Algorithmic Accountability: A Legal and Economic Framework

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

Credit Risk Footprints and Algorithmic Discrimination

Adair Morse and Robert Bartlett

June, 2020

How Economists Think about Discrimination

TASTE-BASED DISCRIMINATION

- An individual dislikes members of a particular group and derives utility from discriminating against them (Becker 1957)
- Should not persist in the long run because of competition
  - *Discrimination is costly*
- De facto: discretion persists

STATISTICAL DISCRIMINATION

- A decision-maker (employer, lender) does not observe a business necessity variable (productivity, creditworthiness).
- Uses a proxy for that variable, such as the average for a group of people (Arrow, 1973, Phelps, 1972)
  - *Dicrimination profit maximizes*
- De facto use of statistical discrimination

Mostly ***indirect stat discrimination***:

using averages over a non-protected variable (not “black” but “high school name”) as a proxy for creditworthiness

How the Law Thinks about Discrimination

The mapping of the law to economists' thinking is clear on the below:

1. Make taste-based discrimination illegal

   (And anyway, it is not profit maximizing)

2. Make sure technology does not implement the direct form of Arrow/Phelps discrimination
   - i.e.: allowing lenders to score by a protected category or a "highly correlated" variable
     - Protected category: race, ethnicity, gender, etc.
     - Highly-correlated = hair styles, redlining, etc.

How the Law Thinks about Discrimination

But the law is not quite so simple as 1 and 2:

1. Make taste-based discrimination illegal
2. Make sure technology does not implement the direct form of Arrow/Phelps discrimination

Disparate treatment

What about indirect statistical discrimination??

Disparate Impact?

Proxy Variables for Statistical Discrimination & Accountability

Outline

I. Law / Caselaw

II. Input Accountability Test

III. Application in Credit Data

UK Law - Equality Act 2010

19. Indirect discrimination

(1) A person (A) discriminates against another (B) if A applies to B a provision, criterion or practice which is discriminatory in relation to a relevant protected characteristic of B's.

(2) For the purposes of subsection (1), a provision, criterion or practice is discriminatory in relation to a relevant protected characteristic of B's if—

(a) A applies, or would apply, it to persons with whom B does not share the characteristic,

(b) it puts, or would put, persons with whom B shares the characteristic at a particular disadvantage when compared with persons with whom B does not share it,

(c) it puts, or would put, B at that disadvantage, and

(d) A cannot show it to be a proportionate means of achieving a legitimate aim.

From U.K. to U.S.

My understanding with conversations with the FCA that the enforcement of the Equality Act regarding indirect discrimination maps to enforcement of Civil Rights Act of the U.S.

U.S. Title VII of the Civil Rights Act of 1964

An unlawful practice for an employer

1. "to … discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, sex, or national origin; or
2. to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities … because of such individual's race, color, religion, sex, or national origin."

A long-standing challenge: How do you implement this in a setting where discrimination may be unintentional?

Burden- Shifting Framework

Caselaw that was later codified as implementation law

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*

Aside

- Like the Civil Rights Act of 1964, 1991 and their caselaw, original application is in context of employment decisions.
- However, credit and housing decisions adopted the interpretation of discrimination and this framework explicitly in Equal Credit Opportunity Act and Fair Housing Act

Burden- Shifting Framework

First Burden: Plaintiff must identify a specific employment practice that causes "observed statistical disparities" across members of protected and unprotected groups.

- If plaintiff successful…

Second Burden: The defendant must then "demonstrate that the challenged practice is *job related for the position in question* and 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

Burden- Shifting Framework

Fair Lending laws adopted burden shifting for lending… switch employment language to creditworthiness

First Burden: Plaintiff must identify a specific em
"observed statistical disparities" across members
groups.

- If plaintiff successful…

Second Burden: The defendant must then "demonstrate that the challenged practice is *creditworthiness for the loan in question* and 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

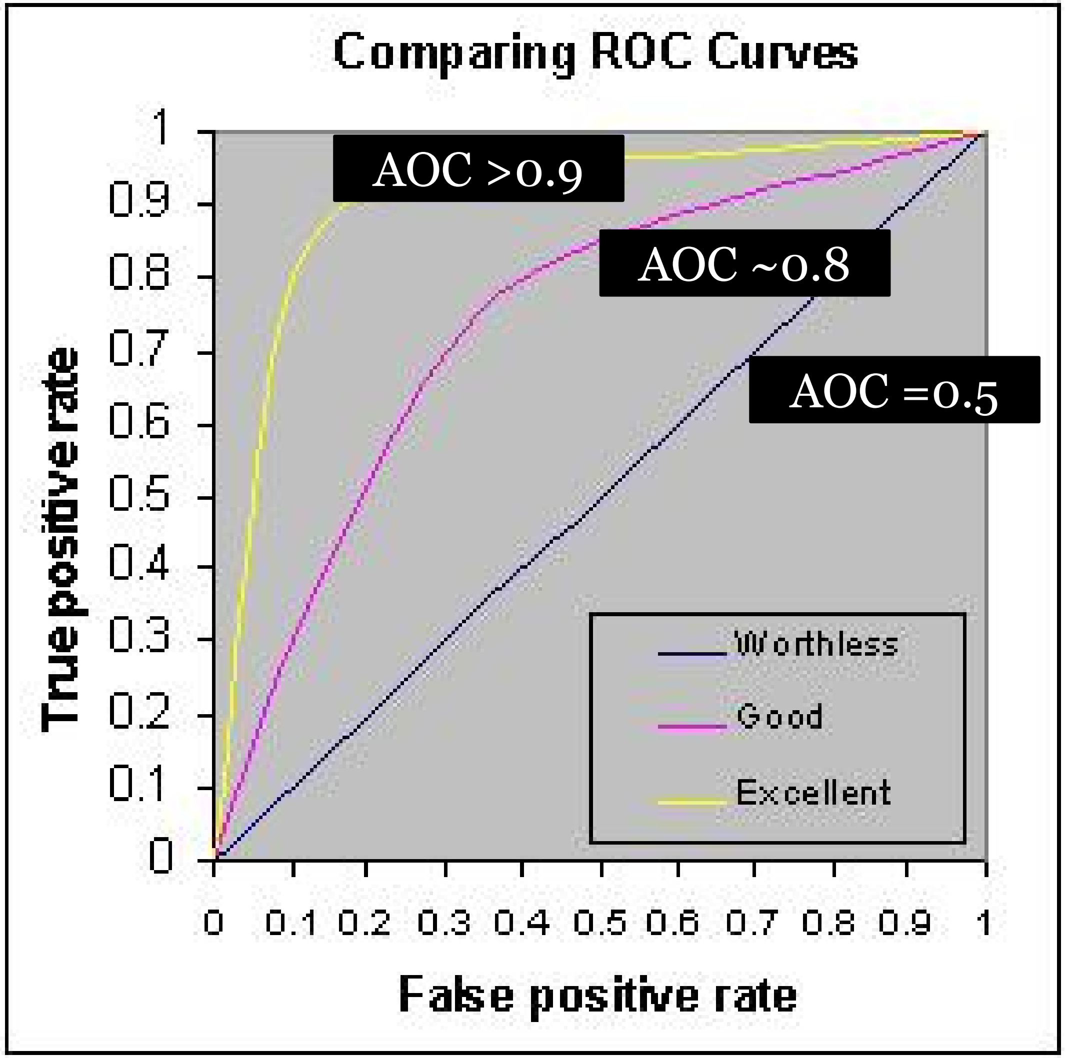
What do Lenders Say they do?

- Lender : a lender (platform, bank, etc.) with 1,000s of variables
- Objective: use machine learning (ML) to do credit scoring without discrimination
- *Corp. Lawyers*: *"To avoid discrimination, apply a 'least discriminatory' approach"*

How?

1. Define "target" (ML term) : = the business necessity for using proxy variables
   - Courts: in lending, target = "creditworthiness" not expected profit of loan
2. Run **predictive accuracy models of default**
   - Noting that default is ex post measure of ex ante credit risk
3. Then, if resulting outcomes are disparately applied against a protected category…
   - Lender needs to be able to show that the algorithm uses the least discriminatory predictive model for a given level of predictive accuracy

Problems with this:
Part 1: An econometrician / data scientist point of view



- ROC curves, think…
  - Run ML model of default on standard credit risk variables plus 1,000s of proxies for missing fundamentals
  - Calculate how predictive model is (goodness of fit)
- Imagine result…
  - *"my best predictive model generates ROC of 0.78"*
  - I can generate many models with interactions of variables /nonparametrics that have similar ROC
  - Which one has least impact on protected group?
- Problem: let's say with just pure cash flow variables the model yields ROC of 0.68. Does the court allow us to increase ROC by 0.10 and then apply the discrimination test?

Problems with this:
Part 2: It's illegal under Burden-Shifting Framework

First Burden: Plaintiff must identify a specific employment practice that causes "observed statistical disparities" across members of protected and unprotected groups.

Second Burden: The defendant must then "demonstrate that the challenged practice is *job related for the position in question* and consistent with business necessity."

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

#1: This is where the least discriminatory approach comes from

#2: But it does not excuse the defendant from satisfying Second Burden

Dothard v. Rawlinson

A California Prison wanted to hire prison guards

- Determined that a job-required necessity is strength (legitimate)
- Could not measure strength of applications, so used proxy of height
- A group of female applicants sued and won

Court:

- Indeed strength is legitimate as target and height predicts performance
- But the strength needed is a specific strength and the height measurement penalizes females beyond the business necessity

Dothard v. Rawlinson: IAT

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

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

Test: $\varepsilon_i \perp gender.....$ $\quad regress: \ \varepsilon_i = \beta_0 + \beta_1 gender$

Proxy height fails ⬄ $\beta_1 \neq 0$

If so, exclude height as only legitimate business necessity

We call this the *Input Accountability Test*

Challenges of the IAT

1. Unobservability of Target
    - Kleinberg, Ludwig, Mullainathan, Sunstein (2019): *training datasets*
    - Calculating thresholds

2. Measurement Error in Target

$$Strength_i^* = Strength_i + \mu_i$$

$$Height = \alpha \cdot Strength_i^* + \zeta_i$$

$$\zeta_i = -\mu_i + \varepsilon_i$$

Note: UnitedHealth is this problem. Also, selective labels problem (De-Arteaga, et al., 2018).

Idea: Structural version

3. Standard errors as n grows large.

Example: UnitedHealth (UH) - insurance co

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, SCIENCE (2019)

UH used an algorithm to inform hospitals about patients' sickness level

- Purpose: Effectively allocation of resources to the sickest patients
- Problem:
  - UH had gauged sickness using historical expense data (cost of care)
  - African-American patients historically spend less for the same illnesses and level of illness
- Result: The algorithm caused African Americans to receive substandard care as compared to white patients

A fix instead of exclude?

Question: Why can't we just fix the scoring by a protect group to de-bias?

- Pope and Sydnor (2011)

Answer: It only works on average, not for individuals. The law is about individuals

Answer: It is illegal. *Ricci v. DeStefano:*

New Haven wanted to discard the results of an "objective examination" that sought to identify city firefighters who were the most qualified for promotion because there was statistical racial disparity in the results against a minority group. A group of white and Hispanic firefighters sued, alleging that the city's discarding of the test results constituted race-based disparate-treatment.

Court ruled for plaintiff… no discarding

Why: Can't use a protected class variable in a decision because (again) it could cause disparities because of the averages part

Implementation: “Footprints & Discrimination”

Motivation

- U.S. household debt: $14 trillion
  - Increase of $1.3 trillion from peak in 2008 (NY Fed)
  - If annual debt turnover is 15%
- Then… new float of recent years ~$2.2 trillion per year
- Of this, how much algorithmically-decided based on 1,000s of proxy variables?
  - Bartlett, et al (2019): 45% of lenders in mortgages have fully automated lending (in 2018)

Jeff Budzik

CTO of ZestFinance:

“The models we put into production for our customers tend to have hundreds or thousands of variables in them. We have one with 2200 variables that’s running an auto lending business”

Footprints & Discrimination

Question

Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2019),

Bartlett, Morse, Stanton, and Wallace (2019)

How can the use of machine learning in credit profiling avoid being inadvertently discriminatory?

Outline of Application :

(1) ROC Analysis

(2) IAT Tests for Gender Discrimination

Data

- Data from a consumer lender in Eastern Europe
- 300,000 consumer loans
- Loans made in stores but not collateralized
- Dataset contains default (the target)

Unique:

- 124 variables (many of the them categorical)
- Can be made "long" into 1,000s of variables even without interactions

Step 1 – Looking for footprints

How well can we do as a ML-er?

Prediction target: Default via area under ROC assessment

Footprints of creditworthiness literature (abridged)

- Berg, Burg, Gombovic, and Puri (2019) :"digital footprints" type of device (tablet, computer, phone), operating system (Windows, iOS, Android), and email provider predicted default rates among the customers of a German lender.
- Bjorkegren and Grissen (2019) mobile phone usage data
- Vissing-Jorgensen (2010) : Consumer goods products people buy

Types of Variables

1. Fundamentals (cash flow, wealth, cost of capital)
2. Occupation
3. Goods
4. Shelter
5. Family Life
6. Soft Info Applying
7. Soft Info Credit

Fundamental Variables

	Mean	StDev		Mean	StDev
Income monthly	168,797	237,125	Missing data Credit Bureau	0.1350	0.3417
Credit Amount	599,028	402,494	# Outstanding Loans	4.3184	10.5095
Payment Amount	27,109	14,494	Prior Loans Delinquent %	0.0054	0.0312
payment_to_credit	0.0537	0.0225	How Delinquent, if any	0.0089	0.0851
payment_to_income	0.1809	0.0946	Ontime Prior Payments, if any	0.1371	0.2522
Homeowner	0.6937	0.4610	Percent of Prior Loans Closed, if any	0.0991	0.2089
Credit Score Max	0.6159	0.1561	Remaining Days on Last Issue	-928.0	644.8
Cedit Score Min	0.3996	0.1874	Days Since Last Issue	-419.3	526.3
# Credit Bureau Requests	0.2313	0.8568	Own Car?	0.3401	0.4737
			Age of Car, if any	0.3418	0.7508

Note: Monetary units are disguised.

Living / Family Variables

	Mean	StDev
Civil Marriage	0.0968	0.2957
Marriage	0.6388	0.4804
Widow	0.0523	0.2227
# Children	0.4171	0.7221
Rural	0.1047	0.3062
Large Metro	0.1572	0.3640

Goods Variables

	Mean	StDev
Purchase Price of Good	538,398	369,447
LTV of Loan to Good	1.1230	0.1240

Occupation Variables

	Mean	StDev		Mean	StDev
Low Skill Worker	0.2058	0.4043	Pensioner	0.1800	0.3842
Drivers Security	0.0824	0.2749	Working - Unnamed	0.5163	0.4997
Office Worker	0.1983	0.3987	Employ Commercial	0.2329	0.4227
Manager /Skilled	0.1658	0.3719	Employment Years	5.3562	6.3202
Prof Services	0.0344	0.1821	Gives Office Phone	0.8199	0.3843

Shelter Variables

	Mean	StDev
Municipal Housing	0.0364	0.1872
Office Housing	0.0085	0.0919
Live with Parents	0.0483	0.2143
Age Building	0.2532	0.3626
N/A Age Building	0.6650	0.4720
Elevators Relative	0.0365	0.0998
N/A Elevators	0.5330	0.4989
Entrances relative	0.0741	0.1028
N/A Entrances	0.5035	0.5000

Soft Application Variables

	Mean	StDev
# Documents	0.9302	0.3443
No Documents	0.0961	0.2947
# Contacts Provided	1.5371	0.7221
Social Network: Defaulters	0.1434	0.4466
Spouse Present	0.0370	0.1887

Prior Credit Proprietary Variables

	Mean	StDev
Previous Good Loan LTV	0.960	0.255
Previous Rejection %	0.223	0.257
# Previous Apps	4.597	4.180

ROC Analysis

Logit (Default) = fundamentals + (iteratively, then all)

1. Occupation
2. Goods
3. Shelter
4. Family Life
5. Soft Info Applying
6. Soft Info Credit

Logit (default) = function of Fundamental Variables)

	Dependent Variable: Default		
Ln Income	-0.151***	Homeowner	-0.0131
	[0.0391]		[0.0148]
Ln Credit Amount	-1.934***	Credit Score Max	-2.084***
	[0.0902]		[0.0464]
Ln Payment Amount	2.269***	Credit Score Min	-2.676***
	[0.0996]		[0.0467]
Payment_to_credit	-32.35***	# Credit Bureau Requests	-0.0112
	[1.540]		[0.00961]
Payment_to_income	-0.372*	Missing data,Credit Bureau	-0.141***
	[0.202]		[0.0245]
		Cut off the prior balances debt vars	
	Observations	307,321	
	Pseudo R-squared	**0.0872**	
	Area under ROC	**0.7217**	

ROC Analysis … Columns adding Proxies

Do the proxies add to the ROC?

How did the Guided ML (Lasso Optimizing) do?

Dependent Variable: Default
Model: Logit

		Variables Included: Fundamentals + ….						
	Funda-mentals	Occu-pation	Goods	Shelter	Family Life	Soft Info App	Soft Info Credit	All
Observations	307,321	307,321	307,045	307,321	307,321	307,321	306,302	306,026
Pseudo R-squared	**0.0872**	**0.0944**	**0.0937**	**0.0885**	**0.0872**	**0.0916**	**0.0904**	**0.108**
Area under ROC	**0.7217**	**0.7297**	**0.7289**	**0.7232**	**0.7217**	**0.7262**	**0.7255**	**0.7434**

Step 2: Which of those Proxy Variables pass the Input Accountability Test?

Example: test the variable "elevators".

- First, start with linear Decomposition: Proxy = fundamentals + residual
- Second: test if residual is correlated with female

Regress:	Elevators = a_1*creditscore+a_1*income+a_2*debt+....a_N*lastFundamantal+ residual
Regress:	Residual = b0 + b1* female
Test:	b1 != 0

- Concern: p-value on b1.... decreases with the number of observations mechanically

- Cannot go down an "economic significance" argument because this is law. There is no sense in the law that "5 people out of 10,000 do not matter"

- d-value approach to the p-value problem as n-> large

D-value : Demidenko (2013)

"The P-value You Can't Buy" American Statistician

- Rather than focus on a comparison of group means, the d-value is designed to examine how a randomly chosen female fared under this proxy variable relative to a randomly chosen male.

P value (under normality): $$p = \Phi\left(-\frac{|b|}{s}\right)$$

D-value (under normality): $$d = \Phi\left(-\frac{|b|}{s\sqrt{n}}\right)$$

Where s is the standard error: $s = \text{stdev}/\sqrt{n}$

<u>Foundations:</u>

- Individual observation comparison of this form are the foundation of the Wilcoxon-Mann-Whitney U Stat (for medians test)

- "D" comes from "discrimination" because the formulation is the same as the area under the ROC curve used for discrimination tests as early as Bamber (1975)

Family Lifestyle

	(1) Civil Marriage	(2) Non-civil Marriage	(3) Widow	(4) # Children	(5) Rural	(6) Large Metro
Coefficient from logit (default)	not signif.	-0.0999***	-0.146***	not signif.	-0.198***	0.0915***
Sign on residual estimation below that would indicate algorithmic bias against females	none	–	–	none	–	+
		Regression:	**Residual = bo + b1* female**			
female	0.0174	-0.0684	0.042	-0.00596	0.0112	0.00604
	[0.00112]	[0.00177]	[0.000833]	[0.00272]	[0.00110]	[0.00136]
Observations	307,321	307,321	307,321	307,321	307,321	307,321
R-squared	0.001	0.005	0.008	0.000	0.000	0.000
Standard errors in brackets						
On d-values below: range +/- 1% around 50% is not concerning						
d-value		47.2%	53.6%		50.7%	50.3%

Family Lifestyle

Having a non-civil marriage lowers default risk. Thus the scoring algorithm rewards those of this category.

	(1) Civil Marriage	(2) Non-civil Marriage	(3) Widow	(4) # Children	(5) Rural	(6) Large Metro
Coefficient from logit (default)	not signif.	-0.0999***	-0.146***	not signif.	-0.198***	0.0915***
Sign on residual estimation below that would indicate algorithmic bias against females	none	–	–	none	–	+
Regression: Residual = b0 + b1* female						
female	0.0174	-0.0684	0.042	-0.00596	0.0112	0.00604
	[0.00112]	[0.00177]	[0.000833]	[0.00272]	[0.00110]	[0.00136]
Observations	307,321	307,321	307,321	307,321	307,321	307,321
R-squared	0.001	0.005	0.008	0.000	0.000	0.000

Standard errors in brackets

On d-values below: range +/- 1% around 50% is not concerning

d-value		47.2%	53.6%		50.7%	50.3%

Family Lifestyle

Having a non-civil marriage lowers default risk. Thus the scoring algorithm rewards those of this category.

	(1) Civil Marriage	(2) Non-civil Marriage	(3) Wi	(4)	(5)	(6) Metro
Coefficient from logit (default)	not signif.	-0.0999***	-0.14	[illegible]	[illegible]	5***
Sign on residual estimation below that would indicate algorithmic bias against females	none	–	–	[illegible]	[illegible]	[illegible]
Regression: **Residual = b0 + b1* female**						
female	0.0174	-0.0684	0.042	-0.00596	0.0112	0.00604
	[0.00112]	[0.00177]	[0.000833]	[0.00272]	[0.00110]	[0.00136]
Observations	307,321	307,321	307,321	307,321	307,321	307,321
R-squared	0.001	0.005	0	[illegible]	[illegible]	000

But the residual of non-civil marriage after orthogonalizing to the credit risk fundamentals is negatively correlated with being female. Thus the use of this variable overly penalizes females.

Standard errors in brackets

On d-values below: range +/- 1% around 50% is not concerning

d-value		47.2%	53.6%		50.7%	50.3%

Is it significant? Yes. The d-value is different from 50% by >1%

Occupation – part 1

	(1) Low Skill Worker	(2) Drivers Security	(3) Office Worker	(4) Manager /Skilled	(5) Prof Services
Coefficient from logit (default)	0.195***	0.308***	Not signif.	Not signif	-0.253***
Sign on residual estimation below that would indicate algorithmic bias against females	+	+	none	none	–
			Regression:	**Residual = b0 + b1* female**	
female	-0.166	-0.157	0.148	0.0571	0.0482
	[0.00150]	[0.000988]	[0.00149]	[0.00139]	[0.000685]
Observations	307,321	307,321	307,321	307,321	307,321
R-squared	0.038	0.076	0.031	0.005	0.016

Standard errors in brackets

On d-values below: range +/- 1% around 50% is not concerning

d-value	42.1%	38.7%			55.1%

Occupation – part 1

	(1) Low Skill Worker	(2) Drivers Security	(3) Office Worker	(4) Manager /Skilled	(5) Prof Services
Coefficient from logit (default)	0.195***	0.308***	Not signif.	Not signif	-0.253***
Sign on residual estimation below that would indicate algorithmic bias against females	+	+	none	none	–

Regression:

	(1)	(2)	(3)	(4)	(5)
female	-0.166	-0.157			
	[0.00150]	[0.000988]			
Observations	307,321	307,321			
R-squared	0.038	0.076			

Standard errors in brackets

On d-values below: range +/- 1% around 50% is not concerning

	(1)	(2)	(3)	(4)	(5)
d-value	42.1%	38.7%			55.1%

Note: these d-values are all different from 50%, but the signs are the opposite of the concern about algorithmic bias against women. In fact, use of these variables discriminates against men.

Occupation – part 2

	(1) Pensioner	(2) Working - Unnamed	(3) Employ Commercial	(4) Employment Years	(5) Gives Office Phone
Coefficient from logit (default)	-2.119***	0.270***	0.168***	-0.0266***	-1.917***
Sign on residual estimation below that would indicate algorithmic bias against females	–	+	+	–	–
		Regression:	**Residual = b0 + b1* female**		
female	0.0494	-0.079	0.00753	0.323	-0.0495
	[0.00140]	[0.00187]	[0.00157]	[0.0235]	[0.00140]
Observations	307,321	307,321	307,321	307,321	307,321
R-squared	0.004	0.006	0.000	0.001	0.004
Standard errors in brackets					
On d-values below: range +/- 1% around 50% is not concerning					
d-value	56.3%	45.7%	48.3%	50.6%	43.7%

Occupation – part 2

Giving an office phone number implies less risky.

	(1)	(2)	(3)	(4) Employment Years	(5) Gives Office Phone
Coefficient from logit (default)	-2.119***	0.270***	0.168***	-0.0266***	-1.917***
Sign on residual estimation below that would indicate algorithmic bias against females	–	+	+	–	–
		Regression:	Residual = b0 + b1* female		
female	0.0494	-0.079	0.00753	0.323	-0.0495
	[0.00140]	[0.00187]	[0.00157]	[0.0235]	[0.00140]
Observations	307,321	307,321	307,321	307,321	307,321
R-squared	0.004	0.006	0.000	0.001	0.004

Standard errors in brackets

On d-values below: range +/- 1% around 50% is not concerning

d-value	56.3%	45.7%	48.3%	50.6%	43.7%

Occupation – part 2

Giving an office phone number implies less risky.

The residual after orthogonalizing "giving office phone" to fundamental variables, is negatively correlated with female, perhaps due to social norms.

Therefore the use of this variable biases against females.

	(1)	(2)	(3)	(4)	(5)
	[illegible]	[illegible]	[illegible]	Employment Years	Gives Office Phone
Coefficient from logit (default)	-2.119***	0.270***	0.168***	-0.0266***	-1.917***
Sign on residual estimation below that would indicate algorithmic bias against females	–	+	+	–	–
				[illegible]ual = bo + b1* female	
female	[illegible]	[illegible]	[illegible]	0.323	-0.0495
	[illegible]	[illegible]	[illegible]	[0.0235]	[0.00140]
Observations	307,321	307,321	307,321	307,321	307,321
R-squared	[illegible]	[illegible]	[illegible]	0.001	0.004

Standard errors in brackets

On d-values below: range +/- 1[illegible]

	(1)	(2)	(3)	(4)	(5)
d-value	56.3%	45.7%	48.3%	50.6%	43.7%

Shelter – part 1

	(1) Municipal Housing	(2) Office Housing	(3) Live with Parents	(4) Age Building	(5) N/A Age Building
Coefficient from logit (default)	0.105***	-0.255***	Not signif	-0.441***	-0.267***
Sign on residual estimation below that would indicate algorithmic bias against females	+	–	none	–	–
			Regression:	**Residual = b0 + b1* female**	
female	0.00467	-0.00134	-0.0149	0.0146	-0.0193
	[0.00071]	[0.000349]	[0.000799]	[0.00136]	[0.00178]
Observations	307,321	307,321	307,321	307,321	307,321
R-squared	0.000	0.000	0.001	0.000	0.000
Standard errors in brackets					
On d-values below: range +/- 1% around 50% is not concerning					
d-value	50.5%	49.7%		50.8%	**49.2%**

Shelter – part 2

	(1) Elevators Relative	(2) N/A Elevators	(3) Entrances relative	(4) N/A Entrances
Coefficient from logit (default)	-0.255**	Not signif	-0.298**	Not signif
Sign on residual estimation below that would indicate algorithmic bias against females	–	none	–	none
	Regression: **Residual = b0 + b1* female**			
female	0.00194	-0.0242	0.00294	-0.0263
	[0.000373]	[0.00186]	[0.000386]	[0.00187]
Observations	307,321	307,321	307,321	307,321
R-squared	0.000	0.001	0.000	0.001
Standard errors in brackets				
On d-values below: range +/- 1% around 50% is not concerning				
d-value	50.4%		50.5%	

Goods & Proprietary Prior Credit

	(1) Goods Price	(2) Goods LTV	(1) previous good loan LTV	(2) Previous Rejection %	(3) # Previous Apps
Coefficient from logit (default)	-5.25 e-07***	0.947***	0.213***	0.617***	-0.0109***
Sign on residual estimation below that would indicate algorithmic bias against females	–	+	+	+	–
	Regression: **Residual = b0 + b1* female**				
female	5437	-0.00547	0.0154	0.0139	0.439
	[563.8]	[0.000463]	[0.000950]	[0.000962]	[0.0156]
Observations	307,045	307,045	307,321	307,321	307,321
R-squared	0.001	0.000	0.001	0.001	0.003
Standard errors in brackets					
On d-values below: range +/- 1% around 50% is not concerning					
d-value	50.7%	49.1%	51.8%	50.4%	51.6%

Soft Info – Application Variables

	(1) # Documents	(2) No Documents	(3) # Contacts Provided	(4) Social Network: Defaulters	(5) Spouse Present
Coefficient from logit (default)	-0.317***	-0.615***	0.0515***	0.160***	-0.0492
Sign on residual estimation below that would indicate algorithmic bias against females	–	–	+	+	none
		Regression:	**Residual = b0 + b1* female**		
female	-0.00753	0.0035	-0.0121	0.0111	-0.0155
	[0.00121]	[0.00102]	[0.00273]	[0.00170]	[0.000715]
Observations	307,321	307,321	307,321	306,302	307,321
R-squared	0.000	0.000	0.000	0.000	0.002
Standard errors in brackets					
On d-values below: range +/- 1% around 50% is not concerning					
d-value	49.6%	50.1%	49.5%	50.7%	

Eliminate & Re-run Default Model

Eliminate 3 of 37 variables for bias

- previous goods loan-to-value
- non-civil marriage
- gives phone for employer

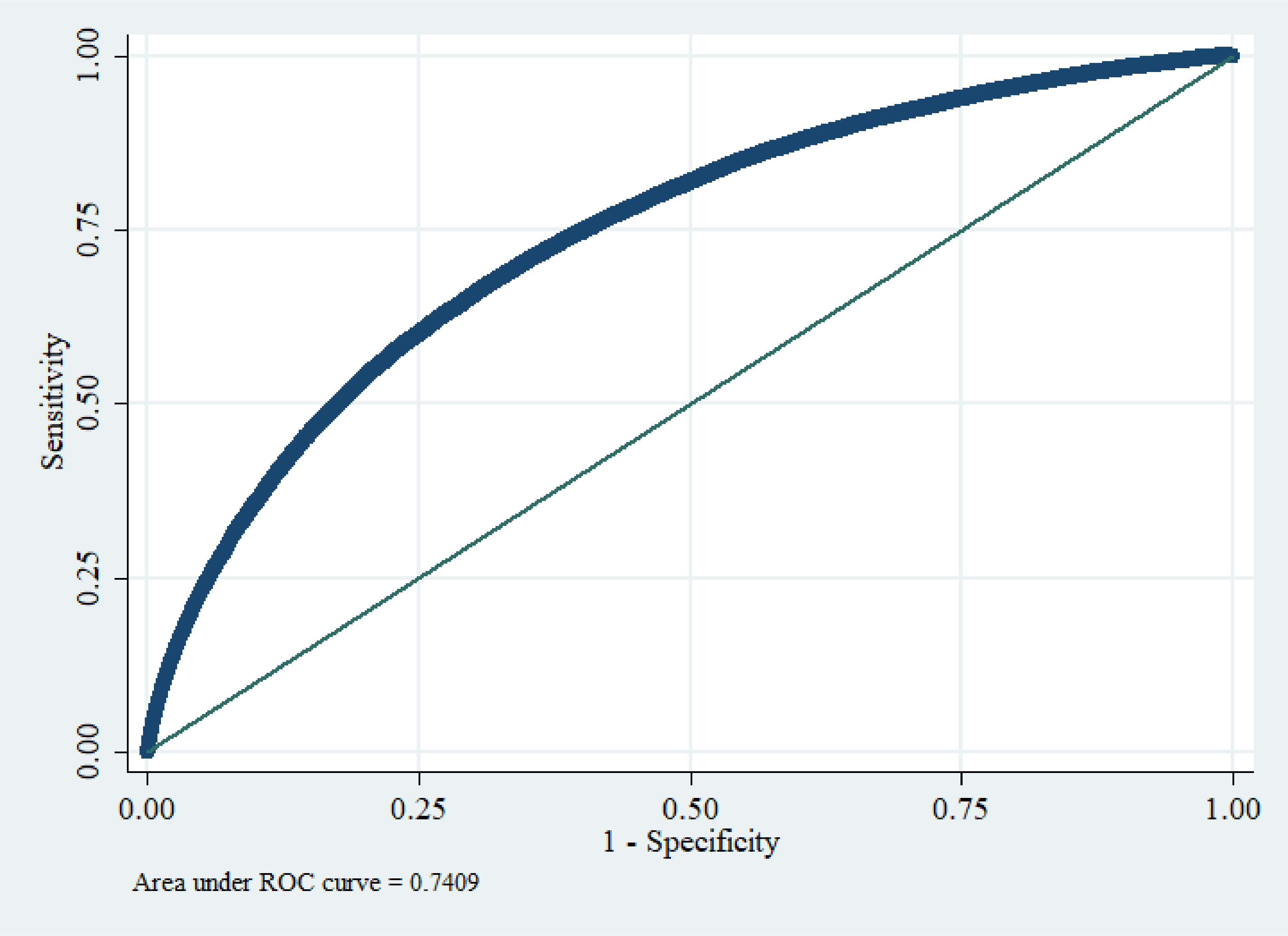
-

How much area under the ROC curve / pseudo r-square is sacrificed?

Re-running Logit (default) dropping biased proxies

Area under ROC drops from 0.7434 to 0.7409

Pseudo rsquared drops from 0.108 to 0.1054



Logistic regression Number of obs = 306,026 Pseudo R2 = 0.1054

	Coef	Z-stat		Coef	Z-stat
lnamt_income_total	0.0909	2.27	occ_lowskilllabor	0.1927	8.29
lnamt_credit	-1.1398	-10.31	occ_drivers_security	0.2659	9.38
lnamt_payment	1.4618	13.44	occ_office_workers	-0.0306	-1.25
payment_to_credit	-24.1704	-14.44	occ_managers_skill	-0.0315	-1.20
payment_to_income	0.7993	3.94	occ_profserv	-0.2464	-5.00
Homeowner	0.0007	0.04	employ_pensioner	-0.2186	-5.24
Max Credit Score	-1.8938	-39.94	employ_workingunnamed	0.2696	8.49
Min Credit Score	-2.4276	-51.09	employ_commercial	0.1472	4.35
# Request Credit Bureau	-0.0141	-1.44	employed_years	-0.0268	-17.82
Missing Requests	-0.1116	-4.5	shelter_municipal	0.1044	2.86
age_car	-0.0398	-4.27	shelter_office	-0.2463	-3.00
amt_goods_price	0.0000	-9.06	shelter_parents	0.0322	1.13
ltv	0.9696	15.64	years_build_medi	-0.3906	-3.32
bb_outstanding_count	0.0005	0.48	na_years_build_medi	-0.2273	-2.51
bb_delinquent	1.8054	6.21	elevators_medi	-0.4098	-4.00
bb_howdelinquent	-0.2077	-1.98	na_elevators_medi	-0.0001	0.00
bb_ontime	-0.1306	-4.22	entrances_medi	-0.2567	-2.19
bb_succ_closed	-0.1768	-3.71	na_entrances_medi	0.0119	0.30
Days outstanding on credit	0.0002	11.93	documents_count	-0.4149	-7.73
Days outstanding on last credit	-0.0001	-2.26	documents_none	-0.7271	-11.51
prev_rej_count_pct	0.6405	20.85	contacts_personal_count	0.0414	4.25
prev_apps_HC_count	-0.0090	-4.59	Network Defaulters	0.1596	11.83

To do’s

1. What is the cost in dollars and counts of people from a wrong prediction due to excluding the variables failing the IAT?
2. What if one does not have all the fundamental variables?
   - Step into the benefit of each grouping of variables
   - Then the cost of failing the IAT is more, presumably
3. Add in the final dataset of credit card transaction data
4. Interactions? More ML?

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

Accomplished (hopefully)

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