

Measuring Bias in Consumer Lending*

Will Dobbie[†] Andres Liberman[‡] Daniel Paravisini[§] Vikram Pathania[¶]

August 2018

Abstract

This paper tests for bias in consumer lending decisions using administrative data from a high-cost lender in the United Kingdom. We motivate our analysis using a simple model of bias in lending, which predicts that profits should be identical for loan applicants from different groups at the margin if loan examiners are unbiased. We identify the profitability of marginal loan applicants by exploiting variation from the quasi-random assignment of loan examiners. We find significant bias against both immigrant and older loan applicants when using the firm's preferred measure of long-run profits. In contrast, there is no evidence of bias when using a short-run measure used to evaluate examiner performance, suggesting that the bias in our setting is due to the misalignment of firm and examiner incentives. We conclude by showing that a decision rule based on machine learning predictions of long-run profitability can simultaneously increase profits and eliminate bias.

Keywords: Discrimination, Consumer Credit

JEL codes: G41, J15, J16

*First version: August 2018. We are extremely grateful to the Lender for providing the data used in this analysis. We also thank Leah Boustan, Hank Farber, Alex Mas, Crystal Yang, and numerous seminar participants for helpful comments and suggestions. Emily Battaglia, Nicole Gandre, Jared Grogan, Bailey Palmer, and James Reeves provided excellent research assistance. All errors and omissions are ours alone.

[†]Princeton University and NBER. Email: wdobbie@princeton.edu

[‡]New York University. Email: aliberma@stern.nyu.edu

[§]London School of Economics and CEPR. Email: d.paravisini@lse.ac.uk

[¶]University of Sussex. Email: v.pathania@sussex.ac.uk

I. Introduction

There are large disparities in the availability and cost of credit within many developed countries. In the United States, for example, blacks pay higher interest rates and are more likely to be rejected for a mortgage compared to observably similar whites, even after accounting for observable differences in credit history and earnings (e.g., Charles, and Hurst 2002, Bayer, Ferreira, and Ross 2017). There are also large disparities in credit usage and interest rates by ethnicity and gender within many European countries that cannot be explained by observable differences in creditworthiness (e.g., Alesina, Lotti, and Mistrulli 2013, Deku, Kara, and Molyneux 2016).

These disparities have fueled concerns that lenders may be biased against minorities and women due to prejudice or inaccurate stereotyping, which, in turn, has provided a rationale for a range of policies aimed at preventing discriminatory lending practices such as the Home Mortgage Disclosure Act in the United States and portions of the Equality Act of 2010 in the United Kingdom. Yet, the observed disparities in credit availability and pricing could also be driven by omitted variable bias or statistical discrimination on the part of lenders. For example, lenders may use variables that are not observed by the econometrician that are correlated with both creditworthiness and ethnicity or gender when making lending decisions, such as an applicant’s expected future income, leading to omitted variable bias. Lenders may also be using observable group traits such as ethnicity or gender to form accurate beliefs about the unobservable characteristics of different applicants, commonly known as statistical discrimination (e.g., Phelps 1972, Arrow 1973).¹

To distinguish between bias and these alternative explanations, Becker (1957, 1993) proposed an “outcome test” that compares the success or failure of decisions across groups at the margin. In the context of consumer lending, the outcome test is based on the idea that long-run profits to the lender should be identical for marginal applicants from all groups if loan examiners are unbiased

¹In both the United States and the United Kingdom, it is illegal for lenders to discriminate against minorities or women, regardless of whether that discrimination is driven by bias or statistical discrimination. Lenders are, however, allowed to use a wide range of variables that may not be observed by the econometrician when making lending decisions, so long as these variables have a legitimate business purpose and do not have a disparate impact on a protected class. In practice, however, statistical discrimination and omitted variable bias may be indistinguishable from each other unless there is direct evidence that lenders used ethnicity or gender to make lending decisions. The policy implications of bias and statistical discrimination are also very different. For example, bias stemming from prejudice or inaccurate stereotyping may be best addressed by policies that increase competition or encourage the employment of minority or female loan examiners (e.g., Becker 1957). In contrast, statistical discrimination is likely best addressed through the enforcement of existing laws such as the Equal Credit Opportunity Act or the Fair Housing Act in the United States or the Equality Act in the United Kingdom.

and the disparities across groups are solely due to omitted variables and statistical discrimination. In contrast, marginal applicants from a targeted group (e.g., minorities) will yield higher profits to the lender than marginal applicants from the non-targeted group (e.g., non-minorities) if loan examiners are biased against the targeted group and these disparities cannot be fully explained by omitted variables and statistical discrimination.

Estimating the profitability of marginal applicants and implementing the outcome test has been complicated by three important issues. First, researchers rarely observe long-run profits, and intermediate outcomes such as loan default may not provide an accurate measure of long-term profits (e.g., Han 2004, Agarwal et al. 2015). Second, characteristics observable to the lender but not the econometrician may be correlated with loan decisions, resulting in omitted variable bias when estimating the profitability of marginal applicants from each group. Finally, and most importantly, standard OLS estimates recover the profitability of the average applicant from each group, not the marginal applicant as required by the outcome test (e.g., Ayres 2002). As a result, comparisons based on OLS estimates will not recover the true level of bias in consumer lending decisions unless one is willing to assume that there is an identical distribution of potential profitability across groups.

In this paper, we implement the Becker outcome test using detailed administrative data on loan outcomes from a high-cost lender in the United Kingdom (hereafter, “the Lender”). This setting offers an ideal laboratory to test for bias in consumer credit for at least three reasons. First, we observe detailed data on cash flows to and from the Lender that allow us to construct individual-level measures of profitability. Second, the Lender’s loan examiners must make on-the-spot judgments with only the standard credit information and limited interaction with loan applicants, making their decisions particularly prone to the kind of inaccurate stereotypes or categorical heuristics that can lead to bias. Finally, as in most credit markets, the Lender’s loan examiners are evaluated using a measure based on short-run default, not long-run profits, creating exactly the kind of agency problems that can unintentionally lead examiners to discriminate against groups where short- and long-run outcomes systematically diverge.

We identify the differences in profitability at the margin required for the Becker outcome test using variation in the approval tendencies of quasi-randomly assigned loan examiners. The Lender uses a blind rotation system to assign first-time loan applicants to examiners of the same nationality, effectively randomizing new applicants to examiners within each branch and nationality. Using the

assigned loan examiner as an instrumental variable (IV) for loan take-up, we can recover the causal effect of loan take-up on long-run profits for applicants at the margin of loan take-up. Though IV estimators are often criticized for the local nature of the estimates, we exploit the fact that the outcome test relies on the difference between exactly these kinds of local treatment effects to test for bias.²

In our empirical analysis, we find significant bias against both immigrant and older loan applicants when using the firm’s preferred measure of long-run profits. Following the initial loan decision, we find that marginal immigrant applicants yield profits that are £566 larger than native-born applicants, or nearly four times larger. Marginal older applicants also yield profits that are £348 larger than younger applicants, or more than two times larger. Conversely, marginal female and male applicants yield statistically identical profits, suggesting no bias against (or in favor of) female applicants. Our estimates are nearly identical if we account for other observable differences by group, suggesting that our results cannot be explained by ethnic or age-related differences in baseline credit history or other demographic characteristics. We also find similar results using a marginal treatment effects (MTE) specification that relies on a different set of identifying assumptions. In contrast to these IV and MTE results, however, naïve OLS estimates indicate much more modest levels of bias across all three groups, highlighting the importance of accounting for both infra-marginality and omitted variables when estimating bias in consumer credit decisions.

The second part of the paper explores which form of bias explains our findings. The first possibility is that, as originally modeled by Becker (1957, 1993), prejudice leads loan examiners to discriminate against immigrants and older loan applicants at the margin. However, immigrant applicants are typically matched to ethnically similar loan examiners, and loan examiners tend to be older themselves, institutional features that are inconsistent with most models of ethnic or age-based prejudice. A second possibility is that loan examiners form inaccurate stereotypes on the profitability of lending to immigrant and older loan applicants. Recent work suggests that these types of inaccurate stereotypes can arise if immigrant and older applicants are over-represented in

²Our empirical strategy builds on Arnold, Dobbie, and Yang (Forthcoming), who test for racial bias in bail decisions using the quasi-random assignment of bail judges to identify outcomes for marginal white and marginal black defendants, and Marx (2017), who tests for racial bias in police stops using police officers of different races to identify bounds on the outcomes for marginal white, Hispanic, and black drivers. Our IV strategy is also related to research designs used by Liberman, Paravisini, and Pathania (2016) to study the effects of high-cost credit on credit reputation and future access to credit and Dobbie, Goldsmith-Pinkham, and Yang (2017) to estimate the impact of bankruptcy protection.

the left tail of the distribution of potential profits (e.g., Bordalo et al. 2016). However, we find that immigrant and older applicants are actually under-represented in the left tail of the payoff distribution, exactly the opposite pattern we would expect if inaccurate stereotypes were driving our results.

In this paper, we propose a new explanation for bias in consumer lending: the misalignment of firm and examiner incentives.³ Long-run profits in our setting are largely driven by the number of loans an individual takes out, not whether or not an individual defaults on the first loan. Yet, loan examiners are evaluated using a measure based on first-loan defaults, not long-run profits, due to the perceived infeasibility of waiting for one to two years to measure examiner performance. As a result, loan examiners in our setting have an incentive to minimize first-loan default, not maximize long-run profits, potentially leading them to discriminate against applicants where these outcomes systematically diverge. Consistent with this explanation, we find no evidence of bias against immigrant and older applicants when using first-loan default as an outcome. We also find that long-run profits are systematically higher for a given level of default risk for immigrant and older applicants compared to native-born and younger applicants, respectively, exactly as we would expect if the bias in our setting is driven by the misalignment of incentives. Finally, we find that the decisions made by loan examiners are strikingly consistent with a data-based decision rule minimizing short-term default, but inconsistent with a decision rule maximizing long-term profits.

We conclude by showing that a decision rule based on machine learning (ML) predictions of long-run profits could simultaneously increase profits and eliminate bias. Following Kleinberg et al. (2018), we use the quasi-random assignment of loan examiners to identify the implicit rankings of applicants by loan examiners, which we then compare to the rankings produced by a standard ML algorithm. Consistent with our earlier results, we find that loan examiners systematically misrank loan applicants at the margin of loan take-up, particularly immigrant and older applicants. The Lender would earn approximately 58 percent more per applicant if marginal lending decisions were

³The misalignment of firm and examiner incentives is likely widespread in credit markets (e.g., Heider, and Inderst 2012). Keys et al. (2010) show, for example, that the securitization of subprime mortgage loans prior to the financial crisis reduced the incentives of financial intermediaries to carefully screen borrowers. Berg, Puri, and Rocholl (2013) and Agarwal, and Ben-David (Forthcoming) similarly show that volume incentives distort the incentives of loan examiners, leading to higher default risk, while Hertzberg, Liberti, and Paravisini (2010) show that loan officers underreport bad news due to reputational concerns. The Wells Fargo account fraud scandal, where millions of fraudulent savings and checking accounts were created without customers' consent, is also widely thought to be the result of poorly designed sales incentives among branch employees.

made using the ML algorithm rather than loan examiners. Including all applicants, not just those at the margin of loan take-up, the Lender would earn over 30 percent more per applicant if lending decisions were made using the ML algorithm.

Our results contribute to an important literature testing for bias in consumer credit decisions. Outcome tests based on standard OLS estimates suggest that black mortgage borrowers in the United States have, if anything, slightly higher default rates (e.g., Van Order, Lekkas, and Quigley 1993, Berkovec et al. 1994) and similar recovery rates (e.g., Han 2004) compared to observably similar white mortgage borrowers, suggesting little bias in this market. In contrast, both in-person and correspondence audit studies (e.g., Ross et al. 2008, Hanson et al. 2016) suggest that loan officers treat black and Hispanic mortgage applicants worse than identical white applicants. There is also evidence that racial disparities in credit outcomes narrow as competition increases (e.g., Berkovec et al. 1998, Buchak, and Jørring 2016), a finding that is inconsistent with most models of statistical discrimination.

Our paper is also related to a large literature documenting disparities in the availability and cost of credit by ethnicity and gender. There is considerable evidence that minorities have either less access to credit or are forced to pay more for credit compared to observably similar nonminorities for mortgage loans (e.g., Charles, and Hurst 2002, Bayer, Ferreira, and Ross 2017), auto loans (e.g., Charles, Hurst, and Stephens 2008), small business loans (e.g., Cavalluzzo, and Cavalluzzo 1998, Cavalluzzo, Cavalluzzo, and Wolken 2002, Blanchflower, Levine, and Zimmerman 2003), and consumer loans (e.g., Cohen-Cole 2011). There is also evidence that women pay more for both consumer credit (e.g., Alesina, Lotti, and Mistrulli 2013) and small business loans (e.g., Bellucci, Borisov, and Zazzaro 2010) than observably similar men, and that blacks are more likely to be rejected for peer-to-peer loans than observably similar whites (e.g., Pope, and Sydnor 2011).

The rest of the paper is structured as follows. Section II describes the theoretical model underlying our analysis and develops our empirical test for bias. Section III describes our institutional setting, the data used in our analysis, and the construction of our instrument. Section IV presents the main results. Section V explores potential mechanisms, and Section VI concludes. The Online Appendix provides additional results and detailed information on the outcomes used in our analysis.

II. An Empirical Test of Bias in Consumer Lending

In this section, we motivate and develop our empirical test for bias in consumer lending decisions. We first show that we can test for bias by comparing treatment effects for the marginal loan applicants from different groups, whether that bias is driven by prejudice, inaccurate stereotypes, or the misalignment of firm and examiner incentives. We then show that we can identify these group-specific treatment effects using the quasi-random assignment of loan applications to loan examiners.

A. Model of Examiner Behavior

This section develops a stylized theoretical framework that allows us to define an outcome-based test of bias in consumer lending. We begin with a model of taste-based discrimination that closely follows the theory of discrimination developed by Becker (1957, 1993) and subsequent work applying this theory to test for bias in consumer lending (e.g., Han 2004). We then present two alternative models of examiner behavior based on inaccurate stereotypes and misaligned examiner incentives. Each model suggests that we can test for bias in consumer lending decisions by comparing treatment effects for the marginal loan applicants from different groups.⁴

Taste-Based Discrimination: Let i denote loan applicants and \mathbf{V}_i denote all applicant characteristics considered by the loan examiner, excluding group identity g_i , such as ethnicity or gender. Loan examiners, indexed by e , form an expectation of the long-run profits of lending to applicant i conditional on observable characteristics \mathbf{V}_i and group g_i , $\mathbb{E}[\alpha_i|\mathbf{V}_i, g_i]$.

The perceived cost of lending to applicant i assigned to examiner e is denoted by $t_g^e(\mathbf{V}_i)$, which is a function of observable applicant characteristics \mathbf{V}_i . The perceived cost of lending $t_g^e(\mathbf{V}_i)$ includes both the firm's opportunity cost of making a loan and the personal benefits to examiner e from any direct utility or disutility from being known as either a lenient or tough loan examiner, respectively. Importantly, we allow the perceived cost of lending $t_g^e(\mathbf{V}_i)$ to vary by group $g \in T, R$ to allow for

⁴In our analysis, we abstract away from the effects of market competition on bias in consumer credit markets. In a simple model of taste-based discrimination, market competition indirectly raises the cost of discrimination through the threat of market share losses to competitive entrants, leading to a lower level of bias in equilibrium (e.g., Becker 1957, Peterson 1981). Similar arguments extend to models of discrimination based on inaccurate stereotypes or misaligned examiner incentives. See Berkovec et al. (1998) and Buchak, and Jørring (2016) for empirical work estimating the effects of competition on lending disparities.

examiner preferences to differ for applicants from the target group (e.g., minority applicants) and the reference group (e.g., non-minority applicants), respectively. We do not, however, allow the lender's true opportunity costs of lending to vary by group.

Following Becker (1957, 1993), we define loan examiner e as biased against the target group if $t_T^e(\mathbf{V}_i) > t_R^e(\mathbf{V}_i)$. Thus, biased loan examiners reject target group applicants that they would otherwise approve because these examiners perceive a higher cost of lending to applicants from the target group compared to observably identical applicants from the reference group.

For simplicity, we assume that loan examiners are risk neutral and maximize the perceived net benefit of approving a loan. We also assume that the loan examiner's sole task is to decide whether to approve or reject a loan application given that, in practice, this is the only decision margin in our setting.

Under these assumptions, the model implies that loan examiner e will lend to applicant i if and only if the expected profit is weakly greater than the perceived cost of the loan:

$$\mathbb{E}[\alpha_i | \mathbf{V}_i, g_i = g] \geq t_g^e(\mathbf{V}_i) \quad (1)$$

Given this decision rule, the marginal applicant for examiner e and group g is the applicant i for whom the expected profit is exactly equal to the perceived cost, i.e., $\mathbb{E}[\alpha_i^e | \mathbf{V}_i, g_i = g] = t_g^e(\mathbf{V}_i)$. We simplify our notation moving forward by letting this expected profit for the marginal applicant for examiner e and group g be denoted by α_g^e .

Based on the above framework, the model yields the standard outcome-based test for bias from Becker (1957, 1993).

OUTCOME TEST 1: TASTE-BASED DISCRIMINATION. If examiner e is biased against applicants from the target group, then the expected profitability for the marginal target group applicant is higher than the expected profitability for the marginal reference group applicant: $\alpha_T^e > \alpha_R^e$.

Outcome Test 1 predicts that marginal target and marginal reference group loan applicants should have the same profitability if examiners are unbiased, but marginal target group applicants should yield higher profits if examiners are biased against applicants from the target group. The correct procedure to test whether loan decisions are biased is therefore to determine whether loans

to marginal target group applicants are more profitable than loans to marginal reference group applicants.

Inaccurate Group Stereotypes: In the taste-based model of discrimination outlined above, we assume that examiners agree on the (true) expected net present profit of lending to applicant i , $\mathbb{E}[\alpha_i|\mathbf{V}_i, g_i]$, but not the perceived cost of lending to the applicant, $t_g^e(\mathbf{V}_i)$. An alternative approach is to assume that examiners disagree on their (potentially inaccurate) predictions of the expected profit, as would be the case if examiners systematically underestimate the profitability of target group applicants relative to reference group applicants in the spirit of Bordalo et al. (2016) and Arnold, Dobbie, and Yang (Forthcoming). We show that a model motivated by these kinds of biased prediction errors can generate the same predictions as a model of taste-based discrimination.

Let i again denote applicants and \mathbf{V}_i denote all applicant characteristics considered by the loan examiner, excluding group identity g_i . The perceived cost of lending to applicant i assigned to examiner e is now defined as $t^e(\mathbf{V}_i)$, where we explicitly assume that $t^e(\mathbf{V}_i)$ is independent of the group identity of the applicant.

The perceived profitability of lending to applicant i conditional on observable characteristics \mathbf{V}_i , $\mathbb{E}^e[\alpha_i|\mathbf{V}_i, g_i]$, is now allowed to vary across examiners. We can write the perceived profitability as:

$$\mathbb{E}^e[\alpha_i|\mathbf{V}_i, g_i] = \mathbb{E}[\alpha_i|\mathbf{V}_i, g_i] + \tau_g^e(\mathbf{V}_i) \quad (2)$$

where $\tau_g^e(\mathbf{V}_i)$ is a prediction error that is allowed to vary by examiner e and group identity g_i .

Following Arnold, Dobbie, and Yang (Forthcoming), we define examiner e as making biased prediction errors against target group applicants if $\tau_T^e(\mathbf{V}_i) < \tau_R^e(\mathbf{V}_i)$. Thus, biased loan examiners reject target group applicants that they would otherwise approve because these examiners systematically underestimate the true profitability of lending to target group applicants compared to reference group applicants.

Following the taste-based model, loan examiner e will lend to applicant i if and only if the perceived expected profit is weakly greater than the cost of the loan:

$$\mathbb{E}^e[\alpha_i|\mathbf{V}_i, g_i = g] = \mathbb{E}[\alpha_i|\mathbf{V}_i, g_i = g] + \tau_g^e(\mathbf{V}_i) \geq t^e(\mathbf{V}_i) \quad (3)$$

The prediction error model can be made equivalent to the taste-based model of discrimination outlined above if we relabel $t^e(\mathbf{V}_i) - \tau_g^e(\mathbf{V}_i) = t_g^e(\mathbf{V}_i)$. As a result, we can generate identical empirical predictions using the prediction error and taste-based models.

Following this logic, our model of biased prediction errors yields a similar outcome-based test for bias.

OUTCOME TEST 2: INACCURATE STEREOTYPES. If examiner e systematically underestimates the true expected profitability of lending to target group applicants relative to reference group applicants, then the expected profitability for the marginal target group applicant is higher than the expected profitability for the marginal reference group applicant: $\alpha_T^e > \alpha_R^e$.

Parallel to Outcome Test 1, Outcome Test 2 predicts that marginal target group and marginal reference group applicants should have the same profitability if loan examiners do not systematically make prediction errors that vary with group identity, but marginal target group applicants should yield higher profits if examiners systematically underestimate the true expected profitability of lending to target group applicants relative to reference group applicants. The correct procedure to test whether loan decisions are biased is therefore, once again, to determine whether loans to marginal target group applicants are more profitable than loans to marginal reference group applicants.

Misaligned Examiner Incentives: In both the taste-based and inaccurate stereotypes models of discrimination, we assume that all examiners maximize the perceived long-run profit of lending to applicant i , $\mathbb{E}[\alpha_i | \mathbf{V}_i, g_i]$. A final approach is to assume that loan examiners maximize an intermediate short-run outcome, not long-run profits, as would be the case if the examiner's compensation contract was based (at least in part) on that short-run outcome.⁵ We show that a model motivated by these kinds of principal-agent problems can also generate the same predictions as the taste-based model of discrimination.

Let i again denote applicants and \mathbf{V}_i again denote all applicant characteristics observed by examiners, excluding group identity g_i . The perceived cost of lending to applicant i assigned to

⁵The optimal compensation contract may include intermediate outcomes if, for example, loan examiners are more impatient than the Lender and the intermediate outcome is correlated with long-run profits. Heider, and Inderst (2012) show, for example, that it may be optimal for lenders to compensate examiners only for the number of loans given, not loan performance.

examiner e is again defined as $t^e(\mathbf{V}_i)$, which is assumed to be independent of the group identity of the applicant as in the inaccurate stereotypes model.

We now assume that examiners maximize their private benefit of lending to applicant i , $\beta_e \mathbb{E}[\alpha_i^{SR} | \mathbf{V}_i, g_i] + (1 - \beta_e) \mathbb{E}[\alpha_i | \mathbf{V}_i, g_i]$, which is equal to a weighted average of both expected short-term profits, $\mathbb{E}[\alpha_i^{SR} | \mathbf{V}_i, g_i]$, and expected long-run profits, $\mathbb{E}[\alpha_i | \mathbf{V}_i, g_i]$. We allow the weight on short-term profits, β_e , to vary across examiners to capture the idea that examiners may vary in their risk tolerance, impatience, or liquidity constraints, although this assumption is not critical to our model. The perceived benefit from lending to individual i from the perspective of the loan examiner can be rewritten as:

$$\mathbb{E}[\alpha_i | \mathbf{V}_i, g_i] + \beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i] \quad (4)$$

where $\beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i]$ represents the wedge between examiner e 's perceived benefit of lending to applicant i in the short run and the lender's expected profit of lending to applicant i in the long run. The wedge between the examiner and firm's objective functions is increasing in the weight placed on the short-term outcome, β_e , and is larger for applicants where there is a larger gap between expected short- and long-run profits, $\mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i]$.

We define examiner e as making biased valuations against target group applicants if $\beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i = T] < \beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i = R]$. Thus, biased loan examiners reject target group applicants that they would be willing to approve if these applicants were from the reference group, because the utility of biased loan examiners depends on short-term profits and target group applicants have lower short-run profits for a given level of long-run profits. From the lender's perspective, biased loan examiners therefore underestimate the true benefit of lending to target group applicants relative to reference group applicants.

Following the taste-based model, loan examiner e will lend to applicant i if and only if examiner e 's perceived benefit is weakly greater than the cost of the loan:

$$\mathbb{E}[\alpha_i | \mathbf{V}_i, g_i] + \beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i] \geq t^e(\mathbf{V}_i) \quad (5)$$

We can again show that the misaligned incentives model can be reduced to the taste-based model

of discrimination outlined above if we relabel $t^e(\mathbf{V}_i) - \beta_e \mathbb{E}[\alpha_i^{SR} - \alpha_i | \mathbf{V}_i, g_i = g] = t_g^e(\mathbf{V}_i)$. As a result, we can generate identical empirical predictions using the misaligned incentives, inaccurate stereotypes, and taste-based models. Our model of misaligned incentives therefore yields a similar outcome-based test for bias.

OUTCOME TEST 3: MISALIGNED EXAMINER INCENTIVES. If examiner e systematically undervalues the true expected long-run profitability of lending to target group applicants relative to reference group applicants, then the expected profitability for the marginal target group applicant is higher than the expected profitability for the marginal reference group applicant: $\alpha_T^e > \alpha_R^e$.

Parallel to Outcome Tests 1 and 2, Outcome Test 3 predicts that marginal target group and marginal reference group applicants should have the same long-term profitability if examiners' incentives do not systematically lead to valuation errors that vary by group, but marginal target group applicants should have higher long-run profits if examiners systematically undervalue the true expected long-run profitability of lending to target group applicants relative to reference group applicants. The correct procedure to test whether loan decisions are biased is therefore, once again, to determine whether loans to marginal target group applicants are more profitable in the long run than loans to marginal reference group applicants.

The above theoretical framework presents three different behavioral models with identical empirical predictions: if there is bias against target group applicants, then long-run profits will be higher, in expectation, for marginal target group applicants compared to marginal reference group applicants. In contrast, marginal target and reference group applicants will yield identical long-run profits if the observed disparities in consumer lending are solely due to statistical discrimination. The interpretation of our findings, however, depends on the specific model underlying examiner behavior. We discuss which of the three models is most consistent with our findings in Section V.⁶

⁶In contrast to the three models we consider in this section, models of accurate statistical discrimination suggest that target group applicants may be treated worse than observably identical reference group applicants if either target group applicants are, on average, riskier given an identical signal of profitability (e.g., Phelps 1972, Arrow 1973), or target group applicants have less precise signals of profitability (e.g., Aigner, and Cain 1977). In both of these cases, however, examiners use group identity to form accurate predictions of profitability, both on average and at the margin of loan take-up. As a result, neither form of accurate statistical discrimination will lead to differences in profitability at the margin.

B. Empirical Test of Bias in Consumer Lending

This section explains how we identify the differences in profitability at the margin required for the Becker outcome test using variation in the approval tendencies of quasi-randomly assigned loan examiners, building on work by Arnold, Dobbie, and Yang (Forthcoming) in the context of bail decisions. We begin with a definition of the target parameter and a series of simple graphical examples that illustrate our approach. We then formally describe the conditions under which our examiner IV strategy yields consistent estimates of bias in consumer lending decisions and discuss the interpretation of the estimates.

Overview: Following the theory model, let the average profitability for applicants from group g at the margin for examiner e , α_g^e , for some weighting scheme, w^e , across all loan examiners, $e = 1 \dots E$, be given by:

$$\begin{aligned}\alpha_g^{*,w} &= \sum_{e=1}^E w^e \alpha_g^e \\ &= \sum_{e=1}^E w^e t_g^e\end{aligned}\tag{6}$$

where w^e are non-negative weights which sum to one that will be discussed in further detail below. By definition, $\alpha_g^e = t_g^e$, where t_g^e represents examiner e 's threshold for loan approval for applicants from group g . In our context, profitability can be identified by the treatment effect of loan take-up on long-run profits, as applicants who do not take up a loan yield exactly zero profit. Thus, $\alpha_g^{*,w}$ represents a weighted average of the treatment effects for applicants of group g at the margin of loan take-up across all examiners.

Following this notation, the average level of bias among loan examiners, $B^{*,w}$, for the weighting scheme w^e is given by:

$$\begin{aligned}B^{*,w} &= \sum_{e=1}^E w^e (t_T^e - t_R^e) \\ &= \sum_{e=1}^E w^e t_T^e - \sum_{e=1}^E w^e t_R^e \\ &= \alpha_T^{*,w} - \alpha_R^{*,w}\end{aligned}\tag{7}$$

Equation (7) generalizes the outcome test to the case where there are many examiners and the level of bias across examiners may vary. Following Equation (6), we can then express the target parameter, $B^{*,w}$, as a weighted average across all examiners of bias in lending decisions, measured by the difference in treatment effects for target and reference group applicants at the margin of loan take-up.

Recall that standard OLS estimates will typically not recover unbiased estimates of the weighted average of bias, $B^{*,w}$, for two reasons. The first is that characteristics observable to the loan examiner but not the econometrician may be correlated with loan approval, resulting in omitted variable bias when estimating the treatment effects for different types of loan applicants. The second, and more important, reason OLS estimates will not recover unbiased estimates of bias is that the average treatment effect identified by OLS will not equal the treatment effect at the margin required by the outcome test unless there is either an identical distribution of potential profits for loan applicants from different groups or constant treatment effects across the entire distribution of loan applicants — the well-known infra-marginality problem (e.g., Ayres 2002).

Following Arnold, Dobbie, and Yang (Forthcoming), we estimate the differences in profitability at the margin required for the Becker outcome test, $B^{*,w}$, using variation in the approval tendencies of quasi-randomly assigned loan examiners. Our estimator uses the standard IV framework to identify the difference in local average treatment effects (LATEs) for reference group and target group applicants near the margin of loan take-up. Though IV estimators are often criticized for the local nature of the estimates, we exploit the fact that the outcome test relies on the difference between exactly these kinds of local treatment effects to test for bias. This empirical design allows us to recover a weighted average of the long-run profitability of different groups near the margin, where the weights are equal to the standard IV weights described in further detail below.

Figure 1 provides a series of simple graphical examples to illustrate the intuition of our approach. In Panel A, we consider the case where there is a single unbiased examiner to illustrate the potential for infra-marginality bias when using a standard OLS estimator. The examiner perfectly observes expected profitability and chooses the same approval threshold for all loan applicants, but the distributions of profitability differ by group identity such that reference group applicants, on average, yield higher profits than target group applicants. Letting the vertical lines denote the examiner’s approval threshold, standard OLS estimates of α_T and α_R measure the average profitability for

target and reference group applicants who take up a loan, respectively. In the case illustrated in Panel A, the standard OLS estimator indicates that the examiner is biased against reference group applicants, when, in reality, the examiner is unbiased. Panel B illustrates a similar case where the standard OLS estimator indicates that the examiner is unbiased, when, in reality, the examiner is biased against target group applicants.

To illustrate how our IV estimator identifies the profitability of marginal applicants, the last two panels of Figure 1 consider a case where there are two loan examiners, one that is lenient and one that is strict. In Panel C, we consider the case where the two examiners are unbiased, while in panel D we consider the case where the two examiners are both biased against target group applicants. In both cases, an IV estimator using examiner leniency as an instrument for loan take-up will measure the average profitability of “compliers,” or applicants who take up a loan when assigned to the lenient examiner but not when assigned to the strict examiner. In other words, the IV estimator only measures the profitability of applicants between the two examiner thresholds, ignoring applicants that are either above or below both examiner thresholds. When the two examiners in our example are “close enough” in leniency, the IV estimator will therefore measure the profitability of applicants only at the margin of loan take-up, allowing us to correctly conclude that the examiners are unbiased in the example illustrated in Panel C and biased against target group applicants in Panel D.

Consistency of the IV Estimator: We now briefly review the conditions under which our examiner IV strategy yields consistent estimates of bias in consumer lending decisions. See Arnold, Dobbie, and Yang (Forthcoming) for formal proofs.

Let Z_i be a scalar measure of the assigned examiner’s propensity for loan take-up for applicant i that takes on values ordered $\{z_0, \dots, z_E\}$, where $E + 1$ is the total number of examiners. For example, a value of $z_e = 0.7$ indicates that 70 percent of all applicants assigned to examiner e take up a loan. We construct Z_i using a standard leave-out procedure that captures the approval tendencies of examiners. We calculate a single Z_i for all groups to minimize measurement error in our instrument, but we show in robustness checks that our results are similar (if less precise) if we allow the instrument to vary by group.

Following Imbens, and Angrist (1994), an estimator using Z_i as an instrumental variable for

loan take-up is valid and well-defined under the following three assumptions:

ASSUMPTION 1. (EXISTENCE) $Cov(TakeUp_i, Z_i) \neq 0$

ASSUMPTION 2. (EXCLUSION) $Cov(Z_i, \mathbf{v}_i) = 0$

ASSUMPTION 3. (MONOTONICITY) $TakeUp_i(z_e) - TakeUp_i(z_{e-1}) \geq 0$

where $\mathbf{v}_i = \mathbf{U}_i + \varepsilon_i$ consists of characteristics unobserved by the econometrician but observed by the examiner, \mathbf{U}_i , and idiosyncratic variation unobserved by both the econometrician and examiner, ε_i . Assumption 1 requires the instrument Z_i to increase the probability of loan take-up $TakeUp_i$. Assumption 2 requires the instrument Z_i to be as good as randomly assigned and to only influence profitability through the channel of loan take-up. In other words, Assumption 2 ensures that our instrument is orthogonal to characteristics unobserved by the econometrician, \mathbf{v}_i . Assumption 3 requires the instrument Z_i to weakly increase the probability of loan take up $TakeUp_i$ for all individuals.

Taking Assumptions 1–3 as given, let the true IV-weighted level of bias, $B^{*,IV}$ be defined as:

$$\begin{aligned} B^{*,IV} &= \sum_{e=1}^E w^e (t_T^e - t_R^e) \\ &= \sum_{e=1}^E \lambda^e (t_T^e - t_R^e) \end{aligned} \tag{8}$$

where $w^e = \lambda^e$, the standard IV weights defined in Imbens, and Angrist (1994).

Let our IV estimator that uses examiner leniency as an instrumental variable for loan take-up be defined as:

$$\begin{aligned} B^{IV} &= \alpha_T^{IV} - \alpha_R^{IV} \\ &= \sum_{e=1}^E \lambda_T^e \alpha_T^{e,e-1} - \sum_{e=1}^E \lambda_R^e \alpha_R^{e,e-1} \end{aligned} \tag{9}$$

where λ_g^e are again the standard IV weights and each pairwise treatment effect $\alpha_g^{e,e-1}$ captures the treatment effects of compliers within each $e, e-1$ pair, i.e. individuals such that $\alpha_g^{e,e-1} \in (t_g^{e-1}, t_g^e]$.

Our IV estimator B^{IV} provides a consistent estimate of the true level of bias $B^{*,IV}$ if two conditions hold: (1) the instrument Z_i is continuously distributed over some interval $[\underline{z}, \bar{z}]$, and

(2) the weights on the pairwise LATEs λ_g^e are identical across groups. The first condition ensures that the group-specific IV estimates are equal to the true IV-weighted average of treatment effects for applicants at the margin of loan take-up.⁷ The second condition for consistency ensures that any difference in the group-specific IV estimates is driven by differences in the true group-specific treatment effects, not differences in the IV weights applied to those treatment effects. This equal weights condition holds if there is a linear first stage across groups, as is true in our data (see Figure 2). We also find that the distributions of IV weights by nationality, gender, and age are visually indistinguishable from each other (see Appendix Figure A1) and that the IV weights for each examiner are highly correlated across groups (see Appendix Figure A2), indicating that the equal weights condition is unlikely to be violated in our setting.

Interpretation of IV Weights: We conclude this section by discussing the economic interpretation of our IV-weighted estimate of bias, B^{IV} . Appendix Table A1 presents OLS estimates of IV weights in each examiner-by-branch cell and observable examiner characteristics. We find that our IV weights are uncorrelated with the number of applications, examiner experience, examiner leniency, and examiner gender. We also find that our IV weights are largely uncorrelated with examiner-level estimates of bias against each group obtained from our MTE specification described below, indicating that our IV-weighted estimates of bias are likely to be very similar to estimates based on other weighting schemes. In robustness checks, we also report estimates from an MTE specification that allows us to impose equal weights when calculating the average level of bias across loan examiners at the cost of additional auxiliary assumptions.

III. Background, Data, and Instrument Construction

This section summarizes the most relevant information regarding our institutional setting and data, describes the construction of our examiner leniency measure, and provides support for the baseline assumptions required for our IV estimator. Further details on the cleaning and coding of variables are contained in Online Appendix B.

⁷The maximum estimation bias from using a discrete instrument, as we do in this paper, can be calculated using the empirical distribution of examiner leniency and the worst-case treatment effect heterogeneity among compliers. Using the 10th and 90th percentiles of observed profits as the worst-case treatment effect heterogeneity among compliers, we find that the maximum estimation bias when using a discrete instrument is only £18, indicating that this issue is unlikely to be a significant problem in our setting.

A. Institutional Setting

We test for bias in consumer lending decisions using information from a large subprime lender in the United Kingdom. The Lender offers short-term, uncollateralized, high-cost loans to subprime borrowers. Loan maturities are typically less than six months, and can be as short as a few weeks. Loan amounts range from £200 to £2,000, with an average first-loan amount of just under £300. All loans require weekly payments starting soon after the loan is disbursed, with interest rates that average about 600 percent. By comparison, the typical payday loan in the United States is below \$300 with an APR of 400 to 1,000 percent and a seven- to thirty-day maturity (Stegman 2007). The Lender also allows applicants who remain in good standing after one month the option of “topping up” their initial loan, or increasing their outstanding balance back to the initially approved loan amount. In other words, applicants can convert their initial loan to a line of credit up to the original loan amount after one month. The Lender’s profits are largely driven by the use of these loan top-ups over the next one to two years, with only about 25 percent of the variation in long-run profits coming from the repayment of the original loan amount. In contrast, the number of loan top-ups explains 34 percent of the variation in long-run profits among individuals taking up a loan, while the number of loans explains 41 percent of the variation in long-run profits in this sample. The repayment of the original loan amount also explains very little additional variation in long-run profits once these longer-run measures are included.

The Lender operates 24 branches throughout the United Kingdom to handle all in-person applications, and a virtual branch to handle all online and phone applications. In the physical branches maintained by the Lender, loan applicants are first greeted by a receptionist, who gathers basic information such as the applicant’s name, address, phone number, and nationality. Loan applicants are then randomly assigned to one of the loan examiners working in the branch that day using a blind rotation system. The blind rotation system randomly assigns native-born applicants to the full set of loan examiners working in the branch that day, but only randomly assigns foreign-born and non-English speaking applicants to the set of the loan examiners with the same ethnic background to put these applicants at ease and improve the accuracy of the screening process. Next, the assigned loan examiner reviews the applicant’s credit history, including the applicant’s credit score, outstanding debt, and past repayment behavior. Loan examiners are also encouraged to ask

about the applicant’s income and employment status, as well as any other relevant information, during the initial interview. Following the examiner’s approval decision, approved loan applicants decide whether to take up the loan or not, as well as the total amount to borrow from the maximum allowable credit line. Loans are then disbursed to approved applicants before leaving the store. The process is broadly similar for online and phone applications, although applicants are typically not randomly assigned to loan examiners and, as a result, we do not include these applicants in our analysis.

The assigned loan examiner has complete discretion to approve or reject first-time loan applicants whose credit scores exceed a minimum threshold established by the Lender. Loan examiners are compensated with a combination of a fixed salary plus a bonus that increases with loan volume and decreases with loan default. Loan examiners are not, however, compensated for long-run profits due to the perceived difficulty of waiting for one to two years to measure examiner performance. We explore the potential importance of this compensation contract when explaining our results in Section V.

B. Data Sources and Descriptive Statistics

We use administrative data on all loan applications and loan outcomes at the Lender between May 2012 and February 2015. The loan-level data contain detailed information pulled from a private credit registry at the time of application, including credit scores and information on outstanding debts and past repayment behavior. The data also contain information gathered by the examiner during the interview, including the applicant’s nationality, age, gender, earnings and employment, marriage status, number of dependents, months at his or her current residence, and the stated reason for the loan. Finally, for individuals who take up at least one loan, the data contain information on loan disbursement amounts, interest rates, maturities, payments, top-ups, and defaults for all loans during our sample period. The data are high-quality and complete with one important exception: earnings and employment information is only collected when examiners believe the application is likely to be approved, meaning that it is missing for a relatively large part of our sample. We therefore do not include earnings and employment controls in our baseline results, as the availability of these controls is mechanically correlated with examiner leniency. None of our results are significantly changed if we include these controls, however.

In our main results, long-run profits are defined as the sum of all payments made by the applicant minus all disbursements from the Lender for both the first loan and all subsequent loans during our sample period. In robustness checks, we present results using the net present value of long-run profits for a variety of discount rates. We also control for time-of-application fixed effects throughout to account for the fact that we observe some applicants for more time than others.

We make five restrictions to the estimation sample. First, we drop repeat applications, as these applications are not randomly assigned to examiners and have a nearly 100 percent approval rate. Second, we drop all online and phone applications, as these individuals are also not randomly assigned to examiners during our sample period. Third, we drop loans assigned to loan examiners with fewer than 50 applications, and loan applications where there is only a single applicant in a branch by nationality cell. Fourth, we drop a handful of applications where applicants are younger than 18 years old, older than 75 years old, or where the credit check information is missing. Finally, we drop all applications after December 2014 to ensure that we observe loan outcomes over a reasonable period. The final sample contains 45,687 first-time loan applications assigned to 254 loan examiners between May 2012 and December 2014.

Table 1 reports summary statistics for our estimation sample separately by loan take-up. Forty percent of first-time applicants are immigrants, 56 percent are female, 73 percent have lived at least one year at their current residence, and 42 percent are married. The average age of first-time applicants in our sample is 33.9 years old, with the typical applicant having just under one dependent. Over 91 percent of first-time applicants have a bank account and 29 percent have other loan payments. Twenty-six percent of loans are for emergency expenses, 11 percent are for a large one-time expense, 5 percent are to avoid an overdraft, and 23 percent are for shopping or a holiday.

For the 66 percent of first-time applicants who take out a loan, the average amount is about £290, with an APR of 663 percent and a maturity of 5.5 months. For these first loans, 35 percent end in default, 44 percent result in a top-up, with the remainder ending in the full repayment of the original balance. The average long-run profit for individuals taking out a loan, defined as the sum of all payments made by the applicant minus all disbursements from the Lender for both the first loan and all subsequent loans, is equal to £267. By definition, applicants who do not take out a loan have a 0 percent default rate, 0 percent top-up rate, 0 percent repayment rate, and yield profits of exactly £0.

C. Construction of the Instrumental Variable

We estimate the causal impact of loan take-up for the marginal loan applicant using a leave-out measure of loan examiner leniency as an instrumental variable for loan take-up. As discussed above, first-time loan applicants are assigned to loan examiners of the same nationality using a blind rotation system, effectively randomizing applicants to a subset of examiners within each branch. For example, Polish loan applicants visiting a particular branch on a particular day are randomly assigned to one of the Polish-speaking loan examiners working in that branch on that day. In contrast, native-born loan applicants are randomly assigned to the full set of examiners within each branch, including the Polish-speaking loan examiners. Importantly, the assigned loan examiner is given complete discretion to approve or reject these first-time loan applicants, leading to significant variation in approval rates across examiners.

We measure examiner leniency using a leave-out, residualized measure that accounts for the assignment process used by the Lender following Dahl, Kostøl, and Mogstad (2014) and Arnold, Dobbie, and Yang (Forthcoming). To construct this residualized examiner leniency measure, we first regress loan take-up on an exhaustive set of branch-by-month-by-nationality fixed effects, the level at which loan applicants are randomly assigned to loan examiners. We then use the residuals from this regression to calculate the leave-out mean examiner-by-branch take-up rate for each loan applicant. We calculate our instrument across all applicants assigned to an examiner within a branch to increase the precision of our leniency measure. In robustness checks, we present results that use an instrument that is allowed to vary by nationality, age, and gender, and results that use an instrument based on loan approval, not loan take-up.

Appendix Figure A3 presents the distribution of our residualized examiner leniency measure for loan take-up separately by nationality, age, and gender. Our sample includes 254 examiners, with the typical examiner-by-branch cell including 180 first-time loan applications. Controlling for branch-by-month-by-nationality fixed effects, our examiner leniency measure ranges from -0.165 to 0.195 with a standard deviation of 0.047. In other words, moving from the least to most lenient loan examiner increases the probability of loan take-up by 36.0 percentage points, a 55 percent change from the mean take-up rate of 66.1 percentage points.

D. Instrument Validity

Existence of First Stage: The first baseline assumption needed for our IV estimator is that examiner assignment is associated with loan take-up. To examine the first-stage relationship between examiner leniency (Z_{ite}) and loan take-up ($TakeUp_{ite}$), we estimate the following specification for applicant i assigned to examiner e at time t using a linear probability model:

$$TakeUp_{ite} = \gamma Z_{ite} + \pi \mathbf{X}_{it} + \mathbf{v}_{ite} \quad (10)$$

where, as described previously, Z_{ite} are leave-out (jackknife) measures of examiner leniency. The vector \mathbf{X}_{it} includes branch-by-month-by-nationality fixed effects and the baseline controls in Table 1. The error term \mathbf{v}_{ite} is composed of characteristics unobserved by the econometrician but observed by the examiner, as well as idiosyncratic variation unobserved to both the examiner and econometrician. Robust standard errors are clustered at examiner level.

Figure 2 provides graphical representations of the first stage relationship between our residualized measure of examiner leniency and the residualized probability of loan take-up that accounts for our exhaustive set of branch-by-month-by-nationality fixed effects, overlaid over the distribution of examiner leniency. Appendix Figure A3 presents the same results separately by nationality, age, and gender. Figure 2 and Appendix Figure A3 are a flexible analog to Equation (10), where we plot a local linear regression of residualized loan take-up against examiner leniency. The individual rate of residualized loan take-up is monotonically increasing in our leniency measure for all groups. The first stage relationship between loan take-up and examiner leniency is also linear over nearly the entire distribution of our examiner leniency measure, consistent with the identifying assumptions discussed in Section II.

Column 1 of Table 2 presents formal first stage results from Equation (10) for all applicants. Columns 1-6 of Appendix Table A2 present results separately by nationality, age, and gender. Consistent with the graphical results in Figure 2 and Appendix Figure A3, we find that our residualized examiner instrument is highly predictive of whether an individual receives a loan in both the full sample and within each subgroup. Table 2 shows, for example, that an applicant assigned to a loan examiner that is 10 percentage points more likely to approve a loan is 7.2 percentage points more

likely to receive a loan in the full sample. There are also a number of other applicant characteristics that are highly predictive of loan take-up. For example, women are 2.5 percentage points more likely to receive a loan compared to male applicants in the full sample, while applicants who are ten years older are 0.9 percentage points less likely to receive a loan compared to younger applicants. Loan take-up is also increasing with the credit score.

Exclusion Restriction: The second baseline assumption needed for our IV estimator is that examiner assignment only impacts applicant outcomes through the probability of receiving a loan. This assumption would be violated if examiner leniency is correlated with any unobservable determinants of future outcomes. Column 2 of Table 2 and columns 1-6 of Appendix Table A3 present a series of randomization checks to partially assess the validity of this exclusion restriction. Following the first stage results, we control for branch-by-month-by-nationality fixed effects and cluster standard errors at the examiner level. We find that examiners with differing leniencies are assigned observably identical applicants, both in the full sample and within each subgroup. None of the results suggest that there is systematic non-random assignment of applications to examiners.

The exclusion restriction could also be violated if examiner assignment impacts the profitability of a loan through channels other than loan take-up. The assumption that examiners only systematically affect loan outcomes through loan take-up is fundamentally untestable, but we argue that the exclusion restriction assumption is reasonable in our setting. Loan examiners only meet with applicants one time, and are forbidden, by law, to give advice or counsel applicants, leaving relatively little scope through which the assigned examiner could influence outcomes other than through loan take-up. Thus, it seems unlikely that loan examiners would significantly impact loan applicants other than through the loan approval decision.

Monotonicity: The final baseline assumption needed for our IV estimator is that the impact of examiner assignment on loan take-up is monotonic across loan applicants. In our setting, the monotonicity assumption requires that applicants who receive a loan when assigned to a strict examiner would also receive a loan when assigned to a more lenient examiner, and that applicants not receiving a loan when assigned to a lenient examiner would also not receive a loan when assigned to a stricter examiner. To partially test the monotonicity assumption, Appendix Figure A4 plots examiner leniency measures that are calculated separately for each examiner by nationality, age,

and gender. Consistent with our monotonicity assumption, examiners exhibit similar tendencies across observably different types of applicants. We also find a strong first-stage relationship across various applicant types in Appendix Table A2. None of the results suggest that the monotonicity assumption is invalid in our setting. In robustness checks, we also relax the monotonicity assumption by letting our leave-out measure of examiner leniency differ across applicant characteristics.

IV. Results

In this section, we present our main results applying our empirical test for bias in consumer lending. We then show the robustness of our results to alternative specifications, before comparing the results from our empirical test with standard tests based on OLS specifications.

A. Empirical Test for Bias

We estimate the profitability of marginal loan applicants using the following two-stage least squares specification for applicant i assigned to examiner e at time t :

$$Y_{ite} = \alpha_I^{IV} TakeUp_{ite} + B^{IV} TakeUp_{ite} \times TargetGroup_i + \beta \mathbf{X}_{it} + \mathbf{v}_{ite} \quad (11)$$

where Y_{ite} is the long-run profitability of the loan, as measured by the difference between total loan payments minus total loan disbursements. α_I^{IV} measures the profitability of the marginal loan to the reference group. B^{IV} is our measure of bias, or the difference in profitability between the reference and target group applicants. The vector \mathbf{X}_{it} includes branch-by-month-by-nationality fixed effects and the baseline controls listed in Table 1. As described previously, the error term $\mathbf{v}_{ite} = \mathbf{U}_{ite} + \varepsilon_{ite}$ consists of characteristics unobserved by the econometrician but observed by the loan examiner, \mathbf{U}_{ite} , and idiosyncratic variation unobserved by both the econometrician and examiner, ε_{ite} . We instrument for loan take-up, $TakeUp_{ite}$, with our measure of examiner leniency, Z_{ite} . We similarly instrument for the interaction of loan take-up and target group status, $TakeUp_{ite} \times TargetGroup_i$, with the interaction of our examiner leniency measure and group status, $Z_{ite} \times TargetGroup_i$. Robust standard errors are clustered at the examiner level.

Estimates from Equation (11) are presented in Table 3. Column 1 reports results pooling across all applicants. Columns 2-4 report results with interactions for applicant nationality, age, and gen-

der, respectively. Column 5 reports results with all interactions simultaneously. For completeness, Appendix Figure A3 provides a graphical representation of our reduced form results separately by nationality, age, and gender. Following the first stage results, we plot the reduced form relationship between our examiner leniency measure and the residualized profitability of loan take-up.

We find convincing evidence of bias against immigrants and older applicants using our IV estimator. We find that marginal loan applicants yield a profit of £331 following the initial loan decision (column 1), 24 percent larger than the mean profit level of £267. Marginal native-born applicants yield a profit of only £195, however, £566 less than marginal immigrant applicants (column 2), consistent with bias against immigrant applicants. We similarly find that marginal younger applicants yield a profit of £334, £348 less than marginal older applicants (column 3), consistent with bias against older applicants as well. These estimates imply that marginal immigrant applicants are almost four times more profitable than marginal native-born applicants, while marginal older applicants are more than two times as profitable as marginal younger applicants. In contrast, we find that marginal female and male applicants yield similar profits, suggesting no bias against (or in favor of) female applicants.

B. Robustness

Appendix Tables A4-A8 explore the sensitivity of our main results to a number of different specifications. Appendix Table A4 presents results where we instrument for loan approval, instead of loan take-up as in our preferred specification. Appendix Table A5 presents results where we use a net present value measure of long-run profits for a variety of different discount rates. Appendix Table A6 presents re-weighted estimates with the weights chosen to match the distribution of observable characteristics for target group loan applicants to explore whether differences in characteristics such as credit history or earnings can explain our results.⁸ Appendix Table A7 presents results from an MTE estimator that puts equal weight on each examiner in our sample, instead of the IV weights as

⁸Arnold, Dobbie, and Yang (Forthcoming) show that it is possible to test for bias holding fixed other group differences using a re-weighting procedure that weights the distribution of observables of the target group to match observables of the reference group in the spirit of DiNardo, Fortin, and Lemieux (1996) and Angrist, and Fernández-Val (2013). This narrower test for bias relies on the assumption that examiner preferences vary only by observable characteristics. See Appendix Table A9 for the complier characteristics used to construct the weights and Arnold, Dobbie, and Yang (Forthcoming) for additional details.

in our preferred specification.⁹ Finally, Appendix Table A8 presents estimates where the instrument is calculated separately for each subgroup in the data, relaxing the monotonicity assumption. Results are generally similar to our preferred specification across all alternative specifications, although some of our estimates lose statistical significance. In particular, our MTE estimates in Appendix Table A7 are particularly noisy due to the increased weight put on a handful of imprecise examiner-level estimates compared to our preferred specification. There is also considerably more noise when using smaller cells to calculate the leave-out examiner leniency measure in Appendix Table A8. The estimates are not economically or statistically different across specifications, however, and none of the results suggest that our preferred estimates are invalid.¹⁰

C. Comparison to OLS Estimates

Appendix Table A11 replicates the outcome tests from the prior literature (e.g., Han 2004) that rely on standard OLS estimates of Equation (11). In contrast to our IV test for bias, standard OLS estimates suggest much lower levels of bias against immigrant and older applicants. We find, for example, that the gap between the average native-born and immigrant applicant is only £102 (column 2), 82 percent lower than our IV estimate for marginal applicants in Table 3. The gap between the average younger and older applicant is also only £89 (column 3), 74 percent lower than our IV estimate for marginal applicants. Standard OLS estimates also suggest, incorrectly, that there is bias against female applicants (column 4). Taken together, these results highlight the

⁹While the MTE estimator has the advantage of allowing the researcher to choose any weighting scheme across examiners, it comes at the cost of additional functional form assumptions to interpolate estimates between observed values of the instrument. Following Arnold, Dobbie, and Yang (Forthcoming), we estimate these MTE results using a two-step procedure. In the first step, we estimate the entire distribution of MTEs using the derivative of residualized profits with respect to variation in the propensity score provided by our instrument. To do this, we regress the residualized profit variable on the residualized examiner leniency measure to calculate the group-specific propensity score. We then compute the numerical derivative of a local quadratic estimator to estimate group-specific MTEs (see Appendix Figure A5). In the second step, we use the group-specific MTEs to calculate the level of bias for each individual examiner, and the simple average of these examiner-specific estimates. We calculate standard errors by bootstrapping this two-step procedure at the examiner level.

¹⁰Our test for bias assumes that there are no differences in the true cost of lending to different groups. This assumption would be violated if, for example, there are differences in the systematic risk of lending to different groups. We explore this concern in Appendix Table A10, where we calculate long-run profits using a 10 percent discount rate for applicants from the reference group and a variety of higher discount rates for applicants from the target group. Our test is motivated by a standard CAPM model, where the higher discount rate for target group applicants captures the additional risk of lending to these applicants. For example, the largest discount rate difference we examine – 50 percent – implies that lending to target group applicants is more than seven times riskier than lending to reference group applicants at a market-risk premium of 5 percent. We continue to find evidence of bias against immigrants and older applicants in Appendix Table A10 even when we assume that these applicants have a 50 percent higher discount rate, however, indicating that our results are unlikely to be driven by differences in the systematic risk of lending to different groups.

importance of accounting for both infra-marginality and omitted variables when testing for bias in consumer lending.

V. Potential Mechanisms

In this section, we attempt to differentiate between three alternative forms of bias that could explain our findings: (1) prejudice, (2) inaccurate stereotypes, and (3) misaligned examiner incentives.

A. Prejudice

The first potential explanation for our results is that loan examiners either knowingly or unknowingly discriminate against immigrant and older applicants at the margin. Loan examiners could, for example, harbor explicit biases against immigrant and older applicants that leads them to exaggerate the cost of lending to these individuals (e.g., Becker 1957, 1993). Loan examiners could also harbor implicit biases against immigrant and older applicants, leading to biased lending decisions despite the lack of any explicit prejudice (e.g., Greenwald et al. 2009). However, immigrant applicants are typically matched to loan examiners from the same ethnic background and all loan examiners tend to be older themselves, institutional features that are inconsistent with most models of ethnic or age-related prejudice. We also find no bias against female applicants even among male examiners, another finding that is inconsistent with the simplest models of prejudice, although we note that these results are very imprecise (see Appendix Table A12).¹¹ These results suggest that either prejudice is not driving our results or that loan examiners are prejudiced against immigrant and older applicants despite sharing those same characteristics.

B. Inaccurate Stereotypes

A second potential explanation for our results is that loan examiners are making biased prediction errors, potentially due to inaccurate stereotypes against immigrant and older applicants. Bordalo et al. (2016) show, for example, that representativeness heuristics—probability judgments based on the most distinctive differences between groups—can exaggerate perceived differences between

¹¹We cannot estimate results separately by examiner ethnicity, as immigrant applicants are typically matched to loan examiners from a similar ethnic background. We also cannot estimate results separately by examiner age, as our data do not include this variable.

groups. In our setting, these kinds of group-based heuristics or inaccurate stereotypes could lead loan examiners to systematically underestimate the potential profitability of lending to immigrant and older applicants relative to native-born and younger applicants at the margin.

Following Arnold, Dobbie, and Yang (Forthcoming), we first explore whether our data are consistent with the formation of negative stereotypes that could lead to these kinds of biased prediction errors. Extending Bordalo et al. (2016) to our setting, negative stereotypes against immigrant and older loan applicants should only be present if these applicants are over-represented in the left tail of the predicted profit distribution compared to native-born and younger loan applicants. Appendix Figure A6 presents the distribution of the predicted long-run profits by nationality, age, and gender, where the predicted profits are calculated using the ML algorithm described below and the full set of baseline applicant characteristics in our data. We also present the likelihood ratios, $\mathbb{E}(x|target)/\mathbb{E}(x|reference)$, throughout the risk distribution to illustrate the most distinctive differences between each group. Results for each individual characteristic in our predicted risk measure are also presented in Appendix Table A9. In stark contrast to the predictions of the Bordalo et al. (2016) model, we find that both immigrant and older loan applicants are significantly over-represented in the right tail of the predicted profit distribution. For example, immigrant applicants are 2.1 times more likely than native-born applicants to be represented among the top 25 percent of the predicted profit distribution, while older loan applicants are 2.6 times more likely than younger applicants to be represented among the top 25 percent.

We can also test for biased prediction errors by examining situations where prediction errors of any kind are more likely to occur. One such test for biased prediction errors uses a comparison of experienced and inexperienced loan examiners, as examiners may be less likely to rely on inaccurate group stereotypes as they acquire greater on-the-job experience, at least in settings with limited information and contact. To test this idea, Appendix Table A13 presents subsample results for more and less experienced examiners, where we measure experience using an indicator for being employed by the Lender when our sample period begins. There are no systematic patterns by examiner experience and, if anything, the estimates suggest more bias against immigrants among more experienced loan examiners. Taken together with Appendix Figure A6, these results suggest that inaccurate stereotypes are also unlikely to be driving our results.

C. Misaligned Examiner Incentives

In this paper, we propose a new explanation for bias in consumer lending: the misalignment of firm and examiner incentives. The model of bias based on misaligned examiner incentives developed in Section II has two testable implications. The first is that we should find no bias when implementing the outcome test using default risk, the short-run outcome used to evaluate examiners in our setting. That is, we should find $\alpha_T^{SR} = \alpha_R^{SR}$ for all applicants, including for immigrants and older applicants. The second testable implication of our model is that we should only observe bias for groups where the short-run and long-run outcomes diverge. That is, we should find $E[\alpha_T^{SR} - \alpha_T | \mathbf{V}_i] > E[\alpha_R^{SR} - \alpha_R | \mathbf{V}_i]$ for immigrant and older applicants, but not female applicants. We test both implications below, before turning to a more general investigation of the examiners' decision-making process.

To test whether there is bias when implementing the outcome test using first-loan default, we estimate the default risk of marginal loan applicants using our IV estimator in Table 4. Consistent with the misalignment of examiner incentives, there are no statistically significant differences in the first-loan default risk of marginal loan applicants by nationality, age, or gender. Marginal first-loan loan applicants, in general, default on 44.7 percent of loans, 28 percent more than the average default rate of 35.0 percent. Marginal immigrant applicants, however, are only 2.5 percentage points less likely to default on the first loan than marginal native-born applicants. Marginal older applicants are 14.5 percentage points less likely to default on the first loan than marginal younger applicants, and marginal female applicants are 8.5 percentage points less likely to default than marginal male applicants, with none of these differences being statistically significant.

To test whether immigrant and older applicants are both high-profit and high-default, we plot the distributions of predicted long-run profits and predicted short-run default by nationality, age, and gender in Figure 3. We calculate predicted profits and predicted default risk using the ML algorithm described below. Consistent with the misalignment of examiner incentives, both immigrant and older applicants are visually more likely to default for a given level of long-run profits compared to native-born and younger applicants. In contrast, female and male applicants are equally as likely to default for a given level of long-run profits. To provide a more formal test of this hypothesis, Appendix Table A14 presents OLS results regressing predicted long-run profits on applicant characteristics and a quadratic in predicted short-run default. Controlling for predicted default, immigrant applicants

have predicted profits that are £36 larger than native applicants (column 1), while older applicants have predicted profits that are £31 larger than younger applicants. In contrast, male and female applicants have statistically identical predicted profits for a given level of predicted default.

We conclude this section with a more general investigation of the examiners’ decision-making process. The model of bias based on misaligned examiner incentives suggests that unbiased examiners rank applicants using long-run profits, α_i . In contrast, biased examiners rank applicants using short-run default, α_i^{SR} . We can therefore test for the misalignment of examiner incentives by comparing examiners’ decisions with two different data-based decision rules, one based on short-run default and the other based on long-run profits.

To implement this final test for misaligned examiner incentives, we first estimate predicted short- and long-run outcomes using a ML algorithm that efficiently uses all observable applicant characteristics. In short, we use a randomly-selected subset of the data to train the model using all individuals who receive a loan. In training the model, we must choose the shrinkage, the number of trees, and the depth of each tree. Following common practice (e.g., Kleinberg et al. 2018), we choose the smallest shrinkage parameter (i.e., 0.005) that allows the training process to run in a reasonable time frame. We use a five-fold cross-validation on the training sample in order to choose the optimal number of trees for the predictions. The interaction depth is set to four, which allows each tree to use at most four variables. Using the optimal number of trees from the cross-validation step, predicted outcomes are then created for the full sample.^{12,13}

One important challenge is that we only observe outcomes for applicants who receive a loan, not those who do not. This missing data problem makes it hard to evaluate counterfactual decision rules based on algorithmic predictions or to identify the implicit decision-rule used by loan examiners. To overcome this missing data problem, we follow Kleinberg et al. (2018) and start with the set of loan applicants receiving a loan from the most lenient examiners. From this set of applicants, we then

¹²Appendix Table A15 presents the correlates of our predicted profitability measure. Predicted profitability is increasing in the credit score used by the lender. Predicted profitability is also higher for female applicants, older applicants, and applicants with more dependents.

¹³One potential concern is that our measures of predicted profitability may be biased if loan examiners base their decisions on variables that are not observed by the econometrician (e.g., demeanor during the loan application). Following Kleinberg et al. (2018), we test for the importance of unobservables in loan decisions by splitting our sample into a training set to generate the profitability predictions and a test set to test those predictions. We find that our measure of predicted profitability from the training set is a strong predictor of true profitability in the test set, indicating that our measure of predicted profitability is not systematically biased by unobservables (see Appendix Figure A7).

choose additional applicants to hypothetically reject according to the predicted outcomes calculated using our ML algorithm. For each additional hypothetical rejection, this allows us to calculate the hypothetical change in profitability or default risk for the now smaller set of applicants that would have received a loan. Importantly, the hypothetical change in profitability and default risk can be compared to the outcomes produced by the stricter examiners because applicant characteristics are, on average, similar across examiners due to the quasi-random assignment of applicants to examiners.

Figure 4 presents these results for both long-run profits and short-run default. The solid black curve calculates the change in profitability or default that would have resulted if additional applicants had been rejected in order of the algorithm’s predicted profitability and default rates. Each of the points denotes the different examiner leniency quintiles. We also plot the change in profitability and default rates that would have resulted if we used a decision-rule based on a ML algorithm that does not include nationality, age, and gender; a decision-rule based on the baseline credit scores used to screen loan applicants; and a decision-rule based on a random number generator. We find that the decisions made by loan examiners are strikingly consistent with a data-based decision rule minimizing short-run default, but inconsistent with a decision rule maximizing long-run profits. Long-run profits actually decrease as we move from the most lenient to most strict examiners, worse than a decision-rule based on a random number generator and far worse than a decision rule based on our ML algorithm.¹⁴ In stark contrast, loan examiners are nearly as effective as our ML algorithm in decreasing default rates. The second quintile of examiners, for example, reduce the default rate by 2.3 percentage points relative to the most lenient quintile examiners by increasing the rejection rate by 8.5 percentage points. Our ML algorithm using all characteristics could have decreased the default rate by 4.8 percentage points with the same 8.5 percentage point increase in the rejection rate, or just 2.5 percentage points better than the loan examiners. The ML algorithm using only allowable characteristics could have similarly decreased the default rate by 4.1 percentage points, or 1.8 percentage points more than the loan examiners.¹⁵

¹⁴For example, the second quintile of examiners reduce profitability by £29 per applicant relative to the most lenient quintile examiners by increasing the rejection rate by 8.5 percentage points. In contrast, the ML algorithm using all characteristics could have increased profits by £38 per applicant with the same 8.5 percentage point increase in the rejection rate. The ML algorithm using only allowable characteristics could have increased profits by £33 per applicant, just £6 less than the full algorithm, while the credit score could have only increased profits by £9 per applicant.

¹⁵In contrast, the credit score used by the Lender could have decreased the default rate by only 0.5 percentage points, 1.8 percentage points less than the loan examiners and only 0.6 percentage points more than a random decision rule. The poor performance of the credit score variable is likely driven by the fact that the credit score

Taken together, our results suggest that loan examiners make biased lending decisions due to the misalignment of firm and examiner incentives. In contrast, we find limited evidence in support of the hypothesis that loan examiners are prejudiced or hold inaccurate negative stereotypes. These results are particularly important given the extensive prior evidence on incentive problems in credit markets (e.g., Hertzberg, Liberti, and Paravisini 2010, Keys et al. 2010, Berg, Puri, and Rocholl 2013, Agarwal, and Ben-David Forthcoming). Our results suggest that these kinds of poorly designed incentives can lead to biased and inefficient lending decisions.

VI. Conclusion

In this paper, we test for bias in consumer lending using the quasi-random assignment of loan examiners to identify the profitability of marginal loan applicants. We find evidence that there is substantial bias against immigrant and older loan applicants, ruling out statistical discrimination and omitted variable bias as the sole explanations for the disparities in credit availability for these groups. Our estimates are nearly identical if we account for observable differences by group, indicating that our results cannot be explained by differences in credit history or earnings.

We find several pieces of evidence consistent with our results being driven by the misalignment of firm and examiner incentives, as opposed to prejudice or inaccurate group stereotypes. First, immigrant applicants are typically matched to loan examiners from a similar ethnic background and loan examiners for all applicants tend to be older themselves, institutional features that are inconsistent with most models of ethnic or age-related prejudice. Second, we find that immigrant and older applicants are actually over-represented in the right tail of the payoff distribution, exactly the opposite pattern we would expect if inaccurate stereotypes were driving our results. Third, marginally approved immigrant and older loan applicants are just as likely to default as marginally approved native-born and younger loan applicants. Fourth, immigrant and older applicants are both more likely to default and more likely to yield higher than usual profits. Finally, we find that the decisions made by loan examiners are strikingly consistent with a data-based decision rule minimizing short-run default, but inconsistent with a decision rule maximizing long-run profits.

The findings from this paper have a number of important implications. If the misalignment of

variable purchased by the Lender is calibrated to entire credit market, not the subprime market that the Lender operates in.

firm and lender incentives is an important driver of bias in consumer lending decisions, then a data-based decision rule based on long-run profits could simultaneously increase profits and eliminate bias. In our setting, for example, the Lender would earn approximately £157 more in profit per applicant if marginal lending decisions were made using the machine learning algorithm rather than loan examiners, a 58 percent increase from the mean. Including all applicants, not just those at the margin of loan take-up, the Lender would earn over £53 more per applicant if lending decisions were made using the machine learning algorithm, a 30 percent increase from the mean.¹⁶

We can also test whether a data-based decision rule is biased by comparing the observed and predicted profitability of applicants from different groups.¹⁷ Following the logic of our outcome test developed above, the observed profitability at each level of predicted profitability will be identical across groups if our algorithm is unbiased. In contrast, if our algorithm is biased against, say, non-native applicants, the observed profitability of non-natives at a given level of predicted profitability will be higher than the observed profitability of natives. There is no evidence of bias by nationality, age, or gender (see Appendix Figure A8).

There are two important caveats to our analysis. First, bias in other consumer lending decisions such as mortgage or automobile loans may be very different than the setting examined here. Second, the welfare effects of credit, particularly high-cost credit, are largely unknown (e.g., Agarwal, Skiba, and Tobacman 2009, Melzer 2011, Morgan, Strain, and Seblani 2012, Morse 2011, Gathergood, Guttman-Kenney, and Hunt 2014, Liberman, Paravisini, and Pathania 2016, Zaki 2016). Given these concerns, we are unable to estimate the welfare effects of bias in consumer lending decisions using our data and research design. Developing a framework to assess the precise welfare effects of bias in consumer credit decisions is an important area of future work.

¹⁶Following Kleinberg et al. (2018), we compute the gains from the machine learning algorithm relative to the average examiner in our test sample using a three step process. First, we impute long-run profits for applicants who never took up a loan under the assumption that our algorithm perfectly predicts true long-run profits and that all selection is on observable characteristics. Second, we re-rank all applicants, both rejected and approved, and select a hypothetical group to approve using predicted long-run profits. Finally, we compare the predicted long-run profitability of applicants for both the algorithm and average examiner, both at the margin of loan approval and aggregating over all loan applicants.

¹⁷There is an emerging literature investigating the effects of algorithmic lending on racial disparities in U.S. credit markets. Fuster et al. (2017), for example, show that improvements in algorithmic lending may increase racial disparities in the U.S. mortgage market, even when these improvements increase the overall amount of lending. In contrast, Bartlett et al. (2018) show that there are significantly smaller racial disparities among FinTech mortgage lenders, who are more likely to rely on algorithmic lending rules, compared to traditional mortgage lenders, who are more likely to rely on loan examiners. These articles do not test whether algorithmic lending rules are biased, however.

References

- Agarwal, Sumit, and Itzhak Ben-David.** Forthcoming. “Loan Prospecting and the Loss of Soft Information.” *Journal of Financial Economics*.
- Agarwal, Sumit, Paige Marta Skiba, and Jeremy Tobacman.** 2009. “Payday Loans and Credit Cards: New Liquidity and Credit Scoring Puzzles?” *American Economic Review*, 99(2): 412–417.
- Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Strobel.** 2015. “Regulating Consumer Financial Products: Evidence from Credit Cards.” *Quarterly Journal of Economics*, 130(1): 111–164.
- Aigner, Dennis J., and Glen G. Cain.** 1977. “Statistical Theories of Discrimination in Labor Markets.” *Industrial and Labor Relations Review*, 30(2): 175–187.
- Alesina, Alberto F., Francesca Lotti, and Paolo Emilio Mistrulli.** 2013. “Do Women Pay More for Credit? Evidence from Italy.” *Journal of the European Economic Association*, 11(1): 45–66.
- Angrist, Joshua, and Ivan Fernández-Val.** 2013. “ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework.” *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, ed. D. Acemoglu, M. Arellano, and E. Dekel Vol. 51, 401–434. Cambridge University Press.
- Arnold, David, Will Dobbie, and Crystal S. Yang.** Forthcoming. “Racial Bias in Bail Decisions.” *Quarterly Journal of Economics*.
- Arrow, Kenneth.** 1973. “The Theory of Discrimination.” *Discrimination in Labor Markets*, ed. Orley Ashenfelter, and Albert Rees, 3–33. Princeton University Press.
- Ayres, Ian.** 2002. “Outcome Tests of Racial Disparities in Police Practices.” *Justice Research and Policy*, 4(1-2): 131–142.
- Bartlett, Richard, Adair Morse, Richard Stanton, and Nany Wallace.** 2018. “Consumer Lending Discrimination in the FinTech Era.” *Unpublished Working Paper*.

- Bayer, Patrick, Fernando Ferreira, and Stephen L. Ross.** 2017. "What Drives Racial and Ethnic Differences in High-Cost Mortgages? The Role of High-Risk Lenders." *The Review of Financial Studies*, 31(1): 175–205.
- Becker, Gary S.** 1957. *The Economics of Discrimination*. University of Chicago Press.
- Becker, Gary S.** 1993. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy*, 101(3): 385–409.
- Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro.** 2010. "Does Gender Matter in Bank-Firm Relationships? Evidence from Small Business Lending." *Journal of Banking & Finance*, 34(12): 2968–2984.
- Berg, Tobias, Manju Puri, and Jorg Rocholl.** 2013. "Loan Officer Incentives and the Limits of Hard Information." *NBER Working Paper No. 19051*.
- Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan.** 1994. "Race, Redlining, and Residential Mortgage Loan Performance." *The Journal of Real Estate Finance and Economics*, 9(3): 263–294.
- Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan.** 1998. "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending." *Review of Economics and Statistics*, 80(2): 241–250.
- Blanchflower, David G., Phillip B. Levine, and David J. Zimmerman.** 2003. "Discrimination in the Small-Business Credit Market." *Review of Economics and Statistics*, 85(4): 930–943.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2016. "Stereotypes." *Quarterly Journal of Economics*, 131(4): 1753–1794.
- Buchak, Greg, and Adam Jørring.** 2016. "Does Competition Reduce Racial Discrimination in Lending?" *Unpublished Working Paper*.
- Cavalluzzo, Ken S., and Linda C. Cavalluzzo.** 1998. "Market Structure and Discrimination: The Case of Small Businesses." *Journal of Money, Credit and Banking*, 30(4): 771–792.

- Cavalluzzo, Ken S., Linda C. Cavalluzzo, and John D. Wolken.** 2002. "Competition, Small Business Financing, and Discrimination: Evidence from a New Survey." *The Journal of Business*, 75(4): 641–679.
- Charles, Kerwin Kofi, and Erik Hurst.** 2002. "The Transition to Home Ownership and the Black-White Wealth Gap." *Review of Economics and Statistics*, 84(2): 281–297.
- Charles, Kerwin Kofi, Erik Hurst, and Melvin Stephens.** 2008. "Rates for Vehicle Loans: Race and Loan Source." *American Economic Review*, 98(2): 315–320.
- Cohen-Cole, Ethan.** 2011. "Credit Card Redlining." *Review of Economics and Statistics*, 93(2): 700–713.
- Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg.** 2016. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." *Labour Economics*, 41: 47–60.
- Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad.** 2014. "Family Welfare Cultures." *Quarterly Journal of Economics*, 129(4): 1711–1752.
- Deku, Solomon Y., Alper Kara, and Philip Molyneux.** 2016. "Access to Consumer Credit in the UK." *The European Journal of Finance*, 22(10): 941–964.
- DiNardo, John, Nicole M. Fortin, and Thomas Lemieux.** 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." *Econometrica*, 64(5): 1001–1044.
- Dobbie, Will, and Jae Song.** 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." *American Economic Review*, 105(3): 1272–1311.
- Dobbie, Will, Paul Goldsmith-Pinkham, and Crystal S. Yang.** 2017. "Consumer Bankruptcy and Financial Health." *Review of Economics and Statistics*, 99(5): 853–869.
- Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther.** 2017. "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *Unpublished Working Paper*.

- Gathergood, John, Ben Guttman-Kenney, and Stefan Hunt.** 2014. “How Do Payday Loans Affect Consumers?” *Unpublished Working Paper*.
- Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji.** 2009. “Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity.” *Journal of Personality and Social Psychology*, 97(1): 17–41.
- Hanson, Andrew, Zackary Hawley, Hal Martin, and Bo Liu.** 2016. “Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment.” *Journal of Urban Economics*, 92: 48–65.
- Han, Song.** 2004. “Discrimination in Lending: Theory and Evidence.” *The Journal of Real Estate Finance and Economics*, 29(1): 5–46.
- Heider, Florian, and Roman Inderst.** 2012. “Loan Prospecting.” *The Review of Financial Studies*, 25(8): 2381–2415.
- Hertzberg, Andrew, Jose Maria Liberti, and Daniel Paravisini.** 2010. “Information and Incentives Inside the Firm: Evidence from Loan Officer Rotation.” *Journal of Finance*, 65(3): 795–828.
- Imbens, Guido W., and Joshua D. Angrist.** 1994. “Identification and Estimation of Local Average Treatment Effects.” *Econometrica*, 62(2): 467–475.
- Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig.** 2010. “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans.” *Quarterly Journal of Economics*, 125(1): 307–362.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan.** 2018. “Human Decisions and Machine Predictions.” *Quarterly Journal of Economics*, 133(1): 237–293.
- Liberman, Andres, Daniel Paravisini, and Vikram Pathania.** 2016. “High-Cost Debt and Borrower Reputation: Evidence from the U.K.” *Unpublished Working Paper*.
- Marx, Philip.** 2017. “An Absolute Test of Racial Prejudice.” *Unpublished Working Paper*.

- Melzer, Brian T.** 2011. "The Real Costs of Credit Access: Evidence from the Payday Lending Market." *Quarterly Journal of Economics*, 126(1): 517–555.
- Morgan, Donald P., Michael R. Strain, and Ihab Seblani.** 2012. "How Payday Credit Access Affects Overdrafts and Other Outcomes." *Journal of Money, Credit and Banking*, 44(2-3): 519–531.
- Morse, Adair.** 2011. "Payday Lenders: Heroes or Villains?" *Journal of Financial Economics*, 102(1): 28–44.
- Peterson, Richard L.** 1981. "An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending." *The Bell Journal of Economics*, 12(2): 547–561.
- Phelps, Edmund S.** 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review*, 62(4): 659–661.
- Pope, Devin G., and Justin R. Sydnor.** 2011. "What's in a picture? Evidence of Discrimination from Prosper.com." *Journal of Human Resources*, 46(1): 53–92.
- Ross, Stephen L., Margery Austin Turner, Erin Godfrey, and Robert R. Smith.** 2008. "Mortgage Lending in Chicago and Los Angeles: A Paired Testing Study of the Pre-Application Process." *Journal of Urban Economics*, 63(3): 902–919.
- Stegman, Michael A.** 2007. "Payday Lending." *Journal of Economic Perspectives*, 21(1): 169–190.
- Van Order, Robert, Vassilis Lekkas, and John M. Quigley.** 1993. "Loan Loss Severity and Optimal Mortgage Default." *Real Estate Economics*, 21(4): 353–371.
- Zaki, Mary.** 2016. "Access to Short-Term Credit and Consumption Smoothing Within the Paycycle." *Unpublished Working Paper*.

Table 1: Descriptive Statistics

	All Loans	Loan Taken-Up	Loan Not Taken-Up
	(1)	(2)	(3)
<i>Panel A: Applicant Characteristics</i>			
Immigrant	0.402	0.409	0.388
Age	33.880	33.631	34.365
Female	0.559	0.570	0.537
One-Plus Years at Residence	0.731	0.772	0.652
Married	0.415	0.373	0.496
Number of Dependents	0.973	1.054	0.815
Credit Score	538.166	545.893	522.431
Has Bank Account	0.912	0.938	0.862
Has Other Loan Payments	0.288	0.349	0.168
Loan Amount Requested (£)	409.270	396.519	434.117
Loan for Emergency	0.266	0.267	0.265
Loan for Large One-Time Expense	0.106	0.109	0.100
Loan for Overdraft Avoidance	0.052	0.053	0.052
Loan for Shopping or Holiday	0.230	0.233	0.226
<i>Panel B: Loan Characteristics</i>			
Loan APR (%)	—	663.139	—
Loan Duration (Months)	—	5.498	—
Loan Amount Net of Fees (£)	—	289.897	—
<i>Panel C: Loan Outcomes</i>			
Loan Take-up	0.661	1.000	0.000
Loan Approved	0.752	1.000	0.269
Loan Default	0.231	0.350	0.000
Loan Top-Up	0.289	0.438	0.000
Total Profits (£)	176.465	267.014	0.000
Observations	45687	30192	15495

Note: This table reports descriptive statistics. The sample consists of first-time loan applicants assigned to a loan examiner between 2012 and 2014. We drop online and phone applicants, applicants younger than 18 or older than 75 years old, applicants assigned to examiners with fewer than 50 observations, and applicants that are unique to a store-by-month-by-nationality cell. Loan uses are self-reported at the time of application. Total profits are the sum of all payments made from the applicant to the Lender over the course of their entire relationship minus all disbursements from the Lender to the applicant. Loan default, top-up, and profits are all equal to zero for applicants not taking out a loan. See the data appendix for additional details on the variable construction.

Table 2: First Stage and Balance Tests

	Loan Take-Up	Examiner Leniency
	(1)	(2)
Examiner Leniency	0.71923*** (0.07353)	
Age	-0.00089*** (0.00024)	-0.00001 (0.00002)
Female	0.02448*** (0.00486)	-0.00021 (0.00063)
One-Plus Years at Residence	0.10167*** (0.00865)	-0.00068 (0.00110)
Married	-0.08323*** (0.00692)	-0.00002 (0.00106)
Number of Dependents	0.03306*** (0.00255)	0.00006 (0.00021)
Credit Score (/1000)	1.54585*** (0.05029)	0.00762* (0.00420)
Has Bank Account	0.13765*** (0.01201)	0.00372** (0.00160)
Has Other Loan Payments	0.21579*** (0.00870)	-0.00028 (0.00087)
Loan Amount Requested (£/1000)	-0.09883*** (0.00624)	-0.00045 (0.00071)
Loan for Emergency	0.00366 (0.00553)	-0.00125 (0.00340)
Loan for Large One-Time Expense	0.02028** (0.00871)	0.00092 (0.00387)
Loan for Overdraft Avoidance	-0.00130 (0.00865)	-0.00133 (0.00380)
Loan for Shopping or Holiday	0.00671 (0.00571)	-0.00108 (0.00337)
Dep. Variable Mean	0.661	0.000
Observations	45687	45687
p-value on Joint F-test	[0.000]	[0.420]
Clusters	254	254

Note: This table reports first stage results and balance tests. The regressions are estimated on the sample described in the notes to Table 1. Examiner leniency is estimated using data from other loan applicants assigned to an examiner following the procedure described in Section III. Column 1 reports estimates from an OLS regression of loan take-up on the variables listed. Column 2 reports estimates from an OLS regression of examiner leniency on the variables listed. The p-value reported at the bottom of the columns is for an F-test of the joint significance of the variables listed in the rows. All specifications control for store-by-month-by-nationality fixed effects. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 3: Loan Take-Up and Long-Run Profits

	2SLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Take-Up	331.100*** (60.459)	195.215*** (63.564)	334.254*** (97.115)	168.781* (96.553)	-45.605 (143.324)
Take-Up x Immigrant Applicant		566.015*** (172.909)			603.237*** (187.343)
Take-Up x Older Applicant			347.803** (161.218)		348.719** (160.881)
Take-Up x Female Applicant				-5.399 (125.283)	118.381 (130.406)
Dep. Variable Mean	176.533	176.533	176.533	176.533	176.533
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

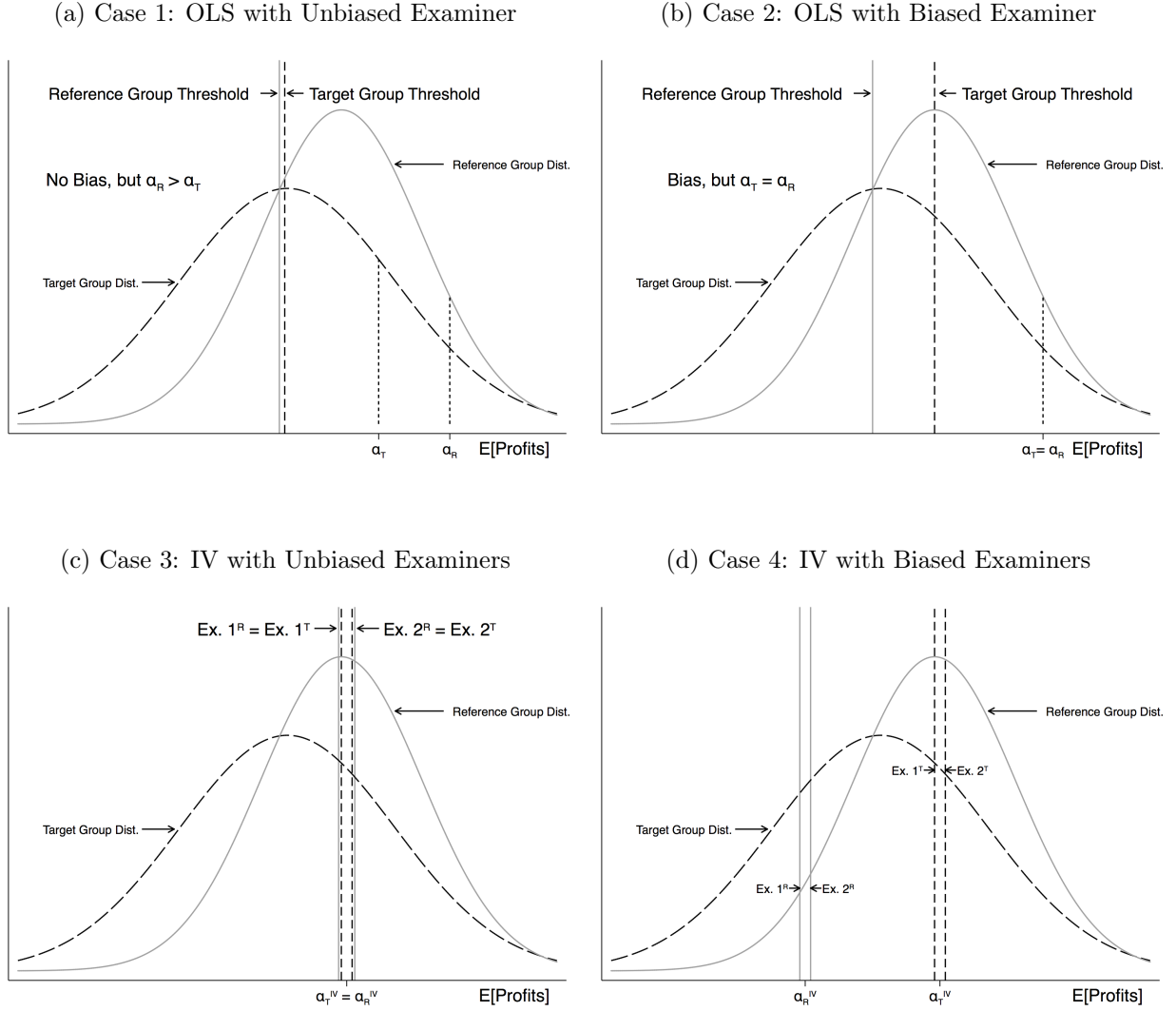
Note: This table reports IV estimates of bias in consumer lending decisions using long-run profits as an outcome. Each column reports two-stage least squares estimates of the impact of loan take-up on long-run profits using the sample described in the notes to Table 1. We instrument for loan take-up using the leave-out examiner leniency measure constructed using the procedure described in Section III, and for the interaction of loan take-up and applicant characteristics using the interaction of leave-out leniency and the same characteristic. All specifications control for store-by-month-by-nationality fixed effects and the baseline characteristics listed in Panel A of Table 1. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 4: IV Estimates for Short-Run Default

	2SLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Take-Up	0.447*** (0.050)	0.453*** (0.057)	0.497*** (0.085)	0.515*** (0.071)	0.588*** (0.111)
Take-Up x Immigrant Applicant		-0.025 (0.114)			-0.058 (0.123)
Take-Up x Older Applicant			-0.145 (0.103)		-0.149 (0.102)
Take-Up x Female Applicant				-0.085 (0.115)	-0.099 (0.120)
Dep. Variable Mean	0.231	0.231	0.231	0.231	0.231
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

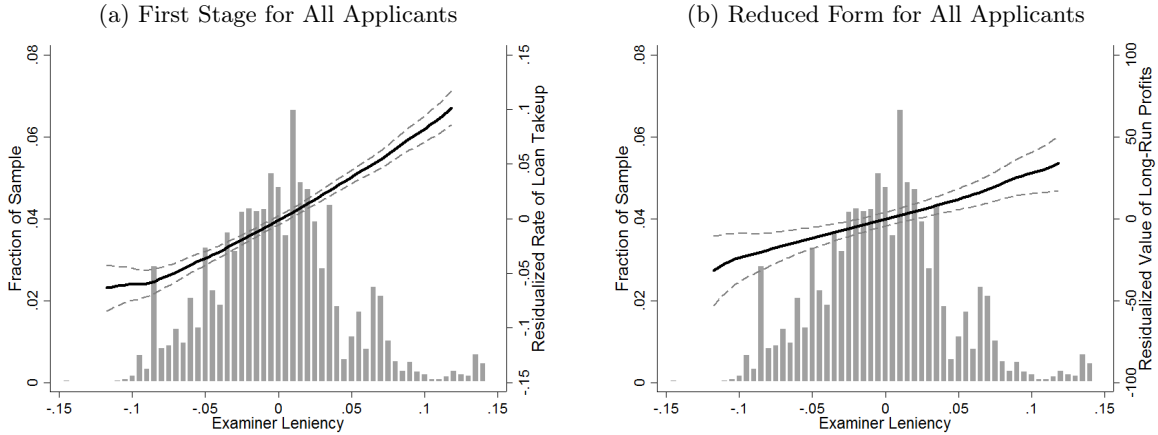
Note: This table reports IV estimates of bias in consumer lending using short-run default as an outcome. Each column reports two-stage least squares estimates using the sample described in the notes to Table 1. We instrument for loan take-up using the leave-out examiner leniency measure constructed using the procedure described in Section III, and for the interaction of loan take-up and applicant characteristics using the interaction of leave-out leniency and the same characteristic. All specifications control for store-by-month-by-nationality fixed effects and the baseline characteristics listed in Panel A of Table 1. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 1: Illustration of Estimation Problem



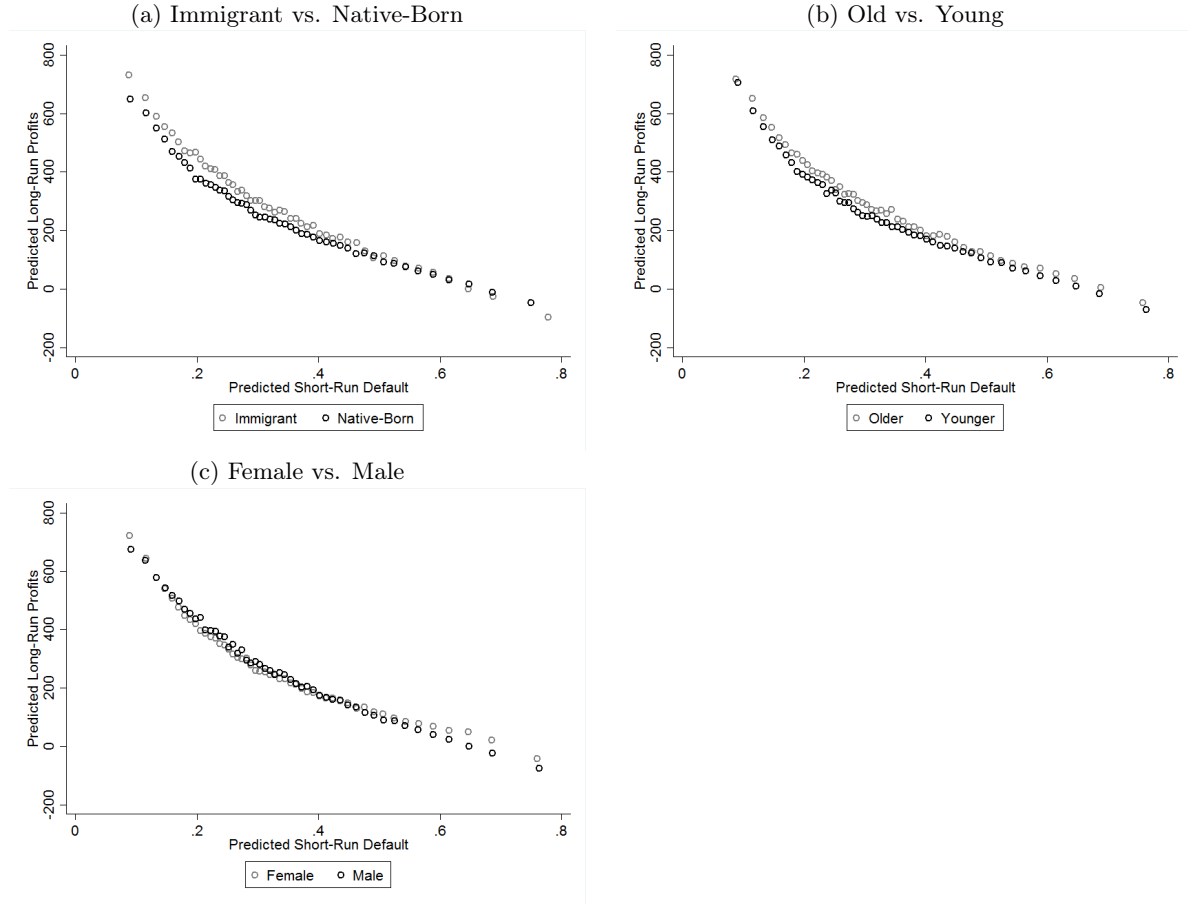
Note: These figures report hypothetical risk distributions for reference and target group applicants. Panel A illustrates the OLS estimator for an unbiased examiner who chooses the same threshold for take-up for both reference and target applicants. Panel B illustrates the OLS estimator for a biased examiner who chooses a higher threshold for loan take-up for target applicants than reference applicants. Panel C illustrates the IV estimator for two unbiased examiners. Panel D illustrates the IV estimator with two biased examiners. See the text for additional details.

Figure 2: First Stage and Reduced Form Results



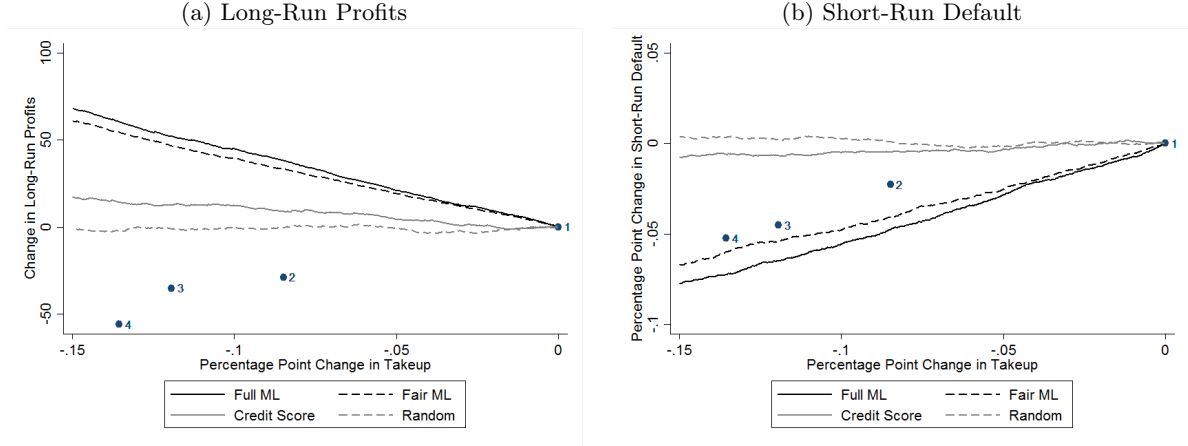
Note: These figures report the first stage and reduced form relationships between applicant outcomes and examiner leniency. The regressions are estimated on the sample described in the notes to Table 1. Examiner leniency is estimated using data from other applicants assigned to a loan examiner following the procedure described in Section III. In the first stage regression, the solid line represents a local linear regression of loan take-up on examiner leniency. In the reduced form regression, the solid line represents a local linear regression of long-run profits on examiner leniency. Loan take-up and long-run profits are residualized using store-by-month-by-nationality fixed effects. Standard errors are clustered at the examiner level.

Figure 3: Joint Distributions of Machine Learning Predictions



Note: These figures report the relationship between predicted long-run profits and predicted short-run default separately by group. Predicted long-run profits and predicted short-run default are obtained using the machine learning algorithm described in Section V. See the text for additional details.

Figure 4: Comparing Additional Profits and Defaults by Ranking Method



Note: These figures examine the performance of different data-based decision rules versus the actual decisions made by stricter loan examiners. The rightmost point in the graph represents the loan outcomes and loan take-up rate of the most lenient bin of examiners. The additional three points on the graph show loan outcomes and take-up rates for the actual decisions made by the second through fourth most lenient bins of examiners. Each line shows the loan outcome and take-up trade-off that comes from denying additional applicants within the most lenient bin of examiners' approval set using different data-based decision rules. The solid black line shows the trade-off when using the machine learning algorithm described in Section V trained using all available variables; the dashed black line for the same machine learning algorithm omitting nationality, gender, and age; the solid gray line for the credit score used to screen applicants; and the dashed gray line for randomly rejecting applicants. Panel A presents these results for long-run profits. Panel B presents these results for short-run default. See the text for additional details.

Appendix A: Additional Results

Appendix Table A1: Correlation Between IV Weights and Examiner Observables

	Applicant Nationality		Applicant Age		Applicant Gender	
	Immigrant IV Weights x 100	Native-Born IV Weights x 100	Old IV Weights x 100	Young IV Weights x 100	Female IV Weights x 100	Male IV Weights x 100
	(1)	(2)	(3)	(4)	(5)	(6)
Examiner Discrimination (1000s)	-0.022 (0.043)	-0.005 (0.039)	0.042 (0.036)	0.027 (0.036)	-0.025 (0.095)	-0.032 (0.099)
Examiner Case Load (100s)	0.007 (0.005)	0.001 (0.003)	0.001 (0.005)	0.002 (0.004)	0.006* (0.004)	0.008* (0.005)
Examiner Average Leniency	-0.076 (0.206)	-0.137 (0.190)	0.090 (0.189)	0.013 (0.191)	-0.059 (0.185)	-0.053 (0.193)
Experienced Examiner	0.014 (0.023)	0.015 (0.021)	0.012 (0.021)	0.016 (0.022)	0.013 (0.021)	0.016 (0.022)
Male Examiner	-0.026 (0.023)	-0.024 (0.021)	-0.025 (0.021)	-0.025 (0.021)	-0.025 (0.021)	-0.025 (0.022)
Dep. Variable Mean	0.394	0.394	0.394	0.394	0.394	0.394
Observations	254	254	254	254	254	254

Note: This table reports the relationship between instrumental variable weights assigned to a specific examiner cell and observables of the examiner cell by group. To ease readability, the weights are multiplied by 100. To compute the IV weights assigned to an examiner cell, we compute continuous IV weights following the procedure described in Cornelissen et al. (2016). We then discretize these continuous weights by assigning each examiner the weight associated with his or her average leniency. Finally, we divide the discretized examiner weights by the sum total to ensure the discretized weights sum to one. Examiner discrimination is calculated using the MTE procedure described in the text. Examiner case load is measured over the entire sample period. Examiner experience is an indicator for being employed by the Lender at the start of the sample period. See the data appendix for additional information on the construction of the variables. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A2: First Stage Results by Applicant Characteristics

	Applicant Nationality		Applicant Age		Applicant Gender	
	Immigrant	Native-Born	Old	Young	Female	Male
	(1)	(2)	(3)	(4)	(5)	(6)
Loan Take-Up	0.552*** (0.095)	0.792*** (0.082)	0.719*** (0.080)	0.716*** (0.090)	0.751*** (0.074)	0.635*** (0.113)
Dep. Variable Mean	0.672	0.653	0.647	0.672	0.674	0.644
Observations	18234	27329	20461	24038	24948	19601
Clusters	254	254	254	254	254	254

Note: This table reports first stage results of examiner leniency on loan take-up estimated separately by group. The regressions are estimated on the sample as described in the notes to Table 1. Examiner leniency is estimated using data from other loan applicants assigned to an examiner following the procedure described in Section III. Observations in subgroups do not always add to the full sample size due to dropping of singleton observations. All specifications control for store-by-month-by-nationality fixed effects and the baseline characteristics listed in Panel A of Table 1. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A3: Balance Tests by Subgroup

	Applicant Nationality		Applicant Age		Applicant Gender	
	Immigrant (1)	Native-Born (2)	Old (3)	Young (4)	Female (5)	Male (6)
Female	0.0003 (0.00066)	-0.00030 (0.00084)	-0.00024 (0.00082)	-0.00038 (0.00082)		
Age	0.0001 (0.00003)	-0.00002 (0.00002)	0.00001 (0.00004)	-0.00001 (0.00008)	-0.00001 (0.00003)	-0.00002 (0.00003)
One-Plus Years at Residence	0.00058 (0.00087)	-0.00183 (0.00155)	-0.00146 (0.00142)	-0.00026 (0.00114)	-0.00124 (0.00143)	0.00016 (0.00097)
Married	0.00110 (0.00113)	-0.00081 (0.00125)	-0.00046 (0.00101)	0.00024 (0.00141)	-0.00015 (0.00109)	0.00021 (0.00141)
Number of Dependents	0.00021 (0.00035)	-0.00005 (0.00023)	0.00025 (0.00028)	-0.00004 (0.00030)	0.00012 (0.00022)	-0.00008 (0.00039)
Credit Score	-0.00000 (0.00001)	0.00001** (0.00001)	0.00001 (0.00001)	0.00001* (0.00001)	0.00001 (0.00001)	0.00001** (0.00001)
Has Bank Account	0.00345* (0.00192)	0.00387** (0.00179)	0.00284 (0.00184)	0.00454*** (0.00167)	0.00313* (0.00174)	0.00473*** (0.00178)
Has Other Loan Payments	0.00038 (0.00120)	-0.00061 (0.00099)	-0.00021 (0.00109)	-0.00026 (0.00098)	0.00008 (0.00100)	-0.00083 (0.00105)
Loan Amount Requested (£)	-0.00000 (0.00000)	-0.00000 (0.00000)	0.00000 (0.00000)	-0.00000 (0.00000)	0.00000 (0.00000)	-0.00000** (0.00000)
Loan for Emergency	-0.00322 (0.00273)	-0.00008 (0.00410)	-0.00165 (0.00347)	-0.00119 (0.00359)	-0.00098 (0.00376)	-0.00137 (0.00319)
Loan for Large One-Time Expense	-0.00138 (0.00276)	0.00317 (0.00523)	0.00148 (0.00439)	0.00038 (0.00384)	0.00287 (0.00457)	-0.00030 (0.00345)
Loan for Overdraft Avoidance	-0.00114 (0.00356)	-0.00120 (0.00425)	-0.00142 (0.00402)	-0.00178 (0.00389)	0.00018 (0.00432)	-0.00305 (0.00345)
Loan for Shopping or Holiday	-0.00175 (0.00253)	-0.00071 (0.00411)	-0.00228 (0.00342)	-0.00035 (0.00354)	-0.00087 (0.00362)	-0.00091 (0.00328)
Dep. Variable Mean	0.001 18234	-0.000 27329	0.000 20461	0.000 24038	-0.000 24948	0.001 19601
Observations						
p-value on Joint F-test	[0.638]	[0.225]	[0.833]	[0.317]	[0.860]	[0.029]
Clusters	254	254	254	254	254	254

Note: This table reports balance tests separately by group. The regressions are estimated on the sample as described in the notes to Table 1. The dependent variable is examiner leniency, which is estimated using data from other loan applicants assigned to an examiner following the procedure described in Section III. The p-value reported at the bottom of the columns is for a F-test of the joint significance of the variables listed in the rows. Observations in the subgroups do not always add to the full sample size due to dropping of singleton observations. All specifications control for store-by-month-by-nationality fixed effects. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A4: IV Estimates Using Loan Approved

	2SLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Approved	351.465*** (90.555)	214.888** (90.611)	359.331** (141.589)	171.385 (136.913)	-6.657 (188.256)
Approved x Immigrant Applicant		612.468** (249.385)			609.037** (275.732)
Approved x Older Applicant			379.361* (225.900)		357.771 (221.049)
Approved x Female Applicant				-12.464 (169.684)	83.156 (169.625)
Dep. Variable Mean	176.533	176.533	176.533	176.533	176.533
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

Note: This table reports robustness results testing for bias in loan approvals, rather than loan take-up. See the notes to Table 3 for details on the sample and empirical specification. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A5: IV Estimates Using Discounted Long-Run Profits

	Discount Rates					
	5 Percent (1)	10 Percent (2)	20 Percent (3)	30 Percent (4)	40 Percent (5)	50 Percent (6)
Loan Take-Up	-4.910 (129.132)	-9.328 (119.363)	-18.714 (102.651)	-27.132 (88.967)	-34.309 (77.775)	-40.510 (68.607)
Take-Up x Immigrant Applicant	495.398*** (176.610)	456.328*** (163.295)	389.173*** (140.305)	333.695*** (121.476)	285.857*** (106.043)	245.942*** (93.625)
Take-Up x Older Applicant	289.628** (142.971)	265.695** (131.549)	223.415** (111.723)	187.240** (95.195)	157.245* (81.648)	132.281* (70.609)
Take-Up x Female Applicant	74.871 (118.907)	73.357 (109.905)	71.596 (94.611)	69.961 (82.193)	67.328 (72.123)	64.662 (63.927)
Dep. Variable Mean	185.789	170.494	143.478	120.337	100.458	83.292
Observations	45687	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254	254

Note: This table reports IV estimates of bias in consumer lending using discounted long-run profits as the outcome. Long-run profits are discounted using the discount rate listed in the column header. See the notes to Table 3 for additional details. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A6: Re-Weighted IV Estimates

	2SLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Take-Up	331.100*** (60.459)	192.886*** (64.867)	380.626*** (96.777)	170.112* (97.563)	-25.766 (141.033)
Take-Up x Immigrant Applicant		365.144* (193.036)			550.917*** (191.163)
Take-Up x Older Applicant			313.343* (168.429)		397.315** (168.378)
Take-Up x Female Applicant				105.317 (130.367)	135.918 (126.731)
Dep. Variable Mean	176.533	176.533	176.533	176.533	176.533
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

Note: This table reports robustness results testing for bias when the estimates are re-weighted so that the observable characteristics of target group applicants match the distribution of observable characteristics of reference group applicants. See Appendix Table A9 for the complier characteristics used to construct the weights and Arnold, Dobbie, and Yang (Forthcoming) for additional details. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A7: Marginal Treatment Effect Estimates

	MTE Estimates			
	(1)	(2)	(3)	(4)
Loan Take-Up	407.290** (172.240)	271.892* (163.177)	234.724* (138.360)	378.160** (157.663)
Loan Take-Up x Immigrant Applicant		355.785 (302.167)		
Loan Take-Up x Older Applicant			369.855* (208.183)	
Loan Take-Up x Female Applicant				54.547 (205.864)
Dep. Variable Mean	176.533	176.533	176.533	176.533
Observations	45687	45687	45687	45687

Note: This table reports robustness results testing for bias using an MTE estimator that puts equal weight on each examiner. We estimate these MTE results using a two-step procedure. In the first step, we estimate the entire distribution of MTEs using the derivative of residualized profits with respect to variation in the propensity score provided by our instrument. To do this, we regress the residualized profit variable on the residualized examiner leniency measure to calculate the group-specific propensity score. We then compute the numerical derivative of a local quadratic estimator to estimate group-specific MTEs that are presented in Appendix Figure A5. In the second step, we use these group-specific MTEs to calculate the level of bias for each examiner and the average level of bias across all loan examiners. We calculate standard errors by bootstrapping this two-step procedure at the examiner level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A8: IV Estimates Using Group-Specific Leniency

	2SLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Take-Up	333.146*** (60.027)	170.890** (68.293)	444.428*** (155.915)	81.893 (157.320)	-70.020 (459.851)
Take-Up x Immigrant Applicant		697.223 (441.663)			635.336 (901.813)
Take-Up x Older Applicant			469.603* (252.789)		382.091 (381.573)
Take-Up x Female Applicant				-79.943 (177.468)	132.286 (437.907)
Dep. Variable Mean	176.465	176.465	176.465	176.465	176.465
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

Note: This table reports robustness results testing for bias using an examiner leniency measure that is estimated separately by group. See the notes to Table 3 for details on the sample and empirical specification. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A9: Representativeness Statistics

	$\frac{\mathbb{E}(x \text{Immigrant})}{\mathbb{E}(x \text{Native-Born})}$	$\frac{\mathbb{E}(x \text{Old})}{\mathbb{E}(x \text{Young})}$	$\frac{\mathbb{E}(x \text{Female})}{\mathbb{E}(x \text{Male})}$
	(1)	(2)	(3)
Immigrant	–	1.100	0.607
Age	1.007	–	0.986
Female	0.678	0.966	–
One Plus Years at Residence	0.801	1.094	1.057
Married	1.404	1.724	0.853
Number of Dependents	0.741	1.219	2.031
Credit Score	1.051	0.991	0.967
Has Bank Account	1.028	0.990	0.986
Has Other Loan Payments	0.685	1.269	1.380
Loan Amount Requested (£)	1.299	1.164	0.860
Loan for Emergency	1.030	1.007	0.887
Loan for Large One-Time Expense	2.120	0.912	0.653
Loan for Overdraft Avoidance	0.714	1.194	0.978
Loan for Shopping or Holiday	0.884	0.905	1.312
Observations	45687	45687	45687

Note: This table reports the mean of the listed variable conditional on target group status, divided by the mean of the listed variable, conditional on reference group status. The means are estimated for the sample as described in the notes to Table 1.

Appendix Table A10: IV Estimates Using Subgroup-Specific Discounted Long-Run Profits

	Additional Target Group Discount Rate					
	5 Percent (1)	10 Percent (2)	20 Percent (3)	30 Percent (4)	40 Percent (5)	50 Percent (6)
Loan Take-Up	0.316 (116.862)	9.067 (114.598)	25.103 (111.401)	40.752 (109.556)	55.102 (108.772)	67.274 (108.668)
Take-Up x Immigrant Applicant	438.152*** (159.390)	420.529*** (155.706)	388.006*** (149.904)	355.720*** (145.771)	329.196** (143.153)	306.801** (141.482)
Take-Up x Older Applicant	255.678** (128.172)	246.511** (125.126)	231.508* (120.574)	218.047* (117.188)	206.795* (114.817)	197.237* (113.145)
Take-Up x Female Applicant	64.050 (108.150)	55.355 (106.575)	38.789 (104.339)	22.466 (103.329)	7.278 (103.333)	-5.606 (103.875)
Dep. Variable Mean	168.226	166.019	162.078	158.714	155.936	153.542
Observations	45687	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254	254

Note: This table reports IV estimates of bias in consumer lending using discounted long-run profits as the outcome. Reference group long-run profits are discounted using a discount rate of 10 percent. Target group long-run profits are additionally discounted by the amount listed in the column header. See the notes to Table 3 for additional details. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A11: OLS Results of Loan Take-Up on Long-Run Profits

	OLS Estimates				
	(1)	(2)	(3)	(4)	(5)
Loan Take-Up	209.483*** (8.278)	176.515*** (8.276)	154.520*** (8.805)	168.964*** (8.318)	59.891*** (8.555)
Take-Up x Immigrant Applicant		101.680*** (11.930)			126.766*** (12.027)
Take-Up x Older Applicant			88.751*** (7.464)		89.506*** (7.389)
Take-Up x Female Applicant				100.296*** (8.393)	123.407*** (8.277)
Dep. Variable Mean	176.533	176.533	176.533	176.533	176.533
Observations	45687	45687	45687	45687	45687
Clusters	254	254	254	254	254

Note: This table reports OLS estimates of bias in consumer lending decisions based on long-run profits. The regressions are estimated on the sample as described in the notes to Table 1. All specifications control for store-by-month-by-nationality fixed effects and the baseline characteristics listed in Panel A of Table 1. Standard errors clustered at the examiner level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A12: Subsample IV Estimates by Examiner Gender

	Female Examiner			Male Examiner				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Loan Take-Up	247.900*** (95.408)	64.419 (386.473)	386.471*** (118.892)	-196.732 (397.715)	22.432 (238.815)	-189.656 (233.538)	-309.748 (459.876)	-200.523 (438.997)
Take-Up x Immigrant Applicant	897.649** (437.807)			886.037** (420.153)	-1,029.177 (1,653.314)			-1,075.787 (1,743.517)
Take-Up x Older Applicant		672.721 (724.215)		693.134 (745.221)		152.503 (118.315)		154.801 (128.643)
Take-Up x Female Applicant			64.869 (140.814)	143.904 (159.527)			223.960 (392.076)	204.134 (412.308)
Dep. Variable Mean	176.268	176.268	176.268	176.268	176.959	176.959	176.959	176.959
Observations	28209	28209	28209	28209	16503	16503	16503	16503
Clusters	157	157	157	157	95	95	95	95

Note: This table reports IV estimates of bias in consumer lending separately by examiner gender. See the notes to Table 3 for details on the sample and empirical specification. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A13: Subsample IV Estimates by Examiner Experience

	Inexperienced Examiner			Experienced Examiner		
	(1)	(2)	(3)	(4)	(5)	(6)
Loan Take-Up	106.365* (64.235)	40.557 (108.859)	212.135** (99.514)	-95.624 (141.114)	361.300* (191.958)	663.668*** (249.027)
Take-Up x Immigrant Applicant	553.248*** (180.698)			531.471*** (189.591)	1,428.279* (769.228)	1,631.579* (872.656)
Take-Up x Older Applicant		379.179** (179.984)		383.466** (183.584)		240.363 (362.937)
Take-Up x Female Applicant			-5.339 (128.921)	59.614 (135.970)		31.410 (310.546)
Dep. Variable Mean	145.135	145.135	145.135	145.135	232.227	232.227
Observations	28969	28969	28969	28969	16214	16214
Clusters	165	165	165	165	89	89

Note: This table reports IV estimates of bias in consumer lending separately by examiner experience. Experienced examiners are those employed by the Lender at the start of the sample period. See the notes to Table 3 for details on the sample and empirical specification. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A14: Predicted Long-Run Profits and Applicant Characteristics

	OLS Estimates			
	(1)	(2)	(3)	(4)
Immigrant Applicant	36.443*** (5.643)			42.558*** (5.414)
Older Applicant		30.996*** (2.459)		35.289*** (2.539)
Female Applicant			-4.039 (3.115)	14.864*** (2.203)
Predicted Default (x100)	-21.022*** (0.515)	-21.439*** (0.560)	-22.150*** (0.561)	-19.930*** (0.542)
Predicted Default Squared (x100)	0.136*** (0.515)	0.143*** (0.006)	0.147*** (0.006)	0.130*** (0.005)
Dep. Variable Mean	264.689	264.689	264.689	264.689
Observations	30192	30192	30192	30192
Clusters	254	254	254	254

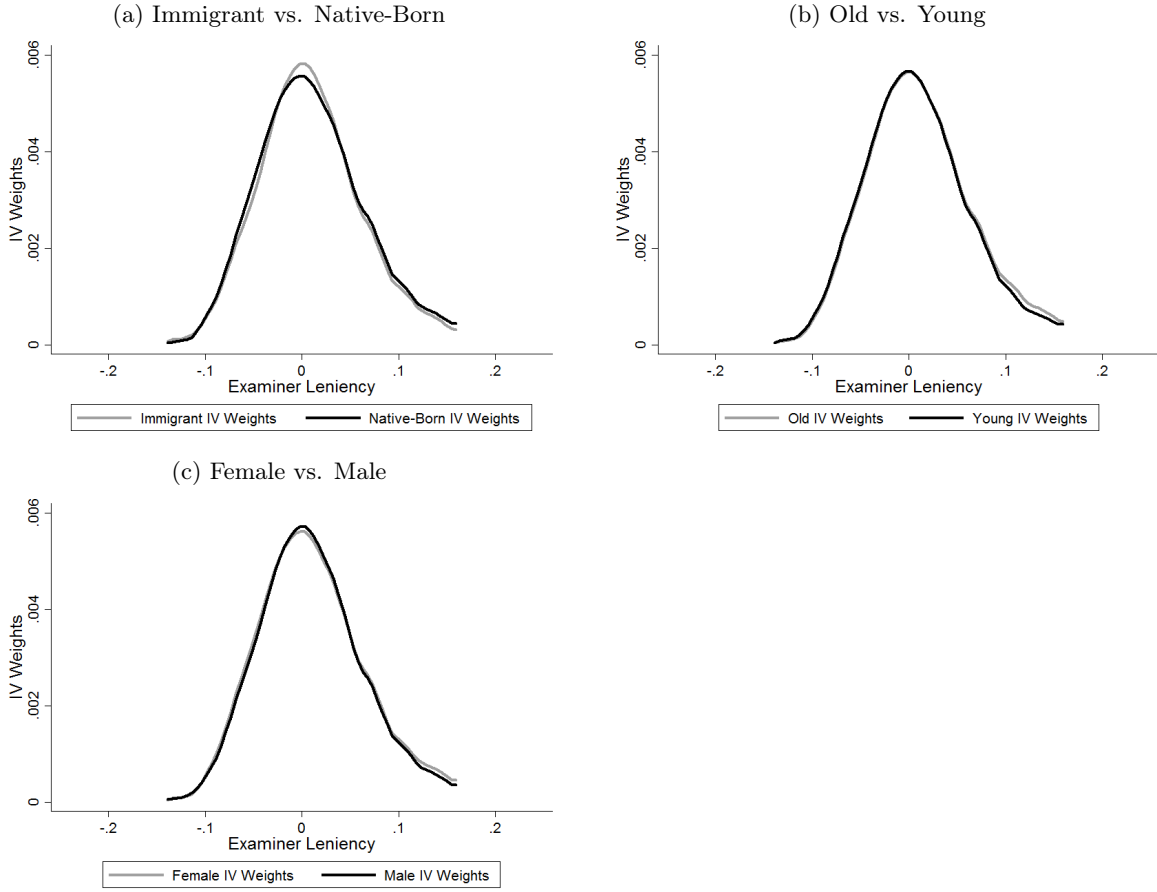
Note: This table reports the correlation between predicted long-run profits and applicant characteristics. The regressions are estimated using only applicants who took out a loan. Predicted long-run profits and predicted short-run default are obtained using the machine learning algorithm described in Section V. Standard errors clustered at the examiner level are reported in parentheses. See the text for additional details. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Table A15: Correlates of the Machine Learning Prediction for Long-Run Profits

	OLS Estimates			
	(1)	(2)	(3)	(4)
Credit Score (/1000)	431.130*** (14.305)	457.298*** (12.266)	364.247*** (9.770)	184.690*** (5.314)
Immigrant		94.743*** (5.448)	45.921*** (4.936)	-14.910*** (4.985)
Age		7.132*** (0.146)	7.363*** (0.120)	6.794*** (0.091)
Female		120.153*** (3.176)	112.701*** (2.726)	116.488*** (2.398)
Married		18.395*** (2.568)	7.000*** (1.898)	10.755*** (1.013)
Disposable Income (£/1000)			-30.307*** (9.529)	12.666*** (2.781)
Months in UK (/1000)			-129.030*** (8.212)	-77.111*** (4.207)
Number of Dependents			13.325*** (1.048)	20.925*** (0.706)
Loan for Emergency			-9.022*** (2.599)	-7.304*** (1.310)
Customer was Referred			113.343*** (3.077)	89.248*** (2.105)
Loan Amount Requested (£/1000)			143.687*** (5.437)	127.304*** (4.389)
Total Income (£/1000)			62.323*** (6.396)	-1.847 (1.298)
Salary (£/1000)				14.471*** (1.218)
Other Loan Payments (£/1000)				234.608*** (9.413)
Debt to Income Ratio (/1000)				1,409.619*** (256.804)
Total Credit Outstanding (£/10 ⁶)				37.077 (71.121)
Number Open Lines of Credit				-0.334*** (0.124)
Months at Current Residence (/1000)				87.465*** (6.018)
Credit Arrears from Other Lenders (£/10 ⁶)				64.362 (51.482)
Dep. Variable Mean	228.790	228.790	228.790	228.790
Observations	45687	45687	45687	45687
Adjusted R-Squared	0.217	0.458	0.659	0.858

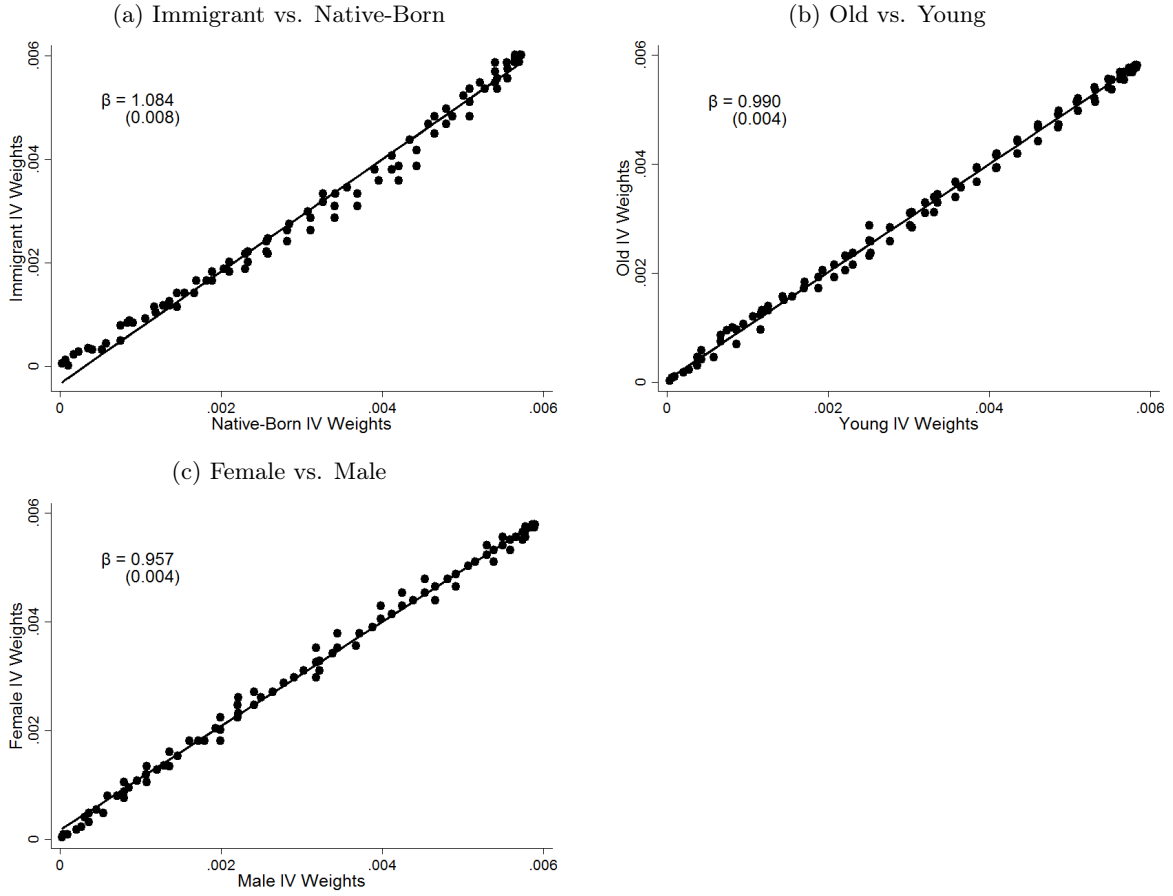
Note: This table reports selected coefficients from an OLS regression of predicted total profits on loan and applicant characteristics. The regressions are estimated on the sample as described in the notes to Table 1. All regressions include store and time fixed effects. Predicted total profits are obtained using the machine learning algorithm described in Section V. Standard errors clustered at the examiner level are reported in parentheses. See the text for additional details. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Appendix Figure A1: Distribution of IV weights



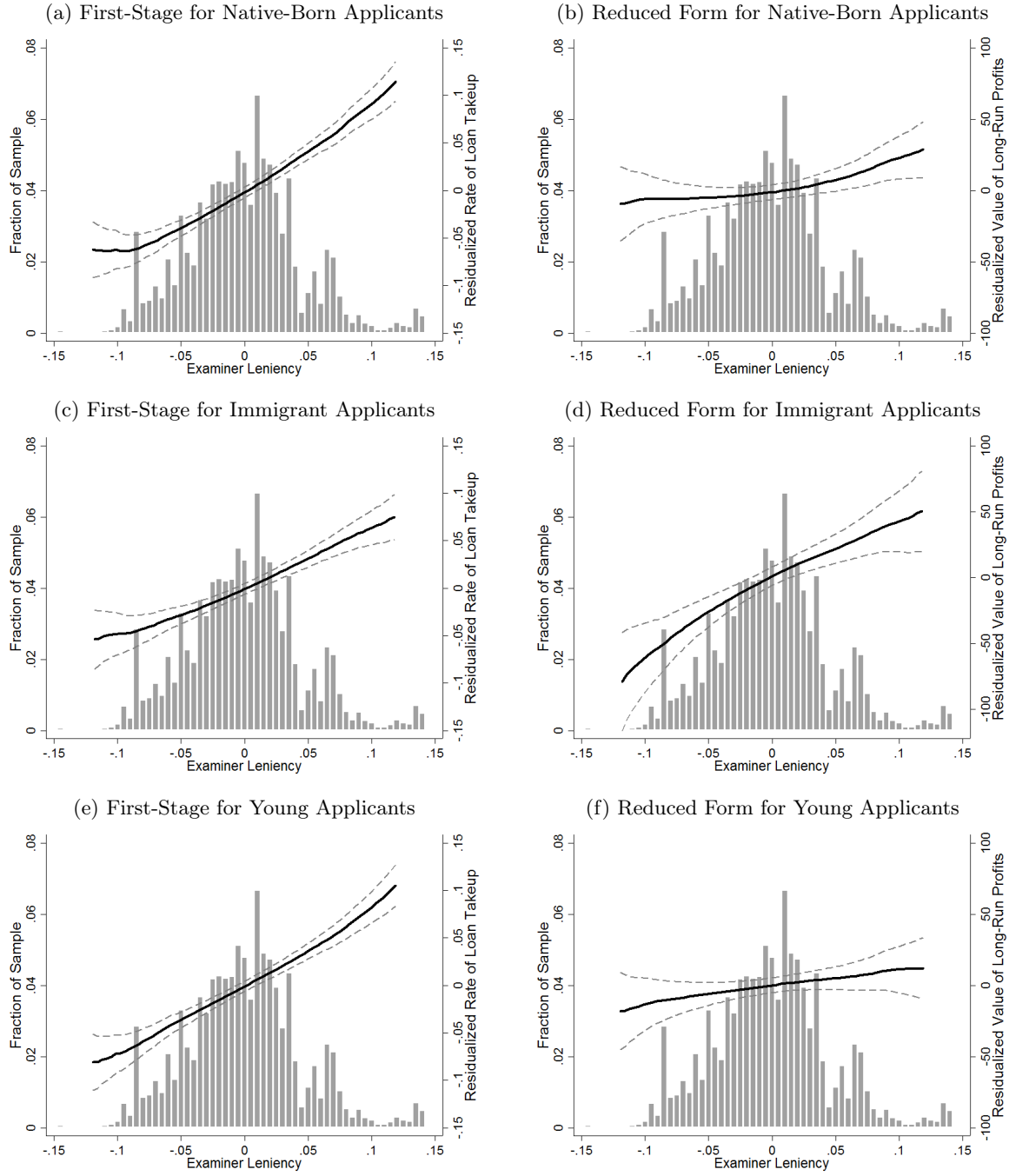
Note: These figures report the distribution of IV weights separately by group. To compute the IV weights assigned to an examiner cell, we compute continuous IV weights following the procedure described in Cornelissen et al. (2016). We then discretize these continuous weights by assigning each examiner the weight associated with his or her average leniency. Finally, we divide the discretized examiner weights by the sum total to ensure the discretized weights sum to one. See the text for additional details.

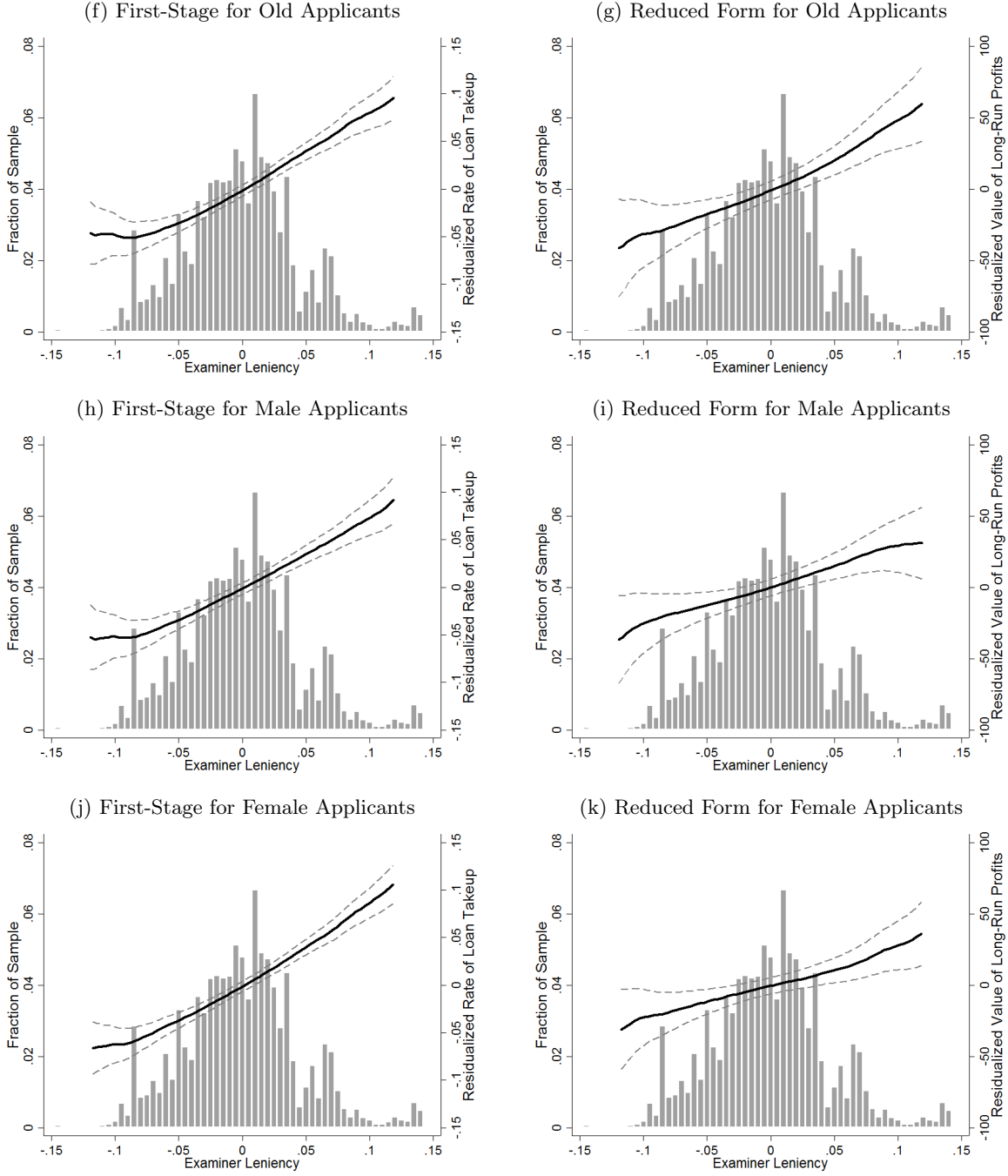
Appendix Figure A2: Correlation Between Subgroup-Specific IV Weights



Note: These figures report the correlation between group-specific IV weights. To compute the IV weights assigned to an examiner cell, we compute continuous IV weights following the procedure described in Cornelissen et al. (2016). We then discretize these continuous weights by assigning each examiner the weight associated with his or her average leniency. Finally, we divide the discretized examiner weights by the sum total to ensure the discretized weights sum to one. The best fit line is estimated using OLS. See the text for additional details.

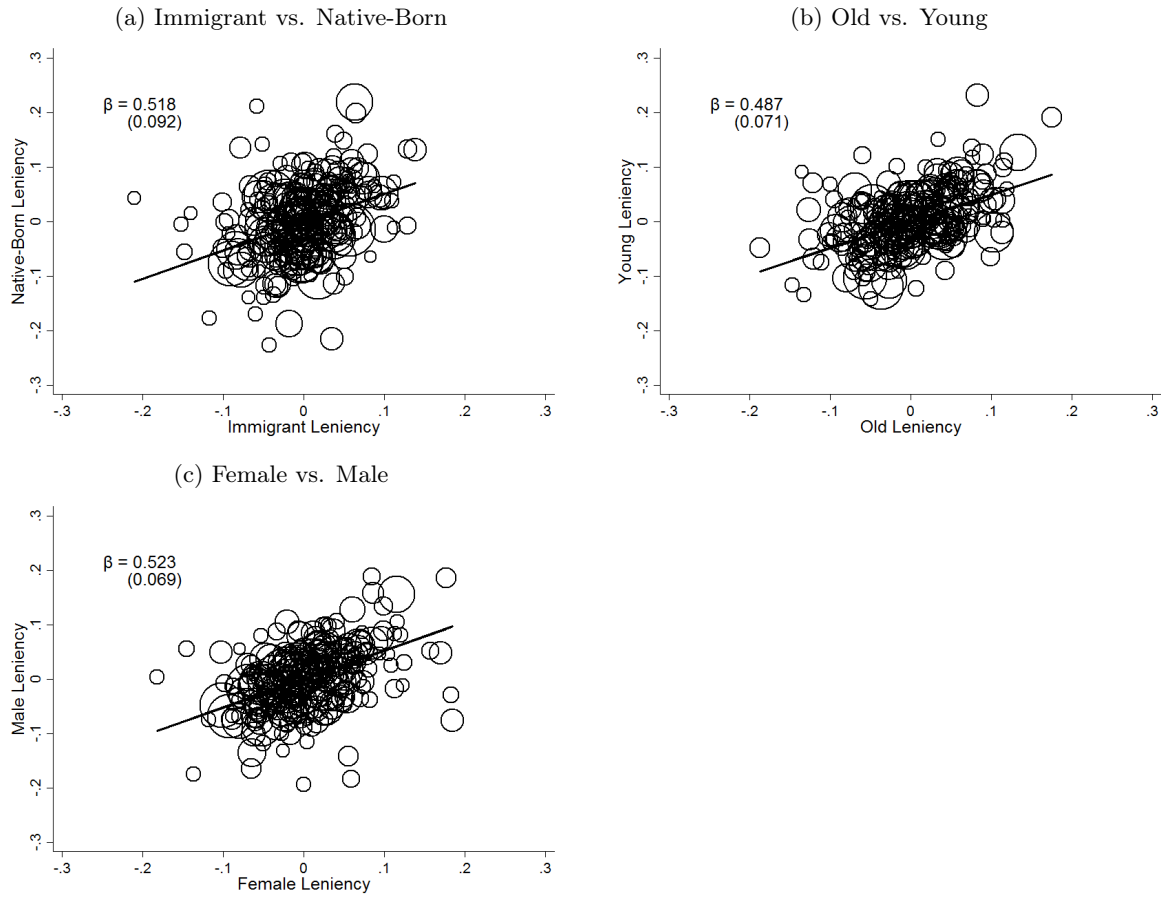
Appendix Figure A3: First Stage and Reduced Form Results by Group





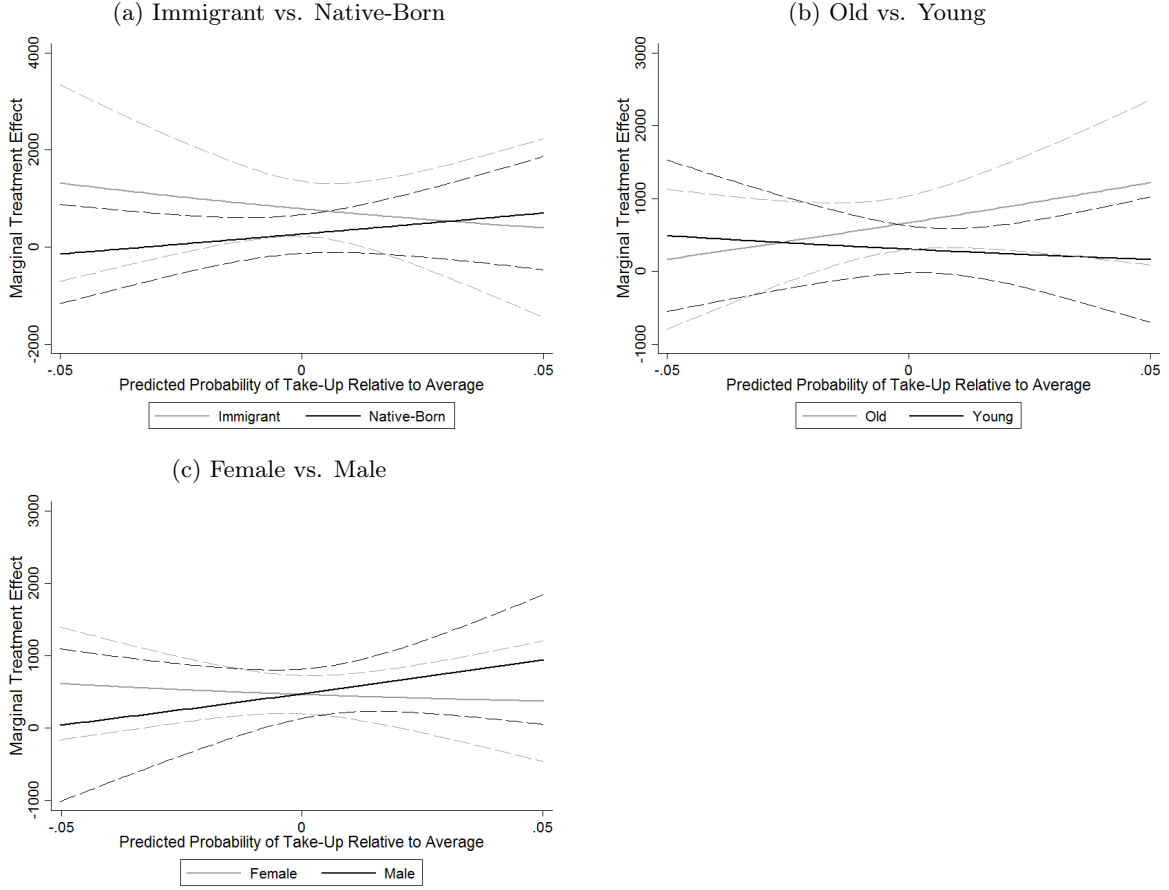
Note: These figures report the first stage and reduced form relationships between applicant outcomes and examiner leniency by group. The regressions are estimated on the sample described in the notes to Table 1. Examiner leniency is estimated using data from other applicants assigned to a loan examiner following the procedure described in Section III. In the first stage regressions, the solid line represents a local linear regression of loan take-up on examiner leniency. In the reduced form regressions, the solid line represents a local linear regression of long-run profits on examiner leniency. Loan take-up and long-run profits are residualized using store-by-month-by-nationality fixed effects. Standard errors are clustered at the examiner level.

Appendix Figure A4: Correlation Between Subgroup-Specific Examiner Leniency Measures



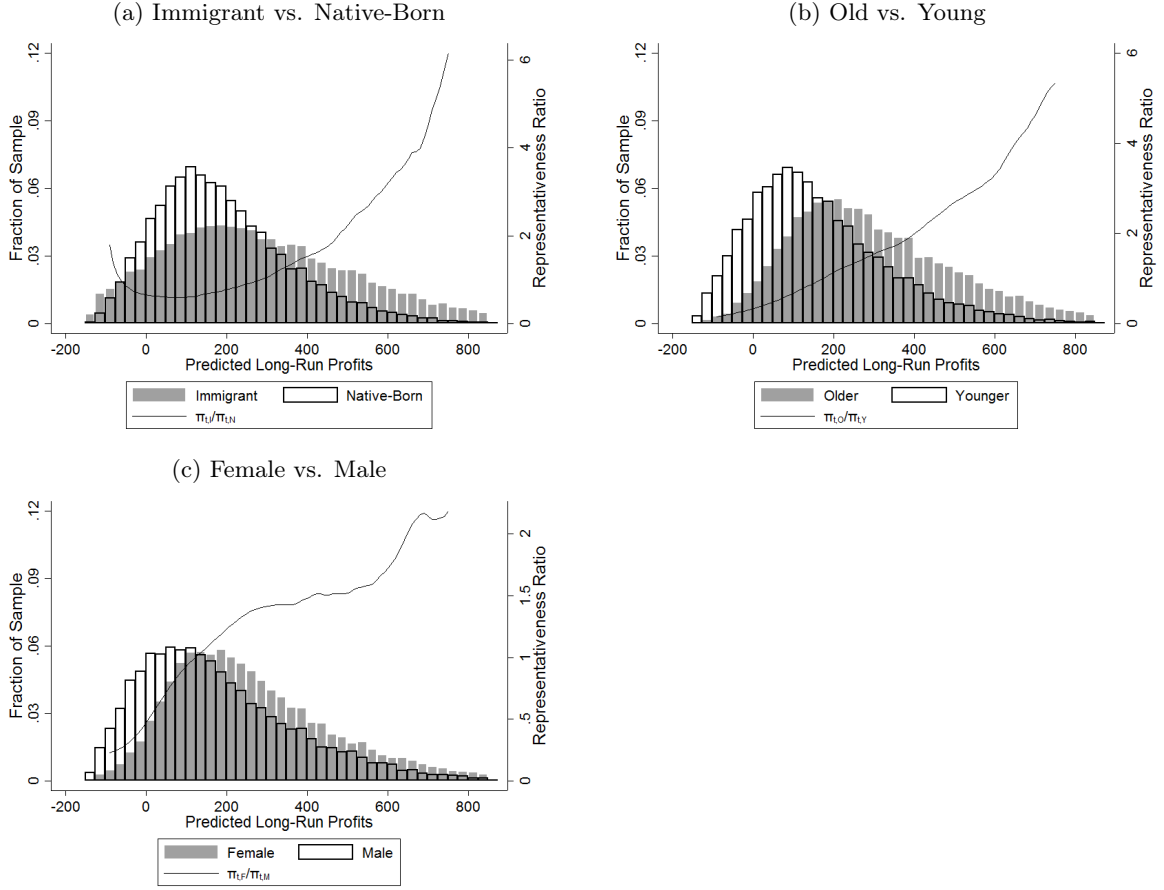
Note: These figures report the correlation between group-specific examiner leniency measures. Examiner leniency by group is estimated using data from other applicants from the same group assigned to a loan examiner following the procedure described in Section III. The best line fit is estimated using OLS.

Appendix Figure A5: Marginal Treatment Effects



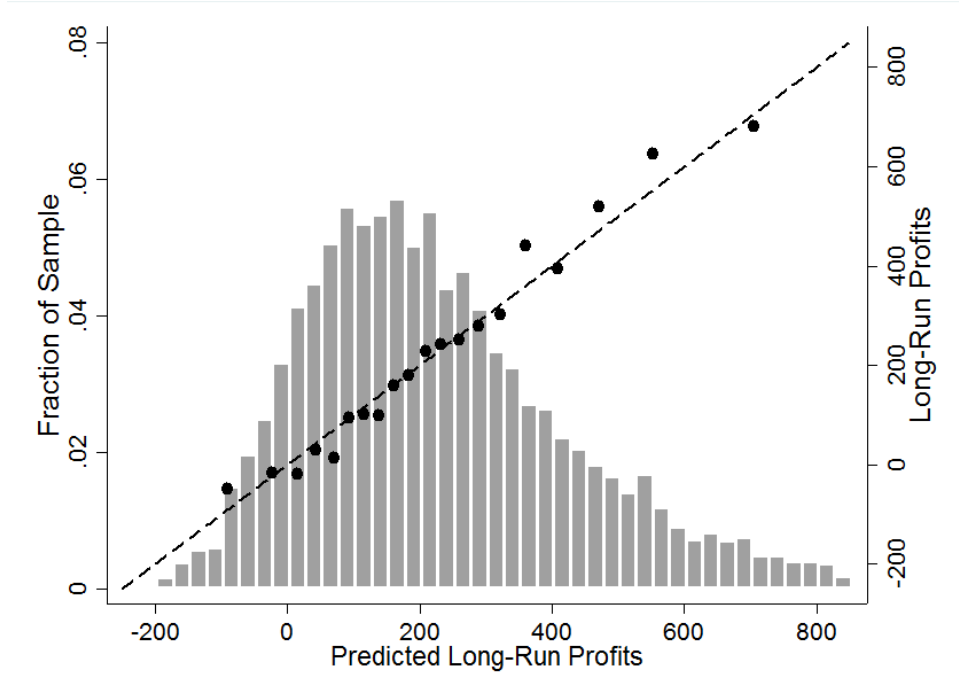
Note: These figures report estimated marginal treatment effects of loan take-up on total profits separately by group. To compute the marginal treatment effects, we first estimate the predicted probability of take-up using only variation in examiner leniency. We then estimate the relationship between predicted probability of take-up and total profits using a local quadratic estimator. We calculate the numerical derivative to estimate the marginal treatment effect at each point in the distribution. The solid lines represent the estimated marginal treatment effects separately for each subgroup, while the dashed lines represent 90 percent confidence intervals, with standard errors that are computed using 500 bootstrap replications clustered at the examiner level. See the text for additional details.

Appendix Figure A6: Predicted Long-Run Profit Distributions by Group



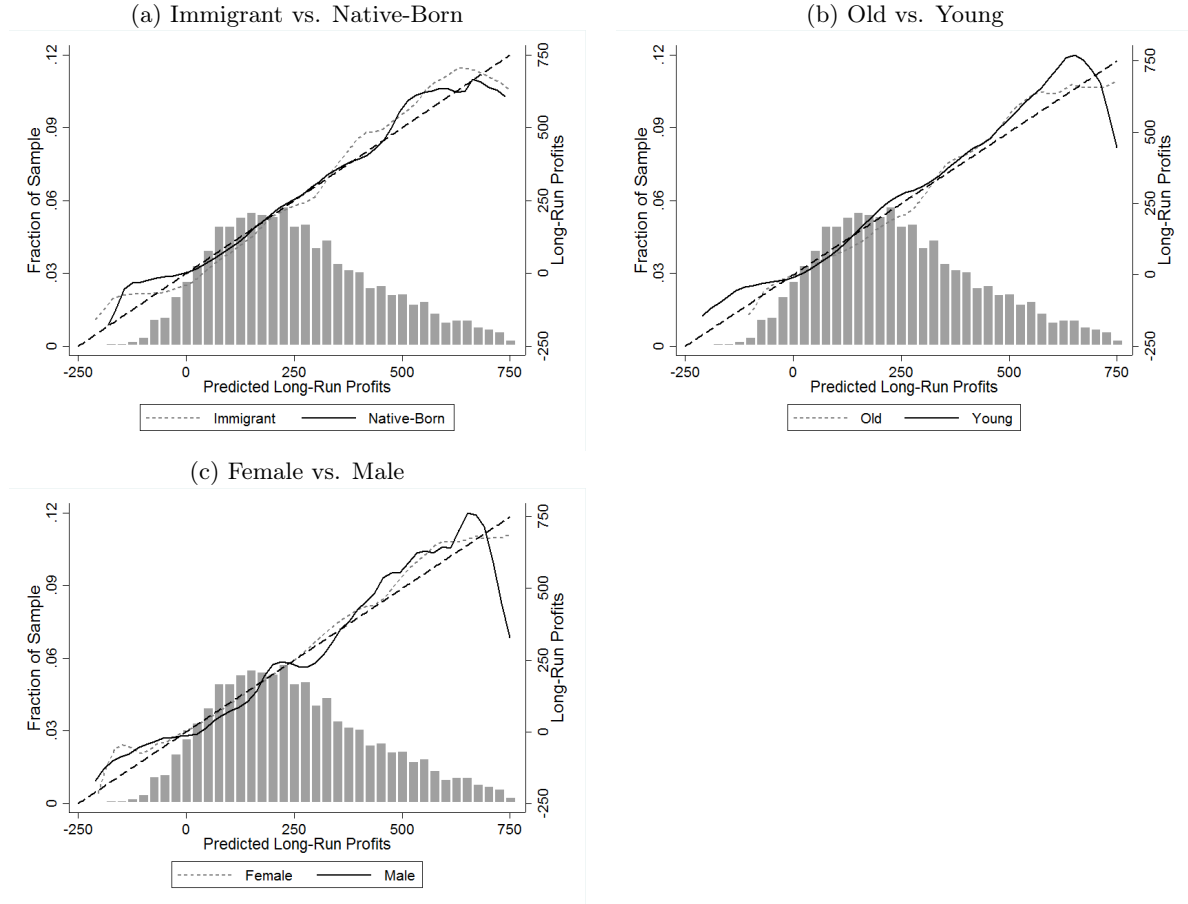
Note: These figures report the distribution of predicted long-run profits separately by group. Predicted long-run profits are calculated using the machine learning algorithm described in Section V. The solid lines report the representativeness ratio for target group versus reference group applicants, or the predicted long-run profits for the target group divided by the predicted long-run profits for the reference group. See the text for additional details.

Appendix Figure A7: Relationship Between Observed and Predicted Long-Run Profits



Note: This figure reports the relationship between observed and predicted long-run profits in the test sample. Predicted long-run profits are calculated using the machine learning algorithm described in Section V. The straight dashed line is the 45 degree line. See the text for additional details.

Appendix Figure A8: Bias in Machine Learning Predictions



Note: These figures report the relationship between observed and predicted long-run profits by group. Predicted long-run profits are calculated using the machine learning algorithm described in Section V. The straight dashed line is the 45 degree line. See the text for additional details.

Appendix B: Data Appendix

Examiner Leniency: We calculate examiner leniency as the leave-out mean residualized take-up decisions of loan examiners within a physical branch location. We use the residual take-up decision after removing store-by-month-by-nationality fixed effects. In our main results, we define loan take-up based on whether the first-time applicant took up their loan with the Lender.

Loan Approved: An indicator for the loan examiner approving the loan application (versus rejecting the loan application).

Loan Take-Up: An indicator for the loan applicant taking out a loan with the Lender (versus not taking out a loan). Loan Take-Up is set to zero if the loan application is rejected or if the application is approved but the applicant decides not to take out a loan.

Loan Top-Up: An indicator for the loan applicant closing the initial loan and taking out a new loan to cover the remaining balance on the initial loan. Loan Top-Up is set to zero for applicants who never take out an initial loan.

Long-Run Profits: We calculate profits as the sum of all payments made from the applicant to the Lender over the course of their relationship, minus the sum of all disbursements made from the Lender to the applicant in pounds. Long-Run Profits are set to zero for applicants who never take out a loan.

Short-Run Default: An indicator for the applicant defaulting on his or her first loan with the Lender. Short-Run Default is set to zero for applicants who never take out a loan.

Immigrant Applicant: An indicator for whether the applicant is an immigrant (versus native-born).

Female Applicant: An indicator for whether the applicant is female (versus male).

Applicant Age: The applicant's age in years. We drop applicants who are younger than 18 years old or older than 75 years old.

Old Applicant: An indicator for whether the applicant is at least 32 years old, the median sample age in the sample (versus less than 32 years old).

Months at Current Residence: The number of months the applicant has spent at their current residence as of the time of their application.

One Year at Residence: An indicator for the applicant having spent at least 12 months at their current residence as of the time of their application (versus less than 12 months at their current residence).

Married Applicant: An indicator for the applicant being married at the time of his or her first loan application (versus unmarried).

Number of Dependents: The applicant's number of dependents at the time of his or her first loan application. The number of dependents variable is winsorized at the 99th percentile of the distribution.

Credit Score: The applicant's credit score from a nationwide credit bureau at the time of his or her first loan application. The credit score variable is set to 0 for the approximately 2.6 percent of the sample with a missing credit score. In all regressions, we include an indicator for whether an applicant is missing the credit score variable or not.

Has Bank Account: An indicator for the applicant having a bank account at the time of his or her first loan application (versus no bank account).

Has Other Loan Payments: An indicator for the applicant having other loan payments at the time of his or her first loan application (versus no other loan payments).

Other Loan Payments: The value of an applicant's other loan payments in pounds at the time of his or her first loan application.

Number Open Lines of Credit: The applicant's number of open lines of credit at the time of his or her first loan application.

Total Credit Outstanding: The value of the applicant's total outstanding credit in pounds at the time of his or her first loan application.

Credit Arrears from Other Lenders: The value in pounds of overdue credit the applicant has to other lenders at the time of his or her first loan application.

Customer was Referred: An indicator for whether the applicant was referred to the Lender.

Loan Amount Requested: The applicant's requested loan amount in pounds. The applicant may take out less than the requested amount.

Loan for Emergency: An indicator for the self-reported reason for the loan being for an emergency.

Loan for Large One-Time Expense: An indicator for the self-reported reason for the loan being large non-recurring expense.

Loan for Overdraft Avoidance: An indicator for the self-reported reason for the loan being to avoid overdraft penalties.

Loan for Shopping or Holiday: An indicator for the self-reported reason for the loan being shopping or holiday expenses.

Loan Amount Net of Fees: The loan amount initially taken out minus all associated Lender fees in pounds. This variable is missing for applicants who did not take out a loan.

Loan APR: The annualized nominal interest rate of the loan. This variable is missing for applicants who did not take out a loan.

Loan Duration: Length of the loan in months if the applicant follows the set payment schedule. This variable is missing for applicants who did not take out a loan.

Male Examiner: An indicator for the examiner being male (versus female). We are missing examiner gender data for two examiners.

Experienced Examiner: An indicator for the examiner being employed by the Lender at the start of the sample period.