Martina Weber

Alternative Finance for Small and Medium-sized Enterprises

Universität Regensburg
Regensburg, Germany

Center of Finance

Universität Regensburg

Center of Finance – Dissertation series No. 1 (2018)

Alternative Finance for Small and Medium-sized Enterprises

Martina Weber

Center of Finance

Faculty of Business, Economics,
and Management Information Systems

Universität Regensburg

Martina Weber

Alternative Finance for Small and Medium-sized Enterprises

A dissertation submitted to the Faculty of Business, Economics, and Management Information Systems in partial fulfillment of the requirements for the degree of Doktor der Wirtschaftswissenschaft (Dr. rer. pol.)

Advisors:
Prof. Dr. Gregor Dorfleitner
Prof. Dr. Klaus Röder

Date of disputation: July 11, 2018

Universität Regensburg
Regensburg, Germany

Acknowledgments

I would like to thank Prof. Dr. Gregor Dorfleitner for his continuous advice and support throughout my dissertation. I appreciate his encouragement, motivation and the opportunities he provided me. Most importantly, I am thankful he convinced me to pursue my Ph.D. in the first place. I gratefully acknowledge Prof. Dr. Klaus Röder to be my second adviser.

Moreover, I am grateful to my co-authors, my colleagues and the fellow Ph.D. students at the University of Regensburg for their feedback, cooperation, and friendship.

Last but not least, I would like to thank my family and friends. This doctoral thesis would not have been possible without their almost unbelievable support. Special thanks go to my Mum, my Dad, Carsten, and Johannes.

Contents

1 Introduction **1**

2 Description-text related soft information in peer-to-peer lending **7**

2.1 Introduction 8
2.2 Literature review 10
2.3 Hypotheses development 11
2.4 Data and methodology 14
2.5 Results 23
2.6 Conclusion 39
2.7 Appendix 41

3 German Mittelstand bonds: Yield spreads and liquidity **45**

3.1 Introduction 46
3.2 Background 48
3.3 Data and methodology 49
3.4 Results 55
3.5 Conclusion 66
3.6 Appendix 69

4 Pricing in the online invoice trading market: First empirical evidence **76**

4.1 Introduction 77
4.2 Related literature 77
4.3 Data and methodology 79
4.4 Results 82
4.5 Conclusion 86

5 Dynamics of investor communication in equity crowdfunding **88**

5.1 Introduction 89
5.2 Theoretical foundation and hypotheses 90
5.3 Data 94

5.4 Results . . . 99
5.5 Conclusion . . . 112
5.6 Appendix . . . 114

6 The portfolio formation behavior of peer-to-business lending investors 116
6.1 Introduction . . . 117
6.2 Theory and hypotheses . . . 119
6.3 Data . . . 123
6.4 Methodology . . . 128
6.5 Results . . . 129
6.6 Conclusion . . . 137
6.7 Appendix . . . 142

7 Conclusion 144

Bibliography 146

List of Tables

2.1 Description of variables . . . 17
2.2 Descriptive statistics of metric variables . . . 19
2.3 Relative frequency distributions of categorical variables . . . 20
2.4 Regression results concerning the funding probability on Auxmoney . . . 24
2.5 Regression results concerning the funding probability on Smava . . . 25
2.6 Regression results concerning the default probability on Auxmoney . . . 28
2.7 Regression results concerning the default probability on Smava . . . 29
2.8 Subsample regressions concerning the solvency indicators for Auxmoney . . . 33
2.9 Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Auxmoney . . . 34
2.10 Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Smava . . . 35
2.11 Classification of keywords . . . 41
2.12 P2P-specific adaptions of the GNU Aspell regarding the spell check . . . 42
2.13 Pairwise Bravais-Pearson correlation coefficients, Auxmoney data set . . . 43
2.14 Pairwise Bravais-Pearson correlation coefficients, Smava data set . . . 44

3.1 LOT liquidity estimate . . . 55
3.2 Descriptive statistics of the Mittelstand bonds in the year of the emission . . . 57
3.3 Summary statistics of the Mittelstand bonds separated by year . . . 58
3.4 Liquidity measure tests . . . 59
3.5 Panel regression results with yield spread as dependent variable . . . 62
3.6 Panel regression results with yield spread as dependent variable; accounting variables added one by one . . . 63
3.7 Panel regression results using three-stage least squared regressions . . . 67
3.8 List of all Mittelstand bonds . . . 69
3.9 Definition of variables . . . 74

4.1 Definition of variables . . . 82
4.2 Descriptive statistics . . . 83

4.3 Regression results of the logit models 84
4.4 Regression results of the tobit models 85

5.1 Definition of variables 96
5.2 Summary statistics updates data set 100
5.3 Summary statistics investment data set 101
5.4 Regression results funding period, FGLS estimation 102
5.5 Regression results funding period, probit estimation 103
5.6 Regression results probability of updates 107
5.7 Regression results probability of updates, Cox proportional hazard model . . . 109
5.8 Time to first update 111
5.9 Probability of updates, fixed-effects logit estimation 115

6.1 Definition of variables 124
6.2 Descriptive statistics of the corporations and the investors, metric variables . . . 125
6.3 Descriptive statistics of the corporations and the investors, categorical variables 125
6.4 Descriptive statistics of the portfolios 127
6.5 Descriptive statistics of $\Delta RAROC$ 127
6.6 Regression results for investment behavior 130
6.7 Subsample regressions for investment behavior 133
6.8 Regression results for $\Delta RAROC$ 134
6.9 Subsample regressions for $\Delta RAROC$ 135
6.10 Robustness test for $\Delta RAROC$, different VaRs 138
6.11 Robustness test for $\Delta RAROC$, different dependent variables in the first stage . . 139
6.12 Probability of default of the loans 142

List of Figures

3.1 LOT liquidity estimate model 53

4.1 MarketInvoice scheme 80

5.1 Time to first update 104

Chapter 1

Introduction

Small and medium-sized enterprises (SMEs) are often referred to as the backbone of the European economy. In 2016, SMEs account for 99.8 % of all the enterprises in the non-financial business sector within the European Union. These SMEs provide approximately 93 million jobs and represent 67 % of the total employment in this sector (Muller et al., 2017). However, SMEs often face difficulties in satisfying their financing needs. Therefore, this dissertation empirically analyzes different sources of financing which represent alternatives to traditional financing systems for SMEs.

The European Commission (2003) defines SMEs as enterprises with a workforce of less than 250 employees, a turnover of less than 50 million EUR and / or a balance sheet total of less than 43 million EUR. Despite the importance of SMEs for the European and the worldwide economy, financial markets often fail to provide SMEs sufficient financing. Previous literature examines several reasons for this financing gap. Berger and Udell (1998) and Gregory et al. (2005) highlight the role of information asymmetries. Due to informational opacity it is often difficult to assess the creditworthiness of SMEs. Small businesses frequently lack a great deal of relevant data such as audited financial statements or credit histories. This results in high cost for the screening and monitoring of SMEs (Ang, 1992) and can also lead to credit rationing (Stiglitz and Weiss, 1981; Berger and Udell, 1998). Besides, many SMEs cannot provide sufficient collateral in the early stages of their development (Berger and Udell, 1998; Avery et al., 1998).

SMEs within in the European Union often rely on bank-related products to cover their financing needs (ECB, 2017). Since the implementation of the Basel II and III regulations, loan conditions have worsened for many SMEs. Nowadays, all firms applying for a loan have to provide a mandatory rating. In particular for poorly or non-rated SMEs this has resulted in higher costs of borrowing as banks face increased equity requirements for such loans (Schindele and Szczesny, 2015; Müller et al., 2011). Hence, many SMEs are in need of a remedy to overcome this financing gap.

In recent years, several alternative forms of finance have been introduced and many new players have entered the market. The need for financing on the one hand, and opportunities due to technical innovations on the other hand have boosted the growth of these new financing instruments. Providers of alternative sources of finance are able to mitigate some of the financing obstacles for SMEs described above by utilizing new technologies (Block et al., 2018b). For example, many new financing instruments take advantage of big data or the wisdom of the crowd in order to reduce agency problems.

With the increasing use of alternative finance the opportunity and the need for further research arises. It is important to analyze the new financing instruments in order to understand the market implications as well as the opportunities and the challenges for SMEs and investors. Furthermore, the increased data availability for many of these new financing forms allows drawing conclusions about both investor and entrepreneurial behavior.

This thesis focuses on several forms of alternative finance for SMEs. In particular, different forms of crowdfunding as well as Mittelstand bonds are examined. In this context, crowdfunding represents an external source of financing in which a relatively large number of individuals collectively raises capital (Belleflamme et al., 2014; Mollick, 2014). Different forms of crowdfunding can be distinguished depending on what the entrepreneurs promise the investors in return for their contributions (Bradford, 2012). In peer-to-peer (P2P) lending and peer-to-business lending the crowd grants a loan to individuals or businesses in order to receive a pre-determined interest rate and the repayment of the loan. By contrast, in equity crowdfunding investors are offered a share of the equity, debt, or mezzanine capital. In general, the contracts in equity crowdfunding include an equity-like profit participation for the investors (Klöhn et al., 2016). The two other forms of crowdfunding, namely donation-based and reward-based crowdfunding, are not examined in detail in this dissertation. While investors do not receive any consideration in donation-based crowdfunding, they receive some form of non-monetary compensation in return for their investments in reward-based crowdfunding (Bradford, 2012).

Academic research in crowdfunding has extensively investigated investors' behavior. Several factors signal quality to investors and thereby increase the likelihood of a successful funding of a crowdfunding campaign. In particular project characteristics such as the share of equity offered (see, e.g., Ahlers et al., 2015; Vismara, 2016; Bernstein et al., 2017) as well as borrower characteristics (see, e.g., Duarte et al., 2012; Herzenstein et al., 2011; Pope and Sydnor, 2011) and the amount of voluntary information throughout the campaign (see, e.g., Mollick, 2014; Kuppuswamy and Bayus, 2017; Block et al., 2018a) can help to explain the funding success of a crowdfunding campaign. Furthermore, literature highlights the importance of media use

(Courtney et al., 2017) and social capital (Colombo et al., 2015). Other streams of literature deal with follow-up fundings, exits and insolvencies of successfully funded crowdfunding projects (see, e.g., Signori and Vismara, 2018; Hornuf et al., 2018), market mechanisms (see, e.g., Wei and Lin, 2016; Chen et al., 2014) or regulation (see, e.g., Klöhn et al., 2016).

Online invoice trading, another form of alternative finance, is also closely related to crowdfunding. SMEs pre-finance their outstanding invoices through a relatively large number of individual and institutional investors. The investors do not assume the default risk of the invoice bought, hence, online invoice trading represents a form of recourse factoring.

In addition, this dissertation investigates Mittelstand bonds as an alternative source of finance for SMEs. Since 2010, SMEs in Germany have had the opportunity to directly access the capital markets by issuing small bonds. These Mittelstand bonds have a volume of less than 100 million EUR and have been issued at five German stock exchanges. However, due to several defaults two stock exchanges had already shut down their segments by 2016. Several research papers examine Mittelstand bonds. Kammler and Röder (2013) investigate the returns of Mittelstand bonds and reveal a negative internal rate of return. Schöning (2014) shows that the coupons of many Mittelstand bonds are well below the risk-adjusted value.

This thesis contributes to the growing literature on alternative sources of financing in manifold ways. In crowdfunding, the dissertation reveals several predictors of the funding success and the probability of default. Particularly the interest rate is important for explaining both the funding success and defaults in P2P lending as well as the probability of default in online invoice trading. Further determinants such as soft factors derived from the description text and the pricing regime are highlighted as well. Using data from crowdfunding platforms, this thesis also contributes to a better understanding of the behavior of both entrepreneurs and investors. With respect to Mittelstand bonds, the dissertations sheds light on the existence and the size of the liquidity premium investors demand for holding these bonds.

This thesis consists of five independent research papers with several co-authors. In the remainder of the introduction, the articles are briefly summarized with respect to the research question, the data, the statistical method, and the contribution. Chapters 2 to 6 of the dissertation present the research papers. The last chapter provides a conclusion.

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

In this article, we analyze the role of soft factors in predicting both the funding success and the probability of default in P2P lending. We use data from two leading European platforms, namely Auxmoney and Smava, and derive several soft factors from the description text of the loan applications. Even though both platforms serve the same market, they have implemented different platform designs. Auxmoney requests only few pieces of mandatory information and does not demand the applicants include a solvency score. By contrast, Smava is more restrictive and only allows loan applications with a minimum credit score.

We use simultaneous IV probit regression to overcome endogeneity concerns and find that the effect of soft factors depends on the platform design. On Auxmoney, the less restrictive platform, orthography, text length, and the mentioning of positive emotion evoking keywords are significantly related to the funding success. On Smava, however, only two keywords are associated with the funding probability. Hard factors appear to be more important in explaining the funding success on the latter platform. Interestingly, our results indicate a negative relationship between the funding success and the interest rate on both platforms. Investors appear to mistrust borrowers who are willing to pay extremely high interest rates.

Analyzing the loans which were closed and granted on both platforms, we find that soft factors are hardly related to the probability of default on either platform. Again hard information, in particular the interest rate, strongly predict the probability of default. Turning to profitability, we find that Auxmoney shows the better risk-return-profile for investors. Overall, our results suggest that soft factors help investors to effectively identify creditworthy borrowers in the absence of hard factors. If hard factors are provided, investors rely on these information to decide which loans to invest in.

German Mittelstand Bonds: Yield Spreads and Liquidity

Since the launch of Mittelstand bonds, the yield spreads observed have increased steadily. In this paper, we empirically analyze the importance of illiquidity of the bonds in explaining the yield spreads. We analyze 92 Mittelstand bonds and employ a cross-sectional model to measure the size of the liquidity premium. We use two different liquidity measures, namely the LOT liquidity estimate and the bid-ask spread. Furthermore, we control for the default risk as well as bond and firm characteristics. Our results suggest that investors do indeed demand a liquidity premium. The size of the liquidity premium for Mittelstand bonds is high and equals approximately twice the size of the liquidity premium of speculative grade US corporate bonds. We estimate a simultaneous equation model as a robustness check.

Pricing in the online invoice trading market: First empirical evidence

The possibility to pre-finance outstanding invoices through online invoice trading is an opportunity for SMEs to raise short-term debt. The central aim of this paper is to analyze whether the risk of payment difficulties is appropriately reflected in the pricing variables of online invoice trading platforms.

To this end, we use data from the UK-based platform MarketInvoice, the market leader in online invoice trading and analyze which factors explain the probability of default and the loss rate, respectively. We apply logit and tobit models and find that the interest rate is a key predictor of both defaults and the loss rate. Moreover, the duration and the percentage funded are also significantly related to the probability of default. Since the platform changed the pricing mechanism form a real-time auction to a fixed price mechanism within the observation period, we have the opportunity to compare two pricing regimes. Our results indicate that the probability of default is higher within the auction period. However, we also show that the net return investors gain are higher within the auction period.

Dynamics of Investor Communication in Equity Crowdfunding

Equity crowdfunding provides non-sophisticated private investors the possibility to invest in (highly risky) start-ups. The start-ups that seek funding via equity crowdfunding can voluntarily communicate with investors by posting updates. Previous research indicates that updates increase investments in the focal crowdfunding campaign (Mollick, 2014; Kuppuswamy and Bayus, 2017; Block et al., 2018a). We investigate the communication behavior of start-ups during and after the funding period and examine whether start-ups strategically post updates with specific language or content.

Therefore, we use hand-collected data from two major German equity crowdfunding portals, namely Companisto and Seedmatch, and apply several statistical models including probit estimations and survival analysis. We find evidence suggesting that start-ups strategically communicate with investors. First, we analyze changes in the communication behavior during and after the funding period. Our results show that start-ups post updates more frequently during the funding period and that start-ups use more linguistic devices that enhance the feeling of group cohesion and group identity. Second, we focus on the funding period and find that the probability of an update during the funding period increases along with strong competition of other contemporary crowdfunding campaigns.

Paralyzed by shock and confused by glut: The portfolio formation behavior of peer-to-business lending investors

In this article, we investigate the investor behavior on a peer-to-business lending platform and find evidence suggesting that investors suffer from two new investment biases—the default shock bias and the deep market bias. In particular, investors refrain from investing in new loans and stop diversifying their portfolio after experiencing a loan default. This behavior results in a worsening of the risk-return profile of their portfolio. Moreover, investors appear to be overwhelmed with a glut of simultaneously active loan campaigns. Consequently, they invest less in new loans which, in turn, has a negative effect on the risk-return profile of their portfolio. Investment experience on the platform reduces the effect of the deep market bias.

We use data from the German crowdlending platform ZenCap which allows retail investors to invest in corporate loans. In contrast to the stock market, peer-to-business investors cannot receive a diversified portfolio at once but they have to invest in new loan campaigns continuously. We analyze the investment behavior using several statistical methods including OLS and logit regressions. In order to examine the risk-return profile of the portfolios we construct Value at Risk (VaR) measures and obtain the risk-adjusted return on capital (RAROC). We analyze changes in the RAROC by applying two-stage least squares (2SLS) estimations.

Four of the papers described above are published in academic journals and the last study is under review at the date of the submission of this thesis. Due to different style-requirements of the journals small formal differences in the presentation of the research papers may be present.

Chapter 2

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

This research project is joint work with Gregor Dorfleitner, Christopher Priberny, Stephanie Schuster, Johannes Stoiber, Ivan de Castro, and Julia Kammler. The paper has been published as: Gregor Dorfleitner, Christopher Priberny, Stephanie Schuster, Johannes Stoiber, Martina Weber, Ivan de Castro and Julia Kammler (2016), Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms, *Journal of Banking & Finance* 64, 169-187.

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.

Keywords peer-to-peer lending, soft information, funding probability, probability of default

JEL Classification G20, G32

2.1 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 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. (2016), 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.

2.2 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 and Hui, 2013; 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., 2016; Everett, 2015; Hildebrand et al., 2017; 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 Wei and Lin (2016) 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., 2016). 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; Faßbender, 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 Faßbender (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. (2016) 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.

2.3 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

[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.

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.

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.*

2.4 Data and methodology

2.4.1 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 Auxomey 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

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

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 2.4.4, 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 3,298 (Auxmoney) and 2,216 (Smava) observations.

2.4.2 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

[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.

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.

2.4.3 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 2.1 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 2.12 in the Appendix.

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.

Table 2.1 Description of variables.
Note: [a] indicates that variable is soley available for Auxmoney or Smava [s].

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:* Wise-clerk.
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 $\leq$ 10 days, Smava $\leq$ 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 2.12 in the Appendix. 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.

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 activities (*Leisure*) and business (*Business*). All keywords and the associated categories are displayed in Table 2.11 in the Appendix.

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., 2016) or gender information (e.g. Faßbender, 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*).

2.4.4 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 Table 2.2 and Table 2.3.

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

[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.

Table 2.2 Descriptive statistics of metric variables.
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 2.1. *Data sources:* Auxmoney, Smava, Datastream.

	DATA	N	MIN	Q25%	MEDIAN	MEAN	Q75%	MAX	SD
variables concerning both platforms									
Volume	AUX	76,945	1,000	1,500	3,000	5,030.07	6,700	30,350	5,054.36
	AUX, CGL	3,298	1,000	1,500	2,000	3,243.01	4,000	20,000	3,141.25
	SMA	10,423	500	3,250	6,250	8,995.32	12,000	50,000	7,967.97
	SMA, CGL	2,216	500	2,500	3,750	5,301.78	6,500	50,000	4,772.00
I	AUX	76,945	0.00	0.10	0.13	0.12	0.14	0.18	0.03
	AUX, CGL	3,298	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	2,216	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	3,298	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	2,208	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	3,298	−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	2,216	−0.23	−0.08	0.01	−0.01	0.04	0.18	0.11
#Lender	AUX, CGL	3,298	1	10	15	20.84	26	123	16.39
	SMA, CGL	2,216	1	6	9	12.34	16	115	10.96
#Words	AUX	76,945	0	13	34	55.94	70	8,441	83.52
	AUX, CGL	3,298	1	44	81	109.40	138	2,129	108.35
	SMA	10,423	0	19	26	41.43	50	531	43.62
	SMA, CGL	2,216	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	1,771	−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	2,216	−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	2,216	23	36	46	47.16	55	93	14.21

Table 2.3 Relative frequency distributions of categorical variables in percentage values.
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=3,298 (AUX) and N=2,216 (SMA). *KeyWord* is shown in absolute frequencies. The variables are defined in Table 2.1. *Data sources:* Auxmoney, Smava, Wiseclerk.

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	5,685	4,831	6,267	20,620	11,893	2,068	2,914	
	AUX, CGL	980	402	354	497	1,440	656	135	141	
	SMA	2,567	430	348	1,022	3,385	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		

looking at the interrelation between the number of lenders (#*Lender*) reported in Table 2.2 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 overcompensation 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

[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.

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.

Table 2.13 and Table 2.14 in the Appendix 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).

2.4.5 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 2.1. 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.

$$y^*_{1i} = \mathbf{m}'_i \boldsymbol{\delta} + \alpha y_{2i} + u_i \tag{2.1}$$

$$y_{2i} = \mathbf{m}'_i \boldsymbol{\gamma} + \pi z_i + e_i \tag{2.2}$$

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.

2.5 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.

2.5.1 Funding probability

Auxmoney

The first four columns in Table 2.4 show the results for the model specifications with *FGL* as an 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, 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

Table 2.4 Regression results concerning the funding probability on Auxmoney.
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 2.1.

	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	

Table 2.5 Regression results concerning the funding probability on Smava.
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 2.1.

	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	

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.

Smava

Table 2.5 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)*.

2.5.2 Probability of default

Table 2.6 and Table 2.7 show the results of the default probability analysis.

Auxmoney

The specifications AD.I to AD.IV in Table 2.6 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 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 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 1b (orthography)* and *Hypothesis 2b (description length)*. Both findings are consistent with the results of Iyer et al. (2016), 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

Table 2.6 Regression results concerning the default probability on Auxmoney.
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 2.1.

	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	−2,695.36	−2,727.55	−2,724.87	−2,734.85	
N	3,298	3,298	3,298	3,298	

Table 2.7 Regression results concerning the default probability on Smava.
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 2.1.

	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	2,208	2,216	2,216	2,208	

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 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.

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. (2016). 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 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.

[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.

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

2.5.3 Effects of control variables on funding and default probability

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

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., 2016). 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 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 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.

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

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.

2.5.4 Robustness checks

We perform several model variations and subsample regressions as robustness checks which are shown in Tables 2.8 and 2.9 for Auxmoney and Table 2.10 for Smava.

Table 2.8 Subsample regressions concerning the solvency indicators for Auxmoney.
Notes: AFR.I and ADR.I are subsample regressions with respective specifications to AF.IV and AD.IV (main results, already shown in Table 2.4 and 2.6) 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 Table 2.4 and 2.6) 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 2.1.

	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		1,311		1,987	

Table 2.9 Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Auxmoney.
Notes: AFR and ADR are regressions with respective specifications to AF.IV and AD.IV (main results, already shown in Table 2.4 and 2.6) using the residual interest rate. ADDQ is a regression with respect to specification AD.IV (main results, already shown in Table 2.6) 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 2.1.

	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

Table 2.10 Robustness checks concerning the residual interest rate and the Wiseclerk data quality for Smava.
Notes: SFR and SDR are regressions with respective specifications to SF.IV and SD.IV (main results, already shown in Table 2.5 and 2.7) using the residual interest rate. SDDQ is a regression with respect to specification SD.IV (main results, already shown in Table 2.7) 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 2.1.

	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	2,208	1,007

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 2.9 for Auxmoney and in Table 2.10 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.

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.

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 2.8. 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.

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 2.9, 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 2.10, Specification SDDQ). Altogether, there is no evidence in favor of an unreported-default bias.

2.5.5 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

[10] Again, with the above conservative calculus (see Section 2.4.1) 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.

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.

2.6 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 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.

2.7 Appendix

Table 2.11 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 2.12 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 2.13 Pairwise Bravais-Pearson correlation coefficients among the explanatory variables concerning the Auxmoney data set.
Notes: The symbol * expresses significance at the 5% level. The variables are defined in Table 2.1.

	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

Table 2.14 Pairwise Bravais-Pearson correlation coefficients among the explanatory variables concerning the Smava data set.
Notes: The symbol * expresses significance at the 5% level. The variables are defined in Table 2.1.

	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*	0.03*	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.	15.	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*	0.00	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. *TurnYear*	−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*	0.22*	0.29*	0.19*	0.22*	0.11*	0.23*	0.23*	0.32*
31. (*#Words*)^2	0.01	0.08*	0.14*	0.21*	0.14*	0.15*	0.08*	0.16*	0.17*	0.19*

	22.	23.	24.	25.	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. *TurnYear*	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

Chapter 3

German Mittelstand bonds: Yield spreads and liquidity

This research project is joint work with Sebastian Utz and Maximilian Wimmer. The paper has been published as: Sebastian Utz, Maximilian Wimmer and Martina Weber (2017), German Mittelstand Bonds: Yield Spreads and Liquidity, *Journal of Business Economics* 86(1), 103-129.

Abstract We estimate a cross-sectional model of the yield spreads of German Mittelstand bonds as a function of liquidity measures as well as a number of variables that control for both the characteristics of the issuing firm and the bond characteristics. Our results show a significant positive effect of illiquidity on the yield spread, which persists after controlling for the risk of the bond. Economically, the size of the liquidity premium of Mittelstand bonds is approximately twice the size of speculative grade US corporate bonds. Our findings are robust to different measures of liquidity and potential endogeneity biases.

Keywords German Mittelstand Bond, Liquidity, Yield Spread, SME, Minibonds

JEL Classification G12, G32

3.1 Introduction

The German Mittelstand is often hailed as the powerhouse of the German economy. It is characterized by being mostly medium-sized, family-owned, and family-run companies, which traditionally lend through relationship banking to cover their financing needs. However, with the phase-in of the Basel II regulations, financing via relationship banking has become more restrictive for many Mittelstand firms, as the new regulations enforce a mandatory rating for all issued loans (Schindele and Szczesny, 2015). Launched in 2010, the possibility to issue Mittelstand bonds with volumes of less than 100 million Euro on the capital market is a remedy for the Mittelstand to close this financing gap. Yet, the observed yield spreads of Mittelstand bonds are high. Longstaff et al. (2005) argue that default risk is the key determinant for the yield spread of corporate bonds over government benchmarks. Notwithstanding, they also find that default risk cannot explain the entire variation of the spread. Indeed, market frictions such as liquidity costs also play an important role (Fisher, 1959; Chen et al., 2007b). The size of the liquidity premium, however, depends on the the credit rating of the issuing firm, i.e. less solvent firms show higher liquidity premia. Since the solvency of Mittelstand firms is often unclear, we empirically examine the size of the liquidity premium that is priced in the spread of Mittelstand bonds. We find that illiquidity is indeed significantly associated with the yield spread after controlling for default risk. Economically, the size of the liquidity premium of Mittelstand bonds is approximately twice the size of speculative grade US corporate bonds. Our results are robust to different measures of liquidity and a potential endogeneity bias.

Mittelstand bonds are a young financing vehicle enabling small and mid cap firms to directly tap capital markets. Since its launch in 2010 the market for Mittelstand bonds has developed rapidly. Five German stock exchanges[1] created segments for Mittelstand bonds and more than 120 bonds with a total volume exceeding 6 billion Euros have been issued in the period to July 2015. However, studies such as Kammler and Röder (2013) report a total loss of capital of 3.71% on the Stuttgart Stock Exchange for Mittelstand bonds by the end of 2012. After the default of several Mittelstand bonds, two stock exchanges (Stuttgart and Dusseldorf) decided to close their segments for Mittelstand bonds. By contrast, the remaining stock exchanges successfully established their Mittelstand segments. For instance, the Frankfurt Stock Exchange reports four new bond emissions in the first half of 2015.

For the analysis of the relationship of bond-specific liquidity and the yield spread, we use two different liquidity estimates, namely the bid–ask spread and the LOT liquidity estimate based on

[1] Namely Stuttgart (bondm), Frankfurt (Entry Standard), Dusseldorf (Der Mittelstandsmarkt), Munich (m:access) and Hannover/Hamburg (Mittelstandsboerse).

Lesmond et al. (1999). While the bid–ask spread is a canonical measure of liquidity (see e.g. Brandt and Kavajecz, 2004; Fleming, 2003), data to calculate that spread is not available for all bonds. Therefore, we use the LOT liquidity estimate as an alternative measure of liquidity. The LOT liquidity estimate reflects the round-trip liquidity costs based on the frequency of zero returns. To analyze the yield spread determinants, we apply fixed effects panel regressions with clustered standard errors and regress the yield spread on the liquidity estimates and bond-specific, firm-specific, and macroeconomic variables. To control for potential endogeneity of the liquidity measures and the credit rating, we apply a simultaneous equation model performing a three-stage least squares estimation technique.

Analyzing a comprehensive sample of 92 Mittelstand bonds, we find that investors demand a higher liquidity premium for more illiquid Mittelstand bonds. Both liquidity measures are significantly positively related to the yield spread in our regressions. A 1% increase in the bid–ask spread leads to an incremental increase in the yield spread in the range of 3.19% to 6.41%. The predicted increase of the yield spread is slightly smaller for a similar increase in the LOT liquidity estimate. Since the within R^2 ranges between 58% and 82%, our models provide high explanatory power for the variation of the yield spread of Mittelstand bonds. Therefore, we confirm that default risk accounts for only part of the variation of the yield spread. Bond liquidity is another key determinant of the spread, which is especially pronounced for Mittelstand bonds.

Our paper has important implications for financial managers of Mittelstand firms. While the observed high yield spreads are commonly perceived as a proxy for default risk, which is exogenous for the firms, we highlight that a significant part of the yield spread is indeed associated to illiquidity. Illiquidity, in turn, results as a consequence of trading costs, search problems, private information, and inventory risk of market makers (Bagehot, 1971; Amihud and Mendelson, 1980) and is therefore, at least partly, endogenous for firms. Thus, by reducing the sources of illiquidity, Mittelstand firms can decrease the yield spreads of their issued bonds and thus reduce their effective cost of capital.

The remainder of this paper is organized as follows. Section 3.2 outlines the theory of this paper and Section 3.3 explains the institutional setting of the Mittelstand bond market, our data, and our methodological framework. We present and discuss our results in Section 3.4. Section 3.5 concludes.

3.2 Background

Due to their opportunity costs, investors expect to be compensated for lending money. On the one hand, they expect to earn the risk-free interest rate as compensation for the time value of money. Moreover, for risk-bearing investments, investors expect to earn an additional return—the risk premium—as compensation for the risk of their investment. The *yield spread* of a corporate bond is the difference between the bond's yield to maturity and the yield to maturity of a benchmark government bond that has exactly the same maturity and currency. Since such a benchmark government bond rarely exists, the benchmark yield is typically interpolated using a benchmark government bond with a lower maturity and a benchmark government bond with a higher maturity. As government bonds are considered to be risk-free, the yield spread measures the risk premium for the investment in a corporate bond.

While default risk, i.e. the risk that the principal of the bond is not repaid in full at maturity, certainly is a crucial determinant of the yield spreads, default risk cannot explain the full variation of corporate bond yield spreads. For instance, Fisher (1959) analyzes the determinants of corporate bond yield spreads for the years 1927, 1932, 1937, 1949, and 1953. He finds that yield spreads are positively influenced by default risk and negatively influenced by marketability—a synonym for liquidity. More recently, Chen et al. (2007b) confirm the existence of a liquidity premium using a comprehensive sample of US corporate bonds over the period from 1995 to 2003 and find that the liquidity premium is higher for speculative grade bonds compared to investment grade bonds.

Generally, the term *liquidity* describes the ease of trading a security (Amihud et al., 2005). In frictionless markets, every security can be traded at no cost all of the time. Therefore, in standard asset pricing theories which are based on the assumption of frictionless markets (e.g. Cochrane, 2001; Duffie, 1996), liquidity does not affect asset prices. However, real markets are far from being frictionless. There are four market imperfections that induce illiquidity to the markets[2]: Exogenous trading costs, search problems, adverse selection due to private information, and inventory risk for market makers. Trading costs and search problems directly adversely influence liquidity by reducing the number of noise traders on the markets. Private information induces the existence of informed and uninformed traders. Since market makers generally lose from trades with informed traders, they need to charge a certain bid–ask spread to gain from trades with uninformed (noise) traders (Bagehot, 1971). Finally, since not all traders are present at all times, market makers need to build up an inventory in order to provide immediate trading to any

[2]We refer to Amihud et al. (2005) for a detailed overview of the sources of illiquidity.

trader. Such an inventory inhibits a price risk which the market makers have to hold and wish to be compensated for by higher bid–ask spreads (Amihud and Mendelson, 1980; Ho and Stoll, 1981).

Given this theoretical framework, we hypothesize that liquidity influences the yield spreads of Mittelstand bonds, too. Due to the relatively small size of Mittelstand firms, we expect a relatively large liquidity premium as private information is adversely related to firm size (Diamond and Verrecchia, 1991; Vega, 2006). To gain evidence on this hypothesis, we continue our paper with an empirical study of a comprehensive sample of Mittelstand bonds that disentangles the influences of default risk and liquidity on the yield spreads.

3.3 Data and methodology

In this chapter, we commence with a brief overview on the development of the Mittelstand bond market and describe our sample of Mittelstand bonds. Afterward, we introduce the two liquidity measures employed in our study in detail.

3.3.1 German Mittelstand bonds

The application of the Basel II rules on all banks in the European Union in January 2007 introduced a mandatory rating for each firm applying for a loan. As a result, the interest rates offered to low-rated firms have increased significantly because of higher equity requirements for such loans (Müller et al., 2011; Schindele and Szczesny, 2015). Mittelstand firms are affected in particular by these adverse conditions due to their relatively low equity ratios compared to large firms (Feiler and Kirstein, 2014). The Basel III accords continue to pursue the aim of the Basel II capital requirements to increase the resilience of banks during crises. The relationship bank system, which was an essential backbone for German Mittelstand firms, is facing serious difficulties in offering reasonable loan conditions for poorly or non-rated Mittelstand firms. Therefore, the Mittelstand needs an alternative source of financing. Since Mittelstand firms are often family-run, they are reluctant to tap equity markets in order to not dilute their ownership and control rights.

Common stock exchanges, so far, allowed only bond emissions with a volume of at least 100 million Euros, which exceeds the required amount of capital for small or mid cap firms in general. As long as the relationship bank system runs properly, small and mid cap firms can avoid costly public bond issues. In the light of the new requirements stemming from the developments according to bank regulations, small and mid cap firms have to reconsider this method of financing. Instead of solely relying on relationship bank loans, they need to tap other debt financing instruments to be able to invest and successfully compete in an international market environment.

Recognizing this funding gap, the Stuttgart Stock Exchange was the first German stock exchange to create bondm, a segment that enables small and mid cap firms to access the public capital market in 2010. Four other German stock exchanges—namely Frankfurt (Entry Standard), Dusseldorf (Der Mittelstandsmarkt), Munich (m:access), and Hannover/Hamburg (Mittelstandsboerse)—followed suit. Yet, the requirements for bond emissions vary considerably between the exchanges. While in Stuttgart, Dusseldorf, and Munich a minimum volume of 25 million Euros or 10 million Euros respectively is obligatory, Hannover/Hamburg and Frankfurt accept any size of emission. Furthermore, a strict rating obligation only exists in Dusseldorf and Munich. The Frankfurt and Stuttgart exchanges accept emissions without ratings for listed companies while the Hannover/Hamburg exchange generally waives the rating obligation. Despite this heterogeneous institutional setting, Mittelstand bonds usually have certain common characteristics. Mostly, these bonds have an issue volume of 15 to 150 million Euros, a maturity of 3 to 10 years, and a fixed coupon.

In our paper we define Mittelstand bonds as corporate bonds that are or were traded in the respective segments on any one of the five stock exchanges. We hand-collect the International Securities Identification Numbers (ISINs) of the Mittelstand bonds from the homepages of the five stock exchanges to form our data set. In sum, we derive a data set of 120 bonds in the period from November 24, 2010, to July 15, 2015, with a total issue volume of more than 6 billion Euros. Since the introduction of Mittelstand bonds, several issuers have declared insolvency. Analyzing the bondm segment up to December 2012, Kammler and Röder (2013) find a total loss of capital of 3.71% and a negative internal rate of return of −3.04% for investments into Mittelstand bonds. Schöning (2014) also uses bondm data to calculate the risk-adjusted interest rate for Mittelstand bonds. He finds that the coupons of many bonds are well below the risk-adjusted value. In the light of this development, the stock exchanges of Stuttgart and Dusseldorf decided to shut down their segments for Mittelstand bonds. By contrast, Frankfurt's Entry Standard continues to be successful. In the first half of 2015 four new bonds with a total volume of 220 million Euros were issued.

3.3.2 Yield spreads and corporate information

We use the ISINs of our sample of 120 Mittelstand bonds to match bond and firm-characteristic data from four different sources. Daily data on the bid–ask spread, the clean price, and the yield spread are obtained from Thomson Reuters Datastream. For our regression analysis we use the yearly average of the daily yield spreads. Bond-specific and macroeconomic factors are crucial for explaining the yield spread and the bond liquidity (Elton et al., 2001; Chakravarty and Sarkar, 1999; Campbell and Taksler, 2003). Therefore, we also download the time to maturity, the age of the bond, the 1-year yield on German Bunds, and the term slope (difference in yields of 10-year and 2-year German Bunds) from Datastream. Additionally, we estimate the bond volatility by calculating the yearly standard deviation of the clean prices.

Default risk is another important bond characteristic (Longstaff et al., 2005). However, Mittelstand bonds are usually not rated by any of the three leading rating agencies but by smaller German agencies instead. Hence, we collect the credit ratings from rating reports when they are accessible on http://anleihen-finder.de, a website that provides data for most Mittelstand bonds. When available, we use the bond rating, otherwise the credit rating of the issuing firm. From the credit ratings, we construct the variable *Rating Scale* which codes a numeric value to each rating class ranging from 1 for A (the best rating in our sample) to 15 for D (default). Furthermore, we double check the ratings of bonds with a clean price below 80% at any point during our sample period. We find that the issuers of 24 bonds in our sample have bankrupted throughout the observation period. We use the day they declared insolvency to manually change the respective ratings to D.

However, as there is no general rating obligation on all five stock exchanges, not all firms and bonds are rated. Since credit ratings are mostly derived from financial ratios, accounting data can provide similar insights into the default risk and the solvency of a firm. In particular, we consider interest coverage, operating income to sales, long-term debt to assets, and debt to capital as firm-specific control variables (Campbell and Taksler, 2003). We define interest coverage as EBIT plus interest divided by interest. Accounting data to calculate these performance measures is obtained from Bureau van Dijk's Dafne, a database with financial information for more than one million German companies. In the case that Dafne data was not available (i.e. for non-German companies) we use Bureau van Dijk's Amadeus (via WRDS) as a second database for financial information. To avoid a potential forward-looking bias, we lag these ratios by one year for our further analysis.

Furthermore, we exclude all bonds that defaulted during our sample period for our regression analysis to avoid a potential bias due to the non-linear increase in the yield spread of firms that are close to default. We also exclude one bond with obviously incorrect clean prices in Datastream. We finally disregard bonds for which no yield spread is available on Datastream and callable bonds after the announcement of the exercise of the call since the clean price usually equals the call price after the announcement. In sum, our final sample comprises 92 German Mittelstand bonds. We list all bonds of our final sample and the main bond characteristics in Table 3.8 in Appendix 3.6.1.

3.3.3 Bid–ask spread

As it describes the round-trip transaction costs for an immediate transaction, the bid–ask spread is a canonical and commonly used measure of liquidity. We obtain data on daily composite bid and ask prices from Datastream. These composite prices are calculated as the average of all available contributors' quotes. The (relative) bid–ask spread is the difference between the ask and bid prices divided by the average of both prices. Yet, data to calculate this spread is not always available. In particular in the beginning of our observation period, data on bid and ask quotes is rare, since the coverage of ask prices in Datastream starts for most bonds only in October 2013. For each bond, we estimate the average yearly bid–ask spread by calculating the mean of all daily spreads, if at least one bid–ask spread is available in the respective bond-year.

3.3.4 LOT liquidity estimate

Our second measure of liquidity is based on the limited dependent variable model of Tobin (1958) and Rosett (1959). Lesmond et al. (1999) use this model to estimate transaction costs based on the frequency of zero returns of equity. In this paper, we refer to this measure as the LOT liquidity estimate and calculate it in the version of Chen et al. (2007b) for corporate bonds. In contrast to bid–ask spreads that are only available for a limited number of firms due to poor data availability, the LOT liquidity estimate requires only the time series of daily returns to endogenously estimate liquidity in terms of transaction costs on a firm level. In a nutshell, the LOT liquidity estimate models illiquidity through the incidence of zero returns. In the presence of transaction costs, not all information will be immediately priced. Only if the value of the information exceeds the costs of trading, will a marginal investor trade on it. On the other hand, if the value of the information is below the costs of trading, a marginal investor will refrain from trading, causing a zero return. The LOT liquidity estimate is defined as the difference between the buy-side and sell-side transaction costs for a marginal investor. It is estimated by modeling

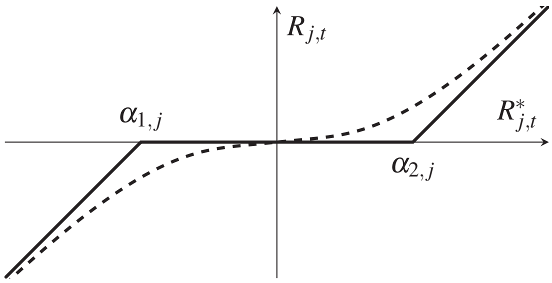


Figure 3.1 LOT liquidity estimate model. This graph details the relationship between the 'true' return $R^*_{j,t}$ (on the x-axis) and the measured return $R_{j,t}$ (on the y-axis). The bold solid line depicts the case of perfect information, the dashed line depicts the measured *expected* return that the investors would price given uncertainty about the true return.

the return generating process of a bond and comparing the thereby computed 'true' returns with observed bond returns. In particular, it estimates the buy-side and sell-side transaction costs by observing the thresholds of the 'true' returns that lead to a trade, i.e. a non-zero observed return.

Liquidity costs cause assets to have lower prices in order to compensate investors for illiquidity (Amihud and Mendelson, 1986). In the case of bonds, the difference between the observed value on the market and the intrinsic 'true' value is the liquidity premium (Amihud and Mendelson, 1986, 1987). Figure 3.1 illustrates the liquidity effects on bond returns. The bold line represents the case of perfect information. In this instance, a marginal trader will only buy (sell) a bond j at time t if she receives information about the bond that has a higher value than the buy-side costs $\alpha_{2,j}$ (sell-side costs $\alpha_{1,j}$). Therefore, the observed return $R_{j,t}$ is zero when the value of the new information, i.e. the 'true' return $R^*_{j,t}$, is between $\alpha_{1,j}$ and $\alpha_{2,j}$. Only if the 'true' return $R^*_{j,t}$ exceeds the buy-side costs $\alpha_{2,j}$ (sell-side costs $\alpha_{1,j}$), does a marginal trader start trading and we observe a return $R_{j,t}$, which is the 'true' return $R^*_{j,t}$ reduced by the buy-side costs $\alpha_{2,j}$ (sell-side costs $\alpha_{1,j}$). Therefore, in the case of perfect information, we have the following relationship of $R_{j,t}$ and $R^*_{j,t}$:

$$
\begin{aligned}
R_{j,t} &= R^*_{j,t} - \alpha_{1,j} && \text{if } R^*_{j,t} < \alpha_{1,j} && \text{and} && \alpha_{1,j} < 0 \\
R_{j,t} &= 0 && \text{if } \alpha_{1,j} \leq R^*_{j,t} \leq \alpha_{2,j} && && \\
R_{j,t} &= R^*_{j,t} - \alpha_{2,j} && \text{if } R^*_{j,t} > \alpha_{2,j} && \text{and} && \alpha_{2,j} > 0.
\end{aligned}
\tag{3.1}
$$

To compute the liquidity cost threshold for each bond, we need a model for the 'true' return $R^*_{j,t}$.

Following the methodology of Chen et al. (2007b), we use a two-factor model to estimate the 'true' return of corporate bonds. The first factor is the long-term interest rate and the second factor the equity market return. This model accounts for the fact that corporate bonds are essentially a hybrid between a risk-free bond and equity. In order to obtain stable estimation coefficients, the risk coefficients are scaled by the duration D of the respective bond (see Jarrow, 1978). In particular, our two-factor model for the 'true' returns is

$$R^*_{j,t} = \beta_{j,1} D_{j,t} \cdot \Delta R_{f,t} + \beta_{j,2} D_{j,t} \cdot \Delta \text{DAX}_t + \varepsilon_{j,t}, \tag{3.2}$$

where $\Delta R_{f,t}$ is the daily change in the 10-year German Bunds rate and ΔDAX_t is the daily return on the DAX 30 composite stock index.

Since the error term $\varepsilon_{j,t}$ in model (3.2) introduces uncertainty about the 'true' return, the expected return that investors price given the uncertainty about the 'true' return slightly differs from equation (3.1). Rosett (1959) models the locus of this curve. The dashed line in Figure 3.1 illustrates the relationship of the measured return and the measured expected return in the case of uncertainty.

With σ_j being the (unknown) standard deviation of the error term $\varepsilon_{j,t}$, we estimate the liquidity cost thresholds $\alpha_{1,j}$ and $\alpha_{2,j}$ of each bond j in year t by maximizing the logarithm of the likelihood function $L(\alpha_{1,j}, \alpha_{2,j}, \beta_{j,1}, \beta_{j,2}, \sigma_j \mid R_{j,t}, \Delta\text{DAX}_t)$

$$\begin{aligned}
\max_{\alpha_{1,j},\alpha_{2,j},\beta_{j,1},\beta_{j,2},\sigma_j} \ln L = & \sum_{t \in \mathscr{R}_1} \ln \frac{1}{\sqrt{2\pi\sigma_j^2}} \\
& - \sum_{t \in \mathscr{R}_1} \frac{1}{2\sigma_j^2} (R_{j,t} + \alpha_{1,j} - \beta_{j,1} D_{j,t} \cdot \Delta R_{f,t} - \beta_{2,t} D_{j,t} \cdot \Delta\text{DAX}_t)^2 \\
& + \sum_{t \in \mathscr{R}_2} \ln \frac{1}{\sqrt{2\pi\sigma_j^2}} \\
& - \sum_{t \in \mathscr{R}_2} \frac{1}{2\sigma_j^2} (R_{j,t} + \alpha_{2,j} - \beta_{j,1} D_{j,t} \cdot \Delta R_{f,t} - \beta_{2,t} D_{j,t} \cdot \Delta\text{DAX}_t)^2 \\
& + \sum_{t \in \mathscr{R}_0} \ln(\Phi_{2,j} - \Phi_{1,j}),
\end{aligned}$$

where $\mathscr{R}_1$ denotes the set of days with negative measured returns $R_{j,t}$, $\mathscr{R}_2$ denotes the set of days with positive measured returns $R_{j,t}$, and $\mathscr{R}_0$ denotes the set of days with zero returns. The term $\Phi_{i,j}$ represents the cumulative distribution function of the standard normal distribution for each bond-year evaluated at $(\alpha_{i,j} - \beta_{j,1} D_{j,t} \cdot \Delta R_{f,t} - \beta_{j,2} D_{j,t} \cdot \Delta\text{DAX}_t)/\sigma_j$. For purposes of liquidity estimation, the critical parameters of the limited dependent variable model are in the intercept terms, $\alpha_{1,j}$ and $\alpha_{2,j}$. We define the LOT liquidity estimate for bond j

$$\text{LOT}_j = \alpha_{2,j} - \alpha_{1,j}$$

by the difference of the buy-side and the sell-side cost estimates per year.

Table 3.1 This table reports upon the number of bonds, average costs of sell trades (α_{1j}), buy trades (α_{2j}), LOT liquidity estimate ($\alpha_{2j} - \alpha_{1j}$), and the t-statistics testing for zero LOT separated by year.

Year	# Obs	$\hat{\alpha}_1$ (%)	$\hat{\alpha}_2$ (%)	LOT (%)	t(LOT)
2010	3	−0.093335	0.178805	0.272140	1.977116
2011	25	−0.338217	0.528104	0.866321	2.441339
2012	51	−0.113481	0.221435	0.334916	3.315454
2013	78	−0.273492	0.400965	0.674457	5.948445
2014	88	−0.712876	0.749138	1.462014	5.244218
2015	88	−0.859537	0.796175	1.655712	5.384455

Daily clean prices, duration, DAX index, and Bunds returns are obtained from Datastream. Table 3.1 reports upon the number of bonds, the average sell-side and buy-side cost estimates, the average LOT liquidity measure, and the average t-statistics testing for zero LOT per year. The LOT liquidity estimates are significantly different from zero in all years. The Spearman correlation of the LOT liquidity measure and the bid–ask spreads is 65.7% over all bond-year observations.

Notice that the LOT liquidity measure accounts for additional information from the return generating process besides zero returns, such as commission costs, opportunity costs, and price impact costs. Potential limitations of the LOT model occur in the case of no or too many zero returns (more than 85%) within one year. In our sample the average yearly percentage of zero returns of the cross-section of all bonds is 18.7%. Furthermore, our data contains at least one zero return observation in each bond-year.

Table 3.9 in Appendix 3.6.2 presents a summary of all variables, their detailed meanings, and their respective data sources.

3.4 Results

Before performing our main regression analysis on the yield spread determinants of Mittelstand bonds, we commence this chapter presenting summary statistics of our sample and several tests regarding the consistency of our two liquidity measures.

3.4.1 Summary statistics

Table 3.2 reports upon summary statistics for the time-invariant bond characteristics and accounting data of the issuing firms of the Mittelstand bonds in our sample. The average issue volume equals 46 million Euros and is small compared to common corporate bonds. Furthermore, the bonds pay relatively high interest with an average coupon of 7.23%. However, the size of the coupons varies noticeably and ranges from 2.00% (*DF Deutsche Forfait AG*) to 11.5% (*Air Berlin AG*). In terms of maturity the bonds do not show much variation. A mean maturity of 5.21 years and a standard deviation of 0.08 years suggest that the bonds are relatively homogenous in this property. Additionally, issuing firms' accounting data at the emission date of the bonds is presented. With −20.6 million Euros in 2011, *Air Berlin AG* has the lowest EBIT in our sample. By contrast, *Porr AG* is highly profitable with an EBIT of more than 88 million Euros. Taking sales and total assets into account, the figures indicate that the firms in our sample differ considerably in size and in profitability. The same pattern can be observed with respect to leverage. While some firms have a very low debt to assets ratio (*Peach Property*: 0.01), other companies are deeply indebted (*FC Schalke 04*: 1.33). Yet, in the case of *FC Schalke 04* the extremely high leverage mostly results from discretionary accounting policies such as the non-capitalization of the fair value of the squad.

Further summary statistics on time-variant measures grouped by year are presented in Table 3.3. The average yield spread and both liquidity measures—the bid–ask spread and the LOT liquidity estimate—tend to increase over the sample period. The average bid–ask spread is particularly high in 2011 (9.24%). However, data to calculate the spread is scarce at the beginning of the sample period and thus there is only one firm with valid bid–ask spread data available in 2010 and 2011. Along with the yield spread and the liquidity measures the rating scale increases over time. This is a first indication that higher liquidity costs are reflected in higher yield spreads.

3.4.2 Bid–ask spread tests

The correlation of 65.7% between the bid–ask spread and the LOT liquidity estimate indicates a relatively strong dependence between our two measures of liquidity. To confirm the consistency of these liquidity estimates we perform further tests. We regress the bid–ask spread on the LOT liquidity estimate and control variables.

Table 3.2 This table reports upon descriptive statistics of the Mittelstand bonds in the year of the emission of the bonds. We report the number of observations (# Obs), the mean value of all bonds, the standard deviation (sd), skewness, kurtosis, minimum value (min), first quartile (q25), median value, third quartile (q75), and maximum value (max).

	# Obs	mean	sd	skewness	kurtosis	min	q25	median	q75	max
Volume (in thousands)	92	462,08.25	46,962.93	2.528	9.072	3455	20,000	30,000	50,000	225,000
Coupon (%)	92	7.27	1.05	−1.685	10.087	2.00	6.75	7.25	8.00	11.50
Maturity (years)	92	5.21	0.08	2.791	13.987	3.00	5.00	5.00	5.00	10.00
EBIT (in thousands)	67	1,255.38	41,194.29	−3.293	18.200	−206,336	−996.55	2,886.30	10,291.74	88,026
Total Assets (in thousands)	72	307,991.40	538,016.50	2.594	8.531	373.79	43,130.74	106,130.40	259,354.20	2,296,470
Sales (in thousands)	61	259,982	689,293.60	5.158	30.950	1,534	31,512.55	79,037	174,968	4,663,798
Debt to assets	64	0.47	0.22	0.735	5.934	0.01	0.36	0.47	0.57	1.33

Table 3.3 This table reports upon summary statistics of the Mittelstand bonds separated by year. Yield spread refers to the difference of a bond yield and an equivalent government benchmark. Bid–ask spread is a proportional spread as described in Section 3.3.3. LOT equals the liquidity estimate as described in Section 3.3.4. Rating scale assigns a numeric value to each rating class starting with 1 for A up to 15 for D. Yield spreads are denoted in basis points (bp). # Obs denotes the number of observations.

Year		2010	2011	2012	2013	2014	2015
Yield Spread (bp)	Mean	474.519	621.519	674.059	687.959	996.540	1236.411
	# Obs	3	24	50	76	86	86
Bid–Ask Spread (%)	Mean	0.300	9.236	3.134	1.312	1.968	2.367
	# Obs	1	1	5	67	85	85
LOT (%)	Mean	0.272	0.895	0.340	0.674	1.462	1.656
	# Obs	3	24	50	78	88	88
Rating Scale	Mean		5.154	5.493	6.092	6.572	6.723
	# Obs	0	13	39	61	70	71

Analyzing stock data from 1997 and 1998, Stoll (2000) finds expanding bid–ask spreads with increasing volatility of stock returns. Furthermore, Brandt and Kavajecz (2004) emphasize the importance of bond volatility in explaining liquidity costs in the US Treasury market. Thus, we include bond volatility as a control variable. Chakravarty and Sarkar (1999) use further bond characteristics to identify determinants of the bid–ask spreads of corporate, municipal, and government bonds. They argue that the age of the bond and credit risk are positively related to the spread. Sarig and Warga (1989) argue that bonds become less liquid with time and therefore use the age of the bond as a measure of liquidity. Their results support the hypothesis of a positive relationship between the age of the bond and the yield spread. Hence, we also include the age of the bond as bond-specific control and use the variable *Rating Scale* to capture the effect of credit risk.

We analyze the bid–ask spread by a fixed effects panel regression as follows:

$$\begin{aligned} \textit{Bid–Ask Spread}_{i,t} = \; & \eta_0 + \eta_1 LOT_{i,t} + \eta_2 \textit{Bond Volatility}_{i,t} \\ & + \eta_3 Age_{i,t} + \eta_4 \textit{Rating Scale}_{i,t} + \varepsilon_t, \end{aligned}$$

where the subscript i,t denotes bond i in year t. We first regress the bid–ask spread on the LOT liquidity estimate only and second on the LOT liquidity estimate including the control variables. The results are reported in the first two columns of Table 3.4.

Table 3.4 This table reports upon liquidity measure tests. The dependent variable is the bid–ask spread in the first and the second model and the yield spread in the third and the fourth model. We apply fixed effects panel regressions and cluster the standard errors at bond level. (B) indicates that we use the bid–ask spread and (L) the LOT liquidity estimate as explanatory liquidity measure. The absolute value of t-statistics are shown in parenthesis. *, **, *** denote significance at a 10%, 5%, and 1% level, respectively.

	Bid–Ask Spread		Yield Spread	
	(L)	(L)	(B)	(L)
Bid–Ask Spread			269.24**	
			(2.45)	
LOT	0.56***	0.51**		197.53***
	(3.23)	(2.06)		(3.07)
Bond Volatility		0.09***		
		(3.38)		
Age of the bond		0.22**		
		(2.03)		
Rating Scale		0.14		
		(0.93)		
Constant	1.38***	0.93	444.90**	686.13***
	(7.49)	(0.97)	(2.04)	(9.60)
# Obs	244	192	239	325
F-Statistic	10.42	6.65	5.98	9.45
Within R^2	0.55	0.68	0.55	0.43

The first model suggests a highly significant positive relationship between both liquidity measures. According to the within R^2 the LOT liquidity estimate explains 55% of the variation of the bid–ask spread. This result is robust to adding control variables in the next model. In line with the above literature, higher bond volatility and higher age of the bond is associated with higher bid–ask spreads. However, the rating is insignificant in our sample.

3.4.3 Yield spread determinants of Mittelstand bonds

Having confirmed the consistency of our liquidity measures, we now move on with our main analysis—the examination of whether illiquidity in fact explains part of the yield spread variation in our sample.

To gain more preliminary insight into the relationship of the yield spread and our liquidity measures, we directly regress the yield spread on the bid–ask spread and the LOT liquidity estimate, respectively. The results are reported in the last two columns of Table 3.4. The coefficients of both liquidity estimates are positive and significant at a 1% level. The regressions including the bid–ask spread and the LOT liquidity estimate show a within R^2 of 55% and 43%, respectively, and thus, a high explanatory power regarding the variation of the yield spreads. This suggests that higher liquidity costs are indeed associated with higher yield spreads.

However, these preliminary findings neglect that there are other determinants for the yield spread that might affect the outcome of the regressions. In order to add rigor to our results, we include an array of bond-specific, firm-specific, and macroeconomic control variables that are other well-documented determinants of yield spreads.

Default risk is the most prominent determinant of the yield spread. Longstaff et al. (2005) analyze a comprehensive data set on credit default swaps and corresponding bond price data and point out that default risk accounts for the majority of the yield spread. Depending on the credit rating, between 51% and 83% of the yield spread can be explained by default risk. Hence, we add the variable *Rating Scale* to capture this effect in our regression. Yet, approximately 23% of our bonds are not rated. Therefore, we include accounting ratios to measure the effect of the default risk for these bonds, too. Campbell and Taksler (2003) argue that while high values of the interest coverage and the income to sales ratio suggest healthy companies, the opposite is true for the two other accounting ratios. Long-term debt to assets and debt to capital describe the leverage of a company. Since highly leveraged firms are more likely to default, we expect the former two accounting variables to be negatively and the latter two to be positively related to the yield spread. Complete accounting ratios are available for only 57% of our observations, however. Nevertheless, we can increase our sample, as there are 44 bond-years without rating but with accounting data.

Campbell and Taksler (2003) document a positive relationship between the time to maturity and the yield spread for investment grade bonds. Chen et al. (2007b) confirm this effect for investment grade bonds. Yet they find the opposite effect for speculative grade bonds. To control for this potential influence we include time to maturity as a control variable in our regression analysis.

Furthermore, the general economic growth plays an important role. Longstaff and Schwartz (1995) argue that increases in the spot rates cause a steeper risk-neutral drift term in the firm value process. Therefore, the probability of default of the firm decreases (see e.g. Merton, 1974) and thus the yield spreads decrease, too. Hence, we add the rate on 1-year German Bunds, our proxy for the risk-free interest rate, as a control variable and expect it to be negatively associated with the yield spread. Collin-Dufresne et al. (2001) argue that the term structure of the yield curve has an effect on the yield spread as well. A decreasing term slope indicates an expected weaker economy and therefore lower recovery rates. In turn, we expect this to lead to higher yield spreads. Thus, we include the term slope as an additional control variable.

We specify our general regression model as follows:

$$\begin{aligned}
\textit{Yield Spread}_{i,t} = \eta_0 &+ \eta_1 \textit{Liquidity}_{i,t} + \eta_2 \textit{Maturity}_{i,t} + \eta_3 \textit{Government Bond}_{i,t} \\
&+ \eta_4 \textit{Term Slope}_{i,t} + \eta_5 \textit{Rating Scale}_{i,t} + \eta_6 \textit{Income/Sales}_{i,t-1} \\
&+ \eta_7 \textit{Debt/Assets}_{i,t-1} + \eta_8 \textit{Interest Coverage}_{i,t-1} \\
&+ \eta_9 \textit{Debt/Capital}_{i,t-1} + \varepsilon_t,
\end{aligned}$$

where the subscript i,t denotes bond i in year t and *Liquidity* refers to either the bid–ask spread or the LOT liquidity estimate, respectively.

We apply three different regression models for each liquidity estimate. Model 1 includes credit rating but not accounting data. Model 2 includes both credit rating and accounting data. Model 3 includes accounting data but not credit rating. Considering the availability of rating and accounting data in our overall sample, Model 1 maintains the largest sample whereas Model 2 has the smallest sample. We run each model for our two liquidity measure specifications, bid–ask spread and LOT liquidity estimate. Using the LOT liquidity estimate maintains larger sample sizes compared to the bid–ask spread due to better data availability. The results of the regressions are presented in Table 3.5.

Table 3.5 This table reports upon panel regression results with yield spread as dependent variable and fixed effects. Standard errors are clustered at bond level. (B) indicates that we use the bid–ask spread and (L) the LOT liquidity estimate as explanatory liquidity measure. The absolute values of the t-statistics are shown in parenthesis. *, **, *** denote significance at a 10%, 5%, and 1% level, respectively.

	Model 1		Model 2		Model 3	
	(B)	(L)	(B)	(L)	(B)	(L)
Bid–Ask Spread	319.48***		579.83***		641.42***	
	(3.17)		(4.90)		(5.60)	
LOT		203.68***		201.41***		264.95***
		(3.28)		(2.76)		(3.00)
Time to Maturity	−70.95	78.09	394.18	136.33	−244.91	35.51
	(0.13)	(0.96)	(0.63)	(1.02)	(0.56)	(0.36)
Government Bond	−48.56	−223.00	−535.33	−316.60	−1862.36	−223.52
	(0.03)	(1.37)	(0.21)	(1.26)	(0.92)	(1.30)
Term Slope	107.53	−303.83	−782.24	−818.78	574.39	−952.95*
	(0.06)	(1.56)	(0.38)	(1.60)	(0.39)	(1.93)
Rating Scale	303.61*	301.12***	483.37***	318.12***		
	(1.96)	(2.99)	(3.23)	(2.80)		
Income to Sales			−89.57	−63.91	−237.32	−147.85
			(0.70)	(0.94)	(1.25)	(1.54)
Debt to Assets			−687.59	38.80	−855.18	−11.66
			(0.92)	(0.05)	(1.04)	(0.02)
Interest Coverage			32.77	−5.63	38.48	−5.82***
			(0.90)	(0.83)	(0.95)	(3.07)
Debt to Capital			145.47	−81.56	2728.49	703.78
			(0.10)	(0.12)	(1.38)	(0.90)
Constant	−1517.63	−1126.31	−3339.77**	−630.79	−1721.13	1251.83**
	(1.20)	(1.50)	(2.08)	(0.84)	(0.92)	(2.45)
# Obs	187	248	106	148	132	188
F-Statistic	7.02	4.56	7.53	3.18	8.91	3.83
Within R^2	0.71	0.59	0.82	0.60	0.75	0.58

Table 3.6 This table reports upon panel regression results with yield spread as dependent variable and fixed effects. Accounting variables are added one by one. Standard errors are clustered at bond level. (B) indicates that we use the bid–ask spread and (L) the LOT liquidity estimate as explanatory liquidity measure. The absolute values of the *t*-statistics are shown in parenthesis. *, **, *** denote significance at a 10%, 5%, and 1% level, respectively.

	Model 1		Model 2		Model 3		Model 4	
	(B)	(L)	(B)	(L)	(B)	(L)	(B)	(L)
Bid–Ask Spread	623.40***		370.14**		374.45***		405.95***	
	(6.10)		(2.50)		(2.77)		(2.82)	
LOT		261.36***		264.09***		274.37***		277.69***
		(3.00)		(2.98)		(3.14)		(3.04)
Time to Maturity	−469.26	−10.55	−539.25***	50.40	−1157.65*	25.31	−481.55***	84.82
	(0.89)	(0.14)	(2.93)	(0.54)	(1.82)	(0.33)	(2.87)	(0.94)
Government Bond	−2377.79	−25.47	−933.08	−119.20	−5021.13*	−262.14	−826.68	−99.99
	(0.97)	(0.17)	(0.89)	(0.83)	(1.90)	(1.50)	(0.82)	(0.72)
Term Slope	1195.04	−925.46*	641.72	−979.41**	3167.37	−909.53*	560.14	−1005.60**
	(0.68)	(1.98)	(1.13)	(2.11)	(1.54)	(1.95)	(1.04)	(2.23)
Income to Sales	−99.99	−189.38**						
	(1.52)	(2.17)						
Debt to Assets			554.89	247.00				
			(0.69)	(0.61)				
Interest Coverage					−25.60	−6.43**		
					(1.52)	(2.49)		
Debt to Capital							1724.20**	820.53
							(2.49)	(1.49)
Constant	159.20	1914.35***	1145.90	1618.96***	586.40	1750.09***	−112.16	946.20**
	(0.25)	(4.09)	(1.29)	(4.35)	(0.92)	(4.10)	(0.12)	(2.06)
# Obs	140	197	144	202	139	196	158	221
F-Statistic	10.79	4.28	8.28	3.33	9.27	4.15	9.41	3.51
Within R^2	0.73	0.56	0.63	0.55	0.65	0.57	0.63	0.52

The liquidity estimates are highly significant in each model irrespective of whether credit rating, accounting variables, or all control variables are used. We find that the results are consistent for both liquidity estimates. In each model the coefficients of the bid–ask spread and the LOT liquidity estimate are positive and significant at a 1% level. A higher value of the liquidity measures indicates higher liquidity costs. Hence, our results do indeed support our main hypothesis that lower bond liquidity is associated with a higher yield spread for Mittelstand bonds.

The economic significance varies slightly between the liquidity estimates. The first model suggests that a 1% increase of the bid–ask spread is related to an incremental 3.19% increase in the yield spread. When using the LOT liquidity estimate instead the associated incremental increase of the yield spread only equals 2.04%. While the coefficient of the bid–ask spread increases in the second model after adding the accounting data the coefficient of the LOT liquidity estimate remains at a similar level compared to the first model. Yet, the coefficients on both liquidity estimates show the highest values in Model 3, in which accounting data instead of the rating information is included. Here, we can report that a 1% increase in the bid–ask spread (LOT liquidity measure) is related to an incremental 6.41% (2.65%) increase in the yield spreads. Comparing our results to the results of Chen et al. (2007b), we observe that the effect of the liquidity measures for Mittelstand bonds is approximately twice as strong as for speculative grade US corporate bonds and four to eight times as pronounced as for investment grade US corporate bonds.

Credit rating, one of our proxies for the default risk, is also highly significant. In both Models 1 and 2, and also for both liquidity specifications, a higher rating scale, and thus a higher default risk, is associated with a higher yield spread. In each model and each specification the coefficient of rating scale is positive indicating that a rating downgrade by one step is related to an increase of the yield spread by 3.00% to 4.90%. All other control variables are insignificant in Models 1 and 2. Yet, in the LOT liquidity specification of Model 3, term slope and interest coverage are significant. Consistent with Collin-Dufresne et al. (2001) the sign of term slope is negative. Interest coverage is also negatively related with the yield spread. As a high interest coverage suggests high financial performance and solvency, this result is intuitive. To further detail the effect of the accounting variables we modify Model 3 and include the accounting variables one by one. The results are presented in Table 3.6. Using bid–ask spread as the liquidity measure specification, debt to capital is the only significant accounting control variable. On the other hand, regressing the yield spread on the LOT liquidity estimate plus the control variables shows that while the accounting variables related to the firm performance (interest coverage and income to sales) are as expected negatively significant, both leverage ratios are insignificant.

Both liquidity measures provide high explanatory power regarding the yield spread of Mittelstand bonds. The values of the within R^2 range from 71% to 82% for the bid–ask spread and from 58% to 60% for LOT liquidity estimate. We observe higher within R^2 for regressions using the bid–ask spread compared to the LOT liquidity estimate. In particular, in Model 2 the bid–ask spread explains in combination with the other control variables 82% of the variation of the yield spread. Using the same control variables, the LOT liquidity estimate provides a within R^2 of 60% and thus, slightly lower explanatory power.

3.4.4 Simultaneous equation model tests

To control for potential endogeneity biases we apply a simultaneous equation model. A possible source of endogeneity are the liquidity estimates. In particular, liquidity costs could be influenced by credit rating. Credit quality is a main driver of adverse selection costs due to private information in the context of corporate bonds. Assuming that private information problems are more severe for bonds with a higher default risk indicates that bonds with a lower credit rating should incorporate higher private information costs. Private information costs, in turn, are a determinant of the liquidity costs. Thus a lower credit rating might lead to lower bond liquidity. Furthermore, the credit rating itself could be a second source of endogeneity. Rating agencies might not only consider accounting data to assess the quality of a bond but also account for market information observed through the yield spread. Hence, a higher yield spread could result in a lower credit rating.

To recognize that the liquidity and the credit rating might be determined endogenously we specify a system of three equations as follows:

$$
\begin{aligned}
\mathit{Yield\ Spread}_{i,t} &= \eta_0 + \eta_1 \mathit{Liquidity}_{i,t} + \eta_2 \mathit{Maturity}_{i,t} + \eta_3 \mathit{Government\ Bond}_{i,t} \\
&\quad + \eta_4 \mathit{Term\ Slope}_{i,t} + \eta_5 \mathit{Rating\ Scale}_{i,t} + \varepsilon_t \\
\mathit{Liquidity}_{i,t} &= \eta_0 + \eta_1 \mathit{Bond\ Volatility}_{i,t} + \eta_2 \mathit{Rating\ Scale}_{i,t} \\
&\quad + \eta_3 \mathit{Yield\ Spread}_{i,t} + \varepsilon_t \\
\mathit{Rating\ Scale}_{i,t} &= \eta_0 + \eta_1 \mathit{Income/Sales}_{i,t-1} + \eta_2 \mathit{Debt/Assets}_{i,t-1} \\
&\quad + \eta_3 \mathit{Interest\ Coverage}_{i,t-1} + \eta_4 \mathit{Debt/Capital}_{i,t-1} \\
&\quad + \eta_5 \mathit{Yield\ Spread}_{i,t} + \varepsilon_t
\end{aligned}
$$

where the subscript i,t denote bond i in year t and *Liquidity* refers to the bid–ask spread or the LOT liquidity estimate. As both the liquidity measures and the credit rating are endogenous in our framework, we use a three-stage least squares estimation technique to examine how these variables simultaneously impact the yield spread. The results are presented in Table 3.7. We estimate a separate model for each liquidity measure. The first column of each model shows a GLS-type estimation using the yield spread as dependent variable and the instrumented values instead of the endogenous variables. The last two columns represent the first-stage regression of the endogenous variables.

The results highlight that a possible bias due to endogeneity does not affect the previously examined relationship between liquidity and yield spread. In our sample the first-stage regressions cannot confirm an influence of the credit rating on the liquidity measures. However, the yield spread is significantly associated with both liquidity estimates. Regressing the rating scale on the yield spread and other variables also indicates that a higher yield spread is related to a higher rating scale and thus a lower bond quality. Nonetheless, when accounting for these endogeneities using the simultaneous equation model we find that the coefficients of both liquidity measures remain positive and significant at 1% level in the third-stage regressions. Therefore, after controlling for potential endogeneity bias we can indeed conclude that lower liquidity leads to a higher yield spread for the Mittelstand bonds.

Credit rating has a positive sign in the third stage of both models, yet, the coefficients are insignificant. Moreover, none of the other control variables is significantly associated with the yield spread. The R^2 indicates that 65% of the variation in the yield spread can be explained by Model (B). However, performing the same regression with the LOT liquidity estimate instead of the bid–ask spread only explains 5% of the variation.

3.5 Conclusion

The decision on the capital structure is critical for firms all over the world. US firms frequently tap capital markets to raise money and cover only one quarter of their funding requirements with traditional bank loans. By contrast, the German Mittelstand relied heavily on loans via relationship banking. Yet, in the light of tighter regulation, a trend towards other funding sources is clearly observable. Issuing bonds is a promising option to structure debt for Mittelstand firms.

Since the launch of the Mittelstand bonds market in 2010, yield spreads have increased steadily. In the next few years many of the early issued bonds will mature. Therefore, the near future

Table 3.7 This table reports upon simultaneous equation tests using three-stage least squared regressions. The instrumental variable indicates the dependent variable of each regression. The first column of each model represents a GLS-type estimation using the instrumented values instead of the endogenous regressors. The last two columns represent the first-stage regression. (B) indicates that we use the bid–ask spread and (L) the LOT liquidity estimate as explanatory liquidity measure. The absolute values of the t-statistics are shown in parenthesis. *, **, *** denote significance at a 10%, 5%, and 1% level, respectively.

	Model (B)			Model (L)		
Instrumental Variable	Yield Spread	Bid–Ask Spread	Rating Scale	Yield Spread	LOT	Rating Scale
Bid–Ask Spread	573.68***					
	(4.92)					
LOT				729.52***		
				(4.42)		
Time to Maturity	3.72			−2.67		
	(0.09)			(0.12)		
Government Bond	−767.65			−176.39		
	(0.63)			(0.56)		
Term Slope	−248.24			87.00		
	(0.60)			(0.35)		
Rating Scale	54.92	−0.10		91.64	−0.13	
	(0.54)	(0.63)		(0.81)	(0.88)	
Bond Volatility		0.11***			0.02	
		(2.90)			(0.58)	
Yield Spread		1.03E−3***	1.35E−3***		1.30E−3***	1.76E−3***
		(4.86)	(4.61)		(4.45)	(6.39)
Interest Coverage			−0.07			0.03
			(0.80)			(0.77)
Income to Sales			0.06			−0.07
			(0.21)			(0.26)
Debt to Assets			−2.61***			−0.97
			(2.77)			(1.34)
Debt to Capital			3.71***			2.49***
			(2.91)			(2.64)
Constant	−64.13	0.92	3.47***	−410.00	0.50	2.94***
	(0.08)	(1.13)	(3.42)	(0.73)	(0.65)	(4.10)
# Obs	106	106	106	148	148	148
R^2	0.65	0.67	0.35	0.05	0.19	0.18
p-Value	0.00	0.00	0.00	0.00	0.00	0.00

will show whether Mittelstand firms will be able to reschedule their debt. Rescheduling debt requires issuing new bonds to pay back the existing bond. The new bonds, however, will require a coupon that is adjusted to the contemporaneous level of the yield spread. Pessimists claim that given the current level of yield spreads, many Mittelstand firms will not be able to afford such new bonds. Moreover, with the default of more Mittelstand bonds, the perceived risk of this investment class will increase, increasing the yield spreads further and thus closing the vicious circle.

Our research provides important insights into this debate. We show that the effect of illiquidity on the yield spread is especially pronounced for Mittelstand bonds. This finding could open a back door towards the future of Mittelstand bonds. While, given a fixed investment and operating policies, default risk of Mittelstand firms is mostly exogenous[3], liquidity is endogenous for the firms. As an example, firms could increase the liquidity of their bonds by decreasing the adverse selection costs due to private information, for instance by more timely and comprehensive reporting. As our research shows, even small increases in the liquidity of Mittelstand bonds can lead to substantial decreases in the yield spreads.

[3]Notice that as discussed in Section 3.3.1 Mittelstand firms are very reluctant to increase their equity on public markets in order to not dilute the founding family's ownership and control rights.

3.6 Appendix

3.6.1 List of all Mittelstand bonds

Table 3.8 This table reports upon characteristics of all German Mittelstand bonds om our final sample.

ISIN	Borrower Name	Maturity (years)	Coupon (%)	Volume (in thousands)	Issue Date
AT0000A0U9J2	Scholz AG	5	8.5	182,500	03/08/12
AT0000A0XJ15	Porr AG	4	6.25	50,000	12/04/12
AT0000A185Y1	UBM Realitätenentwicklung AG	5	4.875	200,000	07/09/14
DE000A11QGQ1	KTG Agrar AG	5	7.25	50,000	10/15/14
DE000A11QHZ0	HanseYachts AG	5	8	20,000	06/03/14
DE000A11QJA9	Vedes AG	5	7.125	20,000	06/24/14
DE000A12T1W6	Beate Uhse AG	5	7.75	30,000	07/09/14
DE000A12UAA8	KSW Immobilien GmbH & Co. KG	5	6.5	25,000	10/07/14
DE000A12UD98	Studierendengesellschaft Witten Herdecke EV	10	3.6	7,500	12/02/14
DE000A13SAD4	Neue ZWL Zahnradwerk Leipzig GmbH	6	7.5	25,000	02/17/15
DE000A161F97	Katjes International GmbH & Co. KG	5	5.5	60,000	05/15/15
DE000A1CR0X3	Albis Leasing AG	5	7.625	50,000	10/04/11
DE000A1ELQU9	KTG Agrar AG	5	6.75	50,000	06/01/09
DE000A1EWGX1	Duerr AG	5	7.25	225,000	09/28/10
DE000A1EWL99	Nabaltec AG	5	6.5	30,000	10/15/10
DE000A1EWNF4	Hahn Immobilien Beteiligungs AG	5	6.25	20,000	10/01/12
DE000A1G9AQ4	Enterprise Holdings LTD	5	7	35,000	09/26/12
DE000A1H3EY2	MAG IAS GmbH	5	7.5	50,000	02/08/11
DE000A1H3F20	Albert Reiff GmbH & Co. KG	5	7.25	30,000	05/27/11
DE000A1H3GE9	Joh. Friedrich Behrens AG	5	8	30,000	03/15/11

continued.

ISIN	Borrower Name	Maturity (years)	Coupon (%)	Volume (in thousands)	Issue Date
DE000A1H3J67	German Pellets GmbH	5	7.25	75,000	04/01/11
DE000A1H3V53	ENO Energy GmbH	5	7.375	25,000	06/30/11
DE000A1H3VN9	KTG Agrar AG	6	7.125	200,000	06/06/11
DE000A1H3YJ1	Semper Idem Underberg GmbH	5	7.125	70,000	04/20/11
DE000A1H3YK9	Valensina GmbH	5	7.375	85,000	04/28/11
DE000A1HJLL6	S&T AG	5	7.25	15,000	05/22/13
DE000A1HLTD2	Metalcorp Group BV	5	8.75	50,000	06/27/13
DE000A1HPZD0	VST Building Technologies AG	6	8.5	15,000	10/02/13
DE000A1HSNV2	Porr AG	5	6.25	50,000	11/26/13
DE000A1K0169	Bastei Luebbe GmbH & Co. KG	5	6.75	30,000	10/26/11
DE000A1K0FA0	Eyemaxx Real Estate AG	5	7.5	25,000	07/26/11
DE000A1K0FF9	GIF Gesellschaft für Industrieforschung MBH	5	8.5	3,455	09/20/11
DE000A1K0NJ5	MITEC Automotive AG	5	7.75	50,000	03/30/12
DE000A1K0QA7	Royalbeach Spielwaren und Sportartikel Vertriebs GmbH	5	8.125	25,000	10/28/11
DE000A1K0SE5	Textilkontor Walter Seidensticker GmbH	6	7.25	30,000	03/12/12
DE000A1K0U44	Procar Automobile Finanz Holding GmbH & Co. KG	5	7.75	12,000	10/14/11
DE000A1KQ367	Uniwheels Property Germany GmbH	5	7.5	44499	04/19/11
DE000A1KQ3C2	Senivita Sozial Gemeinnuetzige GmbH	5	6.5	15,000	05/17/11
DE000A1KQ8K4	Peach Property Group Deutschland GmbH	5	6.6	50,000	07/18/11
DE000A1KQZL5	MS Spaichingen GmbH	5	7.25	23,000	07/15/11

continued.

ISIN	Borrower Name	Maturity (years)	Coupon (%)	Volume (in thousands)	Issue Date
DE000A1KRBM2	Katjes International GmbH & Co. KG	5	7.125	45,000	07/19/11
DE000A1MA9E1	Golfino AG	5	7.25	12,000	04/05/12
DE000A1MASJ4	Singulus Technologies AG	5	7.75	60,000	03/23/12
DE000A1ML257	KTG Energie AG	6	7.25	50,000	09/28/12
DE000A1ML4T7	Fussballclub Gelsenkirchen Schalke 04 EV	7	6.75	35,000	06/11/12
DE000A1MLSJ1	Ekosem-Agrar GmbH	5	8.75	50,000	03/23/12
DE000A1MLWH7	Eyemaxx Real Estate AG	5.56	7.75	15,000	04/11/12
DE000A1MLYJ9	Friedola Gebr Holzapfel GmbH	5	7.25	13,000	04/11/12
DE000A1PGQL4	BDT Media Automation GmbH	5	8.125	17,380	10/09/12
DE000A1PGQR1	Rene Lezard Mode GmbH	5	7.25	15,000	11/26/12
DE000A1PGRG2	Travel24 com AG	5	7.5	25,000	09/17/12
DE000A1PGUT9	posterXXL AG	5	7.25	15,000	07/27/12
DE000A1PGWZ2	Steilmann Boecker Fashion Point GmbH & Co. KG	5	6.75	40,000	06/27/12
DE000A1R07C3	Constantin Medien AG	5	7	65,000	04/23/13
DE000A1R07G4	Deutsche Rohstoff AG	5	8	100,000	07/11/13
DE000A1R09H8	Timeless Homes GmbH	7	9	10,000	07/02/13
DE000A1R0RZ5	Ekosem-Agrar GmbH	6	8.5	78,000	12/07/12
DE000A1R0VD4	Homann Holzwerkstoffe GmbH	5	7	100,000	12/14/12
DE000A1R0YA4	Rudolf Woehrl AG	5	6.5	30,000	02/12/13
DE000A1R1A42	Adler Real Estate AG	5	8.75	35,000	04/03/13
DE000A1R1BR4	Alno AG	5	8.5	45,000	05/14/13
DE000A1R1CC4	DF Deutsche Fortfait AG	7	2	30,000	05/27/13
DE000A1RE1V3	Berentzen Gruppe AG	5	6.5	50,000	10/18/12

continued.

ISIN	Borrower Name	Maturity (years)	Coupon (%)	Volume (in thousands)	Issue Date
DE000A1RE5T8	Laurel GmbH	5	7.125	20,000	11/16/12
DE000A1RE7P2	Jacob Stauder GmbH & Co. KG	5	7.5	10,000	11/23/12
DE000A1RE8B0	Euroboden GmbH	5	7.375	15,000	07/16/13
DE000A1REWV2	Karlsberg Brauerei GmbH	5	7.375	30,000	09/28/12
DE000A1REXA4	Eterna Mode Holding GmbH	5	8	55,000	10/09/12
DE000A1RFBP5	Immobilien Projekt Salamander Areal Kornwestheim	7	6.75	30,000	12/06/12
DE000A1TM2T3	Eyemaxx Real Estate AG	6	7.875	15,000	03/26/13
DE000A1TM8Z7	Stern Immobilien AG	5	6.25	20,000	05/23/13
DE000A1TNA70	Sanha GmbH Co. & KG	5	7.75	37,500	06/04/13
DE000A1TNAP7	German Pellets GmbH	5	7.25	72,000	07/09/13
DE000A1TND44	More & More AG	5	8.125	13,000	06/11/13
DE000A1TND93	Paragon AG	5	7.25	20,000	07/02/13
DE000A1TNFX0	Peine GmbH	5	8	15,000	07/05/13
DE000A1TNG90	Karlie Group GmbH	5	6.75	30,000	06/25/13
DE000A1TNGG3	Cloud NO 7 GmbH	4	6	35,000	07/03/13
DE000A1TNHC0	Bioenergie Taufkirchen GmbH & Co. KG	7	6.5	15,000	07/30/13
DE000A1TNJY0	Gamigo AG	5	8.5	15,000	06/20/13
DE000A1X3HZ2	Helma Eigenheimbau AG	5	5.875	35,000	09/19/13
DE000A1X3MA5	Alfmeier Praezision Baugruppen und Systemloesungen AG	5	7.5	30,000	10/29/13
DE000A1X3MD9	Gebr. Sanders GmbH & Co. KG	5	8.75	18,000	10/22/13
DE000A1X3MS7	Sympatex Holding GmbH	5	8	13,000	12/03/13
DE000A1X3VZ3	Ferratum Capital Germany GmbH	5	8	25,000	10/21/13

continued.

ISIN	Borrower Name	Maturity (years)	Coupon (%)	Volume (in thousands)	Issue Date
DE000A1YC1F9	Neue ZWL Zahnradwerk Leipzig GmbH	5	7.5	25,000	03/04/14
DE000A1YC7Y7	GEWA 5 TO 1 GmbH & Co. KG	4	6.5	35,000	03/24/14
DE000A1YCRD0	Hoermann Finance GmbH	5	6.25	50,000	12/05/13
DE000A1ZWPT5	Enterprise Holdings LTD	5	7	85,000	03/30/15
DE000AB100A6	Air Berlin PLC	5	8.5	200,000	11/10/10
DE000AB100B4	Air Berlin PLC	7	8.25	225,000	04/19/11
DE000AB100C2	Air Berlin PLC	3	11.5	150,000	11/01/11

3.6.2 List of all variables

Table 3.9 This table describes the variables employed in our study.

Variable Name	Abbreviation	Description	Data Source
Age of the Bond	Age	Time period since the issuance of the bond (in years)	Datastream
Bid–Ask Spread	Bid–Ask Spread	Calculated as the ask quote minus the bid quote divided by the average of both quotes	Datastream
Bond Volatility	Bond Volatility	Yearly standard deviation of the clean prices of the bond	Datastream
DAX Return	ΔDAX	Daily return on the DAX 30 composite stock index	Datastream
Debt to Assets	Debt / Assets	Long-term debt to assets	Dafne / Amadeus
Debt to Capital	Debt / Capital	Total debt to capital	Dafne / Amadeus
Duration	D	Modified Duration to final date	Datastream
Government Bond	Government Bond	1-year German Bunds rate	Datastream
Income to Sales	Income / Sales	Operating income to sales	Dafne / Amadeus
Interest Coverage	Interest Coverage	EBIT plus interest to interest	Dafne / Amadeus
LOT Liquidity Estimate	LOT	Liquidity measure based on Lesmond et al. (1999)	Datastream
Rating Scale	Rating Scale	Numeric value for each rating class ranging from 1 for A (the best rating in our sample) to 15 for D (default)	Rating Reports
Risk-free Bond	$R_{f,t}$	10-year German Bunds rate	Datastream
Term Slope	Term Slope	Difference in rates of 10-year and 2-year German Bunds	Datastream
Time to Maturity	Maturity	Remaining life of a Mittelstand bond	Datastream

continued.

Variable Name	Abbreviation	Description	Data Source
Yield Spread	Yield Spread	Spread of the yield of a Mittelstand bond over an equivalent government benchmark bond	Datastream

Chapter 4

Pricing in the online invoice trading market: First empirical evidence.

This research project is joint work with Gregor Dorfleitner and Jaqueline Rad. The paper has been published as: Gregor Dorfleitner, Jacqueline Rad and Martina Weber (2017), Pricing in the online invoice trading market: First empirical evidence, *Economics Letters* 161, 56-61.

Abstract In recent years, online invoice trading has gained importance in providing SMEs with short-term financing. In this paper, we present first empirical evidence concerning the question whether the risk of payment difficulties is appropriately reflected in the pricing variables. To this end, we investigate predictors of default of online invoice trading platforms. We analyze both the probability of default and the loss rate and find that the interest rate, the duration and the percentage funded have good predictive power. Furthermore, we show that the pricing mechanism (auction vs. fixed prices) helps to explain defaults on online invoice trading platforms.

Keywords Invoice trading, Factoring, FinTech, MarketInvoice, auction, efficiency, PD forecast

JEL Classification G21, G23, L31, M14

4.1 Introduction

Invoice trading is a fast and easy way in which small and medium sized enterprises (SMEs) can raise short-term debt by pre-financing their outstanding invoices through individual or institutional investors. In this study, we empirically analyze whether the risk of payment difficulties is appropriately reflected in the prices of online invoice trading platforms. To this end, we use a novel data set stemming from an invoice trading platform to investigate which factors predict defaults, i.e. events in which the investors do not fully receive the invested amount plus interest rate.

SMEs often face difficulties in obtaining sufficient sources of financing. In addition to the traditional factoring market and other forms of financing such as bank loans and overdraft facilities, online invoice trading platforms can help SMEs to raise working capital. Generally, these web-based platforms are hosted by FinTechs. In recent years, the market for online invoice trading has grown substantially. In the UK, the market volume more than tripled between 2013 and 2015 (Zhang et al., 2016). While in 2013 the volume amounted to £97m, the market exceeded this figure considerably with nearly £325m in 2015. From a global perspective, online invoice trading is likely to continue to grow further.

We are the first to analyze this new market of invoice trading on web-based platforms. We use data of the world's largest invoice trading platform MarketInvoice henceforth also called *the platform*, which is based in the UK. To investigate the determinants of repayment difficulties, we focus on crystallized losses and the loss rate of the invoices and apply both logit and tobit models. We find that the interest rate, the duration and the percentage of the invoice funded are related to the default probability. Within our observation period, the platform applies two different market mechanisms to set the prices of the invoices, namely an auction and a fixed-price mechanism. We show that the default probability is lower within the fixed-price regime. However, the gross yield as well as the return for investors are higher within the auction period. On online invoice trading platforms, an invoice is generally sold to several investors. Hence, our study also contributes to the growing amount of literature upon different forms of crowdfunding.

4.2 Related literature

In general, factoring is a short-term supply of financing whereby companies sell their accounts receivables at a discount in exchange for immediate cash. In recourse factoring and usually also in invoice trading, the buyer pre-finances the invoice but does not resume the credit risk for a potential default of the invoice.

Klapper (2006) states that factoring is a growing source of financing for SMEs all around the world. However, she finds evidence proving that the factoring market is larger in countries with good economic development and growth as well as in countries with a sound provision of credit information on companies. Soufani (2002b) focuses on the UK factoring market and examines parameters influencing the decision of factoring companies to purchase accounts receivables. Additionally, Soufani (2002a) investigates the choice of companies to use factoring as a source of financing.

In online invoice trading, an invoice is generally sold to one or more investors. Hence, the concept is closely linked to other forms of crowdfunding upon which a vast amount of literature has been published. In particular, previous research deals with determinants of defaults in crowdfunding. Several studies find that the interest rate and other loan characteristics such as the credit score are highly important in explaining the default probability in crowdfunding (see for example Dorfleitner et al., 2016; Serrano-Cinca et al., 2015; Emekter et al., 2015). Yet, further academic work shows that other pieces of information such as the perceived creditworthiness of the borrower (see Duarte et al., 2012) or online friendships (see Lin et al., 2013) are also associated with defaults.

In crowdfunding, prices are generally either set by the platform (fixed price) or emerge in market-based auctions. In the early days of crowdfunding, many platforms preferred the auction mechanism. However, most of these platforms have changed from auctions to fixed prices over time. Wei and Lin (2016) examine the latter price regime change on the US crowdlending platform Prosper. They find that loans are more likely to be funded under a fixed-price regime. However, the default rates are higher when the platform posts the prices, which is also reflected by higher interest rates. While Huang (2017) also focuses on the price regime change on Prosper, Chen et al. (2014) analyze whether the auction on Prosper leads to the lowest payments of borrowers under the assumption of strategic and rational agents.

Further research deals with similar market mechanisms in other crowdfunding forms. Hornuf and Schwienbacher (2017) compare the funding dynamics of equity crowdfunding portals with a fixed price and those using an auction mechanism. In contrast to platforms with fixed prices, the funding patterns of platforms using auctions are U-shaped. Franks et al. (2016) focus on the lending-based crowdfunding platform FundingCircle. Amongst others, they find that auctions generate additional information that helps to predict defaults. Furthermore, there are several studies that compare fixed-price regimes with auction mechanism in other markets (e. g. Wang (1993); Hammond (2013); Einav et al. (2018); Chen et al. (2007a)).

4.3 Data and methodology

4.3.1 Institutional background and data

A transaction on the platform constitutes the pre-financing of invoices in the sense of recourse factoring. The investors purchase the accounts receivables but do not assume the risk of the debtor's insolvency. Figure 4.1 visualizes how a transaction is proceeded.

After the seller has uploaded the invoice and after the platform has verified it, investors can purchase either the invoice or fractions of it. Dependent on the seller's industry, the duration of an invoice and the stage of the seller's business, the investors fund a fraction of up to 90 % and more of the invoice face value. This fraction is called the *advance*. A transaction is frequently split between 20 or more investors. Subsequent to the funding, the seller immediately receives the advance value. Within the payment period, the investors accrue interest on a daily basis until the invoice is repaid. At maturity, the debtor (the seller's customer) repays the full face value of the invoice to the platform. Then the platform pays back the advance value and all accrued interest to the investors. Finally, the seller receives the non-advanced remainder less interest (see MarketInvoice Limited, 2017b, 2016). In case the debtor does not fully pay the invoice, the platform demands that the seller repurchases the invoice. Therefore, only cases in which neither the seller nor the debtor repay the entire advance value plus interest result in crystallized losses for investors. This marks a big difference to the field of crowdlending, where the credit risk solely depends on the risk of the debtor.

Since the end of 2013, the interest rate on the face value of the invoice as well as the maximum advance rate have been predetermined by a platform-internal risk-based pricing model. Within this pricing mechanism, the seller receives the invested amount regardless of whether or not the maximum advance rate is reached. Before December, 2013, this interest rate and the percentage funded were set through a real-time auction mechanism. Before the start of an auction, the seller defines the minimum advance value and the maximum interest rate he or she is willing to pay as well as the duration of the auction. The investors bid based on information about these seller requirements and a rating of the invoice provided by the platform. The bids that satisfy the minimum requirements defined by the seller and that are best in the sense of a high advance value and a low interest rate are executed at the end of the auction at a unique interest rate and at an advance that is as high as possible.[1] According to a statement of the platform, the auction system was no longer suitable because of the rapidly growing volume of invoices and, therefore, they changed to a fixed-price mechanism. Furthermore, the platform states that data on more than 18,200 invoices enables the platform to develop a fixed-price model, which assesses the assets in a timely manner.

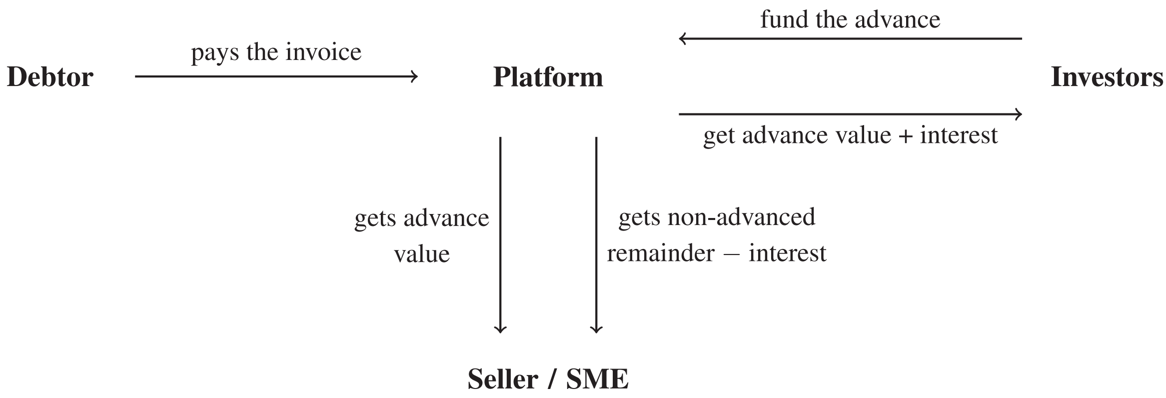


Figure 4.1 A transaction on the platform.

To sell invoices companies have to fulfil several requirements defined by the platform such as a turnover of at least £100,000 and a business activity of more than six month. Furthermore, only limited or LLP companies are allowed to use the invoice trading platform. The invoices can be bought by accredited institutional investors, family offices and also self-certified sophisticated investors as well as certificated high net worth individuals.

The dataset used in this paper was obtained from MarketInvoice and contains all completed fundings from March 2011 until mid-May 2017. We only consider closed transactions and therefore exclude all information on invoices that still await repayment and have not yet resulted in crystallized losses. After data cleansing, our data set includes 19,566 observations.

[1] The realized advance emerges as the sum of the investment volumes of the successful bids divided by face value, while it needs to be less or equal to the maximum advance the successful bidders are willing to accept. Consider for example the auction of an invoice worth £10,000, which the seller wants to be financed at a maximum interest rate of 10 % and a minimum advance rate of 60 %. We consider four bidders, each offering a volume of £2500. Bidder A offers an interest rate of 8 % and a maximum advance rate of 75 %, B 7 % and 85 %, C 9 % and 85 %, and D 8.5 % and 60 %, respectively. At the end of the auction, the invoice is sold to A, B, and C for an interest rate of 9 %. The advance rate equals 75 %. Bidder D is not successful even though his offered interest rate is lower than the winning interest rate of 9 %. D only accepts an advance rate of 60 %, thus, selling the invoice to A, B, and D would violate this condition.

4.3.2 Methods and Explanatory Variables

To study the pricing in the invoice trading market, we focus on repayment difficulties of sold invoices. First, we consider defaults with a binary variable (*Default*) indicating whether or not a seller has failed to fully repay the invoice. To this end, we estimate logit models with Eicker–Huber–White heteroskedastic-consistent standard errors. Second, we investigate the determinants of the percentage of an invoice which has not been paid back (*Loss Rate*). In the majority of the cases, we do not observe any form of payment difficulties and our dependent variable equals zero. In the case of a default, the loss rate ranges between 0 and 1. Thus, we apply a tobit estimation.

In both approaches we include several explanatory variables. The interest rate investors demand for funding the invoice (*Gross Yield*) serves as a proxy for the perceived risk of a given invoice.[2] The lower the creditworthiness of the seller and the debtor, the higher the yield investors demand. Furthermore, the platform also considers the default risk when deciding on the advance rate. Whether investors do so in the auction mechanism is unclear. Additionally, we consider the maturity of the invoice (*Time to pay*) as well as the loan amount (*Advance Value*). All dependent and explanatory variables are defined in Table 4.1.

4.3.3 Descriptive analysis

Table 4.2 presents the descriptive statistics of the categorical and the metric variables. The data reveals that defaults are rare in this form of factoring. Only 384 out of 19,566 sold invoices resulted in crystallized losses for the investors. The maximum loss investors faced of a single invoice equals almost £500,000. With a mean of £167 and a median of zero, the outstanding amount is highly skewed with a fat right tail. On average, the investors funded 81 % of the invoice value resulting in a mean advance value of £48,609. Taking into account the outstanding amount and the advance value of defaulted transactions reveals an average loss given default (LGD) rate of 14.7 %. The interest investors gain for funding the invoices ranges between 4.03 % and 48.16 % p. a. On average, the rate equals 12.28 % p. a. Furthermore, the data shows that the majority of the invoices are sold with a fixed price set by the platform. An auction only took place for approximately 9.4 % of the transactions.

[2]The rating provided by the platform is highly correlated with the interest rate (correlation coefficient of 0.92). In the further analysis we therefore only focus on the gross yield.

Table 4.1 Definition of variables.

Advance (%)	The percentage of the invoice value which is funded by the investors and paid out to the seller upfront.
Advance Value	The amount of money which is funded by the investors and paid out to the seller upfront, in GBP.
Auction	Dummy variable that indicates whether the price of the transaction is set via an auction. The platform used an auction model for all transactions until December 2013.
Default	Dummy variable that indicates whether both the debtor and the seller are unable to fully repay the invoice amount plus interest. In particular, a transaction is regarded as default when the seller has gone into liquidation or administration and the process reveals that the seller does not have enough assets to cover the outstanding liabilities. Furthermore, invoices are marked as default when 180 days have passed without resolution (MarketInvoice Limited, 2016).
Gross Yield (%)	The gross yield an investor receives on average p. a. as a percentage of the advance value. The gross yield equals the annualized interest rate, whereby the interest rate equals the total fee the seller has to pay for the invoice discounting service.
LGD	The loss given default (LGD) equals the share of the invoice that is not paid back in the event of a default.
Loss Rate	The percentage of an invoice that is not paid back. Calculated as *Outstanding* over *Advance Value*.
Outstanding	The residual debt that both the debtor and the invoice seller have not repaid before settlement of the transaction.
Time to pay	The difference between the Advance Date and the Expected Payment Date.

When focusing on the subsamples, we find some crucial differences between the auction period and the fixed-price period. We test for a difference of means using a t-test allowing unequal variances. We find that both the average default rate and the loss rate are significantly higher within the auction period. However, the average annual gross yield is significantly higher during this period as well. It is noteworthy that the advance rate is also significantly higher during the auction period. This phenomenon can be explained by the auction mechanism itself (illustrated in footnote 1), which maximizes the advance given the restrictions of the seller and the bidders. The LGD does not significantly change between the two pricing mechanisms.

4.4 Results

Table 4.3 shows the average marginal effects of our first model, the logit regressions, with the default dummy as dependent variable. We present the estimation results for our second model, the tobit regressions with the loss rate as dependent variable, in table 4.4, respectively. In both models we estimate Eicker–Huber–White heteroskedastic-consistent errors and test for potential

Table 4.2 Descriptive statistics. The variables are defined in Table 4.1.

Categorical Variables	n	Yes	Mean	S.D.	Auction	Fixed Price	Difference
Default	19,566	384	0.020	0.139	0.034	0.018	0.0153***
Auction	19,566	1,848	0.094	0.292			

Metric Variables	n	Mean	S.D.	Min	Median	Max
Outstanding	19,566	167.45	5299.50	0.00	0.00	488,898.80
Loss Rate	19,566	0.003	0.0482	0.00	0.00	1.00
LGD	384	0.147	0.3118	0.00	0.00	1.00
Gross Yield	19,566	12.276	4.070	4.03	11.35	48.16
Advance (%)	19,566	81.329	7.874	3.37	85.00	97.00
Time to pay	19,566	45.589	27.353	0.00	41.00	404.00
Advance Value	19,566	48,609.44	88,409.53	78.55	21,524.40	1,670,064.00

	Auction (n=1848)		Fixed Price (n=17,718)		Difference
Subsamples	Mean	S.D.	Mean	S.D.	
Outstanding	202.686	2272.091	163.775	5520.490	38.911
Loss Rate	0.005	0.052	0.003	0.048	0.003**
LGD	0.156	0.239	0.146	0.324	0.010
Gross Yield	16.812	5.152	11.802	3.627	5.009***
Advance (%)	83.950	6.147	81.055	7.983	2.894***
Time to pay	38.577	23.263	46.320	27.643	−7.744***
Advance Value	47,132.430	71,639.660	48,763.490	89,979.120	−1631.065

issues due to multicollinearity[3]. Specification (1) includes all explanatory variables as well as year dummies to capture variation over time. In specification (2), we incorporate a dummy for the auction period instead of the year dummies. The last two specifications, (3) and (4), focus on subsamples for the different pricing mechanisms.

Table 4.3 Average marginal effects of the logit models. The dependent variable is the default dummy. The regression is performed using Eicker–Huber–White heteroskedastic-consistent errors. Standard errors are in parentheses. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 4.1.

Sample	(1) Complete	(2) Complete	(3) Auction	(4) Fixed Price
Gross Yield	0.0010***	0.0015***	0.0035***	0.0012***
	(0.0003)	(0.0002)	(0.0007)	(0.0002)
Time to pay	0.0002***	0.0002***	0.0006***	0.0002***
	(0.0000)	(0.0000)	(0.0002)	(0.0000)
Advance (%)	−0.0004***	−0.0002	0.0011	−0.0003***
	(0.0001)	(0.0001)	(0.0007)	(0.0001)
ln(Advance Value)	−0.0026***	−0.0021***	−0.0006	−0.0026***
	(0.0009)	(0.0008)	(0.0034)	(0.0008)
Auction		0.0059*		
		(0.0030)		
Year Dummies	Yes	No	No	No
Observations	19,566	19,566	1848	17,718
Pseudo-R^2	0.07	0.0515	0.1396	0.0387

We find that the gross yield reveals good predictive power in explaining both the default probability and the loss rates. An increase in the gross yield is associated with a higher probability of default and a higher loss rate. This is consistent with previous research. Stiglitz and Weiss (1981) explain that borrowers who accept high interest rates perceive their own creditworthiness as being poor and are therefore more likely to default. Interestingly, both in logit and tobit models we find that the gross yield has the highest average marginal effect for the auction period. The magnitude of the effect is smaller for the fixed-price regime.

Furthermore, the results show a significant, positive relationship between the maturity of the invoice and the probability of default. Sellers are less likely to default on invoices with a short expected time until payment. The same relationship holds true for the loss rates.

The effect of the advance rate is more complex. In our first model, we find that the advance rate is significantly negatively associated with the probability of default in the first and the last specification. Within the fixed-price regime the platform sets the advance rate and considers,

[3]In all specifications, the variance inflation factors are below a value of 1.40.

Table 4.4 Average marginal effects of the tobit models. The dependent variable is the loss rate. The regression is performed using Eicker–Huber–White heteroskedastic-consistent errors. Standard errors are in parentheses. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 4.1.

Sample	(1) Complete	(2) Complete	(3) Auction	(4) Fixed Price
Gross Yield	0.0472***	0.0503***	0.0549***	0.0365*
	(0.0131)	(0.0124)	(0.0112)	(0.0189)
Time to pay	0.0079***	0.0080***	0.0084***	0.0082***
	(0.0011)	(0.0010)	(0.0018)	(0.0014)
Advance (%)	−0.0110	−0.0065	0.0353***	−0.0182**
	(0.0075)	(0.0082)	(0.0121)	(0.0082)
ln(Advance Value)	0.0417	0.0492	−0.0867	0.0594
	(0.0548)	(0.0544)	(0.0740)	(0.0684)
Auction		0.5354***		
		(0.1546)		
Year Dummies	Yes	No	No	No
Observations	19,566	19,566	1848	17,718
Pseudo-R^2	0.0809	0.0711	0.1822	0.0407

among other things, the risk of the invoice not being repaid. Thus, the significant negative relationship reflects the ability of the platform to assess the creditworthiness of the sellers. In our second model, the advance rate is significantly negatively related to the loss rates under the fixed-price regime as well. Yet, for the auction subsamples we do not find such a relationship. We interpret this result in the following way. From an investor's perspective, the interest rate is a more important pricing parameter than the advance rate since a high interest rate always has a positive impact on the expected return, while the investors only benefit from a low advance rate in the case of a default. Furthermore, the risk can effectively be reduced by diversifying the portfolio of invoices. Thus, investors prefer to price the risk by demanding high interest rates rather than by demanding low advance rates. Moreover, the auction mechanism fosters high advance rates, which also can be observed empirically (see Table 4.2). An analysis of the pairwise Bravais-Pearson correlation coefficients undermines this view. The correlation between the gross yield and the advance is considerably higher in the fixed-price period (−0.48) than in the auction period (−0.33) indicating that the advance indeed is a more relevant measure for credit risk in the fixed-price period than before. However, the significantly positive sign of the advance rate in the auction subsample of the second model cannot be explained just by this consideration. We regard this result as an indication for the presence of sophisticated investors which tended to demand high interest rates for risky invoices by at the same time allowing high advances in order to maximize their return. Indeed, this notion is in line with an analysis of the investors' net returns[4]. We find that during the auction period the average net return amounts to 7.52 % p.a. This rate decreases considerably to only 4.18 % p.a. when prices are set by the platform.

In our first model, the results show a significant negative relationship between the amount of the loan and defaults in all specifications except the auction-period subsample. Larger and more creditworthy companies can be assumed to obtain larger loans more easily. Hence, the advance value may be a proxy for the size of the seller. Additionally, this result indicates that sellers of invoices do not tend to engage in fraudulent behavior. Otherwise they would tend to strategically default on larger loans. We do not observe any significant relationships between the advance value and the loss rates in our second model.

Moreover, the auction dummy is significantly associated with the defaults and the loss rates. Both the probability of default and the loss rates are higher under the auction mechanism indicating that riskier invoices were sold within the auction period. In line with Wei and Lin (2016), we show that the market mechanism resulting in higher interest rates is associated with higher default rates. This is also consistent with Stiglitz and Weiss (1981) who point out that high interest rates are related to high default rates. Better economic conditions in the years of the fixed-price period may be another reason for higher loss rates during the auction period. Large parts of the UK observed less company insolvencies in recent years than in the auction period (United Kingdom Statistics Authority, 2017).

4.5 Conclusion

In this paper, we empirically analyze the determinants of defaults and the loss rates in online invoice trading. In line with previous literature, we find that the interest rate plays an important role in explaining the default probability in the context of invoice trading (see Stiglitz and Weiss, 1981; Dorfleitner et al., 2016; Wei and Lin, 2016). Furthermore, the duration and the percentage funded are associated with the defaults.

As the platform has changed the pricing mechanism from a real-time auction to a fixed-price regime, we have the opportunity to compare these two market mechanisms. We find that the default probability is higher within in the auction period. Furthermore, we show that the average

[4]Net returns equal the returns after losses and fees. We generally assume a fee of 25 % of the gains (MarketInvoice Limited, 2017a). Furthermore, we assume that sellers of defaulted invoices do not pay interest. To control for the macroeconomic environment we also deduct the respective average LIBOR rate.

net return investors gained was higher during the auction period than afterwards. So, one can interpret the change from the auction to the fixed-price regime as an attempt to make the pricing fairer for the companies selling the invoices.

As the market for online invoice trading is growing quickly the opportunity and the need for future research arises. In this study, we only focus on data of the market leader for online invoice trading. A rich data set of different online invoice trading providers could enable greater insights into the market. Furthermore, as sellers use invoice trading repeatedly, reputation effects may arise. Further research should focus on the question to what extent the reputation of sellers lowers the information asymmetries and therefore affects the prices for sold invoices. A more detailed analysis of the different market mechanisms and the consequences for both investors and sellers could be of interest as well.

Chapter 5

Dynamics of investor communication in equity crowdfunding

This research project is joint work with Gregor Dorfleitner and Lars Hornuf. The paper is forthcoming in the *Journal of Electronic Markets*.

Abstract In crowdfunding, start-ups can voluntarily communicate with their investors by posting updates. We investigate whether start-ups strategically use updates, which have previously been shown to increase investments. To this end, we use hand-collected data from 751 updates and 39,036 investment decisions from two major German equity crowdfunding portals: Seedmatch and Companisto. We find evidence of strategic communication behavior of start-ups during an equity crowdfunding campaign. During the funding period, start-ups more frequently post updates with linguistic devices that enhance the group identity and the group cohesion as well as updates on the business development. Furthermore, the probability of an update during the funding period increases along with strong competition of other contemporary crowdfunding campaigns.

Keywords Investor Communication, Entrepreneurial Finance, Sentiment Analysis, Linguistic Devices

JEL Classification G21, G24, G32, L11, L26

5.1 Introduction

In recent years, equity crowdfunding has gained increasing importance in providing start-ups with funding. In contrast to traditional early-stage financing sources such as venture capital and banks, equity crowdfunding has introduced the possibility for non-sophisticated private investors to invest in start-ups. However, there are crucial differences between the information rights and the experience of venture capitalists and crowd investors. In this article, we therefore analyze the communication behavior of start-ups in equity crowdfunding during and after the funding period and investigate whether entrepreneurs use voluntary disclosure strategically.

Recent academic research in equity crowdfunding analyzes follow-up fundings, crowd exits and insolvencies of successfully funded equity crowdfunding campaigns (Hornuf and Schmitt, 2016; Signori and Vismara, 2018; Hornuf et al., 2018). However, the majority of the literature investigates determinants of the funding success of a campaign. A correlation between the success of campaigns and the size and education of the management team as well as particular project characteristics—e.g. the share of equity offered or disclosure of financial projections—has been shown (Ahlers et al., 2015; Vismara, 2016; Bernstein et al., 2017). Furthermore, the posting of voluntary information in the form of updates during the campaign increases the likelihood of funding success (Mollick, 2014; Block et al., 2018a).

Both the crowdfunding and the corporate finance literature (Diamond and Verrecchia, 1991; Healy and Palepu, 2001; Merton, 1987) provide evidence of a positive impact of voluntary disclosure on the funding success or the company value, respectively. Yet Block et al. (2018a) find that the effect of updates on the success of equity crowdfunding campaigns depends on the content of the published information. Not all updates have a positive impact on the invested amount and the number of investments. Updates with verifiable and business-related information about the development of the start-up are most relevant, while the business model, team, and promotional activities evoke less interest among investors. Furthermore, the language of updates, i.e. the sentiment, can influence the perception of the investors. In our paper we reverse the research question of Block et al. (2018a). Instead of investigating the effect of updates on the funding success, we examine whether start-ups take into account these previously shown relationships and strategically post updates containing a specific language or content. To this end, we investigate the sentiment, the language, and the content of updates. First, we analyze changes in the communication behavior during and after the funding period. Second, we focus solely on the funding period and investigate which circumstances increase the likelihood of start-ups posting an update.

We use hand-collected data from two major German equity crowdfunding portals—Seedmatch and Companisto—to investigate the communication behavior of start-ups. Therefore, we use the data set of Block et al. (2018a) and expand it with further updates posted during[1] and after the funding period. Through analyzing the language and the content of 751 updates as well as 39,036 individual investment decisions, we find evidence that start-ups use updates during the funding period strategically. The frequency of updates is significantly higher over the course of the funding period than afterwards and start-ups use more linguistic devices that create a feeling of group cohesion and group identity. We also find some evidence for the hypothesis that start-ups strategically post updates with specific content during the funding period. Moreover, during the funding period the probability of an update increases along with strong competition of other contemporary equity crowdfunding campaigns.

Our study thus contributes to answering the question of whether start-ups rationally use investor communication to ensure successful funding and to what extent and in what way they change their communication behavior after the funding is ensured. While the answer to the first question could help to improve the entrepreneurial behavior in crowdfunding campaigns, the latter aspect may be important for both the decision making of investors and in the context of investor protection. Knowledge about the possibly strategic communication behavior of start-ups can help investors to optimize their investment decisions.All these issues are highly relevant for the continued development of the regulatory framework for equity crowdfunding.

The remainder of the paper is structured as follows. In Section 2 we describe our hypotheses regarding changes in the communication behavior of start-ups and the determinants of updates during the funding period. Section 3 provides an overview of the data set and the key variables. Section 4 presents descriptive statistics and analyzes the use of updates in equity crowdfunding. Section 5 concludes.

5.2 Theoretical foundation and hypotheses

In crowdfunding, updates are a form of voluntary disclosure for start-ups. There are several reasons why start-ups publish updates both during and after the funding period of the campaign, thereby informing (potential) investors about developments regarding the start-up.

[1]On Seedmatch, entrepreneurs can post updates on two different parts of the webpage. In contrast to Block et al. (2018a) we take into account both of these possibilities to publish updates. In this way, we make use of additional 80 updates during the funding period.

Generally, the managers of a company are assumed to have comparatively good knowledge of the firm value and the expected future performance of the company than investors. These information asymmetries between managers and shareholders can be reduced by providing additional information through voluntary disclosure. Furthermore, updates can serve as a signal of quality (Mollick, 2014; Block et al., 2018a). According to Spence (2002), signals can further reduce information asymmetries between the involved parties. Lower information asymmetries, can in turn, reduce the cost of capital for companies (Diamond and Verrecchia, 1991; Healy and Palepu, 2001; Merton, 1987). Rational entrepreneurs can therefore be expected to publish updates during the funding period of a campaign. Previous research shows that updates are indeed important for the funding success of a crowdfunding campaign (Xu et al., 2014; Mollick, 2014; Kuppuswamy and Bayus, 2017; Block et al., 2018a; Hornuf and Schwienbacher, 2017).

Subsequent to the funding period, communication with investors is rational as well. The reasons for participation in crowdfunding are diverse. Hamari et al. (2016) describe internal motivations such as perceived sustainability and enjoyment as well as external motivations including reputation and economic benefits. Therefore, monetary motives may not necessarily be the only reason either for entrepreneurs or for investors to engage in equity crowdfunding. Particularly in crowdfunding, the support and feedback of the crowd both in the development and promotion of products and services can be considered as being important for the future success of the start-up. If these non-monetary incentives play a role for investors and the start-up, we expect the entrepreneur to communicate with the investors both during and after the campaign.

In crowdfunding, no regulations concerning the form or the content of voluntary disclosure exist, and usually no third party verifies the published information. Therefore, entrepreneurs can easily make use of the signaling effect of updates and strategically post updates with specific content or language during the funding period. In this way, they can signal quality to investors and thereby gain investments. As the business development of the start-up is not different at any specific time during or directly after the funding period ends per se, the availability of the disclosable hard information should not significantly change during and after the funding period. Hence, if either the language or the content of updates significantly differs between these two periods, we conclude that start-ups strategically post updates to encourage investors. In the following, we derive several hypotheses regarding such strategic communication behavior of start-ups.

The financial disclosure literature indicates that an optimistic and positive tone of reports is associated with increased firm performance (Li, 2010; Davis et al., 2012; Henry, 2008). For example, Henry (2008) investigates the effect of language used in earnings press releases on the stock price. He shows that press releases written in a positive tone are associated with higher abnormal returns. The results remain stable even after controlling for the financial results of the company. Positivity is also closely linked to the concept of passion in the literature on entrepreneurship. Empirical evidence suggests that the optimism, passion, and self-confidence of an entrepreneur increase the likelihood of obtaining venture capital and indirectly raise the prospects of future growth (Baum and Locke, 2004; Cardon et al., 2009; Chen et al., 2009). Start-ups might prefer to use updates with a more positive tone during the funding period to show that they are passionate and optimistic. As the business development of start-ups should not be better in the funding period than afterwards per se, a more positive tone during the funding period suggests strategic communication behavior of start-ups.

Hypothesis 1: During the funding period, updates have a more positive tone than after the funding period.

Furthermore, Allison et al. (2013) use the warm-glow theory of Andreoni (1990) to explain funding success on Kiva, a crowdfunding platform for micro loans. The warm-glow theory suggests that individuals receive utility by helping others. By examining the credit applications of micro loans, Allison et al. (2013) show that credit applications containing linguistic devices that evoke warm-glow effects experience faster funding. Gerber and Hui (2013) find similar motives for other forms of crowdfunding. They point out that investors are motivated by the desire to help others and to be part of a community. By publishing updates with specific linguistic devices that evoke a feeling of cohesion start-ups may try to use this coherence. Using emotional language and the first person plural can create a feeling of group identity and improve the group cohesion (Zheng, 2000; Sexton and Helmreich, 2000; Tausczik and Pennebaker, 2010). Furthermore, using the past tense can create a psychological distance (Tausczik and Pennebaker, 2010) and therefore, we expect start-ups to strategically employ the first person plural, more frequently use emotional language, and increase the use of the present tense in updates during the funding period. Such communication behavior of start-ups tends to indicate a strategic use of language in updates since, on average, the disclosable hard information should not significantly change during and after the funding period.

Hypothesis 2: During the funding period, updates contain more linguistic devices that evoke a feeling of group cohesion than they do after the funding period.

As there are no rules concerning the content of updates in crowdfunding, start-ups can generally publish any type of information in updates. Yet it is not surprising that not all forms of updates promote the funding success of a campaign (Xu et al., 2014; Block et al., 2018a). In particular, updates informing about new developments of start-ups such as new funding sources, the development of the respective business, and updates containing information about cooperations increase the funding success within the funding period. By contrast, updates with information that was previously available such as information about the entrepreneurial team or the business model are not significantly associated with an increase in investments (Block et al., 2018a). If entrepreneurs wish to target the investment spirit, start-ups can be expected to publish disproportionately more updates disclosing information about new developments during the funding period than after the funding period. Again, as there should not per se be a higher density of these new developments within the funding period than later on, posting relatively fewer of these updates after the end of the campaign provides evidence of strategic communication behavior of start-ups.

Hypothesis 3: During the funding period, entrepreneurs publish more updates with information on new funding sources, the business development, and updates with information about cooperations of the start-up.

On most of the equity crowdfunding platforms, start-ups define a funding goal before the campaign begins. The funding goal represents the threshold of the invested amount of money the start-ups need to obtain to be successfully funded. Therefore, start-ups have a strong incentive to obtain investments worth a minimum of the amount of the funding goal. Hornuf and Schwienbacher (2017) highlight the L-shape of investments under a first-come-first-served mechanism in equity crowdfunding. Vulkan et al. (2016) find that the chances for successful funding decrease after the campaign has begun. Hence, start-ups that are almost at the end of the funding period and have attracted investments below the funding goal are expected to act promptly in order to gain more backers. They may post more updates to trigger the investments needed to reach the funding goal, even if the probability of disclosable information does not change over the course of the funding period.

Hypothesis 4: Start-ups are more likely to post an update when the funding goal of the campaign has not been achieved and the remaining funding period is short.

During the funding period, start-ups may also consider the competitive environment of their equity crowdfunding campaigns. Many parallel equity crowdfunding campaigns or so-called *blockbusters*, popular campaigns with an extremely large number of backers, may lure investors away from the focal crowdfunding campaign. When competition is strong, start-ups may be more likely to post an update to draw attention to their own campaign. However, previous research indicates that blockbusters not only accelerate investments in the focal campaign but also increase them in other crowdfunding campaigns (Kickstarter, 2012). This is because blockbusters usually enjoy extensive media coverage and new backers may be attracted to crowdfunding in general. With data from the reward-based crowdfunding portals Kickstarter and Indiegogo, Doshi (2016) shows that, on average, the invested volume increases in the blockbusters' project category. Depending on the project category, blockbuster can also create spill-over effects to other project categories. Darrough and Stoughton (1990) analyze voluntary disclosure in competitive markets. They highlight that under some assumptions such as low entry costs to the market, strong competition favors voluntary disclosure to deter the entry of competitors. In the context of equity crowdfunding, Hornuf and Schwienbacher (2017) and Block et al. (2018a) find a positive relationship between a strong competition of campaigns and the funding success of a particular campaign. Overall, the probability of disclosing voluntary information in the form of updates can be expected to increase in a highly competitive environment. Since the available disclosable hard information should not depend on the competitive environment, such communication would again suggest a strategic behavior of start-ups.

Hypothesis 5: Start-ups are more likely to post an update when the number of competing investments in contemporary equity crowdfunding campaigns is high.

5.3 Data

5.3.1 Data sources

For the empirical analysis we hand-collect data from two German equity crowdfunding portals—Seedmatch and Companisto—during the period from June 7 2012 to April 27 2015. The portals Seedmatch and Companisto are the market leaders for equity crowdfunding in Germany and account for around 75% of the total equity crowdfunding capital raised in Germany during the observation period. We obtain all data directly from the platforms. Habitually, start-ups do not only use equity crowdfunding portals to post their updates but also publish the information on social media platforms or in newsletters. After the campaign, the equity crowdfunding portals retain a page with a project overview as well as all key characteristics of the campaign and the possibility to post updates. As start-ups also seek visibility when the campaign has concluded, we expect the start-ups still to use all communication channels including the equity crowdfunding portals, in order to post their updates. For the further analysis we use two different data sets.

To analyze changes in the communication behavior of start-ups we focus on the updates posted during and after the funding period and examine all campaigns run on Seedmatch and Companisto that include at least one update. In total, our first data set (updates data set) includes 751 updates of 97 equity crowdfunding campaigns. With 64 campaigns the majority of the 97 campaigns were run on Seedmatch. Yet start-ups running equity crowdfunding campaigns on Companisto appear to post more updates. Approximately 52% of the updates in our data set were posted on this portal. Several start-ups run multiple equity crowdfunding campaigns, hence the 97 campaigns belong to 88 unique start-ups. Most of these start-ups operate either in the information & communication or in the wholesale & retail sectors.

We additionally obtained a second data set with daily investment data for 71 campaigns (investment data set) to further investigate the determinants of updates during the funding period. Investment data refers to the daily investments of all backers as well as the total invested amount on each campaign day. We were able to retrieve investment data for 26,456 investments belonging to the entire 36 campaigns on Companisto. We also retrieved daily investment data for 12,580 investments and 35 campaigns on Seedmatch. Seedmatch removes all investment data from the website once the funding has been completed and hence the amounts invested by individual backers are no longer publicly available. Due to this limited availability of data we could not include all campaigns run on Seedmatch in the investment data set. Importantly, during the funding period only 57 campaigns include updates, which were also considered in the updates data set. We also obtain investor data for 14 campaigns that refrained from posting any updates during the funding period. Overall, eight start-ups ran multiple equity crowdfunding campaigns; thus the 71 campaigns belong to 63 unique start-ups. In a final step, as in Kuppuswamy and Bayus (2017) and Hornuf and Schwienbacher (2017), we construct a panel data set in which the time dimension is equal to the days of the campaign and the cross-sectional dimension is represented by the campaigns. The investment data set contains 5,176 campaign days and 314 updates posted on these days.

5.3.2 Dependent variables and key explanatory variables

To test our hypotheses, we define different dependent variables. For each day of the funding period, we identify whether the start-up posted an update or not (*Update*). Furthermore, we consider all updates posted during and after the funding period and examine the content and the language of these updates. We apply a coding process to examine the information contained in the updates. Following Block et al. (2018a), we use nine categories to describe the content of the updates: *Team*, *BusinessModel*, *Certification*, *Product*, *Cooperation*, *Campaign*, *NewFunding*, *Business*, and *Promotions*. A detailed description of all the categories is included in Table 5.1. The categories are not mutually exclusive; different categories can apply to one update. Furthermore, updates without relevant content are not included in any of the categories.

Table 5.1 List and Definition of all Variables.
The data is retrieved from the German equity crowdfunding portals Seedmatch and Companisto.

Variable	Description
	Updates and Update Categories
Update	Dummy variable equal to 1 if the start-up publishes an update on day *t*, and 0 otherwise.
Business	Dummy variable equal to 1 if the update on day *t* discloses information about the customers or financials (e.g. number of customers, amount of sales), and 0 otherwise.
BusinessModel	Dummy variable equal to 1 if the update on day *t* discloses information about the business model, the relevant market or future plans and strategies, and 0 otherwise.
Campaign	Dummy variable equal to 1 if the update on day *t* discloses information on the funding of the campaign (e.g. number of investors, archived funding amount, change of funding limit), and 0 otherwise.
Certification	Dummy variable equal to 1 if the update on day *t* discloses information on external certification of the company or product (e.g. press coverings, awards, patents), and 0 otherwise.
Cooperation	Dummy variable equal to 1 if the update on day *t* discloses information on cooperation projects or collaborations of the start-up, and 0 otherwise.
Emotional	Dummy variable equal to 1 if the update on day *t* contains emotional language, and 0 otherwise.
NewFunding	Dummy variable equal to 1 if the update on day *t* discloses information on additional funding sources of the start-up such as business angels, venture capitals or government grants, and 0 otherwise.
Product	Dummy variable equal to 1 if the update on day *t* discloses information on the product or the product development, and 0 otherwise.
Promotions	Dummy variable equal to 1 if the update on day *t* discloses information about promotions for the crowd (discounts, rewards), invites the crowd to participate on events or appeals to the crowd to support the start-up (e.g. recommendations and network), and 0 otherwise.
Team	Dummy variable equal to 1 if the update on day *t* discloses information about the entrepreneurial team (e.g. work experience, age and education), and 0 otherwise.
	Sentiment and Language of Updates

Table 5.1 continued.

Variable	Description
Negative	Percentage of words that evoke negative emotions within the update text (e.g. hurt, ugly and nasty). Obtained by the software LIWC.
Past	Percentage of words that refer to the past within update text (e.g. went, had and ran). Obtained by the software LIWC.
Positive	Percentage of words that evoke positive emotions within the update text (e.g. love, nice and sweet). Obtained by the software LIWC.
WC	The total number of words that appear in the update text.
We	Percentage of words that refer to first person plural within the update text (e.g. we, us and our). Obtained by the software LIWC.
	Timing of Updates and Competitive Environment
Alarm	Dummy variable equal to 1 if the funding goal is not reached and more than three quarters of the funding period have passed or three quarters of the extended funding period, and 0 otherwise.
FIN	Dummy variable equal to 1 if the update is published during the funding period, and 0 otherwise.
Interval	Time interval between the publications of updates in a particular campaign, in days.
#Investments	Total number of all investments made on day *t* across all campaigns on three major and one minor German equity crowdfunding portal (Companisto, Seedmatch, Innovestment, and United Equity).
Time	Total number of days passed from the start of the campaign before publishing the first update. Updates on the first campaign day are either considered (subsample 2) or not (subsample 1).
Update1Day	Dummy variable equal to 1 if an update is published on the first day of the campaign, and 0 otherwise.
	Controls
Amount	Total amount of money invested by the crowd until day *t* in a particular campaign, in Euro.
FundingGoal	The minimum funding goal as defined by the start-up and the portal on day 0, in Euro.
EquityShare	Funding Goal over pre-money valuation.
Industry	Dummy variables for the industry in which the start-up operates in, either information & communication; wholesale & retail; manufacturing; professional, scientific & technical activities; financial & insurance activities or accommodation & food service activities.
%Invested	*Amount* over funding goal at day *t* in a particular campaign.
PostFunded	Dummy variable equal to 1 if the invested sum of money of the campaign has exceeded the funding goal on day *t* in a particular campaign, and 0 otherwise.
Portal	Dummy variable equal to 1 if the campaign is run on the portal Companisto, and 0 otherwise.
VDAX	Volatility index on the German stock index DAX on day *t*. Source: Datastream.

To ensure the reliability of our coding scheme, a second, independent researcher rated the updates. At first, we provide the second researcher with a coding manual containing a detailed description of each category. The researcher rated approximately 20% of the updates. In a following discussion, we adapt our coding scheme and come up with the final description of the ten categories. Thereafter, both raters coded all updates again (Reis and Judd, 2014). To measure the inter-rater agreement, we calculate the Cohen's Kappa coefficient (Cohen, 1960; Fleiss et al., 2003). Over all categories we have a Cohen's Kappa of 0.85. Depending on the category, the inter-rater reliability ranges from 0.77 to 0.94 indicating excellent agreement[2] between the two raters (Landis and Koch, 1977).

To further evaluate the sentiment and the language of the updates, we use the text analysis software Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001; Wolf et al., 2008). LIWC counts the words in the updates and compares them with dictionaries of different linguistic and psychological categories (for example positive or negative emotions). The software calculates the percentage of total words for each category. Thus we can measure the sentiment of the updates (*Positive* and *Negative*) and the usage of the past tense (*Past*) as well as the usage of first person plural (*We*).

In general, the start-ups have 60 days to gain enough investments to reach their funding goals (*funding period*) subsequent to the publication of the campaign on the crowdfunding platform. However, for each campaign the start-ups can extend the funding period one time only for another 60 days (Klöhn and Hornuf, 2012). To investigate changes in the communication behavior, we derive the variable *FIN*.

By using daily investment data, we define several key explanatory variables. We measure the success of a campaign using two different proxies. On the one hand, we create the dummy *Alarm*. *Alarm* accounts for the start-ups that urgently require further investments, in the sense that the hitherto invested amount has not yet reached the funding goal and the remaining time of the funding period is short. On the other hand, we use the variable *Amount*. Moreover, we measure the competitive environment of a campaign (*#Investments*).

We also include several further control variables based on prior research. Hornuf and Schwienbacher (2017) show that investments in equity crowdfunding decrease under a first-come-first-served mechanism once the funding goal has been surpassed. Therefore, we include a dummy variable *PostFunded*. In another paper, Hornuf and Neuenkirch (2017) demonstrate that a high level of stock market volatility is associated with higher premia for the equity crowdfunding portal Innovestment. The authors conclude that equity crowdfunding is a substitutional, as opposed to a supplementary asset class, when stock markets are volatile. Thus we also include the German VDax (*VDAX*) as a control variable. To capture portal-specific effects, we include a dummy variable for the equity crowdfunding portal Companisto (*Portal*). Finally, we control for the industry of the start-up, the year, and the day of the week (see, for example, Vismara (2016);

[2] According to Landis and Koch (1977) a Cohen's Kappa between 0.61 and 0.8 indicate substantial agreement, values above 0.81 indicate almost perfect agreement.

Block et al. (2018a); Hornuf and Neuenkirch (2017)). A description of all variables is presented in Table 5.1.

5.4 Results

5.4.1 Summary Statistics

Table 5.2 presents summary statistics for the updates data set. The majority of the 751 updates is published during the funding period. However, we also consider 299 updates that are subsequently posted.[3] The bulk of the updates discloses information on promotions of the start-ups and / or describes the business model. By contrast, only few updates contain emotional language and disclose information either about the entrepreneurial team or new funding sources. Some start-ups use updates extensively to communicate with their investors. In total, the start-up *Riboxx* posted 29 updates since its campaign start in July 2014. On average, 33 days pass before a subsequent update is posted in a particular campaign. However, the length of this interval differs between the two portals. On Companisto an average of 28 days pass between the posting of an update. On Seedmatch though, this interval is, on average, 39 days. The length of the updates varies considerably as well. The shortest update only consists of one word ("Danke", meaning *thanks*) while the longest contains 1,293 words. Furthermore, the updates employ a relatively positive tone. Approximately 3.9% of the words are positive and, by contrast, only around 0.3% are negative.

Summary statistics for the investment data set are shown in Table 5.3. More than 80% of the campaigns have at least one update during the funding period. On average, a start-up posts 4 updates during this time. However, the number of updates differs between the campaigns. Some start-ups refrain from posting a single update while others use this tool for communication extensively. For example, the start-up *MyParfume* posted 14 updates during the funding period. Yet, the campaign length of *MyParfume* is above the average of 72 days (123 days).

As soon as the campaign becomes active and backers have the possibility to invest, start-ups are able to communicate with their investors via updates. Most of the start-ups post their first update at the beginning of the funding period (see Figure 5.1). Several start-ups even post updates on the very first day of the campaign. These updates are rarely linked to the progress of the

[3] In our data set, the average funding period is with 72 days considerably shorter than the average period following successful funding (573 days).

Table 5.2 Summary Statistics Updates Data Set.
97 campaigns. All variables are defined in Table 5.1. Corr denotes the pairwise Bravais-Pearson Correlation Coefficients with *FIN*.

UPDATES DATA SET							
Binary Variables	Yes	Mean	Median	SD	# Obs.	Corr	
FIN	452	0.602	1	0.4898	751		
Business	184	0.245	0	0.4303	751	0.1597	
BusinessModel	345	0.155	0	0.3629	751	0.0292	
Campaign	143	0.190	0	0.3928	751	0.2420	
Certification	283	0.376	0	0.4849	751	−0.1141	
Cooperation	170	0.226	0	0.4187	751	0.0174	
Emotional	117	0.156	0	0.3629	751	0.1019	
NewFunding	51	0.067	0	0.2517	751	0.0574	
Product	292	0.388	0	0.4878	751	0.4878	
Promotions	347	0.462	0	0.4988	751	−0.1247	
Team	87	0.116	0	0.3203	751	−0.0201	
Metric Variables	Mean	Median	SD	Min.	Max.	# Obs.	Corr
Positive (in %)	3.981	3.54	4.0948	0.00	100.00	751	0.0094
Negative (in %)	0.262	0	0.4700	0.00	4.26	751	−0.0153
We (in %)	3.988	3.87	2.4807	0.00	26.67	751	0.0643
Past (in %)	1.524	1.34	1.2019	0.00	8.00	751	−0.0910
Interval	32.882	16	67.8472	0.00	662.00	650	−0.3532
WC	256.163	222	176.9025	1.00	1,293.00	751	−0.0272

campaign. As described in Mollick (2014), start-ups may strategically post updates soon after the campaign commencement to show that they are well prepared for the campaign and thus indicate a high campaign quality.

The majority of the equity crowdfunding campaigns managed to reach their funding goal quickly, but, 6 campaigns were not able to achieve the funding goal before three quarters of the funding period had elapsed. Overall, 47 investments were made on an average campaign day. By comparison, an average of 7.56 investments were made each day in a particular campaign.

5.4.2 Univariate Analysis: Changes in communication behavior after the funding period

To investigate modifications in the communication behavior during and after the funding period, we apply a univariate analysis. As we observe several updates per campaign, we have to consider the correlation between updates within the same campaign. For the continuous dependent

Table 5.3 Summary Statistics Investment Data Set.
71 campaigns. All variables are defined in Table 5.1. Corr denotes the pairwise Bravais-Pearson Correlation Coefficients with *Update* and *Time*, respectively.

INVESTMENT DATA SET							
Binary Variables	Yes	Mean	Median	SD	# Obs.		
Update	314	0.061	0	0.2387	5,176		
Business	94	0.018	0	0.1335	5,176		
BusinessModel	156	0.030	0	0.1709	5,176		
Campaign	83	0.016	0	0.1256	5,176		
Certification	106	0.020	0	0.1416	5,176		
Cooperation	82	0.015	0	0.1248	5,176		
Emotional	62	0.012	0	0.1088	5,176		
NewFunding	23	0.004	0	0.0665	5,176		
Product	136	0.026	0	0.1599	5,167		
Promotions	146	0.028	0	0.1655	5,176		
Team	45	0.009	0	0.0928	5,176		
Metric Variables	Mean	Median	SD	Min.	Max.	# Obs.	Corr
Alarm	0.02	0	0.1349	0	1	5,176	−0.0049
#Investments	47	30	72.1444	0	1160	5,176	0.0586
Amount	497,352	141,500	1,254,637	1,260	7,497,250	5,176	−0.0557
FundingGoal	112,459	50,000	211,229	25,000	1,000,000	5,176	−0.0643
VDAX	18.20	17.63	3.2100	12.70	32.08	5,176	−0.0120
Portal	0.55	1	0.4977	0	1	5,176	0.0272
PostFunded	0.86	1	0.3489	0	1	5,176	0.0059
EquityShare	0.02	0.02	0.0246	0.0045	0.23	5,176	0.0040
%Invested	4.51	2.99	4.1910	0.0075	20	5,176	−0.0010
%Invested1Day	1.5876	0.7987	2.3098	0.0075	14.9975	71	−0.1948
Update1Day	0.2535	0	0.4381	0	1	71	−0.3133

variables, we use a Feasible Generalized Least Squares (FGLS) estimation. According to Cameron and Miller (2015), a FGLS estimator can lead to efficiency gains compared with OLS when accounting for dependencies within groups. We perform a modified Hausman test and, in case the Hausman test leads us to dismiss the random effects estimator, we apply fixed effects. Otherwise we retain random effects. For binary dependent variables (i.e. the update categories), we use a probit regression with standard errors clustered at campaign level. To test whether there are differences between the funding period and the subsequent time, we use a dummy for the funding period (*FIN*) as an explanatory variable. In case the coefficient of this dummy is significantly different from zero with a positive (negative) value, significantly more (less) updates of this category are posted within the funding period. Table 5.4 and 5.5 present the results.

We find several significant changes in the communication behavior of start-ups over time. To

Table 5.4 Regression Results Funding Period, FGLS-Estimation.
This table reports upon regression results using the Updates Data Set and random-effects model (dependent variables: *Positive*, *Negative*, *Interval*, *WC*) and fixed-effects model (dependent variables: *We*, *Past*). Heteroskedasticity-robust standard errors are shown in parentheses. For regression models with random-effects the overall-R^2 and for those with fixed-effects the Within-R^2 are shown. ** and *** denote significance at a 5%- and 1%-level.

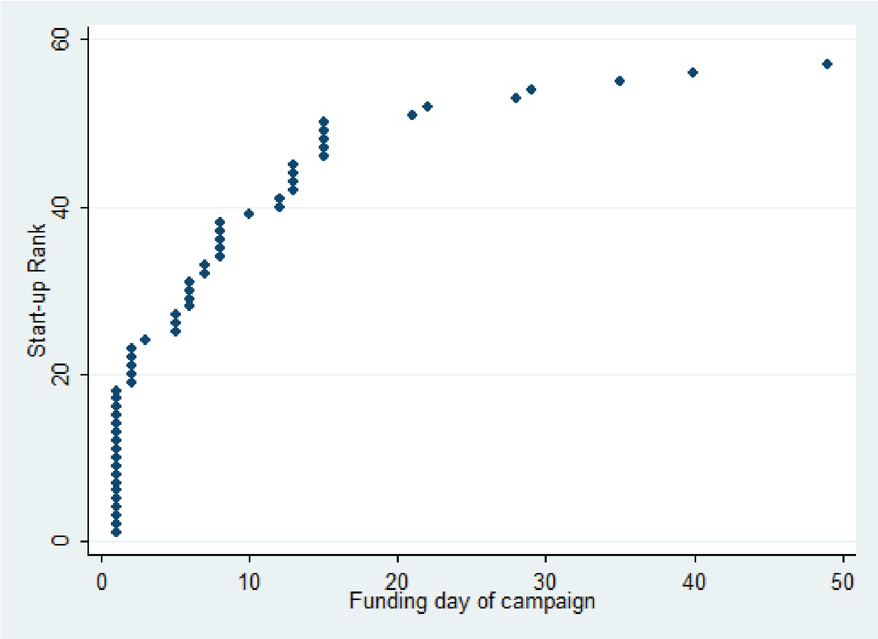
	Hypothesis 1		Hypothesis 2		Additional	
	Positive	*Negative*	*We*	*Past*	*Interval*	*WC*
FIN	0.0784	−0.0635	0.8012***	−0.3879**	−55.8470***	−5.7234
	(0.3662)	(0.0535)	(0.2519)	(0.1518)	(8.7571)	(14.6006)
Constant	3.9339***	0.3250***	3.5056***	1.7572***	75.8570***	250.3666***
	(0.2440)	(0.0552)	(0.1516)	(0.0913)	(9.7145)	(15.2357)
# Obs.	751	751	751	751	650	751
R^2	0.0001	0.0002	0.0187	0.0163	0.1248	0.0007

Table 5.5 Marginal Effects Funding Period, Probit-Estimation.
This table reports upon average marginal effects using the Updates Data Set and probit regressions. Cluster- and heteroskedasticity-robust standard errors are shown in parentheses. *, **, and *** denote significance at a 10%-, 5%-, and 1%-level.

	Hypothesis 2	Hypothesis 3			
	Emotional	*Cooperation*	*NewFunding*	*Business*	*Promotions*
FIN	0.0782**	0.0150	0.0307	0.1442***	−0.1256**
	(0.0342)	(0.0421)	(0.0233)	(0.0484)	(0.0520)
# Obs.	751	751	751	751	751
$Pseudo-R^2$	0.0124	0.0003	0.0069	0.0238	0.0113

	Additional				
	Team	*BusinessModel*	*Certification*	*Product*	*Campaign*
FIN	−0.0130	0.0298	−0.1112**	0.0237	0.2105***
	(0.0295)	(0.0442)	(0.0565)	(0.0480)	(0.0341)
# Obs.	751	751	751	751	751
$Pseudo-R^2$	0.0006	0.0006	0.0098	0.0004	0.0668

Figure 5.1 Time to first Update
Number of days until the first update is published. Investment Data Set.



begin with, the frequency of updates differs significantly between the funding period and the period thereafter. During the funding period on average 56 days less go by than after the funding period until a subsequent update is published. This result indicates that for many start-ups, obtaining funding is indeed the primary goal of an equity crowdfunding campaign. Yet since entrepreneurs continue to communicate with investors after the successful funding, non-monetary motivations play a role in equity crowdfunding as well.

The sentiment of the updates is not significantly different between the funding period and the period thereafter. The updates contain neither less positive nor more negative words once the funding has been completed. Hence we find no evidence to support our first hypothesis that start-ups use a positive tone in updates during the funding period in order to encourage investors. However, the results suggest that start-ups use different devices to the sentiment of the update to reach out to the crowd. We observe a significant positive relationship between the funding period and updates that use emotional language (*Emotional*). Furthermore, the updates during the funding period contain significantly more first person plural and less past tenses than updates subsequent to the funding period. The latter relationship may exist due to the fact that most start-ups run equity crowdfunding campaigns to obtain seed finance. Many of these start-ups started their businesses recently and may not have had past events to report upon in the equity crowdfunding campaign. Overall, the results support our second hypothesis, which is that updates during the funding period contain more linguistic devices evoking a feeling of group cohesion and improving group identity than updates posted after the end of the campaign.

We also investigate whether the usage of updates with a specific content differs between the funding period and the period thereafter. Since the latter period is, on average, longer than the funding period, overall more disclosable information should be available in the later period. However, two update categories, namely *Business* and *Campaign* have a significant positive relationship with *FIN*. The fact that significantly more updates containing information about the development of the businesses are published during the funding period represents evidence in favor of our third hypothesis, being that entrepreneurs strategically use updates about new developments of the start-up. The positive relationship between *FIN* and *Campaign* is not surprising, either. This effect is driven by the fact that start-ups post more information about the campaign progress, such as the achieved funding amount or the number of backers on a particular day, during the funding period than after the successful funding. The two other categories that we hypothesize are posted more often during the funding period, *NewFunding* and *Cooperation*, are not significantly associated with *FIN* in our analysis. This is possibly due to the fact that start-ups indeed require relevant hard information to be able to publish updates in these categories. It is less difficult, for example, to publish easily obtainable sales figures (*Business*) in the funding period than information about a new strategic cooperation that may simply not exist.

Overall, we find some evidence in favor for our third hypothesis. One out of three categories, namely *Business*, shows the expected correlation with the funding period. Therefore, the data weakly supports our third hypothesis. Indeed the different results for *Business*, *NewFunding* and *Cooperation* indicate that start-ups strategically change the content of updates during the funding period and thereafter.

We also find that start-ups post significantly fewer updates about external certification and promotions during the funding period than thereafter. In many cases, the start-ups do not have a fully developed product at the time of the equity crowdfunding campaign. Therefore, many start-ups are not able to post updates about external certification during the funding period. Furthermore, the funding period is shorter than the following period. Due to this extended time period, the probability of a disclosable hard information rises.

Block et al. (2018a) point out that the length of the update text is not significantly associated with investments. In line with this result, we do not find any evidence to suggest that updates during the funding period contain more words than subsequent updates.

5.4.3 Multivariate Analysis: Communication dynamics during the funding period

Why do entrepreneurs post an update on a specific campaign day? To answer this question, we estimate several statistical models. Our dependent variables are binary and equal to one if an update or an update of a specific category is posted on a particular campaign day and zero otherwise. We begin with panel models and apply a Hausman test. We have to dismiss the random-effects model as it is inconsistent for our data. However, the fixed-effects logit model only uses variation within the campaign and therefore implies heavy losses of observations depending on the update category. Furthermore, coefficients for time-invariant regressors cannot be estimated. Thus we use a pooled probit regression as a main model and include the fixed-effects model as a robustness check. Table 5.6 presents the results for the pooled probit with 'posting of an update' and 'posting of an update with a specific content' as dependent variables, respectively.

In a first step, we examine whether start-ups are more likely to post updates when they have not reached the funding goal and the remaining funding period is short. We find that effects differ for the update categories. While we observe a significant positive relationship between the *Alarm* dummy and emotional updates as well as those updates which disclose information about the business and campaign development, all other categories are insignificant. We cannot estimate average marginal effects for updates about new funding sources and the entrepreneurial team as these updates are never posted when the *Alarm* dummy equals one. The probability of an update increases for the significant categories, *Campaign*, *Business*, and *Emotional*, by between 1.7% and 2.8% when the *Alarm* dummy equals one. On the one hand, the significant positive effect of the *Alarm* dummy on emotional updates and those disclosing information on the business development suggest strategic communication behavior of start-ups. Since the availability of disclosable information should not change over the course of the campaign, a significant change in the communication behavior indicates a strategic posting of updates. On the other hand, *NewFunding* and *Cooperation*, the two other categories that increase investments according to Block et al. (2018a) are not significant in our data. However, this may again be due to the fact that start-ups need disclosable information in order to publish updates within these categories. Overall, we only find weak evidence to support our forth hypothesis which is that start-ups are more likely to post an update when the funding goal is not reached and the remaining funding period is short.

In a second step, we focus on the competitive environment of equity crowdfunding campaigns. We observe a significant positive relationship between the total number of investments in equity crowdfunding campaigns on the overall market during the previous day and the probability

Table 5.6 Probability of Updates, Pooled Probit Estimation.
This table reports upon average marginal effects of the pooled probit regression using the investment data set. The natural logarithm of *Amount* is used and *#Investments* is displayed in 1,000. Cluster- and heteroskedasticity-robust standard errors are shown in parentheses. *, **, and *** denote significance at a 10%-, 5%-, and 1%-level, respectively.

	Update	*Team*	*BusinessModel*	*Certification*	*Product*	*Cooperation*	*Campaign*	*NewFunding*	*Business*	*Promotions*	*Emotional*
Alarm	0.0160		0.0207	0.0096	−0.0168	0.0065	0.0166*		0.0280***	0.0151	0.0275*
	(0.0220)		(0.0190)	(0.0169)	(0.0143)	(0.0060)	(0.0100)		(0.0075)	(0.0132)	(0.0149)
$\#Investments_{t-1}$	0.1147***	0.0285***	0.0378*	−0.0020	0.0534***	0.0320**	0.0219	0.0092	0.0271	0.0663***	0.0306**
	(0.0335)	(0.0094)	(0.0227)	(0.0268)	(0.0186)	(0.0141)	(0.0151)	(0.0059)	(0.0197)	(0.0221)	(0.0125)
$\ln Amount_{t-1}$	−0.0011	−0.0020	−0.0016	0.0003	−0.0001	−0.0000	0.0010	−0.0001	0.0009	−0.0007	−0.0006
	(0.0032)	(0.0017)	(0.0018)	(0.0018)	(0.0022)	(0.0014)	(0.0018)	(0.0005)	(0.0015)	(0.0028)	(0.0013)
VDAX	0.0009	−0.0006	−0.0013	0.0002	0.0003	−0.0001	0.0007	−0.0004	−0.0001	0.0001	−0.0004
	(0.0012)	(0.0005)	(0.0011)	(0.0008)	(0.0008)	(0.0010)	(0.0006)	(0.0003)	(0.0007)	(0.0008)	(0.0007)
PostFunded	−0.0017	−0.0103**	−0.0054	0.0180**	−0.0036	−0.0120***	0.0093	0.0012	0.0005	0.0031	0.0220**
	(0.0094)	(0.0049)	(0.0071)	(0.0090)	(0.0071)	(0.0046)	(0.0059)	(0.0032)	(0.0068)	(0.0074)	(0.0098)
Portal	0.0041	0.0168***	0.0072	0.0072	−0.0023	0.0161***	−0.0148***	−0.0005	0.0036	0.0024	−0.0055
	(0.0121)	(0.0055)	(0.0075)	(0.0064)	(0.0069)	(0.0062)	(0.0050)	(0.0023)	(0.0055)	(0.0066)	(0.0051)
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-Week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	5,174	5,174	5,174	5,174	5,174	5,174	5,174	5,174	5,174	5,174	5,174
$Pseudo-R^2$	0.0597	0.1362	0.0527	0.0673	0.0618	0.0834	0.0678	0.0769	0.0346	0.0757	0.1013

of an update in the focal campaign. An increase in the total number of investments by 1,000 is associated with an 11.47% increase in the probability of an update. In times of a highly competitive environment, start-ups therefore attempt to draw attention to their campaigns by posting updates, thereby attracting backers. This relationship also holds for most of the update categories. Updates of the categories *Team*, *BusinessModel*, *Product*, *Cooperation*, *Promotions* and *Emotional* are significantly positively associated with the total number of investments on the market. Overall, the results therefore support our fifth hypothesis which is that the likelihood of an update increases along with market competition.

Our second proxy for the campaign success, the amount invested prior to the previous day, is not significantly related to the probability of an update. With respect to the other control variables, we observe a significant relationship between the probability of an update and the ability to reach the funding goal (*PostFunded*) as well as the *VDAX* for some categories. The portal on which the equity crowdfunding campaign is run also plays a role for some of the update categories. The sign, however, differs between the categories under consideration. While significantly more updates about the entrepreneurial team and collaborations of the start-up are posted on Companisto, significantly less updates are disclosed concerning the campaign development.

To consider endogeneity on a campaign level, we perform a fixed-effects logit regression. The results are presented in Table 5.9. They show a significant positive relationship between the probability of an update of the *Business* category and the *Alarm* dummy. Furthermore, we can confirm the previous results regarding the significant positive impact of competing investments on updates in general and on those that disclose information about *Team*, *Product*, *Cooperation*, *Business*, *Promotions*, and *Emotional* in particular. In this way, we show that our main results are not driven by unobserved time-invariant variables.

As an alternative model, we apply survival analysis and perform a Cox proportional hazard model with the number of days before the update is posted as a dependent variable. By using this model we are able to analyze the duration, i.e. the time that elapses before an update (or an update with a particular content) is published considering various covariates. The Cox model applies a semi-parametric method to estimate the impact of the covariates on the hazard rate. In this context, the hazard rate represents the chance of an update being published on the next day when taking into consideration the time period that has already passed. As we have so-called multiple-failure data, i.e. each campaign can exhibit more than one update, we cluster the standard errors at campaign level. The results are shown in Table 5.7. In this analysis we report on hazard ratios, which can be interpreted as semi-elasticity or multiplicative effect.

The results are similar to those of the pooled-probit model. We can confirm the positive

Table 5.7 Probability of Updates, Cox Proportional Hazard Model.
The table reports upon hazard rates of the Cox Proportional Hazard Model using the investment data set. Dependent variable is defined by the duration to an update (or update category) in days. ln*Amount·t* and *#Investments·t* are interaction terms between the explanatory variables ln*Amount* and *#Investments* and the time passed. The natural logarithm of *Amount* is used and *#Investments* is denoted in 100. Cluster-robust standard errors are shown in parentheses. *, **, and *** denote significance at a 10%-, 5%-, and 1%-level.

	Update	*Team*	*BusinessModel*	*Certification*	*Product*	*Cooperation*	*Campaign*	*NewFunding*	*Business*	*Promotions*	*Emotional*
Alarm	0.7371		1.4153	1.0272	0.3588	0.6588	1.6056		2.4031*	1.0877	9.9387*
	(0.1773)		(0.6158)	(1.0164)	(0.2407)	(0.4654)	(0.7668)		(1.1077)	(0.4623)	(12.4366)
#Investments	1.5872***	2.4810***	1.8896***	1.0648	1.9002***	1.1273	1.1086	1.3210	2.1069***	0.9903	1.7229***
	(0.1703)	(0.4206)	(0.1965)	(0.2222)	(0.1979)	(0.2592)	(0.2548)	(0.3412)	(0.2833)	(0.1664)	(0.2084)
#Investments·t	0.9975***	0.9965***	0.9976***		0.9967***				0.9968***		0.9958***
	(0.0005)	(0.0010)	(0.0006)		(0.0008)				(0.0009)		(0.0012)
ln*Amount*	0.6112*	0.1543***	0.3343***	0.2029***	0.2794***	0.2269***	0.7969	0.2829	0.2522***	0.4341***	0.9095
	(0.1636)	(0.0567)	(0.0732)	(0.0461)	(0.0627)	(0.0583)	(0.1721)	(0.2463)	(0.0889)	(0.0866)	(0.2529)
ln*Amount·t*	0.9956**										
	(0.0019)										
PostFunded	0.4278***	0.7108	0.5201	1.6949	0.6052	0.2565**	0.3364***	1.0156	0.7186	0.5493	2.0761
	(0.1239)	(0.3700)	(0.2118)	(1.2275)	(0.2436)	(0.1509)	(0.1337)	(1.1181)	(0.4672)	(0.2142)	(2.1669)
VDAX	1.0075	0.9891	0.9488	1.0018	0.9681	0.9890	0.9895	0.9556	0.9969	0.9697	0.9681
	(0.0294)	(0.0836)	(0.0376)	(0.0485)	(0.0357)	(0.0672)	(0.0535)	(0.1356)	(0.0441)	(0.0392)	(0.0651)
Portal	1.7839**	20.9530***	1.6994	4.0142***	1.8455**	5.6336***	0.4938**	0.8727	1.9614	1.5112	0.8261
	(0.4483)	(13.8625)	(0.5835)	(1.8277)	(0.4619)	(2.0697)	(0.1626)	(1.4189)	(1.1101)	(0.4483)	(0.3156)
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-Week	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# Obs.	380	380	380	380	380	380	380	380	380	380	380

relationship between updates which disclose information about the business development as well as emotional updates and the *Alarm* dummy. Furthermore, we find a positive relationship between the total number of investments and the probability of an update being posted for most of the update categories. We test the proportionality assumption of the Cox model for all explanatory variables. In case the assumption is violated, we include an interaction term of the explanatory variable with time (t). The interaction term *#Investments·t* indicates that the effect of competition of contemporary equity crowdfunding campaigns is not constant but decreases over time both for updates in general and for those that disclose information about the entrepreneurial team, the business model, the product, the business development, as well as emotional updates.

Using the Cox proportional hazard model, our second proxy for the success of the campaign, *Amount*, is significantly negatively associated with the probability of an update. Start-ups with a lower amount of funding are more likely to post an update. This result provides further evidence of the strategic communication behavior of start-ups. Again, the interaction term between *Amount* and the time period suggests a decreasing effect of *Amount* on the probability of an update over time.

5.4.4 Further Analysis

Colombo et al. (2015), Kuppuswamy and Bayus (2017), and Vulkan et al. (2016) highlight the fact that collective attention at the beginning of the campaign is crucial: crowdfunding campaigns that attract investors in the early phase of the funding period are significantly more successful. Our descriptive analysis of the data has also shown that start-ups tend to post updates soon after the campaign starts. In order to analyze the communication behavior of the first stage of the funding period in more detail, we consider the duration before the first update is posted. As a main model, we use a Cox proportional hazard model. An advantage of the survival analysis in this context is that we deal with right censoring. We do not only consider the campaigns with a first update but also those campaigns that did not post an update during the funding period. The results are presented in column 1 and 3 of Table 5.8. Furthermore, we apply a negative binomial model to investigate the number of days before the first update is published. Columns 2 and 4 in Table 5.8 show the results for the negative binomial estimations.

The updates posted on the very first day of a campaign are not usually linked to the progress of the campaign. Hence we use two different subsamples: one in which we omit updates posted on the first day (model 1 and 2) and one in which we include these updates (model 3 and 4).

Table 5.8 Time to first Update.
This table reports upon hazard rates of a Cox Proportional Hazard Model and marginal effects of a negative binomial regression using the investment data set. The dependent variable constitutes the time elapsed before the first update is published (*Time*). Model 1 and 2 do not consider updates posted on the very first day of campaign, models 3 and 4 include updates posted on the first day. In models 1 and 3 we estimate a Cox Proportional Hazard Model in models 3 and 4 a negative binomial regression. *FundingGoal* is denoted in 10,000 EUR. *, **, and *** denote significance at a 10%-, 5%-, and 1%-level.

	Model 1	Model 2	Model 3	Model 4
$\#Investments_{t=1}$	1.0019	−0.0270		
	(0.0031)	(0.0223)		
$\%Invested_{t=1}$	0.8957	−0.6231		
	(0.1824)	(1.4694)		
Update1Day	2.3052*	−1.2878		
	(1.0285)	(2.9089)		
Portal	1.1230	−4.1340	2.2760**	−10.2619***
	(0.5329)	(3.1047)	(0.9069)	(3.6149)
EquityShare	224.4835	−24.4252	886.9319	−62.9241
	(1335.9686)	(47.3784)	(5361.6060)	(56.9074)
FundingGoal	224.4835	0.1046	0.9704	0.4514
	(1335.9686)	(0.3569)	(0.0436)	(0.4271)
Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes
# Obs.	71	57	71	57
$Pseudo-R^2$	0.0850	0.072	0.0832	0.0786

In the first two models we do not only consider explanatory variables that are determined before the commencement of the campaign but also two variables indicating the success of the campaign and the competitive environment on the first day of the campaign. However, the results suggest that neither the number of competing investments on the first day nor the portion of the funding goal reached on the first day are significantly associated with the time until the update is posted. This indicates that the competitive environment and the success of a campaign are less important for the posting of the first update. In models 1 and 2 we also include a dummy variable indicating whether or not an update has been posted on the first campaign day. Interestingly, by using the Cox proportional hazard model we find that start-ups which post an update on the first day of a campaign make subsequent updates significantly quicker. Hence start-ups which start to communicate with investors early on appear to communicate more frequently later as well.

When also considering updates posted on the first campaign day (model 3 and 4), we find that the portal is significantly associated with the time before the first update is posted. In particular, for campaigns run on the platform Companisto, the time before the first update is published is significantly shorter. This finding can be considered as evidence of the fact that portals are critical concerning the promotion of entrepreneurs who run successful equity crowdfunding campaigns.

5.5 Conclusion

Entrepreneurship literature has extensively analyzed the interactions between venture capitalists or angel investors and entrepreneurs as well as the strategic behavior of each party (for example Sahlman, 1990; Schwienbacher, 2007; Mohamed and Schwienbacher, 2016). However, up to now, little has been known about the strategic behavior of entrepreneurs in crowdfunding. In this paper, we investigate the communication behavior of start-ups during and after an equity crowdfunding campaign. Such an investigation is important because, in contrast to venture capitalists, crowd investors neither obtain information from an insider from the board of directors nor do they receive news through contractual obligations such as specific covenants. Furthermore, in crowdfunding, the form and the content of voluntary disclosure is not specifically regulated through ad hoc disclosure requirements. Platforms often do not verify the disclosed information and thus entrepreneurs can strategically publish information through updates.

We use a data set of German equity crowdfunding campaigns in order to examine five hypothesis related to our research questions. The empirical evidence from our first hypothesis shows that the probability of an update increases along with stronger competition from parallel equity crowdfunding campaigns. There is only weak evidence in favor of a higher probability of updates when the campaign comes to an end and the reaching the funding goal becomes increasingly difficult. While the latter can be regarded as a sign of regular communication, the first finding indicates that start-ups indeed place their updates in such a way as to attract more attention. Regarding the question to what extent and how the communication behavior changes after the funding has been granted, we consider three hypotheses. While the hypothesis being that the tone of the updates is more positive during funding periods than thereafter, cannot be confirmed, we find evidence that during the funding period start-ups use a language that evokes warm-glow effects among potential investors and a feeling of group cohesion. Furthermore, we find some evidence to support the hypothesis that start-ups strategically post updates about the business development during the funding period. Moreover, they clearly post fewer updates after the funding has been ensured. All in all, this evidence indicates that during the funding period, the start-ups strategically place their updates with respect to frequency, content and the purpose to evoke emotions.

From these findings, we deduce the following implications for actors in the field. Given that equity crowdfunding often falls outside traditional securities regulation and, in particular, outside the securities prospectus regime as well as the market abuse regulation, securities regulators and platform providers should be wary about the content that start-ups post during an equity crowdfunding campaign. If equity crowdfunding further increases in importance, rules for investor communication may become necessary. For investors who primarily seek to maximize their return and who are not attracted by non-monetary motives, the strategic communication behavior may lead to sub-optimal investment decisions. This could be due possibly to blurred informational content of some updates which may be targeted at receiving funds and which do not accurately reveal real information. Whether a specific type of communication behavior of start-ups indeed leads to lower returns for investors should be investigated once the respective data becomes available. This is particularly relevant, given that little is know about the truthfulness of the information communicated by the start-ups. If start-ups systematically and strategically post fraudulent updates with the aim to increase investments, regulators have to consider enhancing investor protection in the context of equity crowdfunding. Our research suggests that companies that seek funding through an equity crowdfunding campaign should not rely too heavily on their strategic behavior as it can be revealed through systematic investigation.

Our paper also has clear limitations. With 97 campaigns (updates data set) and 71 campaigns (investment data set), our samples barely allow us to conduct extensive subsample analyses for different industries or founder teams. For example, larger founder teams may have better capacities and could be more creative in strategically posting updates. At the same time, they might also provide better checks and balances when it comes to the content of information disclosure. We use solely data from German platforms. Yet major international equity crowdfunding platforms such as Crowdcube and Seedrs have similar business models and also allow for updates both during and after the funding period. Therefore, the findings from our German data set can in principle be applied to many equity crowdfunding platforms worldwide, at least in the sense of an anticipated behavior.

Future research may focus on the learning process of entrepreneurs. Entrepreneurs with experience from multiple crowdfunding campaigns could apply a more sophisticated communication strategy than first-timers. Furthermore, the effects of mandatory disclosure in equity crowdfunding could also be of interest. In the context of venture capital, Cumming and Knill (2012) find evidence for a positive effect of strict disclosure requirements on both the supply and the performance of venture capital.

5.6 Appendix

Table 5.9 Probability of Updates, Fixed-Effects Logit Estimation.
The table reports upon odd's ratios of the fixed-effects logit estimation using the investment data set. The natural logarithm of *Amount* is used and #*Investments* is denoted in 1,000. *, **, and *** denote significance at a 10%-, 5%-, and 1%-level.

	Update	*Team*	*BusinessModel*	*Certification*	*Product*	*Cooperation*	*Campaign*	*NewFunding*	*Business*	*Promotions*	*Emotional*
Alarm	1.0796	0.0000	1.6330	1.2615	0.4786	0.9925	1.9731	0.0000	4.6661**	2.1227	$1.156 \cdot 10^{7}$
	(0.5572)	(0.0040)	(1.1140)	(1.4469)	(0.5096)	(0.7899)	(1.7112)	(0.0062)	(2.8930)	(1.8249)	$(1.0813 \cdot 10^{10})$
$\#Investments_{t-1}$	6.1381***	39.2974***	3.0519	2.3047	7.3930***	7.1817**	2.2531	4.0563	5.6256*	7.9231***	6.7288*
	(3.4624)	(45.0491)	(2.7057)	(2.8926)	(5.4082)	(6.0548)	(2.9512)	(6.1292)	(5.4165)	(5.2887)	(6.8310)
$\ln Amount_{t-1}$	0.9674	0.8000	0.9574	0.9991	0.9353	0.9878	1.0819	1.0592	1.0713	0.9269	0.9108
	(0.0552)	(0.1424)	(0.0803)	(0.0931)	(0.0800)	(0.0987)	(0.1183)	(0.1927)	(0.1026)	(0.0815)	(0.1306)
VDAX	1.0271	0.9175	0.9874	1.0404	1.0084	0.9895	0.9958	1.0040	1.0111	1.0307	0.9246
	(0.0291)	(0.0758)	(0.0415)	(0.0519)	(0.0413)	(0.0540)	(0.0553)	(0.0982)	(0.0496)	(0.0446)	(0.0694)
PostFunded	0.7614	0.1115***	0.6288	2.6651*	0.7796	1.0560	0.7010	1.1589	1.4739	0.6132	1.9475
	(0.1839)	(0.0802)	(0.2177)	(1.4163)	(0.2633)	(0.4815)	(0.3667)	(1.4006)	(0.6196)	(0.2231)	(2.2276)
# Obs.	4,387	1,963	3,923	3,318	3,675	2,810	2,918	1,388	3,278	3,662	2,160
$Pseudo-R^2$	0.0056	0.0578	0.0038	0.0069	0.0072	0.0064	0.0036	0.0053	0.0120	0.0108	0.0201

Chapter 6

Paralyzed by shock and confused by glut: The portfolio formation behavior of peer-to-business lending investors

This research project is joint work with Gregor Dorfleitner and Lars Hornuf and corresponds to a working paper with the same name.

Abstract We study the investor behavior on a leading peer-to-business lending platform and find evidence of two new investment biases—a default shock bias and a deep market bias. First, we find investors to stop investing in new loans and to cease from diversifying their portfolio after experiencing a loan default. This default shock significantly worsens the risk-return profile of investors' loan portfolios. Second, investors are unable to cope with a glut of loan campaigns. Similar to the default shock bias, investors cease from investing in new loans and consequently underdiversify their portfolios as more loans become available on the platform. Deeper markets also result in a deterioration of investors' risk-return profiles. Third, investment experience on the platform reduces the effect of the deep market bias.

Keywords Behavioral Finance, Investment Bias, Peer-to-business lending, crowdlending, RAROC, Diversification

JEL Classification G11, G41, G40

6.1 Introduction

In this paper, we analyze two new investment mistakes: a *default shock bias* and a *deep market bias*. We refer to a *default shock bias* when investors cease to diversify their portfolio after experiencing a default in their existing portfolio. A *deep market bias* refers to an investment bias that causes investors to underinvest in their portfolios because too many investment opportunities are currently available on the market. Our research aptly provides two novel explanations concerning why investors underdiversify their portfolios. By using data from peer-to-business lending, we study whether experience of a loan default and the availability of many investment opportunities affects the investment behavior of retail investors and consequently deteriorates the risk-return profile of their investment portfolios.

A natural benchmark for the two investment mistakes under scrutiny constitutes sophisticated lenders such as banks. The literature on credit risk modeling and bank management evidences that banks build portfolios based on the principle of diversification and that they use quantitative credit risk models to steer their loan portfolios (Hull, 2015). Such models explicitly consider default probabilities and losses given default of loans as well as their contribution to the portfolio risk and their profitability. Moreover, banks do not usually adopt their investment strategies at all after experiencing a loan default, as defaults are a well-anticipated part of their business model.

Although in sum retail investors behave in line with what has been referred to as 'the wisdom of the crowd' (Kelley and Tetlock, 2013), on an individual level they have been shown to make several investment mistakes (Calvet et al., 2009). In addition, their returns are often driven by sentiments (Kumar and Lee, 2006; Bollen et al., 2011). Furthermore, the evidence shows that retail investors, among others, underdiversify their portfolios (Goetzmann and Kumar, 2008; Calvet et al., 2009), adhere to a local bias (Seasholes and Zhu, 2010) and the disposition effect (Shefrin and Statman, 1985; Odean, 2002). Because retail investors are more prone to exhibiting all sorts of biases, they may also be more likely than professional investors to suffer from a default shock bias and a deep market bias. The digitalization of financial services and the recent advent of new financial technologies (fintechs) could principally render investment mistakes less likely. Digital innovations have the potential to support retail investors in their investment decisions. However, up to now, many new investment tools still had to prove their value. Through investigating a robo-advice tool from India, D'Acunto et al. (2018) show that ex ante well-diversified investors possess smaller portfolios, once they use the robo-advicing tool. In crowdfunding markets, evidence on the performance of retail investors is mixed. While in the peer-to-peer lending context Lin and Viswanathan (2016) evidence that investors suffer from a home bias, investors in equity crowdfunding appear to generate comparatively high returns (Signori and Vismara, 2017). By analyzing data from the crowdlending platform Funding Circle, Mohammadi and Shafi (2017) show that institutional investors perform much better than individual lenders in using the observable information on the platform website.

Peer-to-business lending is peculiar in many respects. First, unlike in peer-to-peer lending, where borrowers seek to refinance their personal debt or capital needs for consumption purposes, peer-to-business lending involves the financing of corporations. In order to make this type of business model sustainable, borrowers must provide sophisticated information upon their current financial situation. While anyone can provide capital for these loan projects, it requires at least some degree of financial literacy to understand the projects that seek funding. Second, investments in peer-to-business lending are possible with sums as small as 100 EUR. This makes losses relatively easy to digest. Consequently, investors should continue making investments and improve the diversification of their portfolio independent of a default in their portfolio and the number of available investment opportunities.

Furthermore, we address the question of whether the *default shock bias* and the *deep market bias* are reduced by the experience of the investors. Through the use of Swedish data, Calvet et al. (2007, 2009) have shown that financially more sophisticated and better educated investors are less likely to underdiversify their portfolios of stocks and mutual funds as well as to suffer from risky share inertia and the disposition effect. Given that peer-to-business lending is an activity that does not rely on financial advice and that investors themselves have to actively identify and choose investment projects on these markets, we expect more experienced investors to have a better risk-return profile and suffer from both biases to a lesser extent.

We start with deriving our three hypotheses concerning why investors may suffer from the investment biases outlined above and why more experienced investors may suffer less from them. Thereafter, we describe our data and outline the methods we apply. We then commence with a series of tests of whether the default shock bias and a deep market bias exist. Our findings are robust to different model specifications and dependent variables. In particular, we test whether the loan defaults and a glut of investment opportunities reduce the probability of an investment taking place at all, the number of new investments, the amount of new investments, and the amount of new investments relative to the investor's existing portfolio. If investors indeed suffer from a default shock bias and a deep market bias, we expect all of these measures to decrease and the risk-return relationship of the overall portfolio to worsen if a loan in the portfolio is defaulted or the available investment opportunities are comparatively large. This is confirmed by various tests. To measure the risk-return profile, we construct Value at Risk (VaR) measures and determine the risk-adjusted return on capital (RAROC). We then further examine whether the risk-return profiles improve as investors gain more experience and suffer less from these biases. Our data supports the conjecture that experienced investors are less prone to the deep market bias. Our results are robust to different RAROCs based on different VaR estimates.

Our paper is among the first to investigate investor behavior in peer-to-business lending. It attempts to identify some of the mistakes especially less experienced investors make. Given that the market leader Funding Circle has recently passed the mark of 5 bn USD lent worldwide, this is a relevant market segment of crowdlending that is largely underresearched. Our paper does not only help to understand existing investment biases better, but also informs policy makers and regulatory initiatives like the 'FinTech Action plan' that was recently proposed by the European Commission (2018).

6.2 Theory and hypotheses

Peer-to-business lending represents a new asset class for retail investors. Before the rise of peer-to-business lending platforms, investing in small corporate loans was almost exclusively available to institutional investors. Investors on online lending platforms cannot immediately obtain a diversified portfolio. They have to invest continuously in new loan projects over time in order to benefit from diversification. Furthermore, many retail investors have no experience in corporate loan investments and do not receive professional investment advice. Before we outline our hypotheses on possible behavioral biases, we provide a theory on how investors should rationally build a loan portfolio in peer-to-business lending.

6.2.1 A short theory of rational loan portfolio formation

Due to the nature of their business, banks can be regarded as being professional investors in loan portfolios. They use credit risk models and further tools based on these models to control risks and the risk-return relationship in their loan portfolios (Hull, 2015). Besides the purely regulatory requirements, which instruct them how to calculate the VaR, they typically run their own internal credit risk models for portfolio steering decisions (Hull, 2015).

The risk capital of financial institutions is a scarce resource that is meant to cover the losses from lending activities. Because the risk capital should be used efficiently, it has become a standard approach to consider the RAROC, which measures the portfolio return over the risk capital employed, when it comes to the optimal portfolio formations of financial institutions (Hull, 2015).[1] While the numerator of the RAROC is based on the expected profit of the portfolio, that is the interest charged minus the refinancing costs and expected loan losses, the denominator is essentially based on the VaR.[2] It can also be employed for the decision on expanding or reducing certain lines of business (Buch et al., 2011). Investors make investments that increase the RAROC and refrain from making those that reduce the RAROC. The literature on risk capital allocation is concerned with the question of how to allocate the overall risk capital to the existing business lines in order to calculate a RAROC for every line of business (Perold, 2005). However, in our setting retail investors only invest in peer-to-business loans, so that the risk capital allocation to several business lines is not required.

Banks hold large loan portfolios, which are generally well-diversified (Casu et al., 2006). From a RAROC perspective this is perfectly rational. While the numerator (as quantity relative to the portfolio size) remains roughly the same if we were to add or remove average profitable loans, the denominator decreases with an increasing number of loans at least as long as the portfolio is not well-diversified. If a bank holds only a small portfolio, it is therefore advisable to diversify into new loans, as the expected return does not change if the loan has an average interest margin, but the RAROC will increase due to a smaller VaR.

For banks, loan losses—even if the vast majority of loans reveals no defaults—are everyday business and are considered ex ante in the numerator and the denominator of the RAROC. It is also part of their regular business to extend new loans independent of whether old loans are paid back or default (Roy, 2016). If the number of defaults is higher than anticipated in the calculations leading to the RAROC, the bank will not usually cease to extend loans but instead update its credit risk model.

6.2.2 Hypotheses

Personal experience affects future investment decisions and helps to explain the heterogeneity in portfolio choices. Consistent with reinforcement learning theory (see e.g. Cross, 1973; Kaustia and Knüpfer, 2008), investors tend to repeat investment strategies that have resulted in favorable

[1]It should be noted that the concept exists in several variants, some of which are also called RORAC (return on risk-adjusted capital).

[2]Typically, the unexpected loss is used, which is defined as the portfolio VaR at a 99.5 % or a 99.9 % level minus the expected loss.

outcomes and tend to avoid investment strategies that have resulted in less favorable outcomes. Investment decisions can be affected by both the personal investment experience (see e.g. Kaustia and Knüpfer, 2008; Choi et al., 2009; Chiang et al., 2011; Andersen et al., 2018) and broader economic circumstances an investor has experienced such as a recession or particular labor market conditions (see e.g. Malmendier and Nagel, 2011; Knüpfer et al., 2017; Laudenbach et al., 2017). Andersen et al. (2018) highlight that stock investors who have suffered losses from defaults in the financial crisis subsequently change their risk-taking behavior. We conjecture that even in comparatively good economic conditions a default in the crowdlending portfolio may be a reason to alter the investment behavior. Investors, who experience a loan default may draw the conclusion from this event that they have made a mistake in trusting the platform and the lender. Consequently, they may reduce their exposure or stop participating in the new crowdlending market altogether. As diversification has to be achieved over time in this new asset class, investors tend to have small portfolios and may therefore be more greatly affected by a default in their loan portfolio compared with investors of well diversified stock portfolios. However, such behavior may well be irrational and constitute a bias if refraining from investing deteriorates the risk-return profile of the crowdlending portfolio. We therefore conjecture:

Hypothesis 1: Investors suffer from a *default shock bias* that decreases their readiness to further diversify their portfolio and thereby deteriorates their risk-return profile.

It has often been shown that human decision making capacity deteriorates if individuals receive too much information. This phenomenon has been referred to as information overload and has been found in various domains such as organization science, marketing, accounting, and management information systems (see Eppler and Mengis (2004) for an excellent review of the literature). By using an experimental setting, Tuttle and Burton (1999) show that information overload also exists if individuals analyze investments. In particular, they show that the human capacity to process information limits on the amount of information that can be processed per unit of time. Moreover, there is also extensive evidence of the fact that consumers suffer from choice overload (Iyengar and Lepper, 2000; Iyengar et al., 2004; Dhar, 1997; Shafir et al., 1993), which results in a type of behavior in which individuals either choose the default option or no option at all if confronted with too many prospects.

In the realm of crowdfunding, investors can inform themselves about active projects on the respective Internet platform. Being aware of the fact that investors may be overstrained in searching the entirety of active projects that suit their portfolio, many platforms have implemented filters that enable them to search, for example, for specific business segments or project volumes. Nevertheless, investors still confront the task to identify the relevant filters to place and which projects to look for. Empirical research on crowdfunding evidences that the number of projects that is active on a particular day impacts crowd support. This result has been shown to be relatively robust and holds for reward-based crowdfunding (Kuppuswamy and Bayus, 2017), equity-crowdfunding (Hornuf and Schwienbacher, 2017), and peer-to-business lending (Cumming and Hornuf, 2018). Crowd support can deteriorate for two reasons—either because potentially new investors do not join the community or existing investors amend their investment behavior. In our empirical analysis, we test whether changes in the behavior of existing investors affect crowd support and conjecture that the more loan campaigns become available, the less likely it is that investors invest in a loan. We refer to markets with many investment opportunities as deep markets and hypothesize:

Hypothesis 2: Investors suffer from a *deep market bias* that decreases their readiness to further diversify their portfolio and thereby deteriorates their risk-return profile.

Investors tend to make better investment decisions as they gain more experience (Korniotis and Kumar, 2011; Nicolosi et al., 2009). Moreover, experienced investors are less prone to behavioral biases. Feng and Seasholes (2005) and Dhar and Zhu (2006) show that the disposition effect, which describes the tendency of investors not to realize losses, decreases with investor experience and sophistication. Calvet et al. (2009) find that more experienced investors are less likely to suffer from the disposition effect and to underdiversify their portfolios. Moreover, in an earlier paper, Calvet et al. (2007) evidence that more sophisticated households incur higher returns because they invest more efficiently and more aggressively. We study whether experience affects the investment performance in general, as well as the magnitude of the default shock bias and the deep market bias in particular. We thus hypothesize:

Hypothesis 3a: Investors with more experience tend to improve their risk-return profile.

Hypothesis 3b: Investors with more experience suffer less from the *default shock bias* and the *deep market bias*.

6.3 Data

6.3.1 Summary statistics of loan data

For the following analysis, we use data from Zencap, which is the first and the largest German peer-to-business lending platform. We obtain data from the time of the inception of the platform in March 2014 until the merger of Zencap with the platform Funding Circle in November 2015. Since the merger, Funding Circle has been the world's leading crowdlending platform for corporate loans.

The platform facilitates loans for small and medium-sized corporations. These corporations post their loan projects with the requested principal amount as well as several firm characteristics and financial information on the platform. Corresponding to the estimated default risk, the platform assigns a rating ranging from A (best) to E (worst) as well as a corresponding interest rate. The investors can invest in the loan campaign within a pre-defined funding period.[3] If the invested amount reaches the principal amount at the end of the funding period, a loan campaign is successful and the loan is funded. After receiving the funds, the borrowers re-pay their loans in the form of fixed monthly annuities, due at the middle of each month.

In the observation period, 414 borrowers applied for a loan via the platform. Tables 6.2 and 6.3 show the descriptive statistics of these corporations and of the investors backing the loans. On average, the platform assigned a nominal interest rate of 7.38 % to these corporations. Not all of the corporations were profitable. The net income ranges between a minimum of −346,300 EUR and a maximum of over 1 m EUR. In total, 367 loan applications were successful. The platform does not provide any particular information on the repayment status of the loans but states that within the observation period only a handful of loans defaulted. We made use of the forum P2P-kredite.com (http://www.p2p-kredite.com) to research which loans defaulted within the observation period. We observe five borrowers who declared insolvency before November 2015, equaling 1.36 % of all successfully funded corporations. However, we do not consider the defaults of successfully funded corporations which declared insolvency after the end of the observation period. In total, 2,129 investors backed the loans on the platform. Overall, 89 % of the investors are male. Moreover, the average investor is 41 years old.

[3]In general, the funding period lasts 21 days. However, corporations can extend the funding period to a maximum of 61 days.

Table 6.1 Definition of variables.

	Borrower Variables
EBIT	EBIT of the corporation running the loan project, in EUR.
Employees	Number of employees in the corporation running the loan project.
Equity	Equity of the corporation running the loan project, in EUR.
Foundation Year	Foundation year of the corporation running the loan project.
Lenders	Total number of lenders backing a loan project.
Loan Duration	Duration of the loan in month.
Net Income	Net income of the corporation running the loan campaign.
Rating Class	Rating class of the campaign as assigned by the platform. Ranging from A+ to C−.
	Investor Variables
Age	Age of the investor.
Gender	Gender of the investor (1 = female).
	Portfolio Variables
#Campaigns	Number of simultaneously active loan projects on the platform on each valuation date.
Distance	Mean distance of all the active loan projects on the valuation date and the investor in km.
ExpReturn	Expected Return of the portfolio calculated as the weighed average of the expected returns of the loan projects the investor invested in on each valuation date. The expected return for each loan is calculated as the nominal yield times the invested amount minus the expected loss as described by the rating class.
Experience	Experience of the investor on the platform in months. Calculated as the date on valuation date minus the account creation date.
InsolvDummy	Dummy Variable indicating whether the borrower declared insolvency previous to the valuation date (1 = insolvency).
#Inv	Total number of investments and investor invested in until the valuation date.
Min Distance	Minimal distance of all the active loan projects on the valuation date and the investor in km.
NewInvAmount	Amount the investor newly invested on each valuation date, in EUR.
NewInvDummy	Dummy variable indicating whether the investor made at least one new investment on the valuation date.
#NewInv	Number of investments an investor newly invested on the valuation date.
NewInvRel	Amount newly invested over the portfolio value of the previous valuation date.
Nominal Yield	Nominal yield of the loan project as assigned by the platform.
Principal Amount	Principal amount of the loan, in EUR.
RAROC	Risk-adjusted return on capital (RAROC). Calculated as the expected return of the portfolio divided by the value at risk of the portfolio.
Success Dummy	Dummy Variable indicating whether the campaign was successful (1 = successful).
VaR	Value at Risk (VaR) of the portfolio. The VaR is a relative measure and is calculated as the risk capital over the total invested amount on each valuation date. The 99.5 % VaR is used if not stated otherwise. For a detailed calculation of the VaR, see the Appendix.

Table 6.2 Descriptive statistics of the metric variables of the corporations and the investors. The variables are defined in Table 6.1.

Variable	N	Mean	SD	Min	Max
Borrower Characteristics					
Nominal Yield	414	0.0738	0.0188	0.0408	0.1564
Principal Amount	414	72,183.57	46,889.03	10,000	250,000
Loan Duration	414	34.01	13.80	6	60
Success Dummy	414	0.89	0.31	0	1
Foundation Year	414	2001	17.19	1784	2014
Equity	414	160,649.70	519,186.70	$-1,214,900$	7,492,967
EBIT	414	93,367.33	127,433.80	$-379,600$	1,291,700
Net Income	414	66,744.35	101,940.60	$-346,300$	1,112,533
Employees	414	17.55	28.51	1	300
Lenders	414	84.70	47.28	4	302
Investor Characteristics					
Age	2,129	40.89	12.57	18.44	107.16

Table 6.3 Descriptive statistics of the categorical variables of the corporations and the investors. In general, borrower variables have 414 observations. For *InsolvDummy*, only the 367 successfully funded corporations are considered. Investor variables have 2,129 observations. The variables are defined in Table 6.1.

Borrower Characteristics					
Rating Class	A+	A	B	C	C−
Absolute frequency	86	21	196	96	15
Relative frequency	20.77 %	5.07 %	47.34 %	23.19 %	3.62 %
Success Dummy	1 (Yes)	0			
Absolute frequency	367	47			
Relative frequency	88.65 %	11.35 %			
InsolvDummy	1 (Yes)	0			
Absolute frequency	5	362			
Relative frequency	1.36 %	98.63 %			
Investor Characteristics					
Gender	1 (Yes)	0			
Absolute frequency	239	1,890			
Relative frequency	11.23 %	88.77 %			

6.3.2 Construction and summary statistics of portfolio data

Because we are interested in investor specific VaRs and RAROCs and loan repayments only take place once per month, we create 20 valuation dates in the middle of each month starting from 15th April 2014 and ending 15th November 2015. We set the valuation dates immediately after the repayment dates. In a first step, we determine how much investors have invested in each loan at each valuation date, also considering the payback from the monthly annuities. Thus, we derive a loan portfolio for each investor on each valuation date and construct a monthly panel data set with valuation dates forming the time dimension and the investor as a cross-sectional dimension. In a next step, we use this monthly panel data to calculate aggregated portfolio variables. All portfolio variables are weighted with the amount invested in each loan on each valuation date. Furthermore, we observe the new investment decisions the investors make on each valuation date.

Table 6.4 shows the descriptive statistics of the portfolio variables. The investment behavior varies greatly within the group of investors and the valuation dates. On 36 % of the valuation dates, investors decide to invest in at least one new project. The mean amount the investors invest equals 559 EUR and is well above the minimum investment of 100 EUR even though we have many observations with zero investments. The number of new investments in loan projects ranges from 0 to 41. On average, investors hold portfolios consisting of 10 different loans. Within the observation period, 99 % of all investors hold 78 loans or less. However, in order to achieve a well-diversified portfolio, that is one in which the portfolio risk scales linearly with portfolio size, a portfolio of several hundred loans would be necessary.[4]

We measure the effect of the investment decisions on the risk and the return of the portfolio through several variables. The VaR, at a 99.5 % confidence level, measures the relative loss risk of the portfolio. The descriptive analysis of VaR shows that, on average, 99.5 % of the losses will not exceed 42 % of the portfolio value. The average expected return of the portfolios equals 6 %. To measure the risk-return profile we combine the mentioned two measures and obtain the RAROC. The RAROC is calculated as the expected return of a portfolio over the VaR. A detailed explanation of the VaR determination of the portfolios and the calculation of the RAROC can be found in the Appendix. To analyze the risk-return profile over time, we observe changes in the RAROC ($\Delta RAROC$). An increase of the RAROC indicates an improvement of the risk-return profile. By contrast, a decrease implies a deterioration of the risk-return profile.

[4] Dorfleitner and Pfister (2014) show that in order to have constant per unit risk, which can be interpreted as having a well diversified portfolio, the minimum number of loans ranges from roughly 200 (VaR at 95 % level and loan default probability of 0,5 %) to more than 500 (VaR at 99.9 % level and loan default probability of 10 %).

Table 6.4 Descriptive statistics of the portfolios. Monthly panel data for 2,219 investors and 20 valuation dates. The variables are defined in Table 6.1. As many investors created their accounts after the beginning observation period, several portfolio variables are not available for all investors on all valuation dates.

Variable	N	Mean	SD	Min	Max
#Campaigns	42,580	34.75	16.63	6	61
Distance	42,580	320.32	260.84	149.22	7972.41
Min Distance	42,580	55.31	259.53	0	7703.35
Experience	22,087	6.7015	4.82	0	19.87
NewInvAmount	22,087	558.50	2014.29	0	130,000
NewInvDummy	22,087	0.36	0.48	0	1
#NewInv	22,087	1.31	2.99	0	41
NewInvRel	18,405	0.22	1.31	0	74.00
#Inv	20,398	10.32	17.18	1	315
ExpReturn	20,398	0.06	0.01	0.03	0.13
VaR	20,398	0.42	0.32	0.04	1
RAROC	20,398	0.23	0.17	0.03	1.24
ΔRAROC	18,265	0.02	0.05	−0.27	0.43
InsolvDummy	20,398	0.01	0.04	0	1

As described in Table 6.8, the change of the RAROC is, on average, positive when investors choose new investments, yet, in approximately 17 % of the cases new investments result in a negative change of the RAROC. Not investing in new loans can also have a positive, negative or zero effect on the risk-return profile.

Table 6.5 Descriptive statistics of the change of the RAROC.

Variable	N	Mean	SD	Min	10th Pctile	Median	90th Pctile	Max
NewInvDummy = 1								
ΔRAROC	6,422	0.0461	0.0714	−0.2676	−0.0465	0.0453	0.1305	0.4257
NewInvDummy = 0								
ΔRAROC	11,843	−0.0007	0.0115	−0.1928	−0.0033	0	0.0004	0.2254

Furthermore, we determine the experience of each investor on the platform on each valuation date. Therefore, we calculate the difference between the valuation date and the date of the account creation for each investor. Moreover, the descriptive analysis suggests that on 1 % of the valuation dates, an investor has already experienced at least one of the five insolvencies that occurred on the platform. We also investigate the competitive environment of the campaigns. On average, 34 loan campaigns were active on a given day. While the average distance between these active loan projects and an investor amounts to more than 300 km, the distance between the closest active loan campaign and an investor is, on average, only 55 km.

6.4 Methodology

To examine which factors impact on the investment behavior, we estimate the effects of several covariates on the investment decision of each investor *i* at each valuation date *t*. We specify the following regression equation:

$$\begin{aligned} InvestmentDecision_{i,t} = & \alpha + \beta_1 \cdot InsolvDummy_{i,t} + \beta_2 \cdot \#Campaigns_{i,t} + \beta_3 \cdot Distance_{i,t} \\ & + \beta_4 \cdot Experience_{i,t} + \beta_5 \cdot (Experience_{i,t})^2 + \beta_6 \cdot \#Inv_{i,t-1} \\ & + \beta_7 \cdot Gender_i + \beta_8 \cdot Age_{i,t} + \beta_9 \cdot TimeFE + \varepsilon_{i,t} \end{aligned} \tag{6.1}$$

where *Investment Decision* represents one of four different dependent variables, namely *NewInvDummy*, *log(NewInvAmount)*, *#NewInv* and *NewInvRel*. *NewInvDummy* is a binary variable equaling one if the investor makes at least one new investment by the next valuation date. Therefore, we apply a logit regression for this specification. For the natural logarithm of the newly invested capital (*NewInvAmount*) as well as the newly invested capital related to the total portfolio value of the previous valuation date (*NewInvRel*), we estimate OLS regressions. Finally, since the number of new investments is a count variable and our data suffers from overdispersion, we use a negative binomial model to examine the effect of the covariates on *#NewInv*.

To test our hypotheses we include the explanatory variables $InsolvDummy_{i,t}$, $\#Campaigns_{i,t}$, as well as $Experience_{i,t}$. To account for potential non-linear effects of experience, we also add the squared term of experience. Prior research indicates that crowdfunding investors suffer from a local bias (Lin and Viswanathan, 2016). Therefore, we include the average distance between the investors and the active loan campaigns at each valuation date ($Distance_{i,t}$) as a control variable. Furthermore, the size of the portfolio is shown to be related to the investment behavior (Kaustia and Knüpfer, 2008; Goetzmann et al., 2015; Laudenbach et al., 2017). We thus add the number of loans in the portfolio at the last valuation date ($\#Inv_{i,t-1}$). Moreover, we control for the age and the gender of the investor, who have been shown to influence risk taking and the investment behavior (see e.g. Barber and Odean, 2001; Agnew et al., 2003; Bajtelsmit et al., 1999). In all models, we cluster the standard errors at investor level. Furthermore, we include time fixed effects (*Time FE*).

In a next step, we analyze the determinants of changes of the risk-return profile of the portfolios. We use our monthly panel data to investigate the changes of the RAROC for the investor i on the valuation date t. To address potential endogeneity concerns resulting from the simultaneous determination of the investment decision and the change of the RAROC, we implement an instrumental variable approach. We estimate a two-stage least squares (2SLS) model. In the first stage, we predict the *Investment Decision* using equation (6.1). In this context, the hypothesis-related variables *InsolvDummy*, *#Campaigns*, and the control variable *Distance* serve as instruments. In the second stage, we use the predicted value of the investment decision ($\widehat{InvestmentDecision}_{i,t}$) instead of the investment decision and estimate the following regression equation:

$$\begin{aligned}\Delta RAROC_{i,t} =& \alpha + \beta_1 \cdot \widehat{InvestmentDecision}_{i,t} + \beta_2 \cdot Experience_{i,t} + \beta_3 \cdot (Experience_{i,t})^2 \\ &+ \beta_4 \cdot \#Inv_{i,t-1} + \beta_5 \cdot Gender_i + \beta_6 \cdot Age_{i,t} + \beta_7 \cdot TimeFE + \varepsilon_{i,t}\end{aligned} \tag{6.2}$$

We control for the relevance of our instruments by ensuring that the F-statistic of the first stage is above a value of 10 (Bascle, 2008; Staiger and Stock, 1997). Furthermore, we test the exogeneity condition of the instruments with the use of a test of overidentifying restrictions (Bascle, 2008).

6.5 Results

6.5.1 What are the determinants of new investment decisions?

In the first part of our analysis, we investigate which factors drive investors towards making new investment decisions. Table 6.6 displays the results. Moreover, we test for the influence of experience in more detail in Table 6.7.

For the first three specifications, we find that *InsolvDummy* is significantly negatively related to new investment decisions. This suggests that investors indeed tend to recoil from new investments after experiencing a default. The economic significance is large as well. All else equal, the second specification suggests that investors who have experienced a default invest approximately 93 % less[5] than investors who have never encountered such a negative event.

[5] Calculated as $e^{-2.71} - 1 = -93.35\%$.

This result provides strong evidence in favor of our first hypothesis, being that investors inherit a default shock bias and tend to stop investing in an asset class after experiencing a default within this asset class. Unlike banks, investors appear to be surprised that loans can default and thus, they irrationally change their investment behavior after experiencing such an insolvency.

Table 6.6 Regression Results for the Investment behavior using logit (dep. variable: *NewInvDummy*.), ordinary least squares (dep. variable: log(*NewInvAmount*), and *NewInvRel*) and negative binomial model (dep. variable: *#NewInv*). Standard errors are clustered on investor level and shown in parentheses. For the logit and the negative binomial model, marginal effects at means are displayed instead of coefficients. In the last specification, we additionally include an interaction term between #Campaigns and the experience (Exp·#Camp). The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

Dependent Variable	NewInvDummy	log(NewInvAmount)	#NewInv	NewInvRel	log(NewInvAmount)
InsolvDummy	−3.4795***	−2.7129***	−7.4936***	−0.0394	−1.114***
	(0.6550)	(0.6898)	(1.3640)	(0.0405)	(0.1336)
#Campaigns	−0.0400***	−0.2721***	−0.0808***	−0.2135***	−0.2474***
	(0.0077)	(0.0614)	(0.0157)	(0.0751)	(0.0623)
log(Distance)	−0.0199	−0.0908	−0.0135	0.0018	−0.1091
	(0.0267)	(0.1803)	(0.0676)	(0.0371)	(0.1784)
Experience	−0.0457***	−0.2736***	−0.1776***	−0.1477***	−0.3321***
	(0.0046)	(0.0285)	(0.0135)	(0.0144)	(0.0405)
$(\text{Experience})^2$	0.0012***	0.0084***	0.0056***	0.0066***	0.0099***
	(0.0003)	(0.0014)	(0.0008)	(0.0007)	(0.0016)
$\#\text{Inv}_{t-1}$	0.0154***	0.0671***	0.0413***	−0.0006**	0.0724***
	(0.0009)	(0.0056)	(0.0018)	(0.0003)	(0.0061)
Gender	−0.0899***	−0.5629***	−0.2327***	−0.0508	−0.5654***
	(0.0207)	(0.1042)	(0.0614)	(0.0362)	(0.1036)
Age	0.0031***	0.0272***	0.0076***	0.0025***	0.0278***
	(0.0005)	(0.0033)	(0.0012)	(0.0007)	(0.0033)
Time FE	yes	yes	yes	yes	yes
Exp·#Camp					0.0099***
					(0.0016)
Constant		6.6454***		4.6493***	6.5136***
		(1.5027)		(1.4550)	(1.4953)
N	18,315	18,315	18,315	18,315	18,315
Pseudo-/ Adj. R^2	0.1606	0.1754	0.0820	0.0370	0.1816

Furthermore, our results show that the number of simultaneously active loan campaigns is significantly negatively related to new investment decisions in all four specifications. This indicates that investors invest less when they have a great number of investment possibilities. In the data, one additional active loan project decreases the invested capital by almost 24 %.[6] This result is consistent with our second hypothesis.

Moreover, we find that the effect of experience is U-shaped. Initially, experience is negatively associated with new investment decisions in all four specifications. In particular, investors appear to invest more when they are relatively new on the platform, which contradicts the hypothesis 3a. However, the squared term of experience indicates a reversal point, given the significantly positive coefficient of the squared term. This result suggests that experience has a negative influence on new investment decisions but the effect diminishes with more experience. At a certain tipping point the relationship reverses. As the squared terms in non-linear models cannot intuitively be interpreted, we plot the predictive margins for different levels of experience for the logit and the negative binomial models and find that the negative effect of experience weakens for more experienced investors. There are several plausible explanations for this type of investment behavior. On the one hand, investors could start using the platform by investing a certain amount of capital in several loan projects. Subsequently, they observe how their investments develop without investing further capital. If they are satisfied with their investments, they start investing again. This point could mark the reversal point in our analysis. On the other hand, the investment behavior could be driven by liquidity. Again, investors commence by investing a certain amount of capital. Then they wait until they receive sufficient repayments from their loan investments before they invest again. However, we test for the influence of the aggregated cash repayments and do not find evidence of a relationship between aggregated liquidity and the investment behavior.[7]

Turning to control variables, we find several interesting effects as well. The number of previously invested loan projects is significantly positively related to new investment decisions for the first three specifications, suggesting that investors who already have invested in more loans are more likely to invest again. Furthermore, the data suggests that women and young people have a lower propensity to undertake new investments. The first relationship could be explained by the fact that women generally tend to trade less and are more risk averse than men (see e.g. Barber and Odean, 2001; Agnew et al., 2003; Dwyer et al., 2002). That younger people have a lower propensity to invest is in line with the literature, which shows that older people tend to trade more (see e.g. Agnew et al., 2003). Additionally, age may serve as a proxy for wealth, indicating that wealthier individuals invest more (see e.g. Bajtelsmit et al., 1999). The coefficient of the mean distance of the active loan projects, however, is not significantly different to zero. This

[6] Calculated as $e^{-0.27} - 1 = -23.82\%$.

[7] Results can be provided upon request.

holds true for all four dependent variables. Thus, we do not find evidence of the fact that the geographical distance plays a role in peer-to-business lending.

To test hypothesis 3b, we divide our data set into two subsamples based on the mean experience of the investors: the first subsample contains all observations of investors with little experience on the platform (experience of less than 200 days), and the second subsample includes all observations with experienced investors (experience of 200 days or more). Table 6.7 provides the results. In the section concerning robustness checks, we repeat the analysis using different levels of experience to create our subsamles.

We find that the influence of the default shock bias persists not only for new investors but also for investors with longer experience. The coefficient of *InsolvDummy* is negative and highly significant in all specifications. Moreover, the results show a different impact of the simultaneously active loan campaigns on the investment decision depending on the level of experience. While new investors tend to invest less when they have more investment opportunities, experienced investors appear to act more rationally and invest more when they have greater investment choices. To examine the effect of experience on the deep market bias in more detail, we include an interaction term between the number of active campaigns and experience in our main model. Results are shown in the last column of Table 6.6. The significant positive sign of the coefficient of the interaction term suggests that experienced investors suffer from the deep market bias to a lesser extent. Overall, the results provide evidence for hypothesis 3b.

6.5.2 What drives changes in the risk-return profile of the portfolios?

In addition to the analysis of the investment behavior, we examine how these investment decisions, as well as experience and further controls, impact the risk-return profile of the investors. Note that new investment decisions also include the decision not to make a new investment on a valuation date. We estimate three different models. The first model includes all hypothesis-related variables, but not the new investment decision. In the second model, we add the investment decision. In the third model, which we regard as being our main model, we estimate a 2SLS model, considering potential endogeneity concerns due to simultaneous determination of the investment decision and the change in the RAROC. Table 6.8 provides the results.

The first model indicates that *InsolvDummy* and *#Campaigns* have a significant negative impact on the change of the RAROC. However, this effect vanishes if we include the new investment

Table 6.7 Subsample-Regression Results for the investment behavior using logit (dep. variable: *NewInvDummy*.), ordinary least squares (dep. variable: log(*NewInvAmount*), and *NewInvRel*) and negative binomial model (dep. variable: *#NewInv*). Standard errors are clustered on investor level and shown in parentheses. For the logit and the negative binomial model, marginal effects at means are displayed instead of coefficients. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

	Experience $<$ 200 days				Experience $\geq$ 200 days			
	NewInvDummy	log(NewInvAmount)	#NewInv	NewInvRel	NewInvDummy	log(NewInvAmount)	#NewInv	NewInvRel
InsolvDummy	−3.4298***	−3.4769***	−20.0320***	−0.9532***	−2.7082***	−2.6925***	−5.4842***	−0.1879***
	(0.8596)	(1.1408)	(7.6326)	(0.1398)	(0.5020)	(0.6788)	(0.9784)	(0.0452)
#Campaigns	−0.0468***	−0.3531***	−0.1844***	−0.2422***	0.0172***	0.0242***	0.0458***	0.0013***
	(0.0089)	(0.0618)	(0.0288)	(0.0742)	(0.0048)	(0.0025)	(0.0118)	(0.0004)
log(Distance)	0.0074	0.1267	0.1328	0.0309	−0.0331	−0.2795	−0.0691	−0.0558
	(0.0360)	(0.2551)	(0.1681)	(0.0601)	(0.0366)	(0.2215)	(0.0847)	(0.0399)
$\#Inv_{t-1}$	0.0562***	0.4960***	0.3562***	−0.1550***	0.0108***	0.0559***	0.0248***	−0.0003
	(0.0056)	(0.0361)	(0.0286)	(0.0191)	(0.0008)	(0.0048)	(0.0015)	(0.0002)
Gender	−0.1190***	−0.7121***	−0.5198***	−0.0705	−0.0864***	−0.4957***	−0.1708**	−0.0404***
	(0.0265)	(0.1390)	(0.1380)	(0.0701)	(0.0300)	(0.1396)	(0.0772)	(0.0123)
Age	0.0024***	0.0202***	0.0109***	0.0099***	0.0026***	0.0223***	0.0048***	0.0001
	(0.0007)	(0.0043)	(0.0032)	(0.0015)	(0.0006)	(0.0043)	(0.0014)	(0.0005)
Time FE	yes	yes	yes	yes	yes	yes	yes	yes
Constant		3.4067*		5.2040***		0.4696		0.3406
		(1.8247)		(1.4906)		(1.2900)		(0.2461)
N	9094	9094	9094	9094	9221	9221	9221	9221
Pseudo-/ Adj. R^2	0.0396	0.0821	0.023	0.0299	0.1926	0.2049	0.0950	0.0041

decision in the second and the third model. Therefore, we conclude that the first model suffers from an omitted variable bias and *#Campaigns* as well as *InsolvDummy* only capture the significant effect of the investment behavior in the first model. In both the second and the third model, the investment behavior is highly significantly associated with the change of the RAROC. In particular, the 2SLS model, indicates that *InsolvDummy* and *#Campaigns* significantly affect the investment decision. The investment decision, in turn, is significantly related to the change of the RAROC. The evidence suggests a positive relationship between a high amount of newly invested capital and improvements of the risk-return profile of the investors.

Table 6.8 Regression Results for the change of the RAROC. Standard errors are clustered on investor level and shown in parentheses. The last two columns show a two-stage-least-squares (2SLS) estimation with log(*NewInvAmount*) being the dependent variable of the first stage and $\Delta RAROC$ the dependent variable of the second stage. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

			2 SLS	
Dependent Variable	ΔRAROC	ΔRAROC	log(NewInvAmount)	ΔRAROC
log(NewInvAmount)		0.0067***		0.0084***
		(0.0001)		(0.0017)
InsolvDummy	−0.0197***	−0.0005	−2.8556***	
	(0.0069)	(0.0036)	(0.7578)	
#Campaigns	−0.0028***	−0.0010	−0.2725***	
	(0.0010)	(0.0009)	(0.0614)	
log(Distance)	−0.0007	−0.0001	−0.0909	
	(0.0017)	(0.0012)	(0.1806)	
Experience	−0.0085***	−0.0066***	−0.2720***	−0.0062***
	(0.0004)	(0.0003)	(0.0286)	(0.0006)
$(\text{Experience})^2$	0.0003***	0.0003***	0.0084***	0.0003***
	(0.0000)	(0.0000)	(0.0014)	(0.0000)
$\#\text{Inv}_{t-1}$	0.0003***	−0.0001***	0.0671***	−0.0002*
	(0.0000)	(0.0000)	(0.0056)	(0.0001)
Gender	−0.0044***	−0.0006	−0.5652***	0.0004
	(0.0010)	(0.0007)	(0.1046)	(0.0012)
Age	0.0002***	−0.0000	0.0272***	−0.0001
	(0.0000)	(0.0000)	(0.0033)	(0.0001)
Time FE	yes	yes	yes	yes
Constant	0.0938***	0.0492***	6.6492***	0.0291***
	(0.0206)	(0.0181)	(1.5036)	(0.0024)
N	18,265	18,265	18,265	18,265
Adj.R^2	0.0809	0.2279	0.1747	0.2183

Furthermore, we find that the coefficient of *Experience* is significantly negative indicating that experience deteriorates the risk-return profile. However, the coefficient of the squared experience is significantly different to zero with a positive sign. Together with the effect of experience throughout the first stage, this indicates a reversal point subsequent to approximately 340 days, after which more experience is associated with an improvement of the RAROC.[8]

With subsample regressions, we examine the effect of covariates on the change of the RAROC depending on the level of experience. Similar to the analysis of the investment behavior, we separate the observations with respect to less experienced investors (experience less than 200 days) and experienced investors (experience of 200 days or more). Table 6.9 shows the results for the 2SLS estimation.

Table 6.9 Subsample-Regression Results for the change of the RAROC. Standard errors are clustered on investor level and shown in parentheses. We apply two-stage-least-squares (2SLS) estimation with log(*NewInvAmount*) being the dependent variable of the first stage and Δ*RAROC* the dependent variable of the second stage. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

	Experience $<$ 200 days		Experience $\geq$ 200 days	
Dependent Variable	log(NewInvAmount)	ΔRAROC	log(NewInvAmount)	ΔRAROC
log(NewInvAmount)		0.0142***		0.0047***
		(0.0024)		(0.0009)
InsolvDummy	−3.6777***		−2.8923***	
	(0.9218)		(0.7786)	
#Campaigns	−0.3759***		0.0453***	
	(0.0608)		(0.0041)	
log(Distance)	0.0274		−0.2789	
	(0.1887)		(0.2221)	
#Inv$_{t-1}$	0.1208***	−0.0012***	0.0558***	−0.0000
	(0.0059)	(0.0003)	(0.0048)	(0.0001)
Gender	−0.5302***	0.0005	−0.4992***	0.0018*
	(0.1222)	(0.0018)	(0.1406)	(0.0010)
Age	0.0312***	−0.0002**	0.0222***	−0.0000
	(0.0038)	(0.0001)	(0.0043)	(0.0000)
Time FE	yes	yes	yes	yes
Constant	6.5068***	0.0072***	0.0767	−0.0003
	(1.5346)	(0.0025)	(1.2957)	(0.0012)
N	9,094	9,094	9,171	9,171
Adj.R^2	0.144	0.151	0.204	0.139

In both subsamples, the investment decision is significantly positively associated with a change in the RAROC. This is consistent with the results from our main models. However, we find interesting effects regarding gender and age. The data suggest that experienced women tend to enhance their risk-return profile compared with experienced men. Moreover, for inexperienced investors we find a negative association between age and the change of the RAROC.

[8] A value of $x = 340$ marks the minimum of the term $(\beta_{Exp.;2ndStage} + \beta_{Exp.1stStage} \cdot \beta_{NewInvAmount}) \cdot x + (\beta_{Exp.^2;2ndStage} + \beta_{Exp.^2;1ndStage} \cdot \beta_{NewInvAmount}) \cdot x^2$, where x represents the experience in days.

Overall, we find strong evidence in favor of our first two hypotheses. Analyzing both the investment behavior and changes in the risk-return profile suggests that investors suffer from a default shock bias. Investors tend to recoil from new investments after having experienced a loan default in their portfolio. This behavior, in turn, is associated with a deterioration of the risk-return profile of the portfolio. Moreover, investors appear to suffer from a deep market bias and invest less when they have a broad selection of loan campaigns in which to invest. Again, this investment behavior appears to be irrational as it is related to a worsening of the risk-return profile.

In order to examine hypotheses 3a and b, we investigate the effect of experience on both the investment behavior and the risk-return profile. Our results suggest that the association between the investment behavior and experience is U-shaped. Investors appear to invest less with increasing experience up to a certain tipping point. After this, the relationship reverses. We observe the same pattern for the relationship between experience and the risk-return profile. Regarding hypothesis 3b, we obtain some evidence suggesting that the effects of the biases are weakened with increased experience. Experienced investors invest significantly more when they have greater investment choices. This investment behavior is related to an improvement of the risk-return profile.

6.5.3 Robustness Checks

We perform several tests and estimate further specifications in order to ensure the robustness of our results. To this end, we repeat all subsample regressions with different levels of experience for both the analysis of the investment behavior and the change within the risk-return profile. The results remain largely unchanged, defining less experienced investors as being investors with less than 250 days and less than 180 days of experience, respectively.[9] In the analysis of the investment behavior, we note small differences for experienced investors in the last specification. In particular, the significance of the coefficient of *#Campaigns* decreases from a 1% significance level in our main model to a 5% level for investors with more than 250 days experience. Furthermore, the coefficient of $\#Inv_{i,t-1}$ is negatively significant at a 10 % level for investors with more than 180 days experience. With regard to the investigation of changes in the RAROC, we find that the coefficient of age, which is not significant in our main model, becomes negatively significant at a 10 % level when defining experienced investors as being investors with more than 250 days of experience.

[9] Results can be provided upon request.

Moreover, we calculate the change of the RAROC based on different VaR levels. Table 6.10 presents the results using a 95 %, a 97 % and a 99 % VaR, respectively. As the probability of default for A-rated bonds equals 0.6 % according to the platform, the VaR of an investor's portfolio can amount to zero for these VaRs. As we cannot calculate the RAROC for a VaR of zero, the number of observations decreases with a lower probability threshold of the VaRs. We find that the results remain robust when using the different RAROCs. In particular, we find a positive influence of new investment decisions in all specifications. Additionally, experience retains its previously exhibited relationship with the change of the RAROC.

As another robustness check for the results of the 2SLS model, we use other measures for the investment behavior as dependent variables in the first stage. The results of the second stage are displayed in Table 6.11. Our findings hardly change when we use different endogenous variables to measure the investment behavior. The predicted values of *NewInvDummy*, *#NewInv* and *NewInvRel* are all significantly positively related to the $\Delta RAROC$. Moreover, the effect of experience is similar to the effect found in the main model. We find evidence that the change of the RAROC first decreases with longer experience, but this relationship reverses after a tipping point is reached. However, the exogeneity condition is not valid for the instruments in the case of *NewInvRel*. Hence, the estimator could be inconsistent using this endogenous variable.

Furthermore, we use different instruments for our 2SLS model. First, we reduce the number of instruments and only use one of the variables *InsolvDummy*, *#Campaigns* and *Distance* as an instrument. The results remain robust. The hypothesis-related variables are significantly related to the investment decision in the first stage, but not significantly related to the change of the RAROC in the second stage. Second, we use the age and the gender as instruments and include all previous instruments in the second stage. Again, *InsolvDummy*, *#Campaigns* and *Distance* are not significantly associated with the $\Delta RAROC$ in the second stage. The effect of the other variables remains unchanged as well. In particular, both the impact of experience and the squared term of experience remain largely unchanged in all models.

6.6 Conclusion

We find evidence to support the fact that investors suffer from two new investment biases—the default shock bias and the deep market bias. Furthermore, we show that experienced investors are less prone to making these investment mistakes. We use data from a peer-to-business lending platform, which allows retail investors to invest in corporate loans. Before the rise of

Table 6.10 Robustness Test for RAROC. The RAROC is calculated on the basis of different VaRs (95 %, 97 % and 99 %). Standard errors are clustered on investor level and shown in parentheses. We apply two-stage-least-squares (2SLS) estimation with log(*NewInvAmount*) being the dependent variable of the first stage and Δ*RAROC* the dependent variable of the second stage. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

	RAROC with 95 % VaR		RAROC with 97 % VaR		RAROC with 99 % VaR	
	log(NewInvAmount)	ΔRAROC	log(NewInvAmount)	ΔRAROC	log(NewInvAmount)	ΔRAROC
log(NewInvAmount)		0.0098***		0.0096***		0.0102***
		(0.0015)		(0.0014)		(0.0018)
InsolvDummy	−8.5361***		−6.6425***		−3.1251***	
	(0.9055)		(0.8637)		(0.8737)	
#Campaigns	−0.2849***		−0.2876***		−0.3244***	
	(0.0688)		(0.0686)		(0.0680)	
log(Distance)	−0.0015		0.0335		−0.0859	
	(0.2415)		(0.2119)		(0.1873)	
Experience	−0.2564***	−0.0158***	−0.2856***	−0.0124***	−0.2830***	−0.0081***
	(0.0380)	(0.0010)	(0.0349)	(0.0008)	(0.0299)	(0.0007)
$(\text{Experience})^2$	0.0068***	0.0007***	0.0081***	0.0005***	0.0087***	0.0004***
	(0.0019)	(0.0000)	(0.0017)	(0.0000)	(0.0015)	(0.0000)
$\#\text{Inv}_{t-1}$	0.0554***	−0.0004***	0.0587***	−0.0003***	0.0654***	−0.0003***
	(0.0048)	(0.0001)	(0.0050)	(0.0001)	(0.0055)	(0.0001)
Gender	−0.5312***	−0.0007	−0.5213***	0.0015	−0.5619***	−0.0000
	(0.1511)	(0.0023)	(0.1395)	(0.0018)	(0.1125)	(0.0013)
Age	0.0309***	−0.0000	0.0298***	−0.0000	0.0274***	−0.0001**
	(0.0044)	(0.0001)	(0.0040)	(0.0001)	(0.0035)	(0.0001)
Time FE	yes	yes	yes	yes	yes	yes
Constant	6.6353***	0.0805***	6.5440***	0.0625***	7.7218***	0.0407***
	(1.8416)	(0.0049)	(1.7133)	(0.0040)	(1.6197)	(0.0029)
N	11,673	11,673	13,397	13,397	16,981	16,981
Adj.R^2	0.1678	0.1679	0.1688	0.1743	0.1734	0.1949

Table 6.11 Robustness Test for RAROC. The first stage is estimated using different measures of the investment behavior as depended variable. The endogenous variables are *NewInvDummy* in specification (1), *NewInvRel* in specification (2), and *#NewInv* in specification (3). Standard errors are clustered on investor level and shown in parentheses. We apply two-stage-least-squares (2SLS) estimation with $\Delta RAROC$ as the dependent variable of the second stage. The symbols *, **, and *** express significance at the 10 %, 5 %, and 1 % level, respectively. The variables are defined in Table 6.1.

Dependent Variable	(1) ΔRAROC	(2) ΔRAROC	(3) ΔRAROC
Investment Behavior	0.0557***	0.0133**	0.0118***
	(0.0107)	(0.0056)	(0.0038)
Experience	−0.0062***	−0.0065***	−0.0054***
	(0.0006)	(0.0009)	(0.0011)
$(\text{Experience})^2$	0.0003***	0.0003***	0.0002***
	(0.0000)	(0.0000)	(0.0000)
$\#\text{Inv}_{t-1}$	−0.0002*	0.0003***	−0.0006**
	(0.0001)	(0.0000)	(0.0003)
Gender	0.0002	−0.0037***	−0.0014
	(0.0011)	(0.0010)	(0.0012)
Age	−0.0000	0.0001***	0.0000
	(0.0000)	(0.0000)	(0.0000)
Time FE	yes	yes	yes
Constant	0.0250***	0.0288***	0.0302***
	(0.0029)	(0.0040)	(0.0031)
N	18,265	18,265	18,265
Adj.R^2	0.1978	0.1241	0.1840

crowdfunding, this asset class was virtually exclusively available to banks and other institutional investors. Hence, the retail investors who are active on such platforms can be assumed to have only limited experience concerning this asset class.

Rational investors should strive to diversify their loan portfolio as much as possible in order to achieve the best possible risk-return profile of their portfolio. In contrast to the stock market, on which investors can invest in exchange traded funds and have a huge universe of stocks to choose from, it is not possible to immediately obtain a diversified loan portfolio in peer-to-business lending. Investors have to invest continuously in several subsequent loan campaigns that are posted on the platform over time. We show that investors are paralyzed by the shock they appear to suffer from experiencing a default of a loan in their portfolio. Thereafter they invest in less new loans and thus stop to diversify their portfolio. This default shock bias results in a deterioration of the risk-return profile of their portfolio. Moreover, a glut of simultaneously active loan campaigns is negatively related to new loan investments. Investors appear to be unable to cope with the information provided if many campaigns are simultaneously active and tend to invest less. The deep market bias also results in a worsening of the risk-return profile of the portfolio. Furthermore, we find that experienced investors suffer from the deep market bias to a lesser extent. However, increased experience is not directly associated with improvements in the risk-return profile. By contrast, our results suggest that this relationship is U-shaped. Investors deteriorate their risk-return profile up to a certain level of experience until this relationship reverses.

One could argue that the default shock bias could also represent a rational investor behavior. In the case that investors gain additional information through experiencing a loan default, the new investment behavior could well be rational. In particular, investors might realize that the true default probability is higher than the expected default probability. Alternatively, investors may suspect that the borrower or the platform are engaging in fraudulent behavior. However, we see no reason to conclude that investors indeed gain such superior information through experiencing a loan default. Analyzing the rating classes of the loans and the average default probability provided from the platform indicates that approximately eight defaults can be expected from the 367 successfully funded loans, while only as few as five were observed during our investigation period. Furthermore, we do not find any indication of fraudulent behavior.

In the stock market, investors are confronted with a huge universe of different stocks. However, there is only little evidence to suggest that investors suffer from a deep market bias in this context. In fact, to the contrary, stocks are a common and popular asset class. Solely Iyengar et al. (2004) show that the participation rate in 401(k) retirement plans tends to decrease along with a high number of investment possibilities. A potential explanation why investors generally remain unfazed by the great amount stocks could be the wide range of investment advice for stock investments. Both the information and the choice overload could be reduced, for example, with the aid of either investment advisers or investment journals. Furthermore, the investment opportunities are less dynamic in a stock market and, therefore, investors usually have no time pressure under which to decide which stocks to buy. In peer-to-business lending, investors can only invest within the funding period. Thereafter the investment opportunity typically ceases to exist.

Both the default shock bias and the deep market bias result in fewer investments in peer-to-business lending. In particular, the platforms providing this form of investment should be concerned with trying to diminish the effects of these investment biases. A credit risk tool provided by the platform could help investors to better understand risk of defaults and also to filter relevant information in order to decrease the information load for investors. Furthermore, such a tool could help investors to invest in a way that optimizes the risk-return profile of their portfolio.

Our paper has several limitations. First, our observation period is rather short. A longer period would lead to more loan observations and enable a greater insight into the learning effects over time. Furthermore, more demographic information on the investors such as the income could help to examine the effect of the behavioral biases and experience in more detail. Moreover, an analysis of the question of whether investors suffer from these investment biases when investing in other asset classes that are predominated by institutional investors appears to be promising. Equity crowdfunding, which enables retail investors to have the opportunity to provide equity to start-ups, could be a relevant setting in this context.

6.7 Appendix

Calculation of the RAROC

In order to measure the risk-return profile of investors in peer-to-business lending, we derive the RAROC. Therefore, we divide the expected return of the portfolio by the risk capital. Due to the low interest rates for retail investors within the observation period, refinancing costs should be negligible in this context. Hence, we calculate the expected portfolio return as interest charged minus the expected loan losses (*ExpReturn*). Moreover, the VaR is a common proxy for the risk capital (see, e.g. Prokopczuk et al., 2004). Thus, we estimate the RAROC as follows:

$$RAROC = \frac{\text{Expected Portfolio Return}}{\text{Risk Capital}} = \frac{ExpReturn}{VaR}. \tag{6.3}$$

To estimate the RAROC, we calculate the value at risk of each portfolio. In a first step, we obtain the average default correlation between the borrowers in a portfolio. Düllmann and Scheule (2003) use data on German SMEs and empirically estimate their asset correlation depending on the size of the corporation and the probability of default (PD). By using a Maximum-Likelihood-Estimator, they obtain asset correlations for small corporations between 0.009 and 0.04. Our data set mainly contains small corporations. Therefore, we generally assume an asset correlation of 0.025 for corporations in different industries. As the asset correlation of corporations within the same industry tends to be higher, we choose a higher value of 0.04 for interindustry correlation. Furthermore, the platform provides an estimate of the average PD for each rating class (see Table 6.12).

Table 6.12 Probability of Default of the loans. The values of the average expected PD are derived from the peer-to-business lending platform. Note: The platform did not provide an average PD for rating class C−. Therefore, we interpolate with the average PDs of rating classes C and D.

Rating Class	A+	A	B	C	C−
Expected PD in %	0.6	1.5	2.3	3	4

With the asset correlation and the PD of the corporations, we next determine the probability of two borrowers defaulting simultaneously. This probability is given by the joint probability of default (JPD). Assuming a bivariate Gaussian distribution, the JPD is calculated as follows:

$$JPD_{i,j} = \Phi(c_i, c_j, \rho_{i,j}^{asset}) \tag{6.4}$$

with c_i being the $\Phi^{-1}(PD_i)$. We follow Frye (2008) and calculate the default correlation between two borrowers as

$$Default\ correlation = \rho_{i,j} = \frac{JPD - PD_i \cdot PD_j}{\sqrt{PD_i \cdot (1 - PD_i) \cdot PD_j \cdot (1 - PD_j)}} \tag{6.5}$$

where PD_i and PD_j are the default probabilities of each loan, according to their rating classes.

As most portfolios comprise several loans, we estimate an average default correlation using the following formula:

$$\rho_{av} = \frac{2}{(\sum_{i=1}^{n} v_i)^2 - \sum_{i=1}^{n} v_i^2} \cdot \sum_{i<j} \rho_{ij} \cdot v_i \cdot v_j \tag{6.6}$$

where v_i and v_j represent the invested amounts in the loans i and j.

In a second step, we follow Ieda et al. (2000) and first calculate the 99.5 % VaR for a homogeneous portfolio. Therefore, we examine the probability that n out of N borrowers will default and calculate the smallest m such that

$$\sum_{n=0}^{m} \int_{-\infty}^{\infty} \left\{ \Phi\left(\frac{c - \sqrt{\rho}u}{\sqrt{1-\rho}}\right) \right\}^n \left\{ 1 - \Phi\left(\frac{c - \sqrt{\rho}u}{\sqrt{1-\rho}}\right) \right\}^{N-n} \phi(u)du \binom{N}{n} \geq 0.995. \tag{6.7}$$

In this context we calculate the c on basis of the average PD of each portfolio. We obtain the value at risk of a homogenous portfolio by multiplying m with the average invested amount in a loan.

In a final step, as suggested by Ieda et al. (2000), we obtain the VaR of a heterogeneous portfolio by multiplying the VaR of a homogenous portfolio with

$$\frac{\sqrt{\rho + \left(\frac{\sqrt{\sum v_i^2}}{\sum v_i}\right)^2 (1-\rho)}}{\sqrt{\rho + \frac{1-\rho}{N}}}. \tag{6.8}$$

In this way, we correct for the fact that investors hold heterogeneous portfolios and for benefits through diversification. We divide the VaR of the portfolio by the total amount invested at each valuation date in order to measure the portfolio risk relative to the portfolio value.

Chapter 7

Conclusion

This thesis provides several contributions to the literature on alternative forms of finance. The five research papers deal with different aspects of both the financing instruments and the players on the market for alternative finance.

In the first study, we provide new insights into drivers of the funding success and the probability of default in P2P lending. Soft factors derived from the description text of loan applications help to predict the funding success if no hard information is available on the crowdlending platform. We conclude that investors do indeed take soft factors into consideration when deciding which loans to fund. However, the importance of soft factors depends on the platform design. By contrast, the probability of default is hardly related to soft factors.

The second research paper investigates the yield spread of Mittelstand bonds. We find a significant positive relationship between illiquidity and the yield spread, which persists after controlling for the default risk as well as firm and bond characteristics. The size of the liquidity premium equals approximately twice the size of the liquidity premium of speculative grade US corporate bonds. This results has important implications for Mittelstand firms issuing the bonds. They can reduce the observed yield spread of their bonds by decreasing sources of illiquidity.

The third article reveals several predictors of the probability of default and the loss rate in online invoice trading. Our results suggest that the interest rate plays a key role in explaining defaults. Additionally, the duration and the percentage funded are also significantly associated with the probability of default. We find that even though the probability of default is higher within the auction period, investors gained higher net returns during this pricing regime.

In the fourth study, we empirically analyze the communication behavior of entrepreneurs in equity crowdfunding. We find evidence for strategic communication behavior. Entrepreneurs post updates more frequently during the funding period and use more linguistic devices that enhance the feeling of group cohesion and group identity within the funding period. Moreover, our results suggest that during the funding period the probability of an update increases with the strength of the competition from other contemporary crowdfunding campaigns. Such strategic communication behavior of entrepreneurs may result in sub-optimal investment decisions by the investors as some updates only target receiving further funds rather than revealing real informational content. Since equity crowdfunding falls outside traditional securities regulation, the need for stricter regulation may arise.

In the last research paper, we examine two new investment biases—the default shock bias and the deep market bias. Using data from a peer-to-business lending platform we show that retail investors invest in less new loans and thereby underdiversify their portfolio after experiencing a loan default. Furthermore, investors appear not to be able to cope with a glut of simultaneous active loan campaigns. They cease to further diversify their portfolios. Both the default shock bias and the deep market bias result in a worsening of the risk-return profile of the portfolio. We show that experienced investors suffer from the deep market bias to a lesser extent.

This dissertation empirically investigates different sources of alternative finance for SMEs. Since many SMEs fail to obtain sufficient financing, the need for a complement to the traditional financing sources arises. While SMEs are drivers of employment and GDP growth all over the world (Ayyagari et al., 2007), the financing gap is especially pronounced in developing countries. Every year, the unmet financing needs in emerging countries equal approximately 5.2 trillion USD (IFC, 2017). Governments all over the world have already stepped in and introduced a range of policies to close the financing gap for SMEs. In this context, many countries have adopted initiatives in order to promote alternative finance and thereby to foster the access to finance for small businesses (OECD, 2018). This development is likely to boost further growth in sources of alternative finance.

Whether alternative finance indeed helps to significantly reduce the financing gap remains to be seen. Future research should evaluate the success of the new financing instruments for providing SMEs with sufficient financing. Moreover, most of the providers of alternative finance build up their business models on new technologies and claim increased transparency compared to traditional providers of financing. Thus, many of them make their data publicly available. This new data provides several promising avenues of research. On several platforms the investor and the entrepreneurial behavior constitutes a natural experiment and may also allow one to draw conclusions about causal relationships.

Bibliography

AGNEW, J., P. BALDUZZI, AND A. SUNDÉN (2003): "Portfolio choice and trading in a large 401(k) plan," *American Economic Review*, 93, 193–215.

AGRAWAL, A., C. CATALINI, AND A. GOLDFARB (2013): "Some simple economics of crowdfunding," in *Innovation Policy and the Economy, Volume 14*, National Bureau of Economic Research, NBER Chapters.

AHLERS, G. K., D. CUMMING, C. GÜNTHER, AND D. SCHWEIZER (2015): "Signaling in equity crowdfunding," *Entrepreneurship Theory and Practice*, 39, 955–980.

ALLISON, T. H., A. F. MCKENNY, AND J. C. SHORT (2013): "The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect," *Journal of Business Venturing*, 28, 690–707.

AMIHUD, Y. AND H. MENDELSON (1980): "Dealership markets: Market making with inventory," *Journal of Financial Economics*, 8, 31–53.

——— (1986): "Asset pricing and the bid–ask spread," *Journal of Financial Economics*, 17, 223–249.

——— (1987): "Trading mechanisms and stock return: An empirical investigation," *The Journal of Finance*, 42, 533–553.

AMIHUD, Y., H. MENDELSON, AND L. H. PEDERSEN (2005): "Liquidity and asset prices," *Foundations and Trends in Finance*, 1, 269–364.

ANDERSEN, S., T. HANSPAL, AND K. M. NIELSEN (2018): "Once bitten, twice shy: The power of personal experiences in risk taking," *Journal of Financial Economics*, forthcoming.

ANDREONI, J. (1990): "Impure altruism and donations to public goods: A theory of warm-glow giving," *Economic Journal*, 100, 464–477.

ANG, J. S. (1992): "On the theory of finance for privately held firms," *Journal of Small Business Finance*, 1, 185–203.

AVERY, R. B., R. W. BOSTIC, AND K. A. SAMOLYK (1998): "The role of personal wealth in small business finance," *Journal of Banking & Finance*, 22, 1019–1061.

AYYAGARI, M., T. BECK, AND A. DEMIRGUC-KUNT (2007): "Small and medium enterprises across the globe," *Small Business Economics*, 29, 415–434.

BAGEHOT, W. P. (1971): "The only game in town," *Financial Analysts Journal*, 27, 12–14.

BAJTELSMIT, V. L., A. BERNASEK, AND N. A. JIANAKOPLO (1999): "Gender differences in defined contribution pension decisions," *Financial Services Review*, 8, 1–10.

BARASINSKA, N. AND D. SCHÄFER (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, 436–452.

BARBER, B. M. AND T. ODEAN (2001): "Boys will be boys: Gender, overconfidence, and common stock investment," *The Quarterly Journal of Economics*, 116, 261–292.

BASCLE, G. (2008): "Controlling for endogeneity with instrumental variables in strategic management research," *Strategic Organization*, 6, 285–327.

BAUM, J. R. AND E. A. LOCKE (2004): "The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth," *Journal of Applied Psychology Copyright 2004 by the American Psychological Association*, 89, 587–598.

BECKER, G. S. (1971): *The Economics of Discrimination*, Chicago, London: University of Chicago Press, 2 ed.

BELLEFLAMME, P., T. LAMBERT, AND A. SCHWIENBACHER (2014): "Crowdfunding: Tapping the right crowd," *Journal of Business Venturing*, 29, 585–609.

BERGER, A. N. AND G. F. UDELL (1998): "The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle," *Journal of Banking & Finance*, 22, 613–673.

BERGER, S. C. AND F. GLEISNER (2009): "Emergence of financial intermediaries in electronic markets: The case of online P2P lending," *BuR – Business Research*, 2, 39–65.

BERNSTEIN, S., A. KORTEWEG, AND K. LAWS (2017): "Attracting early-stage investors: Evidence from a randomized field experiment," *The Journal of Finance*, 72, 509–538.

BHATT, N. AND S. TANG (2002): "Determinants of repayment in microcredit: Evidence from programs in the United States," *International Journal of Urban and Regional Research*, 26, 360–376.

BÖHME, R. AND S. PÖTZSCH (2011): "Collective exposure: Peer effect in voluntary disclosure of personal data," in *Financial Cryptography and Data Security – 15th International Conference*, ed. by G. Danezis, Springer Verlag, vol. 7035, 1–15.

BÖHME, R. AND S. PÖTZSCH (2010): "Social Lending aus der Perspektive des Datenschutzes," in *SICHERHEIT 2010 – Sicherheit, Schutz und Zuverlässigkeit*, ed. by F. Freiling, Bonner Köllen Verlag, vol. LNI 170, 317–328.

BLOCK, J., L. HORNUF, AND A. MORITZ (2018a): "Which updates during an equity crowdfunding campaign increase crowd participation?" *Small Business Economics*, 50, 3–27.

BLOCK, J. H., M. G. COLOMBO, D. J. CUMMING, AND S. VISMARA (2018b): "New players in entrepreneurial finance and why they are there," *Small Business Economics*, 50, 239–250.

BOLLEN, J., H. MAO, AND X. ZENG (2011): "Twitter mood predicts the stock market," *Journal of Computational Science*, 2, 1–8.

BRADFORD, C. S. (2012): "Crowdfunding and the fedaral securities laws," *Columbia Business Law Review*, 2012, 1–150.

BRANDT, M. W. AND K. A. KAVAJECZ (2004): "Price discovery in the U.S. treasury market: The impact of orderflow and liquidity on the yield curve," *The Journal of Finance*, 59, 2623–2654.

BRUTON, G., S. KHAVUL, D. SIEGEL, AND M. WRIGHT (2015): "New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer-to-peer innovations," *Entrepreneurship Theory and Practice*, 9–26.

BUCH, A., G. DORFLEITNER, AND M. WIMMER (2011): "Risk capital allocation for RORAC optimization," *Journal of Banking & Finance*, 35, 3001–3009.

CALVET, L., J. CAMPBELL, AND P. SODINI (2007): "Down or out: Assessing the welfare costs of household investment mistakes," *Journal of Political Economy*, 115, 707–747.

CALVET, L. E., J. Y. CAMPBELL, AND P. SODINI (2009): "Measuring the financial sophistication of households," *American Economic Review*, 99, 393–398.

CAMERON, A. AND P. TRIVEDI (2010): *Microeconometrics using Stata*, College Station, Tex.: Stata Press, rev. ed.

CAMERON, A. C. AND D. L. MILLER (2015): "A practitioners guide to cluster-robust inference," *Journal of Human Resources*, 20, 317–372.

CAMPBELL, J. Y. AND B. B. TAKSLER (2003): "Equity volatility and corporate bond yields," *The Journal of Finance*, 58, 2321–2349.

CARDON, M. S., J. WINCENT, J. SINGH, AND M. DRNOVSEK (2009): "The nature and experience of entrepreneurial passion," *Academy of Management Review*, 34, 511–532.

CASU, B., C. GIRADONE, AND P. MOLYNEUX (2006): *Introduction to banking*, Pearson Education Ltd.

CHAKRAVARTY, S. AND A. SARKAR (1999): "Liquidity in U.S. fixed income markets: A comparison of the bid-ask spread in corporate, government and municipal bond markets," Staff Reports 73, Federal Reserve Bank of New York.

CHEN, J., X. CHEN, AND X. SONG (2007a): "Comparison of the group-buying auction and the fixed pricing mechanism," *Decision Support System*, 43, 445–459.

CHEN, L., D. A. LESMOND, AND J. WEI (2007b): "Corporate yield spreads and bond liquidity," *The Journal of Finance*, 62, 119–149.

CHEN, N., A. GHOSH, AND N. S. LAMBERT (2014): "Auctions for social lending: A theoretical analysis," *Games and Economic Behavior*, 86, 367–391.

CHEN, X.-P., X. YAO, AND S. KOTHA (2009): "Entrepreneur passion and preparedness in business plan presentations: A persuasion analysis of venture capitalists' funding decisions," *Academy of Management Journal*, 52, 199–214.

CHIANG, Y.-M., D. HIRSHLEIFER, Y. QIAN, AND A. E. SHERMAN (2011): "Do investors learn from experience? Evidence from frequent IPO investors," *The Review of Financial Studies*, 24, 1560–1589.

CHOI, J. J., D. LAIBSON, B. C. MADRIAN, AND A. METRICK (2009): "Reinforcement learning and savings behavior," *The Journal of Finance*, 64, 2515–2534.

COCHRANE, J. H. (2001): *Asset Pricing*, Princeton University Press, New Jersey.

COHEN, J. (1960): "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, 20, 37–46.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND S. MARTIN (2001): "The determinants of credit spread changes," *The Journal of Finance*, 56, 2177–2207.

COLOMBO, M. G., C. FRANZONI, AND C. ROSSI-LAMASTRA (2015): "Internal social capital and the attraction of early contributions in crowdfunding," *Entrepreneurship Theory and Practice*, 39, 75–100.

COURTNEY, C., S. DUTTA, AND Y. LI (2017): "Resolving information asymmetry: Signaling, endorsement, and crowdfunding success," *Entrepreneurship Theory and Practice*, 41, 265–290.

CROSS, J. G. (1973): "A stochastic learning model of economic behavior," *The Quarterly Journal of Economics*, 87, 239–266.

CUMMING, D. AND L. HORNUF (2018): "Peer-to-business crowdlending," *SSRN Working Paper 2422111*.

CUMMING, D. AND A. KNILL (2012): "Disclosure, venture capital and entrepreneurial spawning," *Journal of International Business Studies*, 43, 563–590.

D'ACUNTO, F., N. PRABHALA, AND A. ROSSI (2018): "The promises and pitfalls of robo-advising," *The Review of Financial Studies*, forthcoming.

DARROUGH, M. N. AND N. M. STOUGHTON (1990): "Financial disclosure policy in an entry game," *Journal of Accounting and Economics*, 12, 219–243.

DAVIS, A. K., J. M. PIGER, AND L. M. SEDOR (2012): "Beyond the numbers: Measuring the information content of earnings press release language," *Contemporary Accounting Research*, 29, 845–868.

DHAR, R. (1997): "Consumer preference for a no-choice option," *Journal of Consumer Research*, 24, 215–231.

DHAR, R. AND N. ZHU (2006): "Up close and personal: Investor sophistication and the disposition effect," *Management Science*, 52, 726–740.

DIAMOND, D. W. (1984): "Financial intermediation and delegated monitoring," *The Review of Economic Studies*, 51, 393–414.

DIAMOND, D. W. AND R. E. VERRECCHIA (1991): "Disclosure, liquidity, and the cost of capital," *The Journal of Finance*, 46, 1325–1359.

DORFLEITNER, G. AND T. PFISTER (2014): "Justification of per-unit capital allocation in credit risk models," *International Journal of Theoretical and Applied Finance*, 17, 1450039/1–1450039/29.

DORFLEITNER, G., C. PRIBERNY, S. SCHUSTER, J. STOIBER, M. WEBER, I. DE CASTRO, AND J. KAMMLER (2016): "Description-text related soft information in peer-to-peer lending – Evidence from two leading European platforms," *Journal of Banking & Finance*, 64, 169–187.

DOSHI, A. R. (2016): "The impact of high-performance outliers on two-sided platforms: Evidence from crowdfunding," *SSRN Working Paper 2422111*.

DOWLING, M. AND B. M. LUCEY (2005): "The role of feelings in investor decision-making," *Journal of Economic Surveys*, 211–237.

DUARTE, J., S. SIEGEL, AND L. YOUNG (2012): "Trust and credit: The role of appearance in peer-to-peer lending," *The Review of Financial Studies*, 25, 2455–84.

DUFFIE, D. (1996): *Dynamic asset pricing theory*, Princeton University Press, New Jersey, 2nd ed.

DÜLLMANN, K. AND H. SCHEULE (2003): "Asset correlation of German corporate obligors: Its estimation, its drivers and implications for regulatory capital," *Working Paper*.

DWYER, P. D., J. H. GILKESON, AND J. A. LIST (2002): "Gender differences in revealed risk taking: Evidence from mutual fund investors," *Economics Letters*, 76, 151–158.

ECB (2017): "Survey on the access to finance of enterprises in the euro area—April to September 2017," Tech. rep., European Central Bank.

EINAV, L., C. FARRONATO, J. LEVIN, AND N. SUNDARESAN (2018): "Auctions versus posted prices in online markets," *Journal of Political Economy*, 126, 178–215.

ELTON, E. J., M. J. GRUBER, D. AGRAWAL, AND C. MANN (2001): "Explaining the rate spread on corporate bonds," *The Journal of Finance*, 56, 247–278.

EMEKTER, R., Y. TU, B. JIRASAKULDECH, AND M. LU (2015): "Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending," *Applied Economics*, 47, 54–70.

EPPLER, M. J. AND J. MENGIS (2004): "The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines," *The Information Society*, 20, 325–344.

EUROPEAN COMMISSION (2003): "Commission recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises," *Official Journal of the European Union*, (2003/361/EC).

——— (2018): "FinTech action plan: For a more competitive and innovative European financial sector," Tech. Rep. COM(2018) 109/2, European Commission.

EVERETT, C. R. (2015): "Group membership, relationship banking and loan default risk: The case of online social lending," *Banking and Finance Review*, 7, 15–54.

FASSBENDER, D. (2011): *P2P-Kreditmärkte als Finanzintermediäre*, München: GRIN Verlag.

FEILER, M. AND U. KIRSTEIN (2014): *Mittelstandsanleihen – ein Erfolgsmodell für alle Parteien*, Wiesbaden: Springer Gabler.

FENG, L. AND M. S. SEASHOLES (2005): "Do investor sophistication and trading experience eliminate behavioral biases in financial markets?" *Review of Finance*, 9, 305–351.

FERSHTMAN, C. AND U. GNEEZY (2001): "Discrimination in a segmented society: An experimental approach," *The Quarterly Journal of Economics*, 116, 351–377.

FIGUEREDO, L. AND C. K. VARNHAGEN (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, 441–458.

FISHER, L. (1959): "Determinants of risk premiums on corporate bonds," *Journal of Political Economy*, 67, 217–237.

FLEISS, J. L., B. LEVIN, AND M. C. PAIK (2003): *Statistical methods for rates and proportions; 3rd ed.*, Wiley Series in Probability and Statistics, Hoboken, NJ: Wiley.

FLEMING, M. J. (2003): "Measuring Treasury market liquidity," *Federal Reserve Bank of New York Economic Policy Review*, 9, 83–108.

FRANKS, J., N. SERRANO-VELARDE, AND O. SUSSMAN (2016): "Marketplace lending, information aggregation, and liquidity," *SSRN Working Paper 2869945*.

FREEDMAN, S. AND G. Z. JIN (2008): "Do social networks solve information problems for peer-to-peer lending? Evidence from Prosper.com," *SSRN Working Paper 1304138*.

FRYE, J. (2008): "Correlation and asset correlation in the structural portfolio model," *The Journal of Credit Risk*, 4, 75–96.

GAO, Q. AND M. LIN (2015): "Lemon or cherry? The value of texts in debt crowdfunding," *Working Paper*.

GERBER, E. AND J. HUI (2013): "Crowdfunding: Motivations and deterrents for participation," *ACM Transactions on Computing-Human Interaction*, 20, 1–32.

GFK (2010): *GfK-Konsumklima November und Ausblick auf das Weihnachtsjahr 2010*, Nürnberg: Gesellschaft für Konsumforschung.

GIUDICI, G., R. NAVA, C. R. LAMASTRA, AND C. VERECONDO (2012): "Crowdfunding: The new frontier for financing entrepreneurship," *SSRN Working Paper 2157429*.

GOETZMANN, W. N., D. KIM, A. KUMAR, AND Q. WANG (2015): "Weather-induced mood, institutional investors, and stock returns," *The Review of Financial Studies*, 28, 73–111.

GOETZMANN, W. N. AND A. KUMAR (2008): "Equity portfolio diversification," *Review of Finance*, 12, 433–463.

GONZALEZ, L. AND Y. K. LOUREIRO (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.

GREGORY, B. T., AMATTHEW W. RUTHERFORD, S. OSWALD, AND L. GARDINER (2005): "An empirical investigation of the growth cycle theory of small firm financing," *Journal of Small Business Management*, 43, 382–392.

HAMARI, J., M. SJÖKLINT, AND A. UKKONEN (2016): "The sharing economy: Why people participate in collaborative consumption," *Journal of the Association for Information Science and Technology*, 67, 2047–2059.

HAMMOND, R. G. (2013): "A structural model of competing sellers: Auctions and posted prices," *European Economic Review*, 60, 52–68.

HEALY, P. M. AND K. G. PALEPU (2001): "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature," *Journal of Accounting and Economics*, 31, 405–440.

HEMER, J. (2011): "A snapshot on crowdfunding," *Fraunhofer Institute for Systems and Innovation Research ISI*.

HENRY, E. (2008): "Are investors influenced by how earnings press releases are written?" *International Journal of Business Communication*, 45, 363–407.

HERZENSTEIN, M., S. SONENSHEIN, AND U. M. DHOLAKIA (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., M. PURI, AND J. ROCHOLL (2017): "Adverse incentives in crowdfunding," *Management Science*, 63, 587–608.

HIRSHLEIFER, D. (2001): "Investor psychology and asset pricing," *The Journal of Finance*, 56, 1533–1597.

HO, T. S. Y. AND H. R. STOLL (1981): "Optimal dealer pricing under transaction and return uncertainty," *Journal of Financial Economics*, 9, 47–73.

HORNUF, L. AND M. NEUENKIRCH (2017): "Pricing shares in equity crowdfunding," *Small Business Economics*, 48, 795–811.

HORNUF, L. AND M. SCHMITT (2016): "Success and failure in equity crowdfunding," *CESifo DICE Report*, 14, 16–22.

HORNUF, L., M. SCHMITT, AND E. STENZHORN (2018): "Equity crowdfunding in Germany and the UK: Follow-up funding and firm survival," *Max Planck Institute for Innovation and Competition Research Paper*, No. 17-09.

HORNUF, L. AND A. SCHWIENBACHER (2017): "Market mechanisms and funding dynamics in equity crowdfunding," *Journal of Corporate Finance*, forthcoming.

HUANG, G. (2017): "Multiunit uniform-price open auctions vs. posted price: The timing of bids and transaction efficiency," *SSRN Working Paper 2913132*.

HULL, J. (2015): *Risk management and financial institutions*, Pearson Education, 4th ed.

IEDA, A., K. MARUMO, AND T. YOSHIBA (2000): "A simplified method for calculating the credit risk of lending portfolios," *Monetary and Economic Studies*, 18, 49–82.

IFC (2017): "MSME finance gap – Assessment of the shortfalls and opportunities in financing micro, small and medium enterprises in emerging markets," Tech. rep., International Finance Cooperation (IFC).

IYENGAR, S. S., W. JIANG, AND G. HUBERMAN (2004): *Pension design and structure: New lessons from behavioral finance*, OUP Oxford, chap. How much choice is too much? Contributions to 401(k) retirement plans, 83–95.

IYENGAR, S. S. AND M. R. LEPPER (2000): "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology*, 76, 995–1006.

IYER, R., A. I. KHWAJA, E. F. P. LUTTMER, AND K. SHUE (2016): "Screening peers softly: Inferring the quality of small borrowers," *Management Science*, 62, 1554–1577.

JARROW, R. (1978): "The relationship between yield, risk, and the return on corporate bonds," *The Journal of Finance*, 33, 1235–1240.

KAMMLER, J. AND K. RÖDER (2013): "Die Performance von Mittelstandsanleihen am Beispiel von Bondm," *Corporate Finance biz*, 4, 55–60.

KAUSTIA, M. AND S. KNÜPFER (2008): "Do investors overweight personal experience? Evidence from IPO subscriptions," *The Journal of Finance*, 63, 2679–2702.

KELLEY, E. K. AND P. C. TETLOCK (2013): "How wise are crowds? Insights from retail orders and stock returns," *The Journal of Finance*, 68, 1229–1265.

KENNEDY, P. (2008): *A guide to econometrics*, Blackwell.

KICKSTARTER (2012): "Blockbuster effects," https://www.kickstarter.com/blog/blockbuster-effects.

KLAPPER, L. (2006): "The role of factoring for financing small and medium enterprises," *Journal of Banking & Finance*, 30, 3111–3130.

KLÖHN, L. AND L. HORNUF (2012): "Crowdinvesting in Deutschland - Markt, Rechtslage und Regulierungsperspektiven," *Zeitschrift für Bankrecht und Bankwirtschaft*, 24, 237–266.

KLÖHN, L., L. HORNUF, AND T. SCHILLING (2016): "Crowdinvesting Verträge—Inhalt, Entwicklung und praktische Bedeutung," *Zeitschrift für Bankrecht und Bankwirtschaft*, 27, 56–66.

KNÜPFER, S., E. RANTAPUSKA, AND M. SARVIMÄKI (2017): "Formative experiences and portfolio choice: Evidence from the Finnish great depression," *The Journal of Finance*, 72, 133–166.

KORNIOTIS, G. M. AND A. KUMAR (2011): "Do older investors make better investment decisions?" *The Review of Economics and Statistics*, 93, 244–265.

KREINER, D. S., S. D. SCHNAKENBERG, A. G. GREEN, M. J. COSTELLO, AND A. F. MCCLIN (2002): "Effects of spelling errors on the perception of writers," *The Journal of General Psychology*, 129, 5–17.

KUMAR, A. AND C. M. C. LEE (2006): "Retail investor sentiment and return comovements," *The Journal of Finance*, 61, 2451–2486.

KUPPUSWAMY, V. AND B. L. BAYUS (2017): "Does my contribution to your crowdfunding project matter?" *Journal of Business Venturing*, 32, 72–89.

LANDIS, J. R. AND G. G. KOCH (1977): "The measurement of observer agreement for categorical data," *Biometrics*, 33, 159–174.

LAUDENBACH, C., B. LOOS, AND J. PIRSCHEL (2017): "Recent experiences and risk taking: Trading responses to changes in the local environment," *Working Paper*.

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

LESMOND, D. A., J. P. OGDEN, AND C. A. TRZCINKA (1999): "A new estimate of transaction costs," *The Review of Financial Studies*, 12, 1113–1141.

LI, F. (2010): "The information content of forward-looking statements in corporate filings - a naive Bayesian machine learning approach," *Journal of Accounting Research*, 48, 1049–1102.

LIN, M., N. R. PRABHALA, AND S. VISWANATHAN (2013): "Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending," *Management Science*, 59, 17–35.

LIN, M. AND S. VISWANATHAN (2016): "Home bias in online investments: An empirical study of an online crowdfunding market," *Management Science*, 62, 1393–1414.

LONGSTAFF, F. A., S. MITHAL, AND E. NEIS (2005): "Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market," *The Journal of Finance*, 60, 2213–2253.

LONGSTAFF, F. A. AND E. S. SCHWARTZ (1995): "A simple approach to valuing risky fixed and floating rate debt," *The Journal of Finance*, 50, 789–819.

LOUGHRAN, T. AND B. MCDONALD (2014): "Measuring readability in financial disclosures," *The Journal of Finance*, 69, 1643–1671.

LUCAS, R. (1972): "Expectations and the neutrality of money," *Journal of Economic Theory*, 4, 103–124.

MALMENDIER, U. AND S. NAGEL (2011): "Depression babies: Do macroeconomic experiences affect risk taking?" *The Quarterly Journal of Economics*, 126, 373–416.

MARKETINVOICE LIMITED (2016): "Investor guide: Frequently asked questions," .

——— (2017a): "How do you calculate the weighted average annualised yields," .

——— (2017b): "marketinvoice SELECT: How it works," .

MERTON, R. C. (1974): "On the pricing of corporate debt: the risk structure of interest rates," *The Journal of Finance*, 29, 449–470.

——— (1987): "A simple model of capital market equilibrium with incomplete information," *The Journal of Finance*, 42, 483–510.

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, 1385–1413.

MOENNINGHOFF, S. C. AND A. WIEANDT (2013): "The future of peer-to-peer finance," *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung: Zfbf.*, 65, 466–487.

MOHAMED, A. AND A. SCHWIENBACHER (2016): "Voluntary disclosure of corporate venture capital investments," *Journal of Banking & Finance*, 68, 69–83.

MOHAMMADI, A. AND K. SHAFI (2017): "How wise are crowd? A comparative study of crowd and institutions," *Academy of Management Annual Meeting Proceedings*, forthcoming.

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

MULLER, P., J. JULIUS, D. HERR, L. KOCH, V. PEYCHEVA, AND S. MCKIERNAN (2017): *Annual report on European SMEs 2016/2017 - Focus on self-employment*, European Union.

MÜLLER, S., K. BRACKSCHULZE, AND M. D. MAYER-FRIEDRICH (2011): *Finanzierung mittelständischer Unternehmen nach Basel III: Selbstrating, Risikocontrolling und Finanzierungsalternativen*, Munich: Vahlen, 2nd ed.

NICOLOSI, G., L. PENG, AND N. ZHU (2009): "Do individual investors learn from their trading experience?" *Journal of Financial Markets*, 12, 317–336.

ODEAN, T. (2002): "Are investors reluctant to realize their losses?" *The Journal of Finance*, 53, 1775–1798.

OECD (2018): "Enhancing SME access to diversified financing instruments," *SME Ministerial Conference*.

PARK, M., M. AIKEN, L. TOBIN, AND M. VANJANI (2010): "Spelling and grammatical errors in electronic meetings," *Issues in Information Systems*, XI, 384–391.

PENNEBAKER, J. W., M. E. FRANCIS, AND R. J. BOOTH (2001): *Linguistic inquiry and word count - LIWC 2001*, Erlbaum.

PEROLD, A. F. (2005): "Capital allocation in financial firms," *Journal of Applied Corporate Finance*, 17, 110–118.

POPE, D. G. AND J. R. SYDNOR (2011): "What's in a picture? Evidence of discrimination from Prosper.com," *Journal of Human Resources*, 46, 53–92.

PROKOPCZUK, M., S. T.RACHEV, G. SCHINDLMAYR, AND S. TRÜCK (2004): "Quantifying risk in the electricity business: A RAROC-based approach," *Energy Economics*, 29, 1033–1049.

PYNTE, J., A. KENNEDY, AND S. DUCROT (2004): "The influence of parafoveal typographical errors on eye movements in reading," *European Journal of Cognitive Psychology*, 16, 178–202.

RAVINA, E. (2012): "Love & loans: The effect of beauty and personal characteristics in credit markets," *SSRN Working Paper 1101647*.

REIS, H. AND C. JUDD (2014): *Handbook of research methods in social and personality psychology*, Cambridge University Press.

RENNEBOOG, L., J. T. HORST, AND C. ZHANG (2008): "Socially responsible investments: Institutional aspects, performance, and investor behavior," *Journal of Banking & Finance*, 32, 1723–1742.

RIVERS, D. AND Q. H. VUONG (1988): "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39, 347–366.

ROSETT, R. (1959): "A statistical model of friction in economics," *Econometrica*, 27, 263–267.

ROY, A. (2016): "Strategic flexibility for reconfiguring loan portfolios in Indian banks," *Strategic Change*, 25, 525–536.

SAHLMAN, W. A. (1990): "The structure and governance of venture-capital organizations," *Journal of Financial Economics*, 27, 473–521.

SARIG, O. AND A. WARGA (1989): "Bond price data and bond market liquidity," *Journal of Financial and Quantitative Analysis*, 24, 367–378.

SCHINDELE, A. AND A. SZCZESNY (2015): "The impact of Basel II on the debt costs of German SMEs," *Journal of Business Economics*, forthcoming.

SCHÖNING, S. (2014): "Risikoadäquate Verzinsung von Mittelstandsanleihen – Eine exemplarische Analyse für das Bondm-Segment der Boerse Stuttgart AG," *Corporate Finance*, 11, 470–481.

SCHWIENBACHER, A. (2007): "A theoretical analysis of optimal financing strategies for different types of capital-constrained entrepreneurs," *Journal of Business Venturing*, 22, 753–781.

SEASHOLES, M. S. AND N. ZHU (2010): "Individual investors and local bias," *The Journal of Finance*, 65, 1987–2010.

SERRANO-CINCA, C., B. GUTIÉRREZ-NIETO, AND L. LÓPEZ-PALACIOS (2015): "Determinants of default in P2P lending," *PLoS ONE*, 10.

SEXTON, J. B. AND R. L. HELMREICH (2000): "Analyzing cockpit communications: The links between language, performance, error, and workload," *Journal of Human Performance in Extreme Environments*, 5, 63–68.

SHAFIR, E., I. SIMONSON, AND A. TVERSKY (1993): "Reason-based choice," *Cognition*, 49, 11–36.

SHEFRIN, H. AND M. STATMAN (1985): "The disposition to sell winners too early and ride losers too long: Theory and evidence," *The Journal of Finance*, 40, 777–790.

SIGNORI, A. AND S. VISMARA (2017): "Does success bring success? The post-offering lives of equity-crowdfunded firms," *Journal of Corporate Finance*, forthcoming.

——— (2018): "Returns on investments in equity crowdfunding," *Journal of Corporate Finance*, forthcoming.

SOLOMON, J. AND R. WASH (2014): "Coordinating donors on crowdfunding websites," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, New York, NY, USA: ACM, CSCW '14, 38–48.

SONENSHEIN, S., M. HERZENSTEIN, AND U. M. DHOLAKIA (2011): "How accounts shape lending decisions through fostering perceived trustworthiness," *Organizational Behavior and Human Decision Processes*, 115, 69–84.

SOUFANI, K. (2002a): "On the determinants of factoring as a financing choice: evidence from the UK," *Journal of Economics and Business*, 54, 239–252.

——— (2002b): "The decision to finance account receivables: The factoring option," *Managerial and Decision Economics*, 23, 21–32.

SPENCE, M. (2002): "Signaling in retrospect and the informational structure of markets," *American Economic Review*, 92, 434–549.

STAIGER, D. AND J. H. STOCK (1997): "Instrumental variables regression with weak instruments," *Econometrica*, 65, 557–586.

STIGLITZ, J. E. AND A. WEISS (1981): "Credit rationing in markets with imperfect information," *American Economic Review*, 71, 393–410.

STOLL, H. R. (2000): "Friction," *The Journal of Finance*, 55, 1479–1514.

TAUSCZIK, Y. R. AND J. W. PENNEBAKER (2010): "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, 29, 24–54.

TOBIN, J. (1958): "Estimation of relationships for limited dependent variables," *Econometrica*, 26, 24–36.

TUTTLE, B. AND F. BURTON (1999): "The effects of a modest incentive on information overload in an investment analysis task," *Accounting, Organizations and Society*, 24, 673–687.

UNITED KINGDOM STATISTICS AUTHORITY (2017): "Insolvency statistics—January to March 2017 (Q1 2017)," .

VAN WINGERDEN, R. AND J. RYAN (2011): "Fighting for funds: An exploratory study into the field of crowdfunding," .

VEGA, C. (2006): "Stock price reaction to public and private information," *Journal of Financial Economics*, 82, 103–133.

VISMARA, S. (2016): "Information cascades among investors in equity crowdfunding," *Entrepreneurship Theory and Practice*, forthcoming.

VULKAN, N., T. ÅSTEBRO, AND M. F. SIERRA (2016): "Equity crowdfunding: A new phenomena," *Journal of Business Venturing Insights*, 5, 37–49.

WANG, R. (1993): "Auctions versus posted-price selling," *American Economic Review*, 83, 838–851.

WEI, Z. AND M. LIN (2016): "Market mechanisms in online peer-to-peer lending," *Management Science*, Accepted Author Manuscript.

WEISS, G. N. F., K. PELGER, AND A. HORSCH (2010): "Mitigating adverse selection in P2P lending – Empirical evidence from Prosper.com," *SSRN Working Paper 1650774*.

WOLF, M., A. B. HORN, M. R. MEHL, S. HAUG, J. W. PENNEBAKER, AND H. KORDY (2008): "Computergestützte quantitative Textanalyse - Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count," *Diagnostica*, 54, 85–98.

WSI (2013): *Beschäftigte mit Weihnachtsgeld 2013*, Düsseldorf: Wirtschafts- und Sozialwissenschaftliches Institut in der Hans-Böckler-Stiftung.

XU, A., X. YANG, H. RAO, W.-T. FU, S.-W. HUANG, AND B. P. BAILEY (2014): "Show me the money! An analysis of project updates during crowdfunding campaigns," *Proceedings of the 32nd Annual ACM Conference for Human Factors in Computing Systems*, ACM: 591–600.

ZHANG, B., P. BAECK, T. ZIEGLER, J. BONE, AND K. GARVEY (2016): "Pushing boundaries: The 2015 UK alternative finance industry report ," Cambridge Centre for Alternative Finance.

ZHENG, T. (2000): "Characteristics of Australian political language rhetoric: Tactics of gaining public support and shirking responsibility," *Journal of Intercultural Communication*, 4.