Non-Pecuniary Preferences, Adverse Selection, and Moral Hazard in P2P Lending: Evidence from Lending Club

Dong-Jie Fang[1], Zong-Wei Yeh[2], Chien-Hsiu Lin[3*], and Shih-Kuei Lin[4]

First version: October 24, 2023 This version: April 30, 2024

Abstract

This paper investigates the non-pecuniary preferences of investors on P2P lending platforms, including pro-environmental and prosocial preferences. We capture these characteristics and writing styles from loan descriptions by text mining approaches and analyze their impacts on funding success and default. Our results show that the lenders on Lending Club do not have non-pecuniary preferences or even avoid investing in pro-environmental and prosocial loans. However, the absence of such preferences leads them into the trap of adverse selection since these loans have lower default probabilities. Furthermore, the negative descriptions of loans can also lead to adverse selection among investors, while borrowers' fraudulent content in descriptions will expose investors to moral hazard.

Keywords: P2P lending; Non-pecuniary preferences; Adverse selection; Moral hazard; Information asymmetry

JEL Codes: D82; G21; Q56.

---

[1] Department of Money and Banking, National Chengchi University; Postal Address: No.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan; E-mail: 108352506@nccu.edu.tw.

[2] Department of Money and Banking, National Chengchi University; Postal Address: No.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan; E-mail: 111352502@nccu.edu.tw.

[3*] Corresponding author. Department of Money and Banking, National Chengchi University; Postal Address: No.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan; E-mail: clin@nccu.edu.tw.

[4] Department of Money and Banking, National Chengchi University; Postal Address: No.64, Sec.2, ZhiNan Rd., Wenshan District, Taipei City 11605, Taiwan; E-mail: square@nccu.edu.tw.

1. Introduction

As the promoter of inclusive finance, P2P lending not only helps solve the economic frictions that financial intermediation faces by reducing transaction costs (Aaron et al., 2017), but also, widens access to financial services for small and medium-sized enterprises (SMEs) by creating new avenues for funding (Lagarde, 2018). Recently, the discussion about the social impact of P2P lending (e.g., positive externalities and benefits to social welfare) has drawn the attention of researchers. Most of the related literature focuses on crowdfunding platforms like Kiva.org, because of their prosocial nature, rather than general microfinance institutions (MFIs), like Lending Club and Prosper (Allison et al., 2015; Dorfleitner et al., 2021; Ravishankar, 2021).

Our research starts from a different but interesting perspective. The P2P loans on prosocial crowdfunding platforms are usually interest-free. However, sacrificing investment returns for charity should not be purely an "all-in" decision. We are particularly interested in how lenders balance their pursuit of financial returns with non-pecuniary preferences on a more generalized online lending platform. To this end, we investigate whether the lenders (investors) on Lending Club, which once was the world's largest P2P lending platform, have non-pecuniary preferences.

The discussion about the non-pecuniary preference in financial markets is becoming increasingly salient in academic research (Zerbib, 2019; Pástor et al., 2021; Baker et al., 2022; Jo et al., 2022). In this paper, we focus on two types of non-pecuniary preferences: pro-environmental and prosocial preferences. The pro-environmental preference is defined as a conscious choice to minimize the negative impact on the environment (Kollmuss and Agyeman, 2002; Steg and Vlek, 2009; Mesmer-Magnus et al., 2012), while the prosocial one refers to a person's intention to benefit others in the society as a whole (Batson and Powell, 2003). In the P2P lending world, lenders with a pro-environmental preference prefer to fund loans related to the environment or having a positive externality on it, while those with a prosocial preference prefer to fund loans for disadvantaged groups.

Previous literature has mainly concentrated on prosocial behavior in P2P lending or even conflated these two features, whereas we conduct a more detailed categorization and exploration of these non-pecuniary preferences by text mining approaches. We extend the Environment, Social, and Governance (ESG) and business ethics dictionaries by Baier et al. (2020) and Loughran et al. (2023) into a new version to capture the non-pecuniary features in loan description; a narrative expression about the loan purpose written by the borrower.

By using funding success as an indicator reflecting the lenders' decisions, we investigate whether lenders exhibit non-pecuniary preferences when investing in loans. Our empirical results show that lenders on Lending Club do not exhibit such preferences. Surprisingly, the investors actually avoid investing in prosocial and pro-environmental loans.

This difference between Lending Club and those prosocial platforms piques our curiosity as to what impact this aversion psychology has on the investment performance of lenders. According to prior literature, information asymmetry increases the credit risk in P2P lending (Lin et al., 2013; Emekter et al., 2015; Hu et al., 2020), even to the extent of failure of many online lending platforms in China (Gao et al., 2021; Shao and Bo, 2022; Wang and Li, 2023). Therefore, beginning by describing the working mechanism of Lending Club as an extensive-form game, we analyze the potential risks led by information asymmetry. By simultaneously examining the impact of non-pecuniary preferences on funding success and default, we find that the absence of such preferences can lead investors to adverse selection. When lenders prefer to choose a loan without prosocial or pro-environmental features, they choose one with a higher default risk.

Besides, when filtering control variables that can greatly predict funding success and default with backward selection, we notice a particular phenomenon that is different from research with data from other platforms. In our case, we find that the lenders on Lending Club prefer to invest in borrowers with lower credit scores, which is opposite to the findings by other researchers, such as Gao et al. (2023), with data from Prosper. This implies that the investors on Lending Club seek higher returns, while also having a higher risk tolerance.

This result has inspired us to explore soft information ascertain whether or not the writing style of the loan description yields different results from those of Gao et al. (2023). We construct the linguistic metrics, including readability, tone, and deception cues, and investigate their impacts on funding success and loan default. The result for default is consistent with the intuition and findings of Gao et al. (2023). That is, a loan with a more readable or more positive description has a lower default probability than others, while one with more deception cues has the opposite effect. However, the story goes differently for funding success. The lenders prefer a more readable description, which aligns with intuitive expectations, but would also fund a loan with a more negative tone or more deception cues, which can increase the probability of facing default.

For these empirical results, we analyze them from two different perspectives. Firstly, from that of lenders, it emerges that emotional contagion has an impact on the investors' funding decisions. Investors are susceptible to negative emotions, leading to them underestimating the credit risk of loans. Similar to the lack of non-pecuniary preferences, this misunderstanding about the product quality (the default probability) makes them fall into the trap of adverse selection. Secondly, from the perspective of borrowers, they can benefit from information asymmetry. As the party with relatively more information, borrowers can use fabrication and deception in loan descriptions to fool investors and increase the chances of funding success. However, borrowers with more deception cues are more likely to default, which is precisely what imposes moral hazard on investors.

This paper contributes to related literature in several ways. First of all, to the best of our knowledge, this is the first work that examines the existence of pro-environmental and prosocial preferences on a general P2P lending platform. Our results indicate that different types of online lending platforms cater to distinct target customers. Secondly, this paper sheds more light on the information asymmetry issues in P2P lending than prior research, including adverse selection and moral hazard. The empirical results show that the absence of non-pecuniary preferences, erroneous estimations influenced by negative emotional contagion, and misjudgments affected by fraudulent content in loan descriptions will lead to investors facing higher credit risk. Last but not least, building upon previous research emphasizing the importance of loan descriptions (Herzenstein et al., 2011; Michels, 2012; Dorfleitner et al., 2016; Han et al., 2018; Gao et al., 2023), we further verify that as uncertified soft information, the writing style of these narratives harbor default risk that leads investors astray.

The rest of the paper is organized as follows. Section 2 reviews related literature, while Section 3 introduces the theoretical foundations and develops the hypotheses.

Section 4 introduces the methodologies, including identifying non-pecuniary preferences and evaluating linguistic metrics. Section 5 presents data description, data preprocessing, and the selection of control variables. Section 6 presents the empirical results and finally, Section 7 concludes the paper.

2. Literature review

This paper relates to three aspects of literature: (1) research exploring the determinants of loan performance in P2P lending, including funding success, default probability, and interest rate of the loan; (2) research focusing on non-pecuniary preferences in financial markets, including pro-environmental and prosocial characteristics; and (3) research on the information asymmetry between borrowers and lenders in P2P lending, especially the adverse selection and moral hazard issues.

2.1. The determinants of loan performance in P2P lending

Prior literature has examined different types of determinants on funding success, default probability, and interest rate of loans in P2P lending. According to Basha et al. (2021), these determinants can be generally summarized into four categories, including (1) loan amount, interest rate (Cai et al., 2016; Emekter et al., 2015; Herzenstein et al., 2011), and credit grade (Greiner and Wang, 2009 and Tao et al., 2017) as financial determinants; (2) gender, race, and age (Pope and Sydnor, 2011; Chen et al., 2017; Chen et al., 2014; Chen et al., 2020; Barasinska and Schäfer, 2014; Gonzalez and Loureiro, 2014) as demographic determinants; (3) social trust (Lin et al., 2013; Yum et al., 2012; Chen et al., 2014), social groups (Freedman and Jin, 2017), social capital/collateral (Liu et al., 2020), loan purpose (Yao et al., 2019), and borrower's post (Chen et al., 2018; Han et al., 2018; Larrimore et al., 2011) as social determinants; and (4) entrepreneurial experience and financial innovation (Atz and Bholat, 2016), complementary or substitutionary effect (Jagtiani and Lemieux, 2018), financial intermediation/disintermediation (Havrylchyk and Verdier, 2018), and macroeconomic factors (Atz and Bholat, 2016; Yoon et al., 2019), as macroeconomic determinants.

These rich research findings provide valuable insights for us when selecting the control variables needed for empirical analysis. Our discussion about pro-environmental and prosocial features of P2P lending loans also provides additional evidence for research in related fields.

2.2. Non-pecuniary preferences: Pro-environmental and prosocial characteristics

The rise of ESG and corporate social responsibility (CSR) in the business world has led to academics focusing attention on investors' non-pecuniary preferences. Research has involved investigating whether investors not only chase wealth, but also, positive externalities. The discussion about the non-pecuniary preferences spans across multiple financial markets, but can be generally divided into two types: pro-environmental and prosocial preferences.

Starting from the definition, pro-environmental behavior refers to that which minimizes negatively impacting the environment (Kollmuss and Agyeman, 2002), benefits the environment (Steg and Vlek, 2009), or improves environmental sustainability (Mesmer-Magnus et al., 2012). In the fields of finance and economics, with the tide of climate change and green finance, much research has been focused on the pro-environmental preference of investors in financial markets. Baker et al. (2022) and Zerbib (2019) found a negative "greenium" in green bond markets, thus indicating the existence of investors' environmental concerns. Pástor et al. (2021) proposed an asset pricing equilibrium model to describe market participants' non-pecuniary preferences for green assets. The empirical studies by Pástor et al. (2022) verified the theoretically motivated green factor driving stock returns. With data of Chinese environmental mutual funds, Jo et al. (2022) investigated the impact of air pollution on investors' behavior. Their findings indicated that investors with environmental awareness earn benefits from non-pecuniary considerations.

Moreover, research about prosocial behavior in financial markets is growing. Prosocial behavior, a term proposed by sociologists as "an antonym for antisocial", is defined as an action intended to benefit people other than oneself (Batson and Powell, 2003). With data of dual-objective Venture Capital (VC) funds, Barber et al. (2021) found that investors would rather sacrifice returns in exchange for some non-pecuniary benefits (social impacts) by investing in dual-objective Venture Capital (VC) funds. With an experiment involving a two-stock laboratory asset market, Draganac and Lu (2022) investigated the impacts from prosocial preferences and image considerations on stock markets. With incentivized experiments, Bonnefon et al. (2022) characterized investors' moral preferences, including value alignment and impact-seeking preference. Their results showed that investors attempt to align their investments with their social values, while there is no evidence supporting the investment decisions driven by chasing social impact. Different from prior research, Dangl et al. (2023) constructed a theoretical framework to investigate the impacts on investors and corporates' investment decisions according to three different types of social preferences: deontological, non-consequentialist, and consequentialist.

When turning attention to research related to P2P lending, most literature has paid attention to the prosocial feature on online crowdfunding platforms (e.g., Galak et al., 2011; Burtch et al., 2014; Allison et al., 2015; Dorfleitner et al., 2021; Ravishankar, 2021). Through data from Kiva.org, an international crowdfunding platform facilitating prosocial, P2P lending, some researchers have achieved interesting results from different perspectives. Galak et al. (2011) found that lenders favor individual borrowers over groups, especially those borrowers who are proximate to themselves in three dimensions of social distance, including gender, occupation, and first name initial. Burtch et al. (2014) provided evidence that, when choosing transaction partners, lenders prefer borrowers with cultural similarity and geographical proximation. Focusing on a set of entrepreneurs seeking financing on Kiva.org, Allison et al. (2015) elicited that lenders appreciate firms more that highlight the prosocial opportunities in narratives than those that emphasize the business opportunities. Dorfleitner et al. (2021) found two main factors in predicting successful funding, including the social underwriting by a third-party trustee and soft information in the description that fosters the lenders' trust.

Different from prior research, this paper has the aim of investigating the non-pecuniary preferences of lenders on a more general P2P lending platform, namely Lending Club, rather than prosocial crowdfunding platforms, which can provide further interesting insights. We not only pay attention to prosocial behaviors, but also, pro-environmental features, which have not been widely discussed in P2P lending as yet.

2.3. Information asymmetry in P2P lending: Adverse selection and moral hazard

In financial markets, information asymmetry between counterparties can lead to adverse selection (Akerlof, 1970; Spence, 2002) and moral hazard (Stiglitz and Weiss, 1981) issues. Due to P2P lending's heavier reliance on unverified hard and soft information, usually lacking platform verification or third-party endorsement, these platforms exhibit more pronounced information asymmetry issues compared to traditional financial institutions (Michels, 2012). The transparency, completeness, and accuracy of information disclosed by borrowers often affect the rational decision-making of investors. Both the feasibility of information acquisition and the investor sophistication influence the investment performance in P2P lending (Vallee and Zeng, 2019).

Previous research has indicated that the long-standing adverse selection and moral hazard issues on P2P lending platforms increase the risk for investors of achieving rational profits on the platform, while also raising the difficulty for borrowers to successfully obtain loans (Yum et al., 2012; Lin et al., 2013; Emekter et al., 2015; Hu et al., 2020). Adverse selection and moral hazard are even considered as contributing factors to the failure of P2P online lending platforms, which is particularly evident in the numerous closures of small and medium-sized online lending platforms in the Chinese market (Gao et al., 2021; Shao and Bo, 2022; Wang and Li, 2023).

In light of this, some research has been focused on exploring methods to mitigate adverse selection and moral hazard on P2P lending platforms, including the roles of social relationships (Lin et al., 2013; Galema, 2020), the loan pricing mechanisms (Vallee and Zeng, 2019), the third-party credit certifications (Hu et al., 2020), and soft information (Liu et al., 2020).

To extend and to complement the topics discussed in the aforementioned literature, this study delves into the information asymmetry associated with investors' non-pecuniary preferences, particularly pro-environmental behaviors, which have not been thoroughly examined. We also incorporate soft information revealed through the writing style of loan descriptions to investigate adverse selection and moral hazard issues.

3. Theoretical foundations and hypotheses development

3.1. The mechanism of P2P lending: An extensive-form game

The working mechanism of P2P lending on Lending Club can be described as an extensive-form game. According to Figure 1, the borrower submits all the required (as well as optional) hard and soft information during the loan application stage. Based on the submitted information, the P2P lending platform (Lending Club in our case) decides whether to approve the loan and assesses the corresponding price, which is the borrowing interest rate.

With all the public information about the loan and the priced interest rate, the lender will decide on whether to fund the loan (as well as the funding ratio). For the funding decision, the lender needs to decide on full or partial funding, which means the funding ratio is determined. Conversely, when the loan reaches maturity, the borrower needs to decide whether to repay as agreed or default. To simplify the default situation, we only consider total default, rather than a possible recovery rate, in this case.

[Insert Figure 1 here]

Accordingly, the lender's funding decision provides us an appropriate context to examine the existence of non-pecuniary (pro-environmental and prosocial) preferences in the P2P lending market, which is one of the major issues in this paper. Moreover, there is the possibility of potential information incompleteness during the decision-making procedure. If the borrower possesses more information about the real quality of the loan, the information asymmetry between the lender and the borrower can lead to adverse selection and moral hazard in practice. For further investigation, we next discuss the decision making by lenders and borrowers, which are funding and default, respectively. After that, we develop our hypotheses.

3.2. Full funding, funding ratio, and default

Contrary to traditional financial intermediations, P2P platforms place more emphasis on the direct interaction between the borrower and the lender. According to the extensive-form game structure, at every decision point during the sequencing movement, two players have to make the best choice to maximize their own benefits (or utilities).

Starting from the respective interests of borrowers and lenders, this research mainly focuses on the following two concerns: (1) From the borrowers' perspective, they care more about securing the maximum possible funding. We aim to scrutinize whether borrowers design their loan descriptions to attract investors through pro-environmental (or prosocial) content or their writing styles. (2) Conversely, from the lenders' perspective, their major concerns are the loan default probabilities. We aim to explore whether lenders can reduce default risk with the help of loan descriptions.

To evaluate funding success and measure whether the information provided by the borrower can attract investor's funding, we generate two measuring indicators, *Full*

Funding and *Funding Ratio*, by following Hu and Song (2017), Nowak et al. (2017), and Wang and Tong (2020).

(1) Full Funding

Based on the loan amount applied by the borrower and the final amount invested by the lender, *Full Funding* measures whether the loan amount is fully funded by investors, which means the borrower has obtained the requested amount of money. Under this scope, *Full Funding* is a binary variable that equals 1 when the entire loan amount is met and equals 0 otherwise.

We use logistic regression to analyze the influence of explanatory variables on investors' willingness to invest, including loan characteristics, such as the non-pecuniary preferences and the writing style of the description. This allows us to estimate the probability that a given loan will get fully funded based on the explanatory variables. The regression is set as follows:

$$\mathbb{P}(Full\ Funding_i = 1) = f(\boldsymbol{X}_i'\boldsymbol{\beta} + \boldsymbol{Y_i}'\boldsymbol{\gamma} + \varepsilon_i), \tag{1}$$

where $\mathbb{P}(Full\ Funding_i = 1)$ denotes the probability that the *i*-th borrower's loan is fully funded, $\boldsymbol{X_i}$ is a vector of control variables chosen through the backward selection, with *Full Funding* as the response variable, $\boldsymbol{Y_i}$ is a vector of explanatory variables including non-pecuniary preferences, readability, tone, and deception cues, and the logistic function $f(Z_i) = e^{Z_i}/(1 + e^{Z_i})$.

(2) Funding Ratio

Additionally, to reflects the extent to which the loan amount requested has been successfully filled by investors, we examine the *Funding Ratio*, a variable representing the proportion of the loan amount requested that has been funded by investors. This metric provides insights into the level of investors' participation and the overall interest in the loan listing. A higher *Funding Ratio* indicates a higher level of investor confidence and support for the borrower's loan application. When the *Funding Ratio* is less than 1, it means that the borrower has not received the full loan amount.

The funding ratio is a variable restricted to a range between 0 and 1. Specifically, when an investor's investment desire exceeds the borrower's applied loan amount, the ratio is truncated to 1. Under such a circumstance, it is unable to observe instances where the true investment desire exceeds 1, even if such a situation may exist. In other words, since Lending Club uses a post-prices mechanism, rather than the auction adopted by other online crowdfunding platforms (e.g., Kiva) (Wei and Lin, 2017), we cannot ascertain the actual amount an investor might be willing to invest when it exceeds the borrower's requested amount.

In dealing with such a situation, we employ a Tobit regression model for our analysis. This is a regression model designed to handle censored variables, treating truncated data as underlying continuous variables and viewing the censoring points as particular thresholds. In our research, the Tobit model enables us to estimate investors' latent investment desires, even when their actual investment amounts are truncated. The rate at which the investor desires to invest, f_i^*, is a latent variable as follows:

$$f_i^* = \boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta} + \boldsymbol{Y}_{\boldsymbol{i}}'\boldsymbol{\gamma} + \varepsilon_i, \tag{2}$$

where $\boldsymbol{X}_{\boldsymbol{i}}$ is a vector of control variables chosen by backward selection, with *Funding Ratio* as the response variable, and $\boldsymbol{Y}_{\boldsymbol{i}}$ is a vector of explanatory variables.

(3) Default

As is mentioned above, the default is defined as a binary variable in this paper. It equals to 1 when a loan is not repaid by the borrower at the maturity and 0 otherwise. We implement a logistic regression model to examine the impact of different explanatory variables on a loan's default probability. The regression model can be represented as:

$$\mathbb{P}(Default_i = 1) = f(\boldsymbol{X}_{\boldsymbol{i}}'\boldsymbol{\beta} + \boldsymbol{Y}_{\boldsymbol{i}}'\boldsymbol{\gamma} + \varepsilon_i), \tag{3}$$

where $\mathbb{P}(Default_i = 1)$ denotes the *i*-th borrower's default probability, $\boldsymbol{X}_{\boldsymbol{i}}$ is a vector of control variables selected through backward selection, with *Default* as the response variable, $\boldsymbol{Y}_{\boldsymbol{i}}$ is a vector of explanatory variables, and the logistic function $f(Z_i) = e^{Z_i}/(1 + e^{Z_i})$.

3.3. Hypotheses

As is mentioned before, the majority of the existing literature has examined the role of soft information centers around prosocial crowdfunding platforms, such as Kiva.org. From our perspective, online lending platforms like Lending Club provide a more comprehensive reflection of online lenders' non-pecuniary investment preferences. While investors on Lending Club may be willing to sacrifice some returns in favor of selecting loans with prosocial or pro-environmental attributes, they can still maintain a certain level of investment returns, as opposed to the interest-free scenario on Kiva.

However, due to the distinct customer bases of different types of P2P lending platforms, investors on platforms like Lending Club prioritize financial returns more than those on crowdfunding ones. In addition, considering the results from Wang and Tong (2020) and Gao et al. (2023) and our findings when screening control variables in Section 5, we believe that investors on Lending Club, in contrast to Prosper, tend to pursue higher returns associated with higher risks, despite many similarities being shared by these two platforms. Therefore, we propose our first hypothesis as follows.

Hypothesis 1. Lenders on Lending Club do not exhibit non-pecuniary preferences, no matter whether pro-environmental or prosocial, and are even inclined to avoid investing in loans with such characteristics.

Furthermore, borrowers applying for prosocial or pro-environmental loans are often perceived as vulnerable groups in society. Consequently, despite having credit ratings close to other borrowers, they are more likely to be labeled with a high default risk by investors. This potential aversion among investors can lead to adverse selection resulting from their own biases and exacerbated by information asymmetry.

Hypothesis 2. Lenders can fall into adverse selection due to their lack of non-pecuniary preferences.

In addition, loan descriptions can provide investors with more information, but can also exacerbate information asymmetry between borrowers and lenders. As a form of soft information, borrowers can lie or modify their tone in loan descriptions to increase their chances of funding success. This inaccurate information and deliberately crafted expression can potentially mislead lenders, thus exacerbating information asymmetry and leading to issues such as adverse selection and moral hazard.

Hypothesis 3. Lenders may fall into adverse selection and moral hazard due to the soft information reflected in the writing style of the loan description.

Thus, through empirical analysis, we seek to verify and analyze the existence of adverse selection and moral hazard, with respect to the perspectives of lenders and borrowers, separately.

4. Methodologies

4.1. Identifying non-pecuniary preferences

During the P2P lending process, the borrower can provide a description to disclose in detail the loan purpose and any further information that is not reflected in quantitative questionnaires. The investor can make the lending decision based not only on the structural information, such as credit rating, education level, and annual income, but also, on these written narratives.

On the one hand, the loan description data allows researchers to analyze the impact of unstructured data on loan status. Various studies have suggested that borrowers who offer these descriptions can significantly boost their loan approval rates (see Michels, 2012; Peng et al., 2016), thus augmenting the incentive for borrowers to fill them out. On the other hand, these textual materials also provide us a good field to identify whether a loan matches an investor's non-pecuniary preference.

With the burgeoning development of natural language processing (NLP) techniques, a proliferation of recent research has honed in on analyzing textual characteristics within loan descriptions (see Allison et al., 2013; Bruton et al, 2015; Dorfleitner et al., 2016; Hirshleifer, 2001) or using such narratives to investigate borrower's behavior (see Dorfleitner et al., 2016; Herzenstein et al., 2011; Hirshleifer, 2001; Iyer et al., 2016; Larrimore et al., 2011; Wang et al., 2017). However, to the best of our knowledge, there has been limited research directly focusing on the identification of such characteristics in text mining, not to mention distinguishing between pro-environmental and prosocial behaviors. We deem to it appropriate commence our approach with the NLP methodology of the Environment, Social, and Governance (ESG) dictionary.

Baier et al. (2020) (BBK) were the first to develop an ESG-related dictionary. They devised a catalog of 482 terms by examining the annual reports and proxy statements of 25 companies listed in the S&P 100 Index with the largest market capitalization. They, then, grouped these terms into Environmental, Social, and Governance categories. Acknowledging that language usage evolves over time, Loughran et al. (2023) (LMO) extended BBK's ESG dictionary into a business ethics version, by integrating additional terms and more subcategories[5].

However, these terms primarily apply to professional financial reports and may not seamlessly align with the narratives provided by general borrowers on P2P lending platforms. Therefore, we extend the ESG and business ethics dictionaries by Baier et al. (2020) and Loughran et al. (2023) into a new version to capture the non-pecuniary preferences, which includes two subcategories: pro-environmental and prosocial preferences.

We introduce new words (and bigrams) into the dictionary to make it more suitable for our purpose and online lending circumstances. Within the pro-environmental category, we enrich LMO's Environmental words list by adding four bigrams (*warmer climate, natural gas, fuel efficient,* and *fuel saving*) to capture the households' demand of saving energy. For the prosocial category, we have incorporated six words (*tuition, family, unemployed, parent, resident,* and *military*) into BBK's Social words list to better present individual social relations, rather than corporate social responsibilities.[6] According to the approach mentioned above, we classify a loan as a pro-environmental or prosocial one, if the narrative description mentions corresponding lexicons in our dictionary.

4.2. Evaluating linguistic metrics

Beyond analyzing the impact of non-pecuniary preferences on loan performance and lender's decision making, we are also curious about the borrower's strategy reflected in the writing style. By employing NLP techniques, we can extract writing characteristics from the description crafted by borrowers, thereby capturing their cognition and behavior.

The research closest to this topic is Gao et al. (2022). With data from Prosper, a crowdfunding platform in the United States, they provided evidence that the readability,

[5] Based on Baier et al.'s (2020) ESG dictionary, Loughran et al. (2023) expanded the ESG categories by adding more words and developing their own business ethics dictionary by totally incorporating nine different subcategories. See Loughran et al. (2023) for more details.

[6] Loughran et al. (2023) worked from BBK's Social words list to create their Human Rights word list, which does not exactly match the application environment of our research. Thus, we directly modify BBK's Social words list to better suit the P2P lending circumstance.

tone, and deception cues of loan descriptions are critical elements impacting a loan's funding success and default. Their results show that loans with higher readability, more positive tone, or fewer deception cues can attract more investor attention (receive more bids, higher bid amounts, and lower interest rates) and are less likely to be defaulted.

(1) Readability

Following Gao et al. (2022), but based on data from Lending Club, we evaluate the linguistic metrics, including readability, tone, and deception cues of the loan descriptions. The detailed exposition of these linguistic measures is provided as follows.

The readability of an article measures the ease of understanding and comprehension due to the writing style. As with Gao et al. (2022), we evaluate the readability as a composition of three elements: spelling errors, grammatical errors, and the Gunning FOG Index (Gunning, 1969). The first two components assert that fewer spelling or grammatical errors can enhance the readability of a text. By utilizing *LanguageTool*, a rules-based grammar-checker proposed by Naber (2003), we identify the spelling errors and grammatical mistakes. These two components are calculated as a proportion of incorrect words used in the text.

To verify whether readers can effortlessly comprehend the text, the FOG index measures reading complexity by counting hard words that contain at least three syllables. This index is constructed by following equation:

$$FOG = 0.4 \times (average\ sentence\ length + 100 \times average\ hard\ words), \quad (4)$$

where average sentence length is the average number of words contained in each sentence and average hard words is the average number of hard words used in the whole description. By utilizing the *Natural Language Toolkit* in Python, we compute the number of syllables for each word and determine the hard words.

After all the calculations, we standardize each component and sum them up. To intuitively evaluate the readability, we further standardize the sum and take the negative value, meaning a text is easier to understand with a higher readability score.

(2) Tone

Textual sentiment analysis has been extensively explored in finance and accounting research. A widely used approach for evaluating the sentiment is counting the words that correspond to specific terms in a pre-defined dictionary (see Henry, 2008; Loughran and McDonald, 2011; Loughran and McDonald, 2016). However, Huang et al. (2014) and Fan et al. (2021) argue that the outcomes from these lexicon approaches largely hinge on the selection and scoring of words within each dictionary, and they primarily use overlapping information pertinent to sentiment. Thus, these dictionary-based methodologies may fall short in extracting all the latent features or capturing biased information from the text data.

On the other hand, Huang et al. (2014) demonstrate that machine learning methods classify the textual tone more accurately than dictionary-based methods. By learning from large volumes of labeled data, machine learning methods can capture domain-specific emotive terms that may not be included in conventional dictionaries. Furthermore, these models can account for the influence of context on emotive words, thus enabling more accurate sentiment judgment. For instance, a word may carry positive sentiment in one context, while projecting negative sentiment in another. This flexibility endows machine learning methods with superior precision in sentiment analysis compared to lexicon approaches.

In this study, we employ the FinBERT model proposed by Huang et al. (2022) to extract sentiment scores from the loan descriptions. FinBERT is a BERT model pre-trained on financial communication texts, including Forms 10-K & 10-Q, earnings call transcripts, and analyst reports, outperforming other traditional machine learning algorithms in analyzing financial texts.[7] The output of FinBERT classifies the text into three categories, 'Positive', 'Negative', and 'Neutral', along with corresponding sentiment scores. Thus, we define the tone measure according to the classifier output. If the classifier output is 'Positive', the tone is equal to the corresponding sentiment score. If it is 'Negative', the tone is the sentiment score value with a negative sign. If the classifier shows 'Neutral', the tone is set to zero. To facilitate interpretation, we standardize the sentiment score by scaling it to have a mean of zero and a variance of one.

[7] With a sample of researcher-labeled sentences from analyst reports, the FinBERT developed by Huang et al. (2022) outperforms dictionary-based approaches and other machine learning algorithms. Their empirical results also show that the FinBERT better estimates the textual informativeness of earnings conference calls by at least 18%.

(3) Deception cues

The loan description provides borrowers with significant flexibility to provide additional information for a successful application. However, at the same time, borrowers can lie in the description to increase the chance of loan approval. Therefore, detecting deception cues can effectively reflect the strategies the borrowers take when submitting their applications. As a form of linguistic patterns and indicators, deception cues usually reveal inconsistencies in written texts, fabricated information, or misleading claims. Previous research has explored deception cues in various domains, such as online dating profiles (Toma and Hancock, 2012) and analyzing linguistic features in annual reports to detect fraud (Cecchini et al., 2010; Goel et al., 2010; Humpherys et al., 2011).

To determine whether deception cues influence the decision strategies by investors and borrowers, in this paper, we construct the measure of deception cues by following Gao et al. (2022). We focus on five key cues that have demonstrated significant associations with deception: exclusion words, motion verbs, first-person pronouns, third-person pronouns, and negative emotion words (Newman et al., 2003; Hancock et al., 2007; Pennebaker et al., 2003; Toma and Hancock, 2012), which can be summarized into three aspects as follows.

First of all, compared with telling the truth, lying is more cognitively taxing, which implies that liars would prefer straightforward statements. Related literature shows that using exclusion words (e.g., “but”, “except”, and “without”) demands greater cognition, while using motion verbs (e.g., “walk”, “move”, and “go”) requires less. Hence, liars usually use less exclusion words and more motion verbs.

Secondly, drawing upon the theory of linguistic distancing, researchers have observed that liars tend to distance themselves from their lies. That is, deceivers aim to create psychological separation between themselves and the deceptive information they present. This tendency is reflected in their written communication through an increased use of third-person pronouns (e.g., “he”, “him”, and “her”) and a decreased use of first-person pronouns (e.g., “I”, “we”, and “me”).

Last but not the least, studies have shown that deception elicits negative emotions, such as anxiety, shame, and guilt, in individuals. As a result, deceivers exhibit an elevated usage of negative emotional words (e.g., “hate”, “sorry”, and “worthless”), reflecting the emotional strain associated with engaging in deception.

To construct our measure of deception cues, we follow a three-step process. At first, we count these cues according to the LIWC dictionary (Pennebaker et al., 2001) and scale each of them by the total number of words in the loan descriptions, representing the relative frequency of each cue within the text. In the next step, we standardize each cue by transforming them with zero mean and unit variance. Finally, we construct the deception cues as the following equation and standardize the composed measure once more.

$$DeceptionCues = Motion + ThirdPerson + Neg.Emotion \\ -FirstPerson - Exculsion, \quad (5)$$

where the terms on the right-hand side are the standardized counts of corresponding cues.

5. Data

As aforementioned, the loan data used in this paper is obtained from Lending Club, one of the largest P2P lending platforms in the United States. The data set contains numeric variables (e.g., loan amount, FICO score, and debt-to-income ratio), categorical variables (e.g., credit grades), and unstructured data (such as loan description). To mitigate the potential impact on data reliability from the Lending Club governance scandal in 2016[8], we limited our data period from September 2007 to June 2016. In addition, since our research is primarily concerned with assessing the impact of loan description on loan performance, we excluded data without textual descriptions. After applying these data selection filters, we obtained a final dataset of 119,379 records, which formed the empirical basis of this study.

To address the research questions, we constructed the variables of interest for this study, as mentioned in Section 3.2. Table 1 provides the definitions of the response and explanatory variables. Regarding the explanatory variables, we focused on the non-

[8] In April 2016, an internal investigation of Lending Club found that $22 million in subprime loans sold in March and April 2016 to a single investor went against the investor’s expressed terms. The Securities and Exchange Commission (SEC) investigated Lending Club’s disclosures to investors after the CEO Renaud Laplanche resigned on May 9th, 2016.

pecuniary characteristics and the writing style as implicated information in the loan description. On the other hand, the response variables are based on the lending decisions made by investors (*Full Funding* and *Funding Ratio*) and the credit performance of borrowers (*Default*).

[Insert Table 1 here]

Panel A and B in Table 2 present the descriptive statistics of the response and explanatory variables. According to the results, approximately 73.2% of the loans were fully funded by lenders, while the overall average funding ratio reached 98.4%. This suggests that even, if borrowers failed to get the full loan, a substantial majority still managed to secure a significant proportion of their desired funds. In terms of loan defaults, the overall default rate was 15.5%.

[Insert Table 2 here]

Regarding control variables, related literature has found some financial (see Herzenstein et al., 2011, Emekter et al., 2015, Cai et al., 2016, and Tao et al., 2017) and demographic variables (see Pope and Sydnor, 2011, Chen et al., 2014, Barasinska and Schäfer, 2014, Gonzalez and Loureiro, 2014, Chen et al., 2017, and Chen et al., 2020) that serve as determinants of funding success and default probabilities in P2P lending (Basha et al., 2021). In order to filter important determinants from all the structural variables in the raw data, we adopt a backward selection approach to select the control variables.

Before the variable selection, we needed to perform the data preprocessing first, including label encoding, data cleaning, and data transformation. In label encoding, there is a categorical variable named *grade* in the raw dataset, which is a credit rating assigned by Lending Club according to the borrower's information. The credit rating scale descends from A to G, with A being the highest and G the lowest. To facilitate further analysis, we assigned a numerical value to these credit ratings, ranging from 1 (A, the highest) to 7 (G, the lowest).

In data cleaning, we firstly, eliminated variables where the proportion of missing values exceeds 10%. Then, we used the variance inflation factor[9] to remove variables

[9] Previous literature typically sets a threshold value of 10, above which indicates a high correlation.

with substantial correlation. For the remaining variables with missing values, we filled in with the mean of the variable. Whilst there are more complex imputation methods in statistics, such as multiple imputation or the k-nearest neighbors algorithm (k-NN), filling the missing values with the mean can simplify the research process and reduce the computational burden. Besides, since the proportion of missing values was not large, the impact of the missing value filling method on the overall analysis was deemed to be relatively acceptable.

In data transformation, all remaining variables were normalized and winsorized. After the data preprocessing, there remained 46 variables in total, which could be classified into three categories based on their attributes: borrowing information, personal information, and credit history.

Subsequently, we performed backward selection by engaging logistic regressions with *Full Funding* and *Default* as the response variables. We used the Bayesian information criteria (BIC) to evaluate the model selection performance and to identify control variables with respect to funding and default issues. Table 3 presents the regression results for the chosen control variables through the backward selection. When using *Full Funding* as the response variable, a total of 29 control variables were selected, including 3 from borrowing information, 9 from personal information, and 17 from credit history. When using *Default* as the response variable, a total of 22 variables were selected, including 3 from borrowing information, 10 from personal information, and 9 from credit history. Panels C, D, and E in Table 2 present the descriptive statistics of the control variables and the definitions of these variables are summarized in Appendix A.

[Insert Table 3 here]

According to the results of backward selection, we find that, regardless whether from a funding or default perspective, variables from loan information play crucial roles and provide intuitive insights. For example, a larger loan amount or a longer loan term tends to reduce the funding success rate and increase the default rate. In particular, a larger loan amount requires the borrower to have a more robust financial capacity to repay. A longer loan term implies that the borrower requires a longer turnaround time for capital utilization. The increased credit and liquidity risks make it more difficult to raise the funds.

However, the impact of the credit scores that assigned by Lending Club tells a different story. Though a lower credit score of the loan implies a higher default probability, the result in our backward selection shows that the loans with lower credit scores are more likely to be successfully funded. This finding is consistent with that of Wang and Tong (2020), who elicited that investors participating in Lending Club are, to some extent, seeking high returns by investing in loans with high default risk. But on the other hand, this counterintuitive finding is contrary to the results from Gao et al. (2023)[10], who used data from Prosper, another large P2P lending platform in the United States. The differences between results from Gao et al. (2023) and ours enlighten our proposed hypotheses and provide more possibilities. The empirical results of our linguistic metrics could be different from those of Gao et al. (2023), since investors on Lending Club seem to be less risk-averse than those on Prosper.

Besides, it is crucial to emphasize that in the Lending Club setting, the loan interest rate assigned by the platform is highly correlated[11] with the borrower's credit score. This implies that when we control for credit score in the empirical analysis, we are essentially controlling for the impact of financial returns and risks (primarily credit risk) across different loans.

6. Empirical results

6.1. Do investors have non-pecuniary preferences?

First of all, we examined whether lenders on Lending Club have non-pecuniary preferences by running regressions for the following equation:

$$Y_i = \alpha_0 + Nonpecuniary_i + \boldsymbol{Controls_i} + \varepsilon_i, \tag{6}$$

where the response variable Y_i is *full funding*, *funding ratio*, or *default*, respectively. The dummy variable $Nonpecuniary_i$ is equal to 1 when words indicating non-pecuniary preferences are detected in the corresponding loan description, and 0 otherwise. $\boldsymbol{Controls_i}$ contains the control variables filtered through the backward selection. Here, we should note that the set of control variables contains different

[10] Gao et al. (2023) found that on Prosper, a loan with lower credit score is less likely to get funded successfully, which means a lower probability of being fully funded, fewer number of bids, and lower proportion of funds borrowed (total $ amount bid / $ amount requested).

[11] According to the data used in this paper, the correlation coefficient between the loan interest rate and the credit score is 0.94.

variables depending on the response variable.

We also investigated the pro-environmental and prosocial preferences in detail by separating $Nonpecuniary_i$ into two dummy variables, as follows:

$$Y_i = \alpha_0 + Proenvironmental_i + Prosocial_i + \boldsymbol{Controls_i} + \varepsilon_i, \tag{7}$$

where $Proenvironmental_i$ (or $Prosocial_i$) is a dummy variable equal to 1 when words indicating pro-environmental (or prosocial) preference are detected in the loan description, and 0 otherwise. All other settings are the same as Equation (6).

[Insert Table 4 here]

The regression results are shown in Table 4. No matter whether for the whole non-pecuniary preferences or for the separated pro-environmental and prosocial preferences, the corresponding dummy variables significantly negatively impact funding success, both *Full Funding* and *Funding Ratio*.

Since financial return and credit risk are controlled by the *Credit Score*, this result indicates that our Hypothesis 1 is correct. The lenders on Lending Club do not have such non-pecuniary preferences as those investors on prosocial crowdfunding platforms do. More specifically, when facing two twin loans that have close return and risk, but only differ in terms of expressing prosocial or pro-environmental demand, the lenders would rather fund a loan without such descriptions.

Furthermore, the lenders would even show aversion towards these loans since the negative relationships are significant. A borrower who expresses prosocial/pro-environmental demand may find it difficult to secure a loan on Lending Club.

6.2. From the lender's perspective: Adverse selection

"The cure is worse than the disease."

Drawing upon the finding in Section 6.1, we want to determine whether an investor's lack of non-pecuniary preferences will mean putting themselves into dangerous situation. The results of columns (5) to (6) in Table 4 show that the non-pecuniary preferences have a significantly negative impact on default in general. However, whilst the pro-environmental feature has a slightly significant positive relation with default, the prosocial feature holds a significantly negative relation. By linking the results of columns (1) to (4), we find that investors' absence of non-pecuniary preferences can expose themselves to higher default risk, which is a result of adverse selection.

To explain this idea, we come back to the twin-loans story discussed before. When almost all factors that can effectively predict loan defaults have already been controlled, lenders have to choose between the twin loans depending on their subjective assessment of the default risk. Lenders can regard the loan with prosocial features as a riskier one, with one potential reason being that the borrowers of prosocial loans are usually considered as disadvantaged groups in society. However, this choice of borrowers actually exposes them to greater risks. Compared with loans without prosocial features, loans with them are actually less likely to default.

Above all, Hypothesis 2 is confirmed. Due to the lack of non-pecuniary preferences, lenders can fall into adverse selection. This misjudgment is mainly caused by the information asymmetry between borrowers and lenders. Lenders, due to their subjective judgment, may believe they possess more information about the loan default, which is actually, indeed, a misunderstanding.

Next, let's turn our attention to the impact of writing style on funding success and default. We ran the regression with the linguistic metrics as follows:

$$\begin{aligned} Y_i = \alpha_0 + Preferences_i + Readability_i + Tone_i + Deceptioncues_i \\ + \boldsymbol{Controls_i} + \varepsilon_i, \end{aligned} \tag{8}$$

where $Preferences_i$ is $Nonpecuniary_i$, the sum of $Proenvironmental_i$ and $Prosocial_i$, or directly set to be zero for purely testing the writing styles. The linguistics metrics, $Readability_i$, $Tone_i$, and $Deceptioncues_i$ are constructed according to Section 4.2. All other settings are same as for Equation (6).

[Insert Table 5 here]

The results of Equation (5) are summarized in Table 5. We can observe that the impact of three linguistic variables on default is consistent with our intuitive expectations and aligns with the findings of Gao et al. (2023). A loan with a more readable and more positive description has lower default probability, while one with more deception cues in the description has higher default risk.

However, the impacts on funding success are different when it comes to the tone and deception cues. First of all, the tone significantly negatively affects the funding success, which means a more negative loan description can help the borrower win funding. Then, another tricky adverse selection happens. The lender prefers to choose a loan with more negative expression, which is more likely to default in the end.

As a matter of fact, research has confirmed the role of linguistic style and emotional contagion in product sales and advertising (Lee and Theokary, 2021). The packaging of tone is a form of linguistic art. In Lending Club's world, a relatively negative description can lead the lender to misunderstand the true quality of the product, which is the default probability of the loan in our case. This misunderstanding implies the imperfect information and information asymmetry for investors, resulting in the occurrence of adverse selection.

6.3. From the borrower's perspective: Moral hazard

"Smooth talker with a forked tongue."

So far, from the perspective of the lender, we have discussed the adverse selection caused by the absence of non-pecuniary preferences and the tone of the loan description. Now, we turn to analysis of the deception cues from the borrowers' perspective. According to the results in Table 5, these cues positively impact funding success, while they also increase the default probability. In this scenario, moral hazard comes into play.

Compared with hard information, soft information is easier to fabricate and harder to be quantitatively evaluated with intuition. Borrowers successfully defraud the loan by deceiving or concealing the true default risk in their loan descriptions, being unable to repay the loans at maturity, thus exposing lenders to additional risk.

The reason for this moral hazard is the information asymmetry between both parties. The borrowers have the benefits of possessing more private information and fabricating inaccurate information. Therefore, for lenders who are in a relatively disadvantaged position, it is crucial to accurately discern the authenticity of the information.

Overall, the empirical results of the linguistic metrics confirm Hypothesis 3. Lenders on Lending Club come across the adverse selection caused by the tone and the moral hazard caused by the deception cues in the loan descriptions. In addition, the differences in empirical results between Gao et al. (2023) and this paper support the idea that investors on Lending Club seek higher returns and exhibit a higher risk tolerance than those on Prosper, which gradually leads them into the traps caused by information asymmetry.

6.4. Robustness Test

To examine the robustness of our findings in non-pecuniary preferences, we employed propensity score matching (PSM) to conduct additional regression analysis. PSM is a statistical technique that relies on counterfactual reasoning to establish causal relationships, which has been widely used in empirical finance research (e.g., Alda, 2020; Jang et al., 2022; Mu et al., 2023) to mitigate the influence of selection bias in observational studies. This matching process ensures the balance between the two groups of observed covariates, thereby reducing the impact of selection bias on estimating treatment effects.

To achieve this, we calculated propensity scores for the control variables selected in Section 5 and conducted one-to-one matching. The matched samples consisted of 28,262 observations for *Full Funding* and *Default*. Within the matched sample, half of the observations satisfy the non-pecuniary preferences, while the other half does not. To further validate the similarity between the treatment and control groups, we performed diagnostic tests on the matched sub-samples.

Normand (2001), Austin (2011), and Stuart et al. (2013) propose the use of standardized mean differences (SMD) as an indicator of matching performance, which is defined as the following equation:

$$SMD_i = \frac{\bar{X}_{i,treament} - \bar{X}_{i,control}}{\sqrt{\frac{s^2_{i,treament} + s^2_{i,control}}{2}}}, \tag{9}$$

where $\bar{X}_{i,treament}$ and $\bar{X}_{i,control}$ represent the sample means of the i-th covariate in the treatment and control groups, respectively, while $s^2_{i,treament}$ and $s^2_{i,control}$ represent the sample variance of the i-th covariate in the treatment and control groups, respectively. In particular, a threshold of 0.1 is commonly employed for this indicator. An SMD below 0.1 indicates a negligible difference in the mean of a covariate between the treatment group and control group.

In addition, achieving the balance between the treated and untreated groups in PSM involves ensuring that the distributions of covariates in the sub-samples are similar. Therefore, as suggested by Rubin (2001), we not only diagnosed the similarity of means using SMD, but also examined the variance ratio to assess balance in the variances of each covariate within the sub-samples. We employed a conventional threshold between 0.5 and 2 to compare the variance ratios in the balance test.

Table 6 displays the results of these balance tests. According to conventional criteria, our matching mechanism exhibits excellent performance, both in terms of *Full Funding* and *Default*. Almost all covariates pass the mean and variance diagnostic tests. Only *pub_rec_bankruptcies* slightly exceed the upper threshold in the variance ratio. However, these minor discrepancies do not overshadow the overall solidity of our results. These results suggest that we have reason to believe that the covariate means and variances across each subsample are balanced. Consequently, this indicates an absence of selection bias between borrowers, with and without, non-pecuniary characteristics.

[Insert Table 6 here]

Table 7 shows the logistic regression results with PSM. As mentioned above, PSM mitigates the confounding effects between the treatment and control groups, thus allowing for a more precise measurement. The regression results reveal that the non-pecuniary feature is significantly negative for both *Full Funding* and *Default*, which corroborates the findings of our previous empirical investigation. This lends further credence to our argument that our results aren't biased by sample selection issues.

[Insert Table 7 here]

7. Conclusion

In this paper, we have investigated the non-pecuniary preferences of lenders on P2P lending platforms. By extracting soft information in loan descriptions with NLP techniques, we have found that investors on Lending Club do not have pro-environmental or prosocial preferences, or even avoid investing in loans with such characteristics. However, the absence of non-pecuniary preferences can lead lenders into the trap of adverse selection since such loans have a lower default probability compared with others. In addition, we have also elicited that the writing style in loan descriptions can increase investors' default risk due to information asymmetry. From the perspective of lenders, the funding preference for loans with negative descriptions leads lenders to the adverse selection. From the perspective of borrowers, it imposes a moral hazard on investors since, whilst the loans with more deception cues are more likely to secure funding, they also come with a higher default probability.

Our findings can provide suggestions for the operation of P2P lending platforms. Firstly, the platform can increase the visibility of prosocial and pro-environmental loans on the lender's interface or provide member benefits for investors funding for these types of loans. This could not only reduce the potential default risk for investors, but also, enhance the platform's ESG performance and fulfill corporate social responsibility. Secondly, the platforms or investors could enhance the review of loan descriptions by themselves or add paid third-party verification, which would increase the accuracy of information and reduce information asymmetry.

References

Akerlof, G. (1970) The market for lemons: quality uncertainty and the market mechanism, *Quarterly Journal of Economics, 84*, 488–500.

Alda, M. (2020). ESG fund scores in UK SRI and conventional pension funds: Are the ESG concerns of the SRI niche affecting the conventional mainstream?. *Finance Research Letters*, *36*, 101313.

Allison, T. H., McKenny, A. F., & Short, J. C. (2013). The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect. *Journal of Business Venturing*, *28*(6), 690-707.

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

Atz, U., & Bholat, D. (2016). *Peer-to-peer lending and financial innovation in the United Kingdom.* Staff Working Paper (No. 598), Bank of England.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399-424.

Baker, M., Bergstresser, D., Serafeim, G., & Wurgler, J. (2022). The pricing and ownership of US green bonds. *Annual Review of Financial Economics, 14*, 415-437.

Barasinska, N., & Schäfer, D. (2014). Is crowdfunding different? Evidence on the relation between gender and funding success from a German peer-to-peer lending platform. *German Economic Review, 15*(4), 436-452.

Barber, B. M., Morse, A., & Yasuda, A. (2021). Impact investing. *Journal of Financial Economics, 139*(1), 162-185.

Basha, S. A., Elgammal, M. M., & Abuzayed, B. M. (2021). Online peer-to-peer lending: A review of the literature. *Electronic Commerce Research and Applications, 48*, 101069.

Baier, P., Berninger, M., & Kiesel, F. (2020). Environmental, social and governance reporting in annual reports: A textual analysis. Financial Markets, Institutions, and Instruments, *29*(3), 93-118.

Batson, C. D., & Powell, A. A. (2003). Altruism and prosocial behavior. In T. Millon & M. J. Lerner (Eds.), *Handbook of psychology: Personality and social psychology*, Vol. 5, pp. 463–484). John Wiley & Sons, Inc.

Bonnefon, J. F., Landier, A., Sastry, P. R., & Thesmar, D. (2022). The moral preferences of investors: Experimental evidence. *NBER Working Paper*.

Burtch, G., Ghose, A., & Wattal, S. (2014). Cultural differences and geography as determinants of online prosocial lending. *Mis Quarterly, 38*(3), 773-794.

Bruton, G., Khavul, S., Siegel, D., & Wright, M. (2015). New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer–to–peer innovations. *Entrepreneurship Theory and Practice*, *39*(1), 9-26.

Cai, S., Lin, X., Xu, D., & Fu, X. (2016). Judging online peer-to-peer lending behavior: A comparison of first-time and repeated borrowing requests. *Information & Management, 53*(7), 857-867.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Making words work: Using financial text as a predictor of financial events. *Decision support systems*, *50*(1), 164-175.

Chen, S., Gu, Y., Liu, Q., & Tse, Y. (2020). How do lenders evaluate borrowers in peer-to-peer lending in China?. *International Review of Economics & Finance, 69*, 651-662.

Chen, X., Huang, B., & Ye, D. (2018). The role of punctuation in P2P lending: Evidence from China. *Economic Modelling, 68*, 634-643.

Chen, D., Lai, F., & Lin, Z. (2014). A trust model for online peer-to-peer lending: a lender's perspective. *Information Technology and Management, 15*, 239-254.

Chen, D., Li, X., & Lai, F. (2017). Gender discrimination in online peer-to-peer credit lending: evidence from a lending platform in China. *Electronic Commerce Research, 17*, 553-583.

Dangl, T., Halling, M., Yu, J., & Zechner, J. (2023). Social Preferences and Corporate Investment. *SSRN working paper.*

Draganac, D., & Lu, K. (2022). Prosocial preference, social image motivation, and green asset bubble. *SSRN working paper*.

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

Dorfleitner, G., Oswald, E. M., & Zhang, R. (2021). From credit risk to social impact: On the funding determinants in interest-free peer-to-peer lending. *Journal of Business Ethics, 170*, 375-400.

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

70.

Fan, J., Xue, L., & Zhou, Y. (2021). How Much Can Machines Learn Finance From Chinese Text Data?. *SSRN working paper.*

Freedman, S., & Jin, G. Z. (2017). The information value of online social networks: Lessons from peer-to-peer lending. *International Journal of Industrial Organization, 51*, 185-222.

Galak, J., Small, D., & Stephen, A. T. (2011). Microfinance decision making: A field study of prosocial lending. *Journal of Marketing Research, 48*(SPL), S130-S137.

Galema, R. (2020). Credit rationing in P2P lending to SMEs: Do lender-borrower relationships matter?. *Journal of Corporate Finance, 65*, 101742.

Gao, Q., Lin, M., & Sias, R. (2023). Words matter: The role of readability, tone, and deception cues in online credit markets. *Journal of Financial and Quantitative Analysis*, *58*(1), 1-28.

Gao, M., Yen, J., & Liu, M. (2021). Determinants of defaults on P2P lending platforms in China. *International Review of Economics & Finance, 72*, 334-348.

Goel, S., Gangolly, J., Faerman, S. R., & Uzuner, O. (2010). Can linguistic predictors detect fraudulent financial filings?. *Journal of Emerging Technologies in Accounting*, *7*(1), 25-46.

Gonzalez, L., & Loureiro, Y. K. (2014). When can a photo increase credit? The impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance, 2*, 44-58.

Greiner, M. E., & Wang, H. (2009). The role of social capital in people-to-people lending marketplaces. *ICIS 2009 proceedings, 29*.

Gunning, R. (1969). The fog index after twenty years. *Journal of Business Communication*, *6*(2), 3-13.

Han, J. T., Chen, Q., Liu, J. G., Luo, X. L., & Fan, W. (2018). The persuasion of borrowers' voluntary information in peer to peer lending: An empirical study based on elaboration likelihood model. *Computers in Human Behavior, 78*, 200-214.

Henry, E. (2008). Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication (1973)*, *45*(4), 363-407.

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.

Hirshleifer, D. (2001). Investor psychology and asset pricing. *The journal of Finance*, *56*(4), 1533-1597.

Hu, M., Li, X., & Shi, Y. (2020). Adverse Selection in Credit Certificates: Evidence from a Peer-to-Peer Lending Platform. *SSRN working paper.*

Hu, J., & Song, W. (2017). The rational sense and balanced behavior of P2P investors: empirical analysis based on transaction data from Renrendai. com. *Journal of Financial Research*, *7*, 86-104.

Huang, A. H., Wang, H., & Yang, Y. (2022). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, forthcoming.

Huang, A. H., Zang, A. Y., & Zheng, R. (2014). Evidence on the information content of text in analyst reports. *The Accounting Review*, *89*(6), 2151-2180.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, *50*(3), 585-594.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, *62*(6), 1554-1577.

Jagtiani, J., & Lemieux, C. (2018). Do fintech lenders penetrate areas that are underserved by traditional banks?. *Journal of Economics and Business, 100*, 43-54.

Jang, G. Y., Kang, H. G., & Kim, W. (2022). Corporate executives' incentives and ESG performance. *Finance Research Letters*, *49*, 103187.

Jo, H., Kim, H. E., & Sim, M. (2022). Environmental preference, air pollution, and fund flows in China. *Pacific-Basin Finance Journal, 72*, 101723.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior?. *Environmental education research, 8*(3), 239-260.

Lagarde, C. (2018). Central banking and fintech: A brave new world. *Innovations: technology, governance, globalization, 12*(1-2), 4-8.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., & Gorski, S. (2011). Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, *39*(1), 19-37.

Lee, M. T., & Theokary, C. (2021). The superstar social media influencer: Exploiting linguistic style and emotional contagion over content?. *Journal of Business Research, 132*, 860-871.

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

Liu, Z., Shang, J., Wu, S. Y., & Chen, P. Y. (2020). Social collateral, soft information and online peer-to-peer lending: A theoretical model. *European Journal of Operational Research, 281*(2), 428-438.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65.

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, *54*(4), 1187-1230.

Loughran, T., McDonald, B., & Otteson, J. R. (2023). How have corporate codes of ethics responded to an era of increased scrutiny?. *Journal of Business Ethics*, *183*(4), 1029-1044.

Mesmer-Magnus, J., Viswesvaran, C., & Wiernik, B. M. (2012). The role of commitment in bridging the gap between organizational sustainability and environmental sustainability. In S. E. Jackson, D. S. Ones, & S. Dilchert (Eds.), *Managing human resources for environmental sustainability* (pp. 155–186). Jossey-Bass/Wiley.

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

Mu, W., Liu, K., Tao, Y., & Ye, Y. (2023). Digital finance and corporate ESG. *Finance Research Letters*, *51*, 103426.

Naber, D. (2003). A rule-based style and grammar checker.

Normand, S. L. T., Landrum, M. B., Guadagnoli, E., Ayanian, J. Z., Ryan, T. J., Cleary, P. D., & McNeil, B. J. (2001). Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores. *Journal of Clinical Epidemiology*, *54*(4), 387-398.

Nowak, A., Ross, A., & Yencha, C. (2018). Small business borrowing and peer-to-peer lending: Evidence from lending club. *Contemporary Economic Policy*, *36*(2), 318-336.

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2021). Sustainable investing in equilibrium. *Journal of Financial Economics, 142*(2), 550-571.

Pástor, Ľ., Stambaugh, R. F., & Taylor, L. A. (2022). Dissecting green returns. *Journal of Financial Economics, 146*(2), 403-424.

Peng, H. F., Zhao, H. Y., & Zhou, Y. (2016). Could the loan description affect the loan cost and the loan success rate? Textual analysis based on the loan description.

Journal of Financial Research, *4*, 158-173.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*(2001), 2001.

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

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*, 169-188.

Shao, S., & Bo, H. (2022). Behavioural aspects of China's P2P lending. *The European Journal of Finance, 28*(1), 30-45.

Spence, M. (2002). Signaling in retrospect and the informational structure of markets. *American economic review, 92*(3), 434-459.

Stiglitz, J. and Weiss, A. (1981) Credit rationing in markets with imperfect information, *American Economic Review, 71*, 393–419.

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, *66*(8), S84-S90.

Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of environmental psychology, 29*(3), 309-317.

Tao, Q., Dong, Y., & Lin, Z. (2017). Who can get money? Evidence from the Chinese peer-to-peer lending platform. *Information Systems Frontiers, 19*, 425-441.

Toma, C. L., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, *62*(1), 78-97.

Vallee, B., & Zeng, Y. (2019). Marketplace lending: A new banking paradigm?. *The Review of Financial Studies, 32*(5), 1939-1982.

Wang, J., & Li, R. (2023). Asymmetric information in peer-to-peer lending: empirical evidence from China. *Finance Research Letters, 51*, 103452.

Wang, C., & Tong, L. (2020). Lender rationality and trade-off behavior: Evidence from Lending Club and Renrendai. *International Review of Economics and Finance*, *70*, 55-66.

Wang, W., Zhu, K., Wang, H., & Wu, Y. C. J. (2017). The impact of sentiment orientations on successful crowdfunding campaigns through text analytics. *IET Software*, *11*(5), 229-238.

Wei, Z., & Lin, M. (2017). Market mechanisms in online peer-to-peer lending. *Management Science, 63*(12), 4236-4257.

Yao, J., Chen, J., Wei, J., Chen, Y., & Yang, S. (2019). The relationship between soft information in loan titles and online peer-to-peer lending: evidence from RenRenDai platform. *Electronic Commerce Research, 19*, 111-129.

Yoon, Y., Li, Y., & Feng, Y. (2019). Factors affecting platform default risk in online peer-to-peer (P2P) lending business: an empirical study using Chinese online P2P platform data. *Electronic Commerce Research, 19*, 131-158.

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

Zerbib, O. D. (2019). The effect of pro-environmental preferences on bond prices: Evidence from green bonds. *Journal of banking & finance, 98*, 39-60.

Figures

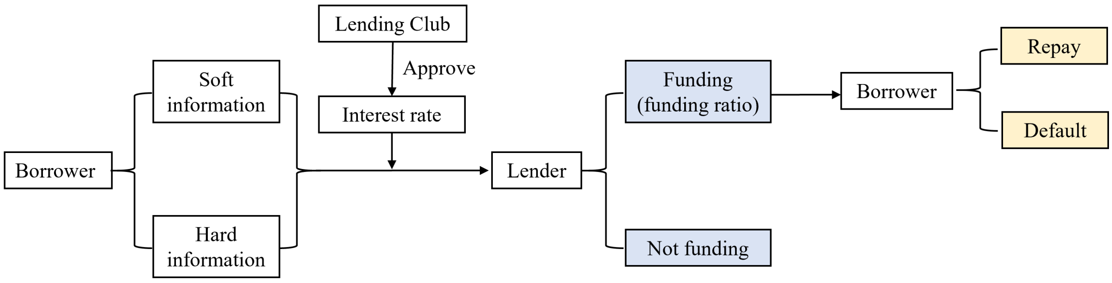

Figure 1. The mechanism of P2P lending: An extensive-form game

Note: This figure presents the working mechanism of P2P lending on Lending Club, which can be described as an extensive-form game.

Tables

Table 1. Definitions of variables

This table presents the definitions of response variables, including Full Funding, Funding Ratio, and Default, as well as explanatory variables, including *Nonpecuniary*, *Proenvironmental, Prosocial*, *Readability*, *Tone*, and *DeceptionCues*.

Variable	Definition
Response variables	
Full Funding	Measuring whether a loan is fully funded by the investor: 1, if the loan is fully funded, and 0, if not.
Funding Ratio	Measuring the proportion of a loan funded by investors: a value between 0 and 1.
Default	Measuring whether a loan is defaulted by the borrower: 1, if the loan is defaulted, and 0, if not.
Explanatory variables	
Nonpecuniary	Measuring the non-pecuniary characteristics of a loan and assigning a binary score of 1, if the loan exhibits such features and 0, if it does not.
Proenvironmental	Measuring the pro-environmental characteristics of a loan and assigning a binary score of 1, if the loan exhibits such features and 0, if it does not.
Prosocial	Measuring the prosocial characteristics of a loan and assigning a binary score of 1, if the loan exhibits such features and 0, if it does not.
Readability	Measuring the readability in loan description, which is composed of *Spelling Error*, *Grammar Error*, and *Lexical Complexity.*
Tone	Measuring the tone in loan description, which is calculated by FinBERT, as proposed by Huang et al. (2022).
DeceptionCues	Measuring the deception cues in loan description, which is composed of *Exclusive Words, Motion Words, First Person Pronouns, Third Person Pronouns,* and *Negative Emotion Words.*

Table 2. Descriptive statistics

This table presents the descriptive statistics of the variables. Panel A contains the response variables *Full Funding*, *Funding Ratio,* and *Default*. Panel B contains the explanatory variables *Nonpecuniary*, *Readability*, *Tone,* and *DeceptionCues*. Panels C, D, and E present the control variables selected by the backward selection approach, which are divided into three categories: Loan Information, Personal Information, and Credit History. The definitions of the control variables are listed in Appendix A.

Variables	N	Mean	Median	Std. Dev.	Min.	Max.
Panel A. Response variables						
Full Funding	119,379	0.732	1.000	0.443	0.000	1.000
Funding Ratio	119,379	0.984	1.000	0.094	0.000	1.000
Default	119,379	0.155	0.000	0.362	0.000	1.000
Panel B. Explanatory variables						
Nonpecuniary	119,379	0.118	0.000	0.323	0.000	1.000
Readability	119,379	0.000	0.267	1.000	-22.714	1.700
Tone	119,379	0.000	-0.447	1.000	-2.738	1.844
DeceptionCues	119,379	0.000	0.004	1.000	-7.492	17.431
Panel C. Loan information						
loan_amnt	119,379	13.686	12.000	7.597	3.000	31.000
term	119,379	3.417	3.000	0.812	3.000	5.000
grade	119,379	2.595	2.000	1.294	1.000	7.000
Panel D. Personal information						
annual_inc	119,379	72.088	62.000	53.305	4.000	7141.778
avg_cur_bal	119,379	14.116	12.656	13.434	0.000	958.084
Dti	119,379	16.234	16.010	7.510	0.000	34.990
fico_score	119,379	703.827	697.000	32.889	627.000	847.500
open_act_il	119,379	2.151	1.592	2.355	0.000	38.371
percent_bc_gt_75	119,379	49.823	50.000	32.912	0.000	100.000
revol_bal	119,379	15.723	11.696	19.133	0.000	1746.716
revol_util	119,379	0.556	0.577	0.246	0.000	1.393
total_cu_tl	119,379	1.539	1.378	1.116	0.000	9.961
total_il_high_credit_limit	119,379	37.197	33.263	32.285	0.000	1214.546
total_rev_hi_lim	119,379	29.874	24.400	26.690	0.000	2013.133
num_bc_tl	119,379	8.925	8.000	4.413	0.000	65.000
num_op_rev_tl	119,379	7.895	7.035	3.583	0.000	58.000
Panel E. Credit history						
delinq_2yrs	119,379	0.218	0.000	0.664	0.000	22.000
inq_last_12m	119,379	2.914	2.159	6.543	0.000	70.000
inq_last_6mths	119,379	0.833	1.000	1.052	0.000	8.000
mo_sin_old_il_acct	119,379	123.391	123.000	41.654	0.000	482.000
mo_sin_old_rev_tl_op	119,379	175.633	162.595	77.183	0.000	760.000
mo_sin_rcnt_rev_tl_op	119,379	13.625	11.000	13.318	0.000	264.000
mo_sin_rcnt_tl	119,379	8.693	7.531	7.825	0.000	211.000
mths_since_last_record	119,379	81.725	82.301	16.047	0.000	130.000
mths_since_recent_bc	119,379	24.861	19.585	25.334	0.000	527.000
mths_since_recent_inq	119,379	7.636	7.132	5.494	0.000	56.206
mths_since_recent_revol_delinq	119,379	37.736	37.550	11.724	0.000	146.000
num_accts_ever_120_pd	119,379	0.337	0.000	0.832	0.000	25.000
num_rev_tl_bal_gt_0	119,379	5.596	5.282	2.398	0.000	29.000
num_tl_120dpd_2m	119,379	0.002	0.000	0.021	0.000	2.000
num_tl_30dpd	119,379	0.002	0.000	0.038	0.000	3.000
num_tl_90g_dpd_24m	119,379	0.081	0.000	0.323	0.000	22.000
pct_tl_nvr_dlq	119,379	95.841	98.202	6.356	16.000	100.000
pub_rec_bankruptcies	119,379	0.079	0.000	0.282	0.000	7.000
tot_coll_amt	119,379	0.340	0.000	1.378	0.000	95.806

Table 3. Backward selection with logistic regressions

This table displays the results of a regression analysis using logistic regression and backward selection, with the BIC as the model evaluation measure. The sample includes a total of 119,379 funded loans, with loan descriptions provided from September 2007 to June 2016. Panel (A) presents the analysis results for *Full Funding*, in which backward selection yielded 29 variables, including: 3 from Borrowing Information, 9 from Personal Information, and 17 from Credit History. Panel (B) illustrates the analysis results for *Default*, with backward selection yielding 22 variables: 3 from Borrowing Information, 10 from Personal Information, and 9 from Credit History. The values of the standard errors are reported in squared brackets. Statistical significance levels are denoted as *** for 1%, ** for 5%, and * for 10% significance. More details on the variable definitions are shown in Appendix A.

Panel A. *Full Funding*			
intercept	1.149***[0.007]	Credit history [17]	
Borrowing information [3]		*delinq_2yrs*	0.056***[0.014]
loan_amnt	-0.398***[0.009]	*inq_last_12m*	-0.273***[0.017]
term	-0.212***[0.009]	*mo_sin_old_il_acct*	0.070***[0.009]
grade	0.194***[0.012]	*mo_sin_old_rev_tl_op*	0.090***[0.010]
Personal information [9]		*mo_sin_rcnt_rev_tl_op*	0.078***[0.011]
annual_inc	0.171***[0.010]	*mo_sin_rcnt_tl*	-0.103***[0.011]
dti	0.159***[0.010]	*mths_since_last_record*	0.338***[0.012]
fico_score	-0.457***[0.017]	*mths_since_recent_inq*	-0.461***[0.015]
open_act_il	0.118***[0.010]	*mths_since_recent_revol_delinq*	0.095***[0.010]
percent_bc_gt_75	-0.115***[0.009]	*num_accts_ever_120_pd*	-0.070***[0.011]
revol_bal	-0.264***[0.019]	*num_rev_tl_bal_gt_0*	-0.067***[0.009]
revol_util	0.063***[0.016]	*num_tl_120dpd_2m*	-0.360***[0.008]
total_il_high_credit_limit	-0.208***[0.010]	*num_tl_30dpd*	-0.230***[0.010]
total_rev_hi_lim	0.171***[0.018]	*num_tl_90g_dpd_24m*	-0.077***[0.011]
		pct_tl_nvr_dlq	-0.080***[0.013]
		pub_rec_bankruptcies	-0.046***[0.009]
		tot_coll_amt	-0.135***[0.010]
Observation	119,379		
Adj. R^2 (%)	13.282		
Panel B. *Default*			
intercept	-1.887***[0.009]	Credit history [9]	
Borrowing information [3]		*inq_last_6mths*	0.102***[0.009]
loan_amnt	0.124***[0.011]	*delinq_2yrs*	0.065***[0.009]
term	0.314***[0.010]	*mo_sin_old_rev_tl_op*	-0.035***[0.010]
grade	0.257***[0.011]	*mo_sin_rcnt_tl*	-0.094***[0.011]
Personal information [10]		*mths_since_last_record*	-0.172***[0.010]
annual_inc	-0.265***[0.017]	*mths_since_recent_bc*	-0.056***[0.011]
avg_cur_bal	-0.049***[0.012]	*pct_tl_nvr_dlq*	0.084***[0.011]
dti	0.121***[0.010]	*pub_rec_bankruptcies*	0.038***[0.008]
revol_bal	0.194***[0.026]	*tot_coll_amt*	0.046***[0.008]
num_bc_tl	-0.101***[0.012]		
num_op_rev_tl	0.144***[0.012]		
percent_bc_gt_75	0.088***[0.011]		
total_cu_tl	-0.209***[0.010]		
total_rev_hi_lim	-0.305***[0.028]		
open_act_il	-0.057***[0.009]		
Observation	119,379		
Adj. R^2 (%)	8.264		

Table 4. Regression results on non-pecuniary preferences

This table presents the regression results of Equations (5) and (6). The dummy variables *Nonpecuniary*, *Proenvironmental*, and *Prosocial* are generated based on the dictionary method proposed in this study, primarily identifying whether the borrower's loan description contains relative words. If the description contains words related to the pro-environment, then *Pro-environmental* = 1. If it contains words related to prosocial, then *Pro-environmental* = 1. If either *Proenvironmental* = 1 or *Prosocial* = 1, then *Nonpecuniary* = 1. Columns (1) to (2) show the logistic regression of *Full Funding* on non-pecuniary preferences, columns (3) to (4) display the Tobit regression of the *Funding Ratio* on non-pecuniary preferences, and columns (5) to (6) illustrate the logistic regression of the *Default* on non-pecuniary preferences. The sample includes a total of 119,379 loans from September 2007 to June 2016. Control variables are included. The values of the standard errors are reported in squared brackets. Statistical significance levels are denoted as *** for 1%, ** for 5%, and * for 10% significance.

	Full Funding		***Funding Ratio***		***Default***	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.201***	1.202***	2.697***	2.699***	-1.877***	-1.876***
	[0.008]	[0.008]	[0.024]	[0.024]	[0.010]	[0.010]
Nonpecuniary	-0.407***		-0.329***		-0.090***	
	[0.021]		[0.002]		[0.027]	
Proenvironmental		-0.260***		-0.021***		0.119*
		[0.050]		[0.004]		[0.062]
Prosocial		-0.423***		-0.034***		-0.130***
		[0.022]		[0.002]		[0.029]
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observation	119,379	119,379	119,379	119,379	119,379	119,379
Adj. R^2 (%)	13.552	13.571	59.520	59.583	8.271	8.290

Table 5. Regression results on linguistic metrics

This table presents the regression results of Equation (7). The dummy variables *Nonpecuniary*, *Proenvironmental*, and *Prosocial* are generated based on the dictionary method proposed in this study, primarily identifying whether the borrower's loan description contains related words. If the description contains words related to the pro-environment, then *Pro-environmental* = 1. If it contains words related to prosocial, then *Pro-environmental* = 1. If either *Proenvironmental* = 1 or *Prosocial* = 1, then *Nonpecuniary* = 1. Linguistic metrics comprise *Readability*, *Tone*, and *DeceptionCues*. Columns (1) to (2) show the logistic regression of *Full Funding* on non-pecuniary preferences, columns (3) to (4) display the Tobit regression of the *Funding Ratio* on non-pecuniary preferences, and columns (5) to (6) illustrate the logistic regression of the *Default* on non-pecuniary preferences. The sample includes a total of 119,379 loans from September 2007 to June 2016. Control variables are included. The values of the standard errors are reported in squared brackets. Statistical significance levels are denoted as *** for 1%, ** for 5%, and * for 10% significance.

	Full Funding			***Funding Ratio***			***Default***		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	1.153***	1.195***	1.1996***	2.710***	2.693***	2.694***	-1.889***	-1.875***	-1.874***
	[0.007]	[0.008]	[0.008]	[0.026]	[0.024]	[0.024]	[0.010]	[0.010]	[0.010]
Nonpecuniary		-0.333***			-0.030***			-0.129***	
		[0.022]			[0.002]			[0.028]	
Proenvironmental			-0.204***			-0.019***			0.061
			[0.051]			[0.004]			[0.063]
Prosocial			-0.349***			-0.031***			-0.169***
			[0.023]			[0.002]			[0.029]
Readability	0.101***	0.078***	0.076***	0.004***	0.002***	0.002***	-0.061***	-0.070***	-0.071***
	[0.007]	[0.007]	[0.007]	[0.001]	[0.001]	[0.001]	[0.008]	[0.008]	[0.008]
Tone	-0.058***	-0.046***	-0.046***	-0.004***	-0.003***	-0.003***	-0.044***	-0.039***	-0.038***
	[0.007]	[0.007]	[0.007]	[0.001]	[0.001]	[0.001]	[0.009]	[0.009]	[0.009]
DeceptionCues	0.037***	0.036***	0.036***	0.002***	0.002***	0.002***	0.012	0.012	0.012
	[0.007]	[0.007]	[0.007]	[0.001]	[0.001]	[0.001]	[0.008]	[0.008]	[0.008]
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	119,379	119,379	119,379	119,379	119,379	119,379	119,379	119,379	119,379
Adj. R^2 (%)	13.533	13.701	13.711	59.060	59.634	59.681	8.352	8.371	8.384

Table 6. Diagnostic test results

This table presents the diagnostic test results for PSM designed to evaluate the balance quality of control variables in relation to *Full Funding* and *Default*. The sample includes a total of 119,379 funded loans, with loan descriptions provided from September 2007 to June 2016. Panel A shows the outcomes for *Full Funding*, wherein the covariates have been selected via backward selection using logistic regression. Similarly, Panel B shows the results for *Default*, with covariate variables selected by the same process. Balance indicators have been contrasted against a benchmark value of 0.1 for standardized mean differences (SMD) and a range of 0.5-2 for adjusted variance ratios (Var. Ratio), reflecting thresholds habitually employed in previous research. * indicates balance indicators that have successfully passed the diagnostic tests.

Panel A. *Full Funding*					
	SMD	Var.Ratio		SMD	Var.Ratio
loan_amnt	0.019*	0.930*	*mo_sin_old_il_acct*	0.020*	0.982*
term	0.006*	1.058*	*mo_sin_old_rev_tl_op*	0.020*	1.113*
grade	0.015*	0.969*	*mo_sin_rcnt_rev_tl_op*	0.011*	0.834*
annual_inc	0.005*	1.014*	*mo_sin_rcnt_tl*	0.018*	1.076*
Dti	0.001*	1.016*	*mths_since_last_record*	0.007*	1.294*
fico_score	0.003*	1.071*	*mths_since_recent_inq*	0.020*	1.305*
open_act_il	0.003*	1.155*	*mths_since_recent_revol_delinq*	0.004*	1.359*
percent_bc_gt_75	0.006*	1.012*	*num_accts_ever_120_pd*	0.006*	0.988*
revol_bal	0.004*	1.723*	*num_rev_tl_bal_gt_0*	0.005*	1.041*
revol_util	0.010*	1.518*	*num_tl_120dpd_2m*	0.003*	1.020*
total_coll_amnt	0.002*	1.143*	*num_tl_30dpd*	0.001*	1.234*
total_il_high_credit_limi	0.004*	1.298*	*num_tl_90g_dpd_24m*	0.017*	1.379*
total_rev_hi_lim	0.010*	1.166*	*pct_tl_nvr_dlq*	0.018*	1.146*
delinq_2yrs	0.008*	1.373*	*pub_rec_bankruptcies*	0.005*	2.107
inq_last_12m	0.004*	1.863*			

Panel B. *Default*					
	SMD	Var.Ratio		SMD	Var.Ratio
loan_amnt	0.006*	0.892*	*delinq_2yrs*	0.013*	1.007*
term	0.010*	1.016*	*mo_sin_old_rev_tl_op*	0.019*	0.826*
grade	0.001*	0.889*	*mo_sin_rcnt_tl*	0.002*	1.288*
annual_inc	0.007*	1.012*	*mths_since_last_record*	0.010*	1.570*
avg_cur_bal	0.022*	1.263*	*mths_since_recent_bc*	0.003*	0.848*
Dti	0.007*	2.581*	*pct_tl_nvr_dlq*	0.005*	1.473*
revol_bal	0.002*	1.025*	*pub_rec_bankruptcies*	0.019*	1.153*
num_bc_tl	0.024*	1.996*			
num_op_rev_tl	0.024*	1.256*			
percent_bc_gt_75	0.001*	1.103*			
tot_coll_amt	0.057*	0.909*			
total_cu_tl	0.006*	1.110*			
total_rev_hi_lim	0.007*	1.004*			
open_act_il	0.007*	1.967*			
inq_last_6mths	0.001*	0.998*			

Table 7. Logistic regression results with PSM

The table presents the logistic regression results for *Full Funding* and *Default* using the matched data. The matched sample consists of 28,262 observations, equally divided between samples with or without non-pecuniary preferences. Both sample sets have undergone PSM and have passed diagnostic tests to ensure a consistent distribution of covariates. Control variables are included. The values of the standard errors are reported in squared brackets. Statistical significance levels are denoted as *** for 1%, ** for 5%, and * for 10% significance.

	Full Funding	***Default***
Intercept	0.8571***	-1.9499***
	[0.021]	[0.020]
Nonpecuniary	-0.3964***	-0.0410**
	[0.028]	[0.017]
Controls	Yes	Yes
Observation	28,262	28,262
R^2	0.1771	0.0868

Appendix A

Table A1. Definitions of the control variables

Variables	Definition
Panel A. Borrowing information	
loan_amnt	The listed amount of the loan applied for by the borrower.
term	The number of payments on the loan. Values are in months.
grade	The loan grade assigned by Lending Club, which is categorized from A (highest) to G (lowest) based on the borrower's credit rating, is utilized in our study. In this research context, we represent these loan grades numerically, ranging from 1 to 7.
Panel B. Personal information	
annual_inc	The self-reported annual income provided by the borrower during registration.
avg_cur_bal	Average current balance of all accounts.
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested Lending Club loan, divided by the borrower's self-reported monthly income.
fico_score	The borrower's average FICO at loan origination.
open_act_il	Number of currently active installment trades.
percent_bc_gt_75	Percentage of all bankcard accounts > 75% of limit.
revol_bal	Total credit revolving balance.
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
total_cu_tl	Number of finance trades.
total_il_high_credit_limit	Total installment high credit/credit limit.
total_rev_hi_lim	Total revolving high credit/credit limit.
num_bc_tl	Number of bankcard accounts.
num_op_rev_tl	Number of open revolving accounts.

Table A1. Definitions (continued)

Panel C. Credit history	
delinq_2yrs	Number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years.
inq_last_12m	Number of credit inquiries in the past 12 months.
inq_last_6mths	Number of inquiries in the past 6 months.
mo_sin_old_il_acct	Months since oldest bank installment account opened.
mo_sin_old_rev_tl_op	Months since oldest revolving account opened.
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened.
mo_sin_rcnt_tl	Months since most recent account opened.
mths_since_last_record	Months since the last public record.
mths_since_recent_bc	Months since most recent bankcard account opened.
mths_since_recent_inq	Months since most recent inquiry.
mths_since_recent_revol_delinq	Months since most recent revolving delinquency.
num_accts_ever_120_pd	Number of accounts that are ever past due 120 or more days.
num_rev_tl_bal_gt_0	Number of revolving trades with balance >0.
num_tl_120dpd_2m	Number of accounts currently 120 days past due.
num_tl_30dpd	Number of accounts currently 30 days past due.
num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months.
pct_tl_nvr_dlq	Percent of trades never delinquent.
pub_rec_bankruptcies	Number of public record bankruptcies.
tot_coll_amt	Total collection amounts ever owed.