Measuring Applicant Quality to Detect Discrimination In Peer-to-Peer Lending

Georg Weizsäcker (Humboldt University Berlin and DIW Berlin)
Christian Zankiewicz (DIW Berlin)

Discussion Paper No. 13

March 25, 2017

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Measuring Applicant Quality to Detect Discrimination in Peer-to-Peer Lending*

Georg Weizsäcker Christian Zankiewicz†

February 8, 2017

Abstract

We measure the quality of applications for online peer-to-peer lending in Germany and relate it to gender discrimination. The data context allows summarizing application quality as a single numeric measure, the expected internal rate of return. The measure serves as a control variable and is interacted with the applicants' gender. We find that women enjoy higher funding rates than men, mainly because they are less punished when they offer a low application quality. The evidence is consistent with the hypothesis that the predominantly male lenders have a less precise understanding of women's applications than of men's applications.

JEL-Classification: D14, D84
Keywords: Gender discrimination, household finance, irrational beliefs

*We are grateful for many discussions with workshop audiences and students in Berlin, with Manuel Bagues, Dorothea Kübler, and especially with Tobias Schmidt who participated in the early stages of this project and with Nataliya Barasinska who provided us with the data and invaluable guidance on how to use them. Support by German Science Foundation through CRC TRR 190 and by the European Research Council through Starting Grant 263412 is gratefully acknowledged.

†both authors: Humboldt University, Spandauer Straße 1, 10178 Berlin, and DIW Berlin, Mohrenstraße 58, 10117 Berlin, Germany. Email: weizsaecker@hu-berlin.de, czankiewicz@diw.de

1. Introduction

Many interesting aspects of gender discrimination concern the interaction of an applicant's gender and quality. Are women's success chances more or less correlated with quality than men's? The answer to this question may well differ from what the unconditional success rates of men and women suggest. Employers, sponsors or other potential discriminators may, for example, favor women over men overall but reward the quality of women's applications less than for men; or they may show the opposite pattern. Allocations and incentives can be severely influenced by such "slope-discriminating" behavior that may appear on top of potential "level-effect" discrimination.

A major difficulty for the analyst is to measure the quality of an application. In most contexts, only proxies for applicant quality are available and their interpretation is often all too flexible. The statistical connection between a given proxy and quality may be nonlinear (more generally, difficult to capture) and it may be different for men and women. It may also be subject to differential measurement error and to selection effects. A substantial portion of these measurement problems arises simply because the objective function of the potential discriminator is unknown. For example, job applicants promise a high-dimensional array of outcomes. It is usually beyond the analyst's power to assess the potential employers' valuations of these outcomes.

In this paper, we focus on a narrow financial context, peer-to-peer lending between German households on the online platform *smava.de*. Here, the applicant is a borrower who describes a project and makes a take-it-or-leave-it offer to all potential lenders. To predict the outcome of a loan contract, a potential lender merely has to assess the anticipated repayment stream. The quality of such a loan application can therefore be reasonably reduced to a single number: its expected internal rate of return. We observe all characteristics of the offered contract and of the applicant that are available to the potential lender, allowing us to assess this measure of predictable quality in detail. We argue that other lender objectives, over and above the expected internal rate of return, are unimportant under standard assumptions. The nature of the interaction between lender and borrower precludes any other relation between them. Risk considerations are also minimal, due to the platform's specific insurance mechanism.

Using our inferred measure of quality, we analyze the applicant's chances of success, with a particular focus on the interaction between gender and quality. We address measurement error by modeling the applicant's quality with a detailed structure and by including a statistical correction method (the SIMEX procedure of Cook and Sefanski, 1994). We find significant effects of both slope discrimination and level discrimination. Women have higher success rates than men, conditional on quality, but this gender difference is driven by a larger increase of men's success rate in quality: women appear to get the benefit of the doubt, such that low-quality applications of women are almost equally successful as high-quality applications of women and men. Low-quality applications of men, in contrast, are much less likely to be successful. In terms of chances of a project being fully funded, the success of a below-median-quality application by a man

is about half of that of an analogous application by a woman.
Within the larger literature on discrimination, this result is noteworthy in that it confirms a particular feature of (some kinds of) statistical discrimination: if the potential discriminators—here, the predominantly male lenders—find it harder to judge a woman than to judge a man, then weak male applicants should have lower success. This feature is usually not found in discrimination studies, largely because a male-favoring level effect outweighs it. We also show that simpler proxies of quality yield different conclusions. One natural candidate proxy for quality is the applicant's credit rating. Its correlation with success, like that of our own quality measure, suggests that women enjoy positive level discrimination in our context, but not a positive slope discrimination. An alternative proxy, the nominal interest rate that the applicant offers, shows an even stronger interaction between gender and quality, such that female applicants do not benefit from higher quality at all. In this sense, one can read our study as cautioning that the choice of proxies for quality is highly important for the conclusion.
A note on causality is in order. We describe correlations, no more. An advantage for simple interpretations of our results is that the available control variables cover essentially all information that is available to the lenders. The lenders therefore plausibly do not condition their choice on other factors. However, we cannot rule out that borrower selection on unobservables (such as a borrower's previous attempts to secure funding) influences the results.
In the next section, we briefly review the literature on discrimination with a particular focus on slope discrimination. Section 3 describes our data context and Section 4 our econometric modelling. Section 5 shows the results and Section 6 concludes.

2. Discrimination Literature: Brief Summary and Connections

As indicated above, we distinguish between level discrimination and slope discrimination. In the extant literature, level discrimination is the much more frequent object of investigation, among the two. It features in studies on differential chances of getting a job offer depending on one's race (see e.g. Nunley et al., 2014), gender (Kuhn and Shen, 2012), religion (Wright et al., 2013), sexual orientation (Ahmed, Anderson and Hammarstedt, 2013), age (Riach and Rich, 2010) or attractiveness (Rooth, 2009). The discrimination literature is not limited to hiring decisions, of course, as referee decisions in sport (Price and Wolfers, 2010), bidder's choices on Ebay (Kricheli-Katz and Regev, 2016), rental market decisions (Carlsson and Eriksson, 2014) and retail market decisions (Zussman, 2013) have all shown to be prone to level discrimination, too. As for the reasons for discrimination, the most prominent proposed theories are Becker's (1957) taste-based or Phelps' (1972) and Arrow's (1973) statistical discrimination (see also Aigner and Cain, 1977, for influential early work). Between them, it is no-table that taste-based discrimination does more straightforwardly predict level discrimination while the discussion of statistical discrimination much more often includes predictions of slope discrimination. Especially the sub-literature on screening discrimination (Cornell and Welch, 1996, and subsequent work) is related to our study. Its main assumption is that the employer can assess certain groups of applicants better than other groups. An application by a man may, e.g., be less punished for its low quality than in the case of a woman. This specific prediction is, however, essentially unconfirmed in the empirical discrimination literature (see e.g. the explicit statement in Bagues and Perez-Villadoniga, 2013).

More general than in our context, slope discrimination concerns the differential valuation of certain achievements, qualifications or characteristics, evidenced for example in criminal background checks (Pager, 2003) or in the quality of a resumé (Bertrand and Mullainathan, 2004), both of which can be differently important for different applicant groups. Pager (2003) shows that a prior conviction decreases a white job applicant's chances for a call-back by half whereas an African-American ex-convict experiences a reduction in call-back chances by two-thirds, compared to having no criminal record. Bertrand and Mullainathan (2004) show that having a high-quality resumé is less effective for African-American job applicants than for whites and that the gap between African-American and Whites increases with resumé quality.[1] A possible explanation is that employers may use screening procedures such as not reading further if the job applicant has an African-American name. But many differential reactions of employers are, of course, also in line with Bayesian updating as the information can be differently informative about the performance on the job. Attributions to a particular mechanism are, thus, quite speculative in the absence of a suitable measurement of applicant quality. The analyzed proxies for quality leave room for interpretation and one may regard their measurement, and the evidence for their correlations with the employers' objectives, as incomplete.

Only few discrimination studies focus on a sophisticated measurement of output productivity. Gallen (2015) uses Danish matched employer-employee data of five industries to estimate the relative productivity of men and women. The gender productivity gap is measured by estimating the efficiency units lost in a firm-level production function for a female worker compared to a male worker, holding other individual characteristics constant. The study finds that 75% of the gender wage gap (-0.16) can be explained by gender productivity differences. Hellerstein et al. (1999) follow a similar methodological path but focus solely on the manufacturing industry. They find a strikingly large gender pay gap of (-0.45) but a much more modest productivity gap (-0.16). Azmat and Ferrer (2015)

[1]While Nunley et al. (2014) replicate Bertrand and Mullainathan's (2004) result that differential success chances do not decrease between African-American and white job applicants when additional information about productivity is added to the resumé, Kaas and Manger (2012) find the opposite result, with discrimination of job applicants with Turkish-sounding names in Germany vanishing when a reference letter containing productivity information was added to the application.

look at productivity data of lawyers and find that roughly half of the gender earnings gap can be explained by productivity differences. From a distance, our paper's main difference to this literature is that we focus on loan applications. We, too, determine a relatively sophisticated quality measure, the expected rate of return, and ask whether this measure has a different correlation with success for men and women. This arguably reduces the room for interpretation significantly.

Prior research on peer-to-peer credit markets includes studies that focus on the borrowers' financial strength (such as credit grade), paid interest, or loan duration as covariates of funding success (Iyer et al., 2009; Herzenstein et al., 2008; Avery et al., 2004), studies that focus on personal characteristics such as race, age, gender, or beauty (Theseira, 2008; Pope and Sydnor, 2008; Ravina, 2008). Gender is usually just used as a control variable in these studies. In contrast, Barasinska and Schäfer (2010), which is also the peer-to-peer lending study most similar to our paper, specifically look for gender discrimination in the same data of the platform *smava.de*. Their regression results show no significant relationship between an applicant's gender and the likelihood of being fully funded. The fact that we find very different results despite using the same data is another indication that the method of measuring quality is key in analyses of discrimination.[2]

3. The Peer-to-Peer Lending Process

3.1. Participation and Information Conditions

The data set consists of loan applications posted between March 2007 and March 2010 on the German platform *smava.de*.[3] In contrast to traditional bank lending, peer-to-peer lending refers to direct lending between private persons. The platform sets the rules according to which the lending process is carried out and supplies the infrastructure of the web-based intermediation procedure. Although funds are exchanged (almost) directly between persons no further communication between peers is possible. Borrowing and lending is carried out anonymously. Informational asymmetries are high since investors have only access to a limited set of verified information provided by the platform.

In order to participate in the lending process, be it as lender or as borrower, adult members of Germany's general population can register at *smava.de* and verify their identity via the "postident" procedure administered by the German postal service provider. *smava.de* requests personal information such as one's name, gender, age and place of residence, of which only age and gender are published on the platform in an uncensored way. Borrowers and lenders choose

[2]Other studies using the online lending platform *smava.de* as data source are Barasinska (2011), Kraus (2013), and Pötzsch and Böhme (2010).

[3]The data are suitable for our purposes only until March 2010, whereafter a loan application can involve joint liability of two persons living in the same household, with only aggregated financial information being shown on the platform.

a username and the user's address is opacitized in the sense that only the federal state of residence is published.
The platform also collects financial information of potential borrowers to infer their "Schufa rating" and "KDF indicator", both being published on the platform. The Schufa rating is the main German credit rating for individuals, issued by the German national credit bureau. It reflects a person's default risk on a scale from A (lowest risk) to M (highest risk). The KDF indicator measures a person's financial burden from outstanding consumer loans ranging between 1 (lowest financial burden) and 4 (highest financial burden).[4]
The available verified information is complemented by voluntary provided unverified information. That is, on the platform one can chose one of several categories indicating one's employment status and, after a successful registration, one can voluntarily upload a profile picture and write a text describing oneself in greater detail.
Registrants at *smava.de* have to be older than 18 years and must permanently reside in Germany. A person can register only either as lender or as borrower, not as both. For eligibility to borrow, one must prove a monthly income of € 1000 at least, a Schufa rating that is no worse than category H and a debt-to-disposable income ratio no higher than 67%.

3.2. The Terms Of An Investment

A successfully registered borrower can post a loan application that describes the requested loan amount (in multiples of € 500 with a maximum of € 50.000), the term of the loan (either 36 or 60 months), the annual interest rate he or she is willing to pay (in multiples of 0.1 percentage points), the purpose of the loan (an unverified choice from a fixed list of 17 categories) and a voluntary added text providing further (unverified) details about the loan application.
Based on the observable loan and borrower specific information a potential lender can decide whether or not to invest in a given loan application. A possible investment has to be in multiples of € 250 and must not exceed the requested total loan amount or € 25.000.[5] In contrast to other peer-to-peer lending platforms such as Prosper.com where loans are auctioned, *smava.de* works on a first-come-first-served basis. As soon as the requested loan amount equals the aggregated supply of funding, or after 14 days, the loan application is closed. Lenders can not underbid each other. However, loan applicants can raise the offered interest rate during the bidding period to attract more funding. The final rate is the ef-

[4]The specific calculations used by *smava.de* are as follows. The sum of all monthly payments to loans (including loans from *smava.de*) is divided by the monthly disposable income (without savings), resulting in the debt-to-disposable income ratio. This measure is assigned to categories from 1 to 4, constituting the KDF indicator. The relevant information as well as the indicator are continuously updated. Debt payments, income, and savings are not observable for other peers apart from the summary in the KDF indicator.

[5]Most investors provide only a small fraction of the requested loan amount so that loans are usually financed by many different investors.

fective rate for all lenders.
If less than 25% of the requested amount are funded after the loan is closed, the application is withdrawn from the platform and the received bids (if any) are transferred back to the lenders. In this case, the potential borrower can change the terms of the loan application and post a modified application.
If the supplied loan amount covers at least 25% of the requested sum, lenders are contractually bound to realize their bids accordingly. Upon acceptance of the funds by the loan applicant the loan contract is legally valid. After a legal loan contract is achieved, *smava.de* charges lenders and the borrower a fee.[6]

3.3. The Risk Of An Investment

After a loan is paid out, the borrower is contractually bound to repay the funds in constant monthly annuities. A borrower can repay early but, in this case, has to compensate the lenders for the missed interest payments. If a borrower fails to repay, lenders incur a loss.[7] That is, loans arranged by *smava.de* are not secured by collateral or third parties. But two insurance mechanisms apply. First, the claim to the resulting debt from a default is sold to a collecting agency which, if successful, recovers an average share of roughly 20% of the missed payments for the investors. Second, and more importantly, a risk sharing mechanism is installed, functioning as follows. Risk sharing is effective via loan pools. Lenders are pooled by two characteristics of the loans they have invested in: the loans' Schufa ratings and their durations until maturity. For example, all lenders who put their money in loan applications with a repayment duration of 60 months and a Schufa rating "B" are assigned to one pool. With two different loan durations and eight applicable Schufa ratings, the grouping results in a total of 16 different pools. The monthly principal re-payments are aggregated within each of the pools and each lender receives a share proportional to his or her relative investment in the respective pool. If a loan defaults, the resulting loss is subtracted from the pooled re-payments. Thus, each member of the pool partially compensates for the loss. The resulting re-payment rates per pool range from 99% for pools of loans with Schufa rating A to 84% for those with a Schufa rating H.[8] Interest payments are exempt from the pooling procedure and transferred directly to the investors if the loan did not default in the previous month. Overall, the risk sharing mechanism essentially insures against the loss of principal, at the cost of the effective pay-out lying below the nominal interest rate.

[6]The structure of the fees for borrowers and investors has been change by *smava.de* once during the period that is relevant for our study. Prior to February 2009, there was no fee for lenders and borrowers paid simply 1% of the borrowed amount. Afterwards, lenders were charged €4 per bid and borrowers 2% of the borrowed amount (or at least €40) for loans with a 36 month duration and 2.5% (or at least €40) for loans with 60 months duration. We account for this change in our loan return calculations and in our regressions via fixed effects.

[7]A failure of repayment is declared as soon as the monthly payment is 60 days late.

[8]The data represent historical average payment rates over the period from April 2007 to January 2010, published by *smava.de*.

As we show in the next section, more than 80% of completed contracts show repayments that corresponds to an internal rate of return lying within the interval [0.05, 0.1] and the remaining 20% all lie within the interval [-0.05, 0.2]. We regard the distribution of repayments as small enough to justify an analysis under the assumption of investor risk neutrality: a lender's main challenge is to identify the probability of default, which would induce a loss of interest payments from the time of default onwards. Expected-utility calculations under standard assumptions confirm that risk considerations can be neglected.[9]

3.4. The Data

Our data set consists of 4144 closed loan applications submitted by 3400 borrowers, including the online-published personal and financial information of the borrowers.

On average 73% of the loan applicants are male, with one half living in the northern states of Germany. Average borrower age is 44 and the mean requested sum is roughly € 8.000, offered at an average nominal 9.9% annual interest. The distribution over the 4 KDF-indicator categories is 1:17%, 2:24%, 3:33%, 4:26%. The distribution of individual borrowers' Schufa ratings is A:15%, B:16%, C:9%, D:10%, E:11%, F:12%, G:16%, H:11%.

At the end of our observational period, 5671 lenders were registered on the platform and submitted on average 10 funding decisions. Only 625 (11%) of lenders are female, precluding us from performing a meaningful interaction of our analysis with lender gender. The exact descriptives are shown in Table 1.

4. Empirical Strategy

In the following, we analyze whether and how lenders evaluate loan applications differently for male versus female applicants. For exposition, we start by presenting two null hypotheses that describe lender behavior as aiming at maximal expected returns, conditional on the variables that are observable to him or her. The main idea behind these hypotheses is that if we were to include a perfect measurement of an application's quality as a control, then lender behavior should not show any partial effects of applicant gender. (While our measurement is not perfect, we argue that the measurement error is small and

[9]For example, consider as a benchmark a lender who evaluates a safe 1-period investment of € 8000 (the empirical average of requested loans in our data set) that repays at an interest of 6.6% (the empirical average of realized internal rate of return). To investigate the potential importance of risk aversion, consider increasing the pay-out variance from zero to the *maximal* extent that is possible such that (i) the expected return is held constant and (ii) the possible return realizations are within the range observed in our data set. Under a log utility evaluation, such an increase in risk affects the investor's expected utility by less than half of what a one-percentage-point reduction in the safe interest rate (down to 5.6%) would inflict. In other words, even an unrealistically high degree of risk would have minuscule effects on lenders' expected utility.

correctable, see Subsection 4.3.) Deviations from these nulls allow multiple interpretations, including the possibility of taste-based discrimination or differential consideration of information about male versus female applicants. But they rule out statistical discrimination in combination with rational expectations.[10]
In specific, we ask: "Is there a significant correlation between the received funding and a loan applicant's gender?" (Hypothesis 1, on level discrimination) and "Does the correlation between expected return and received funding differ by the applicant gender?" (Hypothesis 2, on slope discrimination).
To measure a loan application's quality, we use the expected internal rate of return conditional on the available information, E(IRR). We describe this measure in detail in Subsection 4.2. The dependent variable of the analysis is a loan application's received funding share, categorized in the three relevant categories that the platform imposes, {category 1 - funding <25%, category 2 - funding $\geq$25% and <100%, category 3 - funding =100%}.[11] We fit their incidence in an ordered logit frame work, see Subsection 4.1). Using the β_X coefficient estimators of explanatory variable X in these regressions, we formulate our hypotheses.

Hypothesis 1. *A loan applicant's gender is not related to funding success, i.e.,* $\beta_{Gender} = 0$.

Hypothesis 2. *A loan application's interaction of expected return with the applicant's gender is not related to funding success, i.e.,* $\beta_{Gender*E(IRR)} = 0$.

4.1. Ordered Logit

In our econometric specification we model each loan application i as being of perceived utility U_i to the aggregate pool of investors at *smava.de*. We assume that the observable attributes of a loan application determine U_i according to

$$U_i = \beta' x_i + \epsilon_i \tag{1}$$

where x_i are loan application i's observable attributes, β' contains the coefficients of these attributes and ϵ_i is a random disturbance term (with a standard logistic distribution in our application).
To capture equation (1) econometrically, we use the ordered logit framework. We define a variable $Y_i \in \{1,2,3\}$ that measures the funded share of a loan application in three observable categories, as described above. The categorization

[10] Under the assumption that lenders aim at maximal *subjectively* expected returns, the deviations must stem from false subjective expectations that may depend on the applicant's gender. See Weizsäcker (2010) for analogous tests of rational expectations for different types of information sets in social learning experiments.

[11] We have more fine-grained information on the extent of funding but since almost 90% of the loan applications received either 100% of the requested amount or nothing at all, a finer analysis makes little difference.

depends on whether U_i passes a threshold κ_m, with $m = 1{,}2$. The parameter vector β is then to be estimated together with the thresholds, κ_m, via maximum likelihood with

$$P(Y_i > m) = \frac{exp(X_i\beta - \kappa_m)}{1 + exp(X_i\beta - \kappa_m)}. \tag{2}$$

4.2. Expected Return Calculation

A loan's expected internal rate of return (irr) is unobservable, both for us and lenders, and we therefore use the available loan and borrower specific information to come up with an estimate. For each loan there exist $T + 1$ different outcome scenarios, with T being the total number of monthly payments during the repayment period. The possible scenarios range from an immediate default before the first payment is due to a complete repayment as planned.[12]

Each of the possible outcome scenarios occurs with a certain probability. The probability that the borrower defaults in the first month is denoted $p_1 = Pr\{D = 1\}$, with D as a discrete random variable indicating the month of default, and analogously for defaults in other months, up to the best case scenario being $p_{T+1} = 1 - Pr\{D \leq T\}$. These probabilities are unknown but can be estimated based on past defaults of borrowers in our data set. We estimate the default risks of all months, $p_1 ... p_{T+1}$ with a discrete time hazard model, using as explanatory variables the observable characteristics of the borrower and the terms of the loan. With the help of this information, for each loan we calculate how likely each possible default scenario occurs.[13]

Next, we use the contractual repayment structure of *smava.de* to determine the monetary values of each possible outcome scenario. All loans are annuity loans, i.e., repayments consist of a principal repayment part and an interest payment part. Due to the collective insurance mechanism implemented by *smava.de* (described in Subsection 3.3), lenders receive an insured part of the principal repayments no matter whether the borrower defaults or not.[14] Overall, repayment of a loan that defaults in month D is

$$\text{Payoff}(D) = Rate_{pool} \times \sum_{t=1}^{T} Principal_t + \sum_{t=1}^{D-1} Interest_t, \tag{3}$$

with $Rate_{pool}$ as the insured fraction of the annuity repayments. Using the payoff we can solve for the irr for each of the possible outcome scenarios. It is given by

[12] It is not possible to leave out a certain monthly repayment and continue repaying later on. If a borrower defaults once, it's over.

[13] For further details see Appendix A.1.

[14] As described above, the degree of insurance depends on the investment pool that the investor belongs to. At the time of investment investors do not know the precise default rate in their pool, mainly as it may vary over time. *smava.de* publishes a continuously updated estimate of the pool-specific default rate and in our calculation we make the simplifying assumption that the ensured part is known to be constant at this level.

solving the condition for a break-even flow of payments,

$$\text{Investment} + \text{Fee} = \sum_{t=1}^{T} \frac{Payoff(D)_t}{(1+irr)^t}, \tag{4}$$

where Investment is the amount invested, Fee is the amount charged for using the platform and $Payoff(D)_t$ is the is the part of $Payoff(D)$ received in period t.

The *expected* internal rate of return to an investment, $E(irr)$, is the sum of the possible returns irr_t that would arise from the possible $T+1$ outcome scenarios weighted by the respective default probabilities, $p_1, ..., p_{T+1}$:

$$E(irr) = \sum_{t=1}^{T+1} p_t \times irr_t, \tag{5}$$

with E(IRR) $= 12 \times E(irr)$ as the annual return measure. Figure 1 illustrates the resulting density of E(IRR) over the posted loan applications in our data set.

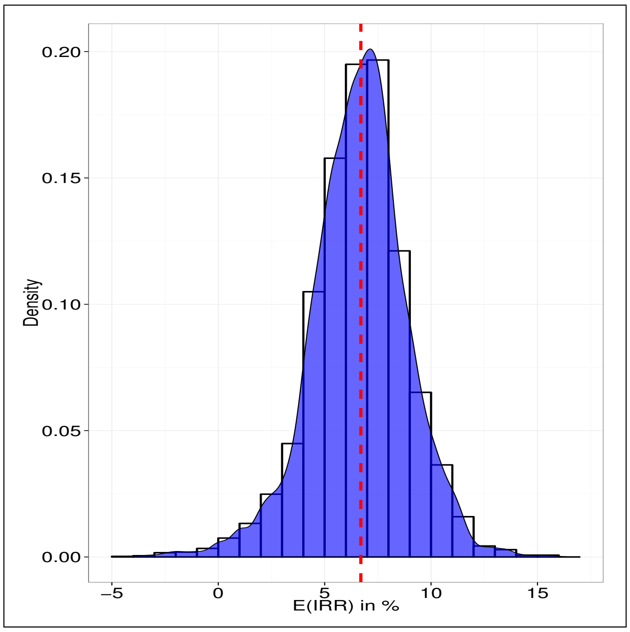

Figure 1: Density of E(IRR) over the posted loan applications in our data set, with the horizontal dashed line indicating the mean return at 6.6%.

4.3. Correcting for Measurement Error

Incorporating likely quantities of not directly observable variables in an econometric analysis comprises the risk of biased results, an unreliable coverage level of confidence intervals and a reduction of statistical power due to measurement error. To account for measurement error, the literature proposes different approaches;[15] the *simulation-extrapolation method* (SIMEX) developed by Cook and

[15]This includes the semi-parametric correction technique by Sepanski et al. (1994), the two-stage bootstrap method by Haukka (1995), the regression splines approach of

Stefanski (1994) is the most appropriate method for our setting, due to its robustness to distributional assumptions, its applicability to relatively small samples and its performance in several simulation studies (see, e.g., Later, Fung and Krewski (1999)). SIMEX allows retrieving asymptotically unbiased and efficient estimates for regression based models.

The method requires, importantly, an estimate of the measurement error variance (or, as in our case, of the estimation error variance), σ_u^2. Since we construct the quality measure ourselves, we know the measurement process which is (potentially) subject to estimation error and can use a Monte-Carlo approach to arrive at the estimate.[16]

Equipped with an estimate of the measurement error, the basic idea of the SIMEX method is fairly straightforward: if a causal relationship is biased by measurement error, then adding more measurement error should increase the degree of this bias. That is, the measurement error variance σ_u^2 is increased to $(1+\Gamma)\sigma_u^2$ where Γ controls the amount of added measurement error. By adding (or simulating) successive levels of measurement error, one can therefore estimate the empirical relationship between the expected value of the coefficient and Γ, and then extrapolate back to the unbiased estimate. That is, for different values of $\Gamma = (0.5, 1, 1.5, 2)$ we create $b = 1, ..., B$ pseudo data sets via simulations

$$\mathrm{E}(\mathrm{I\tilde{R}R})_{b,i} = \mathrm{E}(\mathrm{I\tilde{R}R})_i + \sqrt{\Gamma}\mathrm{Normal}(0, \sigma_u^2)_{b,i} \tag{6}$$

where $\mathrm{E}(\mathrm{I\tilde{R}R}) = \mathrm{E}(\mathrm{IRR}) + u$ and $u \sim N(0, \sigma_u^2)$, with E(IRR) as the true value. Then we refit the pseudo data to obtain the bth pseudo estimate

$$\hat{\theta}_b(\Gamma) = \hat{\theta}(\{Y_i, \mathrm{E}(\mathrm{I\tilde{R}R})_{b,i}\}_1^n) \tag{7}$$

and take the average over all B simulations

$$\hat{\theta}(\Gamma) = B^{-1} \sum_{b=1}^{B} \hat{\theta}_b(\Gamma). \tag{8}$$

We then write the estimates as a function of Γ and since $\Gamma = -1$ is the case of no measurement error, the extrapolating back to $\Gamma = -1$ estimates the parameter for this case.[17] The employed extrapolation function is a quadratic polynomial

Berry et al. (2002), the regression calibration method by Hardin et al. (2003), the so-called indirect method by Jiang and Turnbull (2004) and the adjusted estimator of Cameron and Trivedi (2005).

[16]We re-estimate our measure with different parameter specifications which we randomly sample from their confidence intervals that are estimated from our data set (see Appendix A.1). This accounts for the uncertainty in the employed parameter estimates, which might result in measurement error. The resulting data allows us to estimate the variance parameter(s) of the estimation error of our quality measure. We estimate the mean estimation error variance for the full sample and individually for each loan application. The reported results below use the former, full-sample estimator to run the SIMEX analysis. None of the conclusions change qualitatively if we use the individual-loan estimators.

[17]Although the SIMEX approach with its simulation character seems a natural fit for the bootstrap to obtain standard errors of the SIMEX-parameter estimates, for com-

which is usually the default setting in most statistical packages and has a good performance in many cases.

5. Results

Towards assessing Hypotheses 1 and 2, we first give a short data overview and then estimate the above-described ordinal logit model and apply the SIMEX correction.[18] Finally, we compare our choice of explanatory variables with natural other candidates.

5.1. Data Overview

Female loan applicants get on average 2.7% more of their requested loan amount funded then males (Two-sample t test, p-value<0.001). There are, however, many other characteristics of a loan application besides applicant gender that could explain this finding. Table 1 gives a descriptive overview of the main variables that we observe, separated by gender. It shows that female loan applicants offer on average higher interest rates, request lower amounts, have higher (better) Schufa ratings and are older than male applicants.

Loan/Borrower Characteristics	Female	Male
E(IRR) in %	6.70	6.60
Offered Loan Rate in %	10.14	9.78***
Loan Duration in months	50.34	49.92
Requested Loan amount in 1000€	7.47	8.17***
Schufa Rating in Scores from 1-8	4.51	4.36**
KDF Rating in Scores from 1-4	2.69	2.69
Borrower Age in Years	47.01	43.21***
Borrower Residence in North Germany in %	50.01	49.30
Length of Project Description in Letters	395.04	371.57**
No. of Observations	**1114**	**3030**
Lender Characteristics	Female	Male
Lender Age in Years	45.33	41.49***
Lender Residence in North Germany in %	30.08	32.69*
No. of Observations	**625**	**5046**

Significance levels: $*$: 10% $**$: < 5% $***$: < 1% (one-sided)

Table 1: Two-sample data average and t test results for average gender differences of borrowers and lenders.

plex models like ours it is simply not feasible. We instead use the jackknife method developed by Stefanski and Cook (1995) which has a much smaller computational burden and has shown to deliver valid estimates in simulations.

[18] The ordinal logit model is estimated with the *polr* procedure using the *mass* package in R and implemented in a modified version of the current *simex* R package (see, e.g., Lawrence (2009)). More information concerning the implementation in R available upon request.

5.2. Econometric Analysis

We specify three different econometric models and estimate each of them with and without measurement error correction by means of SIMEX. Model (1) simply controls for a loan applicant's gender with the dummy variable $borrower_{male}$ valued 1 if the respective person is male and zero otherwise. Model (2) additionally accounts for the loan application's expected return by means of the mean centered variable E(IRR) and its interaction effect with the gender variable ($borrower_{male}*$ E(IRR)). Model (3) adds further controls and their respective interactions E(IRR): the applicant's age $borrower_{age}$ (which is a mean centered continuous variable) and place of residence $borrower_{north}$ (a dummy variable valued 1 if the respective person lives in one of the northern part of Germany, and zero otherwise).[19] Moreover, Model (3) includes controls for all other variables that we were able to cleanly extract from out data set: the requested loan amount, the loan rate, the (categorical) purpose of the loan, the loan applicant's Schufa rating, the loan applicant's KDF indicator, the loan applicant's occupation, the length of the loan description and *smava.de*'s fee structure. Table 2 list the results.

[19]We define the two geographical areas such that the number of inhabitants and the economic output are approximately equal in both clusters. Federal states assigned to the northern part of Germany are Bremen, Hamburg, Berlin, Schleswig Holstein, Mecklenburg Western Pomerania, Saxony-Anhalt, Brandenburg, Lower Saxony, and Northrhine-Westphalia while the remaining federal states form the southern part.

Dependent Variable: $share\_funded$			
Explanatory Variable	(1)	(2)	(3)
$borrower_{male}$	-0.203**	-0.134	-0.166
	(0.09)	(0.09)	(0.11)
E(IRR)		0.049	0.169*
		(0.03)	(0.09)
$borrower_{male}*$ E(IRR)		0.185***	0.110**
		(0.04)	(0.04)
$borrower_{age}$			-0.015***
			(0.01)
$borrower_{age}*$ E(IRR)			-0.001
			(0.00)
$borrower_{north}$			0.116
			(0.09)
$borrower_{north}*$ E(IRR)			-0.033
			(0.04)
Fixed Effects	No	No	Yes
	SIMEX-Model(1)	SIMEX-Model(2)	SIMEX-Model(3)
$borrower_{male}$	-0.203**	-0.018	-0.192*
	(0.09)	(0.09)	(0.10)
E(IRR)		0.104*	0.503***
		(0.05)	(0.14)
$borrower_{male}*$ E(IRR)		0.502***	0.318***
		(0.06)	(0.06)
$borrower_{age}$			-0.013***
			(0.00)
$borrower_{age}*$ E(IRR)			-0.006***
			(0.00)
$borrower_{north}$			0.112
			(0.09)
$borrower_{north}*$ E(IRR)			-0.098
			(0.05)
Fixed Effects	No	No	Yes

Significance levels: $*$: 10% $**$: < 5% $***$: < 1%
Fixed effects and κ coefficients are reported in Table 3 in the appendix.

Table 2: Ordinal logit results with (below) and without (above) correction for measurement error via the SIMEX method.

Hypothesis 1 - level effect: All model specifications exhibit a negative coefficient for a loan applicant being male. However, only for the two Model (1) versions is this correlation strongly significant while for Model (2) and Model (3), at most a marginal significance is detected. In Model (1) it could obviously be the case that gender picks up the influence of other variables, like higher E(IRR), consistent with the gender differences reported in Table 1. Controlling for E(IRR) in the saturated Model (2) and adding various other controls in Model (3) does not change the the sign of the gender coefficient but reduces its significance. Overall, the results of model estimations with and without measurement error correction do not allow us to fully reject the null of Hypothesis 1 but at least cast serious doubt at its validity. Women appear to be at an advantage overall, in our data set.

Hypothesis 2 - slope effect: The evidence of a significant interaction between quality and gender is much stronger. For all of our model specifications does the coefficient for the interaction term between gender and E(IRR) show a positive sign at a strong significance both in the case of no measurement error correction and after a SIMEX application. The SIMEX correction increases the size of main effects and interaction effects of E(IRR).[20]

For Model (2), the gender difference in the importance of E(IRR) is so strong that the coefficient of women's E(IRR) is only marginally significant and vanishes if measurement error controls are omitted. In other words, the application's predictable quality is a much better indicator of the funding probability for men than for women. Thus, the regression results allow us to reject Hypothesis 2. A natural interpretation is that the effect reflects irrational expectations, as indicated in the literature of statistical discrimination and screening discrimination: The predominantly male lenders have a worse understanding of women than of men and therefore can identify low-quality men but not low-quality women.

Our data allow us further to check for slope effects of two other borrower characteristics with a discrimination potential: a borrower's age and place of residence. While the expected return is not evaluated differently for borrowers living in the north or south of Germany, a borrower's age seems to matter. The coefficient for the interaction with E(IRR) in SIMEX-Model (3) shows a significant negative effect. The size of the predicted age effect is, however, clearly smaller than the gender difference even if a change of one standard deviation in age (13.6 years) is considered. Further analogous estimations of SIMEX-Model (2) for the two variables confirm these findings (see Table 4 in Appendix B).

Using the predicted values from SIMEX-Model (2), Figure 2 graphically summarizes our findings regarding gender discrimination at *smava.de*. For loan applications in the +/- 5-percentage-point range around the average of expected returns, the likelihood of funding success and the E(IRR) measure correlate much stronger if the application was submitted by a man rather than by a woman. This difference vanishes for loan applications with expected returns exceeding the sample average by more than 5 percentage points. However, the relevance of this group for our analysis is limited since it accounts for only 1% of our data set.

[20]This is not surprising since the SIMEX approach eliminates some of the noise due to measurement error for the E(IRR) variable and thus decreases its bias towards zero.

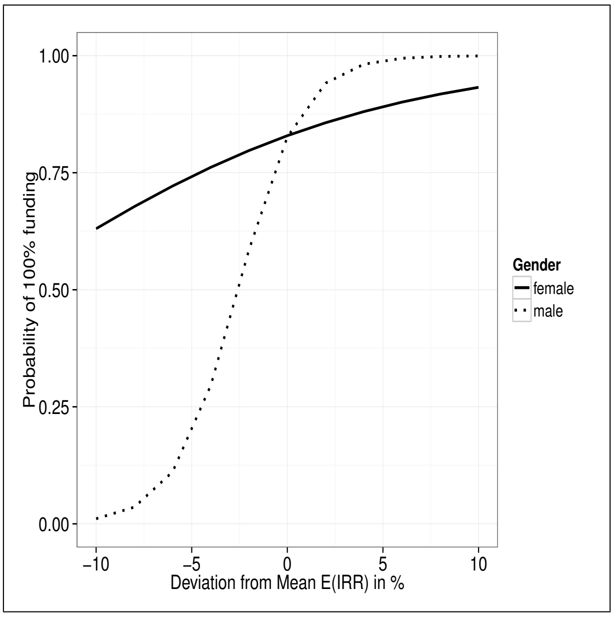

Figure 2: Predicted values for gender differences in the probability of full funding over expected return deviations from the mean.

5.3. Comparison of candidate proxies for quality

We close our empirical analysis by comparing our measure of quality, E(IRR), with other possible proxies of an application's quality. Would our conclusions change if one uses different to measures of applicant quality? The following figures indicate that the answer is affirmative. As a benchmark, Figure 3 depicts the predicted values of Model (2), i.e. using the same explanatory variables as in Figure 2 but without the SIMEX correction.[21] Comparisons of the two figures shows that both reveal the same qualitative insights but that the positive slope for men's E(IRR) is stronger with the SIMEX correction. In contrast, for Figure 4, the role of E(IRR) is taken by another natural candidate for a quality proxy, namely the offered loan rate that is contained in the description of an offer. With this measure of quality, one would arrive at the conclusion that females are, if anything, harmed by offering higher quality. The gender difference in the relation between quality and funding success would thus be even bigger than when using E(IRR) as a quality measure (coefficients: *loan_rate* (-0.002), *borrower_sex* (-0.505*), interaction term (0.030)) . Figure 4 uses yet another measure of quality, the Schufa credit rating. Here, we detect no slope difference across genders at all, only a level effect appears (coefficients: *Schufa* (-0.116**), *borrower_sex* (-0.141), interaction term (-0.016)).

In sum, the comparison of the different quality measures serves as a robustness check but also as a warning: while the rough pattern of results may be similar, the interpretation may change significantly depending on the choice of quality controls. Additionally, the significance of the gender level effect depends on the

[21] For the other two measures of quality that we study, the SIMEX correction is unavailable and we therefore take Figure 3 as a benchmark.

model specification. While the *borrower_sex* coefficient is negative for all three specifications only for the offered loan rate measure does it show to be significant. This makes it even more pressing to measure the quality or value of an application in the best possible way, and if possible account for the precision of the measurement.

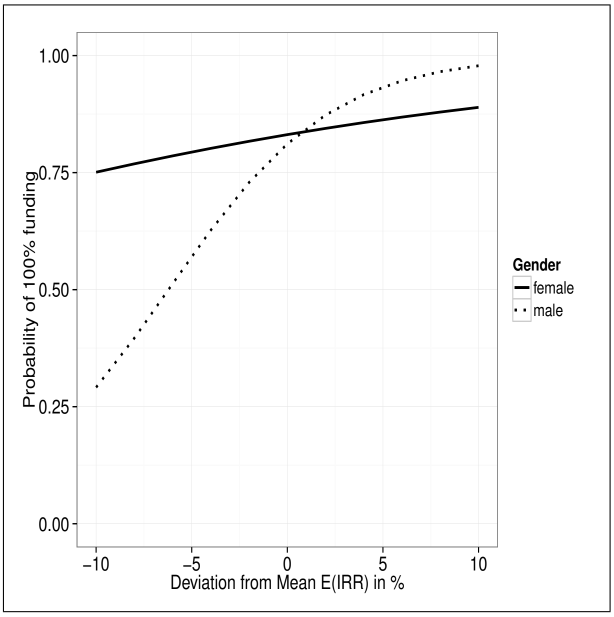

Figure 3: Predicted values for gender differences in the probability of full funding over expected return deviations from the mean.

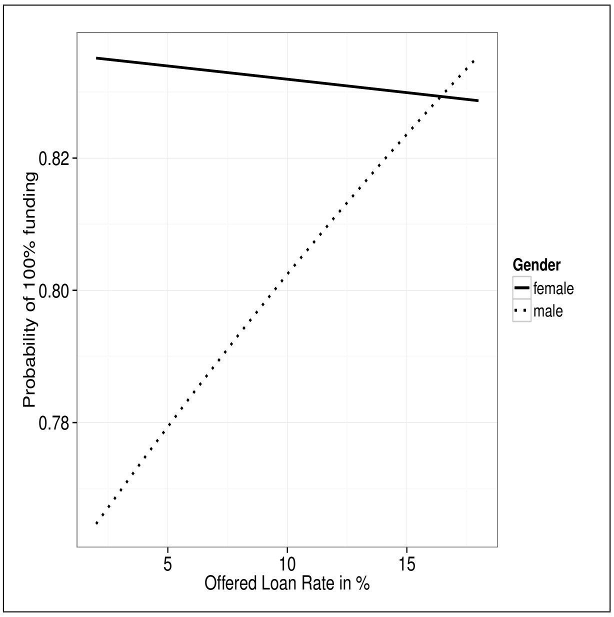

Figure 4: Predicted values for gender differences in the probability of full funding over offered loan rate.

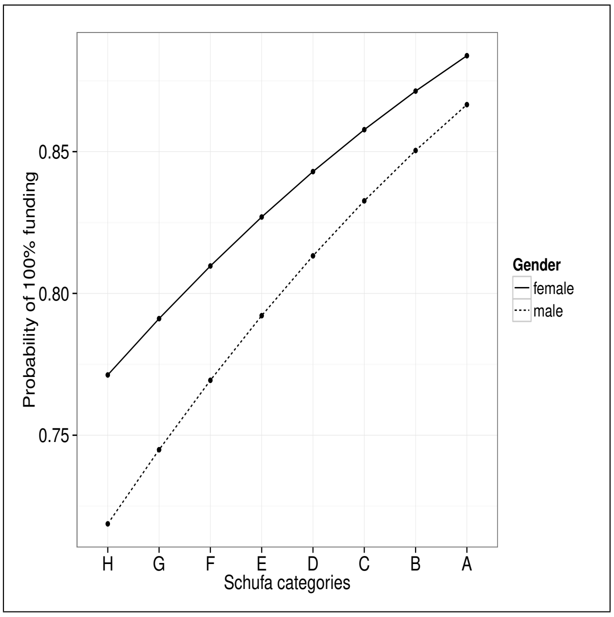

Figure 5: Predicted values for gender differences in the probability of full funding over Schufa rating.

6. Conclusion

The paper makes mainly methodological contributions in that it discusses in novel ways that the choice of proxy for applicant quality, and its measurement, play a role in the conclusions that one may draw from empirical analyses of discrimination contexts. Of course we are aware that the paper's substance, not the method, may be of primary interest to most readers. The finding that low-quality applications of men are penalized more than those of women has immediate analogues to many other contexts, and is in line with the straightforward predictions of statistical discrimination: men can judge men better than they can judge women and this is bad for weak male applicants.

Our choice of data context, peer-to-peer lending, has specific and noteworthy characteristics. First, the "rules of the game" are transparent, not only to the agents but also to the analyst. Especially the fact that the information conditions are highly controlled is a key advantage in order to detect gender differences while controlling for key information on the applications' qualities. Second, the context *is* very different from, e.g., labor market contexts in terms of the prevailing stereotypes and tastes regarding the two genders. Although it is hard to compare across settings, one may find it plausible that an investor who uses peer-to-peer lending has a much more positive view of female loan applicants, compared to an employer's view of female job applicants. We agree to this and clearly do not claim to have found a general women-favoring pattern in discrimination behavior.

References

[1] Ahmed, A.M., Andersson, L., and Hammarstedt, M. (2013). *Are Gay Men and Lesbians Discriminated against in the Hiring Process?* Southern Economic Journal, 79(3): 565-585.

[2] Aigner, D.J., and Cain, G.G. (1977). *Statistical Theories of Discrimination in Labor Markets*. Industrial and Labor Relations Review, 30(2): 175-187.

[3] Arrow, K. (1973). *The Theory of Discrimination*. Discrimination in Labor Markets, 3(10): 3-33.

[4] Azmat, G., and Ferrer, R. (2016). *Gender Gaps in Performance: Evidence from Young Lawyers*. Journal of Political Economy, Forthcoming.

[5] Bagues, M., and Perez-Villadoniga, M.J. (2013). *Why do I like people like me?* Journal of Economic Theory, 148(3): 1292-1299.

[6] Becker, G.S. (1957). *The Economics of Discrimination*. University of Chicago Press.

[7] Berry, S., Carroll, R.J., and Ruppert, D. (2000). *Bayesian Smoothing and Regression Splines for Measurement Error Problems*. Journal of the American Statistical Association, 97: 160-169.

[8] Bertrand, M., and Mullainathan, S. (2004). *Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*. American Economic Review, 94(4): 991-1013.

[9] Cameron, A.C., and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

[10] Carlsson, M., and Eriksson, S. (2014). *Discrimination in the Rental Market for Apartments*. Journal of Housing Economics, 23: 41-54.

[11] Cook, J.R., and Stefanski, L.A. (1994). *Simulation-Extrapolation Estimation in Parametric Measurement Error Models*. Journal of the American Statistical Association, 89: 1314-1328.

[12] Cornell, B., and Welch, I. (1996). *Culture, Information, and Screening Discrimination*. Journal of Political Economy, 104(3): 542-571.

[13] Fung, K.Y., and Krewski, D. (1999). *On measurement error adjustment methods in Poisson regression*. Environmetrics, 10(2): 213-224.

[14] Gallen, Y. (2016). *The Gender Productivity Gap*. Northwestern University Working Paper.

[15] Hardin, J.W., Schmiediche, H., and Carroll, R.J. (2003). *The Regression Calibration Method for Fitting Generalized Linear Models with Additive Measurement Error*. Stata Journal 3.4: 1-11.

[16] Hellerstein, J.K., Neumark, D., and Troske K.R. (1999). *Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production*. Journal of Labor Economics, 17(3): 409-446.

[17] Haukka, J.K. (1995). *Correction for covariate measurement error in generalized linear model - A bootstrap approach*. Biometrics, 51(3): 1127-1132.

[18] Kaas, L., and Manger, C. (2012). *Ethnic Discrimination in Germany's Labour Market: A Field Experiment*. German Economic Review, 13(1): 1-20.

[19] Kuhn, P., and Shen, K. (2013). *Gender Discrimination in Job Ads: Evidence from China*. The Quarterly Journal of Economics, 128(1): 287-336.

[20] Kricheli-Katz, T., and Regev, T. (2016). *How many cents on the dollar? Women and men in product markets*. Science Advances, 2(2): 1-8.

[21] Lawrence, C.N. (2009). *Accounting for the "Known Unknowns": Incorporating Uncertainty in Second-Stage Estimation*. Texas A&M International University Working Paper.

[22] Nunley, J. M., Pugh, A., Romero, N., and Seals, R.A. (2014). *An Examination of Racial Discrimination in the Labor Market for Recent College Graduates: Estimates from the Field*. Auburn University Working Paper.

[23] Pager, D. (2003). *The Mark of a Criminal Record*. American Journal of Sociology, 108(5): 937-975.

[24] Phelps, E.S. (1972). *The Statistical Theory of Racism and Sexism*. American Economic Review, 62(4): 659-661.

[25] Price, J., and Wolfers, J. (2010). *Racial Discrimination Among NBA Referees*. The Quarterly Journal of Economics, 125(4): 1859-1887.

[26] Riach, P.A., and Rich, J. (2010). *An Experimental Investigation of Age Discrimination in the English Labor Market*. Annals of Economics and Statistics, 99/100: 169-185.

[27] Rooth, D.O. (2009). *Obesity, Attractiveness, and Differential Treatment in Hiring: A Field Experiment*.Journal of Human Resources, 44(3): 710-735.

[28] Sepanski, J. H., Knickerbocker, R., and Carroll, R. J. (1994). *A Semiparametric Correction for Attenuation*. Journal of the American Statistical Association, 89: 1366-1373.

[29] Wright, B.R.E., Wallace, M., Bailey, J., and Hyde, A. (2013). *Religious Affiliation and Hiring Discrimination in New England: A Field Experiment*. Research in Social Stratification and Mobility, 34: 111-126.

[30] Wenxin, J., and Trunbull, B. (2004). *The Indirect Method: Inference Based on Intermediate Statistics - A Synthesis and Examples*. Statist. Sci. 19(2): 239-263.

[31] Weizsäcker, G. (2010). *Do We Follow Others When We Should? A Simple Test of Rational Expectations*. American Economic Review, 100(5): 2340-2360.

[32] Zussman, A. (2013). *Ethnic Discrimination: Lessons from the Israeli Online Market for Used Cars*. The Economic Journal, 123(572): 433-468.

A. Appendix A

A.1. Default Probability Estimation

A.1.1. Discrete Time Hazard Model

We denote D as the random variable indicating the time period of a loan's default, with values d_1<d_2<d_3<..., and write the default probabilities as $p(d_j) = Pr\{D = d_j\}$. We then define a loan's survival function at a specific d_j as the probability that the loan didn't default until $D \geq d_j$ resulting in $S(d_j) = Pr\{D \geq d_j\} = \sum_{D \geq d_j} p(d_j)$. Next, we define the discrete time hazard rate, $h(d_j)$, as a loan's conditional probability of defaulting at time d_j given the loan hasn't defaulted up to that point, so that

$$h(d_j) = Pr\{D = d_j | D \geq d_j\} = \frac{p(d_j)}{S(d_j)}. \tag{9}$$

To estimate the discrete time hazard rate conditional on our observable loan characteristics and the current time period, we follow Cox (1972) and use the logistic hazard model approach. The model is fitted by running a logistic regression on a set of pseudo observations generated from our original data set. That is, we generate default indicators y_{ij} that are valued 1 if loan i defaulted at time j and zero other wise. This results in a number of default indicators per loan that is equal to the number of actually observed monthly payments of that loan. Each of these generated indicators is merged with the loan specific covariates, $\mathbf{x_i}$, and a consecutively numbered time index j. Since $Pr\{y_{ij} = 1 | D \geq d_j\} = Pr\{D = d_j | D \geq d_j\} = h_i(d_j)$ the hazard rate can be estimated with the help of the usual maximum likelihood procedures in a binary response data case, with the log likelihood equal to:

$$\text{logL} = \sum_{i=1}^{N} \sum_{j=1}^{D} \left[y_{ij} \log(h_i(d_j, X_i)) + (1 - y_{ij}) \log(1 - h_i(d_j, X_i)) \right], \tag{10}$$

where the hazard rate is assumed to have a logistic functional form. The loan specific characteristics such as requested loan amount, offered loan rate, loan duration, loan purpose as well as the borrower characteristics age, gender, schufa rating, financial burden, employment status and place of residence are captured by X_i. Using the predicted values for $\hat{h}_i(d_j, X_i)$ we are able to calculate the survival function

$$\hat{S}(d_j, X_i) = \prod_{j=1}^{d_j - 1} (1 - \hat{h}_i(d_j, X_i)) \tag{11}$$

and with the help of Equation (9) we can solve for the default probabilities, $\hat{p}(d_j)$.

A.1.2. Estimation Results

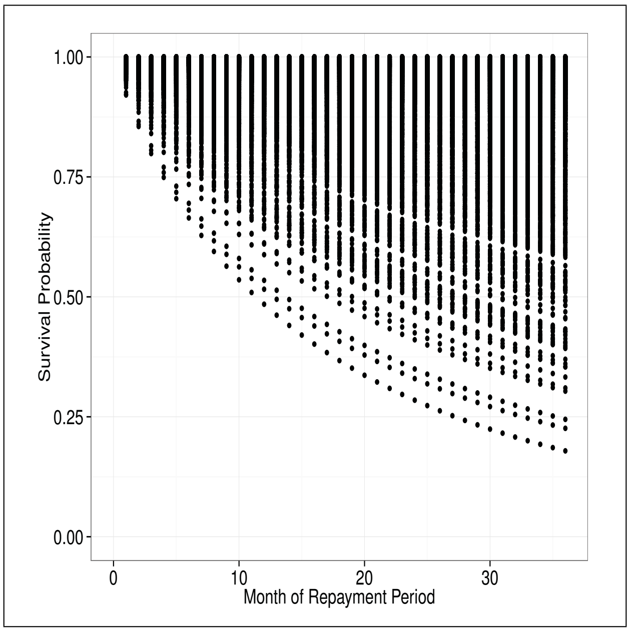

Figure 6: Evolution paths of the survival probability over time for all loans with a duration of 36 months.

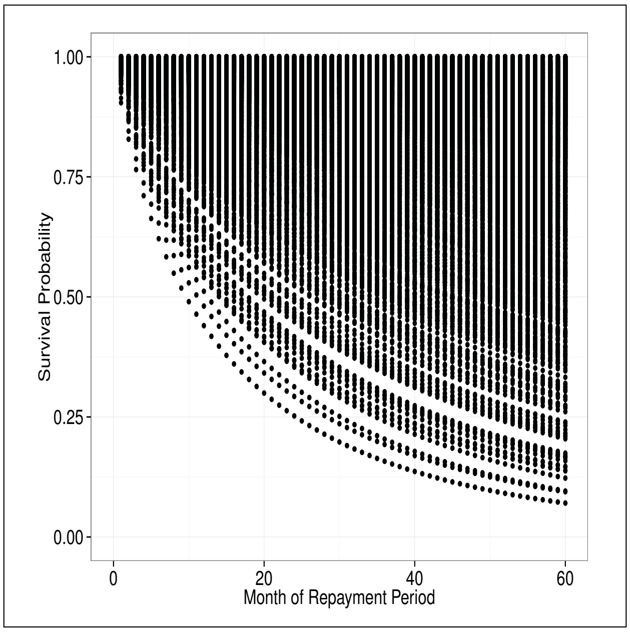

Figure 7: Evolution paths of the survival probability over time for all loans with a duration of 60 months.

B. Appendix B

B.1. Additional Tables

Dependent Variable: *share_funded*		
Explanatory Variable	Model (3)	SIMEX-Model(3)
loan_size	-0.031***	-0.031***
	(0.02)	(0.02)
*job*1	0.087	0.100
	(0.22)	(0.31)
*job*2	0.062	0.005
	(0.32)	(0.28)
*job*3	0.061	-0.049
	(0.25)	(0.21)
*job*4	-0.166	-0.079
	(0.30)	(0.22)
*job*5	-0.050	0.024
	(0.23)	(0.15)
schufaB	-0.683***	-0.594***
	(0.18)	(0.15)
schufaC	-0.245	0.074
	(0.22)	(0.25)
schufaD	-1.196***	-0.898***
	(0.22)	(0.26)
schufaE	-1.244***	-0.861***
	(0.23)	(0.27)
schufaF	-1.706***	-1.107***
	(0.24)	(0.35)
schufaG	-2.236***	-1.291***
	(0.30)	(0.34)
schufaH	-2.640***	-1.305***
	(0.39)	(0.41)
*kdf*2	-0.518***	-0.648***
	(0.15)	(0.41)
*kdf*3	-1.211***	-1.358***
	(0.15)	(0.17)
*kdf*4	-2.188***	-2.417***
	(0.16)	(0.19)
loan_rate	0.178***	-0.006***
	(0.04)	(0.02)
loan_duration_long	-0.076***	-0.970***
	(0.12)	(0.16)
description_length	0.000***	0.000***
	(0.00)	(0.00)
business_loan	0.388***	0.287***
	(0.12)	(0.11)
fee_new	0.274**	0.273**
	(0.13)	(0.14)
κ_1	-3.298	-4.839
	(0.62)	(0.84)
κ_2	-2.604	-4.140
	(0.62)	(0.83)
Significance levels: $*$: 10%	$**$: < 5%	$***$: < 1%

Table 3: Fixed effects.

	SIMEX-Model(1)	SIMEX-Model(2)	SIMEX-Model(3)
E(IRR)	0.104*	0.679***	0.583***
	(0.05)	(0.21)	(0.19)
$borrower_{male}$	-0.018		
	(0.09)		
$borrower_{male}*$ E(IRR)	0.502***		
	(0.06)		
$borrower_{age}$		-0.016***	
		(0.00)	
$borrower_{age}*$ E(IRR)		-0.010***	
		(0.00)	
$borrower_{north}$			0.010
			(0.09)
$borrower_{north}*$ E(IRR)			-0.181***
			(0.06)
Fixed Effects	No	No	No

Significance levels: $*$: 10% $**$: < 5% $***$: < 1%

Table 4: Ordinal logit results with correction for measurement error via the SIMEX method.