View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Deutsches Institut für Wirtschaftsforschung

DIW BERLIN

Discussion Papers

1125

Nataliya Barasinska

Does Gender Affect Investors' Appetite for Risk? Evidence from Peer-to-Peer Lending

Berlin, May 2011

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2011

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Does Gender Affect Investors' Appetite for Risk? Evidence from Peer-to-Peer Lending

Nataliya Barasinska*

April 18, 2011

Abstract: This study investigates the role of gender in financial risk-taking. Specifically, I ask whether female investors tend to fund less risky investment projects than males. To answer this question, I use real-life investment data collected at the largest German market for peer-to-peer lending. Investors' utility is assumed to be a function of the projects expected return and its standard deviation, whereas standard deviation serves as a measure of risk. Gender differences regarding the responses to projects' risk are tested by estimating a random parameter regression model that allows for variation of risk preferences across investors. Estimation results provide *no* evidence of gender differences in investors' risk propensity: On average, male and female investors respond similarly to the changes in the standard deviation of expected return. Moreover, no differences between male and female investors are found with respect to other characteristics of projects that may serve as a proxy for projects' risk. Significant gender differences in investors' tastes are found only with respect to preferred investment duration, purpose of investment project and borrowers' age.

JEL: G11, G21, J16
Keywords: gender, investment choice, risk preferences

---

*I thank Dorothea Schäfer for an in-depth discussion of the paper and valuable comments. Financial support from the European Commission (7^{th} Framework Programme, Grant Agreement No. 217266) is gratefully acknowledged. Correspondence address: DIW Berlin, Mohrenstrasse 58, 10117 Berlin. Tel.: +49 30 897 89 691. Email: nbarasinska@diw.de. The usual disclaimer applies.

1 Introduction

The financial crisis of the early 21st Century triggered, among many other things, a heated public debate about the role of gender in the financial behavior of individuals.[1] One conjecture voiced in the debate is that excessive risk-taking in the financial markets is to be blamed on the prevalence of males in the decision-making positions in the financial industry. As Neelie Kroes, the EU competition commissioner, put it: *"... the collapse of Lehman Brothers would never have happened if there'd been Lehman Sisters with them."*[2] Such claims rely primarily on the popular gender stereotype that males seek greater risk and are overconfident in financial matters than females. An important question is whether gender stereotype reflects the true state of things. Does investor gender really affect risk-taking propensity? The literature investigating this question is extensive, however, no conclusive answer has been provided. So far, most evidence is based on household surveys or laboratory experiments. In contrast, direct evidence on real-life investment behavior is scarce and essentially limited to studies of professional investors.

This study contributes to literature by examining financial behavior of males and females using real-life data. The aim of the study is to test gender differences in the propensity for risk taking by retail investors who participate in a new segment of financial markets known as *peer-to-peer (p2p)* lending. Peer-to-peer lending means direct lending and borrowing between individuals ("peers") without intermediation of a traditional financial institution like a bank. The data are collected from the largest German p2p marketplace *Smava.de*. In this marketplace, individuals lend funds for a variety of purposes ranging from large consumer expenditures to small business investments. The loans are neither collateralized nor guaranteed and lenders can incur losses if borrowers default. Hence, p2p-lenders can be seen as investors who fund risky projects.

Relying on the μ-σ approach, I assume that utility attached by investors to a risky project depends on the project's expected return μ and its standard deviation σ. The more risk averse an investor is, the more his/her utility decreases in response to a small increase in σ. This relationship serves as a basis for the test of gender differences in risk propensity. The aim of test is to answer the question: *Are female investors participating in the German p2p-lending more risk-averse than male investors?* If female investors indeed exhibit higher risk aversion than male investors, their utility will decrease more than the utility of males in response to a marginal increase in return's standard deviation, *ceteris paribus*. Inference about the effect of σ on utility is derived from investors' actual choices.

Advantages of using the p2p-lending data for the analysis are threefold. Firstly, all participating investors are exposed to the same market-related factors: there is only one type of financial product, same investment rules apply for every one and all participants have access to the same information. Therefore, it can be argued that differences in the observed investment choices stem exclusively from investor-related factors. Secondly, a complete history of investment choices of each participant including the characteristics of the investment alternatives is observable. Thirdly, investors' gender is observable to researcher. All these features make p2p data well suited for a study of gender effects on the propensity for risk-taking of retail investors.

[1] See e.g. Economist (2009), Bennhold (2009) and Oakeshott (2009)

[2] Indeed, all members of the executive management at *Lehman Brothers* at the time of the collapse were male. The bank is not an exception: The three German banks – Deutsche Bank, Kommerzbank and HypoVereinsbank – have all-male executive management teams.

Estimation of investors' responses to the riskiness of investment projects relies on mixed logit regression – a qualitative choice model that accommodates repeated choice data. Repeated choice arises because during the observation period majority of investors conducted more than one investment. This advantageous feature of the data eliminates problems stemming from the fact that not all investor-specific factors are observable to researcher.

Results of regression analysis provide *no* evidence of gender differences in investors' risk propensity: On average, male and female investors respond similarly to changes in projects mean-variance profile. Moreover, no differences between male and female investors are found with respect to other characteristics of investment projects that may serve as proxy for projects' riskiness. Significant gender differences in investor taste are found only with respect to preferred investment duration, purpose of investment project and borrower age.

The remainder of the paper is organized as follows. In Section 2, I review studies examining the role of gender in individuals' propensity for risk taking in financial decisions. Information about p2p-credit markets and lending mechanism at *Smava.de* is provided in Section 3. In Section 4, I formulate the research hypothesis. Section 5 is devoted to empirical analysis. Here, I firstly describe the econometric model and the data employed to test the research hypothesis. Then, I report and discuss the main estimation results. The last section concludes.

2 Literature Review

Academic research on the role of gender in the financial behavior of individuals has a long history. Nonetheless, the question regarding the effect of gender on the propensity for risk-taking remains unanswered.

A large group of studies, especially those that analyze financial behavior of individuals in the population at large, suggest that females are on average more risk averse than males, *ceteris paribus*, (Jianakoplos and Bernasek, 1998; Sunden and Surette, 1998; Bernasek and Shwiff, 2001). However, these studies rely on household survey data providing only general information about investments, while such important parameters as expected return, risk or transaction costs are not known. Hence, the level of risk taken by an individual investor cannot be measured exactly. Moreover, in the most survey-based data, financial assets are aggregated at household level making it difficult to identify who is actually responsible for an investment decision in a multi-person household.

A few empirical studies try to overcome these limitations by focusing on professionally trained investors, mostly managers of investment funds, who take risky financial decisions in the course of their jobs. Intuition suggests that males and females who deliberately and actively engage in risky financial activity and have the same professional training should, on average, exhibit similar risk propensity. This should hold even when in population at large females are found to be less risk tolerant than males. Nonetheless, studies of behavior of professional investors provide mixed evidence. Johnson and Powell (1994) and Atkinson et al. (2003) find no differences in the behavior of male and female managers. In contrast, Olsen and Cox (2001), Beckmann and Menkhoff (2008) and Niessen and Ruenzi (2007) show that female managers follow less risky investment styles than their male counterparts. Noteworthy, the latter group of studies has one methodological feature in common. The studied funds are very heterogenous ranging from pure bond-funds to pure equity-funds so that the sampled individuals work in very different settings and face different investment tasks. However, this may preclude unbiased evidence on individual-specific factors of investment decisions.

So far, a careful control over the factors related to investment task could only be assured in laboratory experiments. A number of experimental studies investigate gender differences in risk preferences in objective probability lotteries with both real and hypothetical outcomes (Powell and Ansic, 1997; Schubert et al., 1999; Holt and Laury, 2002; Dohmen et al., 2005; Fehr-Duda and Schubert, 2006; Eckel and Grossman, 2008).[3] Although a majority of the studies confirm the gender stereotype, there are some notable exceptions. For instance, Schubert et al. (1999) find that risk propensity of males and females depends strongly on whether experiments involve abstract gambles or contextually framed lotteries. In the latter setting females and males do not exhibit significant differences in risk propensity. Interesting evidence is provided by Holt and Laury (2002) who show that the effect of gender varies with the level of payoff. Females behave more risk averse than males when lotteries involve low payoffs. However, when lotteries involve high payoffs, no differences between males and females are documented. Thus, experimental evidence on gender differences should be enjoyed carefully as gender differences in financial behavior seems to be sensitive to contextual framing and to the level of payoffs.

This study contributes to the existing literature in several important ways. First, it complements experimental evidence by resolving the concerns regarding the consistency of behavior in a laboratory with behavior in real life. Furthermore, unlike most studies based on observational data, the study analyzes risk-taking in a situation where all investors make decisions about the same type of investment product. Finally, the study provides rare evidence on the behavior of retail investors with detailed information about investments' characteristics available.

3 German Market for Peer-to-Peer Lending *Smava*

3.1 What is Peer-to-Peer Lending?

The term "peer-to-peer lending" refers to direct lending between private persons without intermediation of traditional financial institutions like banks. Classical examples of p2p loans are loans granted among friends or family members. The novelty of the modern p2p lending is the emergence of internet-based marketplaces (so called "platforms") where funds are transferred from surplus and deficit agents and the agents do not know each other personally. The surplus agents, i.e. lenders, provide funds with interest. The deficit agents, i.e. borrowers, are contractually bound to repay the principal and the interest. They can, however, default on their debt obligations and inflict losses on lenders.

The first p2p platform, *Zopa*, was founded in 2005 in the UK. Since then, more than 30 independent market places started in the USA and continental Europe. Currently, the total amount of p2p loans originated by the largest platforms in the USA and Europe – Prosper, Lending Club, Zopa, Smava and Auxmoney – amounts to €600 million.[4] Compared to the volume of the traditional consumer credit market, peer-to-peer lending is still a niche prod-

[3] A concise overview of these studies is provided by Croson and Gneezy (2009).
[4] Own calculations of the author based on official reports of the four platforms.

uct. Nevertheless, its phenomenon attracts significant attention of general public, financial industry professionals and academics.[5]

3.2 How does *Smava* function?

This study focuses on the largest German p2p platform *Smava.de*. The platform was launched in March 2007. By the end of March 2010, a total of 4,148 loan applications were posted on the platform. This leads to a total volume of ca. € 25 million, the result of 3,354 signed loan contracts (Figure 3)[6] The average amount of loan is approximately € 8 thousand.

The market functions in the following way. Individuals who want to invest or borrow on the platform must register and prove their identity. Investing is allowed to private individuals who are at least 18 years old and residents of Germany. Borrowing is allowed to private persons who comply with a range of requirements. First, applicants must be at least 18 years old and have a monthly income of at least € 1,000. Secondly, only those whose individual financial burden does not exceed 67 % are eligible to borrow at the platform. Financial burden is measured as a ratio of monthly payments on all outstanding consumer debts (including loans taken at *Smava*) to the borrower's personal monthly disposable income. Mortgage payments are treated as expenditures and subtracted from the disposable income. Income by other household members, as well as household savings, are not taken into account. Depending on the obtained ratio, borrowers are rated on a scale from 1 to 4 and assigned the so called KDF-indicator as described in Table 2. Finally, the platform accepts only applicants with credit scores ranging from A to H. This rating, commonly referred to as a "Schufa-rating", is assigned to individuals by Schufa, the German national credit bureau, and measures individual's creditworthiness on a 12-point scale from A (the best) to M (the worst). Each rating score corresponds to an estimate of probability that a borrower defaults on his obligations (see Table 3). Applicants' identity is verified via *postident* procedure, a procedure through which individuals prove their identity through verification procedures carried out by employees of Deutsche Post at their local post office. The verified identity is not revealed to other market participants; instead both investors and borrowers operate at the platform under usernames.

After successful registration, borrowers post loan applications on the platform's web page. A loan application specifies the amount of money the applicant wants to borrow, for how long and what nominal annual interest rate he or she is willing to pay. Two restrictions are imposed by the platform on loan applications: the requested loan amount must be between € 500 and € 50,000; and the loan duration must be either 36 or 60 months. In addition, applicants may provide a description of the loan purpose, of their own personality and upload a picture. These additional pieces of information are provided voluntarily and are not verified by the platform.

Investors can browse through the applications and choose which borrower they want to finance. When an investor decides to provide funds to a particular borrower, he or she submits an electronic order. By submitting the order an investor "signs" a binding contract in which he/she commits to provide certain amount of money to the chosen borrower. The

[5] For the general information see e.g. FTD (2009), Sviokla (2009) and Kim (2009); on financial industry analysis see Meyer (2009); and on academic research see Pope and Sydnor (2008), Freedman and Jin (2008), Garman et al. (2008) and Duarte et al. (2009).

[6] Own calculations of the author.

minimum acceptable order is € 250, the maximum is € 25,000. All orders must be multiples of 250. Often several investors submit offers to the same loan and each provides a fraction of the amount requested in the application. The number of investors tends to increase with the size of requested loan. So far, the average number of investors per loan was 15 and the median order is € 250.

An important distinguishing feature of *Smava.de* is that loans are *not* auctioned. In contrast to many other peer-to-peer lending sites, orders at this platform are accepted on the "first-come, first-served" basis, i.e. until the requested loan amount is covered to 100%. Investors cannot underbid offers from other investors by offering money at a lower interest rate. Money can only be provided under the terms specified in loan applications, i.e. under the interest rate and for the duration set by applicants.

Each application remains open for orders during 14 days, starting with the day when it was posted. If after this period less than 25% of the requested amount is raised, the application is canceled and the raised money (if any raised) is returned to investors.[7] The applicant can post the application again, eventually, offering more attractive conditions, e.g. a higher interest rate. In case of a successful brokerage, the platform charges investors with € 4 per order. Borrowers' fee depends on loan maturity and is 2 % of the borrowed sum (or at least € 40) when the loan is due in 36 months and 2.5 % of the borrowed sum (or at least € 60) if the loan matures in 60 months.[8]

Loans procured at the platform are installment credits that are not collateralized or guaranteed by third parties. Borrowers are only contractually bound to repay the debt and the interest in fixed monthly payments. To safeguard the investors from total loss, the platform utilizes two risk-reducing instruments. These instruments are described in more detail in the following sections of the paper.

3.3 What Information Do Investors Have?

Investing at the platform is characterized by substantial informational asymmetries between investors and borrowers. The asymmetries emerge mainly because borrowers' identity is not known and investors are provided with a limited set of information about the borrowers. Investors have access only to information that is collected and disclosed by the platform. Hence, ultimately the decisions of investors are built upon the provided information set.

Loan specific information observable to investors comprises the following details. Investors can observe in real time when a loan request is posted, what bids are submitted by the other investors (if any), when the submissions were made, and what share of the requested sum remains unfunded. Investors can also see the loan conditions set by borrowers: nominal annual interest rate, loan amount and maturity. Further, borrowers have to specify the purpose of loan by choosing an item from a menu of 17 categories. Figure 4 plots the distribution of applications over the categories. In addition to specifying the loan purpose, borrowers can also provide a relatively detailed description of the projects they need money for. This additional information should increase borrowers' trustworthiness and reduce informational asymmetries between the parties. However, the description of loan purpose is voluntarily and is not always provided.

---

[7] About 8% of loan applications in the data set did not raise any money; 5% raised less than 25% of requested amount; 6% raised $\geq$ 25% but less than 100%; 81% managed to raise 100% of requested amount.

[8] Smava changed the terms of the platform several times, but no changes were made during the time period under observation.

The borrower-specific information observed by investors can be subdivided into "hard" and "soft" information. Hard information includes verified data that each borrower is obliged to provide. The data set comprises borrowers' age, sex, employment status, place of residence, credit rating, debt burden measured as debt-to-income ratio, number of delayed payments and defaults on previous *Smava* loans. Availability of hard information is crucial for investors, because it allows them to estimate the expected rate of return on investments and the probability of the borrower defaulting.

Although all pieces of hard information are verified, informational imperfections are still high. In particular, the platform provides only a rough estimate of borrowers' personal financial burden. The actual income and savings are not observable. Furthermore, nothing is known about the income and wealth of other household members. The available "hard" information is complemented by "soft" information. The latter is voluntarily provided by borrowers and is not verifiable. The "soft" data may include information on borrowers' education, hobbies, family status etc.

3.4 What Risks Do Investors Face?

Loans procured at the platform are neither secured by collateral nor guaranteed by third parties. Hence, investors can incur a loss if borrowers default on their obligations. To prevent total losses, the platform uses two instruments. Firstly, in case of default the claim to outstanding debt is sold to a collecting agency. Between 15 and 20 percent of invested capital can be recovered in this way. Secondly, a significantly larger part of capital can be recovered due to a risk sharing mechanism via loan pools.

Risk sharing via pools is accomplished by assigning investors into groups. Specifically, all investors who finance loans of the same duration and rating are assigned into one group. For example, all investors who granted loans to borrowers with rating "A" for 60 months belong to the same pool. Due to existence of 8 rating classes and 2 durations, there are 16 pools in total. Monthly redemption payments done by borrowers of the same pool are lumped together and each investor gets an amount proportional to his/her investment. Interest payments are not pooled together but transferred directly to investors. When some loans from the pool default, the losses are subtracted from the pool and the remainder is then divided among all members of the pool proportionally to their investments. In effect, all members of the pool including those who actually invested in the defaulted loan get a fraction of the usual monthly payment. This faction is called the *pool's payment rate*. For, example there are 100 investors in a pool and each granted a € 250-loan to different borrowers. If two loans get default, the pool's payment rate reduces to 98% which means that every member of the pool gets only 98% of the stipulated redemption payment. If another loan defaults, the pool's payment rate decreases to 97% and so on. The payment rate can, however, be improved when members of a pool invest in new loans of the same duration and rating and the old defaulted loans reach their maturity. The platform provides investors with a prediction of average payment rate for each pool (see Table 4). The described risk sharing mechanism assures that affected investors do not lose 100% of the invested capital. The flip side of the coin is that the losses are covered by withholding a part of cash inflows from the unaffected investors and, hence, reducing their profits.

Loans that are repaid prior to maturity present another source of risk. When a loan is repaid early, investors loose a part of expected interest payments. There is no penalty for early payments and hence investors get no compensation for the foregone interest. A further source of risk is associated with delayed payments. A delayed payment ties up the money and prevents investors from reinvesting it in new projects. Because no penalty for delayed payments is imposed on borrowers, lenders are not compensated for postponed reinvesting. Hence, delayed payments inflict losses in the form of foregone investment opportunities.

4 Research Hypothesis

The goal of the paper is to answer the question: Do females investing in p2p loans exhibit higher risk aversion than males? To answer this question, I analyze the choices of male and female investors.

At the considered market, the set of investment alternatives faced by investors is comprised of loans requested by loan applicants. In the following, I refer to loans as investment projects. An investor ranks his/her preferences over all available investment project depending on how much utility he/she expects to obtain from each project. Specifically, I assume that investors have a two-parameter utility function $U(\mu, \sigma)$. That is, utility attached by an investor to a project depends on a linear combination of the project's expected return μ and its standard deviation σ. Thus, investors rank their preferences over different projects depending on the utility expected from them. If investors are rational, they choose to fund projects yielding the greatest utility. Hence, investor decision problem can be specified as choosing the projects with such combination of μ and σ that maximizes investor utility.

Under these assumptions, investors' propensity for risk-taking can be measured relying on the μ-σ approach. [9] The intuition behind the μ-σ approach is that investors trade-off between the expected return and its standard deviation whereas the latter represents risk. Investors like return and place a positive weight on μ so that $U(\cdot)$ increases in μ. Investors' attitude towards σ depends on the investors' individual risk preferences. Let γ denote a constant reflecting investors' risk preferences. Risk preferences can be explicitly included into investors' utility function: $U(\mu, \gamma\sigma)$. Then the marginal effect of σ on the utility is given by γ and varies across investors with different risk preferences. Specifically, if investors are risk-averse, γ is negative and $U(\cdot)$ decreases in σ. Moreover, the larger the weight the larger the decrease in the utility. For risk-neutral investors, γ is zero and $U(\cdot)$ does not vary with σ. For risk-loving investors, γ is positive, so that $U(\cdot)$ increases in σ and the larger the γ the larger the increase.

This relationship between risk preferences and the marginal effect of σ on the utility, provides the basis for the test of differences in risk preferences between male and female investors. Specifically, if females investing on the considered platform are on average more risk averse (or less risk loving) than males, then γ should systematically differ between males and females, *ceteris paribus*. Let γ_m and γ_f denote respectively the risk preferences of male and female investors. Different risk preferences of males and females mean that $\gamma_m \neq \gamma_f$. To see what implications does this difference have, consider the following five situations:

[9] μ-σ approach is frequently criticized for its restrictive assumptions regarding the functional form of utility (Meyer, 1987; Bigelow, 1993) or distribution of returns (Chamberlain, 1983). However, in contrast to situations where mixtures of distributions are considered, in situations where preferences are to be ordered over a set of simple distributions (as is the case in this study), μ-σ approach can be employed under less restrictive assumptions (Meyer and Rasche, 1992).

Another restrictive property of μ-σ approach is its assumption that investors derive utility only from monetary payoffs of investment projects. However, recent studies show that individuals attach significant value to social returns of an investment (see e.g. Bollen, 2007; Benson and Humphrey, 2008). This circumstance is accounted for in the empirical part of the paper.

Situation 1. Both male and female investors are on average risk averse, whereas females are on average more risk averse than males: $\gamma_m,\ \gamma_f < 0$ and $|\gamma_m| < |\gamma_f|$. Then, an increase in σ has a negative effect on the utility of both gender groups. However, the utility of females decreases more than the utility of males when σ increases. Hence, the difference between the marginal effects of σ for females and males is negative:

$$\frac{\partial U_{Female}}{\partial \sigma} < \frac{\partial U_{Male}}{\partial \sigma} < 0, \quad \frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0.$$

Situation 2. Female investors are on average risk averse, whereas males are on average risk neutral: $\gamma_f < 0,\ \gamma_m = 0$. Then, an increase in σ has a negative effect on the utility of female investors but no effect on the utility of male investors. Respectively, the difference between the marginal effects of σ for females and males should be negative:

$$\frac{\partial U_{Female}}{\partial \sigma} < 0, \quad \frac{\partial U_{Male}}{\partial \sigma} = 0, \quad \frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0$$

Situation 3. Female investors are on average risk neutral, whereas males are on average risk loving: $\gamma_f = 0,\ \gamma_m > 0$. Then, an increase in σ has no effect on the utility of female investors but has a positive effect on the utility of male investors. Respectively, the difference between the marginal effects of σ for females and males should be negative:

$$\frac{\partial U_{Female}}{\partial \sigma} = 0, \quad \frac{\partial U_{Male}}{\partial \sigma} > 0, \quad \frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0$$

Situation 4. Both male and female investors are on average risk loving, whereas females are on average less risk loving than males: $0 < \gamma_f < \gamma_m$. Then, an increase in σ has a positive effect on the utility of both gender groups. However, the utility of females increases less than the utility of males. Hence, the difference between the marginal effects of σ for females and males should be negative:

$$0 < \frac{\partial U_{Female}}{\partial \sigma} < \frac{\partial U_{Male}}{\partial \sigma}, \quad \frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0.$$

Situation 5. Female investors are on average risk averse, whereas males are on average risk loving: $\gamma_f < 0,\ \gamma_m > 0$. Then, an increase in σ has a negative effect on the utility of female investors but a positive effect on the utility of male investors. Respectively, the difference between the marginal effects of σ for females and males should be negative:

$$\frac{\partial U_{Female}}{\partial \sigma} < 0, \quad \frac{\partial U_{Male}}{\partial \sigma} > 0, \quad \frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0$$

Thus, difference between the marginal effects of σ on the utility of females and males is negative in any situation where females are either more risk averse or less risk loving than males. Similarly it can be shown that the marginal effect of σ on utility is the same for both genders when males and females are, on average, equally risk prone. Furthermore, the difference in the marginal effects between males and females is positive when females are more risk prone than males. So, to answer the research question whether females investing in the considered market are on average more risk averse (or less risk loving) than males, the following hypothesis has to be tested:

Hypothesis: *Ceteris paribus, the difference between the marginal effect of σ on the utility of a female investor and the marginal effect of σ on the utility of a male investor is negative,*

$$\frac{\partial U_{Female}}{\partial \sigma} - \frac{\partial U_{Male}}{\partial \sigma} < 0.$$

Hence, gender differences in risk propensity can be tested by estimating the marginal effect of one standard deviation of a project's expected return on the utility of investors. Inference about the utility attached by investors to different projects can made based on the observed investment choices. The empirical test is described in the remainder of the paper.

5 Implementation of the Test

5.1 Econometric Model

Let J_n^t denote the set of investment alternatives faced by investor n in choice situation $t \in T_n$. J_n^t comprises all investment projects that are available at the market at time t, when investor n submits his/her order on one of the projects. The utility that investor n attaches to investment project $j \in J_n^t$ can be decomposed in a deterministic part $\boldsymbol{\beta}_n' \mathbf{x}_{njt}$ which is a linear combination of the project's characteristics observable to researcher and an unobserved part, ε_{njt}:

$$U_{njt} = \boldsymbol{\beta}_n' \mathbf{x}_{njt} + \varepsilon_{njt}, \tag{1}$$

where $\mathbf{x}_{njt}$ is a K-dimensional vector of the characteristics of investment project j. The main characteristics of a project are the expected return and its standard deviation. Besides them, each project is characterized by a number attributes summarized in Table 6. $\boldsymbol{\beta}_n$ is a vector of parameters reflecting investor's n valuation of (or taste for) each attribute $k \in K$. ε_{njt} is a stochastic term representing the random part of utility; it is *iid* over investors and choice situations. It is assumed that investor preference is completely defined by the projects' attributes, that is, utility is derived from the attributes associated with investment projects rather than from projects *per se*. In line with this assumption, Equation 1 has no alternative-specific constants.

Vector $\boldsymbol{\beta}_n$ is explicitly allowed to vary over individuals. I assume that $\boldsymbol{\beta}_n$ is normally distributed with mean $\boldsymbol{b}$ and standard deviation $\boldsymbol{\sigma}_\beta$: $\boldsymbol{\beta}_n \sim N(\boldsymbol{b}, \boldsymbol{\sigma}_\beta)$.[10] This feature reflects the possibility that there is taste variation in the population and any given attribute of an investment project may receive different valuation from different investors. For example, utility derived from an investment project with a given expected return and standard deviation should vary over individuals depending on their risk preferences. However, preferences are not observed. Therefore, the model should accommodate *random* taste heterogeneity emerging due to unobserved investor-specific factors. Furthermore, a part of taste variation may also stem from observable differences among individuals such as, for example, age, income or gender. This kind of taste heterogeneity is *systematic* and can be explicitly modeled by taking investors' characteristics into account. Due to the research aim of this paper, I only focus on how valuation of projects' attributes depends on investor gender.

The two types of taste heterogeneity – random and systematic – are incorporated into Equation 1 by expressing vector $\boldsymbol{\beta}_n$ as a function of investors' gender and the unobserved individual-specific effects:

$$\boldsymbol{\beta}_n = \boldsymbol{b} + \boldsymbol{\gamma} Female_{njt} + \boldsymbol{\eta}_n,$$

where vector $\boldsymbol{b}$ has k-elements each representing the average valuation placed by male investors on project attribute $k \in K$. $Female_{njt}$ is a dummy variable equal 1 if investor is female and 0 if male.

[10] I assume that coefficients of corresponding to different projects' attributes are not correlated. That is, the off-diagonal elements of matrix $\boldsymbol{\sigma}_\beta^2$ are zero.

Vector $\boldsymbol{\gamma}$ has K-elements each capturing the difference between the average effect of project attribute k on the utility of a female investor and the marginal effect of project attribute k on the utility a male investor. For instance, $\gamma_{SD[Return]}$ is one of the elements of $\boldsymbol{\gamma}$ that shows the difference between the effect of returns' standard deviation on the utility of females and the effect on the utility of males. With respect to the research hypothesis, $\gamma_{SD[Return]}$ is of central interest. A negative and statistically significant estimate $\widehat{\gamma_{SD[Return]}}$ means that females are more risk averse (or less risk tolerant) than males.

$\boldsymbol{\eta}_n$ is a K-dimensional vector with elements represents the effect of unobserved factors associated with investor n on his/her valuation of the project's attributes. Technically, $\boldsymbol{\eta}_n$ is a deviation of β_n from its mean: $\boldsymbol{\eta}_n = \boldsymbol{\beta}_n - \boldsymbol{b}$. Therefore, it is by construction normally distributed with zero mean and standard deviation $\boldsymbol{\sigma}_\beta$. $\boldsymbol{\eta}_n$ is allowed to vary across investors but is assumed to be constant over choice situations for a given investor.

After specification of taste heterogeneity, Equation 1 can be rewritten as

$$U_{njt} = \boldsymbol{b}'\mathbf{x}_{njt} + \boldsymbol{\gamma} Female_{njt}\mathbf{x}_{njt} + \boldsymbol{\eta}_n\mathbf{x}_{njt} + \varepsilon_{njt}. \tag{2}$$

Now, the random portion of utility consists of $\boldsymbol{\eta}'_n\mathbf{x}_{njt} + \varepsilon_{njt}$. Due to the common effect of $\boldsymbol{\eta}_n$, the random term is correlated over investment alternatives and choice situations for a given investor.

So far, the equation describing investor choice is specified so that expected utility entered the equation as a dependent variable. Yet, expected utility of an investor is his/her private information that is not observable to a researcher. What is observed is the choice set faced by an investor and the actual choice made. Assuming that investors are utility maximizers, it can be argued that the chosen project provides an investor with the greatest expected utility. Therefore, inference about factors affecting an investor's utility can be made by analyzing the relationship between observable attributes of investment alternatives and the investor's choice. Such analysis can be done by estimating a discrete choice model (Train, 2009).

Consider a data set where unit of observation is an investment project. Each time an investor makes an investment, he/she contributes $N_{nt} = J_n^t$ observations to the data set, whereby J_n^t is the number of projects entering the choice set of investor n in choice situation t. Now, define a new binary variable y_{njt} as follows

$$y_{njt} \begin{cases} = 1 & \text{if project } j \text{ is chosen by investor } n \text{ in situation } t \\ = 0 & \text{if project } j \text{ is not chosen} \end{cases}$$

The probability that investor n chooses project j in choice situation t given projects' attributes is

$$Pr[y_{njt} = 1] \quad = \quad Pr[U_{njt} > U_{nit}, \, \forall\, j \neq i]$$

Brownstone and Train (1998) show that in the case when coefficient vector $\boldsymbol{\beta}_n$ entering the utility equation is randomly distributed with parameters $\boldsymbol{b}$ and $\boldsymbol{\sigma}_\beta$, the choice probability becomes

$$Pr[y_{njt} = 1] \quad = \int L_{njt}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n | \boldsymbol{b},\, \boldsymbol{\sigma}_\beta)) d(\boldsymbol{\beta}_n)$$

where $L_{njt}(\boldsymbol{\beta}_n)$ is given by a standard logit:

$$L_{njt} \quad = \frac{exp(\boldsymbol{b}'\mathbf{x}_{njt} + \boldsymbol{\gamma} Female_{njt}\mathbf{x}_{njt} + \boldsymbol{\eta}_n\mathbf{x}_{njt})}{\sum_i exp(\boldsymbol{b}'\mathbf{x}_{nit} + \boldsymbol{\gamma} Female_{njt}\mathbf{x}_{nit} + \boldsymbol{\eta}_n\mathbf{x}_{nit})}$$

Revelt and Train (1998) extend the model to situation where researcher observes repeated choices for a given decision-maker. Specifically, they show that the probability of a sequence of choices made by an individual is given by

$$Pr[y_{nj} = 1] \quad = \int \prod^{t} L_{njt}(\boldsymbol{\beta}_n) f(\boldsymbol{\beta}_n | \boldsymbol{b}, \boldsymbol{\sigma}_{\beta})) d(\boldsymbol{\beta}_n) \tag{3}$$

Models of this form are known in the literature as mixed logit (Train, 2009). As shown by McFadden and Train (2000) mixed logit models present a very flexible type of discrete choice models that allows efficient estimation of the parameters $\boldsymbol{b}$ and $\boldsymbol{\sigma}_{\beta}$ by means of maximum simulated likelihood.[11]

5.2 The Data Set

Data used to estimate Model 3 are collected from the publicly available electronic archives of *Smava.de*. The data set contains observations on the electronic orders submitted by investors between March 2007 and March 2010. The number of investors registered at the end of observation period was 5,671. The total number of submitted orders is 54,455. On average, each investor submitted 10 orders, meaning that on average each investor made a choice in 10 choice situations (the median is 4, the maximum is 292). In each choice situation, investors faced an average of 17 different investment projects (the median number of alternatives is 13, minimum is 1 and maximum is 84). Figure 8 plots the distribution of choice sets over the number of alternatives entering them.

The majority of investors participating on the platform are male. There are only 625 female investors, 11% of all registered investors.[12] Summary statistics in Table 5 reveal some differences in the profiles of male and female investors. Males started investing at the p2p market 1 month earlier than females and hence can be said to be somewhat experienced than females. Female investors are, on average, 4 years older than male investors. The average amount invested per loan and the total amount invested at the platform by female investors is somewhat smaller than the respective amounts invested by male investors. However, the difference is statistically not significant.

For each submitted order the data includes information about the chosen loan application and the other applications entering the choice set of each investor. Attributes of loan applications that enter vector $\mathbf{x}_{njt}$ in Equation 2 are captured in the following variables. *Amount* is a continuous variable showing how much money a borrower requested in the application. Since the amount is always a multiple of 250 the variable is scaled by factor $\frac{1}{250}$ when used in regression analysis. *Duration* is a dummy variable equal 1 if loan is

[11] Compared to other discrete choice models such as multinomial logit or probit models, mixed logit models exhibit a number of useful properties. For instance, in contrast to mulninomial logit, mixed logit accommodates temporal correlation in error terms and relaxes the restrictive property of independence from irrelevant alternatives (IIA) (Train, 2009). Vis-a-vis multinomial probit model estimation of mixed logit is computationally less demanding. Numerical methods of integration currently used for probit models (for instance, Gaussian quadrature) operate effectively only when the number of alternatives times the number of choice situations is no more than four or five (Train, 2009). Yet, the dimension of the data in hand is much higher. The number of choice situations alone amounts on average to 84, while the number of alternatives in a choice set is on average 17.

[12] The predominance of male investors at the platform suggests that some kind of self-selection is taking place. Unfortunately, the data do not allow modeling the selection mechanism and to identify what factors determine the participation decision. Previous research shows that women are usually under-represented in the financial markets. For instance, only 10% of managers in the investment fund industry are females (Beckmann and Menkhoff, 2008; Niessen and Ruenzi, 2007). Moreover, considering the financial markets at large, females are found to be less likely to invest in risky financial assets (Badunenko et al., 2009).

asked for 60 month and 0 if for 36 months. *Offered interest rate* is a continuous variable describing showing the nominal annual interest rate in % offered by a borrower. *Purpose* is a dummy variable equal 1 if a loan is taken for business purposes and 0 if for consumer purpose. *Description* is a continuous variable measuring the length of description of loan made by a borrower. This variable is equal to a logarithm of the number of characters used in the description. *Female* is a dummy variable describing borrowers' gender. It is equal 1 borrower is female and 0 if male. *Age* is a continuous variable showing the age of borrower when he/she posted the application. Variable *Rating* takes on 8 values from "A" (the best creditworthiness) to H (the worst creditworthiness) and measures the creditworthiness of borrowers according to the scale of the German credit agency, Schufa. Dummy variable *Financial burden: low* is equal 1 if borrower's debt-to-income ratio does not exceed 20%. Dummy variable *Financial burden: moderate* equals 1 if debt-to-income ratio lies within the range 20-40% and 0 otherwise. Dummy variable *Financial burden: substantial* equals 1 if debt-to-income ratio lies within the range 40-60%. Dummy variable *Financial burden: high* equals 1 if debt-to-income ratio lies within the range 60-67%. *Employment* is a dummy variable indicating borrowers' employment status. It is equal 1 if borrower is self-employed, and 0 if borrower is either employed or retired.

Information about projects' expected return and variance of returns is not provided to either investors or researchers. Both must calculate these attributes individually. Calculation of expected return and its standard deviation, as applied in this study, is described in the next section.

5.3 Calculation of expected return and its variance

Assuming that the uncertainty pertaining to the payoff of an annuity loan stems only from the probability that a borrower defaults,[13] then investing in an annuity loan can be seen as buying a lottery with $M+1$ possible outcomes where M equals to the number of monthly installments that a borrower is contractually obliged to pay in order to repay the loan. Depending on when a borrower defaults, the number of actually paid installments can vary between 0 (no payments made) and M (all payments completed). Realization of any of $M+1$ outcomes determines what rate of return to investment is obtained. The rate of return, conditional on realization of an outcome, is denoted by R_m.

Probability of each outcome of the lottery is determined by the probability that a borrower defaults and does not pay back a number of installments. Let $T = \{1,2,...,M\}$ be a discrete random variable indicating the sequential number of installment at which a default occurred, i.e. neither the installment in question nor any of the subsequent installments are paid. Let $f(t)$ denote the probability distribution function of T. Then, probability of default occurring with installment t is $Pr(T=t) = f(t)$. The probability distribution function $f(t)$ is not known. However, it can be estimated based on the payment behavior of borrowers observed in the past. In particular, it is helpful to estimate how probability of default with any given installment depends on the observable characteristics of borrowers and loan terms. Procedure used to estimate the probabilities is described in Appendix A. Based on estimated default probabilities, one obtains estimates of the probability of each outcome for any given loan, $\hat{p}_1,...,\hat{p}_{M+1}$.

[13]There are other sources of uncertainty such as the probability of early repayment of a loan or changes in the payment rates of pools. However, the present analysis does not take these into account.

Figure 5 illustrates the possible outcomes and the respective probabilities for a loan with duration 36 months. The duration of 36 months implies that 36 a borrower must pay 36 installments. Respectively, there are 37 possible outcomes. Let R_1 denote the rate of return received by investor if the first outcome is realized. The first outcome is realized if borrower does not pay any installments. The probability of this outcome, p_1, is the probability that default occurs with the first installment, $Pr(T = 1) = f(2)$. The second outcome is realized if borrower pays the first installment but defaults with the second installment. This outcome occurs with probability $p_2 = Pr(T = 2) = f(2)$. And so on. Finally, the last possible outcome emerges if borrower makes all payments, i.e. does not default on any of the installments. The probability of this event $p_{37} = Pr(T \geq 36) = 1 - f(36)$.

The next step is to determine the rate of return, R_m, generated in case of each outcome. Return to an annuity loan can be determined by calculating the internal rate of return from a series of cash flows produced by the loan. Similar to a common annuity loan, cash flow generated by a *Smava*-loan is given by a series of monthly installments paid by borrowers whereby each installment consists of debt redemption and interest on the outstanding debt. With *Smava* loans, even in case of a borrower default, investors receive some money back due to the collective insurance mechanism described in Section 3. Investors always get a fraction of the contractually stipulated redemption regardless of whether a borrower defaults or not. This fraction is determined by the payment rate of the pool the investor belongs to, P_p.[14] Interest is exempt from the insurance mechanism, such that investors do not get any of the contractually stipulated interest if their borrowers default.

Hence, amount A_t received by an investor at the t-th month of a loan duration is

$$A_t = \begin{cases} P_p \times D_t + I_t, & \forall t < T \\ P_p \times D_t, & \forall\, t \geq T. \end{cases}$$

where D_t is the value of contractually stipulated redemption in month t, P_p is the repayment rate of pool p where investor belongs to, I_t is the contractually stipulated interest in month t, and T is the installment at which a default occurred.

Then, return R_m generated by a loan if outcome m is realized is obtained by solving for r in

$$\begin{array}{lcl} Investment{+}Fee = & \sum_{t=1}^{M} \frac{P_p \times D_t}{(1+r)^t}, & \text{if } m = 1 \\ Investment{+}Fee = & \sum_{t=T}^{M} \frac{P_p \times D_t}{(1+r)^t} + \sum_{t=1}^{T-1} \frac{I_t}{(1+r)^t}, & \text{if } 1 < m < M+1 \\ Investment{+}Fee = & \sum_{t=1}^{M} \frac{P_p \times D_t + I_t}{(1+r)^t}, & \text{if } m = M+1 \end{array}$$

where $Investment$ is the amount invested in the loan by a particular investor and *Fee* is the fixed fee charged by the platform for each investment.[15]

[14] In reality, payment rate of each pool varies depending on how many loans in the pool default in a given month. However, because at the moment of investment investors do not know the future rates, they have to rely on some assumptions regarding the process. In my calculations, I make a simplifying assumption that payment rate remains constant at the level as predicted by the platform. See Table 4.

[15] Presence of a fixed fee implies that the return depends on the investment amount. However, because the fee is very small relative to the minimal possible order, the effect should be negligible. So the calculations are done assuming that each investor allocates the minimal possible amount of 250 Euro per loan. Then, it is sufficient to calculate the return for (only) 3,354 loan applications instead of all 54,455 investments done at the platform. Because calculation of return involves computationally intensive optimization procedure, reducing the number of cases is crucial.

The expected return from a loan is given by a weighted sum of returns associated with all $M+1$ outcomes with weights given by the probabilities, $\hat{p}_1, ..., \hat{p}_{M+1}$:

$$E[Return] = \sum_{m=1}^{M+1} p_m \times R_m.$$

Figure 6 plots the distribution of annualized $E[Return]$ over all investment projects posted on the market. The sample mean of annualized $E[Return]$ is 6.8%.

The measure of returns variation given by the standard deviation in return is calculated as follows

$$SD[Return] = \sqrt{\sum_{m=1}^{M+1} p_m \times (R_m - E[Return])^2}$$

Figure 7 plots the calculated $SD[Return]$ against $E[Return]$ for all investment projects posted on the market. The calculated $SD[Return]$ and $E[Return]$ are further used as explanatory variables entering vector $\mathbf{x}_{njt}$ in Equation 2.

5.4 Estimation Results

Results of the estimation of Model 3 are reported in Table 7. Note that there are three different blocks of estimated parameters: $\hat{\boldsymbol{b}}$, $\hat{\boldsymbol{\sigma}}_\beta$ and $\hat{\boldsymbol{\gamma}}$. $\hat{\boldsymbol{b}}$ is an estimate of the vector of random coefficients $\boldsymbol{\beta}_n$; it represents the average effect of respective variables on the expected utility of male investors. $\hat{\boldsymbol{\sigma}}_\beta$ is the estimated standard deviation of $\boldsymbol{\beta}_n$ reflecting the variation of tastes among investors. The estimate $\hat{\boldsymbol{\gamma}}$ is the parameter of primary interest; it shows how the average effect of variables on the utility of female investors differs from the effect of these variables on the utility of male investors.

Results reported under the header "*S1*" are obtained for a reduced specification of the vector of random parameters $\boldsymbol{\beta}_n$ and the vector of explanatory variables $\mathbf{x}_{njt}$. Specifically, $\boldsymbol{\beta}_n = \boldsymbol{b} + \boldsymbol{\eta}_n$ while $\mathbf{x}_{njt}$ includes only two variables $E[Return]$ and $SD[Return]$. This specification does not take into account investors' gender. However, it allows seeing how investors respond to projects' expected return and variation. The estimate of the mean of the coefficient for $E[Return]$ is 0.79 and is statistically significant implying that expected utility of a project increases with $E[Return]$, holding other characteristics of the project constant. The estimate of the standard deviation of the coefficient for $E[Return]$ is 0.528 and is statistically significant. This means that there is considerable variation in investors' responses to the level of projects' return. For a small fraction of investors the coefficient is even negative.[16] This result does not necessarily imply that investors dislike higher returns. Rather it signals that a fraction of investors rely on a decision rule different from the mean-variance principle. Moreover, in the context of peer-to-peer lending, investors may derive significant utility from social returns stemming from awareness that invested money will be used for a socially useful project or help another person out in a difficult situation. Respectively, individuals may engage more willingly in less profitable projects if such projects are associated with substantial social returns.

The estimate of the mean of the coefficient for $SD[Return]$ is negative (-0.267) and statistically significant, which indicates that on average investors dislike variation in returns.

[16]This inference is derived from the properties of normal distribution. Because coefficients are assumed to be normally distributed, 68% of investors fall within the range between $-\sigma$ and $+\sigma$; 95% of investors fall within the range between -2σ and $+2\sigma$; and 99% of investors fall within the range between -3σ and $+3\sigma$

The probability of investing in a project decreases in response to a marginal increase in return's variation. Hence, the majority of investors on the p2p platform seem to be risk-averse. The estimate of the standard deviation of the coefficient for $SD[Return]$ is statistically significant meaning that preferences for returns' variance vary in the population. Moreover, the magnitude of the standard deviation implies that for a considerable number of investors higher variation in returns is associated with higher expected utility. Again, this result may emerge because not all investors consider mean-variance rule as a criterion for investment choice. Or, alternatively, the finding may indicate that a portion of investors are risk-loving.

Results reported under the header "*S2*" are obtained for the same specification of β as before, but this time vector $\mathbf{x}_{njt}$ is extended by including other observable characteristics of loan projects. Previously received results for the effects of $E[Return]$ and $SD[Return]$ remain basically unchanged: Utility of investors is positively related to investments' return and negatively related to the variation of return. However, the magnitude of the estimates of the means of the two coefficients decreased compared to results for the baseline specification. Because of the way the expected return and its variance are calculated in the study, they depend on the attributes of the projects. When the attributes are additionally included in the regression equation together with $E[Return]$ and $SD[Return]$, it can lead to multicollinearity and respectively reduce the significance ascribed to $E[Return]$ and $SD[Return]$. Moreover, the fact that all considered attributes have significant effect on investors' utility indicates that investors attach significant value to the attributes in addition to the impact these factors have on return and its variation. For example, the coefficient estimate for the dummy variable *Loan duration=60 months* is -1.067 meaning that investors prefer short term loans over long term loans. Even when investors realize that, *ceteris paribus*, return is negatively linked to loan duration they may attach additional negative value to long durations simply because they dislike it when their money is tied up for a long time.

Finally, results under the header "*S3*" relate to an extended specification when $\boldsymbol{\beta}_n = \boldsymbol{b} + \boldsymbol{\gamma} Female_{ijt} + \boldsymbol{\eta}_n$. This specification allows the effect of projects' attributes to vary with investors' gender. The main parameters of interest are reported in the lower part of the table under $\hat{\boldsymbol{\gamma}}$. Coefficient estimates for $E[Return]$ and $SD[Return]$ are statistically insignificant meaning that the effect of one standard deviation of project's return on utility of female investors is not different from the effect on utility of male investors. Hence, contrary to the research hypothesis, a marginal increase in returns variability reduces the utility of a female investor as much as it reduces the utility of a male investor. Also borrowers' rating and financial burden – the two characteristics that might be considered by investors as a rough proxy for investments' riskiness – has the same effect for females as for males. Therefore, the results do not confirm that gender has an effect on investor risk taking propensity.

However, some significant gender differences in investor tastes are found with respect to other attributes of investment projects. For instance, females seem to dislike long-term loans more than males. Unlike males, females prefer consumer loans over business loans. The only borrower-specific characteristic where female investors seem to have different tastes than males is borrower age: Utility derived by females increases with borrower age. However, this result may be driven by the fact that female investors participating at *Smava* are, on average, somewhat older than male investors. Noteworthy, the effect of borrowers' gender does not vary with investors' gender. Hence, both male and female investors are more willing to provide funds to female borrowers than to male borrowers.

6 Conclusions

This paper examines the role of investor gender in their propensity for risk taking when investing on an online p2p credit market. A p2p market serves as a channel though which investors directly allocate capital to investment projects without intermediation of a financial institution or advisor. Because payoffs from loans are uncertain, p2p loans can be seen as a form of risky investment. Investors' choices allow making inference about their risk preferences.

A comparison of investment choices of male and female investors participating in p2p-lending does not reveal any significant differences with respect to their risk propensity. Relying on the mean-variance framework, I test whether female investors respond to increasing variance in expected returns differently than male investors. The results of a test show that gender does not matter for investors' risk preferences. A marginal increase in the standard deviation of expected return equally affects the utility of males and females. Moreover, no differences between male and female investors are found with respect to other characteristics of projects that may serve as proxy for projects' riskiness. Hence, the data on peer-to-peer lending do not support the conjecture that women tend to take less risks in investment decisions than their male counterparts.

However, the results should be enjoyed with caution because low participation of females in the market indicates self-selection. If probability of investing at the market is correlated with individual risk-propensity, then obtained results cannot be generalized to the overall population. Nevertheless, the study provides useful evidence on the behavior of individuals who are likely to self-select into risk-taking activities. A conclusion that can be derived from this perspective is that gender seems to play no role in the behavior of individuals who deliberately engage in risk-taking. Hence, the results are consistent with studies showing that, among professionally trained investors, females behave similarly to males with respect to risk (Johnson and Powell, 1994; Atkinson et al., 2003). The present study supports and extends this literature by showing that this relationship holds also in self-selected groups of not-trained retail investors.

Bibliography

References

Atkinson, S. M., S. B. Baird, and M. B. Frye (2003). Do female mutual fund managers manage differently? *Journal of Financial Research 26*(1), 1–18.

Badunenko, O., N. Barasinska, and D. Schäfer (2009). Risk attitudes and investment decisions across european countries: Are women more conservative investors than men? *DIW Discussion Papers 928*.

Beckmann, D. and L. Menkhoff (2008). Will women be women? Analyzing the gender difference among financial experts. *Kyklos 61*(3), 364–384.

Bennhold, K. (2009). Where would we be if women ran Wall Street? *The New York Times* (February 1).

Benson, K. L. and J. E. Humphrey (2008). Socially responsible investment funds: Investor reaction to current and past returns. *Journal of Banking & Finance 32*(9), 1850 – 1859.

Bernasek, A. and S. Shwiff (2001). Gender, risk, and retirement. *Journal of Economic Issues 35*(2), 345–356.

Bigelow, J. P. (1993). Consistency of mean-variance analysis and expected utility analysis: A complete characterization. *Economics Letters 43*(2), 187 – 192.

Bollen, N. P. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis 42*, 683–708.

Brownstone, D. and K. Train (1998). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics 89*(1-2), 109–129.

Chamberlain, G. (1983). A characterization of the distributions that imply mean–variance utility functions. *Journal of Economic Theory 29*(1), 185 – 201.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological) 34*(2), pp. 187–220.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–74.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2005). Individual risk attitudes: New evidence from a large, representative, experimentally-validated survey. *Framed Field Experiments Working Paper* (0019).

Duarte, J., S. Siegel, and L. A. Young (2009). Trust and credit. *SSRN Working Paper Series*.

Eckel, C. C. and P. J. Grossman (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization 68*(1), 1 – 17.

Economist (2009). Of bankers and bankeresses. *Economist* (August 6).

Fehr-Duda, Helga, d. G. M. and R. Schubert (2006). Gender, financial risk, and probability weights. *Theory and Decision 60*, 283–313.

Freedman, S. and G. Z. Jin (2008). Do social networks solve information problems for peer-to-peer lending? Evidence from prosper.com. *NET Institute Working Paper 8-43*.

FTD (2009). Ebay für Kredite. *Financial Times Deutschland* (February 17).

Garman, S., R. Hampshire, and R. Krishnan (2008). A search theoretic model of person-to-person lending. *Working Paper*.

Holt, C. A. and S. K. Laury (2002). Risk aversion and incentive effects. *The American Economic Review 92*(5), pp. 1644–1655.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics 57*(1), 129–38.

Jianakoplos, N. and A. Bernasek (1998). Are women more risk averse? *Economic Inquiry 36*(4), 620–30.

Johnson, J. and P. Powell (1994). Decision making, risk and gender: Are managers different? *British Journal of Management 5*, 123–138.

Kim, J. J. (2009). Peer-to-peer lending refuses to die. *The Wall Street Journal* (January 22).

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics 15*(5), 447–470.

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *The American Economic Review 77*(3), pp. 421–430.

Meyer, J. and R. H. Rasche (1992). Sufficient conditions for expected utility to imply mean-standard deviation rankings: Empirical evidence concerning the location and scale condition. *The Economic Journal 102*(410), pp. 91–106.

Meyer, T. (2009). The power of people: Online p2p lending nibbles at banks' loan business. *Deutsche Bank Research, E-Banking Snapshot* (July 22).

Niessen, A. and S. Ruenzi (2007). Sex matters: Gender differences in a professional setting. *Working Paper*.

Oakeshott, I. (2009). News review interview: Harriet Harman. *The Sunday Times* (August 2).

Olsen, R. A. and C. M. Cox (2001). The influence of gender on the perception and response to investment risk: The case of professional investors. *Journal of Behavioral Finance 2*(1), 29–36.

Pope, D. G. and J. R. Sydnor (2008). What's in a picture? Evidence of discrimination from prosper.com. *SSRN Working Paper Series*.

Powell, M. and D. Ansic (1997). Gender differences in risk behaviour in financial decision-making: An experimental analysis. *Journal of Economic Psychology 18*(6), 605–628.

Revelt, D. and K. Train (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. *The Review of Economics and Statistics 80*(4), 647–657.

Schubert, R., M. Brown, M. Gysler, and H. W. Brachinger (1999). Financial decision-making: Are women really more risk-averse? *The American Economic Review 89*(2), 381–385.

Sunden, A. E. and B. J. Surette (1998). Gender differences in the allocation of assets in retirement savings plans. *The American Economic Review 88*(2), 207–211.

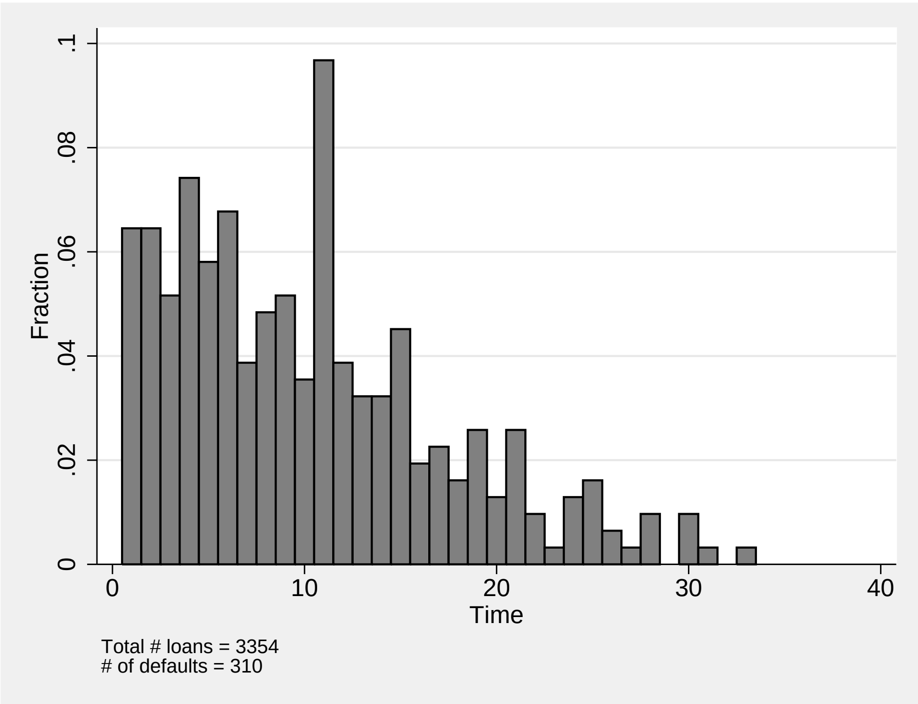
Sviokla, J. (2009). Forget Citibank, borrow from Bob. *Harvard Business Review*.

Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Appendix A: Estimation of the probability of default

Probability of default in a given month of a loan's duration is estimated using the observed payment behavior of borrowers at *Smava*. Information about borrowers' payment behavior is collected from http://www.beobach.de/. Repayment history is observed through the end of December 2010. Figure 1 plots distribution of defaults by the month of default observed in the data. Since the market is young, many credits have not matured yet. Specifically, of 3,354 loans that were granted between March 2007 and March 2010, 386 were repaid (including early repayments) and 310 defaulted by the end of December 2010.

Figure 1: Distribution of defaults by month of default



T is the discrete random variable indexed by a set of positive integers $T = \{1, 2, 3, ..., M\}$ indicating the installment at which a default occurred. Let $f(t)$ denote the probability distribution function of T and $F(t)$ denote the cumulative probability function describing the probability that $T \leq t$. Let $S(t)$ denote the survival function of T describing the probability that default occurs at some time after month t. Essentially, the survival function shows the probability that a borrower serves the debt for at least t months. The relationship of $S(t)$ to $f(t)$ is straightforward: $S(t) = Pr(T > t) = 1 - F(t) = 1 - \sum_{t=1}^{T} f(t)$.

Now, denote the conditional probability that a default occurs in month t conditional on the probability that the debt was timely served during $t-1$ months, as $h(t)$. This conditional probability is known as the discrete-time hazard rate and is linked to the survival probability in the following way

$$h(t) = Pr[T = t | T \geq t] = \frac{f(t)}{S(t-1)}.$$

As shown by Jenkins (1995), $h(t)$ can be estimated using conventional estimation methods for binary response variables. In order to do so, the data are organized such that the unit of observation is the monthly payment and not a loan. Each loan contributes as many observations to the data set as there are monthly payments done by a borrower to repay the loan.

Define a new binary indicator variable y_{it} with $y_{it} = 1$ if loan i defaults in month t and $y_{it} = 0$ otherwise. Note that $Pr(y_{it} = 1|T \geq t) = Pr[T = t|T \geq t] = h_i(t)$. Hence, log-likelihood of observing the data is

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T} \Big[y_{it} \log(h_i(t)) + (1 - y_{it}) \times \log(1 - h_i(t)) \Big].$$

All that is needed now to estimate the hazard rate is a functional specification of $h(t)$. The most commonly used specification is the logistic hazard function (Cox, 1972; Jenkins, 1995). Logistic distribution of the hazard rate implies that $h(t)$ can be estimated by means of a logit regression:

$$h(t|X) = \frac{\exp(\alpha_0 + \alpha_1 \ln(t) + \beta X)}{(1 + exp(\alpha_0 + \alpha_1 \ln(t) + \beta X)}. \tag{4}$$

Time dependence of the hazard rate is operationalized by including a logarithmic function of time, $\alpha_1 ln(t)$ into the model. Such specification of duration dependence implies a monotonically decreasing hazard if $\alpha_1 < 0$, a monotonically increasing hazard if $\alpha_1 > 0$, and a constant hazard if $\alpha_1 = 0$. The effect of observable characteristics included in vector X is captured in parameters' vector β. Vector X includes the following variables: raised loan amount (divided by 250), offered interest rate in % p.a., loan duration, loan purpose, borrower's Schufa-rating with "A" being the best grade, financial burden, employment status, age, gender, place of residence, loan vintage (year and quarter when the first payment is due) and calendar month of payment to capture seasonality effects. Note that only observations on approved loans can be used to estimate equation 4. Estimation results are reported in Table 1. According to the results, $\hat{\alpha}_1 = -0.150$. Thus, $\hat{h}(t|X)$ decreases with the time.

Table 1: Estimation results after discrete-time hazard model

	Coeff.	St.Error
Raised amount	0.009***	(0.00)
Offered interest rate	0.166***	(0.04)
Loan duration		
36 months (ref.)		
60 months	0.147	(0.15)
Rating		
A		
B	0.544*	(0.33)
C	0.250	(0.38)
D	0.422	(0.37)
E	0.590*	(0.35)
F	0.636*	(0.35)
G	0.746**	(0.37)
H	0.889**	(0.42)
Financial burden		
low (ref.)		
moderate	0.870**	(0.35)
substantial	0.863***	(0.33)
high	1.076***	(0.33)
Employment		
Arbeiter/Angestellter (ref.)		
Beamter	-1.135*	(0.61)

Freiberufler	-0.985***	(0.33)
Geschäftsführer	0.044	(0.29)
Gewerbetreibender	0.026	(0.18)
Rentner	0.374	(0.30)
Age	-0.001	(0.01)
Gender		
Male (ref.)		
Female	0.078	(0.14)
Loan purpose		
Aus- & Weiterbildung	-0.001	(0.37)
Auto & Motorrad	0.301	(0.21)
Familie & Erziehung	-0.051	(0.24)
Feierlichkeiten & besondere Anlässe	-0.290	(0.52)
Geschäftserweiterung	-0.439	(0.38)
Gesundheit & Lifestyle	-0.027	(0.44)
Gewerblicher Kreditbedarf	-0.358	(0.46)
Haus, Garten, Heimwerken (ref.)		
Investition	-0.402	(0.66)
Liquidität	0.158	(0.26)
Reisen & Urlaub	-0.268	(0.48)
Sammeln & Seltenes		
Sonstiges	0.297	(0.21)
Sport & Freizeit	0.380	(0.38)
Tierwelt	0.847**	(0.42)
Umschuldung	0.088	(0.25)
Unterhaltungselektronik & Technik	0.415	(0.37)
Place of residence		
Baden-Württemberg	0.036	(0.23)
Bayern	-0.410*	(0.23)
Berlin	0.157	(0.28)
Brandenburg	0.087	(0.37)
Bremen		
Hamburg	0.550	(0.36)
Hessen	0.326	(0.23)
Mecklenburg-Vorpommern	0.242	(0.49)
Niedersachsen	0.241	(0.25)
Nordrhein-Westfalen (ref.)		
Rheinland-Pfalz	0.324	(0.30)
Saarland	-0.724	(1.03)
Sachsen	0.672***	(0.26)
Sachsen-Anhalt	0.436	(0.39)
Schleswig-Holstein	0.841***	(0.26)
Thüringen	0.613*	(0.35)
Season		
Jan (ref.)		
Feb	1.126***	(0.33)
Mar	0.608*	(0.36)
Apr	0.920***	(0.34)
Mai	0.670*	(0.35)
Jun	0.430	(0.36)
Jul	0.930***	(0.33)
Aug	0.149	(0.38)
Sep	0.720**	(0.34)
Okt	0.549	(0.35)
Nov	0.620*	(0.34)
Dec	-0.198	(0.39)
Vintage		
2007q2 (ref.)		
2007q3	0.282	(0.87)
2007q4	-0.316	(0.82)
2008q1	-0.039	(0.81)
2008q2	-0.023	(0.80)
2008q3	0.207	(0.81)
2008q4	0.076	(0.81)
2009q1	-0.235	(0.81)
2009q2	0.142	(0.80)
2009q3	-0.076	(0.81)
2009q4	-0.597	(0.82)
2010q1	-0.765	(0.84)
2010q2	-0.716	(1.07)
ln(t)	-0.150**	(0.06)
Constant	-9.263***	(1.00)

Pseudo-R^2	0.096
N	56589

* p<0.10, ** p<0.05, *** p<0.01
(ref.) = reference category

Using the vector of estimated coefficients, I calculate for each loan application posted at the platform (not only the approved ones) its individual hazard function, $\hat{h}_l(t)$. Based on the determined hazard function, the survival function $\hat{S}_l(t)$ and the probability distribution function $\hat{f}_l(t)$ are calculated for each loan application:

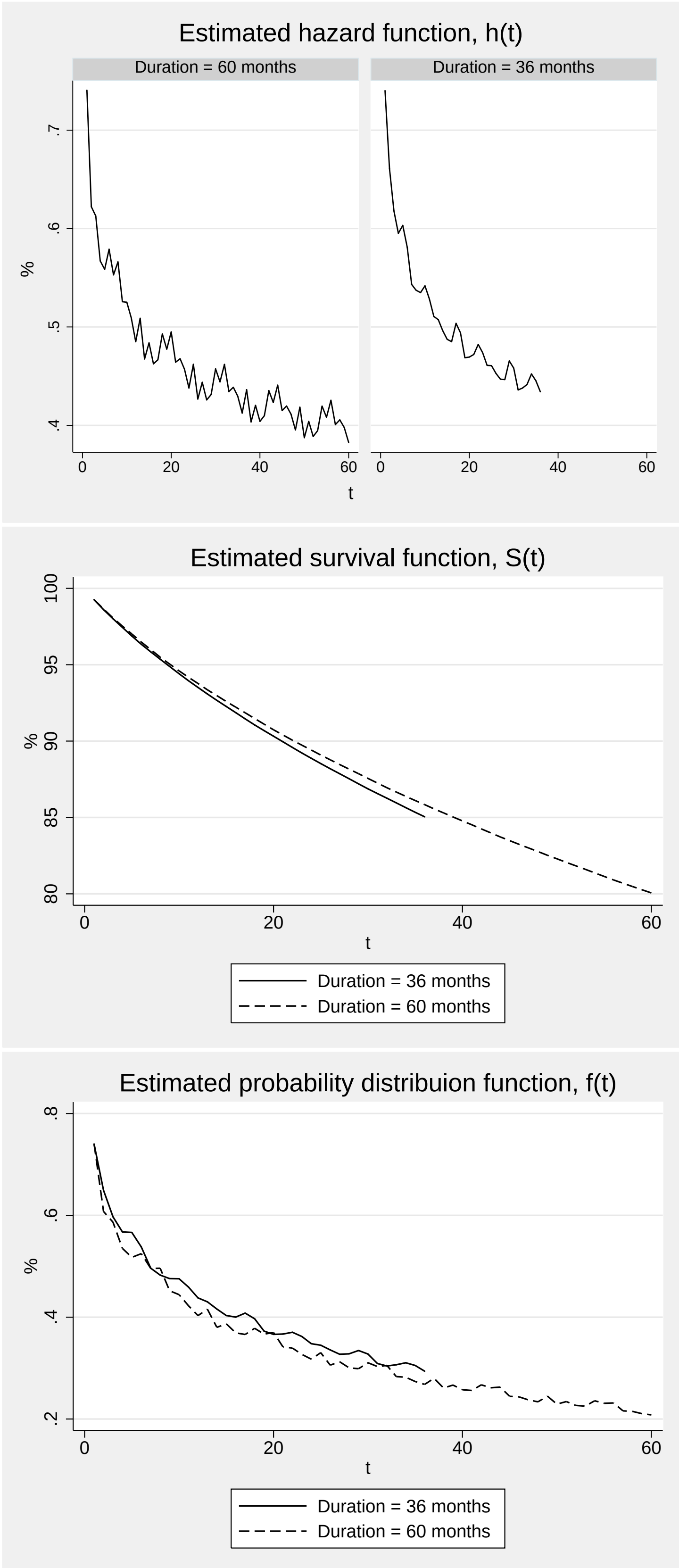
$$\hat{S}_l(t) = \prod_{j=1}^{t}(1 - \hat{h}_l(j)),$$

$$\hat{f}_l(t) = \begin{cases} 1 - \hat{S}_l(t), & \text{if } t = 1 \\ \hat{S}_l(t-1) - \hat{S}_l(t), & \text{if } t > 1. \end{cases}$$

The sample means for the hazard, survival and probability distribution functions are plotted in Figure 2. Estimated probability distribution function of loan l, $\hat{f}_l(t)$, is used to determine $p_{1l}, \ldots, p_{M+1\,l}$ – the probability of each possible outcome from loan l:

$$\hat{p}_{tl} = \begin{cases} \hat{f}_l(t), & \forall\, t < M+1 \\ 1 - \sum_{t=1}^{M} \hat{f}_l(t), & \text{if } t = M+1. \end{cases}$$

Figure 2: Estimated functions



Figures and Tables

Figure 3: Loans procured at *Smava*

This graph plots cumulative distribution of the number and the volume of loans procured at the platform between March, 2007 and March, 2010

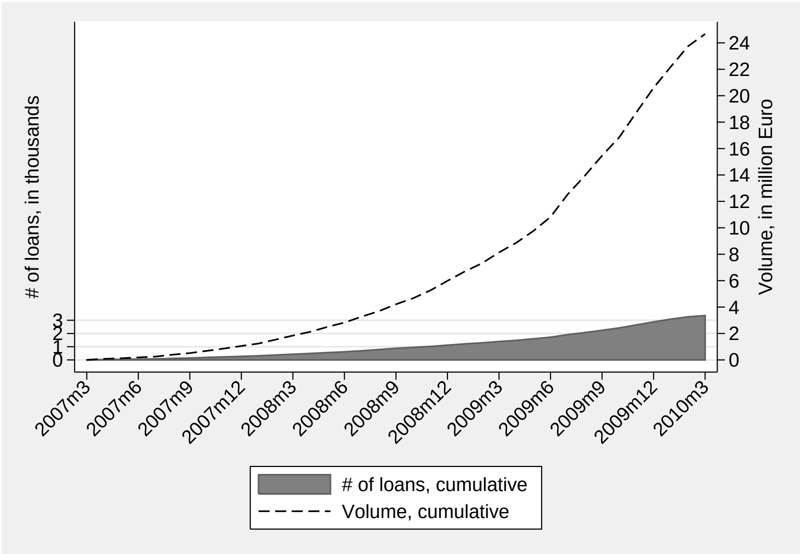


Figure 4: Distribution of loan applications by loan purpose

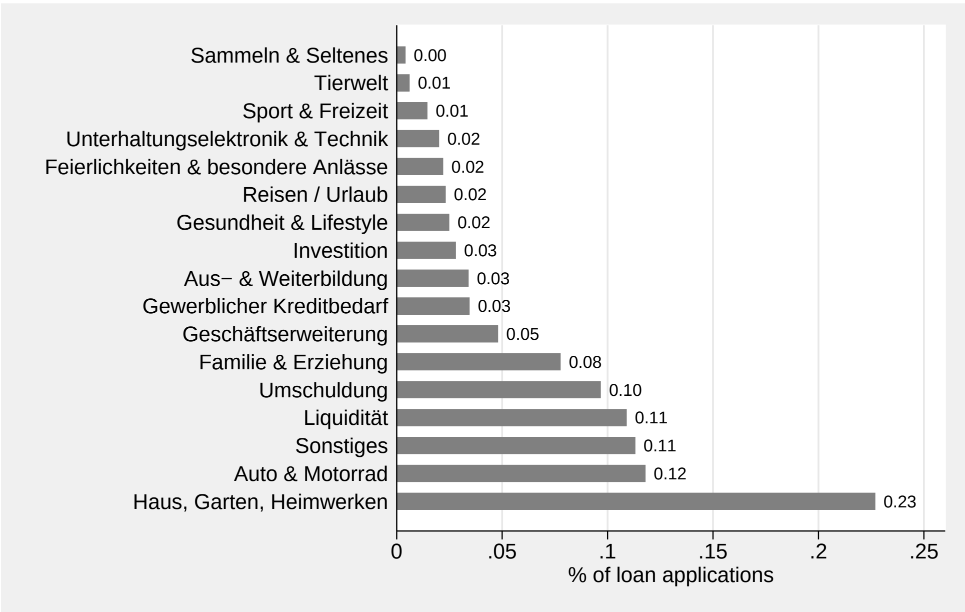


Figure 5: Possible outcomes of investment in a loan with duration 36 months

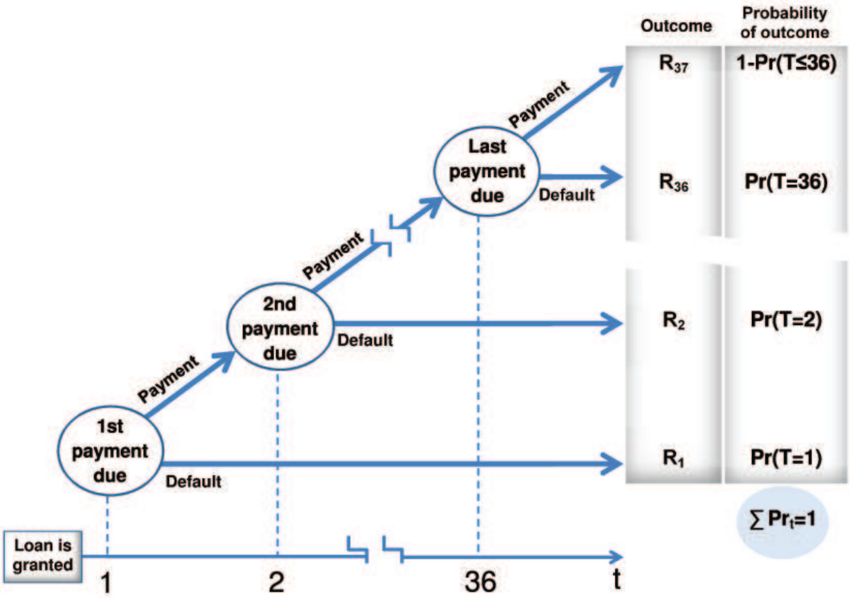


Figure 6: Distribution of expected return over projects

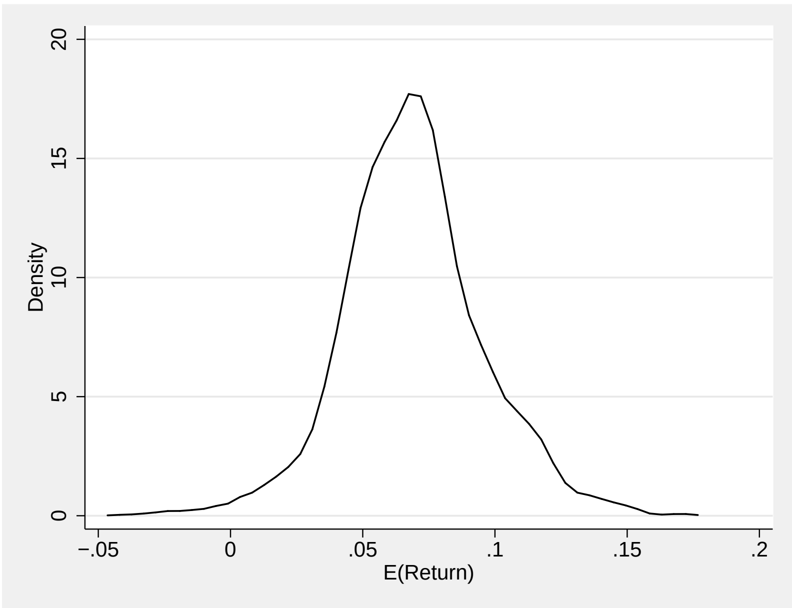


Figure 7: Standard deviation of return plotted against the expected return

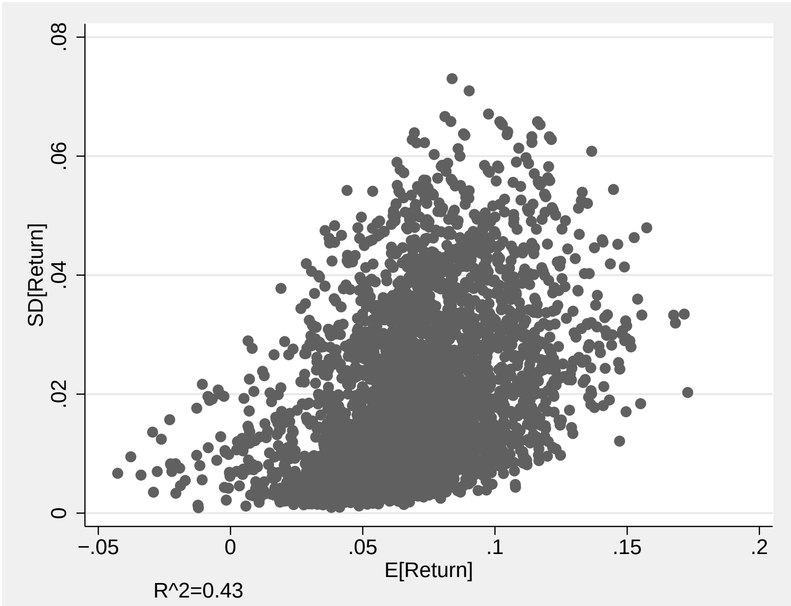


Figure 8: Distribution of choice sets by the number of alternatives in a set

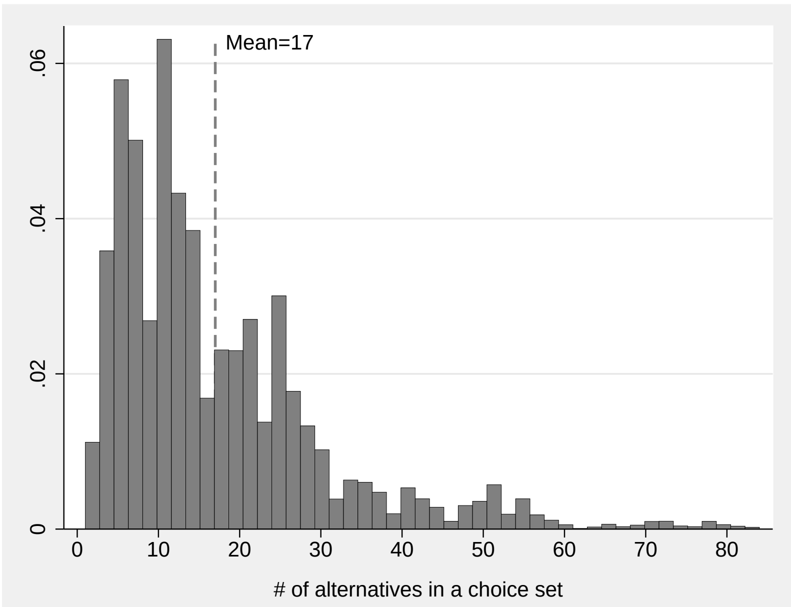


Table 2: KDF-Indicator

KDF-Indikator	Debt-to-disposable income ratio
1	0 bis 20%
2	20 bis 40%
3	40 bis 60%
4	60 bis 67%

Table 3: Creditworthiness rating grades and corresponding PDs

This table shows the rating grades that eligible individuals to borrow at the platform. The rating grades are assigned to borrowers by the German national credit bureau *Schufa*. Each rating grade reflects the probability of a borrower's default given his past credit behavior and current debt obligations.

Rating grade	Probability of default	Fraction of loans with given rating grade
A	1.38	15.91
B	2.46	16.21
C	3.56	10.16
D	4.41	9.94
E	5.57	10.83
F	7.16	12.30
G	10.72	14.77
H	15.02	9.88

Table 4: Historical payment rates in pools

This table shows the predicted and the historical average payment rate for each of the 16 pools. The historical average is calculated over the period from April 2007 to January 2010. Source: http://www.smava.de.

	Loans with duration 36 months							
	A	B	C	D	E	F	G	H
Predicted payment rate	98.8	97.8	96.6	96.1	95.1	93.7	90.6	87.1
Historical average	97.4	95.8	98.4	95.6	95.9	92.4	92.0	89.7
	Loans with duration 60 months							
	A	B	C	D	E	F	G	H
Predicted payment rate	98.5	97.4	95.8	95.4	94.2	92.4	88.8	84.6
Historical average	99.5	97.8	98.5	91.5	95.2	94.2	85.9	84.1

Table 5: Summary statistics of selected variables by investors' gender

	Males N=5,046		Females N=625			
Variable	Mean	St.Dev.	Mean	St.Dev.	t-Test	p-value
Age	41	12.32	45	12.50	-6.31	0.00
Duration of participation	14	8.72	13	7.80	3.81	0.00
# of submitted orders	10	16	8	12	2.24	0.02
Order value, in €	469	372	481	391	-0.80	0.42
Total amount invested	4,436	8456	4,004	7165	1.74	0.25

Table 6: Definition of explanatory variables

Variable Name	Description
Investor-specific characteristics	
Female	= 1 if investor is female, = 0 otherwise
Loan-specific characteristics	
E(Return)	Expected rate of return to investment, in % p.a.
SD(Return)	Standard deviation of the expected rate of return form investment
Amount[a]	Requested loan amount, in €.
Duration=60 months	Dummy variable = 1 if loan duration is 60 months, = 0 if 36 months
Offered interest rate	Offered nominal annual interest rate, in %
Purpose	=1 if business loan, =0 if consumer loan
Description	Length of description of loan purpose, in # of characters
Borrower-specific characteristics	
Age	Age in years
Female	= 1 if borrower is female, = 0 if male
Rating	measures borrowers' creditworthiness, takes 8 values from A (best) to H (worst)
Financial burden: low	= 1 if borrower's debt-to-income ratio does not exceed 20% and 0 otherwise
Financial burden: moderate	=1 if debt-to-income ratio lies within the range 20-40% and 0 otherwise
Financial burden: substantial	=1 if debt-to-income ratio lies within the range 40-60% and 0 otherwise
Financial burden: high	=1 if debt-to-income ratio lies within the range 60-67% and 0 otherwise
Employment	=1 if borrower is self-employed 0 if employed or retired

[a] Since the value is always a multiple of 250, the variable is scaled by factor $\frac{1}{250}$ when used in regressions

Table 7: Estimation results after mixed logit regression

	S1		S2		S3	
	Estimate	St.Error	Estimate	St.Error	Estimate	St.Error
		$\hat{\boldsymbol{b}}$				
E[Return]	0.790***	(0.01)	0.599***	(0.01)	0.614***	(0.01)
SD[Return]	-0.267***	(0.01)	-0.179***	(0.02)	-0.176***	(0.02)
Rating			-0.520***	(0.01)	-0.519***	(0.01)
Loan duration: 36 months (ref.)						
60 months			-1.067***	(0.03)	-1.045***	(0.03)
ln(Amount)			-0.512***	(0.01)	-0.523***	(0.01)
Description			0.201***	(0.01)	0.198***	(0.01)
Offered interest rate			0.405***	(0.01)	0.400***	(0.01)
Loan purpose: consumer loan (ref.)						
business loan			0.095***	(0.02)	0.110***	(0.02)
Employment: employed or retired (ref.)						
self-employed			0.354***	(0.01)	0.350***	(0.01)
Age			-0.006***	(0.00)	-0.006***	(0.00)
Financial burden: low (ref.)						
moderate			0.396***	(0.02)	0.407***	(0.02)
substantial			0.541***	(0.02)	0.554***	(0.02)
high			0.618***	(0.03)	0.624***	(0.03)
Borrower gender: male (ref.)						
female			-0.096***	(0.01)	-0.094***	(0.01)
		$\hat{\boldsymbol{\sigma}}_\beta$				
E[Return]	0.528***	(0.01)	0.544***	(0.01)	0.533***	(0.01)
SD[Return]	0.479***	(0.01)	0.195***	(0.01)	0.199***	(0.01)
Rating			0.188***	(0.01)	0.198***	(0.01)
Loan duration: 36 months (ref.)						
60 months			1.404***	(0.03)	1.377***	(0.03)
ln(Amount)			0.264***	(0.02)	0.251***	(0.02)
Description			0.096***	(0.02)	0.064**	(0.03)
Offered interest rate			0.054***	(0.01)	0.022**	(0.01)
Loan purpose: consumer loan (ref.)						
business loan			0.008	(0.04)	0.013	(0.04)
Employment: employed or retired (ref.)						
self-employed			0.041*	(0.02)	0.048**	(0.02)
Age			0.000	(0.00)	0.000	(0.00)
Financial burden: low (ref.)						
moderate			0.021	(0.03)	0.027	(0.03)
substantial			0.003	(0.02)	0.016	(0.02)
high			0.135***	(0.03)	0.116***	(0.03)
Borrower gender: male (ref.)						
female			0.004	(0.02)	0.011	(0.02)
		$\hat{\boldsymbol{\gamma}}$				
E[Return]					-0.016	(0.04)
SD[Return]					0.029	(0.05)
Rating					-0.020	(0.03)
Loan duration: 36 months (ref.)						
60 months					-0.209**	(0.10)
ln(Amount)					0.056	(0.04)
Description					0.001	(0.02)
Offered interest rate					0.028	(0.04)
Loan purpose: consumer loan (ref.)						
business loan					-0.139*	(0.08)
Employment: employed or retired (ref.)						
self-employed					0.021	(0.05)
Age					0.003**	(0.00)
Financial burden: low (ref.)						
moderate					-0.107	(0.08)
substantial					-0.145	(0.14)
high					-0.102	(0.09)
Borrower gender: male (ref.)						
female					0.016	(0.05)
Log-likelihood	-99629		-89021		-89013	
N	931271		931271		931271	

* p<0.10, ** p<0.05, *** p<0.01; (ref.) = reference category