A Discrimination Report Card

Based on BFI Working Paper No. 2024-40, “A Discrimination Report Card,” by Patrick Kline, University of California, Berkeley; Evan K. Rose, University of Chicago; and Christopher R. Walters, University of California, Berkeley

A new statistical methodology to grade the race and gender callback gaps of large US employers shows that firms assigned the worst grade are estimated to favor white applicants over Black applicants by 24%, while those assigned the best grade favor white applicants by only 3%. Gender discrimination is rare at the interview stage and concentrated in certain industries.

Twenty years ago, Chicago Booth economists Marianne Bertrand and Sendhil Mullainathan published a seminal paper that studied racial discrimination in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. They revealed that equivalent resumes with distinctively white names like Emily and Greg received 50% more callbacks for interviews than those with distinctively Black names like Lakisha and Jamal.

In 2021, Chicago economist Evan Rose, along with Patrick Kline and Christopher Walters, expanded Bertrand’s and Mullainathan’s work to a massive scale. Their experiment, detailed in “Systemic Discrimination Among Large U.S. Employers,” measures the callback rates from over 83,000 fictitious job applications sent to 11,000 entry level job openings at more than 100 Fortune 500 Firms. The research revealed a surprising fact: A small number of companies are responsible for a substantial amount of the contact discrimination measured.

Who are these firms? In this paper, the authors use the results from their experiment to construct a report card that grades the level of discrimination shown in the experiment by each company. Although the authors sent many fictional resumes to each firm, their experimental measures provide only noisy estimates of each firm’s contact discrimination. Because the authors

BFI Blackboard

**Noise:** Variability that arises from unmeasured or unpredictable factors, rather than from the systematic factors that one is seeking to measure (in this case, discrimination)
Although the authors sent many fictional resumes to each firm, their experimental measures provide only noisy estimates of each firm’s contact discrimination. Because the authors sampled only a subset of each firm’s jobs, it is possible that some firms that showed large contact penalties for Black applicants due to chance rather than true systemic patterns. Their approach utilizes a new statistical methodology, developed by the authors in their [working paper], that ensures the grades provide the most information about firm-level contact discrimination while limiting the potential for random chance to lead to mistaken conclusions.

They find the following:

• Patterns of racial discrimination vary substantially across employers, with some discriminating heavily against Black applicants, and others only slightly or not at all. In the author’s baseline results, firms are assigned to three grades. Firms in the grade exhibiting the most discrimination are 24% less likely, on average, to call back a Black applicant compared to a white applicant, while the least discriminatory firms are 3% less likely.

• The three assigned grades account for much of the between-firm differences in discrimination measured in the study, while providing highly accurate guides to which firms discriminate more than others. The authors show that on average, these grades provide misleading information less than 4% of the time.

• More than half of the variation in racial discrimination across firms is explained by their industry. The industries that exhibit the most racial discrimination are car and car parts retailers, such as AutoNation and Napa Auto Parts, and other retailers, such as
as CVS. Building industry information into their procedure yields an even more powerful report card featuring four unique grades.

- Companies also vary enormously in their treatment of applicant gender. Four firms are assigned grades indicating a strong preference for male names and four are assigned a grade signaling a strong preference for female names, while the remaining firms are assigned grades that indicate negligible preferences for gender.

- These grades explain nearly half of the between-firm variation in gender discrimination measured in the study while remaining very reliable. The authors estimate that on average, comparisons implied by the grades would be misleading in less than 2% of cases.

- Industry affiliation explains about half of the variation in gender discrimination across firms, with firms in the apparel sector such as Anne Taylor exhibiting a very strong preference for female names. Incorporating industry into the methodology produces a report card with five distinct grades.

It is unclear what sorts of reforms and organizational practices will improve the fairness and efficiency of corporate recruiting efforts. The authors hope that releasing their firm-level data for use by other researchers will accelerate the pace of research into strategies for mitigating hiring discrimination. It is also possible that the reactions of employees, customers, and leaders of these organizations to the provision of such information will be effective at reducing discrimination. The authors caution, however, that their estimates should not be construed as making a legal assessment that companies in their experiment violated anti-discrimination laws.