Anti-discrimination Laws, AI, & Gender Bias in Credit Lending

Stephanie Kelley
Smith School of Business, Queen's University, Canada

work with Anton Ovchinnikov, Smith School of Business, INSEAD

Motivation

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

Entrepreneur David Heinemeier Hansson says his credit limit was 20 times that of his wife, even though she has the higher credit score

GS Bank Support
@gsbanksupport

We hear you #AppleCard

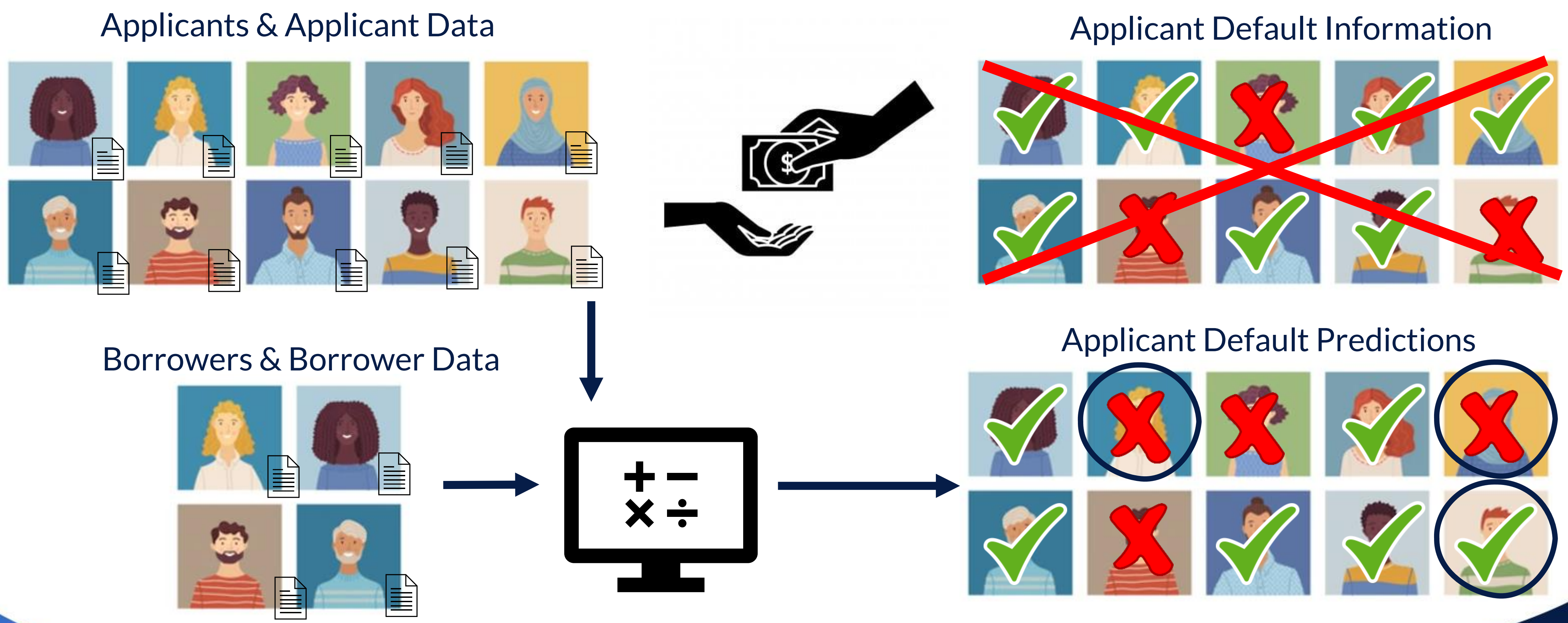
We hear you. Your concerns are important to us and we take them seriously.

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.

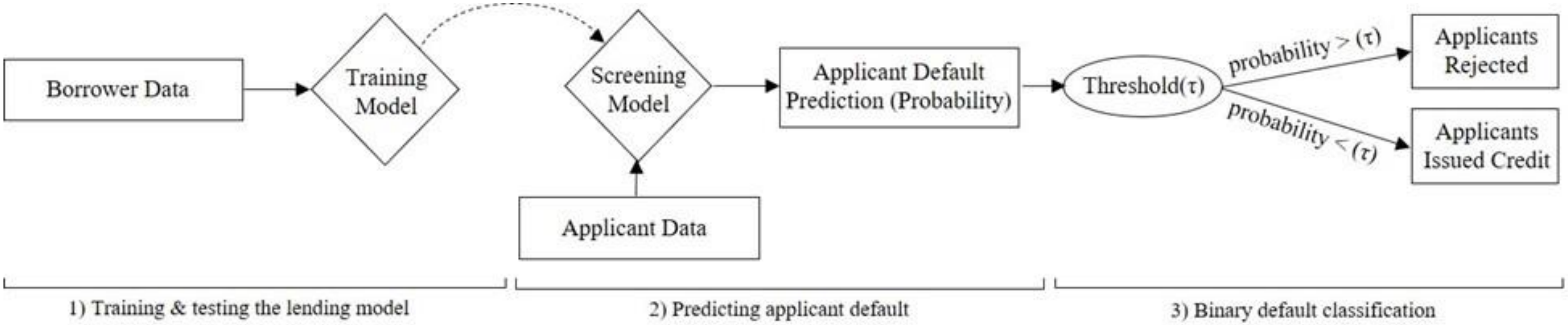
Do certain anti-discrimination laws paradoxically increase gender discrimination (in non-mortgage consumer fintech lending)?

What can firms do to reduce discrimination whilst managing profitability?

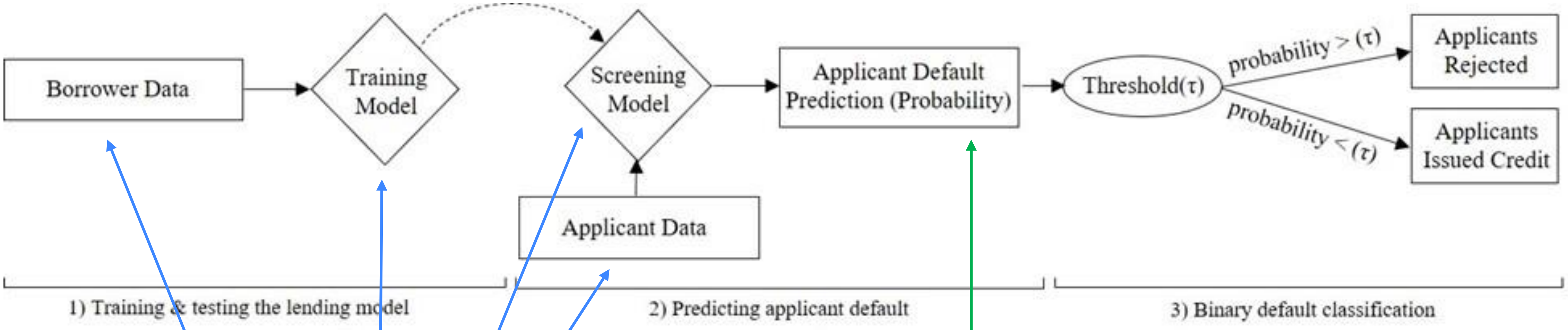
Background: Fintech Lending



Background: Fintech lending



Background: Fintech lending & Discrimination



Δ between model prediction quality for men & women

Legal Definition of Discrimination

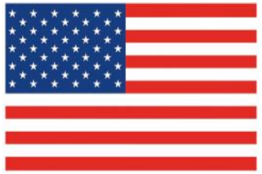
- Directions for data collection, governance and model building
- “Simulated” in study

Societal Definition of Discrimination

- Inequality
- Non-comparative wrong
- One model is better at predicting men vs. women
- “Measured” in study

Background: Legal Definition of Discrimination (Regimes)

What we "simulate" in the paper – affects data collection & modelling process

	Discrimination Across Anti-Discrimination Regimes		
	Regime 1	**Regime 2**	**Regime 3**
Collect gender	✓	✓	✗
Use gender in training model	✓	✓	✗
Use gender in screening model	✓	✗	✗
Legal Jurisdiction Example			

GDPR not applicable

Literature Review

Operations Research:

Empirical study of technology-based businesses (e.g., Cui et al 2018, Cohen & Harsha 2020)
Discrimination in technology-based businesses (e.g., Pope & Sydnor 2011, Ge et al 2016, Chan & Wang 2018, Cui et al 2020 Mejia & Parker 2020)
Algorithmic discrimination (Lambrecht & Tucker 2019, Obermeyer et al 2019)

Financial Economics:

ML discrimination in fintech mortgage lending (Fuster et al 2018, Barlett et al 2019)
LR discrimination in traditional non-mortgage lending (Chandler & Ewert 1979, Andreeva and Matuszyk 2019)

Computer Science:

Excluding protected attributes (Zliobaite & Custer 2016, Kleinberg et al 2018, Lipton et al 2019)
Measures of discrimination (e.g., Chouldechova 2017, Zliobaite 2017)
Drivers of ML discrimination (e.g., Kleinberg et al 2020)
Reducing discrimination
- Pre-processing (e.g., Kamiran & Calders 2012, Chen et al 2018)
- In-processing (e.g., Zafar et al 2017, Perrone et al 2020)
- Post-processing (e.g., Hardt et al. 2016)

Fairness-accuracy trade-off (see Zliobaite 2015)

Key Findings Summary

1. Prohibiting the use of gender leads to increased discrimination, decreased firm profitability, without impacting accuracy in both ML and LR models

2. ML is less discriminatory, more profitable, and more accurate compared to LR

3. Collecting gender, and using it in pre-processing & training allows firms to significantly decrease discrimination, whilst managing profitability
    a. Down-sampling to rebalance gender
    b. Gender-aware hyperparameter tuning
    c. Up-sampling to rebalance gender
    d. Probabilistic gender proxy modeling

The algorithmic mechanism behind the ML discrimination is discussed in the full paper

Data

ID	Gender (A)	Credit Amount ($X_1 \ldots X_{740}$)	Occupation	Social Circle	House Material	……	Known Default (Y)
1	M	14,625	5	0	6	……	1
2	F	1,750	9	1	5	……	0
3	F	8,950	7	2	NA	……	0
4	M	7,461	14	1	1	……	1

- Data from 305,000 borrowers, with known default outcome (Y)
- Data is publicly available on Kaggle: "Home Credit Default Risk" ML competition
- 740 features ($X_1 \ldots X_{740}$) including the protected attribute, gender (A)
- Gender (A) can be predicted with 91% AUC from other features

Modelling Methodology

Logistic Regression Model:

- Used popular LR credit scoring methodology (Hand, 2004)

Machine Learning Model:

- Compared 50+ ML algorithms
- Top performer based on AUC
- Ensemble – xgboost models

- 2 models each to simulate 3 regimes
  1. LR model with gender
  2. LR model without gender
  3. ML model with gender
  4. ML model without gender

- 80% train/20% test

- Output: predicted probabilities

Methodology: Key Metrics & Explanations

Discrimination

- Positive predictive value (Chouldechova, 2017)

$$PPV(\tau) = \frac{TP_M(\tau)}{[TP_M(\tau)+FP_M(\tau)]} - \frac{TP_W(\tau)}{[TP_W(\tau)+FP_W(\tau)]}$$

- Weighted mean group difference (Zliobaite, 2017)

$$WGMD(\tau) = \left[\frac{\sum_{i=1}^{N_W} \hat{Y}_{iW}(\tau)=1}{N_W}\right]_{ModelA} - \left[\frac{\sum_{i=1}^{N_W} \hat{Y}_{iW}(\tau)=1}{N_W}\right]_{ModelB}$$

- Avg across threshold (τ) range (5-30%)
- Across 30 folds

Predictive Quality

- AUC
- 95% CI, Delong method, 2000 bootstraps

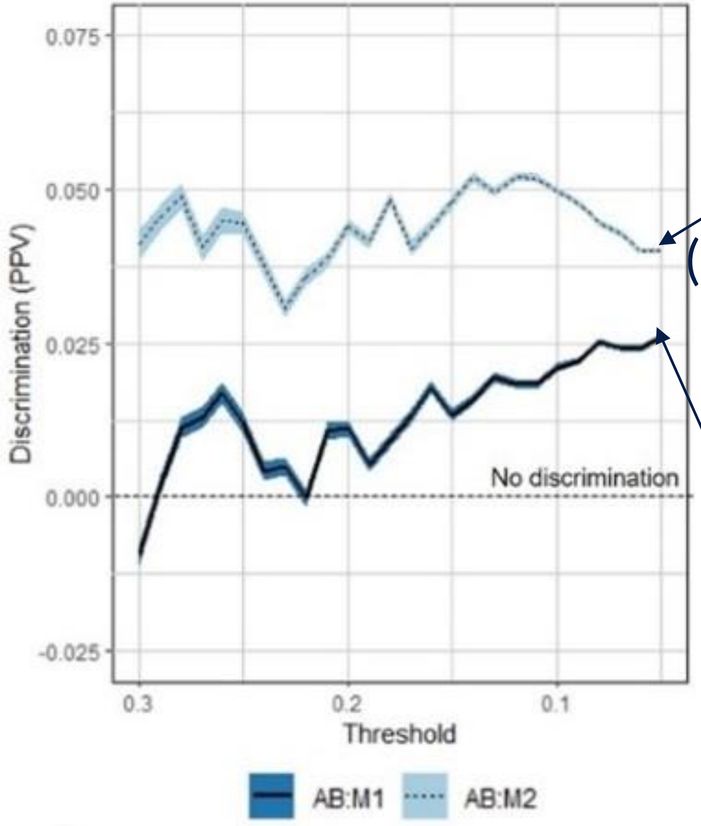
Firm Profitability

- Avg. optimal profit across a range of thresholds
- Across 2,431 Cost:Rev. pairs, 30 folds
- Profit = (Rev * $TN\tau$) – (Cost * $FN\tau$)

Explanations

- ML explainability techniques
- Single-class xgboost toy model
- SHAP values
- SHAP interaction values

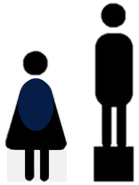
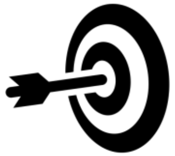
Contribution #1



Model 2
(without gender, Regimes 2 & 3)

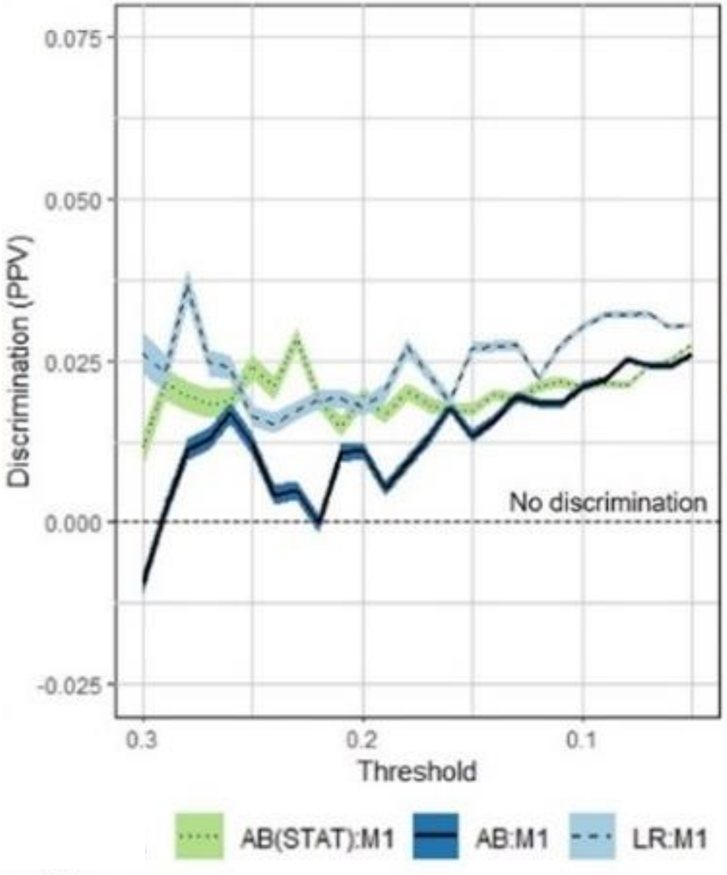
Model 1
(with gender, Regime1)

Impact of excluding gender (ML)

		
Discrimination	Predictive Quality	Firm Profitability
+285.04%	Not stat. sign. different	**-0.25%**

C#1: Prohibiting the use of gender leads to increased discrimination, decreased firm profitability, without significantly impacting accuracy in both ML (and LR models)

Contribution #2

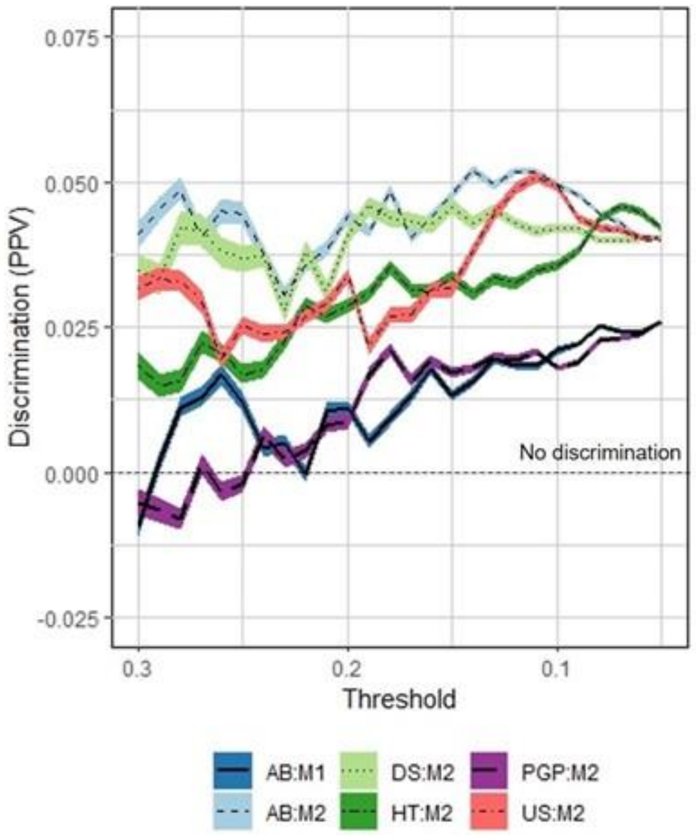


Impact of using ML vs. LR
in models with gender (Regime 1)

Discrimination	Predictive Quality	Firm Profitability
-44.06%	+472 bps	+7.86%

C#2: ML is less discriminatory, more profitable, & more accurate vs LR
(same qualitative findings for models without gender per Regimes 2&3)

Contribution #3



Down-sampling to rebalance (DS:M2)

-4.54% | -175bps | -4.47%

Gender-aware hyperparameter tuning (HT:M2)

-37.73% | -278bps | -4.42%

Up-sampling to rebalance (US:M2)

-24.47% | not sign. | -1.46%

Probabilistic gender proxy modeling (PGP:M2)

-71.09% | not sign. | 0.13%

C#3: Collecting gender and using it in pre-processing & training allows firms to significantly decrease discrimination, whilst managing profitability

Conclusions

Do certain anti-discrimination laws paradoxically increase gender discrimination (in non-mortgage consumer fintech lending)?

- Yes, regimes that force gender exclusion (2 & 3) lead to increased discrimination, decreased predictive quality, & decreased profitability
- ML models help reduce the discrimination & increase profitability vs. LR models, in any Regime

What can firms do to reduce discrimination whilst managing profitability?

- Good: use ML algorithms instead of LR models
- Better: collect & use gender in the training model (4 different pre-processing techniques)
- Best: collect & use gender in both the training and screening models

Thank you

Stephanie Kelley
stephanie.kelley@queensu.ca

full paper available at:
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3719577