ALGORITHMIC DISCRIMINATION AND INPUT ACCOUNTABILITY UNDER THE CIVIL RIGHTS ACTS

Robert Bartlett
(UC Berkeley Law),
Adair Morse, Richard Stanton & Nancy Wallace
(UC Berkeley Finance)

+ a finance application of the ideas

ADAIR MORSE

OCTOBER, 2020

MIT-GOLUB CONFERENCE

SETTING

I. Setting

i. **Lenders** use 1,000s of variables for **algorithmic profiling.**

ii. **Challenge:** How to implement Civil Rights Act for determining what is legal **statistical discrimination**

II. Our Contribution

Going back to **Supreme Court caselaw + Congress law** provides explicit guidance:

- combining legal frame &

- economic fundamental model

We put these pieces together.

III. Economics frame

- Stat discrimination solves a **signal extraction problem**:

- In Lending:

  - We need **proxies for hidden variables** that are part of a model of expected credit risk.

- Starting point:

- What are the hidden variables?

  - Life-cycle or permanent income variables…

  - Income, income growth, wealth, cost of capital, cost of consumption, existing debt, etc

IV. Legal Framework: Burden-Shifting Doctrine

- Original frame from Supreme Court:
  - *Griggs v. Duke Power Co*
- Codified by Congress:
  - *Civil Rights Act of 1991*
- Important Caselaw from Supreme Court:
  - *Ricci v. DeStefano*
  - *Dothard v. Rawlinson*
- Codified Employment Cases to Lending
  - *Equal Credit Opportunity Act; Fair Housing Act*

- **First Burden:** Plaintiff must document "statistical disparities"
  - If plaintiff successful…

- **Second Burden:** The defendant must then "demonstrate that the challenged practice is consistent with business necessity."
  - If defendant successful…

- **Third Burden:** Plaintiff must show that an equally valid and less discriminatory practice was available that the employer refused to use

Business necessity is the target of signal extraction

For any proxy variable to satisfy business necessity, its correlation with race (e.g.) can only be through fundamental variables of expected credit risk.

V: Dothard v. Rawlinson

- A Prison wanted to hire guards
  Strength is business necessity
- Rather than measure strength of applications, use proxy of height
- Some female applicants sued, won.
- **Supreme Court**: Strength is legitimate as target and height predicts performance, but the height measurement penalizes females beyond business necessity

VI: Input Accountability Test

Econometrics version:

- Decompose height into that which predicts the target strength and a residual
- Test if the residual is still correlated with female:

Regress:　　　　$Height_i = \alpha \cdot Strength_i + \varepsilon_i$

Test:　　　　　$\varepsilon_i \perp gender$

VII. Lending Rendition

- Business necessity: credit risk

- Economic fundamental model: Expected cash flow model with life-cycle or permanent income variable targets

- Process: Training dataset/historical date. Decompose the input variable into that which predicts any of the fundamental model targets

- Test: Residual cannot be correlated with race/ethnicity/gender/etc.

Regress:
$$Ivy\ League_i = \alpha_1 \cdot Income_i + \alpha_2 \cdot CreditScore_i$$
$$+\alpha_3 Wealth_i + \alpha_4 Debt_i + \cdots . + \varepsilon_i$$

Test: $\quad \varepsilon_i \perp race \ldots. \qquad regress: \ \varepsilon_i = \beta_0 + \beta_1 race$

Proxy height fails $\Leftrightarrow \beta_1 \neq 0$

Further issues discussed in paper: (i): Unobservability of Target (Kleinberg, Ludwig, Mullainathan, Sunstein (2019), (ii): Measurement Error in Target, (iii): Standard errors as n grows large.

VIII. Contrast Burden-Shifting vs Predictive Accuracy

Lender : Wants to use ML to do credit scoring without discrimination

Corporate Lawyers: "To avoid discrimination, apply a 'least discriminatory' approach"

How?

1. Define the business necessity for using proxy variables
   - Courts: in lending = "credit risk" (not expected profit of loan)
2. Run predictive accuracy models of default
3. Then show that the algorithm uses the least discriminatory predictive model for a given level of predictive accuracy

VIII. Contrast Burden-Shifting vs Predictive Accuracy

Lender : Wants to use ML to do credit scoring without discrimination

Corporate Lawyers: "To avoid discrimination, apply a 'least discriminatory' approach"

How?

1. Define the business necessity for using proxy variables
   - Courts: in lending = "credit risk" (not expected profit of loan)
2. Run predictive accuracy models of default
3. Then show that the algorithm uses the least discriminatory predictive model for a given level of predictive accuracy

> Problem: Least discriminatory approach does not ensure compliance with 2nd burden.
> Court: Predictive accuracy is not sufficient.

IX. Why inputs rather than outputs?

Why not just fix by de-biasing?

    - Pope and Sydnor (2011)

a) Input focus is required under burden shifting

b) Ricci v DeStefano: New Haven firefighters tried to de-bias a qualifications test and the court ruled it is illawful to use a protected class variable in any process, including fixes.

X. Fairness vs Discrimination: Ventilators & Credit Scores

Triage algorithms to allocate ventilators based on **LT survival**

- **SOFA Algorithm**: degree of dysfunction, 6 organs

Problem: Legacy of structural racism and inequality => Black and Latinx Americans higher rates of diabetes, hypertension, etc.

- Under IAT: If LT Survival is business necessity, then whites getting more ventilators is justified.

- Fairness Arguments: Need legislation to re-define the business necessity target

What if **credit scores** are biased against people of color because

- Less chance to build credit histories because of structural inequities

- Turned down for credit because of discrimination, conditional on risk

Giles and Spiess (2019): "discrimination stress testing" in lending => (my relabel) "Fairness stress testing"

Points:

1) Fairness and discrimination law are not the same thing

2) More needs to be done

X. Implementation of IAT in Credit Scoring

Motivation:

Back of the envelope:

- New float of household debt in US ~$2.2 trillion per year

- Bartlett, et al (2019): ~45% algorithmic

Question:

- These lenders want to be empowered to use algorithms pre-tested for discrimination

- How much of a loss of predictive accuracy would IAT impose?

Data

- Consumer lender in Europe with 124 variables (over 700 in long) & default

- Test which variables IAT fails for gender

Test

- Always include fundamental cash flow variables

- Guided ML – parsed to 37 proxy/input variables with highest predictive accuracy
  - Guided= variables make sense

- 3 variables failed

- Predictive accuracy fell

- Area under ROC from 0.7434 to 0.7409 Pseudo R2 from 0.108 to 0.1054

Conclusions

Objectives:

- Get more finance research engaged in the policy debate about algorithmic use in credit scoring
- Debunk the emerging literature that AI poses no danger because it removes discretion, and any biases can be corrected
  - Note: we are very much in favor of technology in credit scoring (Bartlett et al 2019 – fintechs discriminate less), just with accountability

Contributions

1) Demonstrated what the law dictates about inputs & business necessity
2) Provided a really simple test for firms to use ex ante and regulators or courts ex post
3) Showed that at least in our application, the test provides results that are workable to firms