Which wage distributions are consistent with statistical discrimination?

RAHUL DEB, University of Toronto, Canada
LUDOVIC RENOU, Queen Mary University of London, U.K.

In this paper, we propose a general non-parametric model of statistical discrimination in the labor market, and derive a test for statistical discrimination that only requires cross-sectional data on wages. There are two groups whose productivity distributions have identical means, but can otherwise be different. The group identity is observable to employers, but productivities are not. Instead, there are group-dependent statistical experiments that generate signals about the underlying productivity. Signals induce posterior productivity distributions (via Bayes' rule) and, in particular, these can be used to compute posterior estimates (the mean of the productivity conditional on the signal) of the unobserved productivity. Therefore, each group's statistical experiment generates a distribution over posterior productivity estimates. Wages are then determined via a strictly increasing, continuous function of the posterior productivity estimate that, importantly, does not depend on the group. We say that two wage distributions – one for each of the two groups – are consistent with statistical discrimination if they can be rationalized by this model.

We show that two wage distributions are consistent with statistical discrimination if, and only if, neither wage distribution *first-order stochastically dominates* the other. In addition, we show that a rejection of this test can only be attributed to taste-based discrimination (bias for short). In other words, whenever one wage distribution first-order stochastically dominates another, the only explanation is that employers are biased. Our test exploits the information contained in the entire wage distributions, and not just their averages. This is in sharp contrast with the common practice of reporting wage gaps (differences in average wages), which have been, and continue to be, the subject of much public debate.

We generalize our main result to accommodate for different mean productivities. Here, we assume that the null hypothesis is a joint statement that the wage distributions are generated by statistical discrimination alone and the difference in mean productivities is less than an exogenously given bound. Once again, the test takes a simple form: the null is rejected if, and only if, one wage distribution first-order stochastically dominates the other *and* the wage gap is greater than the bound. This shows that the wage gap can be a useful statistic to uncover bias but only when wage distributions are ordered by first-order stochastic dominance.

While we frame our model in the context of the labor market, it can be applied directly or adapted to analyze other contexts such as housing and financial markets, policing or the criminal justice system. Consequently, we view the reduced form nonparametric framework we propose to be one of our main conceptual innovations. We demonstrate this flexibility by adapting our framework to settings with richer data where Becker outcome tests are typically employed.

A full version of this paper can be found at https://cepr.org/publications/dp17676.

ACM Reference Format:
Rahul Deb and Ludovic Renou. 2023. Which wage distributions are consistent with statistical discrimination?. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC '23), July 9–12, 2023, London, United Kingdom.* ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/3580507.3597667

EC '23, July 9–12, 2023, London, United Kingdom

ACM ISBN 979-8-4007-0104-7/23/07…$15.00
https://doi.org/10.1145/3580507.3597667