Journal of Banking & Finance 64 (2016) 169–187

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms

Gregor Dorfleitner [a,b,*], Christopher Priberny [a], Stephanie Schuster [a], Johannes Stoiber [c], Martina Weber [a], Ivan de Castro [d], Julia Kammler [a]

[a] *University of Regensburg, Department of Finance, 93040 Regensburg, Germany*
[b] *CERMi (Centre for European Research in Microfinance), Belgium*
[c] *Humboldt-Universität zu Berlin, Berlin, Germany*
[d] *Technische Universität Berlin, Germany*

ARTICLE INFO

Article history:
Received 12 December 2014
Accepted 20 November 2015
Available online 15 December 2015

JEL classification:
G20
G32

Keywords:
Peer-to-peer lending
Soft information
Funding probability
Probability of default

ABSTRACT

We examine the relation of soft factors that are derived from the description texts to the probability of successful funding and to the default probability in peer-to-peer lending for two leading European platforms. We find that spelling errors, text length and the mentioning of positive emotion evoking keywords predict the funding probability on the less restrictive of both platforms, which even accepts applications without credit scores. This platform also shows a better risk-return profile. Conditional on being funded, text-related factors hardly predict default probabilities in peer-to-peer lending for both platforms.

1. Introduction

Peer-to-Peer (P2P) lending is regarded as being a major innovation in the area of retail banking. In recent years, the number of platforms offering such services as well as the volume of transactions have been steadily increasing. P2P lending, as one facet of crowdfunding and thereby as a form of financial disintermediation, is different to classical banking since a crowd of peers decides whether a loan is granted. Even if classical hard facts such as the solvency of a borrower or the purpose of the loan are relevant for the granting decision, additional information about the borrower's individual situation, the soft information, also enters into the P2P lending decision process. This article examines the relation of soft information which are derived from the description text of the loan application to the probability of successful funding as well as to the default probability of granted loans. To this end, we are the first to compare the transactions and loan applications on the two leading European P2P platforms located in Germany, namely Smava and Auxmoney, with respect to these soft factors. While Smava is more restrictive in admitting loan applications in order to ensure a minimum level of credit quality, Auxmoney does not require credit scores and leaves more room for voluntary information. Our study emphasizes the role of the soft information related to loan description texts written by the loan applicants, in particular orthography, text length and the presence of social and emotional keywords. The major contribution lies in the comprehensive approach, with which we are able to draw the big picture. We use an extensive set of controls, comprising other known soft factors and the extremely important variable interest rate and we simultaneously study the relation to the funding and to the default probability. We even assess the profitability of the investments, which provides a quantitative link between the willingness to fund, the danger of default and the rationality of the investors. Additionally, by considering *two differently designed platforms*, both serving the same market in the same cultural environment, we obtain insights into the question of how the value of soft information depends on the presence or absence of hard facts.

P2P platforms provide lots of data on real transactions. Since, in contrast to bank-based lending, those applications that do not lead to a transaction can also be observed, such platforms constitute a form of natural experiment on loan granting decisions. Thus many

* Corresponding author at: University of Regensburg, Department of Finance, 93040 Regensburg, Germany.

E-mail addresses: gregor.dorfleitner@ur.de (G. Dorfleitner), christopher.priberny@ur.de (C. Priberny), martina.weber@ur.de (M. Weber).

http://dx.doi.org/10.1016/j.jbankfin.2015.11.009
0378-4266/

researchers focus on this relatively new phenomenon. The hitherto best-researched P2P platform is Prosper operating in the U.S. and providing current and historical loan-related information for public download. Based on data from Prosper, previous research finds evidence towards an effect of soft information on funding success, interest rates and default rates. Iyer et al. (2014), for example, show that lenders are able to determine information on the creditworthiness of a potential borrower from soft factors such as the number of friend endorsements or the self-reported purpose of the loan. In addition, several authors examine the effect of including a picture in the loan proposal and aspects of the applicant's appearance (Pope and Sydnor, 2011; Ravina, 2012). Gao and Lin (2015) show that readability, positivity, objectivity and deception cues concerning description texts are related to loan defaults on Prosper. This article contributes to this stream of literature and analyzes the description texts. We put special emphasis on orthography here, as some psychological studies like Figueredo and Varnhagen (2005) and Kreiner et al. (2002) support the conjecture that spelling errors in the description text impair the perception of the creditworthiness of the applicant. Other aspects are the signaling role of the text length and certain keywords appearing in the description. Some keywords that are able to evoke special emotions may have a positive effect on the probability of successful funding.

Our investigation is based on a simultaneous IV probit regression approach to overcome endogeneity issues related to the interest rate and identifies influencing factors on the funding and the default probability. We use 76,945 loan applications from Auxmoney and 10,423 from Smava to examine the funding success and 3,298 closed granted loans from Auxmoney and 2,216 from Smava, for which the event of a default or a non-default can be determined without doubt, in order to research the default probability. We use all data available on each platform archive in October 2013, resulting in the observation periods March 2008 to September 2013 (Auxmoney) and February 2007 to September 2013 (Smava).

Our results show that investors on Auxmoney gain a higher return accompanied by a lower default rate compared to Smava. Smava only allows loan applications with a minimum credit score, and therefore a large share of loans are granted. Our results indicate that soft factors play a minor role in explaining the funding probability and the default probability on Smava and investors rely more on hard facts such as solvency scores or the suggested interest rate. This is in contrast to Auxmoney, on which the provision of a credit score is not mandatory and only a minor share of loans are granted. For this platform, many soft factors related to the description text show significant coefficients in the funding probability regressions, whereas only few of them also have a significant effect in the default regressions. In particular, we find evidence supporting the fact that spelling errors are negatively related and the length of the description text has an inverse u-shaped relation with the probability of successful funding. Keywords evoking positive emotions also significantly relate to the funding success. Another important factor on both platforms is the interest rate suggested in the loan applications. Our findings show that on both platforms, investors associate a higher interest rate with a lower solvency and shrink back from funding those loans.

Concluding, investors appear to be capable of identifying creditworthy borrowers with the help of soft information even though hard facts like credit scores are not provided. However, if hard facts of a certain quality are generally required by the platform then soft information plays a minor role.

The remainder of the paper is structured as follows: In Section 2, we review the relevant literature, while in Section 3 we develop hypotheses concerning soft factors derived from the description texts. In Section 4 we present a description of our data and the used methodology. Section 5 discusses the results on the funding and the default probability—including robustness checks—and compares both platforms. Section 6 concludes.

2. Literature review

Since the start of the first P2P lending platform Zopa in 2006, a considerable amount of academic literature has evolved, in which several strands can be identified. Many of the studies focus on the leading U.S. P2P lending platform Prosper, which has made its data publicly available.

One strand of literature analyzes the economic mechanisms of P2P markets (see Agrawal et al., 2013; Belleflamme et al., 2014; Chen et al., 2014; Gerber et al., 2012; Giudici et al., 2012; Hemer, 2011; Moenninghoff and Wieandt, 2013; Solomon and Wash, 2014) and also discusses legal aspects and other crowdfunding models. Like in bank-based lending, borrowers have an incentive to overplay their financial situation in their application (see Berger and Gleisner, 2009; Weiss et al., 2010). Thus, creditors in P2P markets are dependent on a suitable platform design that helps to overcome asymmetric information (see Diamond, 1984). Freedman and Jin (2008) and Weiss et al. (2010) identify adverse selection effects on P2P platforms. The P2P platform Prosper offers a social network, in which borrowers and lenders can interact. Both, creditors and debtors, benefit from this network which helps to mitigate information asymmetry (see Freedman and Jin, 2008; Berger and Gleisner, 2009; Iyer et al., 2014; Everett, 2010; Hildebrand et al., 2013; Lin et al., 2013). Furthermore, on Prosper the interest rate of a loan used to be conducted by a Dutch auction process until December 19th 2010, when this procedure was replaced by a posted price mechanism. This change is analyzed by Lin and Wei (2013) and Meyer (2013). Both studies indicate a higher funding probability associated with a deteriorated loan quality after this change.

Another strand of literature empirically analyzes the behavior of P2P market actors. There is research on the capability of hard facts to serve as solvency indicators (see Böhme and Pötzsch, 2010; Lin et al., 2013; Weiss et al., 2010). Soft factors can, however, still help to mitigate asymmetric information. By now, there have been several studies that examine the influence of borrowers' soft information with respect to funding success, interest rates and loan defaults. There is evidence regarding a positive effect on the loan conditions when providing a picture in the application (see e.g. Böhme and Pötzsch, 2010; Iyer et al., 2014). However, when examining the content of pictures with respect to skin color, charisma, age and gender, some studies find evidence in favor of taste-based discrimination when it comes to funding success and loan conditions (see Duarte et al., 2012; Herzenstein et al., 2011; Pope and Sydnor, 2011; Ravina, 2012). Taste-based discrimination occurs if people are not treated equally due to prejudices with respect to their appearance (see Becker, 1971; Fershtman and Gneezy, 2001). Duarte et al. (2012) show that borrowers who have a trustworthy appearance face a better chance to have their loan granted. Furthermore, attractive people benefit from better loan conditions and a higher funding probability while showing similar repayment rates (see Ravina, 2012). Several studies show that older people are confronted with a lower funding probability and worse conditions, in the case that their loan is granted (see Böhme and Pötzsch, 2010; Fabender, 2011; Pope and Sydnor, 2011). Concerning young borrowers, findings differ. Böhme and Pötzsch (2010) observe poor loan conditions, but Pope and Sydnor (2011) reveal a higher funding probability for this group. Barasinska and Schäfer (2014) find no gender effect on the funding probability, while Fabender (2011) and Lin et al. (2013) show evidence for taste-based discrimination against men. Gao and Lin (2015) find that the readability of a loan application, a positive sentiment and several deception cues are related to the default probability on Prosper. Iyer et al. (2014) also analyze the description texts on Prosper for a similar short period (February 2007 to October 2008) and reveal the predictive power of soft factors such as the self-reported loan purpose or text characteristics on the default probability. Sonenshein et al. (2011) examine the influence of social accounts, such as whether a text provides an explanation, an acknowledgment or a denial, on a successful funding, based on a Prosper data set consisting of 512 observations posted in June 2006.

3. Hypotheses development

In the following, we utilize the insights of previous literature to derive testable hypotheses regarding the soft factors related to the description text which are considered in this study. Furthermore, our analysis focuses on the two leading P2P platforms in Germany, namely Smava and Auxmoney. By comparing both platforms, insights into the platform structure and the loan granting mechanism can be derived. These factors have not been analyzed regarding their effect on P2P lending on a comprehensive data basis until now.

Orthography. Psychological surveys show that misspellings are often seen as indication of poor cognitive skills of an author (see e.g. Kreiner et al., 2002). More specifically, Figueredo and Varnhagen (2005) find that a text is regarded as being particularly inferior if misspellings are non-homophone, implying that they can be detected by a spell checker. Furthermore, bad orthography makes a text difficult to assess (Pynte et al., 2004) and thus can lower the probability for successful funding. This view is supported by Gao and Lin (2015), who state that the readability of a description text on Prosper is positively appreciated by the lenders. However, in electronic communication an informal writing style is relatively common. Park et al. (2010) explore the influence of misspellings in electronic meetings and find that neither the participants' satisfaction nor their productivity suffers from bad orthographical skills. P2P actors should be quite familiar with the customs of internet communication which could mitigate a possibly negative effect of spelling errors on the funding probability. Summarizing, if the description text of a loan application contains misspellings, this could be interpreted as an indication of a less solvent borrower or the applicant may even appear to be untrustworthy. Therefore, we expect a negative relation to the funding success.

Hypothesis 1a (*orthography*). Loan applications with a high fraction of spelling mistakes within the description text are less likely to be funded.

Even if we can expect—in the case that Hypothesis 1a is valid—that successfully funded loans exhibit a more sophisticated spelling, orthography can still serve as a proxy for education. It is well-known that there is a negative relation between a borrower's level of education and his or her default probability (Bhatt and Tang, 2002). Thus we conjecture that the default probability positively depends on the share of spelling errors.

Hypothesis 1b (*orthography*). Granted loans with a low fraction of spelling mistakes within the description text are less likely to default.

Description length. Closely related to the matter of orthography is the question regarding the length of the description text. First, the longer the text is, the more spelling mistakes could be included. This is why we consider the relative fraction of spelling mistakes in the orthography hypotheses. Second, the description text may contribute to a reduction of information asymmetries (see Michels, 2012) as the loan applicants can describe their individual situation in detail. This makes it easier for lenders to assess an applicant's loan request. Therefore, writing a longer text may serve as a signal of creditworthiness to the lenders and support a higher probability of successful funding.

However, we also do expect that loans with a very long description text are supported less willingly by the investors for two reasons. First, if the description length is far longer than those of other loans, the investors who often only invest small amounts of money into the loans will tend not to be willing to spend the time to read the text and as a consequence tend not to fund such a loan. Second, long-winded description texts can indicate an intricate personality of the applicant. Transferring this characteristic to the context of managing personal finance, the lenders may conclude that the applicant tends not to be concise in this area either. This, in turn, affects the repayment behavior and thus the creditworthiness.[1]

Hypothesis 2a (*description length*). The length of the description text in a loan application is positively related to funding success up to a certain amount of words.

Loan applicants tend to provide information in the description text if these support the funding probability. For this reason, a longer description text can be a signal of creditworthiness and can be expected to result in a lower default probability. However, analogously to the reasoning regarding Hypothesis 2a, above a certain value of the length, there may be reverse effects.

Hypothesis 2b (*description length*). The length of the description text in a loan application is negatively related to the probability of default up to a certain amount of words.

Social and emotional motives. Van Wingerden and Ryan (2011) show in a survey among 124 crowd investors in 2011 that a considerable number of them follow also intrinsic motivations instead of only seeking a financial return. P2P lending is a more emotional matter than e.g. investing money in a bank account, as one directly can observe who is the receiver of the investment. While Gonzalez and Loureiro (2014) observe emotional biases regarding the influence of the loan applicant's picture, such an effect can also be expected for emotionally appealing description texts, be they positive or negative. The lenders may be more willing to invest the money in the case of negative emotions because of the inclination to help (see Renneboog et al. (2008) for a general treatment of such investor behavior, Allison et al. (2013) for the special case of P2P microlending and Böhme and Pötzsch (2011) for weak evidence in P2P lending). In case of positive emotions, potential lenders can reveal the tendency of wanting to participate in the positive issues related to the loan, as Bruton et al. (2015) show for crowdfunding in general, or simply may be subject to the overconfidence bias (Hirshleifer, 2001) due to the positive emotional statements in the text.[2] The description text allows a borrower to explain the loan purpose in detail and to address social motives which can be directly assessed by possible investors. We assess the emotional character of a description text by the emotional keywords used.

Hypothesis 3a (*Social and emotional motives*). Keywords with a social or emotional connotation in the description texts are positively related to the funding success.

As the above mentioned reasons for granting a loan are rather irrational, it can be expected that the risk of loans for which Hypothesis 3a applies is higher as for comparable loans with a similar interest rate. This higher risk can be expected to yield a higher probability of default.

[1] Furthermore, for companies there is corresponding evidence by Loughran and McDonald (2014), who analyze 10-K documents and argue that negative information is often hidden within long texts.

[2] See Dowling and Lucey (2005) for the general role of positive emotions in financial decision making.

Hypothesis 3b (*Social and emotional motives*). Keywords with a social or emotional connotation in the description text are positively related to the probability of default.

4. Data and methodology

4.1. Data

Our unique data set combines data from four sources. Individual loan data was derived from loan applications published online by the P2P lending platforms Auxmoney (www.auxmoney.com) between March 2008 and September 2013 and Smava (www.smava.de) between February 2007 and September 2013. A total of 92 observations from Auxmoney and 24 observations from Smava were excluded from further analysis due to obviously erroneous data. The resulting data sets comprise 76,945 loan applications from Auxmoney and 10,423 from Smava. Neither platform provides information regarding the repayment status of an individual loan. However, there is a vibrant online platform called Wiseclerk (www.wiseclerk.com) that provides tools for P2P investors which allow them to analyze the performance of their P2P loan portfolios. Therefore, investors report their P2P loan portfolio composition and the corresponding loan defaults to Wiseclerk. In the following, we use this data source to extract the information on whether a loan is subject to default. Consequently, we classify closed granted loans without default information as non-defaulted. Note that theoretically, there is a possible bias because P2P investors are not required to report defaulted loans. However, as P2P loans are usually financed by many lenders, it is very likely that defaults are indeed reported on Wiseclerk by at least one of these. For example, if the probability that a lender, who has experienced a default on a loan, reports such an event is assumed to be 0.5, which is a conservatively low value for internet-affine lenders, who also tend to be be intrinsically motivated,[3] then given a number of ten lenders per loan, the probability for an error is only 0.098%. As we will argue below in Section 4.4, we can assume this bias to be so small that it is negligible. Furthermore, the possibility that a loan erroneously is reported as defaulted can be excluded. Lenders will rationally have no incentive for such a costly behavior and in the unlikely case of such an event, creditors will have a high incentive to clear out false statements. Additionally, we receive data on the German stock index (DAX) and the yield curves derived from German government bonds from Thomson Reuters Datastream. The aim of this study is to research both of the following: The probability of the loan being granted by investors is examined via the indicator variable *FGL*, which documents a successful funding. The default probability of granted P2P loans is analyzed utilizing the variable *DEF*, which indicates a loan default. To analyze the latter, the data set is reduced considerably because only granted loans that were closed before December 15th 2013 can be considered. The resulting data sets regarding closed granted loans (CGL) comprise 3298 (Auxmoney) and 2216 (Smava) observations.

4.2. Research design

To carve out the role of soft factors related to the description texts in the lending decision as well as in the default behavior, we utilize data from two P2P platforms that are very distinct with respect to the extent of requiring hard facts and also to the extent of influencing the lending decision. The German P2P platforms Smava and Auxmoney have implemented different designs concerning the procedure of loan applications. Smava—in contrast to Auxmoney—verifies loan applications with respect to several criteria to ensure that listed applications fulfill a minimum level of creditworthiness. As a consequence, the importance of soft information for creditors can possibly be less pronounced there. This could be anticipated by the applicants, who themselves provide only a minimum of information regarding soft factors (see Lucas, 1972). Furthermore, Smava provides a bidding assistant which supports investors by making automated bids on listed applications.[4] The bidding assistant is solely based on hard facts such as the Schufa score or the loan duration and neglects soft information. In addition, some hard facts which have always been mandatory for Smava since the launch of the platform, have not been obligatory on Auxmoney until February 2013. In the case of missing hard facts, investors may rely more strongly on soft information.

With the difference between the probability of successful funding and the default probability being that the first is dependent on the perception of the P2P investors, while the latter is not, we can argue that if there is a relation of soft factors to the default probability at all, there is no reason for it to be different for both platforms. However, there surely is a difference between the platforms with respect to the samples that can be investigated with respect to the likelihood of defaulting. Thus, significances of coefficients could be different due to this effect.

4.3. Explanatory variables

Loan applications usually include a short description text regarding the loan's purpose and/or the personal situation of the applicant. We analyze this description in order to derive several variables, which we use to examine the relations of soft factors derived from the description text in P2P lending. All variables, including other control variables, are defined in Table 1 and those relevant for testing our hypotheses are shortly described in the following.

The orthographic quality of a description text—referring to Hypotheses 1a and 1b—is measured by the variable *SpellError* - which represents the percentage of misspelled words. The variable is derived with a spelling check that is based on the open-source library GNU Aspell but accounts for common terms regarding P2P lending. For this matter, we have treated errors classified by the GNU Aspell which have appeared more than ten times in the analysis manually, regarding the correctness of the spelling. Thereby, we have identified some correct terms that are not included in the GNU Aspell, like abbreviations or names. Detailed information on the spelling check is presented in Table A.11 in the Appendix A.

The length of the description text is proxied by the variable *#Words* which comprises the number of words included. To capture the suggested inversely u-shaped relation of this factor, which is suggested by Hypothesis 2a and 2b, we additionally include this variable in squared form in the regressions. Furthermore, we generate a group of keyword indicator variables (*KeyWord*). To this end, the description text is searched for German keywords regarding the following categories: The indicator variable *Fam* indicates the usage of words associated with family, e.g. wife, children. Other categories are negative aspects (*Neg*, e.g. inhumation), positive aspects (*Pos*, e.g. dream) and separation (*Separ*). We consider this group of keywords as emotional and socially connoted.

Additionally, we consider several already documented effects on P2P lending platforms and address peculiarities of Smava and Auxmoney by implementing several control variables. Therefore, we use a second group of keywords as further controls, namely those describing the loan purpose without potentially raising emotions. These are debt restructuring (*Restruc*), education (*Edu*), leisure

[3] See Van Wingerden and Ryan (2011) for an overview on intrinsic motivation in crowdsourcing.

[4] Note that Auxmoney did not have a bidding assistant within the observation period.

Table 1
Description of variables.

Variable		Description
CEG	CEG signal	Solvency information provided by Creditreform GmbH. 1, if the CEG is the following category: *Green*: green, no negative information, *Yellow*: yellow, twice the amount of the mean probability to default of consumer loans in Germany *Red*: red, negative information, *NA*: no information available. 0, otherwise. *Source:* Auxmoney. [a]
DAX	German stock index DAX	Proxy for economic climate, measured as continuous returns over quarterly averages of the performance index DAX. *Source:* Datastream
DEF	Default indicator	1, if loan is defaulted, e.g. loan is subject to summary proceedings or collection handling. 0, otherwise. *Source:* Wiseclerk
Employment	Employment relationship	1, if loan applicant is an employee (*Employee*), self-employed person (*Selfemp*), civil servant (*CivServant*), pensioner (*Pension*), or does pursue other form of permanent income realization (*Other*). 0, otherwise. *Source:* Smava [s]
FGL	Fully granted loan	1, if enough funds are provided by private investors that loan could be 100% granted. 0, otherwise. Note that in rare cases enough funds were provided by investors but the loan was not retrieved by the loan applicant. *Sources:* Auxmoney, Smava
FedState	Federal state of loan applicant	1, if federal state of loan applicant is Baden-Württemberg (*BW*), Bayern (*BY*), Berlin (*BE*), Brandenburg (*BB*), Bremen (*HB*), Hamburg (*HH*), Hessen (*HE*), Mecklenburg-Vorpommern (*MV*), Niedersachsen (*NI*), Nordrhein-Westfalen (*NW*), Rheinland-Pfalz (*RP*), Saarland (*SL*), Sachsen (*SN*), Sachsen-Anhalt (*ST*), Schleswig–Holstein (*SH*) and Thüringen (*TH*). 0, otherwise. *Source:* Smava. [s]
FundTime	Funding time	Days needed to fully fund the loan. Estimated as period between the first and the last bid regarding 100% funded loans and categorized: *Short* (0 days), *Mid* (Auxmoney ⩽10 days, Smava ⩽5 days) and *Long.* 1, if observation falls in the respective category. 0, otherwise. As no exact application date is provided by both platforms, we use the date of the first bid as a proxy. If no bid is available, the start date is derived based on the incremental identification number of each loan application. Derived from Auxmoney, Smava
I	Interest rate	Loan's nominal interest rate. *Sources:* Auxmoney, Smava
I_{rf}	Risk free interest rate	Yield curve derived from German government bonds with maturities of three (for Auxmoney) and five (for Smava) years. *Source:* Datastream
KDF	KDF indicator	Share of debt service from personal net income, categorized: *1* (0%–20%), *2* (20%–40%), *3* (40%–60%) and *4* (60%–80%).1, if observation falls in the respective category. 0, otherwise. Note that Smava does not allow any share larger than 67%. *Source:* Smava. [s]
KeyWord	Keywords	Keywords associated with the following categories are mentioned in the description text: Family (*Fam*), negative (*Neg*), positive (*Pos*), separation (*Separ*), Leisure (*Leisure*), Business (*Business*), debt restructuring (*Restruc*) and education (*Edu*). We indicate the first four keywords as being related to social and emotional motives. 1, if observation falls in the respective category. 0, otherwise. Multiple references possible. Derived from Auxmoney, Smava
#Lender	Number lenders	Number of lenders derived from biddings on granted loans. *Sources:* Auxmoney, Smava
Male	Gender of loan applicant	1, if loan applicant is male, 0, otherwise. *Source:* Smava. [s]
Mat_Short	Short time to maturity	1, if loan has a short time to maturity, 0 otherwise. A short time to maturity represents 24 month or less for Auxmoney and 36 month or less for Smava. *Sources:* Auxmoney, Smava
Picture	Project picture	1, if a picture regarding funded project is available, 0, otherwise. *Sources:* Auxmoney, Smava
ResRate	Residual interest rate	Loan's nominal interest rate minus risk premium derived from Schufa score and time to maturity. *Sources:* Smava (risk premia, loan's nominal interest rate), Auxmoney
Schufa	Schufa score	Solvency indicator. Category A (excellent solvency) to M (poor) or not provided (*NA*). 1, if observation falls in the respective category. 0, otherwise. Note, that Schufa score is not mandatory for Auxmoney applications. *Sources:* Auxmoney, Smava
SpellError	Spelling error	Share of words in loan description that is misspelled. The spell check is based in the open-source library GNU Aspell, which has been manually extended. More details can be found in Table A.11 in the Appendix A. Derived from Auxmoney, Smava
TurnYear	Turn-of-the-year indicator	1, if loan application took place in December or January. 0, otherwise. *Sources:* Auxmoney, Smava
Volume	Loan volume	The nominal volume of the loan. *Sources:* Auxmoney, Smava
#Words	Number of words	Number of words used in the description text. *Sources:* Auxmoney, Smava

Note: [a] indicates that variable is soley available for Auxmoney or Smava [s].

activities (*Leisure*) and business (*Business*). All keywords and the associated categories are displayed in Table A.12 in the Appendix A.

We capture turn-of-the-year effects with the control variable *TurnYear*, which indicates whether a loan application was started in December or January. Approximately 54% of the German workforce receive a special bonus payment at Christmas, which equals between 20% and 100% of their monthly income (see WSI, 2013). Some people spend this money on Christmas presents, but 41% save at least a fraction of it (see GfK, 2010). As lenders use P2P platforms as an investment opportunity, this capital may increase the supply in German P2P markets in December and January and thus may improve the funding probability at the turn of the year.

Additionally, we add loan and borrower specific controls: the loan volume *Volume* (in logarithmic representation), an indicator for short maturity (*Mat_Short*), the solvency information (*Schufa*, *CEG*, *KDF*) and the interest rate *I* (in logarithmic representation). Note that on both platforms, the interest rate is suggested by the applicant and therefore influenced by his/her personal solvency sentiment. Previous studies proved that a picture (e.g. Böhme and Pötzsch, 2010; Iyer et al., 2014) or gender information (e.g. Fabender, 2011; Lin et al., 2013) have an influence on the likelihood of the loan being granted or the probability of default. Therefore, we include suitable variables (*Picture*, *Male*). Furthermore, we include quarterly returns of the German stock index DAX (*DAX*) to account for macroeconomic effects. In the case of Smava, we additionally control for the federal state (*FedState*) in which the loan applicant's residence is located, the applicant's age (*Age*) and his employment situation (*Employment*).

4.4. Descriptive analysis

The descriptive measures of the metric variables and the relative frequencies of categorical variables for the complete Auxmoney and Smava data sets and the CGL subsamples are shown in Tables 2 and 3.

The share of granted loans is much higher on Smava (89.2%) than on Auxmoney (17.6%). The historical average default rates are within the same range for both platforms, amounting to 12% on Auxmoney and to 13.8% on Smava.[5] Continuing the discussion from above regarding the likelihood of falsely reported non-defaults on Wiseclerk, we can state the following. When looking

[5] Interestingly, the default rates on both platforms decline over time, which we interpret as an indication that the market participants become more experienced with time. Additionally, they also show similar values if we consider the lifetime of the platform, i.e. Smava and Auxmoney have comparable default rates in their second, third year and so on.

at the interrelation between the number of lenders (*#Lender*) reported in Table 2 and the variable *DEF* in a contingency table (not reported here), there are no peculiar deviations in the default rates of those loans which have been granted by only a few lenders. We interpret this finding as a clear indication that the reporting of defaulted loans to Wiseclerk appears to work even if a loan is granted by only a few lenders. We conclude that for a high number of lenders, it is very unlikely that none of them reports a defaulted loan. In case of few lenders, a higher amount of money is at risk, so that it is also very likely that a default is reported.[6]

Note that for each platform the fraction of loans in the CGL sample to the total of granted loans is roughly one fourth. This is a consequence of the fact that in order to avoid a censored-data bias, we have to discard many of the granted loan observations. More precisely, we skip the loan observations with a maturity exceeding the observation period as these are still open and thus the default status cannot be determined without doubt. In particular, this implies that observations from the first part of our observation period are over-represented in the CGL samples. Note that we still use *all* of the corresponding granted loan observations that are not affected by the censored-data problem. As we do not have indications that the mechanism behind the defaulting has changed over time and as we still have enough loans with a longer maturity in the CGL samples (defaulted and non-defaulted ones), we regard this analysis to be relevant for explaining the defaults on both platforms.

The higher ability of the lenders on Auxmoney to identify risky loans cannot be based heavily on traditional solvency measures, like the Schufa score, as a large share of all closed, granted loans on Auxmoney provide no such score (46.3% no Schufa score and 55% no CEG score), whereas for Smava, a Schufa score of at least H or better is mandatory. Therefore, soft information seems to play a role for investors, when deciding whether to grant a specific loan.

On both platforms, the average nominal interest rate is slightly higher for closed granted loans (13.12% on Auxmoney, 10.37% on Smava) than for all loan applications (11.60% on Auxmoney, 8.78% on Smava). For the sample period, we can observe that closed granted loans on Auxmoney outperform Smava regarding risk and return. The higher average interest rate for granted loans can either be a suitable compensation for the higher default risk or an over-compensation in order to make the loan attractive for investors.

Furthermore, we find that the volume of loans on Auxmoney (5,030.07 EUR on average) is smaller compared to Smava (8,995.32 EUR on average) and the same holds for the maturity (36.72 months on Auxmoney, 53.34 months on Smava). Regarding the hypotheses-related variables *SpellError* and *#Words*, we observe differences between both platforms. Description texts are on average longer on Auxmoney (55.94 vs. 41.43 words) and have more spelling errors (7.83% vs. 2.71%) compared to Smava. Contrary to Auxmoney, the orthographical quality is lower in the subsample of closed granted loans compared to the overall sample on Smava. This is a first hint that avoiding spelling errors appears not to be as important on Smava as on Auxmoney.

Tables A.13 and A.14 in the Appendix A show the pairwise Bravais-Pearson correlations among the explanatory variables for the two data sets. All significant correlations show absolute values below 0.8 indicating that no multicollinearity issues arise (see Kennedy, 2008).

4.5. Methodology

The dependent variables *FGL* and *DEF* of our analysis are both binary. Hence, logit or probit regressions appear suitable (e.g. Barasinska and Schäfer, 2014), which only result in unbiased estimators if no endogeneity concerns exist regarding the explanatory variables. In our setting, the interest rate the borrowers are being charged can be subject to endogeneity because these rates are posted by the borrowers themselves while considering their own solvency. We account for this problem by applying simultaneous IV probit regressions (see Rivers and Vuong, 1988) estimated via maximum likelihood with the risk free interest rate as instrumental variable. A suitable instrument should explain a part of the variation of the dependent variable whereas it should not be directly related to the explained variable in the structural equation (see e.g. Cameron and Trivedi, 2010). This is economically sound for the risk free interest rate (I_{rf}), which is defined in Table 1. Consistently with the average maturities on both platforms, we use the yield curve derived from government bonds with a maturity of three years on Auxmoney and a maturity of five years on Smava as proxies for I_{rf}. The regression model shows the following structure regarding the latent variable y^*_{1i} that is linked to the binary explained variable via the probit specification.

$$y^*_{1i} = \mathbf{m}'_i\delta + \alpha y_{2i} + u_i \quad (1)$$

$$y_{2i} = \mathbf{m}'_i\gamma + \pi z_i + e_i \quad (2)$$

The vector $\mathbf{m}'_i$ represents the explanatory variables and z_i the instrumental variable. The terms u_i and e_i are error terms of the structural and reduced form equation, respectively. Conducted Wald tests confirm on the 1% significance level that the IV probit approach is suitable to address endogeneity in our setting.

5. Results

In this section, we first analyze the factors influencing the funding probability and second those regarding the default probability. Additionally, we perform some robustness checks and discuss the differences between both platforms.

5.1. Funding probability

5.1.1. Auxmoney

The first four columns in Table 4 show the results for the model specifications with *FGL* as a dependent variable for Auxmoney. Specifications AF.I to AF.III incorporate the hypotheses-related variables *SpellError* (Hypotheses 1a), *#Words* (Hypotheses 2a) and the keyword indicator variables *KeyWord_Fam*, *KeyWord_Neg*, *KeyWord_Pos*, *KeyWord_Separ* (Hypothesis 3a) separately, each together with the control variables. Specification AF.IV represents the main model including all variables simultaneously. The last column shows the average marginal effects for Specification AF.IV which are used to interpret the effects regarding their magnitude.

As expected, we find a negative and highly significant relationship between the percentage of misspelled words and the funding probability in all relevant specifications. The average marginal effect of *SpellError* shows a value of −0.0021, indicating that a spelling error increase of 1% lowers the funding probability by 0.21% (Note, that *SpellError* is measured in percentage points). At first sight, the impact of this effect is not large, however, the distribution of *SpellError* also has to be taken into account. Thus, a ceteris paribus increase by one standard deviation of *SpellError* corresponds to a decrease of the default probability amounting to 2.9%, which is a considerably large magnitude if compared to the other factors. Thus we can confirm *Hypothesis 1a (orthography)* for Auxmoney.

Regarding the length of the description text, the coefficients of *#Words* in AF.II and AF.IV are positive and highly significant,

[6] Additionally, to dispel remaining doubts we perform some additional checks below by utilizing only those closed granted loans with a high number of lenders.

Table 2
Descriptive statistics of metric variables.

	DATA	N	MIN	Q25%	MEDIAN	MEAN	Q75%	MAX	SD
Variables concerning both platforms									
Volume	AUX	76,945	1000	1500	3000	5030.07	6700	30,350	5054.36
	AUX, CGL	3298	1000	1500	2000	3243.01	4000	20,000	3141.25
	SMA	10,423	500	3250	6250	8995.32	12,000	50,000	7967.97
	SMA, CGL	2216	500	2500	3750	5301.78	6500	50,000	4772.00
I	AUX	76,945	0.00	0.10	0.13	0.12	0.14	0.18	0.03
	AUX, CGL	3298	0.05	0.12	0.13	0.13	0.15	0.17	0.02
	SMA	10,423	0.01	0.06	0.08	0.09	0.11	0.18	0.03
	SMA, CGL	2216	2.50	7.40	9.80	10.37	13.35	18	3.56
SpellError	AUX	76,617	0	0	2.99	7.83	9.09	100	13.87
	AUX, CGL	3298	0	0	2.11	3.51	4.26	100	5.61
	SMA	10,367	0	0	0	2.71	2.86	100	7.77
	SMA, CGL	2208	0	0	1.08	3.27	3.70	100	7.61
DAX	AUX	76,945	−0.23	0.02	0.04	0.03	0.06	0.18	0.07
	AUX, CGL	3298	−0.23	0.02	0.04	0.05	0.10	0.18	0.08
	SMA	10,423	−0.23	−0.01	0.03	0.02	0.06	0.18	0.09
	SMA, CGL	2216	−0.23	−0.08	0.01	−0.01	0.04	0.18	0.11
#Lender	AUX, CGL	3298	1	10	15	20.84	26	123	16.39
	SMA, CGL	2216	1	6	9	12.34	16	115	10.96
#Words	AUX	76,945	0	13	34	55.94	70	8441	83.52
	AUX, CGL	3298	1	44	81	109.40	138	2129	108.35
	SMA	10,423	0	19	26	41.43	50	531	43.62
	SMA, CGL	2216	0	21	38	53.73	71.50	531	52.31
ResRate	AUX	19,035	−0.11	0.02	0.05	0.05	0.09	0.15	0.04
	AUX, CGL	1771	−0.04	0.04	0.07	0.07	0.10	0.15	0.04
	SMA	10,423	−0.04	0.05	0.06	0.06	0.07	0.16	0.02
	SMA, CGL	2216	−0.01	0.06	0.07	0.07	0.09	0.16	0.02
Variables concerning only one platform									
Age	SMA	10,423	20	36	45	46.33	54	95	13.33
	SMA, CGL	2216	23	36	46	47.16	55	93	14.21

Notes: AUX and SMA represent the Auxmoney and Smava data samples. CGL indicates the subsamples of closed granted loans. QXY% refers to the XY% quantile. The variables are defined in Table 1. *Data sources:* Auxmoney, Smava, Datastream.

whereas the coefficients of the squared variable are negative. This constitutes an inversely u-shaped pattern which is consistent with our expectation. According to the average marginal effect, the funding probability increases by 5.2% if the description text is increased ceteris paribus by one standard deviation. However, the funding probability decreases for very long description texts as the coefficients of the squared variable are both significantly negative. This result confirms Hypothesis 2a *(description length)*.

Apart from the orthographical accuracy and the length of the description text, the content can to some extent predict the funding probability. In specification AF.I, almost all coefficients of the keyword variables related to emotional motives are significantly positive. However, if the other factors are taken into account, in AF.IV only *KeyWord_Pos* remains significant. Thus we find that loan applicants using positive keywords have a ceteris paribus 3.3% higher chance of receiving a loan on Auxmoney. Concluding, we have limited evidence to support *Hypothesis 3a (social and emotional motives)*.

Moreover, keywords addressing a business purpose or debt restructuring are significantly related to a higher funding probability. Business activities are supposed to create positive cash flows in the future that can be used for servicing debt. Therefore, investors appear to invest more willingly in such loan applications. A weakly significant negative coefficient is attributed to loans related to leisure activities.

5.1.2. Smava

Table 5 shows the regression results for Smava with *FGL* as dependent variable. Again, the first three regressions (SF.I to SF. III) include all control variables and the hypotheses-related variables separately for each hypothesis. SF.IV is the main specification including all variables simultaneously.

The coefficient of *SpellError* is insignificant in all specifications. This phenomenon may be due to the lower variation of *SpellError* in the Smava sample and to the generally lower level of misspellings (2.71% on average). We derive similar results concerning the text length. Both coefficients for *#Words* are negative, close to zero and not significant in SF.II and SF.IV. Hence, we can neither approve nor reject *Hypothesis 1a (orthography)* and *Hypothesis 2a (description length)*. Thus, spelling errors and text length appear not to be predictive factors for the funding probability on the platform Smava.

Moreover, two of the keyword indicators used in the description text are insignificant, two are significant. The coefficient of *KeyWord_Fam* is negative and a loan application-related to family has a ceteris paribus 3.28% lower chance to be financed. Investors may associate a family with payment obligations, which could affect repayment behavior. The relationship between *KeyWord_Neg* and the funding probability is significantly positive, which indicates some evidence in favor of Hypothesis 3a. Altogether, the opposite signs of the coefficients of *KeyWord_Fam* and *KeyWord_Neg* provide somewhat unclear evidence. Thus, we can neither reject nor confirm *Hypothesis 3a (social and emotional motives)*.

5.2. Probability of default

Tables 6 and 7 show the results of the default probability analysis.

5.2.1. Auxmoney

The specifications AD.I to AD.IV in Table 6 are similar to the model specifications concerning the funding probability and with *DEF* as dependent variable for Auxmoney, but with additional dummy variables related to the time needed to fully fund the loan

Table 3
Relative frequency distributions of categorical variables in percentage values.

Variables concerning both platforms										
FGL		1(Yes)	0							
	AUX	17.6	82.4							
	SMA	89.2	10.8							
Picture		1(Yes)	0							
	AUX	49.0	51.0							
	AUX, CGL	69.9	30.1							
	SMA	11.7	88.3							
	SMA, CGL	19.8	80.2							
TurnYear		1(Yes)	0							
	AUX	18.0	82.0							
	AUX, CGL	16.0	84.0							
	SMA	16,3	83.7							
	SMA, CGL	12.3	87.7							
Mat_Short		1(Yes)	0							
	AUX	32.7	67.3							
	AUX, CGL	73.1	26.9							
	SMA	31.7	68.3							
	SMA, CGL	86.2	13.8							
Schufa		*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
	AUX	0.8	1.1	0.9	1.1	1.5	2.1	3.2	2.5	1.6
	AUX, CGL	2.5	3.3	2.5	3.3	4.4	5.9	7.9	6.2	3.7
	SMA	20.9	18.3	9.4	9.3	9.9	10.6	13.7	7.9	
	SMA, CGL	18.0	17.1	9.1	9.6	10.4	12.1	14.0	9.6	
Schufa		*K*	*L*	*M*	*NA*					
(continued)	AUX	1.0	1.5	7.4	75.3					
	AUX, CGL	2.1	3.9	7.8	46.3					
	SMA									
	SMA, CGL									
Absolute frequencies as multiple references are possible										
KeyWord		*Restruc*	*Edu*	*Neg*	*Business*	*Pos*	*Fam*	*Separ*	*Leisure*	
	AUX	13,137	5685	4831	6267	20,620	11,893	2068	2914	
	AUX, CGL	980	402	354	497	1440	656	135	141	
	SMA	2567	430	348	1022	3385	894	140	286	
	SMA, CGL	624	142	135	244	816	275	55	103	
Variables concerning only the CGL subsamples										
FundTime		*Short*	*Mid*	*Long*						
	AUX, CGL	7.1	48.3	44.6						
	SMA, CGL	50.8	29.9	19.3						
DEF		1(Yes)	0							
	AUX, CGL	12.0	88.0							
	SMA, CGL	13.8	86.2							
Variables concerning only one platform										
CEG		*Green*	*Yellow*	*Red*	*NA*					
	AUX	11.6	10.0	1.3	77.0					
	AUX, CGL	25.6	18.3	1.1	55.0					
Male		1 (Yes)	0							
	SMA	73.1	26.9							
	SMA, CGL	73.0	27.0							
KDF		1	2	3	4					
	SMA	12.6	24.8	38.7	23.9					
	SMA, CGL	20.0	26.7	29.2	24.2					
Employment		*Employee*	*CivServant*	*Selfemp*	*Pension*	*Other*				
	SMA	51.7	4.0	34.9	9.1	0.2				
	SMA, CGL	57.5	4.8	26.5	10.7	0.4				
FedState		*BY*	*BW*	*BE*	*BB*	*HB*	*HH*	*HE*	*MV*	*NI*
	SMA	16.5	12.9	7.5	3.5	0.8	3.3	8.0	1.6	8.7
	SMA, CGL	16.0	12.5	7.9	2.8	0.7	3.2	9.2	1.7	9.1
FedState		*NW*	*RP*	*SL*	*SN*	*ST*	*SH*	*TH*		
(continued)	SMA	19.6	4.3	0.9	4.5	2.1	3.5	2.3		
	SMA, CGL	19.1	4.1	0.9	4.6	2.0	3.8	2.3		

Notes: AUX and SMA represent the Auxmoney ($N = 76{,}945$) and Smava ($N = 10{,}423$) data samples. CGL indicates subsamples of closed granted loans, with $N = 3298$ (AUX) and $N = 2216$ (SMA). *KeyWord* is shown in absolute frequencies. The variables are defined in Table 1. *Data sources:* Auxmoney, Smava, Wiseclerk.

as further control variables to cover the aspect of rational herding. The last column shows the average marginal effects for specification AD.IV.

The coefficients of *SpellError #Words* and the squared value of *#Words* are insignificant in all relevant model specifications, which may be attributable to the fact that the loan applications in the CGL

Table 4
Regression results concerning the funding probability on Auxmoney.

	Funding probability (*FGL*)				
	AF.I	AF.II	AF.III	AF.IV	
				Coeff.	AME
Variables related to hypotheses					
SpellError	−0.0115***			−0.00757***	−0.002082***
	(−15.6)			(−13.0)	
#Words		0.00280***		0.00226***	0.0006212***
		(20.4)		(17.1)	
$(\#Words)^2$		−0.00000139***		−0.00000113***	−0.0000003100***
		(−11.2)		(−10.2)	
KeyWord_ Pos			0.226***	0.120***	0.03288***
			(20.0)	(11.9)	
KeyWord_Neg			0.112***	0.00757	0.002082
			(6.00)	(0.428)	
KeyWord_Fam			0.102***	0.0192	0.005269
			(7.67)	(1.60)	
KeyWord_Separ			0.0741***	−0.0170	−0.004675
			(2.64)	(−0.651)	
Soft controls					
KeyWord_Restruc	0.138***	0.0736***	0.134***	0.0582***	0.01601***
	(9.39)	(5.75)	(9.05)	(4.69)	
KeyWord_Edu	0.143***	0.0383**	0.136***	0.0251	0.006911
	(7.65)	(2.31)	(7.35)	(1.55)	
KeyWord_ Business	0.174***	0.0398**	0.167***	0.0428***	0.01178***
	(10.1)	(2.45)	(9.85)	(2.67)	
KeyWord_Leisure	0.0480*	−0.0447**	0.00957	−0.0398*	−0.01096*
	(1.95)	(−1.98)	(0.397)	(−1.80)	
Picture	0.452***	0.367***	0.429***	0.352***	0.09672***
	(25.9)	(23.3)	(24.3)	(22.4)	
Hard controls					
ln(I)	−0.946***	−1.46***	−1.09***	−1.53***	−0.4203***
	(−6.45)	(−14.6)	(−7.90)	(−16.1)	
ln(Volume)	−0.268***	−0.237***	−0.250***	−0.227***	−0.06249***
	(−17.8)	(−16.6)	(−16.7)	(−16.0)	
Mat_Short	0.239***	0.0960***	0.215***	0.0695**	0.01913**
	(6.07)	(3.06)	(5.50)	(2.31)	
Schufa	Yes	Yes	Yes	Yes	
CEG	Yes	Yes	Yes	Yes	
DAX	0.575***	0.536***	0.586***	0.537***	0.1476***
	(7.68)	(7.92)	(8.01)	(8.04)	
TurnYear	−0.0663***	−0.0735***	−0.0660***	−0.0717***	−0.01973***
	(−4.92)	(−6.25)	(−5.06)	(−6.20)	
CONST	−0.986***	−2.23***	−1.54***	−2.35***	
	(−2.87)	(−9.11)	(−4.83)	(−9.85)	
AIC	133,361.20	132,997.21	133,759.88	131,979.72	
N	76,617	76,945	76,945	76,617	

Notes: Model specifications AF.I to AF.IV are simultaneous IV probit regressions for the funding probability. The column AME AF.IV shows the average marginal effect of the variables on the funding probability with respect to specification AF.IV. Z-statistics are shown in parenthesis. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. The AIC of a model only containing hard controls is 137,108.68. Reference categories: For *Mat* category *Mid*, for *Schufa* category *M*, for *CEG* category *Red*. The variables are defined in Table 1.

subsample show a lower percentage of misspelled words and more words in the loan descriptions as well as lower variation in both variables. Hence, we can neither approve nor reject *Hypothesis 1a (orthography)* and *Hypothesis 2b (description length)*. Both findings are consistent with the results of Iyer et al. (2014), who also do not find a significant relation of spelling errors, but a significantly negative one of the text length, both with the default probability. Indeed, in our regressions the coefficient of *#Words* is also negative with a relatively high *Z*-statistic, albeit not significant.

The social and emotional motives indicator *KeyWord_Separ* is the only indicator which is significant at a 10% level. The positive coefficient suggests that loan applicants using these words have a higher probability of default. Possible problems in their personal lives may affect their repayment behavior. However, as this is the only significant effect we cannot confirm *Hypothesis 3b (social motives indicator)* in general.

All model specifications show a significant positive relationship only between the indicator variable *KeyWord_Business* and the probability of default. This is noteworthy as this dummy variable is also positively significant in the funding regression. Thus we can state a certain inefficiency meaning that the lenders positively appreciate loans for business purposes, which in turn are related to a higher probability of default. This finding is consistent with the weak evidence of Sonenshein et al. (2011) for such behavioral effects. However, this is the only seemingly irrationality that can be found when comparing the funding and the default regressions.[7]

Summarizing, we observe a less strong relation of the description-text related soft factors to the default probability as compared to the funding probability. Only the business keyword is significantly positively related, indicating some inefficiency, while providing a picture expectedly is negatively related to the

[7] Still the behavior can be rational if the interest rate is high enough to cover the expected losses, which is not in our scope.

Table 5
Regression results concerning the funding probability on Smava.

	Funding probability (*FGL*)				
	SF.I	SF.II	SF.III	SF.IV	
				Coeff.	AME
Variables related to hypotheses					
SpellError	−0.00178			−0.00189	−0.0004463
	(−0.998)			(−1.05)	
#Words		−0.000546		−0.000710	−0.0001674
		(−0.799)		(−0.958)	
$(\#Words)^2$		0.00000188		0.00000205	0.0000004800
		(0.818)		(0.867)	
KeyWord_Pos			0.0255	0.0357	0.008431
			(0.857)	(1.14)	
KeyWord_Neg			0.163**	0.173**	0.04077**
			(2.27)	(2.35)	
KeyWord_Fam			−0.145***	−0.139***	−0.03276***
			(−3.12)	(−2.93)	
KeyWord_Separ			0.0737	0.0798	0.01882
			(0.647)	(0.695)	
Soft controls					
KeyWord_Restruc	−0.00381	−0.00240	−0.00781	−0.000465	−0.0001096
	(−0.119)	(−0.0738)	(−0.243)	(−0.0142)	
KeyWord_Edu	0.0332	0.0380	0.0374	0.0483	0.01140
	(0.498)	(0.560)	(0.560)	(0.709)	
KeyWord_Business	0.0675	0.0688	0.0625	0.0752	0.01773
	(1.49)	(1.47)	(1.37)	(1.60)	
KeyWord_Leisure	−0.0372	−0.0336	−0.0284	−0.0218	−0.005148
	(−0.471) (−0.422)	(−0.358)	(−0.273)		
Picture	−0.116***	−0.114***	−0.114***	−0.103**	−0.02439**
	(−2.89)	(−2.74)	(−2.80)	(−2.46)	
Hard controls					
ln(I)	−3.19***	−3.18***	−3.19***	−3.16***	−0.7461***
	(−31.6)	(−30.9)	(−31.5)	(−30.1)	
ln(Volume)	−0.490***	−0.486***	−0.486***	−0.491***	−0.1159***
	(−20.3)	(−20.2)	(−20.2)	(−20.2)	
Mat Short	−0.0724**	−0.0704**	−0.0745**	−0.0754**	−0.01778**
	(−2.11)	(−2.05)	(−2.17)	(−2.18)	
Age	−0.00512***	−0.00524***	−0.00513***	−0.00530***	−0.001249***
	(−3.47)	(−3.54)	(−3.49)	(−3.55)	
Male	−0.109***	−0.112***	−0.105***	−0.106***	−0.02497***
	(−3.41)	(−3.51)	(−3.30)	(−3.28)	
Employment	Yes	Yes	Yes	Yes	
Schufa	Yes	Yes	Yes	Yes	
KDF	Yes	Yes	Yes	Yes	
FedState	Yes	Yes	Yes	Yes	
DAX	−0.00849	−0.0154	−0.00336	0.0184	0.004338
	(−0.0492)	(−0.0893)	(−0.0195)	(0.106)	
TurnYear	0.122***	0.117***	0.115***	0.123***	0.02905***
	(2.98)	(2.85)	(2.81)	(2.98)	
CONST	−1.10***	−1.10***	−1.14***	−0.994**	
	(−2.87)	(−2.83)	(−2.99)	(−2.52)	
AIC	234.61	210.57	178.46	207.60	
N	10,367	10,423	10,423	10,367	

Notes: Model specifications SF.I to SF.IV are simultaneous IV probit regressions for the funding probability. The column AME SF.IV shows the average marginal effect of the variables on the funding probability with respect to specification SF.IV. Z-statistics are shown in parenthesis. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. The AIC of a model only containing hard controls is 216.26. Reference categories: For *FedState* category *BY*, for *Employment* category *Employee*, for *Mat* category *Mid*, for *Schufa* category *H*, for *KDF* category 4. The variables are defined in Table 1.

default probability, a finding that matches the results of other studies. Altogether the market appears to be relatively efficient in the sense that soft factors do not have much prediction power with respect to the default.

5.2.2. Smava

The model specifications SD.I to SD.IV represent the regression results with *DEF* as dependent variable for Smava. SD.I to SD.III separately for each hypothesis include the related variables individually together with controls, while SD.IV includes all relevant variables. The last column shows the average marginal effects for the main specification.

Similarly to the results of Auxmoney, the coefficients of the variables *SpellError* and *#Words* are insignificant. Hence, the orthographical quality and the description length are both not related to the probability of default in our data set on Smava and we can neither approve nor reject Hypothesis 1b *(orthography)* and Hypothesis 2b *(description length)*. Again, as with Auxmoney the findings are consistent with the results of Iyer et al. (2014). Furthermore, we analyze the effects of the social and emotional motives indicator variables. By contrast to Auxmoney, none of the categories is significant. Thus, we can neither confirm nor reject Hypothesis 3b *(social and emotional motives)* on Smava.

Solely the appearance of words referring to education in the description text is significantly negatively related to the

Table 6
Regression results concerning the default probability on Auxmoney.

	Default probability (*DEF*)				
	AD.I	AD.II	AD.III	AD.IV	
				Coeff.	AME
Variables related to hypotheses					
SpellError	0.000708			0.00108	0.0002224
	(0.121)			(0.186)	
#Words		−0.000414		−0.000744	−0.0001526
		(−0.930)		(−1.56)	
(*#Words*)^2		0.000000510		0.000000630	0.0000001300
		(1.27)		(1.52)	
KeyWord_Pos			0.0404	0.0655	0.01343
			(0.647)	(1.02)	
KeyWord_Neg			0.0371	0.0533	0.01093
			(0.410)	(0.581)	
KeyWord_Fam			0.0152	0.0365	0.007483
			(0.213)	(0.500)	
KeyWord_Separ			0.232*	0.248*	0.05075*
			(1.76)	(1.87)	
Soft controls					
KeyWord_Restruc	−0.0716	−0.0697	−0.0826	−0.0742	−0.01521
	(−1.11)	(−1.07)	(−1.27)	(−1.13)	
KeyWord_Edu	−0.132	−0.132	−0.132	−0.118	−0.02417
	(−1.44)	(−1.42)	(−1.44)	(−1.26)	
KeyWord_Business	0.242***	0.251***	0.236***	0.261***	0.05346***
	(3.16)	(3.20)	(3.07)	(3.30)	
KeyWord_Leisure	−0.0788	−0.0920	−0.0987	−0.0968	−0.01985
	(−0.532)	(−0.606)	(−0.660)	(−0.636)	
Picture	−0.129**	−0.126**	−0.130**	−0.126**	−0.02587**
	(−2.01)	(−1.97)	(−2.03)	(−1.96)	
Hard_controls					
ln(I)	4.93***	4.93***	4.85***	4.90***	1.004***
	(7.98)	(7.83)	(7.71)	(7.73)	
ln(Volume)	0.145***	0.146***	0.142***	0.149***	0.03061***
	(2.65)	(2.68)	(2.59)	(2.73)	
Mat_Short	0.283***	0.279***	0.281***	0.281***	0.05760***
	(2.97)	(2.92)	(2.96)	(2.94)	
FundTime_Short	−0.0999	−0.0941	−0.112	−0.100	−0.007725
	(−0.726)	(−0.682)	(−0.810)	(−0.725)	
FundTime_Long	0.147**	0.152**	0.143**	0.152**	−0.009080
	(2.33)	(2.39)	(2.26)	(2.38)	
Schufa	Yes	Yes	Yes	Yes	
CEG	Yes	Yes	Yes	Yes	
DAX	0.0510	0.0317	0.0590	0.0468	−0.005902
	(0.149)	(0.0924)	(0.171)	(0.136)	
TurnYear	0.0436	0.0440	0.0426	0.0421	−0.008491
	(0.539)	(0.545)	(0.526)	(0.520)	
CONST	7.59***	7.63***	7.44***	7.50***	
	(5.14)	(5.12)	(4.94)	(4.99)	
AIC	−2695.36	−2727.55	−2724.87	−2734.85	
N	3298	3298	3298	3298	

Notes: Model specifications AD.I to AD.IV are simultaneous IV probit regressions for the default probability. The column AME AD.IV shows the average marginal effect of the variables on the default probability with respect to specification AD.IV. Z-statistics are shown in parenthesis. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. The AIC of a model only containing hard controls is −2663.78. Reference categories: For *FundTime* category *Mid*, for *Mat* category *Mid*, for *Schufa* category *M*, for *CEG* category *Red*. The variables are defined in Table 1.

probability of default.[8] A possible explanation is that people who are willing to take out a loan for their education have a great incentive to complete their education successfully in order to achieve a higher income afterwards. Consequently, they should have enough money for the repayment.

Summarizing, at Smava the soft factors nearly have no explaining power concerning the default probability, neither the application-text related one nor the conventional ones such as providing a picture.

[8] This is in line with the regressions for Auxmoney, where the coefficient of this variable is also negative, but not significant.

5.3. Effects of control variables on funding and default probability

In the following, the effects of the control variables are briefly presented.

5.3.1. Funding probability

The results suggest that posting a picture is negatively related with the funding probability on Smava. This contradicts both, the significant positive coefficient observed for Auxmoney and the findings of previous research concerning the U.S. P2P platform Prosper (e.g. Iyer et al., 2014). However, there could be an influence of the subject of the pictures, which is not analyzed in our study. Furthermore, only 11.7% of the applicants on Smava upload a picture which is significantly lower than the 49.0% of Auxmoney. Additionally, we find that a higher interest rate decreases the

Table 7
Regression results concerning the default probability on Smava.

	Default probability (*DEF*)				
	SD.I	SD.II	SD.III	SD.IV	
				Coeff.	AME
Variables related to hypotheses					
SpellError	0.00262			0.00212	0.0004409
	(−0.549)			(−0.443)	
#Words		−0.00158		−0.00223	−0.0004641
		(−0.999)		(−1.32)	
$(\#Words)^2$		0.00000237		0.00000365	0.0000007600
		(0.458)		(0.695)	
KeyWord_Pos			0.0781	0.124	0.02584
			(1.05)	(1.59)	
KeyWord_Neg			−0.0824	−0.0342	−0.007113
			(−0.611)	(−0.249)	
KeyWord_Fam			−0.0276	0.00824	0.001714
			(−0.252)	(0.0742)	
KeyWord_Separ			0.0439	0.0832	0.01729
			(0.216)	(0.407)	
Soft controls					
KeyWord_Restruc	0.0940	0.105	0.0761	0.0973	0.02022
	(1.19)	(1.32)	(0.960)	(1.22)	
KeyWord_Edu	-0.429^{***}	-0.401^{**}	-0.436^{***}	-0.405^{**}	-0.08411^{**}
	(−2.60)	(−2.42)	(−2.64)	(−2.46)	
KeyWord_Business	0.0662	0.0853	0.0574	0.0961	0.01998
	(0.594)	(0.755)	(0.516)	(0.854)	
KeyWord_Leisure	−0.0954	−0.0687	−0.111	−0.0881	−0.01831
	(−0.573)	(−0.412)	(−0.663)	(−0.526)	
Picture	0.0264	0.0601	0.0212	0.0560	0.01164
	(0.301)	(0.663)	(0.241)	(0.620)	
Hard controls					
ln(I)	3.35^{***}	3.43^{***}	3.40^{***}	3.53^{***}	0.7338^{***}
	(6.64)	(6.84)	(6.93)	(7.42)	
ln(Volume)	0.0765	0.0846	0.0777	0.0816	0.01696
	(1.34)	(1.49)	(1.38)	(1.44)	
Mat Short	0.332^{**}	0.342^{**}	0.345^{**}	0.363^{**}	0.07551^{**}
	(1.98)	(2.04)	(2.10)	(2.25)	
FundTime Short	−0.0834	−0.0807	−0.0753	−0.103	−0.02151
	(−0.893)	(−0.867)	(−0.814)	(−1.12)	
FundTime Long	0.112	0.117	0.109	0.118	0.02443
	(1.16)	(1.22)	(1.13)	(1.22)	
Age	−0.00191	−0.00247	−0.00166	−0.00249	−0.0005174
	(−0.520)	(−0.671)	(−0.450)	(−0.678)	
Male	0.0705	0.0733	0.0825	0.0682	0.01417
	(0.876)	(0.915)	(1.03)	(0.852)	
Employment	Yes	Yes	Yes	Yes	
Schufa	Yes	Yes	Yes	Yes	
KDF	Yes	Yes	Yes	Yes	
FedState	Yes	Yes	Yes	Yes	
DAX	1.57^{***}	1.55^{***}	1.61^{***}	1.61^{***}	0.3348^{***}
	(3.68)	(3.66)	(3.83)	(3.87)	
TurnYear	0.194^{*}	0.199^{*}	0.199^{*}	0.212^{**}	0.04403^{**}
	(1.81)	(1.87)	(1.86)	(2.00)	
CONST	-10.5^{***}	-10.7^{***}	-10.7^{***}	-11.0^{***}	
	(−7.51)	(−7.72)	(−7.83)	(−8.33)	
AIC	66.32	73.97	65.12	67.75	
N	2208	2216	2216	2208	

Notes: Model specifications SD.I to SD.IV are simultaneous IV probit regressions for the default probability. The column AME SD.IV shows the average marginal effect of the variables on the default probability with respect to specification SD.IV. Z-statistics are shown in parenthesis. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. The AIC of a model only containing hard controls is 75.97 Reference categories: For *FedState* category *BY*, for *Employment* category *Employee*, for *FundTime* category *Mid*, for *Mat* category *Mid*, for *Schufa* category *H*, for *KDF* category 4. The variables are defined in Table 1.

probability of a successful funding on both platforms. This suggests that the investors do suspect that a higher interest rate than suitable for the solvency class is accompanied by a higher default rate. If the average marginal effect of $\ln(I)$ is related to one standard deviation[9] the impact on the funding probability is -18.46% on Auxmoney. Regarding Smava, the average marginal effect related to one standard deviation change equals -26.5% and thus has a even bigger magnitude than on Auxmoney. Thus, according to the average marginal effect analysis the interest rate is an important factor, which again proves that neglecting this variable, as other studies do, would lead to erroneous estimates. The effects of *Volume* and the solvency indicators like the Schufa score are intuitive on both platforms. Regarding the macroeconomic variables we derive ambiguous results. Whereas the results for Auxmoney indicate a significant positive relationship between *DAX* and the funding probability, suggesting that investors tend to finance ceteris paribus more loans in times of a positive economic climate, the same factor has a negative, but

[9] The standard deviation of $ln(I)$ is 43.93% which corresponds to $SD(I) = 3.24\%$ for Auxmoney.

Table 8
Subsample regressions concerning the solvency indicators for Auxmoney.

	Funding probability (*FGL*)				Default probability (*DEF*)			
	No solvency score (NSI)		Solvency score (SI)		No solvency score (NSI)		Solvency score (SI)	
	AFR.I	AME	AFR.II	AME	ADR.I	AME	ADR.II	AME
Variables related to hypotheses								
SpellError	-0.01^{***}	-0.0021^{***}	-0.01^{***}	-0.0026^{***}	0.00	0.0004	−0.00	−0.0000
#Words	0.00^{***}	0.0007^{***}	0.00^{***}	0.0010^{***}	0.00	0.0001	-0.00^{*}	-0.0002^{*}
$(\#Words)^2$	-0.00^{***}	-0.0000^{***}	-0.00^{***}	-0.0000^{***}	−0.00	−0.0000	0.00	0.0000
KeyWord_Pos	0.11^{***}	0.0227^{***}	0.16^{***}	0.0503^{***}	−0.10	−0.0177	0.14^{*}	0.0305^{*}
KeyWord_Neg	−0.01	−0.0021	−0.02	−0.0062	0.09	0.0162	0.07	0.0148
KeyWord_Fam	−0.01	−0.0024	0.07^{***}	0.0205^{***}	0.28^{**}	0.0496^{**}	−0.14	−0.0298
KeyWord_Separ	−0.03	−0.0058	−0.05	−0.0156	0.22	0.0382	0.30^{*}	0.0637^{*}
Soft controls								
KeyWord_Restruc	0.15^{***}	0.0327^{***}	0.00	0.0007	−0.17	−0.0293	−0.00	−0.0008
KeyWord_Edu	0.03	0.0059	0.01	0.0022	−0.03	−0.0058	−0.15	−0.0326
KeyWord_Business	0.06^{**}	0.0138^{**}	0.03	0.0086	0.22	0.0376	0.29^{***}	0.0615^{***}
KeyWord_Leisure	−0.02	−0.0053	-0.09^{**}	-0.0283^{**}	−0.46	−0.0803	0.10	0.0207
Picture	0.48^{***}	0.1042^{***}	0.32^{***}	0.0984^{***}	-0.23^{**}	-0.0393^{**}	−0.07	−0.0142
Hard controls								
ln(I)	-1.04^{***}	-0.2249^{***}	-1.95^{***}	-0.6071^{***}	4.62^{***}	0.8058^{***}	4.83^{***}	1.0270^{***}
ln(*Volume*)	-0.20^{***}	-0.0436^{***}	-0.44^{***}	-0.1354^{***}	−0.02	−0.0029	0.30^{***}	0.0647^{***}
Mat_Short	0.17	0.0358	0.21^{***}	0.0662^{***}	0.30^{*}	0.0515^{*}	0.22^{*}	0.0469^{*}
FundTime_Short					0.08	0.0143	−0.29	−0.0611
FundTime_Long					0.06	0.0100	0.20^{**}	0.0416^{**}
Schufa	Yes		Yes		Yes		Yes	
CEG	Yes		Yes		Yes		Yes	
DAX	0.78^{***}	0.1682^{***}	−0.09	−0.0296	−0.50	−0.0868	0.69	0.1468
TurnYear	-0.08^{***}	-0.0180^{***}	0.07^{***}	0.0218^{***}	−0.21	−0.0369	0.17^{*}	0.0368^{*}
CONST	-2.18^{***}		-2.19^{***}		8.12^{***}		6.13^{***}	
N	55,233		21,384		1311		1987	

Notes: AFR.I and ADR.I are subsample regressions with respective specifications to AF.IV and AD.IV (main results, already shown in Tables 4 and 6) for a subsample containing only loans without solvency indicators. AFR.II and ADR.II are subsample regressions with respective specifications to AF.IV and AD.IV (main results, already shown in Tables 4 and 6) for a subsample containing only loans with solvency indicators. The columns indicated with AME show the average marginal effects. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. Reference categories: For *FundTime* category *Mid*, for *Mat* category *Mid*, for *Schufa* category *M*, for *CEG* category *Red*. The variables are defined in Table 1.

not significant coefficient for Smava. Concerning the turn-of-the-year dummy *TurnYear*, we find a negative effect for Auxmoney and a positive effect for Smava. Remember that the following control variables are only available for Smava. The significant negative coefficient of *Age* validates the findings of Pope and Sydnor (2011) on Prosper. As women have a significantly higher chance of obtaining a loan on Smava, which is shown in the negative coefficient of *Male*, another result of Pope and Sydnor (2011) is also confirmed in the German P2P market. Furthermore, only two federal state dummies have a positive coefficient while most of the other state variables are insignificant in all specifications. Moreover, pensioners and self-employed workers have a better chance of being funded than employees and workers who form the reference category.

5.3.2. Default probability

Similarly to the funding probability, the interest rate shows highly predictive power in explaining the default probability on both platforms. The highly significant positive coefficients of ln(I) and the high magnitudes of this effect are remarkable. Thus, an increase of the interest rate by one standard deviation increases the likelihood of default ceteris paribus by 14.27% on Auxmoney and by 25.99% on Smava. A higher interest rate results in a higher debt service and could therefore be more difficult for borrowers to repay. This finding is consistent with Freedman and Jin (2008), who analyze this issue based on a Prosper data set. Furthermore, the indicator *Picture* is significantly negative on the 5% level in all regressions concerning Auxmoney. Remember, that a picture increases the funding probability and is therefore seen as a positive signal from an investors perspective. The negative coefficient of *Picture* supports this view. Loan applications including a picture have a ceteris paribus 2.59% lower likelihood of defaulting. However, this effect cannot be shown for Smava, for which we observe a positive but insignificant coefficient. Additionally, we find that the length of the funding process affects the default probability. Apparently, loans with a funding period greater than 10 days (*Long*) have a significant higher probability of default on Auxmoney compared to the reference category *Mid*. A possible explanation for this effect is that investors can derive information upon the solvency of a loan applicant to some degree from the application and bid hesitantly for less solvent applicants. Vice versa, for rather solvent borrowers, some kind of rational herding behavior can occur (Lee and Lee, 2012). Concerning Smava, coefficients are similar but not significant. The variable *ln*(*Volume*) has a significantly positive effect on the probability of default only for Auxmoney. The significant positive coefficients of the indicator for a short maturity on both platforms are surprising. However, as many long-term loans have not been closed at the end of our investigation period, they are not included in the CGL subsamples which therefore over-represent short-term loans. Furthermore, some Schufa scores have significant coefficients on both platforms. The coefficients of the macroeconomic controls *TurnYear* and *DAX* are significantly positive in all specifications for Smava. Contrary to the results of Auxmoney, this suggests that loan applications commenced in December or January and/or in times with better economic sentiment predict a higher probability of default.

5.4. Robustness checks

We perform several model variations and subsample regressions as robustness checks which are shown in Tables 8 and 9 for Auxmoney and Table 10 for Smava.

5.4.1. Residual interest rate

Our analyses have so far proven that a highly predictive factor for the funding success and defaults along with descriptive texts is the interest rate. For both platforms our results indicate that a higher interest rate is associated with a lower funding and a higher default probability. Particularly, the first result is not intuitive at first sight. Rational investors are expected to fund loans that pay a higher interest rate for a certain amount of risk more likely. However, as already mentioned before, the interest rate that a loan applicant suggests might include substantial information about his personal solvency sentiment. Our results are already an indication for this. One might argue that this effect might be biased because it is not clear to what extent the interest rate is being set to account for the expected credit risk and what value the actual surplus is. Therefore, we conduct a robustness check that substitutes the interest rate with the residual interest rate (*ResRate*), which is defined as the nominal interest rate minus the risk adjusted market rate according to the Schufa score. In this setting, the *ResRate* captures the effect that a borrower is willing to pay a higher or lower interest rate than the risk adjusted common market rate. Note that using the *ResRate* can still be a source of endogeneity, as other explanatory variables than the Schufa score might influence this measure. Therefore, we apply the IV probit approach again.

The results for the funding and the default probability are shown in the first two columns of Table 9 for Auxmoney and in Table 10 for Smava. For both platforms, the coefficients of *ResRate* are similar and highly significant. Comparably to the main regressions, the effect of *ResRate* is negative concerning the funding success and positive regarding the default event. This is a strong indication for the theory that a higher interest rate offered by a potential borrower is a signal for lower solvency sentiment.

Table 9
Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Auxmoney.

	Application *ResRate*		Data quality
	FGL AFR	*DEF* ADR	*DEF* ADDQ
Variables related to hypotheses			
SpellError	−0.00760***	−0.00361	−0.00716
# Words	0.00319***	−0.00149**	−0.00100*
(# Words)^2	−0.00000172***	0.00000121	0.000000930
KeyWord_Pos	0.169***	0.139	0.0834
KeyWord_Neg	−0.0126	0.110	0.0655
KeyWord_Fam	0.0746***	−0.130	0.0196
KeyWord_Separ	−0.0538	0.240	0.348**
Soft controls			
KeyWord_Restruc	−0.0124	−0.0138	−0.0883
KeyWord_Edu	−0.0203	−0.144	−0.0905
KeyWord_Business	0.0158	0.308***	0.273***
KeyWord_Leisure	−0.115**	0.0771	−0.0490
Picture	0.338***	−0.0717	−0.115
Hard controls			
ln (I)			5.06***
ResRate	−19.5***	33.6***	
ln(Volume)	−0.435***	0.311***	0.123**
Mat_Short	−0.00543	0.318**	0.328***
FundTime_Short		−0.258	0.0565
FundTime_Long		0.167*	0.150**
Schufa	Yes	Yes	Yes
CEG	Yes	Yes	Yes
DAX	−0.109	0.495	0.182
TurnYear	0.0651***	0.198*	0.0642
CONST	2.25***	−4.01***	7.93***
N	18,954	1,771	2,459

Notes: AFR and ADR are regressions with respective specifications to AF.IV and AD.IV (main results, already shown in Tables 4 and 6) using the residual interest rate. ADDQ is a regression with respect to specification AD.IV (main results, already shown in Table 6) for a subsample containing only loans with *#Lender*⩾ 10. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. Reference categories: For *FundTime* category *Mid*, for *Mat* category *Mid*, for *Schufa* category *M*, for *CEG* category *Red*. The variables are defined in Table 1.

Table 10
Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Smava.

	Application *ResRate*		Data quality
	FGL SFR	*DEF* SDR	*DEF* SDDQ
Variables related to hypotheses			
SpellError	−0.00207	0.00198	0.00824
# Words	−0.0000399	−0.00258	−0.00148
(# Words)^2	0.000000850	0.00000428	0.000000690
KeyWord_Pos	0.0357	0.133*	0.111
KeyWord_Neg	0.192***	−0.00973	−0.177
KeyWord_Fam	−0.142***	0.00182	0.226
KeyWord_Separ	0.0578	0.154	0.187
Soft controls			
KeyWord_Restruc	−0.0144	0.0992	0.0260
KeyWord_Edu	0.0533	−0.430**	−0.385*
KeyWord_Business	0.0687	0.0982	0.232
KeyWord_Leisure	−0.00198	−0.0841	−0.181
Picture	−0.101**	0.0549	−0.0596
Hard controls			
ln (I)			2.43***
ResRate	−37.5***	31.5***	
ln(Volume)	−0.484***	0.0794	0.000307
Mat_Short	−0.134***	0.491***	−0.0634
FundTime_Short		−0.0378	−0.000567
FundTime_Long		0.117	0.132
Age	−0.00387***	−0.00133	−0.00230
Male	−0.0981***	0.0565	−0.0372
Employment	Yes	Yes	Yes
Schufa	Yes	Yes	Yes
KDF	Yes	Yes	Yes
FedState	Yes	Yes	Yes
DAX	−0.189	1.13***	1.15**
TurnYear	0.0836**	0.209*	0.248
CONST	7.59***	−3.77***	−6.82***
N	10,367	2208	1007

Notes: SFR and SDR are regressions with respective specifications to SF.IV and SD.IV (main results, already shown in Tables 5 and 7) using the residual interest rate. SDDQ is a regression with respect to specification SD.IV (main results, already shown in Table 7) for a subsample containing only loans with *#Lender*⩾ 9. The symbols *, ** and *** express significance at the 10%, 5% and 1% level. For *FedState* category *BY*, for *Employment* category *Employee*, for *FundTime* category *Mid*, for *Mat* category *Mid*, for *Schufa* category *H*, for *KDF* category 4. The variables are defined in Table 1.

Regarding the hypotheses-related variables, the results are stable and we observe only small changes. In the case of Smava, only the indicator for positive emotions shows a significantly positive relation with the default probability. This finding is a weak evidence supporting *Hypothesis 3b (social and emotional motives)*. For Auxmoney, the family-related keyword indicator becomes highly significant in Specification AFR supporting *Hypothesis 3a (social and emotional motives)* and concerning the default probability *#Words* is now significantly negative on the 10% level. Note that the Auxmoney samples are significantly reduced in this setting because only observations containing a Schufa score can be considered.

5.4.2. Subsample regressions for solvency information on Auxmoney

One important difference between both platforms is that a solvency score (Schufa or CEG score) was not mandatory for

Auxmoney before February 2013. A share of 72.1% of all observations in the Auxmoney data have no solvency score at all. Thus, the question arises whether the soft factors resulting from the description text become more important whenever solvency scores are missing. For this reason, we perform regressions on *FGL* and *DEF* on two disjunct subsamples, one including observations with at least one solvency score (SI) and one without (NSI). The results are presented in Table 8. Surprisingly, the results appear to be reasonably stable. With regard to the funding probability, the family-related keyword indicator becomes significantly positive for the subsample with solvency scores. This is surprising, as we expected soft factors to play a bigger role, whenever hard facts are scarce. The result is also different to Smava, where we do not observe such an effect. With regard to average marginal effects, we observe a similar picture. The hypotheses related average marginal effects do not differ a lot between the two funding related subsamples. When considering the subsample with solvency information, the magnitude of the interest rate is much higher than in the other subsample. This is economically plausible, as it is easier for a potential lender to decide whether an interest rate is suitable in the case that a solvency score is available. The higher average marginal effect of the variable *DAX* in the subsample without any solvency score indicates that investors tend to finance those loans especially in times of economic prosperity.

We observe more coefficient changes with regard to the default probability. If no solvency score is available, *KeyWord_Fam* turns significant and *KeyWord_Separ* insignificant instead. For the other subsample, *#Words* becomes significantly negative and *KeyWord_Pos* significantly positive. Although the average marginal effects of *SpellError* are insignificant in both subsamples, the values differ considerably (0.036% for NSI vs. −0.003% for SI). Furthermore, we find that the average marginal effect of the indicator *Picture* shows more than the doubled amount in the NSI subsample. However, we can not find strong evidence supporting the fact that soft information related to the description text is more important whenever hard facts are not available on Auxmoney.

5.4.3. Data quality Wiseclerk

Last, we perform an additional check to test whether there are any indications for a bias due to unreported defaults on Wiseclerk. To this end, we utilize only those closed granted loans with at least ten lenders, which corresponds to a share of 75% on Auxmoney.[10] The regressions on this subsample show fairly similar results with two additional coefficients now becoming significant, but without a change of the sign (see Table 9, Specification ADDQ). Regarding Smava, using only those loans with at least nine lenders corresponds to the upper 46%.[11] Again, the regressions do not change much (see Table 10, Specification SDDQ). Altogether, there is no evidence in favor of an unreported-default bias.

5.5. Comparisons of both platforms

Last, we compare the results of Auxmoney and Smava. As already evident in the isolated analysis, there are different factors on both platforms which are significantly related to the funding success. Orthography, text length, the social and emotional motive indicator *KeyWord_Pos* and most of the other indicator variables are included in the investors' loan assessment on Auxmoney. Although *KeyWord_Fam* and *KeyWord_Neg* have significant coefficients, the other social and emotional motives indicators as well as the variables *SpellError* and *#Words* are not significantly related to the funding success on Smava. Concluding, one might argue, that the soft factors derived from the description texts are more important for investors in case that hard facts are not available, which is true for most of the observations on Auxmoney. However, the robustness check 'subsample regressions for solvency information on Auxmoney' proves that in the case of Auxmoney measures related to the description text are still highly predictive factors even in those cases in which solvency scores are available. This is a major difference in the investors' behavior on both platforms. One reason might be, that investors on Auxmoney are more used to considering soft information and analyzing the description texts. Furthermore, the bidding assistant and the verification of some provided information on Smava may reduce the incentive for investors to look at other factors than interest rate and the solvency information. Hence, the soft factors are more important on Auxmoney.

While the role of soft information in the funding process differs between the platforms, there is almost no distinction when considering the default probability regressions. Neither the orthography nor the text length are related to the probability of default on Auxmoney and Smava. However, the default rates on both platforms are different, as the default probability is mostly explained by hard factors, e.g. solvency information or the interest rate, which also are distinct between Auxmoney and Smava. Remember that the reason for this finding might be that soft information is indeed used by investors in their granting decision. If this is the case and the factors help to effectively distinct between good and bad loans, the observations of closed granted loans tend to exhibit a corresponding moulding. In the case of Auxmoney, we find several indications for such a pattern. Particularly, the variables *SpellError* and *#Words* are highly significant factors for the funding probability and for both variables, the distributions of the overall sample and of the subsample of closed, granted loans differs a lot.

Furthermore, our results suggest an astonishing finding concerning the interest rates which holds for both platforms. Investors on Smava and Auxmoney seem to mistrust a higher (residual) interest rate and therefore, a higher interest rate is related to a lower funding probability. When considering the defaults, a higher (residual) interest rate indicates a higher default probability on both platforms. Note that we cannot assess the profitability of the investments directly. However, if a loss given default (LGD) of even 90% is assumed, the annual rates of return on the average loan of the CGL subsamples are 0.41% (Auxmoney) and −1.08% (Smava). For an LGD level of 10%, the corresponding values are 5.55% (Auxmoney) and 2.25% (Smava). Hence, investments in loans arranged by Auxmoney, which often lack credit scores, outperform those into Smava loans during the observation period. This shows that investors are able to effectively identify creditworthy borrowers even though hard facts are scarce. Our results indicate that investors then base their granting decision successfully on soft factors that are related to the description texts. Loan applicants without any or without sufficient credit scores are not serviced by banks, which do not gather information regarding soft factors in the same way as P2P platforms. Identifying the borrowers with good solvency amongst the group of these applicants may be profitable. Maybe this is one reason why Auxmoney was able to replace Smava as market leader in Germany.

6. Conclusion

In this article we analyze the role that soft information derived from description texts plays in the funding decision and in predicting the default probability in P2P lending. We especially focus on spelling errors, text length and the presence of social and

[10] Again, with the above conservative calculus (see Section 4.1) assuming a 50% reporting probability of every lender, this means that in this subsample, only less than 0.34 errors can be expected. Thus, we consider this subsample as free of such errors.

[11] With the conservative calculus assuming a 50% reporting probability of every lender this means that in this subsample only less than 0.47 errors can be expected.

emotional keywords in the description text. We are the first to investigate these factors simultaneously for two leading platforms operating in the same target market but with different platform designs. This setting allows us to derive novel insights regarding the behavior of the market participants. We use simultaneous IV probit regressions to account for interest-rate-related endogeneity and data from two differently designed leading European P2P platforms, one with a more (Smava) and one with a less (Auxmoney) restrictive application process. In several robustness checks, we find that our results are resilient.

Our findings are new and partially surprising: Overall, it turns out that there is no such thing as a generalizable stable role that soft factors play in P2P lending and that the value for the investors depends on the platform design and the requirement of credit scores. In particular, spelling errors, text length and keywords evoking positive emotions are significant drivers of the funding probability on Auxmoney, while on Smava only two keywords are. The relation of the text length turns out to be inversely u-shaped. However, these factors appear not to be related to the default probability. When analyzing the (smaller) subsamples of closed granted loans with respect to the probability of default, we find that almost none of the soft factors are significant anymore. Yet, the usual control variables such as solvency scores and especially the interest rate are. Additionally, we identify the interest rate as an important factor that correlates with both, the funding and the default probability. We find that high interest rates show a positive relation with the default probability. This effect is also regarded as a signal for lower solvency by potential investors on both platforms. Altogether the evidence indicates a relatively efficient and rational market. Even though Auxmoney allows borrowers to apply for a loan without providing a credit score, which is not possible in conventional banking, we observe the risk-return profile to be sufficient to ensure an acceptable average return for the investors. As our results are mainly based on correlation analysis, even the confirmed hypotheses do not establish a causal relation. Therefore, a limitation of our research lies in the fact that the reasoning behind the hypotheses cannot be proven.

Summarizing, we can conclude that investors on P2P platforms react to soft information related to the description texts when deciding upon funding. The extent of reacting appears to depend on the platform's hard information requirements for loan applications. By following the soft information the investors do not act irrationally in the sense that the repayment behavior of the granted loans is almost solely dependent on hard facts. Some soft factors may even help to identify debtors with a good level of creditworthiness. Therefore, P2P platforms can indeed provide loans for people who would otherwise not have been able to receive a loan. Yet, this market extension does not come with additional risk for well-diversified investors as long as the interest rate is set in a way which accounts for the hard facts. From this point of view, the present tendency of P2P platforms to standardize the loan application process similar to that of banks is to be considered critically as it partially erodes the benefits of P2P lending.

Acknowledgement

A very early version of this article in German language was entitled 'Der Einfluss von weichen Faktoren im P2P-Lending – Ein Vergleich der Plattformen Smava und Auxmoney' and was awarded the '2nd Prize Postbank Finance Award 2014'. Furthermore, we are grateful to an anonymous referee and Alexander Stolz for helpful comments and thank the participants of the FMA Annual Meeting 2015 in Orlando.

Appendix A. Additional tables

See Tables A.11–A.14.

Table A.11
P2P-specific adaptions of the GNU Aspell regarding the spell check. Words that have been classified by the GNU Aspell as erroneous, but appeared more than ten times in the analysis have been checked manually regarding the correctness of the spelling. Thereby, we identified some terms that were indeed correctly spelled but were not included in the GNU Aspell. Therefore, we replenished the GNU Aspell by terms shown in this Table.

A	Abbezahlung, abgezockt, ABS, Abschluß, Abverkauf, Abzocke, ADHS, AGB, AIS, ALG, Android, Anschluß, Anschubfinanzierung, Antalya, App, Apps, arbeitssuchend, Arvato, Astra, ASU, Aufstockungskredit, ausgelernter, Auskunfteien, Auslegware, auxmoney, Auxmoney, Avant, Avensis, Azubi
B	BAföG, Barclay, Barclaycard, Basisscore, berufsbedingten, Berufsunfähigkeitsversicherung, Besicherung, BHKW, BHW, Bianca, Bio, bißchen, Bistro, Bitcoin, Bj., BJ, BU, Burnout, BWA, BWL
C	Caddy, Carport, Carport, Cashflow, Catering, CDI, CEG, Chevrolet, CHF, Christopher, Clio, CLK, CNC, Coach, Coaching, Combi, Community, Consultant, Controlling, Corsa, Creditreform, Cruiser
D	Dachgeschoß, Dacia, Dämmung, Daniela, daß,.de, Deko, DHL, Disco, Discount, Discounter, Dispo, Dispoausgleich, Dispokredit, Dispokredite, Dispokredites, Dispokredits, Disporahmens, Dispos, Dispozinsen, DJ, Dominic, DPD, dreiköpfige
E	EC, Edit, EEG, EFH, Eigentümergemeinschaft, Einliegerwohnung, Erbengemeinschaft, Erdgeschoß, Ergotherapeutin, Ergotherapie, Erledigungsvermerk, Erwerbsminderungsrente, Erwerbsunfähigkeitsrente, Escort, ESP, Espace, Estrich, ETW, EUR, Event, Events, Exfrau, Exfreund, Exfreundin, Exmann
F	Fabia, Factoring, fahrtüchtig, Fam., festangestellt, festangestellte, FH, Fiesta, Filialleiter, Filialleiterin, Fixum, Focus, Franchise, Franchisegeber, Franchisenehmer, Freelancer, Freiberuflichkeit
G	Gabionen, Galaxy, ganztags, Gerüstbau, Gesellenprüfung, Grunderwerbsteuer, GT
H	Hartz, Herzenswunsch, hochladen, Homeoffice, HTC, HUK, Hyundai
I	Ibiza, iMac, Imkerei, Infoscore, Inkassobüro, Inkassobüros, iPad, iPhone
J	Jasmin, Jennifer, Jenny, Jessica, Julian
K	Ka, Katja, KDF, Kevin, KfW, Kia, Kids, Kitaplatz, KMU, Kontokorrentkredit, kostendeckend, Kostgeld, krankgeschrieben, kV, kWh
L	Label, Laguna, lasern, Laura, LBS, LEGO, Leon, Lifestyle, Limousine, Lounge, Luca, Lupo
M	Macao, MacBook, Maik, mailen, Maklercourtage, Malerbetrieb, Mandy, Manuel, Marco, Marcus, Marina, Mario, Marvin, Master, Masterstudium, Mathias, MBA, Mechatroniker, Merchandising, MfG, Mia, Michelle, Micro, mietfrei, Mike, mittelständige, mittelständigen, Model, monatl., Mondeo, Monique, Mountainbike, MPU, mtl., Münsterland, muß, mußte, müßte, mußten, müßten
N	Nachfinanzierung, nachzahlen, Nancy, Newsletter, Nico
O	Octavia, offenstehende, OP, ÖPNV
P	Partyservice, Passat, PayPal, Photovoltaikanlage, Physiotherapie, Playstation, Polo, Portokasse, Postident, PostIdent, Printmedien, Promoterin, Provisionsbasis, Provisionszahlungen, PTA, Punto, PVC
R	Ranking, Rasenmäher, Ratenhöhe, Ratenkredit, Ratenkredite, Reha, Rene, renovierungsbedürftig, Renovierungskosten, Restaurantfachfrau, Restaurantfachmann, Restaurantleiter, RKV, Roadster, Roller, Ronny, Roswitha, Rover, RSV
S	Santander, Sarah, Schlecker, Schluß, schmerzfrei, schnellstens, Schufa, SCHUFA, Schufaauskunft, Schufaeintrag, Schufaeinträge, Schufascore, Schufawert, Schuldnerberatung, schwerbehindert, schwerbehinderten, Science, Score, Scores, Scorewert, Scoring, Seat, Security, Semesterbeitrag, SEO, Sharan, Shirts, Silvia, Sklerose, Skoda, Sky, smava, Smava, Smavaner, Snacks, Solaranlage, Solaranlagen, Solarenergie, Sollzinsen, Sorgerecht, Speditionskaufmann, Spielothek, Sportback, Stauraum, Steven, Stickmaschine, Style, Suzuki, SWK
T	Tablet, Tacho, Targobank, TDI, Teamleiter, TEUR, Timo, Touran, Touring, Trader, Trading, Tsd., Tuning, Turbo, Twingo
U	Überschuß, Überziehungszins, Überziehungszinsen, UG, Umfinanzierung, Uniklinik, UPS, USD
V	Vanessa, Variant, Vectra, verh., Vespa, Viktor, Vinyl, VIP, vorfinanzieren, vorfinanziert
W	Wärmedämmung, wegzukommen, Wellness, Wellnessbereich, WG, Whirlpool, wußte
X	Xenon
Y	Yamaha
Z	Zafira, zuteilungsreif

Table A.12
Keywords regarding loan purpose and classification.

Category	German keywords
KeyWord_Fam	Ehefrau, Ehemann, Erziehung, Familie, Heirat, Hochzeit, Kind, Kinder, verheiratet, Verlobung
KeyWord_Edu	Ausbildung, Studium, Weiterbildung
KeyWord_Leisure	Reise, Urlaub
KeyWord_Business	Betriebsmittel, Gewerb, Investition, selbstständig, Unternehmen
KeyWord_Restruc	Ablöse, Liquidität, Umschuld, Unterstützung, Dispo, Investition, Finanzamt
KeyWord_Neg	Beerdigung, klag, krank, schwierig, verstorben
KeyWord_Pos	danke, freuen, Traum, dringend, gesund, Wunsch, Vertrauen
KeyWord_Separ	geschieden, scheiden, Scheidung, Trennung

Table A.13
Pairwise Bravais–Pearson correlation coefficients among the explanatory variables concerning the Auxmoney data set.

		1	2	3	4	5	6	7	8
01.	*CEG_NA*	1.00							
02.	*CEG_Green*	−0.66*	1.00						
03.	*CEG_Yellow*	−0.61*	−0.12*	1.00					
04.	*CEG_Red*	−0.21*	−0.04*	−0.04*	1.00				
05.	*DAX*	−0.06*	0.04*	0.03*	0.03*	1.00			
06.	*FGL*	−0.41*	0.38*	0.17*	0.02*	0.05*	1		
07.	*ln(I)*	−0.13*	0.09*	0.08*	0.02*	0.02*	0.15*	1	
08.	*ln(I_rf)*	0.09*	−0.03*	−0.07*	−0.07*	−0.18*	−0.01*	0.04*	1.00
09.	*KeyWord_Business*	−0.11*	0.11*	0.03*	−0.00	−0.00	0.09*	0.05*	0.05*
10.	*KeyWord_Edu*	−0.05*	−0.01*	0.09*	−0.01*	−0.00	0.06*	0.03*	0.01*
11.	*KeyWord_Fam*	−0.03*	0.02*	0.02*	0.01*	−0.01*	0.05*	0.04*	0.04*
12.	*KeyWord_Leisure*	−0.01*	0.01*	0.01*	−0.00	−0.00	0.02*	0.01*	−0.01*
13.	*KeyWord_Restruc*	−0.14*	0.12*	0.06*	−0.01	0.01*	0.11*	0.04*	0.04*
14.	*KeyWord_Neg*	−0.05*	0.04*	0.03*	−0.00	−0.01*	0.05*	0.04*	0.04*
15.	*KeyWord_Pos*	−0.09*	0.04*	0.08*	0.01	−0.02*	0.12*	0.09*	0.07*
16.	*KeyWord_Separ*	−0.01*	0.02*	−0.00	−0.00	−0.01	0.02*	0.02*	0.02*
17.	*Mat_Short*	0.06*	−0.08*	−0.01*	0.04*	0.00	0.15*	−0.07*	0.13*
18.	*Mat_Mid*	−0.06*	0.08*	0.01*	−0.04*	−0.00	−0.15*	0.07*	−0.13*
19.	*Picture*	−0.15*	0.10*	0.09*	0.05*	0.04*	0.20*	0.13*	0.05*
20.	*SpellError*	0.12*	−0.09*	−0.07*	0.00	0.00	−0.12*	−0.07*	−0.05*
21.	*TurnYear*	0.03*	−0.03*	−0.04*	0.06*	0.02*	−0.01*	−0.04*	−0.29*
22.	*ln(Volume)*	−0.18*	0.23*	0.04*	−0.11*	0.01*	−0.05*	0.00	−0.01*
23.	*#Words*	−0.18*	0.13*	0.11*	−0.00	−0.01*	0.19*	0.11*	0.11*
24.	*(#Words)^2*	−0.02*	0.01*	0.01*	−0.00	−0.01	0.01*	0.01*	0.01*
		9	10	11	12	13	14	15	16
09.	*KeyWord_Business*	1.00							
10.	*KeyWord_Edu*	0.02*	1.00						
11.	*KeyWord_Fam*	0.01*	0.01*	1.00					
12.	*KeyWord_Leisure*	0.03*	0.00	0.08*	1.00				
13.	*KeyWord_Restruc*	0.15*	0.03*	0.00	−0.01	1			
14.	*KeyWord_Neg*	0.06*	0.04*	0.08*	0.04*	0.04*	1		
15.	*KeyWord_Pos*	0.07*	0.07*	0.13*	0.03*	0.07*	0.09*	1	
16.	*KeyWord_Separ*	0.01*	−0.00	0.07*	0.01*	0.03*	0.04*	0.04*	1
17.	*Mat_Short*	−0.01*	0.02*	−0.01*	0.02*	−0.03*	−0.00	0.04*	−0.00
18.	*Mat_Mid*	0.01*	−0.02*	0.01*	−0.02*	0.03*	0.00	−0.04*	0.00
19.	*Picture*	0.05*	0.04*	0.05*	0.01*	0.05*	0.04*	0.10*	0.02*
20.	*SpellError*	−0.08*	−0.06*	−0.05*	−0.01	−0.08*	−0.06*	−0.11*	−0.04*
21.	*TurnYear*	−0.01*	−0.02*	−0.01*	−0.01*	−0.01*	−0.00	−0.02*	−0.00
22.	*ln(Volume)*	0.14*	−0.04*	−0.01*	−0.01*	0.12*	0.00	−0.02*	0.00
23.	*#Words*	0.26*	0.16*	0.21*	0.12*	0.17*	0.22*	0.29*	0.14*
24.	*(#Words)^2*	0.04*	0.03*	0.03*	0.04*	0.02*	0.04*	0.03*	0.05*
		17	18	19	20	21	22	23	24
17.	*Mat_Short*	1							
18.	*Mat_Mid*	−1.00*	1						
19.	*Picture*	0.03*	−0.03*	1					
20.	*SpellError*	−0.03*	0.03*	−0.08*	1				
21.	*TurnYear*	−0.00	0.00	0.02*	0.02*	1			
22.	*ln(Volume)*	−0.45*	0.45*	−0.04*	−0.03*	−0.08*	1		
23.	*#Words*	0.01*	−0.01*	0.14*	−0.14*	−0.04*	0.10*	1	
24.	*(#Words)^2*	−0.00	0.00	0.02*	−0.01*	−0.01	0.02*	0.54*	1

Notes: The symbol * expresses significance at the 5% level. The variables are defined in Table 1.

Table A.14
Pairwise Bravais-Pearson correlation coefficients among the explanatory variables concerning the Smava data set.

		1	2	3	4	5	6	7	8	9	10	11
01.	*Age*	1.00										
02.	*DAX*	−0.04*	1.00									
03.	*FGL*	−0.04*	0.23*	1.00								
04.	*Employment_CivServant*	0.01	−0.02	0.01	1.00							
05.	*Employment_Employee*	−0.42*	0.02	0.03*	−0.22*	1.00						
06.	*Employment_Other*	−0.01	0.01	−0.01	−0.01	−0.06*	1.00					
07.	*Employment_Pension*	0.64*	0.01	0.01	−0.07*	−0.35*	−0.02	1.00				
08.	*Employment_Selfemp*	0.04*	−0.02	−0.04*	−0.14*	−0.75*	−0.04*	−0.22*	1.00			
09.	*KDF_1*	−0.01	−0.04*	−0.22*	0.00	0.05*	0.00	0.02	−0.06*	1		
10.	*KDF_2*	0.00	0.00	−0.01	0.02	0.02	– 0.01	0.04*	−0.06*	−0.24*	1	
11.	*KDF_3*	−0.02	0.05*	0.09*	−0.02	0.02	−0.00	−0.01	−0.00	−0.31*	−0.43*	1
12.	*KDF_4*	0.03*	−0.02*	0.09*	−0.00	−0.08*	0.01	−0.04*	0.11*	−0.24*	−0.33*	−0.42*
13.	*ln(I)*	−0.06*	−0.22*	0.05*	−0.08*	−0.10*	0.00	−0.01	0.15*	−0.16*	−0.06*	0.04*
14.	*ln(I_{rf})*	0.06*	−0.22*	−0.29*	−0.02	−0.03*	0.04*	−0.00	0.04*	0.06*	−0.04*	−0.08*
15.	*KeyWord_Business*	−0.04*	−0.05*	−0.05*	−0.06*	−0.18*	– 0.01	−0.10*	0.28*	−0.02	−0.04*	0.01
16.	*KeyWord_Edu*	−0.05*	−0.05*	−0.01	0.02	0.04*	– 0.01	−0.04*	−0.02	0.01	−0.00	−0.01
17.	*KeyWord_Fam*	0.01	−0.06*	−0.06*	0.04*	0.04*	−0.00	0.02	−0.07*	0.01	−0.01	−0.01
18.	*KeyWord_Leisure*	−0.01	−0.01	−0.02	0.01	0.04*	0.00	−0.01	−0.04*	0.02	0.01	−0.01
19.	*KeyWord_Restruc*	−0.04*	−0.05*	−0.02	0.01	−0.07*	−0.02	−0.06*	0.11*	−0.03*	−0.02	−0.01
20.	*KeyWord_Neg*	0.03*	−0.05*	−0.02	−0.02	−0.04*0	0.03*	0.01	0.04*	0	−0.02	−0.01
21.	*KeyWord_Pos*	−0.06*	−0.10*	−0.00	−0.01	0.04*	0.02	−0.02	−0.03*	−0.02	0	−0.00
22.	*KeyWord_Separ*	0.01	−0.02*	−0.00	0.01	0.00	– 0.01	0.00	−0.01	−0.02	0.02	−0.00
23.	*Male*	−0.11*	0.00	−0.03*	0.01	−0.02	−0.01	−0.09*	0.08*	0.01	−0.02*	0.00
24.	*Mat_Short*	0.00	−0.17*	*0.04*	0.05*	0.09*	0.04*	0.04*	−0.14*	0.21*	0.05*	−0.11*
25.	*Mat_Mid*	−0.00	0.17*	0.04*	−0.05*	−0.09*	−0.04*	−0.04*	0.14*	−0.21*	−0.05*	0.11*
26	*Picture*	−0.04*	−0.12*	−0.08*	−0.01	– 0.01	0.00	*0.04*	0.04*	0.01	−0.01	0.01
27.	*SpellError*	0.02	−0.00	−0.02	−0.02	0.01	0.01	0.01	−0.01	0.01	−0.01	0.00
28.	*TurnYear*	0.00	−0.12*	0.07*	0.02*	0.02	−0.02	0.03*	−0.05*	−0.04*	−0.00	0.04*
29.	*ln(Volume)*	0.07*	0.03*	−0.19*	−0.06*	−0.30*	−0.04*	−0.11*	0.42*	−0.09*	−0.03*	0.07*
30.	*# Words*	−0.07*	−0.17*	−0.10*	−0.01	−0.06*	0.00	−0.05*	0.10*	−0.01	−0.02	0.00
31.	*(#Words)^2*	−0.03*	−0.11*	−0.06*	−0.02	−0.05*	0.00	−0.01	0.07*	−0.01	−0.01	0

		12	13	14	17	16	17	18	19	20	21
12.	*KDF_4*	1									
13.	*ln(I)*	0.16*	1								
14.	*ln(I_{rf})*	0.08*	0.26*	1							
15.	*KeyWord_Business*	0.06*	0.06*	0.06*	1						
16.	*KeyWord_Edu*	0.01	0.01	0.07*	0.03*	1					
17.	*KeyWord_Fam*	0.01	−0.01	0.10*	0.05*	0.07*	1				
18.	*KeyWord_Leisure*	−0.01	0.02	0.09*	0.02	0.01	0.08*	1			
19.	*KeyWord_Restruc*	0.06*	0.01	0.10*	0.23*	0.05*	0.03*	−0.00	1		
20.	*KeyWord_Neg*	0.03*	0.06*	0.09*	0.06*	0.04*	0.05*	0.04*	0.08*	1	
21.	*KeyWord_Pos*	0.02	0.06*	*0.02	0.08*	0.05*	0.07*	0.02	0.09*	0.06*	1
22.	*KeyWord_Separ*	−0.00	0.00	0.05*	0.01	0.01	0.06*	−0.00	0.06*	0.03*	0.01
23.	*Male*	0.01	−0.04*	0.01	0.03*	−0.07*	0.01	−0.02	−0.01	−0.06*	−0.05*
24.	*Mat_Short*	−0.10*	0.05*	0.24*	−0.07*	0.010	0.01	0.04*	−0.05*	0.01	0.04*
25.	*Mat_Mid*	0.10*	−0.05*	−0.24*	0.07*	−0.00	−0.01	−0.04*	0.05*	−0.01	−0.04*
26	*Picture*	−0.01	0.04*	0.19*	0.11*	0.04*	0.11*	0.06*	0.06*	0.05*	0.11*
27.	*SpellError*	−0.00	0.02	0.07*	−0.04*	−0.01	−0.03*	0.00	−0.05*	−0.02*	−0.10*
28.	*Turn Year*	−0.01	−0.04*	0.01	−0.02	0.02	−0.01	−0.01	0.04*	0.02*	−0.01
29.	*ln(Volume)*	0.03*	−0.08*	−0.06*	0.21*	−0.00	0.01	−0.04*	0.11*	0.02	−0.01
30.	*# Words*	0.03*	0.10*	0.23*	0.29*	0.29*	0.23*	0.23*	0.23*	0.23*	0.32*
31.	*(#Words)^2*	0.01	0.08*	0.14*	0.21*	0.14*	0.16*	0.08*	0.16*	0.17*	0.19*

		22	23	24	27	26	27	28	29	30	31
22.	*KeyWord_Separ*	1									
23.	*Male*	−0.01	1								
24.	*Mat_Short*	0.02	0.01	1							
25.	*Mat_Mid*	−0.02	−0.01	−1	1						
26.	*Picture*	0.05*	0.00	0.05*	−0.05*	1					
27	*SpellError*	−0.02	0.01	−0.01	0.01	−0.02	1				
28.	*Turn Year*	0.01	−0.03*	*0.07*	0.07*	−0.03*	−0.00	1			
29.	*ln(Volume)*	0.01	0.04*	−0.38*	0.38*	0.05*	−0.03*	0	1		
30.	*# Words*	0.13*	−0.03*	0.00	−0.00	0.32*	−0.10*	−0.00	0.14*	1	
31.	*(#Words)^2*	0.10*	−0.01	−0.02	0.02	0.24*	−0.05*	0.01	0.11*	0.88*	1

Notes: The symbol *expresses significance at the 5% level. The variables are defined in Table 1.

References

Agrawal, A., Catalini, C., Goldfarb, A., 2013. Some simple economics of crowdfunding. Innovation Policy and the Economy, vol. 14. National Bureau of Economic Research, NBER Chapters.

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.

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.

Becker, G.S., 1971. The Economics of Discrimination, second ed. University of Chicago Press, Chicago, London, ISBN 9780226041049.

Belleflamme, P., Lambert, T., Schwienbacher, A., 2014. Crowdfunding: tapping the right crowd. Journal of Business Venturing 29 (5), 585–609.

Berger, S.C., Gleisner, F., 2009. Emergence of financial intermediaries in electronic markets: the case of online p2p lending. BuR – Business Research 2 (1), 39–65.

Bhatt, N., Tang, S., 2002. Determinants of repayment in microcredit: evidence from programs in the united states. International Journal of Urban and Regional Research 26 (2), 360–376.
Böhme, R., Pötzsch, S., 2010. Social lending aus der Perspektive des Datenschutzes. In: Freiling, F. (Ed.), SICHERHEIT 2010 – Sicherheit, Schutz und Zuverlässigkeit. Bonner Köllen Verlag, pp. 317–328.
Böhme, R., Pötzsch, S., 2011. Collective exposure: peer effect in voluntary disclosure of personal data. In: Danezis, G. (Ed.), Financial Cryptography and Data Security – 15th International Conference. Springer Verlag, pp. 1–15.
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, 9–26.
Cameron, A., Trivedi, P., 2010. Microeconometrics Using Stata. StataCorp LP, College Station, Texas, ISBN 9781597180733.
Chen, N., Ghosh, A., Lambert, N.S., 2014. Auctions for social lending: a theoretical analysis. Games and Economic Behavior 86, 367–391.
Diamond, D.W., 1984. Financial intermediation and delegated monitoring. The Review of Economic Studies 51 (3), 393–414.
Dowling, M., Lucey, B.M., 2005. The role of feelings in investor decision-making. Journal of Economic Surveys, 211–237.
Duarte, J., Siegel, S., Young, L., 2012. Trust and credit: the role of appearance in peer-to-peer lending. Review of Financial Studies 25 (8), 2455–2484.
Everett, C.R., 2010. Group Membership, Relationship banking and loan default risk: the case of online social lending, SSRN Working Paper 1114428.
Fabender, D., 2011. P2P-Kreditmärkte als Finanzintermediäre. GRIN Verlag, München.
Fershtman, C., Gneezy, U., 2001. Discrimination in a segmented society: an experimental approach. The Quarterly Journal of Economics 116 (1), 351–377.
Figueredo, L., Varnhagen, C.K., 2005. Didn't you run the spell checker? effects of type of spelling error and use of a spell checker on perceptions of the author. Reading Psychology 26 (4–5), 441–458.
Freedman, S., Jin, G.Z., 2008. Do social networks solve information problems for peer-to-peer lending? Evidence from prosper.com. SSRN Working Paper 1304138.
Gao, Q., Lin, M., 2015. Lemon or cherry? The value of texts in debt crowdfunding, Working Paper.
Gerber, E.M., Hui, J.S., Kuo, P.-Y., 2012. Crowdfunding: Why People are Motivated to Post and Fund Projects on Crowdfunding Plattforms. Northwestern University, Creative Action Lab.
GfK, 2010. GfK-Konsumklima November und Ausblick auf das Weihnachtsjahr 2010, Gesellschaft für Konsumforschung, Nürnberg.
Giudici, G., Nava, R., Lamastra, C.R., Verecondo, C., 2012. Crowdfunding: the new frontier for financial entrepreneurship, SSRN Working Paper 2157429.
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.
Hemer, J., 2011. A snapshot on crowdfunding, Working Papers Firms and Region R2/2011, Fraunhofer Institute for Systems and Innovation Research ISI.
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, 138–149.
Hildebrand, T., Puri, M., Rocholl, J., 2013. Adverse incentives in crowdfunding, SSRN Working Paper 1615483.
Hirshleifer, D., 2001. Investor psychology and asset pricing. The Journal of Finance 56 (4), 1533–1597.
Iyer, R., Khwaja, A.I., Luttmer, E.F.P., Shue, K., 2014. Screening peers softly: inferring the quality of small borrowers, SSRN Working Paper 1570115.
Kennedy, P., 2008. A guide to econometrics. Blackwell.
Kreiner, D.S., Schnakenberg, S.D., Green, A.G., Costello, M.J., McClin, A.F., 2002. Effects of spelling errors on the perception of writers. The Journal of General Psychology 129 (1), 5–17.
Lee, E., Lee, B., 2012. Herding behavior in online P2P lending: an empirical investigation. Electronic Commerce Research and Applications 11, 495–503.
Lin, M., Wei, Z., 2013. Auction vs. posted-price: market mechanism, lender behavior, and transaction outcomes in online crowd-funding, SSRN Working Paper 2328468.
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.
Loughran, T., McDonald, B., 2014. Measuring readability in financial disclosures. Journal of Finance 69, 1643–1671.
Lucas, R., 1972. Expectations and the neutrality of money. Journal of Economic Theory 4 (2), 103–124.
Meyer, A.G.L., 2013. Pricing Mechanisms in Peer-to-peer Online Credit Markets, Job Market Paper. Stanford University, Stanford University, Department of Economics.
Michels, J., 2012. Do unverifiable disclosures matter? Evidence from peer-to-peer lending. The Accounting Review 87 (4), 1385–1413.
Moenninghoff, S.C., Wieandt, A., 2013. The future of peer-to-peer finance. Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung: Zfbf. 65 (6), 466–487.
Park, M., Aiken, M., Tobin, L., Vanjani, M., 2010. Spelling and grammatical errors in electronic meetings. Issues in Information Systems 11 (1), 384–391.
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.
Pynte, J., Kennedy, A., Ducrot, S., 2004. The influence of parafoveal typographical errors on eye movements in reading. European Journal of Cognitive Psychology 16 (1/2), 178–202.
Ravina, E., 2012. Love & loans: the effect of beauty and personal characteristics in credit markets, SSRN Working Paper 1101647.
Renneboog, L., Horst, J.T., Zhang, C., 2008. Socially responsible investments: institutional aspects, performance, and investor behavior. Journal of Banking & Finance 32 (9), 1723–1742.
Rivers, D., Vuong, Q.H., 1988. Limited information estimators and exogeneity tests for simultaneous probit models. Journal of Econometrics 39, 347–366.
Solomon, J., Wash, R., 2014. Coordinating donors on crowdfunding websites. Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing, CSCW '14. ACM, New York, NY, USA, ISBN 978-1-4503-2540-0, pp. 38–48.
Sonenshein, S., Herzenstein, M., Dholakia, U.M., 2011. How accounts shape lending decisions through fostering perceived trustworthiness. Organizational Behavior and Human Decision Processes 115 (1), 69–84.
Van Wingerden, R., Ryan, J., 2011. Fighting for funds: an exploratory study into the field of crowdfunding.
Weiss, G.N.F., Pelger, K., Horsch, A., 2010. Mitigating adverse selection in P2P lending – empirical evidence from prosper.com, SSRN Working Paper 1650774.
WSI, 2013. Beschäftigte mit Weihnachtsgeld 2013, Wirtschafts- und Sozialwissenschaftliches Institut in der Hans-Böckler-Stiftung, Düsseldorf.