Current Approaches to Algorithmic Discrimination

Talia Gillis

Columbia University

November 9, 2021

Major: Computer science

Occupation: Financial analyst

Annual income: $50,000

LOAN OFFERS

Loan interest rate: 16.34% APR

Origination fee: $1,231

Total Cost: $42,288

Major: Computer science

Occupation: Financial analyst

Annual income: $50,000

LOAN OFFERS

Loan interest rate: 21.29% APR

Origination fee: $1,960

Total Cost: $45,785

Marisa Robertson

DHH ✔
@dhh

The @AppleCard is such a ⬛⬛⬛ sexist program. My wife and I filed joint tax returns, live in a community-property state, and have been married for a long time. Yet Apple's black box algorithm thinks I deserve 20x the credit limit she does. No appeals work.

3:34 PM · Nov 7, 2019 · Twitter for iPhone

9.6K Retweets   **29.2K** Likes

Department of Housing and Urban Development

"…where a plaintiff identifies an offending policy or practice that relies on an algorithmic model, a defending party may defeat the claim by: (1) identifying the inputs used in the model and showing that these inputs are not substitutes for a protected characteristic…"

- HUD Proposed Rule on Disparate Impact (August 2019)

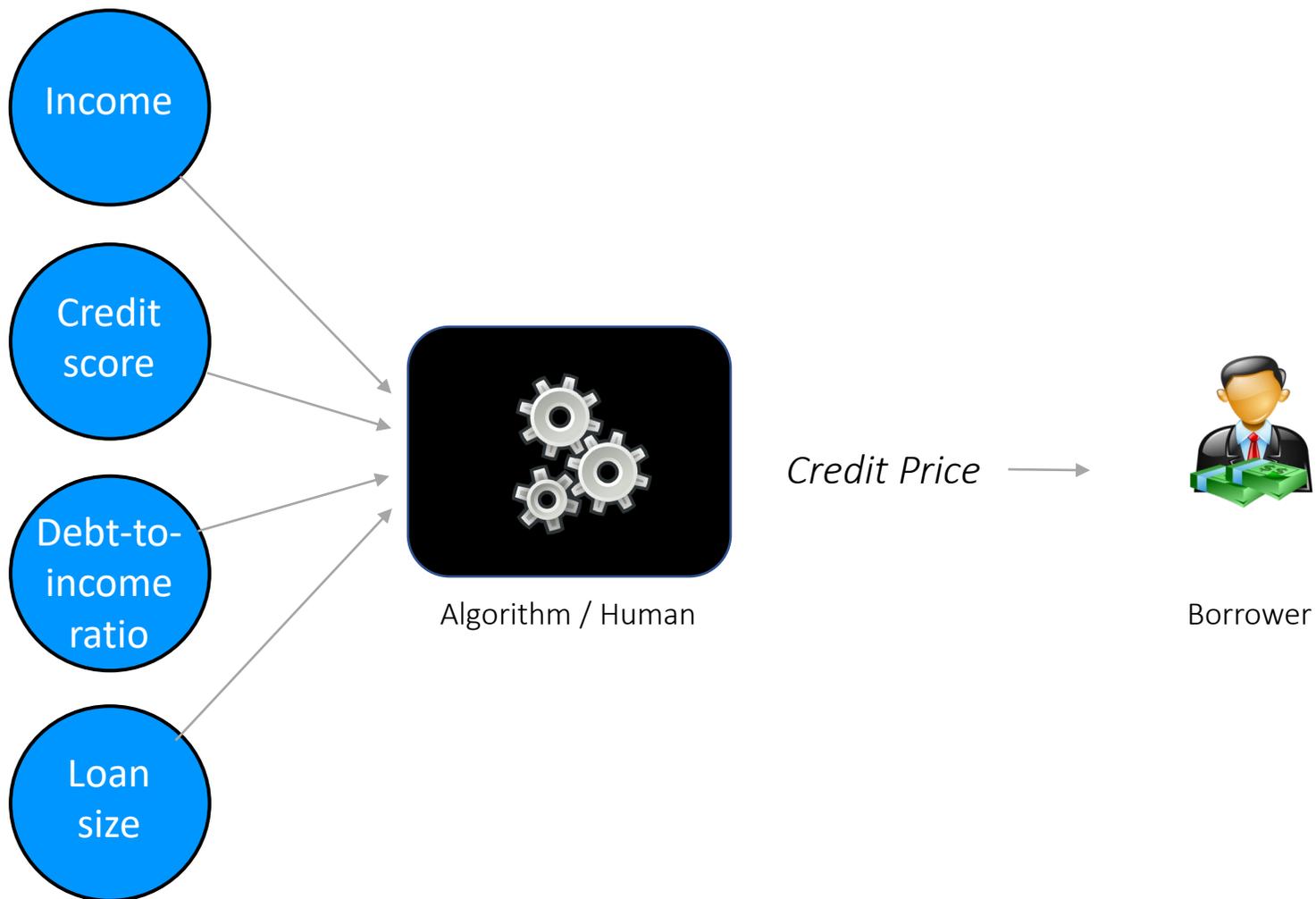# Department of Housing and Urban Development

"...where a plaintiff identifies an offending policy or practice that relies on an algorithmic model, a defending party may defeat the claim by: (1) identifying the inputs used in the model and showing that these inputs are not substitutes for a protected characteristic..."

- HUD Proposed Rule on Disparate Impact (August 2019)

Outline

- Fair lending
- Input-based approaches
- Avoidance approaches
- Paths forward

INPUTS ➡ **DECISION MAKING** ➡ **OUTPUT**

Income

Credit score

Debt-to-income ratio

Loan size

Algorithm / Human

Credit Price

Borrower

Fair lending

- Fair Housing Act (FHA)
  - Discrimination based on- race or color, national origin, religion, sex, familial status, handicap

- Equal Credit Opportunity Act (ECOA)
  - Discrimination based on- race or color, national origin, religion, sex, marital status, age, receipt of income from public assistance programs

Fair lending

- Disparate treatment
- Disparate impact
  - What is the purpose of disparate impact?
    - Covert intentional discrimination (*intent theory*)
    - Preventing disparate impact (*effect theory*)
  - What constitutes a "business justification"?
    - Pricing on default risk or also pricing on demand?
  - What constitutes a less discriminatory alternative?

Simulation Exercise

- Firm uses past default data to set new prices

- Data used to construct simulation:
  - The Home Mortgage Disclosure Act (HMDA) Data (Boston Fed)
  - Simulated default rates

1. Fit a prediction function on training data of 2000 old customers
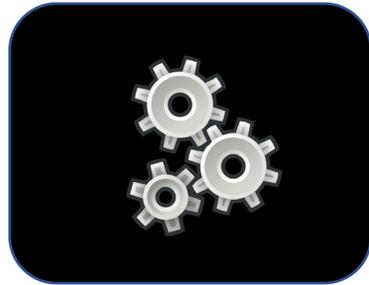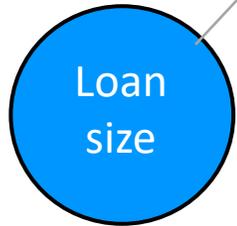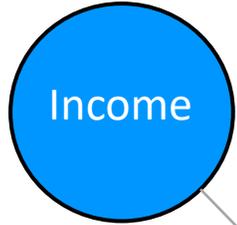2. Inspect predictions on hold-out data of 2000 "new customers"
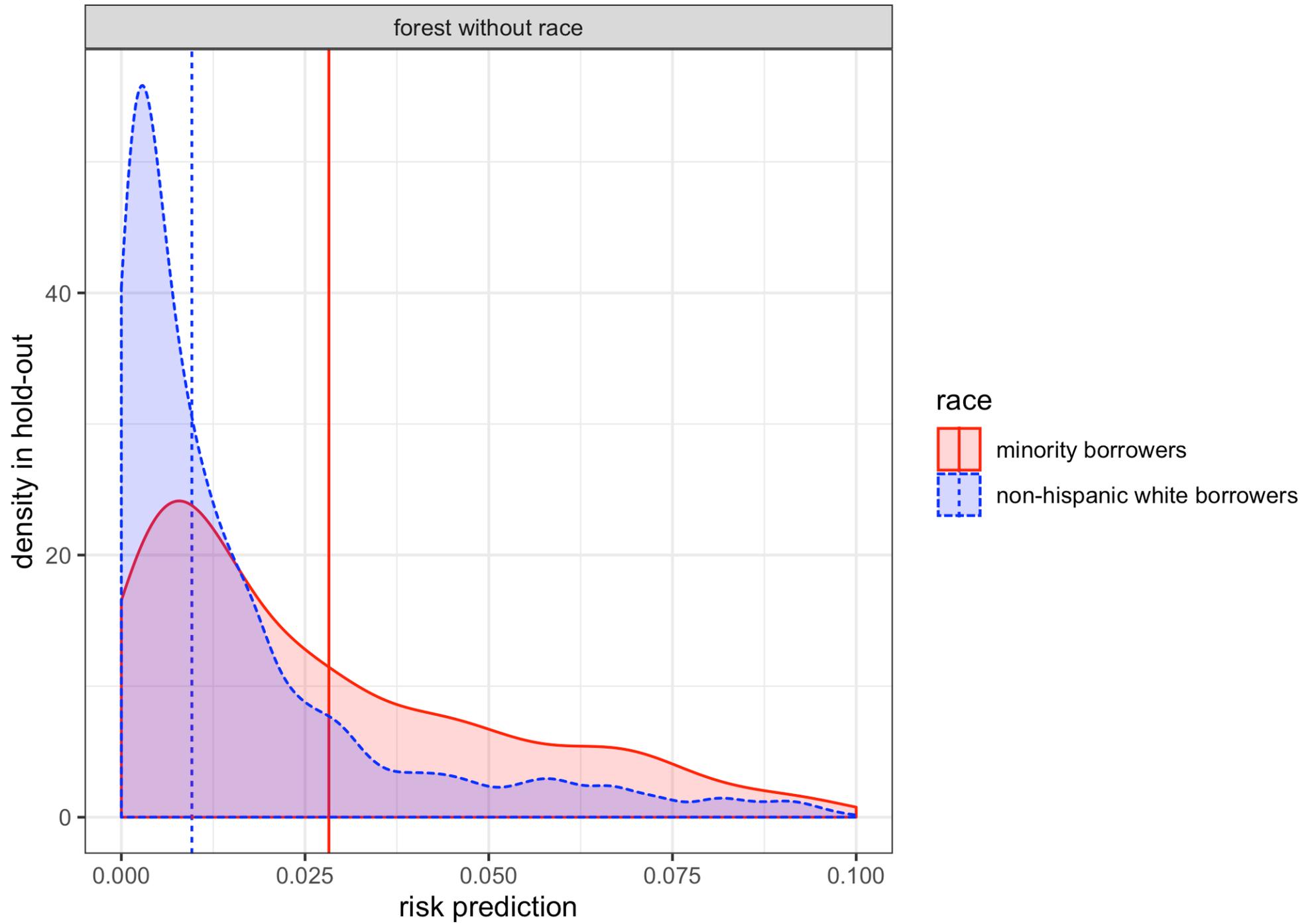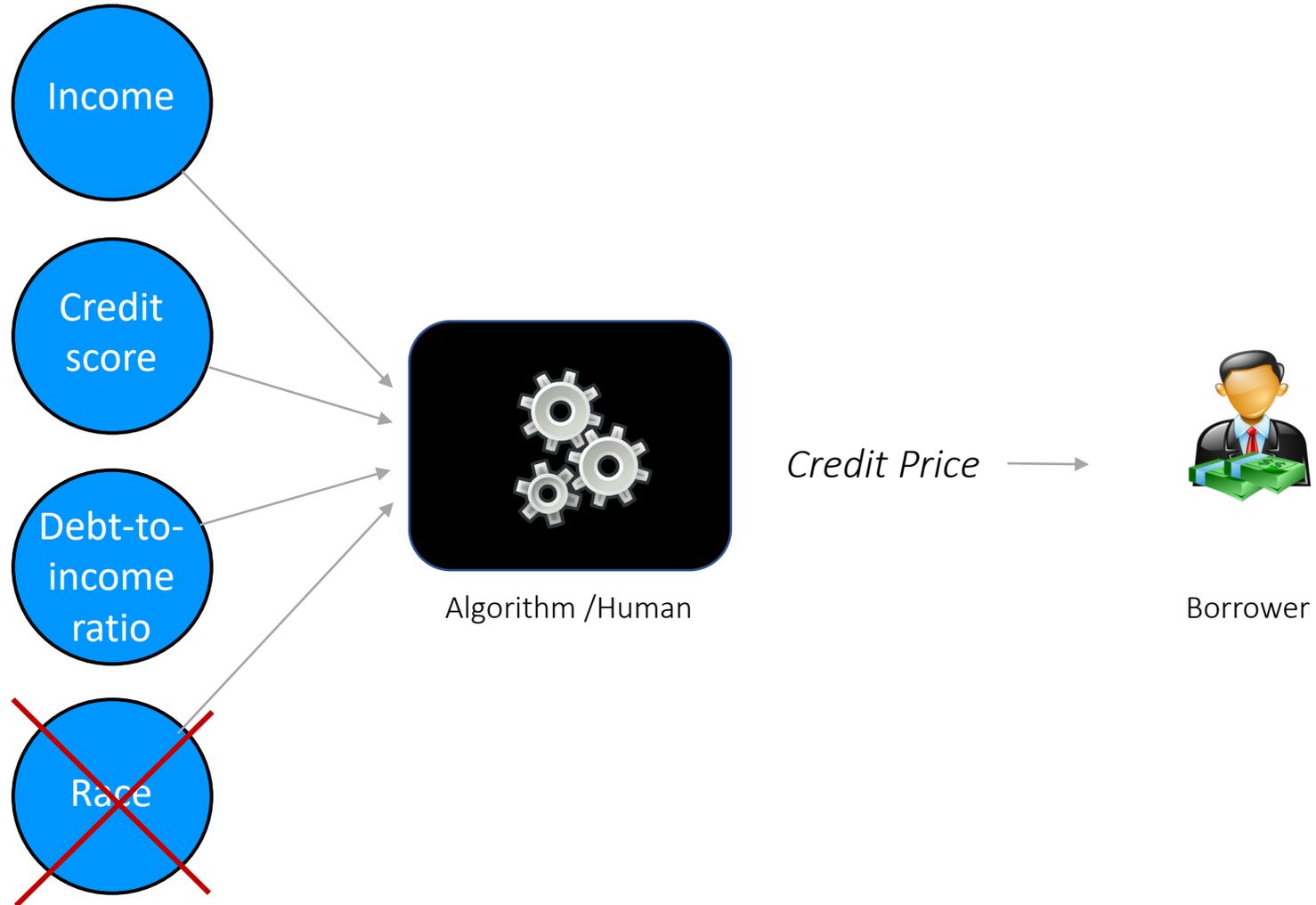
Outline

- Fair lending
- **Input-based approaches**
- Avoidance approaches
- Paths forward

Excluding the protected characteristics

Marisa Robertson

We have not and never will make decisions based on factors like gender. In fact, we do not know your gender or marital status during the Apple Card application process.

GS Bank Support ✓
@gsbanksupport
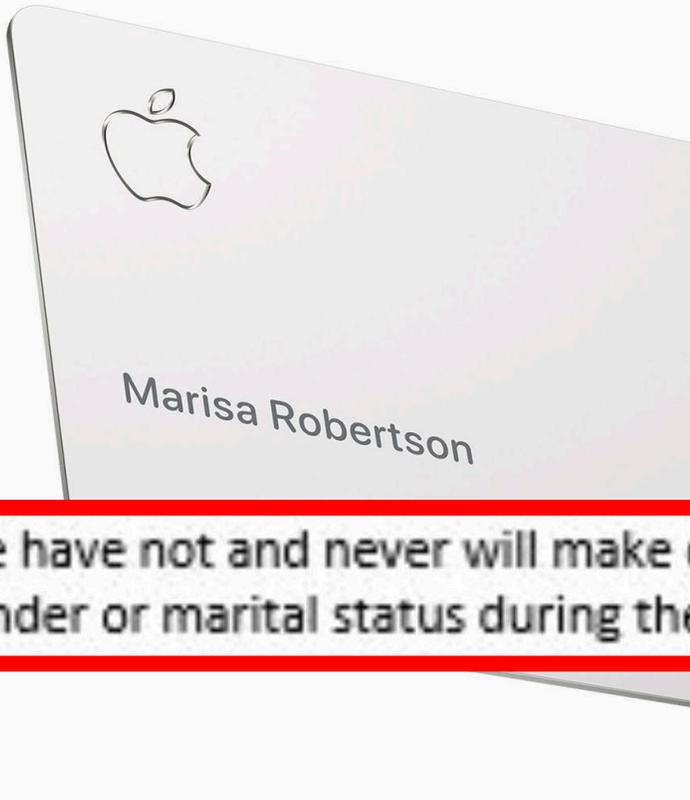
We hear you #AppleCard

We hear you. Your concerns are important to us and we take them seriously.

We have not and never will make decisions based on factors like gender. In fact, we do not know your gender or marital status during the Apple Card application process.

We are committed to ensuring our credit decision process is fair. Together with a third party, we reviewed our credit decisioning process to guard against unintended biases and outcomes.

process is not aware of your marital status at the time of the application.

If you believe that your credit line does not adequately reflect your credit history because you may be in a similar situation, we want to hear from you. Based on additional information that we may request, we will re-evaluate your credit line.

Thank you for being an Apple Card customer.

Carey Halio
Chief Executive Officer
Goldman Sachs Bank USA

5:42 PM · Nov 11, 2019 · Sprinklr Publishing

On the Fairness of Machine-Assisted Human Decisions

Talia Gillis

Columbia University

Bryce McLaughlin

Stanford University

Jann Spiess

Stanford University

October 28, 2021

Preliminary draft, comments welcome!

Abstract

When machine-learning algorithms are deployed in high-stakes decisions, we want to ensure that their deployment leads to fair and equitable outcomes. This concern has motivated a fast-growing literature that focuses on diagnosing and addressing disparities in machine predictions. However, many machine predictions are deployed to assist in decisions where a human decision-maker retains the ultimate decision authority. In this article, we therefore consider how properties of machine predictions affect the resulting human decisions. We show in a formal model that the inclusion of a biased human decision-maker can revert common relationships between the structure of the algorithm and the qualities of resulting decisions. Specifically, we document that excluding information about protected groups from the prediction may fail to reduce, and may even increase, ultimate disparities. While our concrete results rely on specific assumptions about the data, algorithm, and decision-maker, they show more broadly that any study of critical properties of complex decision systems, such as the fairness of machine-assisted human decisions, should go beyond focusing on the underlying algorithmic predictions in isolation.

Excluding the protected characteristics
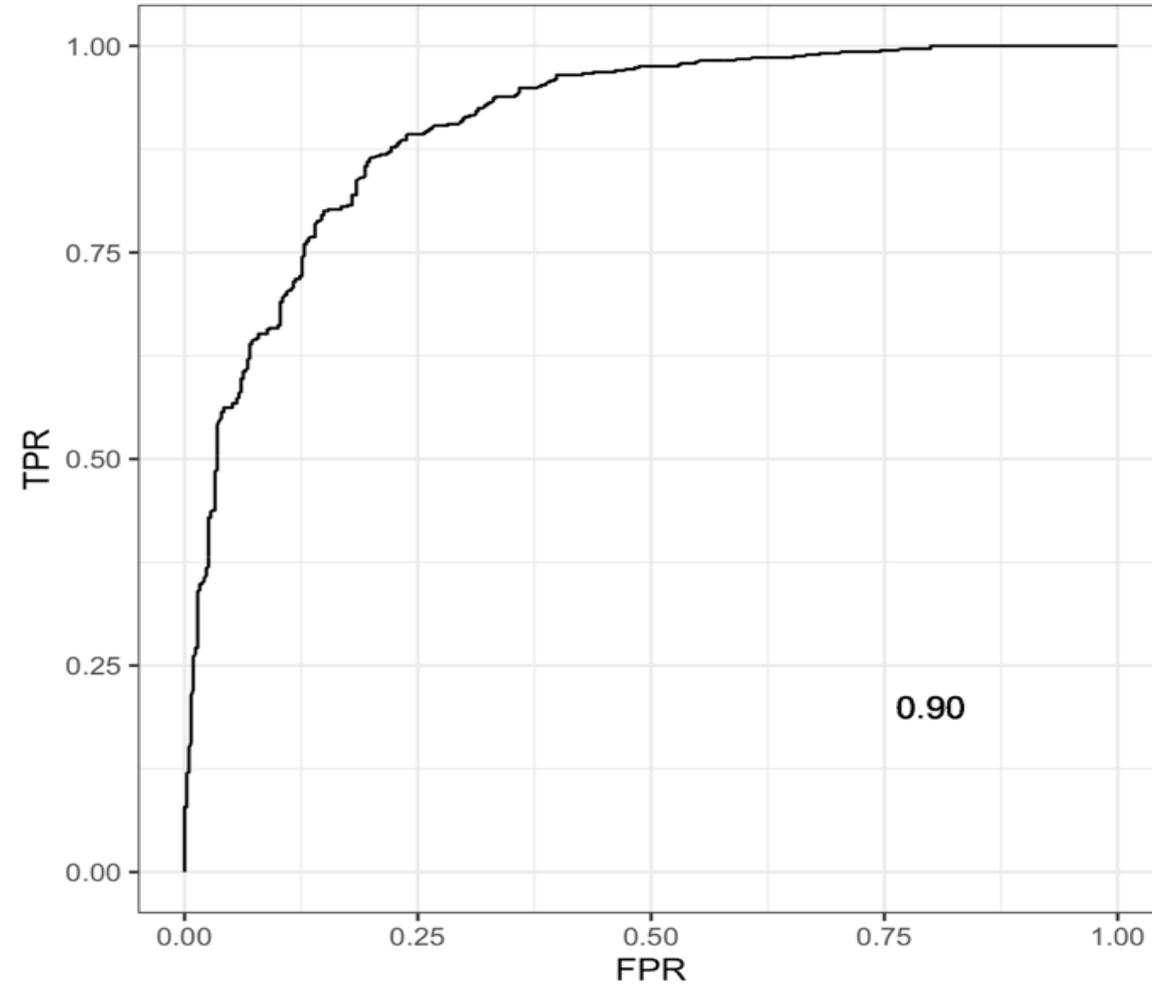
Excluding proxies
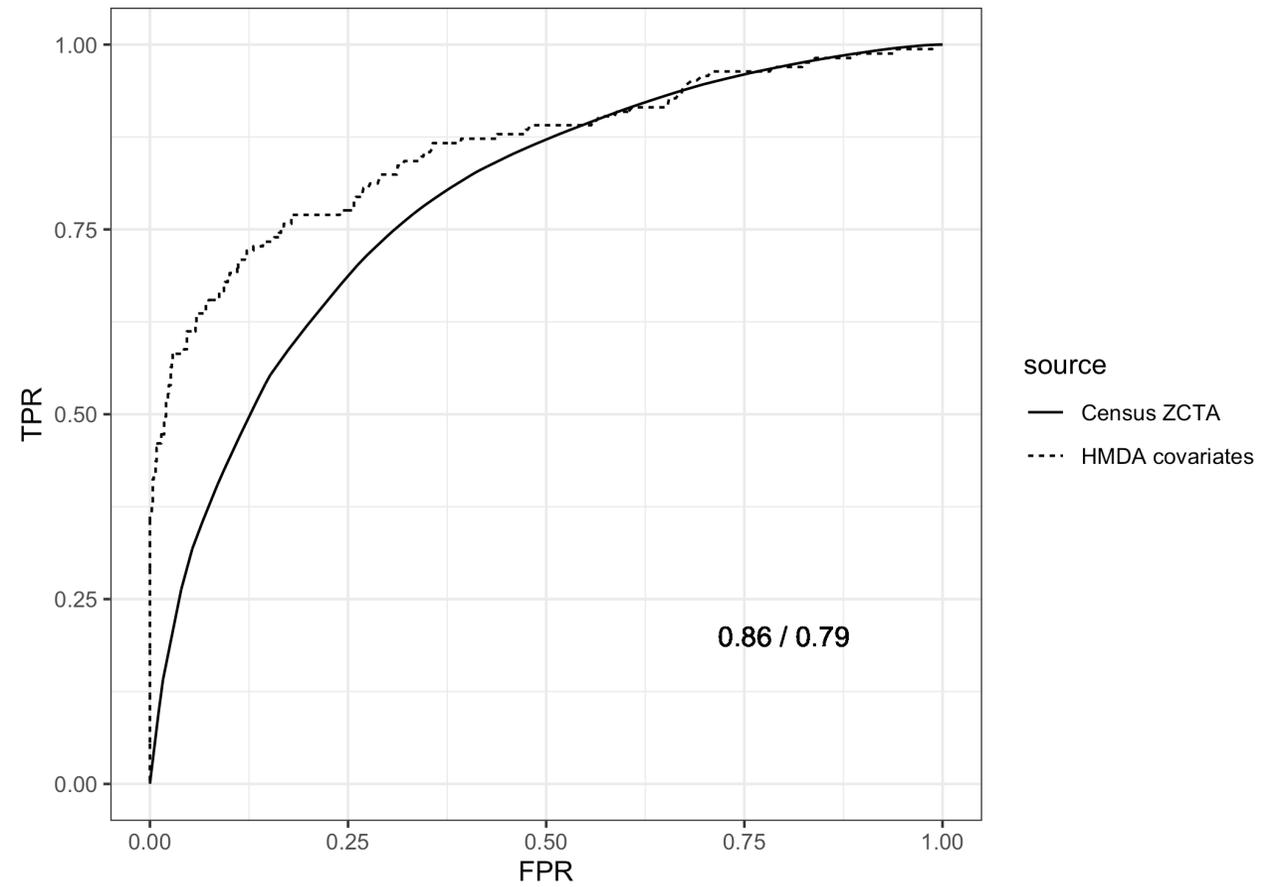
Outline

- Fair lending
- Input-based approaches
- **Avoidance approaches**
- Paths forward

EUROPEAN
COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

EU Proposal for the regulation of AI

- "In particular, **AI systems used to evaluate the credit score or creditworthiness of natural persons should be classified as high-risk** AI systems, since they determine those persons' access to financial resources or essential services such as housing, electricity, and telecommunication services."

- "AI systems used for this purpose **may lead to discrimination** of persons or groups and perpetuate historical patterns of discrimination, for example based on racial or ethnic origins, disabilities, age, sexual orientation, or create new forms of discriminatory impacts."

  - Preamble

EU Proposal for the regulation of AI

- Data accuracy:
  - Article 10: "Training, validation and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high-risk AI system is intended to be used."

- Transparency
  - Article 14: "High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which the AI system is in use."

Outline

- Fair lending
- Input-based approaches
- Avoidance approaches
- **Paths forward**

Incremental approach

- How do the disparities change?

- Consumer Financial Protection Bureau no action letter:
  - "conduct periodic access-to-credit testing to determine how Upstart's model compares to other credit models in enabling credit access, with results provided to the Bureau"

    - November 30, 2020