CEBRIG
CENTRE EMILE BERNHEIM
DE RECHERCHE INTERDISCIPLINAIRE EN GESTION

ULB Solvay Brussels School
Economics & Management

How to identify lending bias when the lender's goal is not profit?

Anastasia Cozarenco & Ariane Szafarz

Outcome tests for discrimination in the credit market typically assume that profit is the lender's goal. This assumption ignores nonprofit and social lending institutions that value prosocial outcomes. These institutions may combine positive and negative discrimination, further complicating the identification of bias. We propose a test for discrimination in lending that is robust to the profit orientation of the lender. Consistent with the Basel framework for credit risk management, our test is based on recovery records. It is applicable to the identification of both positive and negative bias.

Keywords Test for credit discrimination, prosocial lender, social finance, discrimination, affirmative action.

JEL codes C12, C44, J15, J16, G21, D63

Acknowledgments This study was carried out within the framework of the Chair in Social and Sustainable Finance founded by Montpellier Business School in partnership with BNP Paribas and the Caisse d'Epargne Languedoc-Roussillon. The authors thank Cécile Abramowicz, Sylvain Abramowicz, Moez Bennouri, Renaud Bourlès, Jean-Charles Rochet, Patrick Sentis, Ilan Tojerow, and participants to the European Research Conference on Microfinance, the Financial Engineering and Banking Society International Conference, the ASSA Meetings, and the World Finance Conference for their helpful comments and Roxanne Powell for excellent copy-editing. Declarations of interest: none.

CEBRIG Working Paper N°24-007
March 2024

ULB

Université Libre de Bruxelles - Solvay Brussels School of Economics and Management
Centre Emile Bernheim de Recherche Interdisciplinaire en Gestion
ULB CP114/03 50, avenue F.D. Roosevelt 1050 Brussels BELGIUM
cebrig@ulb.be - Tel : +32 (0)2/650.48.64

ULB

How to identify lending bias when the lender's goal is not profit?*

Anastasia Cozarenco
Montpellier Business School and Centre for European Research in Microfinance (CERMi)

Ariane Szafarz
Université Libre de Bruxelles (ULB), SBS-EM, CEBRIG and CERMi

February 2024

Abstract

Outcome tests for discrimination in the credit market typically assume that profit is the lender's goal. This assumption ignores nonprofit and social lending institutions that value prosocial outcomes. These institutions may combine positive and negative discrimination, further complicating the identification of bias. We propose a test for discrimination in lending that is robust to the profit orientation of the lender. Consistent with the Basel framework for credit risk management, our test is based on recovery records. It is applicable to the identification of both positive and negative bias.

Keywords*:* Test for credit discrimination, prosocial lender, social finance, discrimination, affirmative action.

JEL Classifications*:* C12, C44, J15, J16, G21, D63.

* Cozarenco: Montpellier Business School, Montpellier Business School, 2300, avenue des Moulins, 34185, Montpellier cedex 4; France (e-mail: a.cozarenco@montpellier-bs.com); Szafarz: Université Libre de Bruxelles (ULB), SBS-EM, CEBRIG and CERMi, 50, Av. Franklin Roosevelt, 1050, Brussels, Belgium (e-mail: ariane.szafarz@ulb.be). This study was carried out within the framework of the Chair in Social and Sustainable Finance founded by Montpellier Business School in partnership with BNP Paribas and the Caisse d'Epargne Languedoc-Roussillon. The authors thank Cécile Abramowicz, Sylvain Abramowicz, Moez Bennouri, Renaud Bourlès, Jean-Charles Rochet, Patrick Sentis, Ilan Tojerow, and participants to the European Research Conference on Microfinance, the Financial Engineering and Banking Society International Conference, the ASSA Meetings, and the World Finance Conference for their helpful comments and Roxanne Powell for excellent copy-editing. Declarations of interest: none.

1. Introduction

Assessing discrimination in the credit market is difficult for reasons related to the ambiguity of theoretical and legal definitions (Dymski, 2006), which conditions the way in which the discrimination is framed econometrically.[1] Testing methods for credit discrimination remain controversial, and data limitations often lead researchers to use inappropriate methods (Ross, 2002). Recently developed outcome tests are based on the profitability of marginal applicants (Dobbie et al., 2021). Following Becker's (1993) argument that harsher treatment of minority loan applicants should result in higher lender profits from this group, the outcome tests rely on the assumption that profit is the lender's objective. This assumption is restrictive in two ways. First, there are discriminatory practices that do not result in profit losses for the lender but are harmful to rejected applicants (Ferguson & Peters, 2000). Second, for providers of prosocial finance, the purpose is not purely financial (Wry & Zhao, 2018; Cornée & Szafarz, 2014).

This paper proposes a new test procedure based on defaults rather than profits. It improves the methodology of testing for discrimination in lending, thanks to an empirical design that helps navigate the intricacies of lenders' attitudes toward demographic categories. By using repayment records rather than profits to test for discrimination in lending, we operationalize the early theoretical intuition of Ferguson and Peters (1995) (hereafter F&P (1995)) that credit discrimination is detected when an identified subset of applicants suffers from both a lower (or equal) default rate and a higher (or equal) denial rate, provided that at least one inequality is strict. Replacing profit with the lender's recovery rate mimics the practice of using credit scoring models (Roszbach, 2004), which makes sense regardless of the lender's profit orientation. From an econometric point of view, accounting for sample selection is key to studying inequalities between groups (Maasoumi & Wang, 2017). Our novel approach has the advantage of addressing the endogeneity associated with the borrower selection by using data on applicants rather than just clients. The infra-marginality problem is addressed by a (testable) first-order stochastic dominance.[2]

---

[1] In the US, credit discrimination is a crime. The country's legal framework includes the Fair Housing Act of 1968, the Equal Credit Opportunity Act of 1974, and the Home Mortgage Disclosure Act of 1975. Under these laws, lenders must treat all borrowers equally with respect to protected characteristics, such as race and gender. Since 1989, U.S. lenders have been required to report the race and ethnicity of their applicants. This general principle provides scholars with many testing opportunities. Recent contributions on gender discrimination in lending include, for example, Beck et al. (2018), Cozarenco and Szafarz (2018), and Delis et al. (2022).

[2] The infra-marginality problem (Ayres, 2002) arises when groups of applicants have different risk distributions, leading to biased estimates. Scholars typically address this problem using instrumental variables (Arnold et al., 2018). Dobbie et al. (2021) use a two-stage least-squares (2SLS) estimation method in which loan take-up is instrumented by the loan officer's leniency level and a significantly positive effect on profit for one group is

2. Discrimination Tests for Prosocial Lending

Economists typically portray discrimination as a double standard that cannot be justified by the lender's profit maximizing objective. This approach ignores ethical lenders, who are typically motivated by social justice rather than profit maximization (D'Espallier et al., 2021; Adbi, 2023). A broader view of discrimination in the credit market can be obtained by extending the outcome to any socioeconomic motivation, including contributions to the common good. According to this logic, lenders discriminate if their lending to a certain category of applicants is more restrictive without being justified by their social and/or financial mission.

Negative bias in lending is not necessarily intentional, as stereotyping is a common human trait (Nelson, 2014; Bordalo et al., 2016). Depending on the context, positive bias (or affirmative action) can be interpreted as taste-based, statistical, or both. Taste-based bias could result from prosocial lenders who are willing to disregard information about repayment and focus solely on group membership. By contrast, prosocial lenders using statistical discrimination might favor groups with lower credit scores, which are interpreted as a signal of economic hardship. Disentangling these two types of positive bias is difficult. Discrimination implies disparate treatment, but the reverse is not true: Disparate treatment may well be justified by objective credit risk characteristics. If women entrepreneurs were a higher credit risk than men, all else being equal, then differential treatment would be the lender's rational response and would not be considered gender discrimination. Since repayment records are needed to assess credit risk, researchers who do not take them into account have no other choice but to rely on the (strong) assumption of equal creditworthiness across all tested characteristics.

In sum, the main problem plaguing empirical research on credit discrimination is that lender's assessment of creditworthiness is a black box for researchers (Cornée, 2019; Becchetti & Conzo, 2011). One way to identify bias in credit allocation would be to conduct an experiment that holds loan application characteristics fixed while varying gender, as in Fay and Williams (1993) and Brock and De Haas (2023), or to exploit the exogenous variation in credit allocation provided by an explicit staff rotation policy, as in Fisman et al. (2020). Another way

taken as evidence of bias against that group. F&P (1995, p. 744) already had the infra-marginality problem in mind (but not by that name), claiming that the distinction between marginal and average borrowers is key, and by subsequently defining discrimination in lending as "the use of different credit standards across the two components of the population, i.e. a policy that leads to the marginal borrower from each component of the population having a different credit score." This definition is close to the following statement by Ayres (2002, p. 135): "In the mortgage context, a test of disparate treatment would want to ask whether the least qualified whites to whom banks were willing to lend had a higher default rate than the least qualified minorities to whom banks were willing to lend."

would be to have access to the detailed decision process of the lender, and this is the approach we have taken. It is based on regressions explaining the lender's decision, controlling for as many covariates as possible. This approach is reliable if all the relevant variables considered by the lender are also considered by the researcher, as in the situation where there is no personal contact between the lender and the loan applicant.

2.1. Building on the Ferguson and Peters Model

The early theory developed by F&P (1995) addresses the legitimate criticism that tests based only on loan approval fail to account for differences in creditworthiness across the applicant groups. The innovation lies in the introduction of loan default as a second variable of interest in addition to loan approval. In this respect, the F&P (1995) model can be seen as a pioneer in implementing the outcome-testing approach to lending discrimination (Becker, 1971). As F&P put it: "If differences in average credit quality explain differences in denial rates, then equal default rates imply discrimination against the minority population" (F&P, 1995, p. 748). Despite the innovative vision of the F&P article, its econometric implementation remained stalled for a very long time, during which the literature took different paths. This paper aims to fill this gap.

In the F&P model, a bank decides to approve or deny a loan based on the applicant's credit score, which is assumed to be an increasing function of the probability of repayment $\theta \in [0,1]$. A uniform credit policy is defined by a threshold θ^* such that credit is granted to applicants with $\theta \geq \theta^*$, otherwise credit is denied. To simplify the model, F&P assume that all borrowers are charged the same interest rate[3] and that the size of all loans is one dollar. The entire population is characterized by the probability density function $f(\theta)$ and its cumulative distribution function $F(\theta)$. The bank's policy divides this population into successful applicants (i.e., borrowers) and rejected applicants. The borrowers have a mean value of θ, denoted by $\hat{\theta}$, greater than θ^*, and:

$$\hat{\theta} = \int_{\theta^*}^{1} \frac{\theta f(\theta) d\theta}{1 - F(\theta^*)}$$

[3] The authors argue that this assumption is consistent with the mortgage lending practice. In a more general setting, the assumption should impose equality of all loan terms (interest rate, collateral, maturity, etc.). Moreover, defaulted loans may be partially repaid (before the default occurs), which explains why the next section uses the (continuous) recovery rate rather than the (binary) default.

Since an applicant's probability of default is $(1-\theta)$ and the denial rate in the entire population is $F(\theta^*)$, the expected default rate among borrowers is $(1-\hat{\theta})$. Suppose the pool of applicants is divided into two groups with different probability distributions of θ, and let $f_k(\theta)$ and $F_k(\theta)$ be the probability density function and the cumulative distribution function of group $k, k \in \{1,2\}$, respectively. Paying attention to the entire probability distributions meets a key concern in addressing differences between groups (Nelson, 2015).

For simplicity, we order the groups according to their average probability of repayment, which is thus higher for group 1 than for group 2. In addition, F&P (1995) assume that the cumulative distribution function of group 1 first-order stochastically dominates (FOSD) that of group 2, which means that:

$$\forall \theta \in [0,1]: F_1(\theta) \leq F_2(\theta).$$

In this setting, a fair (i.e., nondiscriminatory) bank uses a uniform credit standard θ^* across the two groups, so that applicants in group 1 are more likely to be approved for a loan than applicants in group 2 (see Figure 1), but the marginal borrower in each group has the same credit score. In contrast, if the bank has a "taste for discrimination," as coined by Becker (1993), its sensitivity to group membership goes beyond credit scores. The double standard is evidenced by the use of group-specific values of θ^*. While these values are not directly observable from data on loan denials and default rates, F&P show that some circumstances are sufficient to detect discriminatory lending: either a higher denial rate for group 1 (group 1 is discriminated against) or a lower default rate for group 2 (group 2 is discriminated against). In the remaining situations, inference cannot simply resolve the single versus double standard question.

The FOSD assumption is key to the relevance of the F&P model because it allows the assessment of discrimination in lending with average repayment probabilities ($\hat{\theta}_1 > \hat{\theta}_2$), and thus avoids the problem of infra-marginality.[4] The next section discusses and illustrates this point.

[4] Recent studies on discrimination in lending pay particular attention to the infra-marginality problem (Dobbie et al., 2021). In a different context (racial discrimination against drivers stopped by the police), Simoiu et al. (2017) develop a threshold test that makes the FOSD assumption unnecessary but the translation of their approach to the credit market remains unfulfilled.

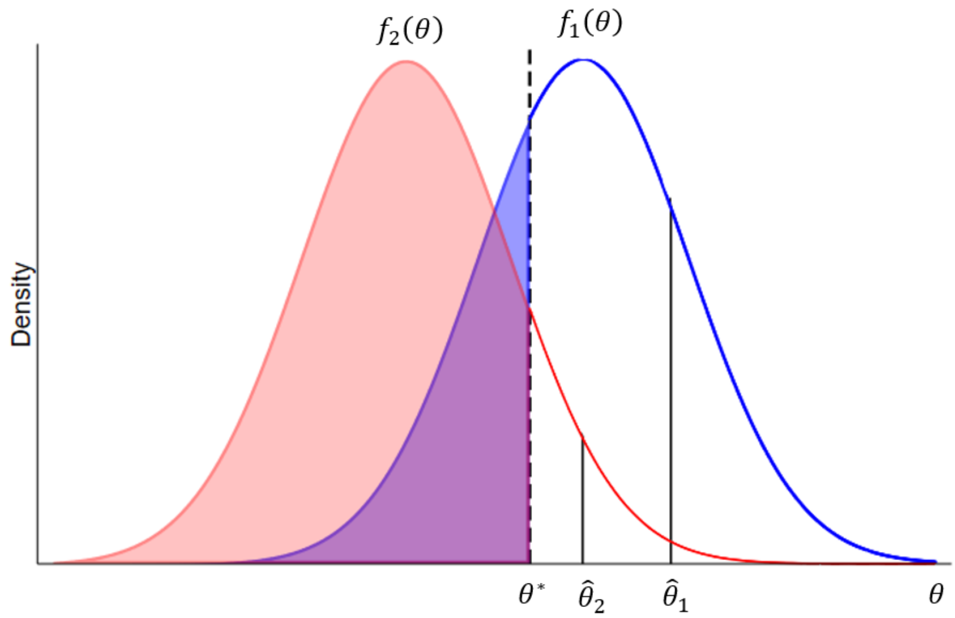

Figure 1. Two groups of applicants with different probability distributions

Notes: The applicants in Group 1 and Group 2 have the density functions $f_1(\theta)$ and $f_2(\theta)$, respectively, where θ is the probability of repayment and the cumulative distribution function of Group 1 first-order stochastically dominates that of Group 2. θ^* represents a uniform lending policy threshold: Loans are granted to applicants with $\theta \geq \theta^*$. $\hat{\theta}_1$ and $\hat{\theta}_2$ are the average repayment rates of the borrowers in Group 1 and Group 2, respectively. For each group, the surface of the shaded area measures the probability of denial.
Source: Adapted from Ferguson and Peters (1995).

2.2 Econometric Implementation

First, consider an unbiased, risk-neutral lender who makes fixed-term loans. For each applicant, this lender has two options: Either the loan is denied, and the future return is zero, or the loan is approved, and the future cash flow depends on the outcome of the loan. If there is no default, the payoff is $LS(1+r)$, where r is the interest rate charged and LS is the loan amount. Default, on the other hand, results in a loss to the lender. The loss given default (LGD) is equal to the write-off of the debt.

In line with F&P (1995), we assume that an unbiased lender forms expectations about the recovery rate rationally, which rules out any discriminatory loan allocation that could be attributed to biased expectations. The lender's decision boils down to approving the loan if the present value of $E[(1+r)LS - LGD]$ is greater than LS, where LGD is the only random component of future cash flows. To formalize this decision rule, we use the relative measure known as the recovery rate:

(1) $$Recovery\ rate = \frac{LS(1+r) - LGD}{LS}$$

Equation (1) gives the following rule:

(2) $$Loan\ Approval = 1 \iff E[Recovery\ rate] \geq 1$$

Equation (2) dictates that any applicant characteristic that is a positive factor for the recovery rate should increase the probability of loan approval. The recovery rate is a key variable in credit scoring. Lenders typically address information asymmetries by assessing the creditworthiness of applicants using credit scoring techniques. In line with the Basel framework for credit risk management, scoring models are based on recovery records (Shaffer, 1996; Boyes et al., 1989). These models assign credit risk levels to customer segments so that applicants who fall into categories with lower estimated creditworthiness are more likely to be denied credit, controlling for personal credit history. Using the recovery rate as the outcome variable is robust to contextual features, such as loan characteristics (duration, collateral, and interest rate) and the legal status of the lender (for-profit, public, nonprofit, or hybrid). Since the recovery rate can only be observed for approved applicants, we need to address the endogeneity concern arising from this selection bias by applying the Heckman (1979) estimation method.

Our two-equation model is used to test for biased loan originators:

(3) $$Recovery\ rate_i = \alpha_R F_i + \beta_{\boldsymbol{R}}' \boldsymbol{X_i} + \varepsilon_i$$

(4) $$Approval_i = \mathbb{1}[\alpha_A F_i + \beta_{\boldsymbol{A}}' \boldsymbol{Z_i} + v_i > 0]$$

where $v \sim N(0,1)$ and $E(\varepsilon|v) \neq 0$. Index i refers to loan applicants; F_i takes the value of 1 if applicant i is a woman and 0 otherwise. The control variables in the vector X include the characteristics of the applicant, while the vector Z is obtained by stacking X with an instrument that affects the approval decision but not the recovery rate, as required by the Heckman estimation method.[5] In practice, we estimate two equations—one for the recovery rate and the other for the approval probability—and compare the signs of the coefficients of interest in the two equations.

According to F&P, there are two situations in which discrimination in lending can be identified. In the first, a group faces a negative bias from the lender if it has a lower relative default rate and a higher or equal relative denial rate. In our setting, this configuration corresponds to $\alpha_R > 0$ and $\alpha_A \leq 0$. The second situation corresponds to a positive bias associated with the combination of a higher or equal relative default rate and a lower relative denial rate, identified as $\alpha_R \leq 0$ and $\alpha_A > 0$. The F&P model leads to the decision rule in Table

[5] When an instrument is difficult to find, the Honoré and Hu (2022) approach for models without exclusion restrictions (i.e., $\boldsymbol{X} = \boldsymbol{Z}$) is a fruitful alternative.

1, regardless of whether women are in group 1 or group 2. This table will guide the practical implementation of our testing procedure. For simplicity, we refer to the group under investigation as "women". The proposed classification detects both negative bias against women and positive bias in favor of women thanks to the signs of the tested parameters, α_R and α_A. This is particularly useful in the context of prosocial lending where the lender may wish to target specific disadvantaged populations.

As an econometric refinement of the F&P model, we refer to strong or weak bias depending on the number of strict inequalities (at a given level of significance). Strong bias means that both criteria are significant (e.g., lower approval rate and higher recovery rate), while weak bias indicates only one significant strict inequality (e.g., higher denial rate with the same recovery rate). The classification in Table 1 detects both negative bias against women and positive bias in favor of women thanks to the signs of the tested parameters, α_R and α_A.

Table 1. Detecting bias in lending with bivariate estimation

	Higher approval rate for women: $\boldsymbol{\alpha_A} > 0$	**Insignificant difference between approval rates:** $\boldsymbol{\alpha_A = 0}$	**Lower approval rate for women:** $\boldsymbol{\alpha_A} < 0$
Higher recovery rate for women: $\boldsymbol{\alpha_R} > 0$	No bias detected	Weak negative bias	Strong negative bias
Insignificant difference between recovery rates: $\boldsymbol{\alpha_R = 0}$	Weak positive bias	No bias detected	Weak negative bias
Lower recovery rate for women: $\boldsymbol{\alpha_R} < 0$	Strong positive bias	Weak positive bias	No bias detected

Notes: This table presents the decision rule of the discrimination test based on the Heckman estimation of the recovery rate and the loan approval rate. α_R and α_A are the coefficients of the gender dummy, which takes the value of 1 for women, in Equations (3) and (4), respectively.

The Heckman estimator is particularly well suited for inference based on observations of denial and default rates. It is consistent with the intuition that biased lending works by making it more difficult for applicants with a particular characteristic that has no effect on objective creditworthiness to obtain credit. The proposed rule subsumes this intuition by addressing both positive and negative forms of discrimination. However, its validity depends on the FOSD assumption. In the real world, this working assumption may or may not be true.[6] Therefore, to

[6] Appendix A illustrates the importance of the FOSD requirement with an example where FOSD is not met and the conclusions in Table 1 no longer hold.

consolidate the results, we introduce a final step and put FOSD to the test. To do this, we estimate the probability densities of the predicted recovery rates using a kernel method and perform a commonly used Kolmogorov–Smirnov test (McFadden, 1989). First, we compute the predicted recovery rate for applicant i:

(5) $$\hat{y}_i = \hat{\alpha}_R F_i + \hat{\beta}_{\boldsymbol{R}}' \boldsymbol{X_i}$$

where $\hat{\alpha}_R$ and $\hat{\beta}_R$ are taken from Equation (3). Then, we use the Epanechnikov (1969) kernel, which is known for its efficiency in smoothing probability density distributions (Zucchini et al., 2003). Finally, we apply the Kolmogorov–Smirnov test to assess the validity of FOSD with the following statistic:

(6) $$D = \max_{\hat{y}}\{G_1(\hat{y}) - G_2(\hat{y})\}$$

where $G_1(\hat{y})$ and $G_2(\hat{y})$ are the empirical cumulative distribution functions for the n_1 female applicants and the n_2 male applicants, respectively. The p-values of the test statistic are obtained from the following closed-form asymptotic distribution (Smirnov, 1933):

(7) $$\lim_{\{n_1,n_2\to\infty\}} \Pr\left\{\sqrt{\frac{\mathrm{n}_1 n_2}{\mathrm{n}_1 + n_2}}\, D_{\{n_1,n_2\}} \leq z\right\} = 1 - 2\sum_{l=1}^{\infty}(-1)^{l-1}\exp\left(-2l^2z^2\right)$$

3. Conclusion

The persistence of bigoted behavior in economic and financial markets suggests that, while competition may mitigate discrimination, it may not be able to crowd out all discriminatory biases. One possible reason may be that some biases do not pay off (Méon & Szafarz, 2011). In this case, biased loan denials fly under the radar of tests based on measures of profitability, because they hurt some applicants without affecting the lender's profits. Another example is prosocial finance, where lenders prioritize a social mission over pure profit. Therefore, profit is not the appropriate lens through which to test for bias in ethical lending and we suggest using recovery rates instead.

Our approach is particularly appropriate when there is no face-to-face interaction between the lender and the borrower. These interactions typically generate "soft information" that is hidden from the researcher, who in turn may incorrectly attribute differences to discrimination rather than unobservable factors (Brock et al., 2012). Therefore, a promising area of application for our testing method concerns discrimination in fintech lending (Fuster et al., 2022). Lending technology for financial inclusion is rapidly evolving with artificial intelligence tools that are prone to gender bias (Chen et al., 2023; Pethig & Kroenung, 2023). Our method advances the research agenda on algorithmic bias detection and the mere legitimacy of using machine learning and algorithmic financial scoring in lending.

Using the recovery rate as the outcome variable has advantages and disadvantages. The two main advantages are its robustness with respect to the lender's objective and its widespread use by lenders worldwide. By being agnostic to the lender's objective, the recovery rate acts as a "one size fits all" outcome, which is clearly more appropriate than profit when dealing with prosocial lenders. In addition, the recovery rate is routinely used as a performance measure in banking studies. Its generality can act as a safeguard against the temptation to manipulate a narrower outcome criterion. On the other hand, the broad spectrum of the recovery rate makes it a relatively crude indicator of performance. Precision could be gained by narrowing the outcome targeted by the lender under review. In practice, this is rarely done because prosocial outcomes are much harder to identify, let alone measure (Värendh Månsson et al., 2020). Therefore, a general, robust method such as the one presented in this paper has significant practical advantages.

References

Adbi, A., 2023. Financial sustainability of for-profit versus non-profit microfinance organizations following a scandal. Journal of Business Ethics 188, 57–74.

Arnold, D., Dobbie, W., Yang, C.S., 2018. Racial bias in bail decisions. Quarterly Journal of Economics 133, 1885-1932.

Ayres, I., 2002. Outcome tests of racial disparities in police practices. Justice Research and Policy 4, 131-142.

Becchetti, L., Conzo, P., 2011. Enhancing capabilities through credit access: Creditworthiness as a signal of trustworthiness under asymmetric information. Journal of Public Economics 95, 265-278.

Beck, T., Behr, P., Madestam, A., 2018. Sex and credit: Do gender interactions matter for credit market outcomes? Journal of Banking and Finance 87, 380-396.

Becker, G.S., 1971. The Economics of Discrimination, second edition. University of Chicago Press, Chicago.

Becker, G.S. 1993. Nobel lecture: The economic way of looking at behavior. Journal of Political Economy 101, 385-389.

Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes. Quarterly Journal of Economics 131, 1753–1794.

Boyes, W.J., Hoffman, D.L., Low, S.A., 1989. An econometric analysis of the bank credit scoring problem. Journal of Econometrics 40, 3-14.

Brock, W.A., Cooley, J., Durlauf, S.N., Navarro, S., 2012. On the observational implications of taste-based discrimination in racial profiling. Journal of Econometrics 166, 66-78.

Brock, J.M., De Haas, R., 2023. Discriminatory lending: Evidence from bankers in the lab. American Economic Journal: Applied Economics 15, 31-68.

Chen, S., Doerr, S., Frost, J., Gambacorta, L., Shin, H.S., 2023. The fintech gender gap. Journal of Financial Intermediation 54, 101026.

Cornée, S., 2019. The relevance of soft information for predicting small business credit default: Evidence from a social bank. Journal of Small Business Management 57, 699-719.

Cozarenco, A., Szafarz, A., 2018. Gender biases in bank lending: Lessons from microcredit in France. Journal of Business Ethics 147, 631–50.

Delis, M.D., Hasan, I., Iosifidi, M., Ongena, S., 2022. Gender, credit, and firm outcomes. Journal of Financial and Quantitative Analysis 57, 359-89.

D'Espallier, B., Hudon, M., Khavul, S., Szafarz, A., 2021. Donors talk: The signaling and imprinting effects of giving to social enterprises. Academy of Management Proceedings, 16014.

Dobbie, W., Liberman, A., Paravisini, D., Pathania, V., 2021. Measuring bias in consumer lending. Review of Economic Studies 88, 2799-2832.

Dymski, G.A., 2006. Discrimination in the credit and housing markets: Findings and challenges. In: Rodgers III W.M. (Ed.) Handbook on the Economics of Discrimination, Edward Elgar Publishing, Cheltenham, pp. 215-259.

Epanechnikov, V.A., 1969. Non-parametric estimation of a multivariate probability density. Theory of Probability & Its Applications 14, 153-8.

Fay, M., Williams, L., 1993. Gender bias and the availability of business loans. Journal of

Business Venturing 8, 363–376.
Ferguson, M.F., Peters, S.R., 1995. What constitutes evidence of discrimination in lending? Journal of Finance 50, 739-748.
Ferguson, M.F., Peters, S.R., 2000. Is lending discrimination always costly? Journal of Real Estate Finance and Economics 21, 23-44.
Fisman, R., Sarkar, A., Skrastins, J., Vig, V., 2020. Experience of communal conflicts and inter-group lending. Journal of Political Economy 128, 3346-3375.
Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., Walther, A., 2022. Predictably unequal? The effects of machine learning on credit markets. Journal of Finance 77, 5-47.
Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153-61.
Honoré, B.E., Hu, L., 2022. Sample selection models without exclusion restrictions: Parameter heterogeneity and partial identification. Journal of Econometrics, https://doi.org/10.1016/j.jeconom.2021.07.017.
Maasoumi, E., Wang, L., 2017. What can we learn about the racial gap in the presence of sample selection? Journal of Econometrics 199, 117-130.
McFadden, D., 1989. Testing for Stochastic Dominance. In: Fomby, T.B., Seo, T.K. (Eds.), Studies in the Economics of Uncertainty, Springer, New York, pp. 113-134.
Méon, P.G., Szafarz, A., 2011. The modern corporation as a safe haven for taste-based discrimination: An agency model of hiring decisions. Labour Economics 18, 487-97.
Nelson, J.A., 2014. The power of stereotyping and confirmation bias to overwhelm accurate assessment: The case of economics, gender, and risk aversion. Journal of Economic Methodology 21, 211-231.
Nelson, J.A., 2015. Are women really more risk-averse than men? A re-analysis of the literature using expanded methods. Journal of Economic Surveys 29, 566-585.
Pethig, F., Kroenung, J., 2023. Biased humans, (un)biased algorithms? Journal of Business Ethics 183, 637-652.
Roszbach, K., 2004. Bank lending policy, credit scoring, and the survival of loans. Review of Economics and Statistics 86, 946-958.
Shaffer, S., 1996. Evidence of discrimination in lending: An extension. Journal of Finance 51, 1551-1554.
Simoiu, C., Corbett-Davies, S., Goel, S., 2017. The problem of infra-marginality in outcome tests for discrimination. Annals of Applied Statistics 11, 1193-1216.
Smirnov, N.V., 1933. Estimate of deviation between empirical distribution functions in two independent samples. Bulletin Moscow University 2, 3-16.
Värendh Månsson, C., Wry, T., Szafarz, A., 2020. Anchors aweigh? Then time to go upstream. Why we need to theorize 'mission' before 'drift'. Academy of Management Review 45, 230-234.
Wry, T., Zhao, E.Y., 2018. Taking trade-offs seriously: Examining the contextually contingent relationship between social outreach intensity and financial sustainability in global microfinance. Organization Science 29, 507-528.
Zucchini, W., 2003. Applied smoothing techniques. Part 1: Kernel density estimation. Temple University, 20 p. [Available online at http://staff.ustc.edu.cn/~zwp/teach/Math-Stat/kernel.pdf.]

Appendix A. The F&P (1995) approach with and without FOSD[7]

To illustrate the importance of the FOSD condition in the application of the F&P (1995) approach, we consider (without loss of generality) the bottom left cell in Table 1 where women are subject to higher approval rate ($\alpha_A > 0$) and lower recovery rate ($\alpha_R < 0$). This configuration should signal a strong positive bias in favor of women.

Figures A1 and A2 show two observationally equivalent situations with these characteristics—namely a higher approval rate for women than for men (73% vs. 64%) and a lower recovery rate for women than for men (39% vs. 45%)—but with and without FOSD, respectively. Each figure shows two density functions of repayment probabilities, in red for women, and in blue for men. In Figure A1, the FOSD causes the male cumulative distribution function to first-order stochastically dominate the female cumulative distribution function, so that $\theta_1^* > \theta_2^*$. In contrast, in Figure A2, we do not have FOSD and $\theta_2^* > \theta_1^*$. Thanks to FOSD, the situation in Figure A1 is consistent with Table 1, and the credit standard for women is lower than that for men, while Figure A2 shows that, without FOSD, this conclusion is no longer valid.

[7] Example inspired from Simoiu et al. (2017).

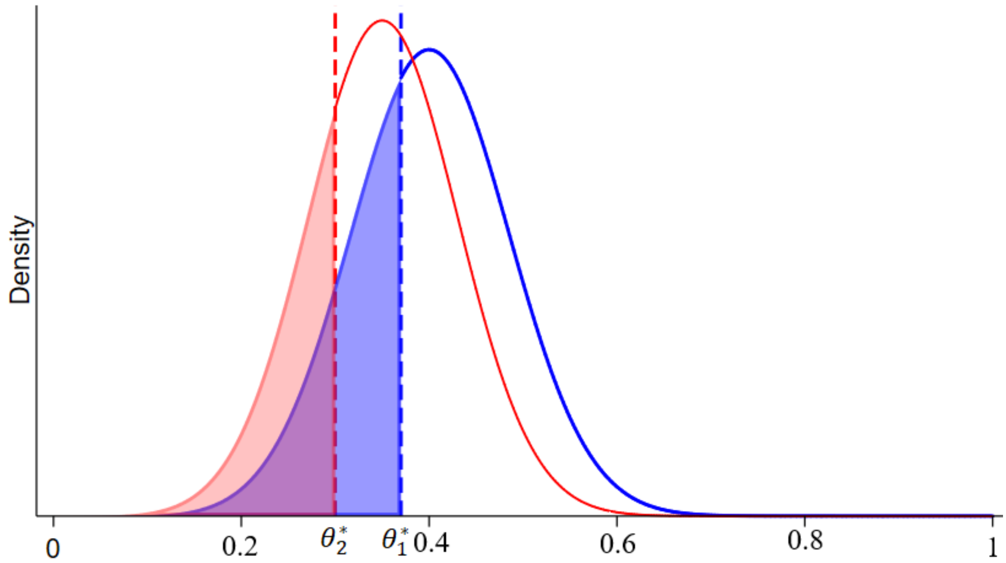

Figure A1. Two groups of applicants with first-order stochastic dominance

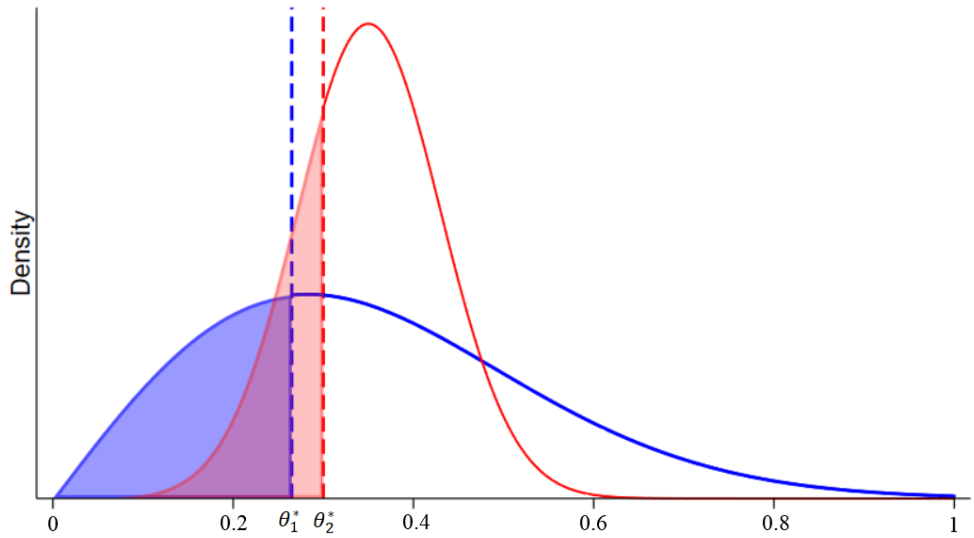

Figure A2. Two groups of applicants without first-order stochastic dominance