Economics, Discrimination and Inequality

Paul Goldsmith-Pinkham

Yale School of Management

April 2021

Fairness in decision-making is an age-old problem

A rich line of work within computer science examines the differential treatment by algorithms of historically disadvantaged and marginalized groups. Much of this work is concerned with fairness of algorithms, which is understood as the absence of discrimination. Many leading notions of fairness – such as predictive parity or balance – are based on some variant of the question *are members of different groups who are of equal "merit" treated equally* by the algorithm?"[1] Research in this space has ranged from translating these fairness notions to various domains to examining when and whether they are simultaneously achievable with other constraints.

Abebe and Kasy (2020)

- This could read "are members of different groups who are of equal merit treated equally by [**society** | **the market** | **judges** | **firms** | **the government**]?"

What is unique about the concern about algorithms

Two dimensions are unique about the growth in algorithms from a fairness perspective:

1. An algorithmic decision-making process makes the subtext text

What is unique about the concern about algorithms

Two dimensions are unique about the growth in algorithms from a fairness perspective:

1. An algorithmic decision-making process makes the subtext text
    - We must specify our inputs and how we manipulate said inputs to get outputs
    - The more automated a process, the more explicit this decision is
        - E.g. less “wiggle room”

What is unique about the concern about algorithms

Two dimensions are unique about the growth in algorithms from a fairness perspective:

1. An algorithmic decision-making process makes the subtext text
   - We must specify our inputs and how we manipulate said inputs to get outputs
   - The more automated a process, the more explicit this decision is
     - E.g. less “wiggle room”
2. By specifying the relevant inputs (“merit”), it starkly outlines the objective function
   - What Abebe and Kasy (2020) describe as “narrow-bracketing”
   - Treats the inputs as [what an economist would describe as] “exogeneous,” or fixed outside the model

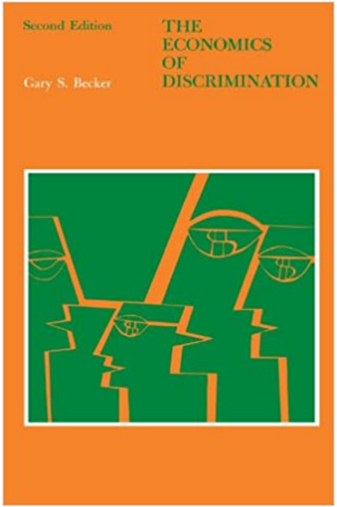
Hence: a debate about algorithmic fairness and discrimination is a debate about societal fairness/discrimination

Models of discrimination in economics - “taste”

- Economics has a rich history of stumbling through these questions
    - Much of this work initially began focused on the labor market, but the ideas have been ported to other subfields

- Gary Becker’s 1957 book, Economics of Discrimination, coined the concept of a “taste” for discrimination and how it would affect the labor market
    - E.g., a set of racist/sexist/prejudiced employers who willing lose out on profitable opportunities to satisfy their bigotry

- Depending on the share of bigoted employers, can show that this will create a wage wedge for employees



Competition does not inherently solve these issues

> ***The erroneous claim that in Becker's (1957) model, market discrimination disappears in the long run. It need not. Entrepreneurs can consume their income in any way they see fit.*** *If a bigoted employer prefers whites, the employer can indulge that taste as long as income is received from entrepreneurial activity, just as a person who favors an exotic ice cream can indulge that preference by being willing to pay the price. Only if the supply of entrepreneurship is perfectly elastic in the long run at a zero price, so entrepreneurs have no income to spend to indulge their tastes, or if there are enough nonprejudiced employers to hire all [black workers], will discrimination disappear from Becker's model.*

Heckman (1998)

Fairness is presupposed in this model

- In the Becker (1957) book, fairness or equality is presupposed – all workers are identically productive, and paid their marginal productivity
- If the workers were not, then the model would rationalize paying one group a lower wage
- The Becker model assumes that workers know who discriminates, and who does not
  - Black (1995) shows when this is relaxed, and workers have to search for work, wage pay gaps emerge even with lots of non-discriminatory firms
  - Effectively, the victims of discrimination have a worse outside option, and that makes it harder for them to negotiate with their (non-prejudiced) employers

Models of discrimination in economics - “statistical”

- Aigner and Cain (1977) discuss statistical discrimination (a uniquely economics term)
  - Essentially, a verison of learning models with particular group attributes

- Consider an employer that wants to know q, but sees

$$y = q + u,\ \ u \sim_{ind.} \mathcal{N}(0, \sigma^2)$$

- Infer $\hat{q} = E(q|y) = (1-\gamma)\alpha + \gamma y$
  - $\gamma = var(q)/var(q) + var(u)$, $\alpha = \bar{y}$
  - With different groups, can have γ and α change
  - Both levels across groups, and “noisiness”

STATISTICAL THEORIES OF DISCRIMINATION IN LABOR MARKETS

DENNIS J. AIGNER and GLEN G. CAIN

ECONOMIC discrimination has been difficult to explain by means of standard neoclassical economic models that assume pervasive competition. Why, after all, should two groups of workers who have the same productivity receive different remuneration? The challenge to explain this phenomenon is posed most sharply by the marked differentials in wages and earnings between blacks and whites and between men and women—differentials that remain substantial despite diligent efforts to control for supply-side productivity traits.

This paper examines that issue from a perspective suggested by Kenneth Arrow, John J. McCall, Edmund S. Phelps, Melvin W. Reder, and A. Michael Spence, all of whom focused on certain implications of employer uncertainty about the productivity of racial (or sex) groups of workers, particularly in the context of hiring and placement decisions.[1] This paper offers several models that clarify the meaning of economic “statistical discrimination,” simplify the theory, and yield plausible empirical implications. On the other hand, the paper also identifies several shortcomings of “statistical discrimination” models; shows that the often-cited Phelps model does not constitute economic discrimina-

Economic discrimination in labor markets is conventionally defined as the presence of different pay for workers of the same ability. This paper analyzes

[1]Kenneth Arrow, “Models of Job Discrimination”

Models of discrimination in economics - “statistical”

- Key issue: what is discrimination here? In economics, discrimination based on q is usually what is viewed as problematic

- In this model, with equal average productivity (e.g. $\alpha_g = \alpha_{g'}$), the group with a higher noisiness will be shrunk more towards the average, but the overall average wage will be the same across groups

- This intuition becomes more complicated as soon as firms are risk-averse: the additional “noise” from predicting high quality will lead firms away from the noisy group

Models of discrimination in economics - “statistical”

- Key issue: what is discrimination here? In economics, discrimination based on q is usually what is viewed as problematic

- In this model, with equal average productivity (e.g. $\alpha_g = \alpha_{g'}$), the group with a higher noisiness will be shrunk more towards the average, but the overall average wage will be the same across groups

- This intuition becomes more complicated as soon as firms are risk-averse: the additional “noise” from predicting high quality will lead firms away from the noisy group

Figure 2 represents a hypothetical model, of course, but it is consistent with our view of reality in two important respects. First, as a consequence of actual employer practices, economic discrimination against blacks, women, and other groups does exist, resulting in group differences in pay despite equal group abilities to perform on the job. However, if the definition of ability includes reliability in test-taking—on grounds, perhaps, that this aptitude conveys useful information to employers—then one could deny that *economic* discrimination exists. Our preference is to retain the term “economic” in describing this type of discrimination, although such discrimination stems from inadequate test instruments rather than employers’ acting upon their tastes for discriminating against black or female workers.

Models of discrimination in economics - “statistical”

- An important question in labor markets is “where did q come from?”
  - A natural answer is schooling
  - Lundberg and Startz (1983) propose endogenizing this choice of schooling

- Given a noisier signal from schooling from one group for the other, there is *less* incentive to invest for one group relative to the other
  - This is true even though the groups were equally “good” before the schooling decision

- Their definition: “Economic discrimination exists when groups with equal average initial endowments of productive ability do not receive equal average compensation in equilibrium.”

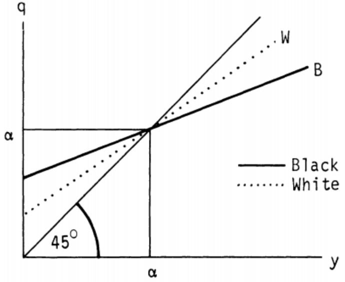


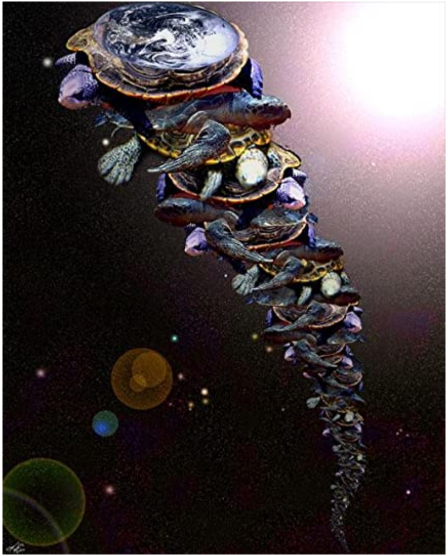
Figure 1B. Predictions of Productivity (q) by Race and Test Score (y), Assuming a Steeper Slope for Whites.

A sales pitch for economists as coauthors

- Features we take as fixed ("exogoneous") are the core features for how we interpret discrimination/fairness:
  - If we think critically about these features, there are important features outside the model that drive them
  - How do we trade these off? (it's turtles all the way down)

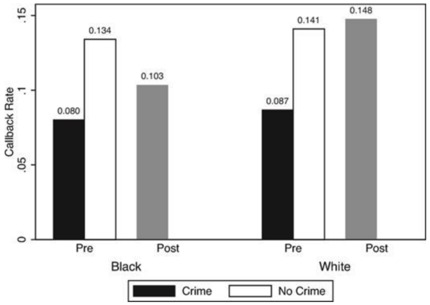
- The equilibrium matters a lot. Is there entry of entreprenuers? How is information aggregated?

- Ignoring these features can have unintended consequences.



Examples of equilibrium - ban the box

- Employers can ask job applicants whether they have been previously convicted of a felony initially on an application
  - This knowledge is a major impediment to employment
- Agan and Starr (2018) show that banning this ability to ask the question causes "pooling", in a downwards way for black applicants
- Key implications:
  - Not unambiguously good (especially for Black applicants)
  - Heterogeneously impacts applicants – those *with* a previous record would get more call backs!



And now a much longer example for thinking about this

Predictably Unequal?
The Effects of Machine Learning on Credit Markets

Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, Ansgar Walther
SNB, Yale SOM, Imperial College ($\times$2)

- Distributional impacts within groups is particularly stark within this paper

Advances in Technology and Inequality

- Machine learning has been rapidly adopted in many industries
- Central application: default prediction in credit markets
  (e.g. Khandani, Kim, and Lo, 2010; Sirignano, Sadhwani, and Giesecke, 2017)
- **This paper: What are the distributional effects of new technology?**

This Paper

Theory: Distributional implications of “better” statistical technology

Mortgage default prediction: Using US administrative data with traditional technology (Logit) and Machine Learning

Distributional consequences of new technology

- Across racial groups: fewer winners in some minority groups; increased dispersion

Equilibrium implications in a model of competitive loan pricing

- Outcomes differ on both extensive and intensive margins

A Lender's Prediction Problem

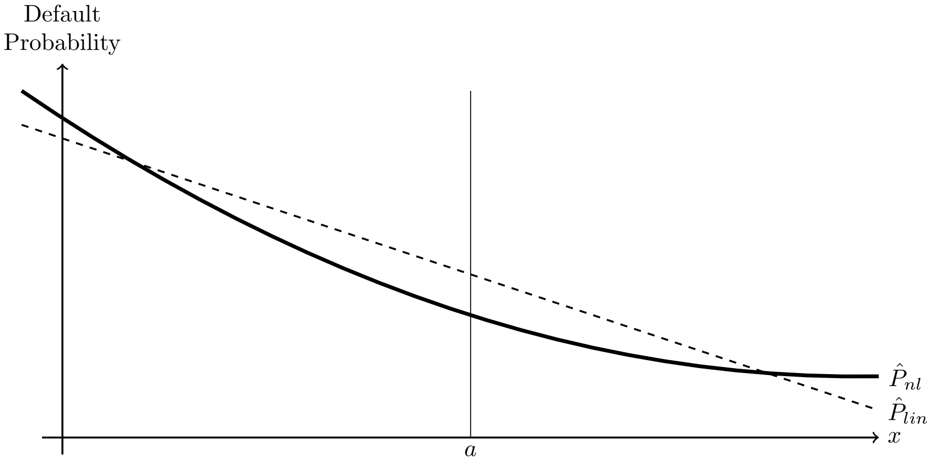
Observe borrowers with characteristics x and default outcome y

Predict $\hat{y} = \hat{P}(x)$ to minimize MSE

- Old technology: Restricted class of functions $\hat{P}$ (e.g. linear)
- New technology: Wider class of permitted functions

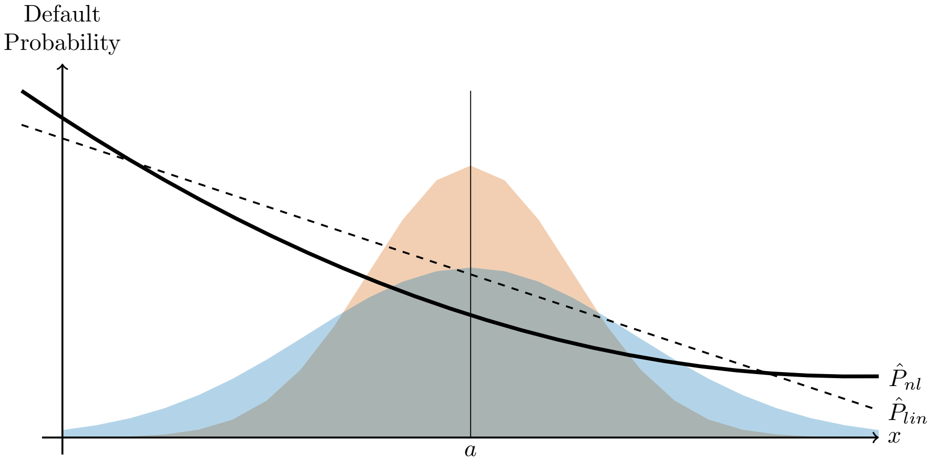
Lemma. Optimal predictions with new technology are a *mean-preserving spread* of those with old technology $\Rightarrow$ *There are winners and losers*

Winners and Losers



Convex quadratic: “extreme” x lose, others gain

Winners and Losers



Two groups: “blue” borrowers lose due to high variance

US Mortgage Data

HMDA

- Application date, applicant income, loan type, size, purpose,
- **race, ethnicity, gender**

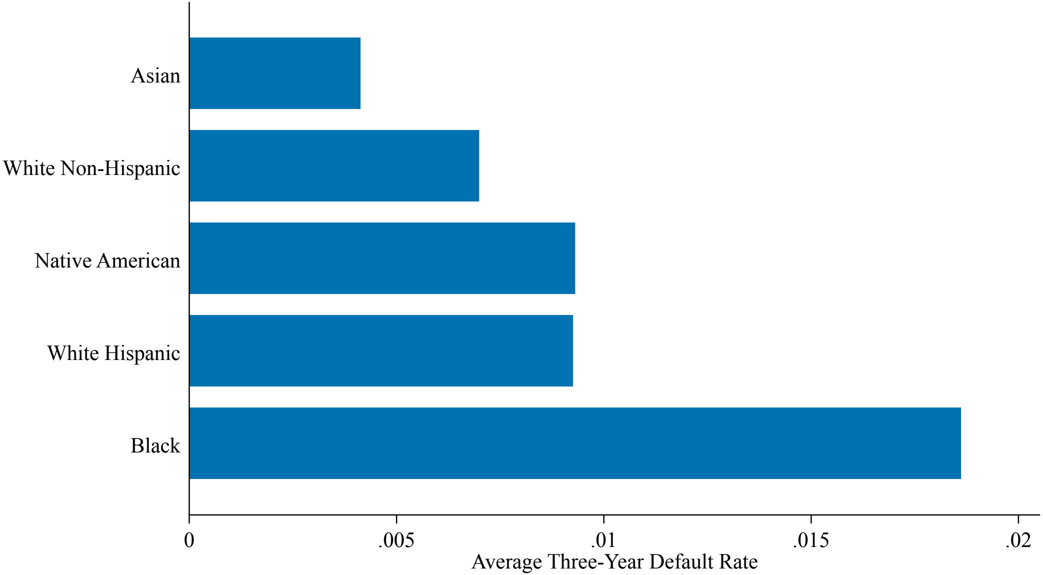
McDash (Black Knight)

- Underwriting, contract and performance: e.g. FICO, LTV, interest rate, **default status**

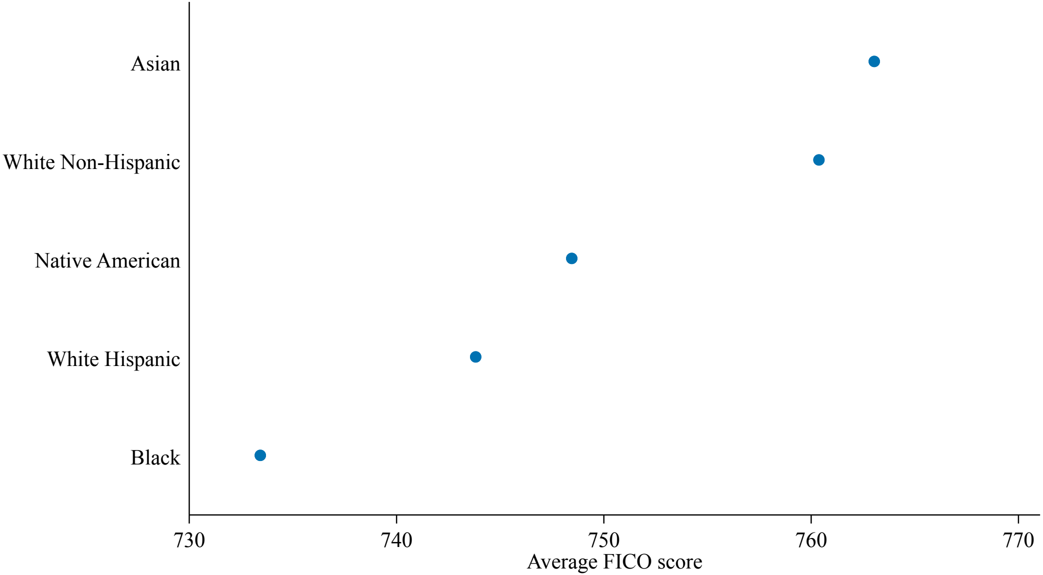
Linked Dataset

- 9.4m mortgage loans from 2009-2013
- Portfolio *and* GSE loans, $<$ \$1m
- **Default**: 90+ days delinquent within 3 years of origination

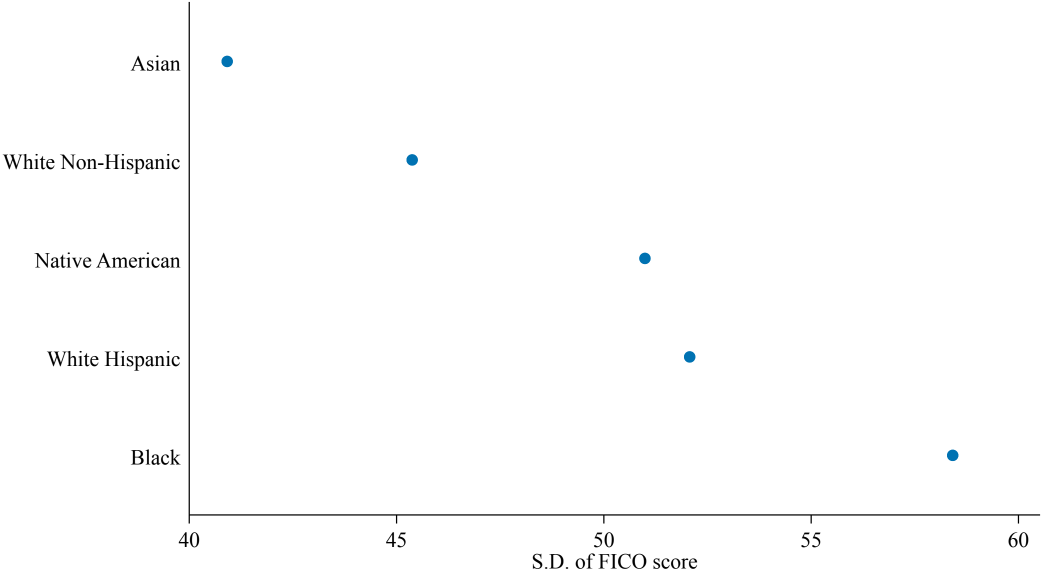
Default Rates Across Race



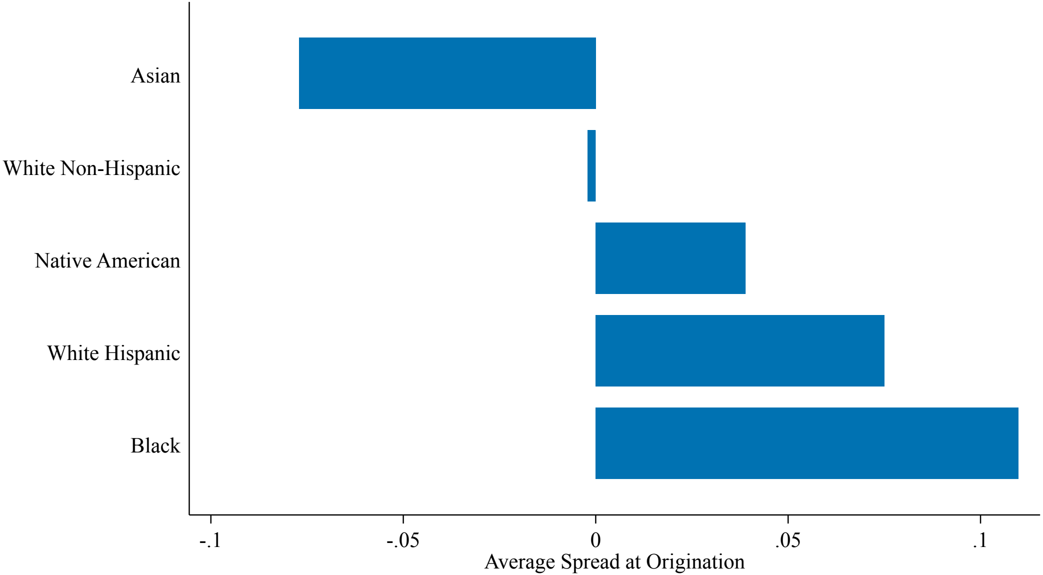
Mean FICO Across Race



S.D. of FICO Across Race



Interest Rates Across Race

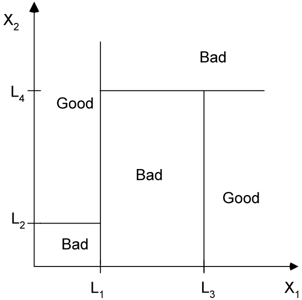


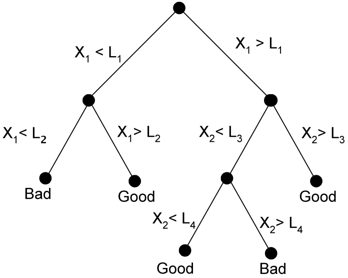
Estimating Probabilities of Default: Technologies

Traditional: Probability of Default = Logit(x) (e.g. Demyanyk and Van Hemert, 2011; Elul et al., 2010)

- Using nonlinear "bin" dummies for FICO, LTV, income

Machine Learning: Decision trees estimate step functions





(from Khandani, Kim, and Lo, 2010)

1. Random forest (w/cross-validation)
2. Calibration (isotonic regression)
   - (Similar if use "XGBoost")

Explanatory Variables

Logit	*Nonlinear Logit*
Applicant Income (linear)	Applicant Income (25k bins, from 0-500k)
LTV Ratio (linear)	LTV Ratio (5-point bins, from 20 to 100%; separate dummy for LTV=80%)
FICO (linear)	FICO (20-point bins, from 600 to 850;) separate dummy for FICO$<$600)
(with dummy variables for missing values)	
Common Covariates	
Spread at Origination "SATO" (linear)	
Origination Amount (linear and log)	
Documentation Type (dummies for full/low/no/unknown documentation)	
Occupancy Type (dummies for vacation/investment property)	
Jumbo Loan (dummy)	
Coapplicant Present (dummy)	
Loan Purpose (dummies for purchase, refinance, home improvement)	
Loan Term (dummies for 10, 15, 20, 30 year terms)	
Funding Source (dummies for portfolio, Fannie Mae, Freddie Mac, other)	
Mortgage Insurance (dummy)	
State (dummies)	
Year of Origination (dummies)	

Model Performance

▸ More Detail

Estimate on *training set* (70%), evaluate on *test set* (30%).

Out-of-sample performance:

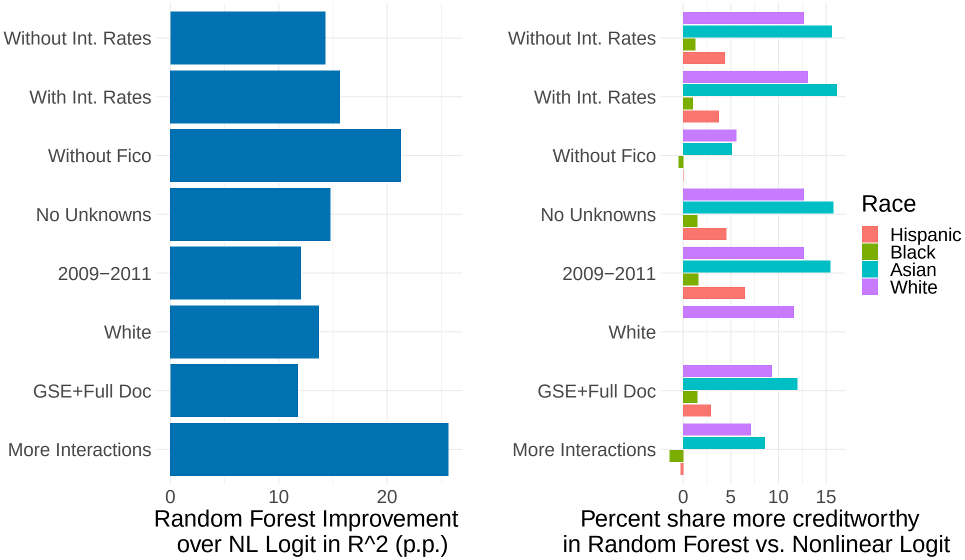
- $R^2 \uparrow$ by 14.30%
- *Precision Score* $\uparrow$ by 5.1%

  How many predicted defaults are true defaults?

- Bootstrap analysis confirms significant differences

$\rightarrow$ Random Forest method better predictor of $\Pr(\textit{default}|X)$

Unequal Effects of New Technology: Alternative Approaches

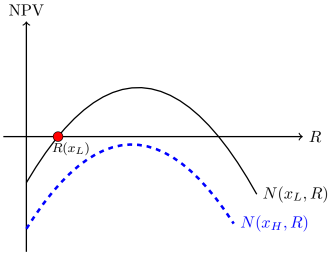


Interest Rates in Competitive Equilibrium

Simple 2-period model:

$$NPV(x,R) = \frac{1}{1+\rho}\left[(1-P(x,R))(1+R)L + P(x,R)\tilde{L}\right] - L$$

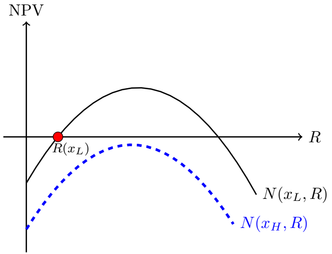
- Equilibrium $R^\star(x)$ solves $NPV = 0$
- Reject x-borrowers if $NPV(x,R) < 0$ for all feasible R



Interest Rates in Competitive Equilibrium

Simple 2-period model:

$$NPV(x,R) = \frac{1}{1+\rho}\left[(1-P(x,R))(1+R)L + P(x,R)\tilde{L}\right] - L$$

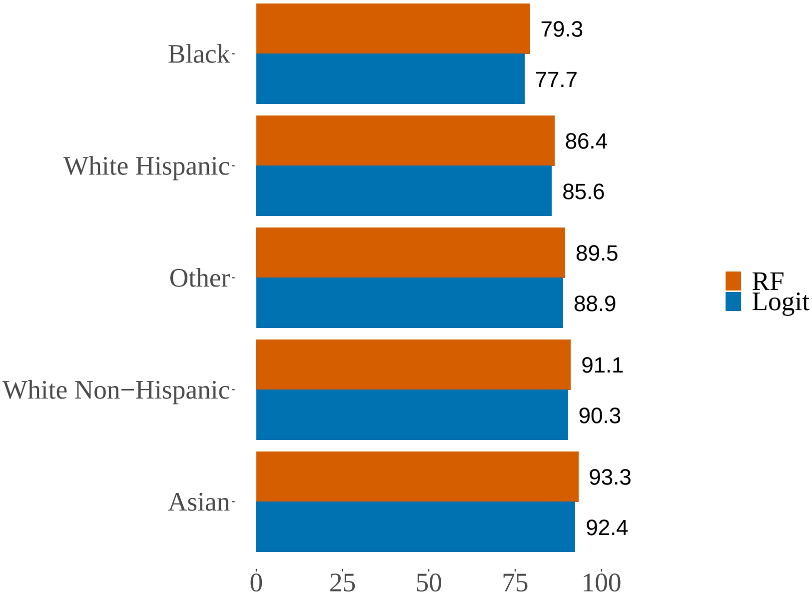


- Equilibrium $R^\star(x)$ solves $NPV = 0$
- Reject x-borrowers if $NPV(x,R) < 0$ for all feasible R

- Calibration:
  - recovery: $\tilde{L} = min((1+R)L, 0.75V) - 0.1L$ (second part: carrying costs, liquidation exp.)
  - WACC: ρ = quarterly average interest rate $-$30bps
  - 3-year PD to lifetime via “standard default assumption” (MBS mkt convention)

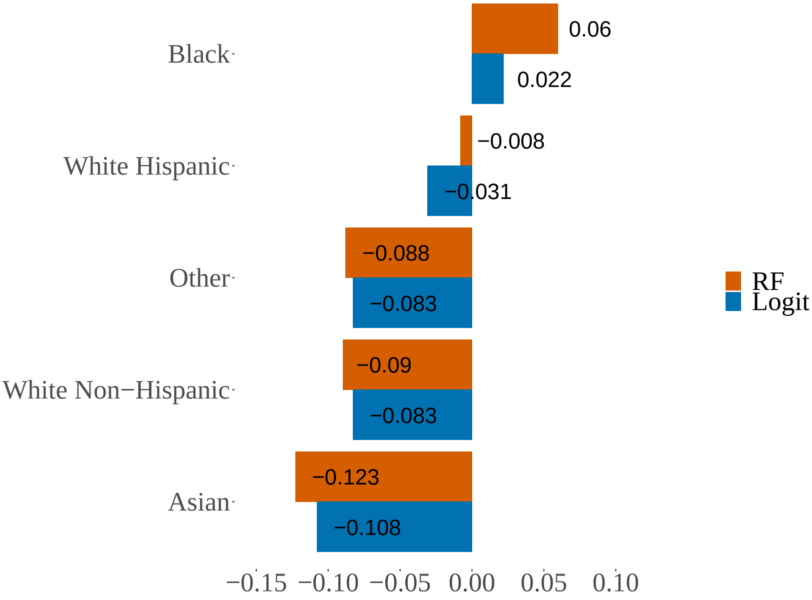
Model Outcomes

- Acceptance rates



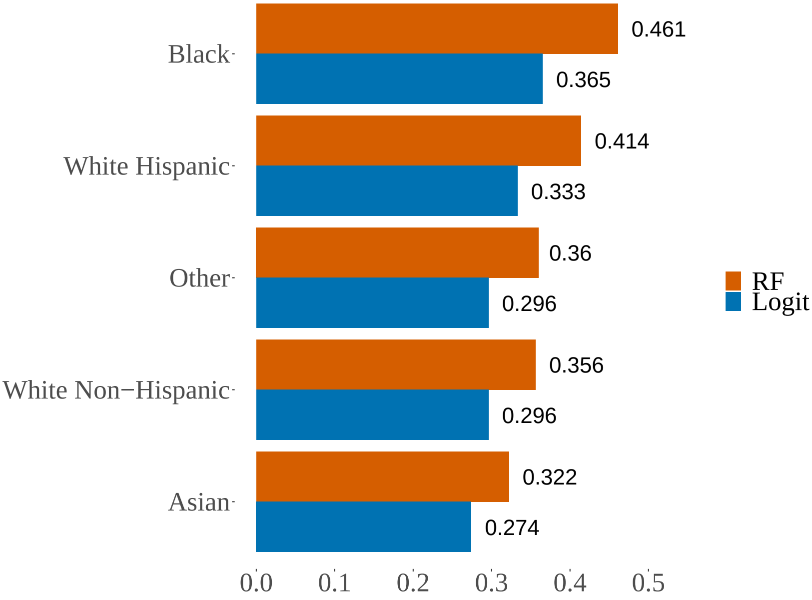
Model Outcomes

- Acceptance rates
- Average SATO ($= R - \bar{R}_t$)



Model Outcomes

- Acceptance rates
- Average SATO $(= R - \bar{R}_t)$
- S.D. of SATO $\rightarrow$ new technology increases dispersion *across* and *within* groups



Conclusion

- Improvements in statistical technology creates
  - Greater predictive power and gains for producers
  - Increased disparity in outcomes for consumers

- Based on US mortgage data, black + hispanic borrowers bear larger changes
  - First-moment effects: More likely to be perceived as high risk
  - Second-moment effects: Greater increase in dispersion of outcomes
  - Improvement comes from more than just “putting race in”

- Equilibrium effects
  - Positive extensive-margin effect of new technology
  - Unequal effects persist at intensive margin

Food for thought

1. Why do we care so much about documenting the “type” of discrimination?
2. An economics model – positive or normative? What do we treat as exogeneous? as endogeneous?
3. The challenge of causal inference when thinking about discrimination across groups
   - Race / ethnicity / gender is not manipulable (Holland (1986))
   - e.g. the bundle of sticks asepct of race (Sen and Wasow)
4. What is the equilibrium? What is sustainable? In governmental settings, policy has huge slackness. Is this true for businesses?
   - If we propose changes or diagnostics, how do we identify the benefits / costs for businesses and soceity?