(-4.34)
PUB_RECORD	-105.033	-56.125*	-0.092*	-0.046	0.055	-0.000	-9.819**	-3.178
	(-1.63)	(-1.76)	(-1.87)	(-1.18)	(1.09)	(-0.01)	(-2.16)	(-0.89)
TIME_SINCE_RECORD	-5.115	7.438	-0.020*	-0.002	0.003	-0.005	-1.647	-0.406
	(-0.29)	(0.89)	(-1.69)	(-0.16)	(0.22)	(-0.49)	(-1.60)	(-0.54)
REV_BAL	-1.635***	-4.271***	-0.001***	-0.003***	-0.000	0.000	0.006	-0.080
	(-5.92)	(-4.87)	(-3.18)	(-3.69)	(-0.56)	(0.58)	(0.35)	(-1.28)
REV_UTIL	202.242***	15.468	0.092***	-0.028	0.035*	0.059***	-2.932*	-4.992***
	(3.91)	(0.56)	(2.98)	(-0.92)	(1.82)	(2.94)	(-1.69)	(-2.71)
OPEN_ACC	7.549***	4.202**	0.004**	-0.002	0.002	0.001	-0.218**	-0.042
	(2.80)	(2.36)	(2.21)	(-0.91)	(1.55)	(0.90)	(-2.03)	(-0.36)
TOTAL_ACC	0.423	0.564	0.001	0.002***	-0.001**	0.001	0.079**	-0.079
	(0.41)	(0.72)	(1.43)	(2.82)	(-2.02)	(1.38)	(2.34)	(-1.56)
Purpose Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	No	No	Yes	Yes	No	No	No	No
State-Year-Month Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	34,137	36,298	34,137	36,298	7,709	9,971	7,709	9,971
(Pseudo) R-squared	0.256	0.276	0.008	0.008	0.048	0.048	0.018	0.026

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Variable definitions can be found in Appendix A

Table 10

Effects of Misspelling - Partitioned by Interest Rate

Dependent Variable:	*FUND_PER_DAY*		*FUND_DAYS*		*DEFAULT*		*RETURNS*	
Partitioning Variable:	*INT_RATE*		*INT_RATE*		*INT_RATE*		*INT_RATE*	
VARIABLES	High	Low	High	Low	High	Low	High	Low
Loan Description Characteristics								
MSPL	-24.840**	-23.689***	-0.019***	-0.025***	0.002	0.010***	-0.160	-0.773***
	(-2.27)	(-3.37)	(-2.91)	(-3.77)	(0.30)	(2.90)	(-0.35)	(-2.76)
WORDS	0.833**	0.373***	0.001***	0.000***	-0.000	-0.000***	0.009	0.013***
	(2.36)	(2.66)	(5.49)	(3.65)	(-1.03)	(-4.21)	(0.99)	(4.19)
FOG	-1.790	-4.536*	-0.001	0.000	0.008**	0.001	-0.773**	-0.084
	(-0.48)	(-1.74)	(-0.28)	(0.01)	(2.34)	(0.85)	(-2.38)	(-0.74)
KINCAID	7.836*	7.172**	0.002	-0.001	-0.009**	-0.002	0.947**	0.121
	(1.83)	(2.35)	(0.78)	(-0.18)	(-2.45)	(-1.06)	(2.54)	(0.95)
Loan Characteristics								
LOAN_AMT	119.695***	101.426***	-0.016***	-0.028***	0.001	0.001	-0.045	-0.013
	(54.48)	(53.85)	(-16.15)	(-21.27)	(0.75)	(1.33)	(-0.45)	(-0.30)
60_MONTH	188.997***	-166.070***	0.074***	-0.229***	0.021	0.034***	-4.145**	-2.527***
	(6.42)	(-4.94)	(4.01)	(-8.41)	(1.16)	(3.24)	(-2.39)	(-3.00)
INT_RATE	-27.802***	39.469***	-0.004	0.030***	0.014***	0.004**	-0.095	0.749***
	(-5.02)	(6.90)	(-1.50)	(5.48)	(2.86)	(1.98)	(-0.20)	(4.17)
Borrower Characteristics								
EMP_LENGTH	11.671***	9.531***	0.011***	0.007***	0.001	0.000	-0.121	0.006
	(4.64)	(5.41)	(7.04)	(4.88)	(0.47)	(0.60)	(-0.71)	(0.11)
INCOME	4.692	2.002	0.005**	0.004***	-0.006***	-0.002***	0.615***	0.113***
	(1.24)	(1.09)	(2.55)	(2.63)	(-3.28)	(-3.94)	(3.32)	(3.14)
VERIFIED	-294.562***	-199.824***	-0.377***	-0.269***	0.010	0.001	-1.004	-0.096
	(-13.35)	(-12.76)	(-20.06)	(-18.67)	(0.56)	(0.19)	(-0.65)	(-0.21)
DTI	-1.554	-2.441***	0.001	-0.001	-0.002*	0.001**	0.172	-0.071**
	(-1.15)	(-2.65)	(0.61)	(-0.92)	(-1.78)	(2.07)	(1.41)	(-2.08)
Borrower Credit History								
DELINQUENT	-92.033	465.584	0.025	0.989***				
	(-0.75)	(1.61)	(0.31)	(3.11)				
DELINQ_AMT	3.252	-16.626	-0.006	-0.012				
	(0.34)	(-1.63)	(-0.82)	(-1.42)				
DELINQ_2YRS	-33.851**	-21.136	-0.022***	-0.038**	-0.006	0.004	0.532	-0.879
	(-2.20)	(-1.36)	(-2.59)	(-2.49)	(-0.60)	(0.52)	(0.49)	(-1.28)
TIME_SINCE_DELINQ	2.525	3.740**	0.002	0.003**	0.002	-0.000	-0.171	-0.009
	(1.22)	(2.45)	(1.34)	(2.12)	(1.14)	(-0.53)	(-1.27)	(-0.15)
CREDIT_HIST	-0.590***	-0.383***	-0.000***	-0.000***	0.000	0.000	-0.008	-0.002
	(-4.56)	(-4.76)	(-4.72)	(-6.48)	(0.36)	(1.11)	(-0.89)	(-0.65)
FICO	-2.065***	0.521	-0.000	0.000	-0.000	-0.000	0.025	0.006
	(-3.02)	(1.42)	(-0.61)	(1.24)	(-0.31)	(-1.28)	(0.63)	(0.41)
INQUIRIES	-27.340***	-18.719***	-0.013**	-0.012*	0.019***	0.019***	-1.545***	-1.394***
	(-3.07)	(-2.97)	(-2.56)	(-1.95)	(3.13)	(5.93)	(-2.72)	(-5.69)
PUB_RECORD	-81.367	-193.543***	-0.104***	-0.100*	0.098*	0.019	-14.071**	-0.162
	(-1.54)	(-3.11)	(-3.61)	(-1.78)	(1.73)	(0.46)	(-2.48)	(-0.06)
TIME_SINCE_RECORD	-4.669	-30.662*	-0.029***	-0.018	0.014	0.001	-2.622**	0.075
	(-0.35)	(-1.69)	(-4.06)	(-1.19)	(1.10)	(0.08)	(-2.22)	(0.09)
REV_BAL	-1.483**	-0.846**	-0.000	-0.000	-0.000	-0.000	0.022	0.004
	(-2.21)	(-2.42)	(-0.80)	(-1.44)	(-0.74)	(-0.82)	(0.53)	(0.25)
REV_UTIL	283.246***	43.281	0.167***	-0.020	0.009	0.058***	-1.606	-4.615***
	(6.30)	(1.17)	(5.84)	(-0.58)	(0.28)	(3.85)	(-0.51)	(-3.71)
OPEN_ACC	13.124***	4.703**	0.004**	0.001	0.002	-0.001	-0.186	0.050
	(5.06)	(2.17)	(2.54)	(0.72)	(1.10)	(-0.82)	(-1.01)	(0.68)
TOTAL_ACC	1.555	0.522	0.003***	0.001*	-0.001*	0.000	0.108	-0.010
	(1.34)	(0.65)	(3.72)	(1.65)	(-1.65)	(0.36)	(1.45)	(-0.39)
Purpose Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homeownership Indicators	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	No	No	Yes	Yes	No	No	No	No
State-Year-Month Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	33,972	35,393	33,972	35,393	5,013	11,089	5,013	11,089
(Pseudo) R-squared	0.321	0.275	0.007	0.011	0.028	0.028	0.027	0.020

State-Year-Month clustered t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Variable definitions can be found in Appendix A

Appendix

Appendix: Variable Definitions

Dependent Variables	
FUND_PER_DAY	The amount, in dollars per day, invested in the loan by investors through LendingClub's website. Calculated as the total number of dollars invested divided by the number of days between the date the loan was listed and the date the loan was funded.
FUND_DAYS	The number of days that it takes the loan to be fully funded.
DEFAULT	Indicator variable for whether the loan ended in default.
RETURNS	Net yield to the investors (before fees) in percentage terms. Calculated as the internal rate of return on the loan assuming uniform payment over length of loan life of all payments received.
Loan Characteristics	
LOAN_AMT	Size of the loan in thousands of dollars.
60_MONTH	Indicator variable for whether the loan is for 60 months. 0 indicates a 36 month loan.
INT_RATE	Stated interest rate of the loan, in percentage terms.
LIST_MONTH	Months since beginning of data set at time of application.
PURPOSE	Unverified purpose of the loan submitted by the lender at time of application (e.g., Credit Card Refinancing, Debt Consolidation, Home Improvement, Business)
Borrower Characteristics	
EMP_LENGTH	Ordered categorical variable for length of employment at current employer at time of application.
INCOME	Borrower submitted monthly income in thousands of dollars per month.
VERIFIED	Indicator variable for whether borrower employment information is confirmed by LendingClub.
DTI	Debt-to-income ratio at time of application.
HOME_OWNERSHIP	Home ownership status of borrower at time of application (e.g., Rent, Own, Mortgage, None)
Borrower Credit History	
DELINQUENT	Indicator variable for whether any of borrowers credit accounts are delinquent at time of application.
DELINQ_AMT	Amount of debt on borrower's credit report that is delinquent at time of application in thousands of dollars.
DELINQ_2YRS	Number of delinquencies reported on borrower's credit report within the last two years at the time of application.
TIME_SINCE_DELINQ	Ordered categorical variable for time since last delinquency on borrower's credit report at time of application, where lower numbers indicate more recent delinquency.
CREDIT_HIST	Length of borrower credit history, in months at time of application.
FICO	Borrower's FICO score at time of application.
INQUIRIES	Number of hard credit inquiries on borrower's credit report within the last six months at time of application.
PUB_RECORD	Number of public records on borrower's credit report at time of application.
TIME_SINCE_RECORD	Ordered categorical variable for time since last public record on borrower's credit report at time of application, where lower numbers indicate more recent record.
REV_BAL	Amount of revolving debt in thousands of dollars on borrower's credit report at time of application.
REV_UTIL	Percentage of credit limit on revolving accounts being used by borrower at time of application.
OPEN_ACC	Number of open credit accounts held by the borrower at time of application.
TOTAL_ACC	Number of credit accounts ever held by the borrower at time of application.
Loan Description Characteristics	
MSPL	Number of misspellings within the text of the borrower submitted loan description.
SQRT_MSPL	Square root of the number of misspellings within the text of the borrower submitted loan description.
PCT_MSPL	Number of misspellings within the text of the borrower submitted loan description divided by the total number of words in the description.
WORDS	Total number of words in the borrower submitted loan description.
FOG	FOG Index of the borrower submitted loan description. FOG is a measure of lexical complexity which is increasing in sentence length and use of multisyllable words.
KINCAID	Flesch-Kincaid ease of reading index, which is increasing in sentence length and syllables per word.

References

Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., Inglis, A., & Lancee, W. (1996). Long-Term Consistency in Speech/Language Profiles: II. Behavioral, Emotional, and Social Outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(6), 815-825.

Berger, A. N., & Udell, G. F. (1995). Relationship Lending and Lines of Credit in Small Firm Finance. *Journal of Business*, 68(3), 351-381.

Bloomfield, R. (2008). Discussion of "Annual Report Readability, Current Earnings, and Earnings Persistence." *Journal of Accounting and Economics*, 45, 248-252.

Boland, J. E., & Queen, R. (2015). If You're House Is Still Available, Send Me an Email: Personality and Assessing Mistakes in Email Messages. *Working Paper*.

Boyes, W. J., Hoffman, D. L., & Low, S. A. (1989). An Econometric Analysis of the Bank Credit Scoring Problem. *Journal of Econometrics*, 40, 3-14.

Brochet, F., Naranjo, P., & Yu, G. (2012). Causes and Consequences of Linguistic Complexity in non-U.S. Firm Conference Calls. *Harvard Business School Working Paper*, No. 13-033.

Campbell, J. Y. (2006). Household Finance. *Journal of Finance*, 61(4), 1553-1604.

Cohen, N. J., Davine, M., Horodezky, N., Lipsett, L., & Isaacson, L. (1993). Unsuspected Language Impairment in Psychiatrically Disturbed Children: Prevalence and Language and Behavioral Characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32(3), 595-603.

Cortese, A. (2014, May 4). Loans That Avoid Banks? Maybe Not. *New York Times*, pp. BU.1.

Devaney, S. A., & Lytton, R. H. (1995). Household Insolvency: A Review of Household Debt Repayment, Delinquency, and Bankruptcy. *Financial Services Review*, 4(2), 137-156.

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.

Dunn, L. F., & Kim, T. (1999). An Empirical Investigation of Credit Card Default. *Working Paper*.

Everard, A., & Galletta, D. F. (2006). How Presentation Flaws Affect Perceived Site Quality, Trust, and Intention to Purchase from an Online Store. *Journal of Management Information Systems*, 22(3), 56-95.

Figueredo, L., & Varnhagen, C. K. (2005). Didn't You Run the Spell Checker? Effects of Type of Spelling Error and Use of a Spell Checker on Perceptions of the Author. *Reading Psychology*, 26(4-5), 441-458.

Gao, Q., & Lin, M. (2013). Linguistic Features and Peer-to-Peer Loan Quality: A Machine Learning Approach. *Working Paper.*

Hargittai, E. (2006). Hurdles to Information Seeking: Spelling and Typographical Mistakes During Users' Online Behavior. *Journal of the Association for Information Systems*, 7(1), 52-67.

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.

Hildebrand, T., Puri, M., & Rocholl, J. (2013). Adverse Incentives in Crowdfunding. *Working Paper.*

Hinshaw, S. P. (1992). Externalizing Behavior Problems and Academic Underachievement in Childhood and Adolescence: Causal Relationships and Underlying Mechanisms. *Psychological Bulletin*, 111(1), 127-155.

Kreiner, D. S., Schnakenberg, S. D., Green, A. G., Costello, M. J., & McClin, A. F. (2002). Effects of Spelling Errors on the Perception of Writers. *The Journal of General Psychology*, 129(1), 5-17.

Lee, Y. (2012). The Effect of Quarterly Report Readability on Information Efficiency of Stock Prices. *Contemporary Accounting Research*, 29(4), 1137-1170.

Lehavy, R., Li, F., & Merkley, K. (2011). The Effect of Annual Report Readability on Analyst Following and the Properties of Their Earnings Forecasts. *The Accounting Review*, 86(3), 1087-1115.

Li, F. (2008). Annual Report Readability, Current Earnings, and Earnings Persistence. *Journal of Accounting and Economics*, 45, 221-247.

Li, F. (2010). Textual Analysis of Corporate Disclosures: A Survey of the Literature. *Journal of Accounting Literature*, 29, 143-165.

Lin, M., & Viswanathan, S. (2014). Home Bias in Online Investments: An Empirical Study of an Online Crowd Funding Market. *Working Paper.*

Mach, T. L., Carter, C. M., & Slattery, C. R. (2014). Peer-to-Peer Lending to Small Businesses. *Working Paper.*

Maier, M. S. (2014). Lending to Strangers: Does Verification Matter? *Working Paper.*

Michels, J. (2012). Do Unverifiable Disclosures Matter? Evidence from Peer-to-Peer Lending. *The Accounting Review*, 87(4), 1385-1413.

Morse, A. (2015). Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending. *Working Paper*.

Nisbett, R. E., & Wilson, T. D. (1977). The Halo Effect: Evidence for Unconscious Alteration of Judgments. *Journal of Personality and Social Psychology*, 35(4), 250-256.

Pape, K., Bjørngaard, J. H., Westin, S., Holmen, T. L., & Krokstad, S. (2011). Reading and Writing Difficulties in Adolescence and Later Risk of Welfare Dependence. A Ten Year Follow-Up, the HUNT Study, Norway. *BMC Public Health*, 11(718), 1-8.

Petersen, M. A. (2004). Information: Hard and Soft. *Working Paper.*

Pope, D. G., & Sydnor, J. R. (2009). What's in a Picture? Evidence of Discrimination from Prosper.com. *Journal of Human Resources*, 46(1), 53-92.

Queen, R. & Boland, J. E. (2015). I Think Your Going To Like Me: Exploring the Role of Mistakes in Email Messages on Assessments of Potential Housemates. *Working Paper*.

Ravina, E. (2012). Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets. *Working Paper.*

Rennekamp, K. (2012). Processing Fluency and Investors' Reactions to Disclosure Readability. *Journal of Accounting Research*, 50(5), 1319-1354.

Russel, D. H. (1955). A Second Study of Characteristics of Good and Poor Spellers. *Journal of Educational Psychology*, 46(3), 129-141.

Rutter, M. & Yule, W. (1975). The Concept of Specific Reading Retardation. *Journal of Child Pyschology and Psychiatry*, 16. 181-197.

Schreiner, M. & Nagarajan, G. (1997). Predicting Creditworthiness with Publically Observable Characteristics: Evidence from ASCRAs and RoSCAs in the Gambia. *Savings and Development*, 21(4), 399-414.

Scott, G. G., Sinclair, J., Short, E., & Bruce, G. (2014). It's Not What You Say, It's How You Say It: Language Use on Facebook Impacts Employability But Not Attractiveness. *Cyberpsychology, Behavior, and Social Networking*, 17(8), 562-566.

Spreen, O. (1988) Prognosis of Learning Disability. *Journal of Child Psychology and Psychiatry*, 56(6), 836-842.

Stiff, C. (2012). Watch What You Write: How Errors in Feedback Influence Consumer Attitudes and Behavior. *Journal of Internet Commerce*, 11(1), 41-67.

Verstein, A. (2011). The Misregulation of Person-to-Person Lending. *University of California, Davis Law Review*, 45, 445-530.

Vignovic, J. A., & Thompson, L. F. (2010). Computer-Mediated Cross-Cultural Collaboration: Attributing Communication Errors to the Person Versus the Situation. *Journal of Applied Psychology*, 95(2), 265-276.