The role of financial literacy on portfolio diversification in peer-to-peer lending

David Mintjes

July 7, 2016

Master Thesis

Erasmus University Rotterdam
Department: School of Economics

Supervisor: Dr. Ruben Cox

Abstract. In general, individual households tend to under-diversify their portfolios, leading to high idiosyncratic risk. The market demands no compensation for this risk because it could be easily reduced by diversification, leading to sub optimal household portfolios. Combining a large survey data set and personal portfolio data from 5 participating intermediaries, I quantitatively provide insights in the Dutch peer-to-peer lending market. The results indicate that financial literacy of crowdfunding investors correlates similarly to characteristics of general investors in other relevant papers. I confirm that a substantial part of Dutch households hold highly under diversified crowdfunding portfolios and that financial literacy positively affects diversification. This effect is strong enough to show significant results after controlling for demographics and behavioral factors. Moreover, I find evidence for the familiarity bias, those who know the fund seeker invest on average more in one project and hold smaller portfolios. Additionally, the average of the approximated portfolio size to net worth (SNWR) is 0.19, almost twice as large as recommended by the regulatory agency. Especially the investor who is low diversified and has a high SNWR bears substantial risks for which the market does not compensate.

Contents

1 Introduction

Peer-to-peer financing, where borrowers are matched with lenders by web-based platforms, is a new phenomenon of the last decade. This peer-to-peer financing, also known as crowdfunding, refers to the use of small amounts of capital to finance ventures or projects, the funding, by a large number of investors, the crowd. Since 2012, the investments made through crowdfunding platforms has been growing exponentially. In 2015, entrepreneurs raised more than 128 million euros of new investments (Douw & Koren, 2016). Experiencing this rapid growth, crowdfunding becomes more and more a viable alternative for traditional financing. This holds for both the fund seeker, which can expect better conditions than via traditional bank loans, as well as the investor, who could get a significant higher return relative to their saving account. In addition to a high return, the online platforms make crowdfunding very accessible for non-professional investors. However, there are also substantial risks involved in transferring money directly without the intervention of a financial institution (Freitas & Rochet, 1997). A few examples are liquidity risks (Diamond & Dybvig, 1983), information asymmetries, lack of monitoring, and bad risk management (Moenninghoff & Wieandt, 2013). Moreover, non-professionals involved in crowdfunding should not underestimate the risk that comes with the subordinated position of bonds, the fixed duration, zero-coupon constructions, and the limited tradability due to a lack of a secondary market for crowdfunding loans.

In addition to the previous risk factors, crowdfunding risk has a large idiosyncratic component, because almost all crowdfunding projects are in the early-stage of funding. This risk could be easily reduced by diversification. All asset pricing models assume that securities are priced by a diversified investor, who demands little compensation for holding idiosyncratic risk. The classical, frightened image of regulators is that non-professional investors put all their savings in one promising project. To which extent is this statement justified? How sophisticated are crowdfunding investors? Which individual characteristics are associated with financial sophistication and under-diversification? Does economic behavior of households vary with financial literacy?

This study contributes to the existing literature in both an academic and a practical manner. First, relevant studies in the field of portfolio diversification (Goetzmann & Kumar, 2008) and the link with financial literacy (Von Gaudecker, 2015) are all applied to the equity market. There are many measurements to estimate diversification in an equity portfolio by looking at Sharpe ratios, normalized portfolio variance (Calvet, Campbell, & Sodini, 2009) or mean-variance relationships (Von Gaudecker, 2015). However, in a lending portfolio of a crowdfunder, it is much harder to measure diversification. I come up with a new formula to proxy to which extend the crowdfunding portfolio reflects the market, in this case all funded projects on the platform . Moreover, I use a unique set up in collecting the data. In the Netherlands, prominent literature often use the Dutch Central Bank Household Survey (DHS) for detailed information about financial literacy and portfolio diversification (Van Rooij, Lusardi, & Alessie, 2011; Von Gaudecker, 2015). Individual transaction data of the DHS is self-reported, which means that respondents are asked to report the number of stocks and mutual funds they poses and the corresponding names and quantities of all assets. My study relies on survey data but uniquely matches the respondent's portfolio with real transaction data of the platform. In contrast to self-reported assets, this feature result in very accurate estimations of household portfolios.

Second, besides the academic component, this research contributes to new insights for regulation of the fast developing peer-to-peer lending market. The rapid growth of the industry together with all risk mentioned earlier seems to have woken several supervisory authorities. Regulators try to develop this regulation in such a way that, on the one side, the consumer is protected, and on the other side, the distortion of the growing phase of the industry is limited.

On January 29th, 2016, the Securities and Exchange Commission (SEC) has adopted new rules related to crowdfunding. The new regulation is designed to assist smaller companies with capital information and provide investors with additional protection (SEC, 2015). More specifically, the regulation allows a company to raise up to 1 million dollars a year through crowdfunding offerings. Individuals are permitted to invest in securities-based crowdfunding transaction subject to certain investment limits, which mainly depends on both net worth and annual income. If either the annual income or net worth is less than 100 000 dollar, the investor can invest up to 2000 dollar or five per cent of the lesser of annual income or net worth. If both net worth and annual income are more than 100 000 dollar, the investor can invest up to 10 per cent of annual income or net worth. This regulation is designed to help the public understand the crowdfunding environment so they can make informed decisions about the risks and rewards in the early-stage businesses.

The Netherlands Autority of Financial Markets (AFM) has also taken steps regarding consumer protection. In responds to the 'Report on crowdfunding' (AFM, 2014), which they published in December 2014, new regulation were announced towards the developing industry. The package of regulation was endorsed at the start of April and mainly focuses on the transaction between the platform, this is the intermediary, and the consumer, in this case the non-professional investor. Platforms are obliged to send a cancellation email to the investor after every transaction. Moreover, capital lenders are required to take a so-called "investor test". Here, questions are asked about investors' wealth, diversification of capital and the awareness of potential risks (AFM, 2015). The "investor test" encountered strong resistance from the platforms. The trade association "Branchevereniging Nederland Crowdfunding" fears that these rules will break down the development of the young industry (Douw, 2015). Because most regulations are not quantitatively supported, insights in crowdfunding behavior would provide better fundamentals for new regulatory actions.

In this study I use a sample that contain 931 observations. I focus on loan-based transactions, also known as crowdlending. Loan-based is by far (77%) the most popular crowdfunding product in the Netherlands. Moreover, donation and reward based crowdfunding is not under supervision of the authority. From all observations, I am able to match 863 individual portfolios across 5 different crowdfunding platforms with the survey data. As far as I know, this set-up results in the most extensive data set in the field of crowdfunding. In cooperation with the AFM, I collected this data in March 2016, just before the new regulations were applied. In October, the AFM will extend my research with follow-up research to see if new regulations have a positive effect on reducing the risks. My study focuses on two key issues. First, I summarise individual characteristics of crowdfunding investors and link these to a proxy for financial literacy. Second, I derive different measurements for diversification in the crowdfunding market and show to what extend diversification can be explained by financial literacy and other household characteristics. I find that there is a lot of variation in economic behavior in the crowdfunding market and across platforms. Based on goals, expectations and motivations I distinguish three types of investors, feel-good funders, return-seeking investors and high-stake investors. My results indicate that financial literacy correlates with gender, education and the financial situation of households, confirming other financial literacy literature that is applied to the equity market.(Van Rooij et al., 2011; Calvet et al., 2009; Lusardi & Mitchell, 2007). Regarding diversification, I show that a substantial proportion of households hold under diversified portfolios, but also a large proportion that act very sophisticated in the crowdfunding market. I find evidence that financial literacy positively affects diversification which corresponds to Von Gaudecker (2015), who investigate equity portfolios. Diversification levels increase with invested capital, age, education, execution only investors and total participating platforms. Investors that participate in the stock market do diversify better. There is significant variation across platforms. I contribute to existing research with a new measurement for diversification, that is especially useful for markets with limited

supply of total assets.
Some limitations in this research do exist. Doing research on household finance is more challenging than in the field of the traditional financial institutions, i.e. deposit takers, investors and insurers. In line with research of Campbell (2006), who concludes that households are unwilling to disclose their financial situation, I find that a large proportion of households report maximum income and wealth, which, in all likelihood, could be upward biased. Moreover, for investors that participate on one platform, I find variation in self-reported assets and assets according to the platform data, which means that households do not accurately report their portfolio data.
This paper is organized as follows: in the next section I provide a literature review on behavioral aspects of non-professionals, the crowdfunding industry, financial literacy and diversification. In section 3, I describe the data set and explain the measurements for financial literacy and diversification. In section 4, the relationship between financial literacy and investor characteristics is described. In section 5, I link financial literacy with diversification. In section 6, I discuss the results and provide several extensions. In section 7, I provide some policy implications for regulators. Section 8 concludes the research, suggests areas for further research and provides some policy recommendations for regulators.

2 Theoretical framework

2.1 Behavioral aspects of non-professional investors

Barberis and Thaler (2003), Campbell (2006) and Barber and Odean (2011) have written extensive working papers in the field of behavioral finance. By combining the relevant literature I can address several forms of portfolio choices that result in under performance of non-professional investors. Most of their theory is applied to the equity market, so I list only subjects that are relevant for the peer-to-peer lending market.

2.1.1 Insufficient diversification of stocks

Many households own relatively few different stocks. According to historical household surveys, the median of stocks held was two until 2001, when it rose to three (Van Horne, Blume, & Friend, 1975), (Kelly, 1995), (Polkovnichenko, 2005). This seems a jaw-dropping result, but we should not forget that households hold a substantial part of equity indirectly, through mutual funds or retirement accounts, and these indirect holdings tend to be much better diversified. Consequently, it is not clear that concentrated individual stockholdings have a large effect on household portfolio risk. In a more recent paper, Von Gaudecker (2015) confirms under diversification and shows that this is incurred by households who neither turn to external investment advice nor have good skills in basic investment operations. The author concludes that policies should provide financial numeracy. Additionally, increasing the uptake of financial advice could help the regulators to reduce welfare losses for households holding sub optimal portfolios.

2.1.2 Naive diversification

Many investors seem to use strategies as simple as allocating *1/n* of their savings to each of the n available investment options, whatever those options are. (Benartzi & Thaler, 2001). In a laboratory setting, they distinguish three allocation options; a) between a stock fund and a bond fund b) between a stock fund and a balanced fund and c) between a bond fund and a balanced fund. They find that people divide their money equally in all three cases, leading to totally different exposure to stocks relative to bonds. Moreover, they find that in 401(k) retirement plans investors will allocate more to stocks because 401(k) plans offer predominantly stock funds.

2.1.3 Local bias

In household-level data, local bias appears in two forms; first, people prefer domestic over foreign assets, and second, choose regional over non-regional companies (French & Poterba, 1991) (Cooper & Kaplanis, 1994). Huberman (2001) tests preferences in telecommunication stocks and finds that individual investors prefer to own stocks of their local telecommunications company. Another research by Zhu (2002), who uses brokerage account data, finds that the regional bias is stronger among investors who do not own international stocks; this suggests a connection between the two forms of local bias. According to Barber and Odean (2011), the main reason for local bias is overestimation of the local economy.

2.1.4 Familiarity bias

In addition to the local bias, many US households have large holdings in the stock of their employer, particular within their 401(k) retirement savings accounts (Mitchell & Utkus, 2002). Some of these holdings result from employer policies. However, Benartzi (2001) shows that, as a result of overestimation of the company's economic conditions, a substantial fraction of unrestricted employee contributions go to employer stock rather than diversified alternatives. This is especially true when the employer stock has recently performed well, suggesting that households extrapolate the past performance of the employer. Consequently, the exposure to the employer in both job perspectives as well as company stocks will result in more exposure to idiosyncratic risk.

2.1.5 Excessive trading of individual investors

Barber and Odean (2000) research the relationship between stock turnover and net return. They find that people who are in the lowest turnover quintile outperformed the S&P 500 Index fund while the 2nd to the 5th quintile all under perform. Barber and Odean conclude that overconfidence is the explanation for excessive trading. Barber and Odean (2002) extended their research with an experiment of online traders and traditional traders. They especially focus on investors that performed well, beating the market by more than 2%, before going online. After going online, they trade more actively and more speculatively, and less profitably, resulting in under performance by more than 3% annually.

2.1.6 Attention effect

Due to limited amount of time and resources, individuals could not search systematically through all assets that are available. They typically buy stocks that have caught their attention. Examples of attention drawing effects are extreme past performance (good or bad), high trading volume stocks, newly announced stocks and media coverage stocks. Engelberg and Parsons (2011) find that local media coverage increases local trading volume by 37%-75%. Moreover, larger firms experience more turnover and unexpected extreme earnings result in more trading volume.

2.1.7 Heterogeneity in portfolio choices

Finally, there is heterogeneity in the strength of these effects across households. Puri and Robinson (2007), for example, show that households with optimistic beliefs about their life expectancy place a higher portfolio weight on individual stocks even though they do not place a higher overall weight on equities. Graham, Harvey, and Huang (2009) find that investors who claim to be comfortable with investment products also tend to trade more frequently and are more internationally diversified. Goetzmann and Kumar (2008) find that older, wealthier, more experienced, and financially sophisticated investors, who diversify with assets other than domestic stocks also hold relatively better diversified domestic stock portfolios.

2.2 A closer look into crowdfunding

Crowdfunding, in general, is a new phenomena of the last decade. The crowdfunding investor and its behavior are not widely discussed in the academic literature. I highlight the most important literature that is relevant for this paper and shed light on the asset characteristics and the attracted investors. Schwienbacher and Larralde (2010) offered one of the first descriptions of crowdfunding, by focusing on a case study of a French music crowdfunding startup. They describe crowdfunding as:

"an open call, essentially through the Internet, for the provision of financial resources either in form of donation or in exchange for some form of reward and/or voting rights in order to support initiatives for specific purposes".

The definition of Schwienbacher and Larralde (2010) corresponds to what the AFM reports on their website:

"Crowdfunding is investing or lending relatively small amounts by a large group of investors (lenders) to an individual or enterprises (fund seekers) via an online platform. Platforms bring lenders and fund seekers together. There are different forms of crowdfunding. Only loan-based and equity based are under supervision of the AFM."

The two definitions partly coincide, but the AFM explicitly distinguishes financial rewards from donations or other reward based crowdfunding. Other literature does the same. Belleflamme, Lambert, and Schwienbacher (2014) look when individuals would choose financial products rather than donation or reward based transactions. They find that pre-ordered products, a form of reward based crowdfunding, are preferred if the initial capital requirement is relatively small compared with market size. Otherwise, people would prefer financial products in return.
Moenninghoff and Wieandt (2013) take a closer look at the risk management side of peer-to-peer lending. A lot of risk that traditional investments do not face are present in crowdfunding. Therefore, peer-to-peer finance requires a risk management mechanism. The authors conclude that the increased use of peer-to-peer finance by institutional investors indicates a movement towards funded risk management where the institution, that provides guarantees, is part of the chain between the household and the platform. The main risk that still inhibits large insititutions from the crowdfunding market is liquidity risk. When the secundairy market for crowdfunding will be introduced, peer-to-peer platforms will transform into institutional-to-peer platforms (Moenninghoff & Wieandt, 2013).
Wardrop, Zhang, Rau, and Gray (2015) address the need for empirical-based research. They state that there is almost no objective, independent and reliable research available to scientifically benchmarks. No reports exist that regularly track the development of key alternative finance markets in European countries. The sector is critically under-studied and often misunderstood because there is no universally accepted taxonomy in Europe to describe and distinguish between the various forms of alternative financing activity. Collins, Swart, and Zhang (2013) and

Wardrop et al. (2015), together with their research partners (e.g. Nesta and UC Berkely, they give empirically insides in the development in the UK relative to Europe. They conclude that the UK outpaces the rest of Europe. They aim for a policy that stimulates and facilitates the growth of the industry, while balancing this against the need for investor protection. This advice corresponds with the conclusions of the AFM in their 'Report on crowdfunding' (AFM, 2014). In this short literature review, I provide short insights in the academical view of crowdfunding. There exist more quantitatively research, but most of them use data of Kickstarter, which is one of the largest platforms in the world.(Mollick, 2014; Marom, Robb, & Sade, 2015). However, these data all rely on reward-based crowdfunding, which represents a totally different market than equity- or loan-based crowdfunding. In the next section, I report literature on financial literacy of investors. In section 2.4, I link financial literacy, diversification, and crowdfunding literature together and list my hypotheses.

2.3 Financial literacy and investment behavior

Financial literacy is widely researched in academics. Calvet et al. (2009) measure the financial sophistication of Swedish households. They construct an index of financial sophistication that explains a set of three main investment mistakes; under diversification, risky share inertia, and the disposition effect (tendency to sell winning stocks and hold the losing stocks). They find that financial wealth, household size, education and financial experience have a positive effect on financial sophistication. The index correlates strongly positive with the share of risky assets held in the portfolio. In their previous paper (Calvet, Campbell, & Sodini, 2006), they conclude that more sophisticated agents tend to invest more aggressively and make smaller mistakes.
Other papers use survey question to measure financial literacy. Lusardi and Mitchell (2007) rely on three main questions. The subjects of the questions are compound interest, inflation and risk diversification. They show that those who score low on these subjects are less likely to plan for retirement and accumulate much less wealth. Another paper of Van Rooij et al. (2011) construct comprehensive survey questions to measure financial literacy. They find that financial literacy affects financial decision-making; investors with higher literacy are more likely to invest in stocks. Von Gaudecker (2015) and Goetzmann and Kumar (2008) link financial literacy to equity portfolio diversification. They both find positive relationships.

2.4 Hypotheses

The previous subsections highlight academic literature on diversification, the crowdfunding market, and financial literacy. First, I describe literature which shows that non-professional investors are insufficiently diversified. Second, I show that crowdfunding is a heterogeneous product, which could be donation based, pre-ordered based, loan based or equity based. The most popular peer-to-peer finance construction in the Netherlands is crowdlending (72% of the total crowdfunding market) (Douw & Koren, 2016). Crowdlending, when investors lend directly small amounts in exchange for a financial return through a commercial loan agreement, shows very similar characteristics as the traditional bond market. Third, I summarise literature on financial literacy, and its link with diversification.
In this paper I test several hypotheses, which will answer the following research question:

What is the role of financial literacy on portfolio diversification in peer-to-peer lending?

First, I want to investigate the level of sophistication of crowdfunding investors. Do they rely on their financial knowledge? And are they heterogeneous in their financial literacy. Other literacy literature show that non-professionals vary in literacy (Calvet et al., 2009; Lusardi & Mitchell, 2007; Goetzmann & Kumar, 2008; Van Rooij et al., 2011). Therefore, I test the following hypothesis:

Hypothesis 1: There is heterogeneity in financial literacy among crowdfunding investors.

If I can reject the null hypothesis of homogeneity in financial literacy, some factors should explain the variety of literacy. Calvet et al. (2006, 2009) find that factors such as education, gender, wealth, and experience positively affect financial literacy of Swedish households. Van Rooij et al. (2011) find similar results for Dutch households. I test the following hypotheses:

Hypothesis 2: There is a positive effect of education, being male, wealth, and experience on financial literacy.

If hypothesis one is true, I want to link the variation in financial literacy to economic behavior in the form of portfolio diversification. As theoretical framework, I focus on the research of Goetzmann and Kumar (2008); Von Gaudecker (2015); Calvet et al. (2009), and expect that this relationship also exists in the peer-to-peer lending market.

Hypothesis 3: Financial literacy positively affects portfolio diversification.

Lastly, I want to extend my research by looking at other factors than financial literacy, that could explain under diversification. In the literature review, I summarise many forms of behavioral influences in the equity market. These could also exist in the crowdfunding market. For regulators, it is very helpful to get insights in the main factors that causes the risk of under diversification. Based on the local bias and familiarity phenomena, I want to test the following hypotheses.

Hypothesis 4: Investors who know the fund seeker have under diversified portfolios and overinvest in these projects.

3 The data

3.1 The sample environment

In cooperation with the Netherlands Authority of Financial Markets (AFM) and several platforms, I use survey data combined with real transaction data directly from the alternative-finance intermediaries. Because supervision on crowdfunding is in the early-stage and mainly country-specific, this study will focus on intermediaries that mainly facilitate funding for Dutch individuals and businesses. In addition to this geographical focus, I only look at loan-based transactions. Loan-based is by far (77%) the most popular crowdfunding product in the Netherlands. Moreover, donation and reward based crowdfunding is not under supervision of the authority. The sample contains data of the four firms, QRST, with largest market capitalization. Additionally, one firm with a substantial different business model, firm U, is included in the data. According to the Crowdfunding Monitor Document[1], the five platforms capture approximately 71 per cent of the market. In this way, the sample is a good proxy for the total crowdfunding market with a

[1]Every 6 months, crowdfunding intermediaries are obligated to report data about total funded projects, total revenue, total defaults on the platform etc. The Crowdfunding Monitor Document is only internally available at the AFM.

good reflection of the market and enough heterogeneity in investor characteristics since these crowdfunding platforms target different investor segments.[2] The aggregate-level data contain specific private financial information. In the data set, I pseudonymize all individual investors and merge all platform data together. In some regressions, dummies for differences among platforms are included, but I do not report the coefficients. The AFM survey opened on 15 March 2016 and closed on 31 March 2016. A total of 943 investors have responded. They gave permission to the AFM to match their survey results with their personal portfolio data and transactions on the platform. A detailed summary of variable definitions is provided in appendix D.

3.2 Data cleaning

Out of the 943 respondents, two people reported zero assets in their portfolio. Since the purpose of this research is to investigate crowdfunders' behavior, I only include investors that participate in at least one project. The 941 email addresses of the participants were sent to the participating platforms as identifier. One small platform (the 6th platform) had only 10 respondents, so I did not asked this platform to send personal transaction data. I find similar result and cannot detect any bias in observable self-reported characteristics when dropping households of the 6th platform, so for consistency between observations in the questionaire data and the transaction data, I drop the 10 observations of platform 6. From the 931 respondents, the five platforms were able to collect 863 personal portfolios in their databases. There are several reasons that not all data were matched. The first reason is that some participants report different email addresses in their survey than on the platform, or, others are inconsistent with capital letters for example. Additionally, some participants who submitted the survey only hold donation-based projects in their portfolio.
Two variables contain some serious outliers. Self-reported assets in the portfolio and self-reported invested capital substantially deviate from the assets and invested capital according to the platform data. Note that the self-reported data and the platformdata should only be the same if the participant is active on only one platform. Even though it is clear that these variables should have the same outcome for participants that are active on one platform, I do not correct for two reasons. First, some projects might not be reported by the platform because it is not (yet) fully funded. Second, it is interesting to look at the estimation of the investor. It might be that he overestimates his portfolio size. In this case, I estimate diversification both for self-reported assets and platform assets. Some people did not understand the question or made typographical errors. Their outcomes causes substantial outliers in the data. I make three corrections for self-reported capital and 22 corrections for self-reported assets. I use the following formula to approximate the correct values.

$$Assets(selfreported) = Capital(selfreported) * \frac{Assets(platformdata)}{Capital(platformdata)} \quad (1)$$

$$Capital(selfreported) = Assets(selfreported) * \frac{Capital(platformdata)}{Assets(platformdata)} \quad (2)$$

3.3 Variable definitions

3.3.1 Measure of financial literacy

Several methods exist to test financial literacy and financial sophistication. Financial sophistication is defined as the ability of an individual to make optimal investment decisions. This paper

[2] More information about investor segments in section 3.3.3.

follows the measurement of Van Rooij et al. (2011), who constructed two sets of questions, basic and advanced questions. The AFM survey contains 3 selected questions that are most relevant for crowdfunding investors. The first question is about the relation between interest rate and bond prices. The second question covers the difference between mutual funds and stocks. The third question covers the working of risk diversification. Table 1 shows the response distribution.

Table 1

Determination of financial literacy. Panel A reports the weighted proportion of households that answer correct, incorrect, "do not know" or "refusal" for each question separately. Panel B reports the Spearman's correlation of the dummies (1=correct, 0=incorrect, do not know or refusal) and the Cronbach's alpha test, measuring internal consistency. Panel C reports the distribution of the total number of correct, incorrect, "do not know" and "refusal" answers on the questions together. The data is from the AFM crowdfunding survey, collected in March 2016.

Panel A: Weighted percentages of correct and incorrect answers (N = 931)

	Question 1	Question 2	Question 3
Correct	59.4	81.1	91.5
Incorrect	17.0	5.6	5.6
Do not know	21.7	12.6	2.9
Refusal	1.9	0.8	-

Panel B: Spearman's correlations and internal consistency. *p <.0.5.

	Question 1	Question 2	Question 3
Question 1	1		
Question 2	0.50*	1	
Question 3	0.28*	0.44*	1
	Cronbach's α=0.667		

Panel C: Summary of responses, weighted number of correct and incorrect answers (N = 931)

	None	1	2	all	mean
Correct	4.2	11.7	32.0	52.1	2.32
Incorrect	74.0	23.9	2.2	-	0.28
Do not know	74.3	16.4	7.0	2.3	0.37
Refusal	97.6	2.0	0.3	-	0.03

There are some drawbacks in this measurement. First, this survey only contains three questions, Van Rooij et al. (2011) uses for example 16 questions, they split these questions into two different levels, and use two different loadings on the questions. Moreover, they prove that wording of questions matters, which indicates that participants have difficultiy with understanding the question. By using different sentences for the same questions they control for this. Second, Cronbach's alpha, that measures internal consistency, is very low (<0.5). This means that the scale reliability of the three questions as a group is insufficient. A possible explanation could be that the answer distribution, correct, incorrect, do not know, and refusal, is limited. Another reason could be the limited amount of questions. Even though there exist limitations in the measurement of financial literacy, it could still be very useful to rank individuals based on financial sophistication.[3] An alternative methodology to determine literacy is factor analytics. However, by using ordered logistic regressions, simply dividing financial literacy tertiles leads to coefficients that are easier to interpret.

3.3.2 Measure of diversification

This paper distinguishes three measurements of diversification, following Van Horne et al. (1975). The first measurement is simply calculated by taking the sum of all crowdfunding projects in the

[3]See the appendix A.1 for details about classifying the households in low, average and high literacy.

portfolio.

$$DIV1 = N_p \tag{3}$$

where N_p is the self-reported number of investment projects in the investors portfolio. Although this method is widely used, this diversification measurement often overstates the level of diversification (Van Horne et al., 1975). DIV1 gives the same score to a portfolio with two assets with 90 per cent in one and 10 per cent in the other, and a portfolio with two assets equally weighted. Therefore, I additionally use a second measure, again from Van Horne et al. (1975). It determines how closely an individual portfolio approximate the market portfolio. Moreover, it controls for portfolio weights.

$$DIV2 = \sum_{i=1}^{n}(w_i - w_m)^2 = \sum_{i=1}^{n}(w_i - \frac{1}{N_m})^2 \tag{4}$$

where n is the number of securities held by the investor. This is not self-reported as in DIV1, but reflects the assets that are held in the portfolio of a specific platform[4]. For this measurement, I only focus on this platform data so I do not need information about the total portfolio size of the consumer. I assume that the consumer diversifies similarly on other platforms. N_m is the number of assets in the market portfolio and w_i is the portfolio weight assigned to each asset i in the investor portfolio, and w_m is the weight assigned to an asset in the market portfolio (w_m = $1/N_m$). A lower value of *DIV2* reflects a higher level of diversification.

Van Horne et al. (1975) assume $w_m \approx 0$ since the proportion of each security in the market portfolio is very small. They calculate diversification as the sum of the squares of the proportions invested in each security. This seems plausible, but, the market portfolio for stocks and the market portfolio for crowdfunding loans are totally different. Since the data are obtained per platform, the market portfolio contains all the available projects within one platform. It follows that a potential investor could only diversify among the projects offered by this platform. This could make the variable more dependent on the platform's offering than on in the investors' diversification. For example, if an intermediary offers 5 projects on his platform, and an investor invest in 2 projects with equally weights, he gets a score of 0.18 ($2 * (0.5 - 0.2)^2$). Now, suppose that the investor could choose among 10 projects instead of 5, his diversification score will increase to 0.36 ($2 * (0.5 - 0.1)^2$), which means a lower level of diversification. If portfolios are not equally weighted, but 60-40, diversification is suboptimal. This will indeed lead to a higher diversification score. ($((0.6 - 0.2)^2) + (0.4 - 0.2)^2 > 2 * (0.5 - 0.2)^2$). A sophisticated investor could optimally diversify his investment equally into all 5 projects, and thereby gets the most efficient diversification score of 0 ($5 * (0.2 - 0.2)^2$). In this case, w_m is equal to w_i.

In addition to the described measurements of Van Horne et al. (1975), this paper will extend the diversification measurement so that it controls for differences in raised capital among projects in the market portfolio. It puts different weights to the projects in the portfolio.

$$DIV3 = \sum_{i=1}^{n}(w_i - w_{proj})^2 = \sum_{i=1}^{n}(w_i - \frac{X_{proj}}{X_{tot}})^2 \tag{5}$$

where X_{proj} is the total funding raised by the project and X_{tot} is the total funding raised by the platform. This formula allows us to reflect the diversification on the market portfolio, large funded projects get more weight than small funded project. Assume that a platform offers three projects with different total fundings: X with 10 000, Y with 30 000, and Z with 50 000. Investor A holds 10 in project X and 40 in project Z. Investor B holds 25 in X and 25 in Y. In this situation, investor A better reflects the market portfolio relative to B. With *DIV2*, investor A

[4] I specifically mean the variable assets according to the platform data, see also the variable description in appendix D

gets a score of 0.24 $((0.2-0.33)^2+(0.8-0.33)^2)$. With *DIV3*, investor A gets a significantly lower score of 0.068. $((0.2-(10/90))^2+(0.8-(50/90))^2)$. Now, suppose that investor C has the exact opposite portfolio of investor A, with 40 in project X and 10 in project Z. Hence, investor C's diversification level is worse due to an overexposure in project X relative to Z. Both investors get the same score according to *DIV2*. However, according to *DIV3*, investor C gets a score of 0.60 $(0.8-(10/90))^2+(0.2-(50/90))^2$, which is worse than the *DIV2* measure and the *DIV3* measure of investor A. Hence, *DIV3* indicates that investor A better reflects the market portfolio.

In table 2, I show that the variable (self-reported) assets is very volatile and strongly positively skewed. Hence, the first diversification measurement is divided in quintiles. The cut-off points are based on equally distribution of the observations. I classify investors with only one asset in their portfolio as very low diversified. The group with two to five assets are low diversified. Since the median is 10 projects, I classify these investors as average diversified. In household finance, it is generally accepted to take 20 assets as a well-diversified portfolio (Statman, 1987; Evans & Archer, 1968). I classify the group with at least 20 assets as highly diversified. More than 46 projects is chosen for very highly diversified such that it equally weight the highest two groups. It is difficult to interpret the absolute values of DIV2 and DIV3. These variables show again strongly positively skewness. When I rank these values into quintiles, the minimum and maximum of assets reported by the platform is not constantly increasing among the quintiles. For example, although a person has only one asset in the portfolio of a small platform that does not offer many projects, according to DIV3 it can be defined as low diversified. Contrary, a person that has 11 assets, one above the median, in its portfolio can still be classified as very low diversified. In general, I show that DIV2 does not only classify asset diversification based on the number of assets in the portfolio but takes into account the single transactions relative to the market portfolio. The market portfolio in this case is the total projects that are available on the platform. DIV3 takes another step further, and put weight on the single transaction relative to the total funding of the project. In panel 2B, the correlations of the different diversification measurements are reported. Although the three measures are related, they capture sufficient differences. Additionally to the correlations between DIV1, 2, and 3, I see that there is variation between self-reported assets and the assets according to the platform data. This is mainly due to people who participate on more than one platform. Another explanation is that people do not precisely know their portfolio size. The correlation between self-reported assets and assets on the platform increases from 0.45 to 0.94 when we only look at people participating on one platform. Out of 494 people that participate on only one platform, 285 did not report the same amount of assets as they actually have according to the platform data. Panel 2B shows also the correlation across DIV1, DIV2 and DIV3. The extra control for weighted investment amounts relative to the total funding seems not to contain a lot of variation, i.e. DIV2 and DIV3 are highly correlated. A possible interpretation is that there is not much variation in the investment amounts of the investor and in the total funding of the projects. Because DIV2 and DIV3 shows the same pattern, the next sections of this paper will focus on the DIV1 and the DIV3 measurements. DIV3 is preferred to DIV2 because it better reflects the market portfolio.

3.3.3 Other explanatory variables

The crowdfunding market is often seen as one industry with homogeneous participants. However, by analyzing the transaction data and the different motivations and expectations, I conclude that there is a lot of variety in investor characteristics. In financial literature, the distinction is often made between a professional investor and a retail investor. However, based on goals, motivation and expectations, and also based on different supervision policy, I distinguish three

Table 2

Determination of diversification. Panel A reports the summary statistics of DIV1, DIV2 and DIV3 and their quintiles. The cut-off points are based on equal distribution. The inverse of div2 and div3 are reported, so that the lowest scores correspond to the lowest diversification class. Panel B reports the correlation among the different diversification classes and the real assets in the portfolios. For all variables, spearman's correlation is reported except the pearson correlation between self-reported assets and platform assets. The data is from the AFM crowdfunding survey, collected in March 2016.

Panel A: Summary statistics and diversification quintiles based on financial literacy

	x̄	σ	p(25)	p(50)	p(75)	min	max	N
Assets (self-reported)	30.5	55.9	2	10	40	1	700	931

DIV1	min (assets$_{self-reported}$)	max (assets$_{self-reported}$)	mean (assets $_{platform}$)	N
Very low diversification	1	1	1	162
Low diversification	2	5	3.2	204
Avg diversification	6	19	10.8	184
High diversification	20	45	31.4	192
Very high diversification	46	700	103.5	189

	x̄	σ	p(25)	p(50)	p(75)	min	max	N
Assets (platform)	16.3	20.3	1	6	26	1	112	863
1/div2	17.1	24.8	1.1	6.0	28.8	1.0	234.7	863

DIV2	min (1/div2)	max (1/div2)	min (assets $_{platform}$)	max (assets $_{platform}$)	mean (assets $_{platform}$)	N
Very low diversification	1.0	1.1	1	1	1	224
Low diversification	1.2	3.2	2	11	2.4	123
Avg diversification	3.3	11.6	3	48	7.3	172
High diversification	11.7	30.4	11	53	22.3	173
Very high diversification	30.6	234.7	24	112	49.0	172

	x̄	σ	p(25)	p(50)	p(75)	min	max	N
Assets (platform)	16.3	20.3	1	6	26	1	112	863
1/div3	18.7	26.3	1.3	6.5	28.4	1.0	202.6	863

DIV3	min (1/div2)	max (1/div3)	min (assets $_{platform}$)	max (assets $_{platform}$)	mean (assets $_{platform}$)	N
Very low diversification	1.0	1.1	1	1	1	176
Low diversification	1.2	4.1	1	11	2.1	170
Avg diversification	4.1	12.2	2	48	7.2	172
High diversification	12.4	34.0	4	57	22.6	173
Very high diversification	34.2	202.6	23	112	48.7	172

Panel B: Correlation of diversification measurement and assets (self-reported and platformdata)

	DIV1	DIV2	DIV3	assets $_{self-reported}$	assets $_{platform}$
DIV1	1				
DIV2	0.77*	1			
DIV3	0.75*	0.98*	1		
assets $_{self-reported}$	0.98*	0.78*	0.76*	1	
assets $_{platform}$	0.79*	0.98*	0.96*	0.45*	1

Note:* p <0.5.

types of crowdfunding investors 1)feel-good funders, 2)return-seeking investors, and 3)high-stake investors. The sample shows a significant group of investors who use their crowdfunding accounts for "playing", shopping for entertaining projects or supporting someone in their own network. This investor segment has no expectations of a high return, getting the initial investment back is not priority. In this case, their real investments, including retirement might be held in other asset-classes. Another group investors use crowdfunding as a different asset-class in their portfolio. They do care if investments are not profitable, they are called return-seeking investors. The last group investors are called high-stake investors. They are the most sophisticated crowdfunders with portfolios larger than 100 000 euro. They invest from a 'business' perspective. Determination of the investor type is based on self-reported motivation, goals, and expectations. The 100 000+ euro is chosen because regulators see this investor as sophisticated, supervision for this group is limited. The determination of the three investor types are extensively described in appendix B. Table 3 reports the most important summary statistics of the three types of crowdfunding investors. My interpretation of the 'feel-good funder' corresponds to an investor that is highly under diversified. This group has median of 1 project in the portfolio. However, even when I group individuals on several characteristics, the diversification variables are strongly rightly skewed. Therefore, I use DIV1 and DIV3 as measurements of diversification in the next analyses.

Table 3

This table reports the summary table of the distribution of the three different types of investors in crowdfunding (demographics & behavior). More specifically, it estimates the mean, median and standard deviation of the continuous variables age, self-reported assets, platform specific assets, self-reported invested capital and platform specific invested capital. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016)

	Crowdfunding investor characteristics								
	Feel-good funder (N=259)			Return-seeking investor (N=641)			High-stake investor (N=31)		
Variable	x̄	Md	σ	x̄	Md	σ	x̄	Md	σ
Age	52.7	53	13.5	49.7	50	13.0	54.1	54	9.7
Assets (self-reported)	13.3	3	29.5	31.3	15	43.9	156.4	100	170.8
Assets (platformdata)	7.8	1	14.0	18.4	9.5	20.9	36.6	34	22.7
Invested cap (self-reported) (thousand of EUR)	7.2	1.2	14.6	19.4	11	21.9	186.1	130	132.3
Invested cap (platformdata) (thousand of EUR)	4.7	1	10.2	12.5	5.7	16.1	52.5	46.3	38.0

People are asked to report their net worth and annual gross income , see appendix A.2. I find that a high proportion of households report a net worth or income that is respectively in the 150 000+ euros and 80 000+ euros group. There are two possible interpretations for this surprising result. First, crowdfunding is done by wealthy investors that use crowdfunding as an extra asset-class. People could prefer to invest primarily in other asset-classes, and only add crowdfunding to their portfolio after they have reached a specific net worth. Second, the AFM recently recommends to invest not more than 10 per cent of their financial wealth in crowdfunding, I call this portfolio size to net worth ratio (SNWR). It could be that households report more wealth and income than they actually have because they fear that the AFM could signal a high SNWR as a threat. Consequently, the unjustified fear that people have, could lead to biased results. There is another obstacle in measuring the SNWR. The assumption should be made that the absolute value of the net worth range will be exactly the mean of the two values. This means that someone who is in the range 10 000 - 25 000 euros net worth will get the absolute value of 17 500 euro. This seems plausible, but it is hard to estimate the absolute value for someone who is in the 150 000+ euros net worth group. The assumption is made that people who are in the 10 000- euros will get 8 000 euros as absolute amount. People that have 150 000+ will be approximated as 180 000 euros. This is based on the growth in differences of absolute financial wealth, which means 8 000, 17 500, 37 500, 65 000, 115 000 and 180 000 euros. Note that 180 000 euros is a arbitrary value and limits the absolute net worth of the top wealthy investors. In table 4, I report the SNWR for all participants relative to their net worth. Based on limiting the net worth of the 150 000+ euros group, I expect that this group is exposed to the $50+\%$ SNWR.

Table 4

Distribution of the portfolio size to net worth ratio (SNWR) and net worth (EUR) (weighted percentages, rows & columns equal to 100). The Pearson chi^2 statistic test the null hypothesis that the distribution of SNWR is independent of net worth (p-values are reported in parentheses) Both invested capital and net worth are self-reported. The data is from the AFM crowdfunding survey, collected in March 2016.

	Net worth						
SNWR	<10 000	10 000-25 000	25 000-50 000	50 000-80 000	80 000-150 000	>150 000	N
≤ 5%	1.3	3.1	4.7	4.6	4.7	10.6	271
5% - 10%	1.0	2.3	2.4	1.9	2.4	6.7	154
10% - 20%	1.0	1.0	2.0	1.9	2.7	6.4	140
20% - 50%	0.9	1.5	3.2	2.8	3.3	9.3	196
>50%	0.3	1.1	1.2	1.1	0.8	3.1	70
Refusal		(N=100)					
N	41	83	126	115	129	337	
	Pearson chi2(30)= 947.08			(p=0.000)			

The wealthiest investors are not the only group exposed to the highest SNWR quintile group. Wealth and SNWR is not strongly correlated, so based on table 4, limiting the 150 000+ group does not directly lead to an upward bias in SNWR, although I cannot fully exclude this effect.

The survey also contain three questions about risk preferences. Individuals are grouped in tertiles based on their risk tolerance, 1) risk averse investors, 2) risk neutral investors, and 3) risk seeking investors.[5]

3.4 Sample robustness check: Selection bias

3.4.1 Selection bias

This research contains two main data sets. First, there is the survey respondents data. The 5 participating platforms together have a total market capitalization of around 71 per cent. I include platform U that focuses on a niche with a slightly different business model. I have 931 respondents that gave permission to use their data. Although this sample construction seems robust, there might be some selection bias. Well-performing investors could have a higher responds rate than low-performing investors. Low-performing investors are consumers that might have several assets in default. In other words, the sample could have an upward bias. I asked platforms to provide three averages of all their investors and relate this to the sample averages. First, I analyse the average transaction amount in euros per project of all investors with respect to the sample. Second, I compare the average number of assets in the portfolio. Third, I contrast the average total capital invested in the portfolio. I use a two-tailed test-statistic with an unknown population distribution according to the following formula:

$$t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{s} \tag{6}$$

[5]Details about the determination of risk tolerance levels can be found in the appendix A.3.

If the test statistic is significant, the mean of the sub sample is significantly different from the average of the platform data. If the test statistic is not significant I conclude that the sample mean is representative for the platform mean. Because these averages are easy to trace back to a specific platform, the outcomes are confidential. I report table E1 in a separate appendix aimed at internal use of the AFM. In table 5 I report weighted averages for all platforms together.

Table 5

This table reports the robustness check to test if the household that participated in the survey are representative for the whole population of investors on the particular platform. Panel A compares the average amount invested per project. Panel B analyses the average portfolio size of the platform with respect to survey participants. Panel C compares the average size of the investment portfolio per investor. T-stat tests the null hypothesis that the survey sample mean is not different from the platform mean. I use weighted averages, so I control for the difference in observations across platforms. Rho indicates the significance level of the t-statistic. The data is from the AFM crowdfunding survey, collected in March 2016 and data about averages specifically requested from platforms

Panel A: avg transaction amount per project N = 14026					
	$\bar{x}$ (platform)	$\bar{x}$ (survey sample)	σ (survey sample)	t-stat	ρ
Weighted average	830	730	711.7	−16.62	0
Panel B: avg assets in portfolio N = 863					
	$\bar{x}$ (platform)	$\bar{x}$ (survey sample)	σ (survey sample)	t-stat	ρ
Weighted average	8.68	16.30	20.3	11.01	0
Panel C: Investmentportfolio per investor N = 863					
	$\bar{x}$ (platform)	$\bar{x}$ (survey sample)	σ (survey sample)	t-stat	ρ
Weighted average	8127	11911	18209.0	6.10	0

To summarize the results, there is a lot of variety across platforms. Platform dummies should be included when looking at diversification. Respondents of three out of five platforms showing similar results as all investors of the platform with respect to the average transaction amount per project. The other two platforms show downward biased, i.e. respondents show lower invested capital per project than investors in general. With respect to portfolio size in panel B, two out of five platforms show similar results. However, due to the other platforms, there is a strong upward bias in portfolio size of respondents relative to all investors. The average portfolio size is almost two times larger for respondents with respect to all investors. In panel C, there is an upward bias in total capital invested per investor. There are two platforms that cause this upward bias and one platform that shows exact opposite results, decreasing the investment portfolio size. In general, the survey shows an upward-bias trend of investors that are more active in investing than the average investor of the platform. Again, platform dummies should be included to control for the variety across platforms.

4 Financial literacy of crowdfunders

4.1 Descriptive statistics: Financial literacy across demographics

Table 6 reports the distribution of financial literacy scores across demographic variables. More specifically, it estimates the single cross-section of gender, as categorical variable, and education, age, income and financial wealth, as ordinal variables on financial literacy. Based on significant Chi^2 tests in the single cross section of education, age, gender, income and wealth, respectively, on financial literacy, I accept hypothesis 1 of heterogeneity in literacy. This research confirms the results of Van Rooij et al. (2011) and Calvet et al. (2009) that financial literacy increases with education. 80 per cent of the people in the lowest educational category have low financial literacy while that is only 28 per cent in the highest educational category. The literacy mean

Table 6

Summary statistics of the investor based on financial literacy. It reports the distribution of financial literacy scores across demographic variables. More specifically, it estimates the single cross-section of gender, as categorical variable, and education, age, income and financial wealth, as ordinal variables on financial literacy. It reports weighted percentages of the proportion of households in each financial literacy tertiles (total row tertiles sum op to 100). The Pearson chi^2 statistic tests the null hypothesis that the distribution of households over the three literacy tertiles is independent of respectively education, age, gender, income, and wealth (p-values are reported in parentheses). The data is from the AFM crowdfunding survey, collected in March 2016.

	Literacy tertiles				
Education	1 (low)	2 (avg)	3 (high)	Mean	*N*
Lager (beroeps)onderwijs	30.0	60.0	10.0	1.8	10
Mavo/Mulo/Vmbo	20.0	20.0	60.0	2.4	30
HAVO/VWO/MBO	25.1	34.8	40.0	2.15	155
HBO	17.3	32.5	50.2	2.33	329
WO	10.6	32.0	52.1	2.48	407
	Pearson chi2(8)= 33.85			(p=0.000)	
	Literacy tertiles				
Age	1 (low)	2 (avg)	3 (high)	Mean	*N*
18-30	23.6	36.1	40.3	2.17	72
31-40	15.2	34.4	50.3	2.35	151
41-50	12.6	30.6	56.8	2.44	222
51-60	12.3	31.4	56.4	2.44	252
61-70	22.0	32.3	46.7	2.25	182
>70	17.3	30.8	51.9	2.35	52
	Pearson chi2(10)= 16.30			(p=0.091)	
	Literacy tertiles				
Gender	1 (low)	2 (avg)	3 (high)	Mean	*N*
Male	11.7	32.3	56.0	2.44	777
Female	37.0	30.5	32.5	1.95	154
	Pearson chi2(2)= 65.53			(p=0.000)	
	Literacy tertiles				
Income	1 (low)	2 (avg)	3 (high)	Mean	*N*
<12 500	26.7	40.0	33.3	2.07	15
12 500 - 33 000	34.9	31.8	33.3	1.98	66
33 000 - 40 000	22.4	38.3	39.3	2.17	107
40 000 - 66 000	14.8	34.3	50.9	2.36	216
66 000 - 80 000	16.1	31.6	52.3	2.36	155
>80 000	8.4	29.6	62.1	2.54	311
Refusal	23.0	24.6	52.5	2.30	61
	Pearson chi2(12)= 50.51			(p=0.000)	
	Literacy tertiles				
Financial Wealth	1 (low)	2 (avg)	3 (high)	Mean	*N*
<10 000	31.7	41.5	26.8	1.95	41
10 000 - 25 000	33.7	27.7	38.6	2.05	83
25 000 - 50 000	19.1	37.3	43.7	2.25	126
50 000 - 80 000	15.7	36.5	47.8	2.32	115
80 000 - 150 000	12.4	36.4	51.2	2.39	129
>150 000	6.5	27.9	65.6	2.59	337
Refusal	27.0	28.0	45.0	2.18	100
	Pearson chi2(12)= 80.56			(p=0.000)	

strongly increases with educational categories, except for the MAVO[6] category. However, the observations are limited for the first two categories. In the regressions I merge the first three educational categories to remain statistical power. Again, in line with Van Rooij et al. (2011), the literacy mean increases till the age of 50 and than decreases slightly. This relation is not very pronounced. A more striking result is the low literacy scores of females relative to males. This is again in line with Van Rooij et al. (2011) and Lusardi and Mitchell (2007). Besides the education, age and gender characteristics, this research is extended by looking at income and wealth distribution. Both variables are negatively skewed, i.e. most observations are in the highest category of self-reported income and self-reported wealth. Especially in the wealth variable, a strong positive relation exists between literacy and wealth. This corresponds to the result of Calvet et al. (2009), who also find a positive relation between financial wealth and financial literacy.

Note that the relations described above are in a single cross-section. Hence, I can not reject the possibility that omitted variables explain the results. Nevertheless, these results are in line with previous research in financial literacy and confirm the validity of my financial literacy measurement.

4.2 Descriptive statistics: Financial literacy and investment behavior

Table 7 describes the single cross-section between financial literacy and investment behavior. I use *DIV1* as ordinal variable. Furthermore it is interesting to look at the investment capital and the portfolio size to net worth ratio (SNWR). The AFM recommends to invest a maximum of 10 per cent of total wealth in crowdfunding projects.[7]

Diversification and financial literacy seem to have a strong positive relationship. The financial literacy mean is constantly increasing among DIV1. More than half of the people that have very small portfolios, i.e. only one asset, are in the lowest financial literacy class. On average, investors that diversify the most, have above average literacy. Because of the high correlation between diversification and invested capital, there exist similar results for the relationship between invested capital and literacy.[8] The single cross-sectional relationship does not exclude the fact that higher financial literacy itself leads to more diversification. In the next section, the effect of financial literacy on diversification will be shown in a multiple regression framework.

Although one would expect that sophisticated people would invest a smaller amount of their net worth (SNWR) in risky crowdfunding transactions, this relation seems not so clear. The AFM recommends to invest not more than 10 per cent of net worth in crowdfunding. Apparently, low literacy is more pronounced in the <10 % SNWR group relative to the >10 % SNWR. It is surprising that similar results exist in the SNWR/literacy distribution as in the distribution of net worth to literacy and the invested capital/literacy distribution. Interpreting these results is difficult and the results should be taken with caution, because of the biases I described in the SNWR determination.[9]

[6] I report educational categories according to the Dutch education system. "Lager (beroeps) onderwijs" = primary education, "Mavo/Mulo/Vmbo" = preparatory intermediate vocational education, "Havo/Vwo/Mbo" = Secundary pre-university or intermediate vocational, "Hbo" = higher vocational, "WO" = university.

[7] I discuss the AFM recommendation of investing no more than 10 per cent SNWR in the discussion 6.4.

[8] More summary statistics of financial literacy among the continuous variables age, total assets and invested capital are reported in appendix A.4.

[9] See also section 3.3.3.

Table 7

This table reports the distribution of financial literacy scores across investment behavior variables. More specifically, it estimates the single cross-section of DIV1, invested capital, and the ratio invested capital/net worth, as ordinal variables on financial literacy. It reports weighted percentages of the proportion of households in each financial literacy tertiles (total row tertiles sum op to 100). The Pearson chi^2 statistic test the null hypothesis that the distribution of households over the three literacy tertiles is independent of respectively DIV1, invested capital, and the portfolio size to net worth ratio (p-values are reported in parentheses). The data is from the AFM crowdfunding survey, collected in March 2016.

	Literacy tertiles				
DIV1 (based on self-reported assets)	1 (low)	2 (avg)	3 (high)	Mean	N
1 (very low)	32.7	29.6	37.6	2.05	162
2 (low)	20.1	35.8	44.1	2.24	204
3 (avg)	15.8	34.2	50.0	2.34	184
4 (high)	6.8	33.9	59.4	2.53	192
5 (very high)	6.4	25.9	67.7	2.61	189
	Pearson chi2(8)= 75.81			(p=0.000)	
	Literacy tertiles				
Invested capital in EUR (self-reported)	1 (low)	2 (avg)	3 (high)	Mean	N
0 - 1,000	25.4	36.1	38.5	2.13	205
1,001 - 5000	21.1	30.6	48.3	2.27	209
5,001 - 10,000	19.1	28.7	52.2	2.33	115
10,001 - 25,000	9.9	37.4	52.8	2.43	182
$>$ 25,000	5.5	26.8	67.7	2.62	220
	Pearson chi2(8)= 58.25			(p=0.000)	
	Literacy tertiles				
SNWR	1 (low)	2 (avg)	3 (high)	Mean	N
$\leq$ 5%	19.6	32.8	47.6	2.28	271
5% - 10%	14.9	32.5	52.6	2.38	154
10% - 20%	11.4	32.9	55.7	2.44	140
20% - 50%	13.8	29.1	57.1	2.43	196
$>$50%	2.9	40.0	57.1	2.54	70
Refusal	27.0	28.0	45.0	2.18	100
	Pearson chi2(10)= 26.13			(p=0.004)	

4.3 The influence of past literacy tests and learning by doing

The questions that I selected to measure financial literacy are not random but are chosen from a larger financial literacy test of Van Rooij et al. (2011). They highlight that their financial literacy proxy is influenced by a group that have "interest in economics." This particular group has no difficulty with answering questions because these are related to their economic hobbies, such as online trading. It follows that "interest in economics" is an endogenous factor of the effect of eduction on literacy. In my questionnaire, I do not ask about hobbies. However, to measure whether fanatic investors better score on literacy, I measure "interest in trading" with: "Does the platform ever refused your investment because your portfolio size was above the limit?". These investors are real obsessed investors that spend much time on choosing projects and creating their portfolios. I want to test the following hypotheses:

Hypothesis 5: Obsessed investors score significantly higher on financial literacy than normal investors.

Another potential endogenous factor in measuring financial literacy I shed light on is the "past literacy" bias. People could already have answered comparable questions in previous tests, for example a mandatory investments test when entering an online trading platform. I measure "past literacy" with the dummy for whether or not the participant is also an execution-only investor[10]. I test the following hypothesis:

Hypothesis 6: Execution-only investors score significantly higher on financial literacy than other crowdfunding investors, who do not have experience in execution-only investing.

In table 8, I estimate the influence of demographics and the three factors "Interest in trading", "financial advice", and "past literacy". Because literacy is an ordinal variable with the values low, average, and high, I use the following ordered logistic regression model;

$$\begin{aligned}\text{Fin\_literacy}_i = {} & \alpha + \beta_1 \text{Age}_i + \beta_2 \text{Education}_i + \beta_3 \text{Gender}_i + \beta_4 \text{Retired}_i \\ & + \beta_5 \text{Self\_employed}_i + \beta_6 \text{Income}_i + \beta_7 \text{Wealth}_i + \epsilon_i\end{aligned}$$

I extend the model by looking at the effect of 'past literacy', 'financial advice' and 'learning by doing' on financial literacy. I use the previous dependent variables as controls.

$$\begin{aligned}\text{Fin\_literacy}_i = {} & \alpha + \beta_1 \text{Incorrect\_capital\_estimation}_i + \beta_2 \text{Past\_literacy}_i \\ & + \beta_3 \text{Advice}_i + \beta_4 \text{Interest\_in\_trading}_i + \beta_5 \text{Controls}_i + \epsilon_i\end{aligned}$$

where financial literacy is an ordinal variable with three values. I use an ordinal logistic regression with proportional odds. This is independent of the category of financial literacy, i.e. the odds ratios are assumed to be constant for all categories.[11] I test for proportionality of odds and show that restricting the ordinal model to one coefficient is valid. Moreover, the Brant test, which tests the parallel regression assumption, is not violated for most variables. Only the "interest in trading" and "gender" variable are not performing very well.[12].

[10] An execution-only investor is an investor that manages its portfolio of assets without relying on financial advice.

[11] Details about the ordinal logistic framework are described in section A.7.

[12] See A.5 for detailed statistics of the Brant test

Table 8

Ordered logistic regression of the effect of demographics, "interest in trading", "advice", and "past literacy" on the financial literacy tertiles. The coefficients are reported as proportional odds ratios. The base regression reports the effect of age, education, dummies for gender, retired, and self-employed, and income and wealth on the tertiles of finacial literacy. The second regression adds the influence of "past literacy" on financial literacy. The third regression adds the "past literacy effect" and "interest in trading" on the financial literacy. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test test the null hypothesis that the parallel regression assumption is valid. The data is from AFM's crowdfunding survey, collected in March 2016.

Variables	(1) OLR (base)		(2) OLR (2)		(3) OLR (3)	
Age dummies (Base group: age $\leq$ 30)						
30 < age $\leq$ 40	1.247	(0.774)	1.255	(0.742)	1.261	(0.758)
40 < age $\leq$ 50	1.461	(1.329)	1.199	(0.591)	1.209	(0.618)
50 < age $\leq$ 60	1.521	(1.478)	1.390	(1.087)	1.391	(1.088)
60 < age $\leq$ 70	0.802	(−0.665)	0.773	(−0.723)	0.794	(−0.647)
age >70	0.954	(−0.104)	0.888	(−0.247)	0.928	(−0.153)
Education dummies(Base group: lager onderwijs)						
Mavo/Mulo/Vmbo	5.773^{**}	(2.574)	4.571^{**}	(2.207)	4.692^{**}	(2.241)
Havo/Vwo/Mbo	2.769^{*}	(1.738)	2.474	(1.514)	2.518	(1.538)
HBO	4.088^{**}	(2.427)	3.985^{**}	(2.342)	4.015^{**}	(2.348)
WO	5.924^{***}	(3.054)	4.828^{***}	(2.655)	4.920^{***}	(2.677)
Female	0.319^{***}	(−6.345)	0.344^{***}	(−5.542)	0.344^{***}	(−5.545)
Retired	1.153	(0.506)	1.090	(0.293)	1.063	(0.207)
Self-employed	0.912	(−0.544)	1.048	(0.263)	1.043	(0.235)
Household income dummies (Base group: income < 12 500)						
12 500 $\leq$ income < 33 000	0.506	(−1.232)	0.697	(−0.621)	0.678	(−0.667)
33 000 $\leq$ income < 40 000	0.627	(−0.869)	0.721	(−0.584)	0.717	(−0.595)
40 000 $\leq$ income < 66 000	0.737	(−0.577)	0.907	(−0.177)	0.903	(−0.184)
66 000 $\leq$ income < 80 000	0.671	(−0.740)	0.909	(−0.169)	0.905	(−0.176)
income $\geq$ 80000	0.862	(−0.277)	1.310	(0.483)	1.312	(0.484)
Refusal	0.949	(−0.0879)	1.489	(0.630)	1.482	(0.622)
Financial wealth dummies (Base group: wealth < 10 000)						
10 000 $\leq$ wealth < 25 000	1.405	(0.924)	1.692	(1.277)	1.688	(1.270)
25 000 $\leq$ wealth < 50 000	1.937^{*}	(1.910)	1.462	(0.991)	1.460	(0.988)
50 000 $\leq$ wealth < 80 000	2.218^{**}	(2.258)	1.753	(1.452)	1.734	(1.423)
80 000 $\leq$ wealth < 150 000	2.430^{**}	(2.535)	1.711	(1.393)	1.645	(1.285)
wealth $\geq$ 150000	4.286^{***}	(4.345)	3.329^{***}	(3.237)	3.162^{***}	(3.079)
Refusal	1.496	(1.031)	1.147	(0.314)	1.127	(0.273)
Incorrect capital estimation			0.709^{*}	(−1.947)	0.723^{*}	(−1.832)
Past literacy (Base group: No)						
Yes			2.578^{***}	(5.802)	2.583^{***}	(5.809)
Advice (Base group: No)						
Yes			1.951^{***}	(2.784)	1.927^{***}	(2.725)
Interest in trading (Base group: No)						
Yes					1.404	(1.233)
Constant cut1	1.302	(0.353)	1.916	(0.827)	1.947	(0.846)
Constant cut2	7.966^{***}	(2.763)	12.95^{***}	(3.232)	13.19^{***}	(3.250)
Observations		931		863		863
chi2		153.47		178.64		180.20
ρ		0.000		0.000		0.000
ρ-value test proportionality of odds		0.285		0.533		0.251
ρ-value Brant test		0.180		0.351		0.263
Pseudo R2		0.0827		0.105		0.106

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

All Chi^2 tests indicate that the model as a whole is statistically significant. As shown in the single-cross sectional relationships in section 4.1, education and gender mainly explain the financial literacy score, which partly confirms hypothesis 2. The odds of being in a higher literacy group are continuously increasing for all education groups except the "Mavo/Mulo/Vmbo" group. This group scores better than expected. However, the sample size of the base group and the "Mavo/Mulo/Vmbo" group is low (<30.). The highest three educational categories do perform much better, and I confirm previous studies that financial literacy increases among education. Additionally, I check whether education interacts with gender, but the results are not significant. I partly confirm hypothesis 2, showing that literacy increases with wealth. Especially the literacy is higher when being in the highest wealth category relative to the lowest wealth category.
The effects are sizable. When looking at demographic characteristics (OLR(1)), having a university degree compared to having almost no education increases the odds with almost 6 of being in at least middle literacy versus low literacy, or in high literacy versus middle or low literacy. The second largest effect is wealth. Having wealth in the highest quintile with respect to the lowest increases the odds with 4.2. The third largest effect is gender, being male increases the odds with 3.1. This is the inverse of the odds of being female (1/0.319).
I extend this research by adding investment behavior variables to the model (OLR(2)). Hypothesis 6, which says that past literacy tests influences the financial literacy score is confirmed. The odds of high literacy versus the combined middle and low literacy increases with 2.58, given that all of the other variables in the model are held constant. When people report they have access to a financial adviser, they score significantly better on financial literacy. In OLR(3), I add the interest in trading variable, but this shows no significant results. I reject hypothesis 5 and conclude that experience in crowdfunding trading does not lead to higher financial literacy. A possible interpretation could be that the financial literacy questions are more relevant to investing in general, and not in crowdfunding specificly. To test this interpretation, I interact wealth with the investment as percentage of total wealth. Again, I do not find that financial literacy is affected by how much people invest. In my analysis, I check whether there is variety in financial literacy across platforms, but I do not find significantly diffrent results when adding platform dummies. In table 9, I show predicted probabilities from the OLR (2) model. By holding all controls at mean value, the table shows the predicted probabilities of being in high, middle, or low literacy by looking at a change in gender, education, past literacy and financial advice.

5 The Extent of diversification in individual portfolios

Most crowdfunding projects are in the early-stage of funding, causing substantial risks with a high idiosyncratic component, which could be easily reduced by diversification. The classical frightened image of consumer protectors is that non-professional investors put all there savings in one promising project and will lose everything in case of a default. According to the descriptives of table 2, an average investor is involved in 16 projects per platform and has 30 projects in its total portfolio. The portfolio of the majority of the crowdfunders seems sufficiently diversified, but a substantial part shows worse performance. It is considerably relevant for the regulatory agency to shed light on the significant factors that causes under diversification. This section focuses on the relationship between financial literacy and the extent of diversification. Calvet et al. (2006) conclude that more sophisticated agents tend to invest more efficiently and make smaller mistakes. Van Rooij et al. (2011) show that investors with higher literacy are more likely to invest in stocks. Based on these results, I expect a strongly positive relation between financial literacy and asset diversification, which I test following hypothesis 3.

Table 9

This table reports the predicted probabilities of the dependent ordinal variable financial literacy, using model OLR(2) from table 8. By holding all controls at mean value, except for gender, education, past literacy and advice respectively, I show the effects on financial literacy, ceterus paribus. The sum of the probabilities of low, average, and high literacy sum up to 1. The data is from the AFM crowdfunding survey, collected in March 2016.

	Literacy tertiles		
	1 (low)	2 (avg)	3 (high)
All at means	0.120	0.359	0.521
Female	0.251	0.443	0.307
Male	0.103	0.334	0.562
Lager (beroeps) onderwijs	0.124	0.361	0.515
Mavo/Mulo/Vmbo	0.112	0.344	0.544
HAVO/VWO/MBO	0.152	0.392	0.456
HBO	0.138	0.377	0.486
WO	0.101	0.326	0.573
Past literacy (Yes)	0.085	0.301	0.614
Past literacy (No)	0.194	0.425	0.381
Access to adviser (Yes)	0.071	0.266	0.664
Access to adviser (No)	0.132	0.370	0.499

5.1 Descriptive statistics: Diversification across demographics

Table 10 shows the summary statistics of investment behavior of the DIV1 quintiles. Although there are a few average diversified people with only 1.000 euros, a strong correlation exists between invested capital and DIV1. This corresponds to the high Spearman's correlation in table 2B. The three investor segments show very different behavior. Feel-good funders are much less diversified than return-seeking investors. High-stake investors, which are determined based on a minimum capital that is invested, are almost all well diversified, which again shows evidence of a strong relationship between invested capital and diversification. The single-cross section of risk tolerance on diversification is controversial. Risk averse persons are less diversified than risk seeking agents and there is no difference between risk seeking and risk neutral investors.[13] There seems a strong relationship between investors that are active in the stock market and diversification. Note that the relations described above are in a single cross-section. Hence, I can not reject the possibility that omitted variables explain part of the variation in diversification. Nevertheless, the results reflect a high correlation between invested capital and a very low correlation of risk tolerance and diversification.

5.2 The effect of financial literacy on portfolio diversification

Table 11 shows the ordered logistic regression of financial literacy and controls on diversification. The control variables are chosen based on previous literature on financial literacy and portfolio diversification[14] The regression model is,

[13]In section 6.5, I discuss risk tolerance in detail.

[14]See appendix A.7 for detailed information about the ordered logistic framework, and see the theoretical framework in section 2 why I choose specific controls in the model.

Table 10

Summary Statistics Investor based on diversification (weighted percentages) Summary statistics of the investor based on diversification.More specifically, it estimates the single cross-section of self-reported invested capital and risk tolerance, as ordinal variables, and investor type, retirement and participating in stock market, as categorical variables on the ordinal variable DIV1. It reports weighted percentages of the proportion of households in each DIV1 quintile(total row quintiles sum op to 100). The Pearson chi^2 statistic test the null hypothesis that the distribution of households over the diversification quintiles is independent of respectively self-reported invested capital, investor type, risk tolerance, retirement and participating in the stock market (p-values are reported in parentheses). The data is from the AFM crowdfunding survey, collected in March 2016

	DIV1 quintiles						
Invested capital in EUR (self-reported)	1 (very low)	2 (low)	3 (avg)	4 (high)	5 (very high)	Mean	*N*
0 - 1,000	63.9	30.7	4.4	0.5	0.5	1.43	205
1,001 - 5000	11.0	47.9	28.7	11.0	1.4	2.44	209
5,001 - 10,000	5.2	18.3	36.5	27.0	13.0	3.24	115
10,001 - 25,000	1.1	8.2	26.9	40.1	23.6	3.77	182
$>$ 25,000	0	2.3	10.9	29.1	57.7	4.42	220
	Pearson chi2(16)= 849.83			(p=0.000)			
	DIV1 quintiles						
Investor type	1 (very low)	2 (low)	3 (avg)	4 (high)	5 (very high)	Mean	*N*
Feel-good funder	34.0	32.8	15.1	10.0	8.1	2.25	259
Return-seeking investor	11.5	18.4	22.6	25.4	22.0	3.28	641
High-stake investor	0	3.2	0	9.7	87.1	4.81	31
	Pearson chi2(10)= 204.25			(p=0.000)			
	DIV1 quintiles						
Risk tolerance	1 (very low)	2 (low)	3 (avg)	4 (high)	5 (very high)	Mean	*N*
Risk averse	22.8	20.4	19.8	18.9	18.3	2.90	334
Risk neutral	16.0	21.1	20.5	19.2	23.5	3.13	313
Risk seeking	12.7	24.7	19.0	24.3	19.4	3.13	284
	Pearson chi2(8)= 15.99			(p=0.042)			
	DIV1 quintiles						
Retired	1 (very low)	2 (low)	3 (avg)	4 (high)	5 (very high)	Mean	*N*
Yes	18.8	25.3	21.8	18.2	15.9	2.86	170
No	17.1	21.2	19.3	21.2	21.3	3.1	761
	Pearson chi2(2)= 4.32			(p=0.364)			
	DIV1 quintiles						
Participating in stock market	1 (very low)	2 (low)	3 (avg)	4 (high)	5 (very high)	Mean	*N*
Yes	13.2	21.6	19.6	22.3	23.3	3.21	750
No	34.8	23.2	20.4	13.8	7.7	2.36	181
	Pearson chi2(4)= 61.87			(p=0.000)			

$$
\begin{aligned}
\text{Diversification}_i = \alpha &+ \beta_1 \text{Fin\_literacy}_i + \beta_2 \text{Invested\_capital}_i + \beta_3 \text{Age}_i + \beta_4 \text{Education}_i \\
&+ \beta_5 \text{Gender}_i + \beta_6 \text{Retired}_i + \beta_7 \text{Self\_employed}_i + \beta_8 \text{Income}_i + \beta_9 \text{Wealth}_i \\
&+ \beta_{10} \text{Participating\_in\_stockmarket}_i + \beta_{11} \text{Access\_to\_financial\_adviser}_i \\
&+ \beta_{12} \text{Investor\_segment}_i + \beta_{13} \text{Short\_investment\_period}_i \\
&+ \beta_{13} \text{Participate\_on\_one\_platform}_i + \epsilon_i
\end{aligned}
$$

where diversification can take the form of the ordinal variable DIV1 (table 11) or DIV3 (table A7). The control variable invested capital equals the self-reported capital for the DIV1 regression because DIV1 is derived from self-reported assets. For the DIV3 regression, I use invested capital on the particular platform, because DIV3 is derived from the platform assets. Although I already control for differences in investor segments, I additionally control for platform differences in the third regression of DIV1 and DIV3. The platform dummies make more sense in DIV1 than in DIV3, since the DIV3 measurement already controls for platform size.[15]

Almost all variables are positively correlated.[16] It is only logical that there is substantial correlation between diversification and invested capital (0.60) and age and retirement (0.64). I intend to control for the "interest in trading" variable and the "high-stake" investor but those variables show multicollineairity. "Interest in trading", which is defined as facing investment limits of 80.000 euros and being a "high-stake" investor, which is derived from the portfolio size of at least 100.000 euros, are not available in the lowest diversification quintiles. Similarly, this holds for the first and second educational categories, so I merge them together. In the ordered logistic regression framework, there should be observations in all the quinitles, otherwise, the proportional odds assumption does not hold. I cannot control for risk tolerance, since the Brant test shows very insignificant results when adding risk tolerance as control variable. Van Rooij et al. (2011) also face problems when controlling for risk tolerance.[17]

From table 11, I find evidence for hypothesis 3 and conclude that financial literacy is an important variable in explaining diversification. Even after controlling for demographics and investment behavior, I still find significant results. Invested capital best explains diversification. However, having average literacy compared to low literacy increases the odds of diversification with 2.4. This effect is sizable. For example, participating in the stock market affects diversification less than literacy. Comparing results of Goetzmann and Kumar (2008), Van Rooij et al. (2011), and Von Gaudecker (2015) I find contradicting results for education, those with a universal degree do under diversify relative to lower educational categories. Even when controlling for platform, which explains much variation in diversification, financial literacy is strongly significant. The F-test that average literacy is equal to high literacy is rejected. I find similar results when using DIV3 as dependent variable.[18]

5.2.1 Predicted probabilities of diversification per literacy group

When transforming the odds ratios into predicted probabilities, the effect of financial literacy on diversification is easier to interpret. Using the controls in OLR(3), I can predict the probabilites of being in a specific diversification class for different financial literacy classes, while holding control variables at their mean. In figure 1, I show that predicted probabilities of low diversification

[15] With platform size, I mean the total projects offered by the platform, which corresponds to the total supply of assets on the platform. See formula 5 for the measurement of DIV3.

[16] The correlation tables are reported in appendix A.6.

[17] In the limitation part in section 6.5, I discuss the risk tolerance variable.

[18] Despite I find strongly similar results for DIV3, I am not able to achieve stable regression results in the order logistic setting. I provide regression statistics and details in appendix A.8.

Table 11

Ordered logistic regression of the effect of financial literacy and controls on DIV1. For validity reasons, I merge the 4th and 5th quintile of DIV1. The base regression reports the effect of financial literacy on DIV1. OLR (2) adds controls. OLR (3) controls for differences in platform. The coefficients are reported as proportional odds ratios. Three cutpoints are made to divide the quartiles and allows one coeffficient for each variable. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test test the null hypothesis that the parallel regression assumption is valid. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

Variables	(1) DIV1 (base)		(2) DIV1 (OLR2)		(3) DIV1 (OLR3)	
Financial literacy dummies (Base group: low literacy)						
Average literacy	2.722***	(5.483)	2.385***	(3.944)	2.418***	(3.773)
High literacy	4.179***	(8.217)	2.687***	(4.559)	2.322***	(3.628)
Invested capital (self-reported)(Base group: invested capital $\leq$ 1000))						
1000 $<$ invested capital $\leq$ 10000			16.31***	(11.90)	9.211***	(8.869)
invested capital $>$ 10 000			185.6***	(18.31)	71.20***	(13.17)
Age dummies (Base group: age $\leq$ 40)						
40<age $\leq$ 60			0.647**	(−2.260)	0.742	(−1.474)
age $>$ 60			0.550**	(−2.053)	0.701	(−1.179)
Education dummies(Base group: $<$ HBO)						
HBO			0.711*	(−1.668)	0.853	(−0.733)
WO			0.543***	(−2.918)	0.791	(−1.060)
Female			0.752	(−1.434)	0.866	(−0.691)
Retired			0.683	(−1.322)	0.636	(−1.497)
Self-employed			0.804	(−1.187)	0.928	(−0.396)
Household income dummies (Base group: income $<$ 33 000)						
33 000 $\leq$ income $<$ 66 000			1.126	(0.441)	1.224	(0.709)
income $\geq$ 66000			0.930	(−0.261)	0.885	(−0.415)
Refusal			1.530	(0.972)	1.407	(0.758)
Financial wealth dummies (Base group: wealth $<$ 25 000)						
25 000 $\leq$ wealth $<$ 80 000			0.916	(−0.368)	0.747	(−1.160)
wealth $\geq$ 80000			0.979	(−0.0846)	0.996	(−0.0150)
Refusal			0.575	(−1.643)	0.675	(−1.128)
Participating in stock market			1.493**	(2.432)	1.702***	(3.024)
Access to financial adviser			0.838	(−0.719)	0.935	(−0.263)
Feel good investor			1.036	(0.215)	0.812	(−1.164)
Invest $<$ 0.5 years			0.477***	(−4.632)	0.531***	(−3.750)
Participate on only one platform			0.349***	(−6.966)	0.197***	(−9.239)
Platform dummies	*NO*		*NO*		*YES*	
Constant cut1	0.562***	(−3.760)	0.524*	(−1.645)	0.146***	(−3.825)
Constant cut2	1.842***	(3.999)	6.255***	(4.415)	2.142	(1.488)
Constant cut3	4.303***	(9.221)	34.19***	(8.354)	16.85***	(5.470)
Observations	931		931		931	
chi2	69.57		877.48		1065.52	
ρ	0.000		0.000		0.000	
ρ-value test proportionality of odds	0.6783		0.1852		0.085	
ρ-value Brant test	0.710		0.084		0.000	
ρ-value literacy coefficients=0	0.000		0.000		0.000	
Pseudo R2	0.0282		0.3562		0.4325	

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

classes are decreasing when literacy is increasing while high DIV classes show increasing predicted probabilities when literacy is increasing. Note that the real predicted probabilities of being in DIV1 are very low due to the fact that other predictors, such as invested capital, are at mean value. By using predicted probabilities, I clearly show that investors with high literacy outperform investors with average literacy and that the higher proportional odds coefficient in OLR(3) of average literacy relative to high literacy is distorted.

6 Discussion and extensions

6.1 Limitations

Campbell (2006) highlights the limitations households face when making investment decisions. For example; fixed costs, uninsurable income risk, borrowing constraints, and contracts that are non-neutral with respect to inflation. Moreover, he highlights challenges regarding doing research and collecting data on household finance. Doing research on household finance is more challenging than in the field of the traditional financial institutions, i.e. deposit takers, investors and insurers. Campbell (2006) mentions several challenges in collecting data and how to measure it correctly. Households are unwilling to disclose their financial situation. Since this research is partly based on survey data from non-professionals I should take self-reported results with caution. I do find evidence that investors do not accurately report their portfolio data. For investors that participate on only one platform, their self-reported assets should be equal to platform-data assets. However, for these investors I find evidence that there is substantial variation in self-reported assets and platform-data assets.

6.2 Real knowledge, cognition or searching for right answer

One could argue that the financial literacy index is not very accurate, and does not reflect investors' real economic knowledge. First, because it only contains three multiple choice questions, guessing is a significant factor. Second, the questions are elementary knowledge in the investors scene, so by having invested a couple of times, people could recognize the subjects and answers, the "cognition" phenomena. Third, the survey is taken in an online environment. By using a search engine, one can easily find the answers on the internet, leading to perfect score on literacy. I do find evidence for the cognition phenomena. The variable "past literacy", which is defined as being an execution-only investor, highly influences literacy. Moreover, I find indirect evidence for the "search engine" phenomena. In table 6, the single-cross section of education on literacy is shown as one would expect. However, the high literacy does not outperform the average literacy, which could be because people who are in the high literacy do not have real knowledge, a signal to the "search engine" phenomena. I would recommend further researchers to extend the literacy test, so one can clearly exclude these biases.

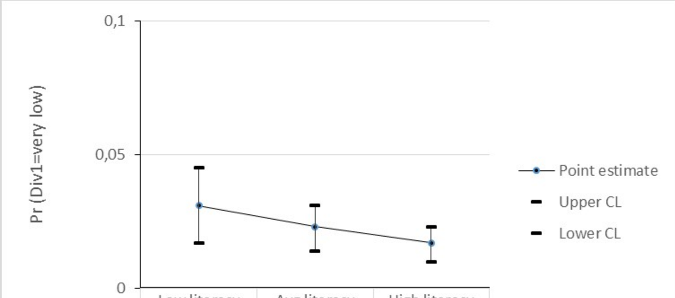
6.3 Local bias, familiarity, and naïve diversification

I extend my research by studying the local bias and familiarity bias in crowdfunding portfolios which I test with hypothesis 4. I detect whether someone makes local investments by looking at their motivation to crowdfund. I select three dummies. Dummy 1, "Motivation1", signals local investment or familiarity when the investor knows the fund seeker and wants to support him. Dummy 2, "Motivation2", signals when the motivation to invest in the last project is knowing the fund seeker as friend, family or other person within the investor's network.[19] I call this

[19]For exact wording of these questions see the appendix C.

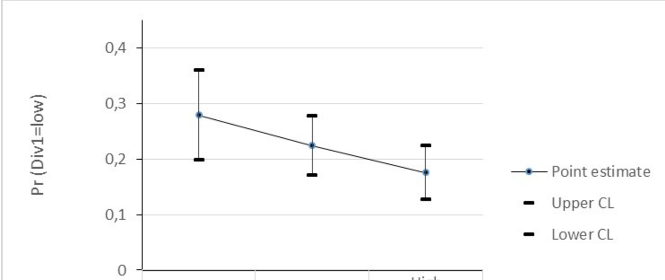
Figure 1

Predicted probabilities of being in one of the 4 quintiles of DIV1, respectively, (a)very low, (b)low, (c)average, (d)(very) high for all three financial literacy tertiles, holding other control variables at their means. Using the same control variables as in model "DIV1 (OLR3)", all estimations are significant at a 1% level. I report the point estimates, the upper confidence and lower confidence of the 95% level. The predicted probabilities of (a),(b), (c), and (d) sum up till 100% for low literacy, average literacy, and high literacy, which is shown in the last figure (e).



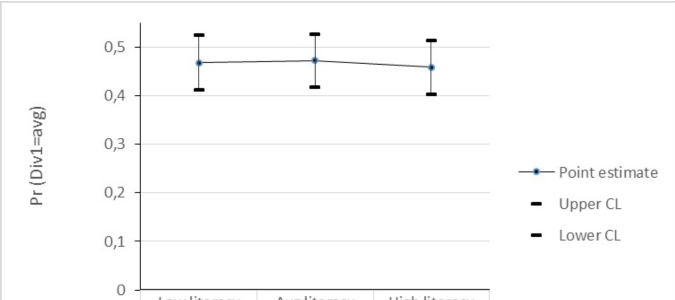
	Low literacy	Avg literacy	High literacy
Point estimate	0,031	0,023	0,017
Upper CL	0,045	0,031	0,023
Lower CL	0,017	0,014	0,01

(a)



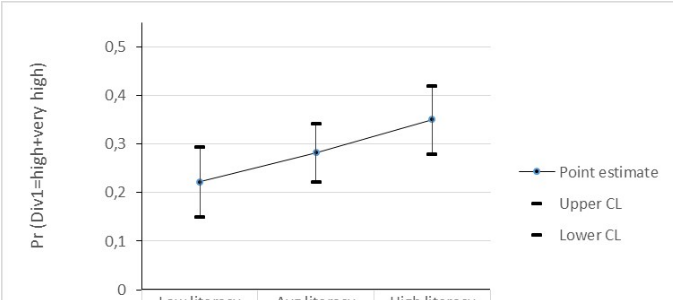
	Low literacy	Avg literacy	High literacy
Point estimate	0,279	0,224	0,176
Upper CL	0,36	0,278	0,224
Lower CL	0,199	0,171	0,128

(b)



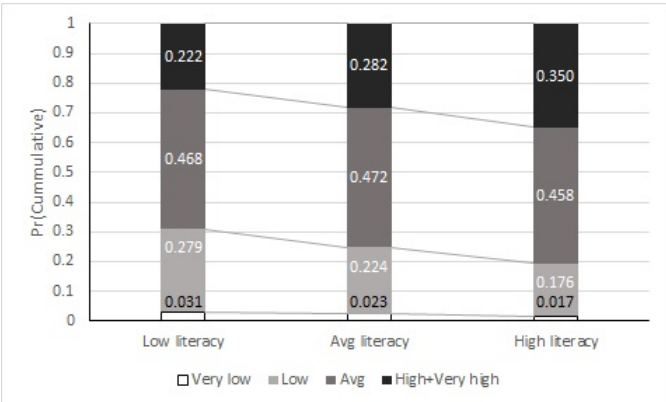
	Low literacy	Avg literacy	High literacy
Point estimate	0,468	0,472	0,458
Upper CL	0,524	0,526	0,513
Lower CL	0,412	0,418	0,402

(c)



	Low literacy	Avg literacy	High literacy
Point estimate	0,222	0,282	0,35
Upper CL	0,294	0,342	0,42
Lower CL	0,15	0,221	0,279

(d)



(e)

motivation two. Dummy 3, "Motivation 1&2", is equal to one if both dummy 1 and dummy 2 are equal to one.

Table 12

This table reports the effect of motivation on behavior. I select two answer options to determine local investment. First, if the investor's motivation to crowdfund is that he knows the crowdfunder, motivation1 = Yes. Second, if the motivation to choose the last project is that the investor knows the fund seeker, motivation2 = Yes. For the last column, motivation 1 and 2 holds. Panel A reports summary statistics. Panel B shows the proportion in the DIV1 quintiles. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

Panel A: Summary statistics, households grouped by motivation

				Crowdfunding investor motivations								
	All			Motivation 1 (Yes) (N=51)			Motivation 2 (Yes) (N=70)			Motivation 1 & 2 (Yes) (N=46)		
Variable	x̄	Md	σ	x̄	Md	σ	x̄	Md	σ	x̄	Md	σ
Assets (self-reported)	30.5	10	55.8	3.4	1	10.8	6.3	1	17.7	1.4	1	0.8
Assets (platformdata)	16.3	6	20.3	2.0	1	5.3	1.5	1	1.5	1.0	1	0.18
Invested cap (self-reported) (thousand of EUR)	21.5	8.0	43.8	3.2	10.8	0.7	5.2	1.0	12.9	1.4	0.5	2.5
Invested cap (platformdata) (thousand of EUR)	11.9	4.2	18.2	2.3	0.5	6.6	1.4	0.5	2.2	1.0	0.5	1.9
Avg investment (platformdata) (thousand of EUR)	0.88	0.55	1.31	1.10	0.5	1.99	1.10	0.5	2.00	0.97	0.5	1.91

Panel B: Weighted percentages of proportion households in DIV1 quintiles, grouped by motivation

Motivation	DIV1 quintiles 1 (very low)	2(low)	3 (avg)	4 (high)	5 (very high)
All	17.4	21.9	19.8	20.6	20.3
Motivation1 (Yes)	70	24.3	2.9	-	2.9
Motivation2 (Yes)	58.6	27.1	8.6	-	5.7
Motivation 1&2 (Yes)	76.1	23.9	-	-	

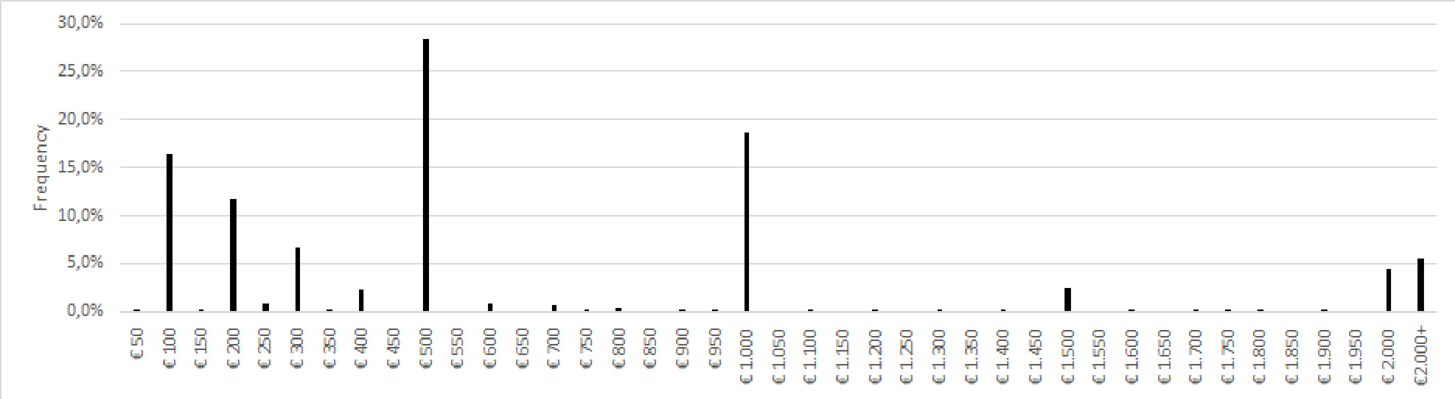
Table 12 clearly shows that people who are motivated by the fact that they know the fund seeker and want to support him, have under diversified portfolios. I accept hypothesis 4 and find evidence that motivation, goals and expectations are linked to economic behavior. In other words, people behave according to what they say. Moreover, the average investment amount is higher for people who are motivated by familiarity. A higher average investment amount in combination with a more under diversified portfolio, lead to more exposure to idiosyncratic risk. Altogether, I show that crowdfunders are very heterogeneous investors, ranging from feelgood-funders that are very unsophisticated and are driven by familiairity, to high-stake investors that have very sophisticated portfolios of more than 50 projects. Additionally, I find evidence for some form of naïve diversification. People often invest the same amounts in different projects. In my sample of 14026 transactions, most transactions are 100 euro (16%), 200 euros (12%), 500 euros (28%) and 1.000 euros (19%). Figure 2 shows the distribution of investment amounts of all crowdfunding transactions.

6.4 Crowdfunding portfolio size to net worth

I show that a substantial part of the crowdfunding portfolios is under diversified. However, one could argue that serious investments are placed in other asset classes, such as equity portfolios, mutual funds or pension funds. In this way, crowdfunding could be used as a "playing account", meant primarily for entertainment and local financial support. If the bulk of real investment accounts is well diversified, the idiosyncratic risk of the crowdfunding projects is negligible for the total household investment portfolio. I test this with the following hypothesis:

Figure 2

Frequency graph of investment amounts in euros. The data is from 863 crowdfunding investors retrieved from the 5 platform databases, collected in April, 2016. n=14026



Hypothesis 7: Crowdfunding is primarily used as playing accounts, serious investments are held in other asset classes.

Almost half of the respondents report how much of their wealth they invest in crowdfunding, how much in other investments and how much they save. By multiplying this self-reported percentage by the categorical variable self-reported wealth, I am able to roughly estimate the real size of the investment portfolio.
I extend the DIV1 OLR(2) model[20] by controlling for other investments. I cannot find reliable results in the regression when dropping half of the households that did not report their asset allocation. For this reason, I include dummies for "unknown allocation" to keep sufficient tests in the ordinal logistic framework. I show that literacy still strongly explains diversification after controlling for other investment portfolios. Second, when looking at the portfolio size to annual income ratio (SIR) and relative to net worth (SNWR), I find that SIR equals 0.36 and SNWR equals 0.19. In other words, a crowdfunding portfolio is roughly 36% of the annual net income of the investor, and 19% of the investor's net worth. I reject hypothesis 7 and shows that a large part of the crowdfunding portfolios represents serious investment accounts. However, these results should be taken cautiously, because I show in table 5 that selection bias might influence these results, i.e. sample averages are higher than averages of all platform investors. Lastly, I find a strong correlation between the part of wealth invested in crowdfunding and diversification.
The outcome of 19% SNWR is surprising. The Dutch crowdfunding market only counts 128 million euros of capital (Douw & Koren, 2016), which is a fraction of the total Dutch capital market. For example, the market capitalization of all Dutch listed companies is more than 600 billion euros (Bank, 2016). So, if one would replicate the market portfolio, according to modern portfolio theory (MPT), only a negligible part should be invested in Dutch crowdfunding projects. Moreover, a sophisticated investor would replicate the world-wide market portfolio, in which the Dutch market portfolio itself is only a small fraction. Altogether, the 10% SNWR recommendation of the AFM is questionable in a risk/return framework of MPT.

[20] I use exactly the same control variables as in 11 OLR (2), except for the "pastliteracy" variable because it shows multicollineairity with other investments.

Table 13

This table shows the effect of financial literacy on diversification when controlling for other investment accounts. Panel A reports the two measurements of other investment portfolios. Panel B shows the ordered logistic regression of the effect of financial literacy and controls on DIV1. The base regression shows the OLR(2) model in table 11. In OLR(2) I control for other investment portfolios relative to wealth, self reported percentages, and in OLR (3), I control for absolute amount of other investments, determined as self-reported percentages * self-reported wealth. The coefficients are reported as proportional odds ratios. Three cutpoints are made to divide the four quartile effect. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test test the null hypothesis that the parallel regression assumption is valid. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

Panel A: Weighted percentages of households that hold other investment portfolios

	<5%	5% - 20%	20% - 50%	>50%	*N*
Other investments relative to NW	9.4	24.5	43.1	22.0	445
	<5k	5k - 15k	15k - 50k	> 50k	*N*
Other investments (absolute amount in EUR)	25.9	22.6	36.8	14.7	402

Panel B: OLR regressions

	(1)		(2)		(3)	
Variables	DIV1 (base)		DIV1 (OLR2)		DIV1 (OLR3)	
Financial literacy dummies (Base group: low literacy)						
Average literacy	2.385^{***}	(3.944)	2.384^{***}	(3.921)	2.423^{***}	(3.995)
High literacy	2.687^{***}	(4.559)	2.678^{***}	(4.513)	2.733^{***}	(4.632)
Other investment relative to wealth (Base group: $\leq$ 5%)						
5% < other inv portfolio $\leq$ 20%			0.882	(−0.296)		
20% < other inv portfolio $\leq$ 50%			0.841	(−0.429)		
other inv portfolio > 50%			0.565	(−1.348)		
Unknown			0.570	(−1.483)		
Other investments (abs amount in EUR) (Base group: other inv <5000)						
5 000 < other inv $\leq$ 15000					0.733	(−0.946)
15 000 < other inv $\leq$ 50000					0.797	(−0.735)
other inv > 50 000					0.557	(−1.571)
Unknown					0.573^{**}	(−2.240)
Other controls (see table 11 OLR2)	*YES*		*YES*		*YES*	
Platform dummy controls	*NO*		*NO*		*NO*	
Constant cut1	0.524^{*}	(−1.645)	0.257^{**}	(−2.545)	0.274^{***}	(−2.874)
Constant cut2	6.255^{***}	(4.415)	3.076^{**}	(2.060)	3.268^{**}	(2.543)
Constant cut3	34.19^{***}	(8.354)	16.90^{***}	(5.157)	17.95^{***}	(6.149)
Observations	931		931		931	
chi2	877.48		879.02		878.27	
ρ	0.000		0.000		0.000	
ρ-value test proportionality of odds	0.1852		0.1935		0.2656	
ρ-value Brant test	0.084		0.114		0.153	
ρ-value test literacy coefficients=0	0.000		0.000		0.000	
Pseudo R2	0.3562		0.3568		0.3565	

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

6.5 Including risk aversion in the model

One dependent variable that is not included in the regression models is risk tolerance. A preference for risk could explain some variation in diversification. However, in theory, risk tolerance does not clearly explains the absolute amount of assets, as it does with differences in risk/return portfolios or stock market participation as Van Rooij et al. (2011) describes. Another important discussion point is the framing effect in questioning risk aversion. The survey contains two almost identical questions[21], only the framing promotes contrary answers. However, these questions do not show contrary results.

Table 14

This table shows the ordered logistic regression of financial literacy on diversification when controlling for risk aversion. The base regression shows the OLR(2) model in table 11. In OLR(2) I add dummies for risk tolerance, and in OLR (3), I add platform dummies. The coefficients are reported as proportional odds ratios. Three cutpoints are made to divide the four quartile effect. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test test the null hypothesis that the parallel regression assumption is valid. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

Variables	(1) DIV1 (base)		(2) DIV1 (OLR2)		(3) DIV1 (OLR3)	
Financial literacy dummies (Base group: low literacy)						
Average literacy	2.385***	(3.944)	2.423***	(3.988)	2.414***	(3.741)
High literacy	2.687***	(4.559)	2.747***	(4.582)	2.319***	(3.554)
Risk tolerance level(Base group: Risk averse)						
Risk neutral			0.943	(−1.178)	0.970	(−0.570)
Risk seeking			1.040	(0.752)	1.045	(0.809)
Other controls (see table 11 OLR2)	*YES*		*YES*		*YES*	
Platform dummy controls	*NO*		*NO*		*YES*	
Constant cut1	0.524*	(−1.645)	0.503	(−1.589)	0.152***	(−3.539)
Constant cut2	6.255***	(4.415)	6.052***	(3.984)	2.248	(1.499)
Constant cut3	34.19***	(8.354)	33.17***	(7.643)	17.69***	(5.275)
Observations	931		931		931	
chi2	877.48		879.12		1066.33	
ρ	0.000		0.000		0.000	
ρ-value test proportionality of odds	0.1852		0.1810		0.0247	
ρ-value Brant test	0.084		0.070		0.000	
ρ-value test risk tolerance coefficients=0			0.4431		0.6677	
Pseudo R2	0.3562		0.3569		0.4329	

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

In the single cross-section in table 10, I show that risk averse people are more under diversified than risk-seeking agents. When including risk tolerance in the OLR (2) model in table 11, the estimates of the financial literacy coefficients do not change appreciably. I cannot reject that the risk aversion coefficients are different from each other. Altogether, the exclusion of my measure for risk aversion does not significantly lead to omitted variable bias in the full model.

6.6 Looking at literacy questions separately

Lastly, I want to focus on the effect of the literacy questions seperately on diversification. Because the literacy questions contain different investment topics, it could be interesting to look at differences in diversification behavior. The topic of the first question is about the inverse relation between interest and bonds, question two is about the difference between risk of a mutual funds and a stock, and the third question is about decreasing risk when the assets in the portfolio increases.[22] Because question three is directly linked to the degree of diversification, I expect that correctly answering the third question mostly explains diversification.

[21] The 1st and 3th question of risk tolerance, see for exact wording appendix C.

[22] See the appendix C for the exact wording of questions.

Table 15

This table shows the ordered logistic regression of financial literacy on diversification when controlling for risk aversion. The base regression shows the OLR(2) model in table 11. In OLR(2) I add dummies for risk tolerance, and in OLR (3), I add platform dummies. The coefficients are reported as proportional odds ratios. Three cutpoints are made to divide the four quartile effect. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test test the null hypothesis that the parallel regression assumption is valid. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

	(1)		(2)		(3)		(4)	
Variables	DIV1(base)		DIV1 (OLR2)		DIV1(base)		DIV1 (OLR4)	
Q1 = Incorrect/DK/Refusal	0.542***	(−4.748)	0.801	(−1.368)				
Q2 = Incorrect/DK/Refusal	0.570***	(−3.421)	0.719	(−1.565)				
Q3 = Incorrect/DK/Refusal	0.488***	(−3.210)	0.554**	(−2.002)				
Q1 = Incorrect					0.766*	(−1.702)	0.875	(−0.692)
Q2 = Incorrect					0.867	(−0.563)	1.198	(0.572)
Q3 = Incorrect					0.494***	(−2.740)	0.591	(−1.543)
Other controls (see table 11 OLR2)	*NO*		*YES*		*NO*		*YES*	
Platform dummy controls	*NO*		*YES*		*NO*		*YES*	
Constant cut1	0.127***	(−1858)	0.0600***	(−5.526)	0.189***	(−17.51)	0.0863***	(−4.984)
Constant cut2	0.418***	(−9.739)	0.861	(−0.295)	0.588***	(−6.965)	1.207	(0.382)
Constant cut3	0.978	(−0.258)	6.732***	(3.747)	1.320***	(3.727)	9.315***	(4.503)
Observations	931		931		931		931	
chi2	72.36		1061.35		10.43		1052.60	
ρ	0.000		0.000		0.0153		0.000	
ρ-value test proportionality of odds	0.8324		0.1339		0.9658		0.2816	
ρ-value Brant test	0.848		0.000		0.974		0.000	
ρ-value F-test Q1,Q2,Q3 coefficients=0	0.0000		0.0086		0.3716		0.0147	
Pseudo R2	0.0294		00.4308		0.0042		0.4273	

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

I leave out the literacy tertiles and create dummies for incorrect answers on the three questions. Table 15 reports the three questions separately, and shows its relation on diversification. The first three dummies are equal to one when the answer is incorrect, don't know, or refusal. The second three dummies are only equal to one when answering incorrectly. In the base regression, I find evidence that incorrectly answering question 3 mostly affects under diversification.

7 Conclusion

One of the main risks in household finance is under diversification. Merging survey data with real transaction data of Dutch households, I quantitatively provide insights in the crowdfunding market in the Netherlands. First, I focus on financial literacy and find evidence that literacy of crowdfunding investors correlates with similar characteristics as general investors in other relevant papers. I demonstrate that a significant proportion of crowdfunding investors have under diversified portfolios. My results show that under diversification is incurred by people who score low on financial literacy. The effects are strong enough to show significant results after controlling for a large set of factors including invested capital, age, education, gender, income, wealth and several other investor characteristics. The effect of literacy is sizable. Having average or high literacy relative to low has a stronger effect on diversification than being an execution-only investor. Moreover, I find several links between motivation and actual investment behavior. Those who report 'feel-good' goals and motivations do under diversify. Similarly, investors that are motivated by local investments and familiarity show under diversified portfolios. Moreover, I find evidence for the familiarity bias, those who know the fund seeker invest on average more in one project and have smaller portfolios. Additionally, the average of the approximated portfolio size to net worth (SNWR) is 19%, almost twice as large as recommended by the regulatory agency.

My results suggest that regulators seeking to reduce welfare loss of consumers should try to increase financial numeracy. The data is obtained just before the AFM announced new regulations

to protect investors. This is the baseline measurement. The follow-up research, in November 2016, will investigate if the adopted rules are effective.
For further research, I recommend researchers to include more questions on financial literacy, so that this could be deeper analyzed. Furthermore, the regulator should not only focus on the risk of under diversification but also on risk perception and responsible asset allocation. From my perspective, I especially see problems for the return-seeking investor, who does not invest execution only, but expects to find a high and safe return on his crowdfunding investment.
Taking all these conclusions together, the results of insufficient diversification and a high SNWR are signalling that the average crowdfunding investor bears substantial idiosyncratic risks. The market does not compensate for this diversifiable risk, leading to low performing portfolios.

References

AFM. (2014, dec). *Crowdfunding, naar een duurzame sector.* https://www.afm.nl/~/profmedia/files/rapporten/2014/crowdfunding.ashx.

AFM. (2015, dec). *Nieuwsbrief crowdfunding.* https://www.afm.nl/nl-nl/professionals/nieuwsbrief/crowdfunding-archief.

Bank, E. C. (2016, jun). *See - securities exchange - trading statistics [data file and codebook].* http://sdw.ecb.europa.eu/browseExplanation.do?node=qview\&SERIES_KEY=181.SEE.A.NL.ENA0.MKP.W.E.

Barber, B. M., & Odean, T. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, *55*(2), 773–806.

Barber, B. M., & Odean, T. (2002). Does online trading change investor behavior? *European Business Organization Law Review*, *3*(01), 83–128.

Barber, B. M., & Odean, T. (2011). The behavior of individual investors. *Available at SSRN 1872211*.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, *1*, 1053–1128.

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

Benartzi, S. (2001). Excessive extrapolation and the allocation of 401 (k) accounts to company stock. *The Journal of Finance*, *56*(5), 1747–1764.

Benartzi, S., & Thaler, R. H. (2001). Naive diversification strategies in defined contribution saving plans. *American economic review*, 79–98.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2006). *Down or out: Assessing the welfare costs of household investment mistakes* (Tech. Rep.). National Bureau of Economic Research.

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2009). Measuring the financial sophistication of households. *The American Economic Review*, *99*(2), 393–398.

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

Collins, L., Swart, R., & Zhang, B. (2013). The rise of future finance: The uk alternative finance benchmarking report. *London, UK: Nesta, University of Berkeley, and University of Cambridge.*.

Cooper, I., & Kaplanis, E. (1994). Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *Review of Financial Studies*, *7*(1), 45–60.

Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *The journal of political economy*, 401–419.

Douw, S. (2015, dec). *Belemmering groei crowdfunding verwacht door meer regels.* http://startcrowdfunding.nl/nieuws/belemmering-groei-crowdfunding-verwacht-door-meer-regels/.

Douw, S., & Koren, G. (2016, feb). *Crowdfunding in nederland 2015.* http://www.dkcrowdfunding.nl/\#\!crowdfunding-in-nederland-in-2015/gz4a7.

Engelberg, J. E., & Parsons, C. A. (2011). The causal impact of media in financial markets. *The Journal of Finance*, *66*(1), 67–97.

Evans, J. L., & Archer, S. H. (1968). Diversification and the reduction of dispersion: an empirical analysis. *The Journal of Finance*, *23*(5), 761–767.

Freitas, X., & Rochet, J. C. (1997). Microeconomics of banking. *Cambridge, Mass./London*, *1*.

French, K. R., & Poterba, J. M. (1991). Investor diversification and international equity markets. *The American Economic Review*, *81*(2), 222–226.

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification*. *Review of Finance*, *12*(3), 433–463.

Graham, J. R., Harvey, C. R., & Huang, H. (2009). Investor competence, trading frequency, and home bias. *Management Science*, *55*(7), 1094–1106.

Huberman, G. (2001). Familiarity breeds investment. *Review of financial Studies*, *14*(3), 659–680.

Kelly, M. (1995). All their eggs in one basket: Portfolio diversification of us households. *Journal of Economic Behavior & Organization*, *27*(1), 87–96.

Long, J. S. (1997). Regression models for categorical and limited dependent variables. *Advanced quantitative techniques in the social sciences*, *7*.

Lusardi, A., & Mitchell, O. S. (2007). Financial literacy and retirement planning: New evidence from the rand american life panel. *Michigan Retirement Research Center Research Paper No. WP*, *157*.

Marom, D., Robb, A., & Sade, O. (2015). Gender dynamics in crowdfunding (kickstarter): Evidence on entrepreneurs, investors, deals and taste-based discrimination. *Investors, Deals and Taste-Based Discrimination (December 6, 2015)*.

Mitchell, O. S., & Utkus, S. P. (2002). *The role of company stock in defined contribution plans* (Tech. Rep.). National Bureau of Economic Research.

Moenninghoff, S. C., & Wieandt, A. (2013). The future of peer-to-peer finance. *Zeitschrift für betriebswirtschaftliche Forschung*, 466–487.

Mollick, E. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, *29*(1), 1–16.

Polkovnichenko, V. (2005). Household portfolio diversification: A case for rank-dependent preferences. *Review of Financial Studies*, *18*(4), 1467–1502.

Puri, M., & Robinson, D. T. (2007). Optimism and economic choice. *Journal of Financial Economics*, *86*(1), 71–99.

Schwienbacher, A., & Larralde, B. (2010). Crowdfunding of small entrepreneurial ventures. *Handbook of entrepreneurial finance, Oxford University Press, Forthcoming*.

SEC. (2015, oct). *Sec adopts rules to permit crowdfunding press release.* `https://www.sec.gov/news/pressrelease/2015-249.html`.

Statman, M. (1987). How many stocks make a diversified portfolio? *Journal of Financial and Quantitative Analysis*, *22*(03), 353–363.

Van Horne, J. C., Blume, M. E., & Friend, I. (1975). The asset structure of individual portfolios and some implications for utility functions. *The Journal of Finance*, *30*(2), 585–603.

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449–472.

Von Gaudecker, H.-M. (2015). How does household portfolio diversification vary with financial sophistication and advice? *The Journal of Finance*, *70*(2), 489–507.

Wardrop, R., Zhang, B., Rau, R., & Gray, M. (2015). Moving mainstream: The european alternative finance benchmarking report. *London: University of Cambridge and Ernst & Young*.

Winship, C., & Mare, R. D. (1984). Regression models with ordinal variables. *American Sociological Review*, 512–525.

Zhu, N. (2002). The local bias of individual investors. *Unpublished paper, Yale University*.

Appendices

A Detailed data description

A.1 Tertiles based on financial literacy scores

The following section shows details about the determination of financial literacy. The survey contains three questions about financial sophistication. The first question is; 'if the interest drops, what should happen with the bond prices?' People could choose between; 'they should increase', 'they should decrease', 'they should not change', 'I don't know', 'I refuse to answer'. The second statement is; 'buying a company stock usually provides a safer return than a stock mutual fund.' Possible answers are; 'true', 'false', 'I don't know', 'I refuse to answer'. The last question is; 'what happens with the risk of losing money if an investor diversifies his investment over different assets?' People could choose between; 'risk will increase', 'risk will decrease', 'risk will be the same', 'I don't know'. Initially, I added also a fourth question that measures the ability to assess default risk of different asset classes. The question was; 'estimate the probability that you will not receive the initial amount of money that you have invested for Dutch government bonds, mutual fund in Dutch stocks, AEX listed stocks, and turbos in the AEX.' Respondents could choose among 'very low', 'low', 'average', 'high', 'very high'. It is correctly answered when people estimate the probabilities in the right order, for example probabilities for government bonds lower than mutual funds. I removed the fourth question because it scores weak when I test for internal consistency. A possible explananation could be that, in contrast to the other three questions, question four is more about risk perception and less about basis financial sophistication.
Following other research, I grade a correct answer with one point. An incorrect answer, 'do not know' and refusals all get zero points. Alternatively, I graded correct answers with two points, 'do not know' and refusal with one, and incorrect answers with zero points. This alternative measurement is motivated by the following hypothesis; unconsciously answering wrong should correspond with lower literacy than consciously answering 'I do not know'. However, I could not find evidence in the literature for this hypothesis. Contrary, other research do treat incorrect, do not know and refusal all as the same.. Another motivation to reject the alternative measurement is that the Cronbach's alpha test, of internal consistency, is much lower for the alternative measurement ($\alpha = 0.27$) with respect to the original one ($\alpha = 0.67$).
The maximum score of literacy is three points and the minimum score is zero points. I rank literacy in tertiles, low, average, and high literacy, see table A1. I show that the largest group has been graded with full points. Based on the difficulty of the question and equal distribution of observations, low literacy corresponds with zero or one point, which reflects everything false or coresponds to answering randomly. Averag literacy corresponds to two points and high literacy with three points.

A.2 The distribution of income and wealth among crowdfunders

The survey uses the standard scale of gross annual income and net worth similarly to other household surveys such as the DNB household survey. The results are reported in table A2.

A.3 Determination of risk preferences

We group individuals into three different risk tolerance levels based on three questions from the survey. People could answer the following questions on a likert scale from (1) totally agree to (7) totally disagree. First, "I am willing to accept a lower return if this provides more certainty."

Table A1

Tertiles based on the financial literacy score. Households get one points for each correct answer, and zero points for "do not know", "refusal" and incorrect answer. All questions correctly answered gives three points, all questions incorrectly answered gives zero points. Cut-off points are based on both equal distribution and the difficulty of the questions. The data is from the AFM crowdfunding survey, collected in March 2016.

Score	0	1	2	3
N	39	109	298	485

Categories based on score	Score interval	*N*	Weighted %
Low financial literacy	0-1	148	15.9
Average financial literacy	2	298	32.0
High financial literacy	3	485	52.1

Table A2

Distribution of gross annual income (EUR) and net worth (EUR) (weighted percentages, rows & columns equal to 100). The Pearson chi^2 statistic tests the null hypothesis that the distribution of income is independent of net worth (p-values are reported in parentheses). Both annual income and net worth are self-reported. The data is from the AFM crowdfunding survey, collected in March 2016.

	Income							
Net worth	<12 500	12 500-33 000	33 000-40 000	40 000-66 000	66 000-80 000	>80 000	Refusal	Total
<10 000	0.5	1.1	0.4	1.0	0.6	0.7	0	4.4
10 000 - 25 000	0.5	1.2	2.3	2.8	1.3	0.7	0	8.8
25 000 - 50 000	0.1	1.3	2.6	3.7	1.8	4.0	0.1	13.6
50 000 - 80 000	0.2	0.5	1.9	3.1	2.1	4.1	0.1	12.3
80 000 - 150 000	0.1	0.7	1.3	4.3	2.7	4.6	0.1	13.7
>150 000	0	1.6	2.2	7.4	6.9	17.1	1.2	36.5
Refusal	0.1	0.5	0.7	1.1	1.4	1.9	5.0	10.7
Total	1.6	7.1	11.5	23.3	16.8	33.3	6.5	100
	Pearson chi2(36)= 454.12			(p=0.000)				

Second, "with respect to peer investors, I usually take more risk". Third, "avoiding risk is more important than a high return." For the risk measurement, the second question will be contrary rated, so strongly agree will get seven points instead of one. A row total score is used to divide investors into three positions. The respondent could get the maximum score of 21, which corresponds to risk seeking, and the minimum score of three, which corresponds to risk averse. Table A3, shows the distribution of the three risk tolerance levels.

Table A3

Determination of risk preferences. Risk scores are based on three likert-scale questions. If households answer totally agree for the first and third question and totally disagree for the second question, they receive three points, and are categorized as risk averse investors. If households answer totally disagree on question one and three and totally agree on question two they will receive 21 points and are categorized as risk seeking investors. Tertiles are based on equal distributions. The data is from the AFM crowdfunding survey, collected in March 2016.

Risk tolerance level	min (Risk score)	max (Risk score)	mean (Risk score)	*N*
Risk averse	3	9	7.25	334
Risk neutral	10	12	11.03	313
Risk seeking	13	20	14.48	284

A.4 Financial literacy across continuous variables

Table A4

This table reports the summary statistics of the distribution of financial literacy scores across continuous variables (demographics & behavior). More specifically, it estimates the mean, median and standard deviation of the continuous variables age, self-reported assets, platform-specific assets, self-reported invested capital and platform-specific invested capital. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016)

				Literacy tertiles								
	all (N=931)			1 (low) (N=148)			2 (avg) (N=298)			3 (high) (N=485)		
Variable	x̄	Md	σ	x̄	Md	σ	x̄	Md	σ	x̄	Md	σ
Age	50.7	52	13.2	50.9	53	14.2	50.3	51	13.3	50.9	52	12.7
Assets (self-reported)	30.5	10	55.8	18.2	3	52.5	25.3	10	52.1	37.4	18	58.2
Assets (platformdata)	16.3	6	20.3	8.5	2	13.7	15.6	6	19.8	19.0	10	21.6
Invested cap (self-reported) (thousand of EUR)	21.5	8.0	43.8	11.4	2.3	36.5	18.8	7.4	46.3	26.3	11	43.7
Invested cap (platformdata) (thousand of EUR)	11.9	4.2	18.2	5.8	1.3	9.5	11.7	4.0	18.4	13.8	5.5	19.6

A.5 Details of brant test of the ordinal logistic regression of financial literacy

Table A5

This table reports the brant test of the ordinal logistic regression model of financial literacy. All variables are reported as in regression OLR(3) in table 8. A significant test statistic provides evidence that the parallel regression assumption has been violated.

	Estimated coefficients from the 2 (j-1) binary regressions		Brant test of Parallel regression assumption	
Variable	Y1	Y2	chi2	ρ
Age dummies (Base group: age $\leq$ 30)				
30 < age $\leq$ 40	0.500	0.181	0.50	0.481
40 < age $\leq$ 50	0.313	0.206	0.06	0.809
50 < age $\leq$ 60	0.463	0.303	0.13	0.714
60 < age $\leq$ 70	-0.349	-0.201	0.09	0.766
age >70	-0.322	0.0688	0.33	0.563
Education dummies(Base group: lager onderwijs)				
Mavo/Mulo/Vmbo	0.680	2.690	2.48	0.115
Havo/Vwo/Mbo	0.501	1.893	1.31	0.253
HBO	0.984	2.331	1.23	0.268
WO	1.313	2.500	0.94	0.332
Female	-1.405	-0.851	4.77	0.029**
Retired	0.197	0.097	0.84	0.359
Self-employed	0.251	-0.014	0.84	0.359
Household income dummies (Base group: income < 12 500)				
12 500 $\leq$ income < 33 000	-1.063	-0.203	1.04	0.308
33 000 $\leq$ income < 40 000	-0.613	-0.387	0.07	0.789
40 000 $\leq$ income < 66 000	-0.455	-0.101	0.18	0.673
66 000 $\leq$ income < 80 000	-0.715	-0.076	0.56	0.455
income $\geq$ 80000	0.0187	0.292	0.10	0.750
Refusal	-0.134	0.455	0.41	0.521
Financial wealth dummies (Base group: wealth < 10 000)				
10 000 $\leq$ wealth < 25 000	0.041	0.685	1.31	0.253
25 000 $\leq$ wealth < 50 000	0.042	0.367	0.35	0.557
50 000 $\leq$ wealth < 80 000	0.256	0.506	0.20	0.658
80 000 $\leq$ wealth < 150 000	0.227	0.492	0.21	0.647
wealth $\geq$ 150000	0.993	1.119	0.05	0.821
Refusal	-0.589	0.306	2.20	0.138
Incorrect capital estimation	-0.158	-0.380	0.77	0.380
Past literacy	1.198	0.794	2.98	0.084*
Advice	0.851	0.489	0.88	0.348
Interest in trading	2.235	0.156	4.02	0.045**

Note: ***p <0.01, ** p <0.05, * p <0.1.

A.6 Correlation tables

Table A6 shows the correlations among all variables used in my regressions. For uniformity reasons, I report Pearson correlation. I find similar results for Spearman correlations.

A.7 The ordered logistic framework

Choosing an ordered logistic setting (OLR) allows me to regress multivariate models with ordinal variable when three strong assumptions hold. First, it explicitly recognizes the ordinality of the variables "financial literacy" and "DIV1/DIV3". Second, it avoids arbitrary assumptions about their scale. Third, this allows me to analyse the effects of continuous as well as the ordinal variables within one regression. An alternative method could be using OLS in a linear setting

Table A6

Pearson correlation for diversification, financial literacy and controls. (1)=DIV1 (2)=DIV3 (3)=Financial literacy (4)=Invested capital(self-reported) (5)=Invested capital(platformdata) (6)=age (7)=education (8)=gender (9)=retired (10)=self-employed (11)=income (12)=wealth (13)=risk tolerance (14)=participating in stock market (15)=interest in trading (16)=investor type (17)=incorrect capital estimation (18)advice (19)=short investment period (20)=only one platform (21)=Other investments relative to NW (22)= Other investments (abs amount) (23)=Q1 (24)=Q2 (25)=Q3.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1)	1													
(2)	0.62*	1												
(3)	0.26*	0.20*	1											
(4)	0.45*	0.29*	0.13*	1										
(5)	0.48	0.60	0.14*	0.53*	1									
(6)	−0.02	0.02	0.01	0.09*	0.16*	1								
(7)	−0.11*	−0.08*	0.15*	−0.01	−0.09*	−0.16*	1							
(8)	−0.22*	−0.13*	−0.25*	−0.10*	−0.14*	−0.03	−0.01	1						
(9)	−0.06	0.00	−0.03	0.01	0.12*	0.63*	−0.12*	−0.06	1					
(10)	0.01	−0.03	0.04	0.14*	0.05	0.00	0.03	−0.06	−0.26*	1				
(11)	0.02	0.00	0.17*	0.03	0.03	0.03	−0.03	0.19*	−0.11	−0.09*	0.02	1		
(12)	0.31*	0.22*	0.25*	0.22*	0.28*	0.22*	0.03	−0.14*	0.12*	0.04	0.42*	1		
(13)	0.07*	0.06	0.19*	0.05	0.03	−0.06	0.02	−0.14*	−0.05	0.07*	0.10*	0.06	1	
(14)	0.20*	0.08*	0.20*	0.05	0.00	−0.04	0.05	−0.06	−0.04	−0.06	0.01	0.09*	0.08*	1
(15)	0.31*	0.26*	0.11*	0.49*	0.47*	0.07*	−0.03	−0.04	0.06	0.03	0.02	0.19*	0.04	0.01
(16)	0.40*	0.26*	0.13*	0.43*	0.37*	−0.07*	−0.05	−0.15*	−0.07*	0.05	0.07*	0.22*	0.06	0.08*
(17)	−0.48*	−0.23*	−0.14*	−0.20*	−0.20*	−0.05	0.02	0.11*	0.03	−0.04	−0.01	−0.17*	−0.06	−0.13*
(18)	−0.01	0.00	0.06	0.05	0.05	0.11*	−0.01	−0.08*	0.09*	0.04	0.09*	0.13*	0.04	−0.46*
(19)	−0.24*	−0.28*	−0.07*	−0.11*	−0.23*	0.01	0.00	−0.02	0.00	0.02	0.03	−0.04	−0.01	−0.04
(20)	−0.41*	−0.07*	−0.11*	−0.25*	0.01	0.08*	−0.05	0.03	0.09*	−0.03	−0.01	−0.13*	−0.03	−0.08*
(21)	−0.27*	−0.23*	−0.24*	−0.20*	−0.25*	0.01	−0.01	0.10*	0.03	0.02	−0.04	−0.21*	−0.10*	−0.20*
(22)	−0.23*	−0.22*	−0.19*	−0.15*	−0.20*	0.03	0.00	0.08*	0.03	0.04	−0.13*	−0.28*	−0.09*	−0.19*
(23)	−0.22*	−0.19*	−0.83*	−0.16*	−0.16*	−0.06	−0.12*	0.18*	−0.03	−0.04	−0.11*	−0.20*	−0.18*	−0.15*
(24)	−0.20*	−0.13*	−0.73*	−0.07	−0.06	0.07*	−0.11*	0.24*	0.12*	−0.03	−0.17*	−0.20*	−0.14*	−0.15*
(25)	−0.16*	−0.07*	−0.45*	0.01	−0.08*	0.02	−0.05	0.18*	−0.03	−0.01	−0.10*	−0.15*	−0.06	−0.14*

	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)
(15)	1										
(16)	0.25*	1									
(17)	−0.13*	−0.14*	1								
(18)	0.09*	0.02	−0.04	1							
(19)	−0.12*	−0.06	0.15*	0.03	1						
(20)	−0.15*	−0.19*	0.48*	−0.05	0.07*	1					
(21)	−0.17*	−0.23*	0.14*	−0.05	0.03	0.07*	1				
(22)	−0.11*	−0.18*	0.09*	−0.05*	0.06	0.04	0.86*	1			
(23)	−0.10*	−0.15*	0.09*	−0.03	0.08*	0.05	0.19*	0.15*	1		
(24)	−0.08*	−0.06	0.11*	−0.07*	−0.1	0.12*	0.19*	0.15*	0.32*	1	
(25)	−0.05	−0.06	0.10*	−0.06	0.05	0.08*	0.10*	0.10*	0.17*	0.28*	1

Note: * p <0.05.

with dichotomous variables. Using ordered logistic regressions has two main advantages. First, the observed values of assets in the portfolio for both self-reported and platformdata, are highly skewed, for which the logistic regression can control. Second, the ordinal logistic setting takes account of ceiling and floor restrictions and the linear model does not.[23] The downside of the ordinal logistic regression is the difficulty to achieve reliable results. When the model includes many independent variables, the tests for proportional odds often fail. Especially the Brant test fails due to the same characteristics in the "high DIV" and the "very high DIV" classes. For this reason, I merge the 4th and 5th quintile together. This allows me to better focus on the lowest quintiles, which clarifies under diversification. One could argue to divide diversification into three groups only, but I lose a lot of explanation for high diversification, as shown in the single-cross sections in table 7. Another adjustment in the model I make is lowering the categories for age, education, wealth and income. Because, in general, the ordered logit model performs worse for a large set of dependent variables, this further improves the test of proportional odds and the Brant test.

A.8 Ordered logistic regression results of DIV3

In this section, I highlight the main results of the regression framework for DIV3. For this model, the methodology is similar as for DIV1, so I again merge the 4th and 5th quintile together. Because especially the results for OLR2 and OLR3 are unreliable, I want to focus on the base regression. With respect to the DIV1 regression, there is still a significant effect of diversification, even stronger for high literacy. I conclude that using two different measurements of diversification (one only takes into account the number of stocks, the other measures the portfolio relative to the market portfolio) there is a strong effect between financial literacy and crowdfunding portfolio diversification. People with high literacy perfomr even better in DIV3 than in DIV1.

[23]Winship and Mare (1984) have written extensive research in the field of ordered logistic regressions, more information can be found in their paper. Another alternative for the OLR regression would be a Tobit regression. Tobit corrects for the left-censoring of the portfolio size variable, but performs worse with respect to the second assumption of ceiling and floor restrictions. Moreover, coefficients in the ordered logistic regression are more valuable than the Tobit regression, since financial literacy is also ordinally. Long (1997) provides more information on Tobit regressions.

Table A7

Ordered logistic regression of the effect of financial literacy and controls on DIV3. For validity reasons, I merge the 4th and 5th quintile of DIV3. The base regression reports the effect of financial literacy on DIV1. OLR (2) adds controls. OLR (3) additionally controls for differences in platform. The coefficients are reported as proportional odds ratios. Three cutpoints are made to divide the four quartile effect. The score test for proportional odds assumption tests the null hypothesis that restricting the ordinal model to one coefficient is valid. The Brant test tests the null hypothesis that the parallel regression assumption is valid. The data is partly from the AFM crowdfunding survey, collected in March 2016, and from the 5 platform databases, collected in April, 2016

Variables	(1) DIV3 (base)		(2) DIV3 (OLR2)		(3) DIV3 (OLR3)	
Financial literacy dummies (Base group: low literacy)						
Average literacy	2.309***	(4.400)	1.662**	(2.133)	1.682**	(2.128)
High literacy	3.178***	(6.384)	1.938***	(2.827)	1.721**	(2.235)
Invested capital (platformdata)(Base group: invested capital $\leq$ 1000))						
1000 < invested capital $\leq$ 10000			36.96***	(15.06)	24.14***	(10.74)
invested capital > 10 000			695.1***	(20.39)	359.8***	(15.08)
Age dummies (Base group: age $\leq$ 40)						
40<age $\leq$ 60			0.498***	(−3.492)	0.524***	(−3.127)
age > 60			0.556*	(−1.888)	0.568*	(−1.764)
Education dummies(Base group: < HBO)						
HBO			1.109	(0.486)	1.163	(0.683)
WO			1.051	(0.224)	1.224	(0.888)
Female			0.787	(−1.116)	0.853	(−0.720)
Retired			0.457**	(−2.487)	0.487**	(−2.232)
Self-employed			0.824	(−0.990)	0.873	(−0.686)
Household income dummies (Base group: income < 33 000)						
33 000 $\leq$ income < 66 000			0.910	(−0.327)	0.923	(−0.268)
income $\geq$ 66000			0.770	(−0.883)	0.721	(−1.058)
Refusal			0.871	(−0.300)	0.897	(−0.231)
Financial wealth dummies (Base group: wealth < 25 000)						
25 000 $\leq$ wealth < 80 000			0.760	(−1.078)	0.682	(−1.447)
wealth $\geq$ 80000			0.722	(−1.210)	0.767	(−0.954)
Refusal			0.677	(−1.076)	0.728	(−0.863)
Participating in stock market			2.004***	(3.796)	2.161***	(4.044)
Access to financial adviser			1.338	(1.097)	1.594*	(1.721)
Feel good investor			1.098	(0.537)	0.990	(−0.0540)
Invest < 0.5 years			0.446***	(−4.847)	0.569***	(−3.176)
Participate on only one platform			0.694**	(−2.364)	0.636***	(−2.745)
Platform dummies	*NO*		*NO*		*YES*	
Constant cut1	0.587***	(−3.307)	0.782	(−0.590)	0.612	(−0.896)
Constant cut2	1.580***	(2.861)	8.174***	(4.768)	6.458***	(3.295)
Constant cut3	3.655***	(7.869)	61.35***	(9.025)	58.35***	(7.059)
Observations	863		863		863	
chi2	41.56		916.33		1000.27	
ρ	0.000		0.000		0.000	
ρ-value test proportionality of odds	0.0637		0.000		0.000	
ρ-value Brant test	0.082		−		−	
Pseudo R2	0.0181		0.3985		0.4350	

Note: Z-scores are reported in parentheses, ***p <0.01, ** p <0.05, * p <0.1.

B Determination of investor segment

The investor segment is determined according to motivation, goals and expectation answers in the questionnaire. The high stake investor is determined based on portfolio size. To be a feel good funder, several criteria has to be met. In figure B1, I show how I distinguish the investor segments.

Figure B1

Determination of the three investor segments; 1) high-stake investor 2)return-seeking investor 3)feel-good funder. For exact wording of the question, see appendix C.

Investor
- → Portfolio volume larger than 100 000
  - → *Yes*, **High-stake investor**
  - → *No*, Motivation; looking for high return / use it as different asset class
    - → *Yes*, **Return-seeking investor**
    - → *No*, Motivation last project; high expected return
      - → *Yes*, **Return-seeking investor**
      - → *No*, Expectations; more return than other investment instruments
        - → *Yes*, **Return-seeking investor**
        - → *No*, Loss; will give financial troubles / no more investing / annoyance
          - → *Yes*, **Return-seeking investor**
          - → *No*, Need to liquidate within 2 years
            - → *Yes*, **Return-seeking investor**
            - → *No*, **Feel-good funder**

C Survey questions and transaction data

C.1 Survey questions

This section reports the exact wording of the survey questions.

General questions

Are you participating on more than one platform?
[ ] One platform
[ ] More than one platform, [...]

In how many projects have you invested? Note, if you participate on more than one platform, please report all your assets. Please only report loan-based and equity-based projects.
[...]

How much capital have you invested in crowdfunding projects. Note, if you participate on more than one platform, please report invested capital of all projects. Please only include capital invested in loan-based and equity-based projects.
[EUR...]

Have you ever been refused by a platform because you had reached your investment limits?
[ ] Yes
[ ] No

Risk tolerance

Do you agree with the following statements? (on a 7-point scale; 1 means strongly agree and 7 means strongly disagree)
I am willing to accept a lower return if this provides more certainty.
[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] 7
With respect to other investors I usual take more risk.
[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] 7
Avoiding risk is more important than a high return.
[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] 7
Looking back at previous crowdfunding investments, I regret my decision to crowdfund.
[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5 [ ] 6 [ ] 7

Goals, expectations and motivations

What is your most important motivation to crowdfund? (open question)
[ ]

Which of the following motivations are important? (max 3 answers possible)
[ ] a. I know the fund seeker and want to support him/her financially.
[ ] b. I want to financially support projects and SME.
[ ] c. Crowdfunding is an innovative feature to invest. I invest for the experience.
[ ] d. I do not trust the traditional sector such as banks and insurance companies.

[ ] e. I am looking for return.
[ ] f. I like to start an investment portfolio. Crowdfunding makes this accessible.
[ ] g. I use crowfunding as extra asset-class, to diversify my risk.

What do you expect regarding the achievable return with crowdfunding? (more than one answer possible)
[ ] a. I expect a high return.
[ ] b. I do not invest for return reasons.
[ ] c. I expect to achieve more return than for my savings account.
[ ] d. I expect to achieve more return than for my investments in other asset accounts, such as bonds or stocks.

Which situation is most applicable? If I lose the initial investment...
[ ] a. I am unhappy, expect a high return.
[ ] b. I get into trouble, I cannot afford losing the money.
[ ] c. I am disappointed, but do not lose sleep over this.
[ ] d. I do not care, it is part of the game.
[ ] e. I would not invest again.

What was the most important reason to invest in your last project? (max 2 answers possible)
[ ] a. I am attracted by the industry.
[ ] b. Friends, family or other connection started the project.
[ ] c. The guaranteed return is high.
[ ] d. The duration of the bond fits my needs.
[ ] e. The crowdfunding platform marks the project with a low risk classification.
[ ] f. The crowdfunding platform provides clear information about the project.
[ ] g. Other, namely [ ]

When do you need to liquidate your investment?
[ ] a. I need to liquidate at any moment.
[ ] b. I can afford to hold the investment one year.
[ ] c. I can afford to hold the investment 2-5 years.
[ ] d. I can afford to hold the investment more than 5 years.

Suppose you invest in a project with a duration of 5 years. After two years, you need to liquidate your investment? Do you expect it is easy to liquidate?
[ ] a. I expect this is easy.
[ ] b. I expect this is hard.
[ ] c. I expect this is not possible.

Which information is most important for you when taking an investment decision? (max 3 answers possible).
[ ] a. The risk classification that the platform provides.
[ ] b. The project information the fund seeker provides.
[ ] c. The general information about the operating procedures of the platform (f.e. information about project screening.
[ ] d. Whether the platform has a licence of the AFM.
[ ] e. The performance of the projects that the platform provides (f.e. success ratios, default

rates)
[ ] f. Information about exit or secondary trading.
[ ] g. Information about costs and return.
[ ] h. Whether or not the project offers information about the performance during the investment period.
[ ] i. Other information

Demographics

What is your gender?
[ ] Male
[ ] Female

What is your age?
[...]

What is your profession?
[ ] a. Self-employed
[ ] b. Employee
[ ] c. Government employee
[ ] d. Disabled, not able to work
[ ] e. Unemployed
[ ] f. Retired
[ ] g. Student
[ ] h. Housewife
[ ] i. Other situation

What is your highest qualification? (according to the Dutch eduaction system)
[ ] a. "Lager (beroeps)onderwijs"
[ ] b. "Mavo/Mulo/Vmbo (theoretische of gemengde leerweg)"
[ ] c. "HAVO/VWO of Middelbaar beroepsonderwijs"
[ ] d. "HBO"
[ ] e. "WO"

Financial situation

What is the gross annual income of your household?
[ ] a. Minimum (less than EUR 12.500)
[ ] b. Below average (EUR 12.500 - EUR 33.000)
[ ] c. Average (EUR 33.000 - EUR 40.000)
[ ] d. Above average (EUR 40.000 - EUR 66.000)
[ ] e. Two times above average (EUR 66.000 - EUR 80.000)
[ ] f. More than EUR 80.000
[ ] g. Refusal

What is your household's total net worth? Excluding real estate.
[ ] a. Less than EUR 10.000
[ ] b. Between EUR 10.000 - EUR 25.000

[ ] c. Between EUR 25.000 - EUR 50.000
[ ] d. Between EUR 50.000 - EUR 80.000
[ ] e. Between EUR 80.000 - EUR 150.000
[ ] f. More than EUR 150.000
[ ] g. Refusal

Which of the following financial products do you have?
[ ] a. Mortgage
[ ] b. Creditcard
[ ] c. Savings account
[ ] d. Personal loan
[ ] e. Investment account
[ ] f. None
[ ] g. Other (f.e. real estate)

Could you report how much of your net worth you allocate to savings and investments?
[...] a. Savings
[...] b. Investments
[...] c. Crowdfunding
[...] d. Other

Which situation is most applicable?
[ ] a. I invest execution-only, and am not able to access a financial adviser.
[ ] b. I have access to a financial adviser, although I do not always exploit this opportunity.
[ ] c. I invest via mutual funds. I do not invest execution-only.
[ ] d. I do not invest.

Financial literacy

How large do you estimate the probability of default in crowdfunding projects?
[ ] a. Very large, more than 20% of the crowfunding projects fail
[ ] b. Large, between 7.5% - 20% of the crowdfunding projects fail
[ ] c. Small, between 2.5% - 7.5% of the crowdfunding projects fail
[ ] d. Very small, less than 2.5% of the crowdfunding projects fail

Suppose you invest in the following assets. Please report how large you estimate the risk that you lose (part of) the initial money.
Dutch government bonds
[ ] Very small [ ] Small [ ] Average [ ] Large [ ] Very large [ ] Don't know
Mutual fund in Dutch stocks
[ ] Very small [ ] Small [ ] Average [ ] Large [ ] Very large [ ] Don't know
Stocks that are AEX listed
[ ] Very small [ ] Small [ ] Average [ ] Large [ ] Very large [ ] Don't know
Crowdfundingprojects
[ ] Very small [ ] Small [ ] Average [ ] Large [ ] Very large [ ] Don't know
Turbo's on the AEX
[ ] Very small [ ] Small [ ] Average [ ] Large [ ] Very large [ ] Don't know

If the interest rate drops, what happens with the bond rates?
[ ] a. They increase
[ ] b. They decrease
[ ] c. Bond rates stay the same
[ ] d. Do not know
[ ] e. Refusal

Is the following statement true or false? Buying a company stock usually provides a safer return than a stock mutual fund.
[ ] a. True
[ ] b. False
[ ] c. Do not know
[ ] d. Refusal

What happens with the risk to lose money if an investor diversifies his investment over different assets?
[ ] a. The risk should increase
[ ] b. The risk should decrease
[ ] c. The risk stays the same
[ ] d. Do not know

C.2 Transaction data

Together with the AFM, I asked the participating platform to provide insights in the transaction data of the respondents. First, we asked three main questions about the assets of the investor.
[ ] a. Project name
[ ] b. Date of investing
[ ] c. Investment amount
Second, we asked the participating platform the information about all the projects that are available on the platform.
[ ] a. Project name
[ ] b. Return
[ ] c. Duration
[ ] d. Risk classification (if available)
[ ] e. Industry (if available)
[ ] f. Collateral (if available)
[ ] g. Total funding
[ ] g. Total lenders

Merging all the data, I create a highly valuable data set. The available information is not only useful for this thesis, but provide much more information about Dutch household and their behavior[24] I strongly recommend further researchers and regulators to follow this methodology to get insights in household portfolios.

[24]For example, I leave out the opportunity to use regressions with time-series.

D Variable description

Description of the variables used in this thesis.

Variables	Description
Investor characteristics	
Age	Age of the investor
Gender	Male/Female
Profession	Profession a)self-employed b)employee c)government employee d)disablement e)job-seeker f)retired g)student h)housewife i)other situation
Education	Highest qualification a)Lager(beroeps) onderwijs b)MAVO c)HAVO/VWO/MBO d)HBO e)WO
Income	Total gross annual income of household (1-6 scale)
Net worth	Total wealth of household (excluding real estate) (1-6 scale)
Feel-good investor	Dummy for specific investor segment (based on goals motivations and expectations)
Return-seeking investor	Dummy for specific investor segment (based on goals motivation and expectations)
High-stake investor	Dummy for specific investor segment (based on goals motivation and expectations)
Past literacy	Dummy = 1 if investor also invest execution-only (has already taken a literacy test)
Participating in stock market	Dummy = 1 if investor also invest execution-only (has already taken a literacy test)
Advice	Dummy = 1 if investor has access to a financial adviser
Interest in trading	Dummy = 1 if experienced investors dealt with investment limits
Risk tolerance	Categorical variable of being risk averse, risk seeking or risk neutral
Motivation 1	Dummy = 1 if general motivation is to support the fund seeker
Motivation 2	Dummy = 1 if motivation of last project is knowing the fund seeker
Motivation 1&2	Dummy = 1 if both motivation 1 is true and motivation 2 is true
Investment characteristics	
Assets (self_reported)	Total projects, current and completed projects of all platforms
Assets (platformdata)	Total projects, according to the transaction data of the specific platform
Only one platform	Dummy = 1 if investor participates on only one platform
Invested capital (self_reported)	Total capital invested, portfolio size in EUR
Invested capital (platformdata)	Total capital invested in the portfolio of the specific platform
DIV1	Diversification according to self reported assets in quintiles
DIV2	Diversification controlled for market portfolio
DIV3	Diversification controlled for market portfolio and total funding
SIR	Portfolio size to annual income ratio (approximation)
SNWR	Portfolio size to net worth ratio (approximation)
Incorrect capital estimation	Dummy for investor who did not accurately estimate their portfolio size (self-reported portfolio size did not corespond with actual transaction data)
Short investment period	Dummy if investor is less than 0.5 year active on platform
Other investments relative to NW	Self-reported percentage of other investments to net worth
Other investments (abs amount)	Approximation of (self-reported) other investment ratio* (self-reported) net worth in EUR
Financial literacy	
Dummy Q1	Dummy for incorrectly answering Q1 ”the inverse relation between the interest and bonds”
Dummy Q2	Dummy for incorrectly answering Q2 ”difference between risk of a mutual funds and a stock”
Dummy Q3	Dummy for incorrectly answering Q3 ”decreasing risks when the stocks in the portfolio increases”
Financial literacy	Categorical variable of having low literacy (0-1 correct answers), average literacy (2 correct answers) and high literacy (all answers correct)