

Towards a Non-Discriminatory Algorithm in Selected Data*

David Arnold[†] Will Dobbie[‡] Peter Hull[§]

March 2021

Abstract

There is a growing concern that algorithmic decision-making can lead to discrimination against legally protected groups, but measuring and reducing such algorithmic discrimination is often challenging because an individual's latent qualification for treatment is often only selectively observed. We develop new quasi-experimental tools to overcome this fundamental selection challenge and both measure and reduce algorithmic discrimination in the setting of pretrial bail decisions. We first show that algorithmic discrimination can be measured and reduced using a small number of moments involving individual qualification. We then show how these moments can be estimated by extrapolating quasi-experimental variation across as-good-as-randomly assigned decision-makers. Estimates from New York City show that both a sophisticated machine learning algorithm and a simple regression-based algorithm discriminate against Black defendants, even though defendant race and ethnicity are not included in the training data. Preliminary analyses suggest algorithmic discrimination can be dramatically reduced or eliminated using our methods, with little loss of predictive accuracy.

*We thank Ashesh Rambachan and Jonathan Roth for insightful comments. Jeremy Albright, Jerray Chang, and Alexia Olaizola provided excellent research assistance. The data we analyze are provided by the New York State Division of Criminal Justice Services (DCJS), and the Office of Court Administration (OCA). The opinions, findings, and conclusions expressed in this paper are those of the authors and not those of DCJS or OCA. Neither New York State, DCJS or OCA assumes liability for its contents or use thereof. *JEL Codes*: C26, J15, K42.

[†]University of California, San Diego. Email: daarnold@ucsd.edu

[‡]Harvard Kennedy School and NBER. Email: will_dobbie@hks.harvard.edu

[§]University of Chicago and NBER. Email: hull@uchicago.edu