

What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making

Peter Hull, Assistant Professor, UChicago's Kenneth C. Griffin Department of Economics

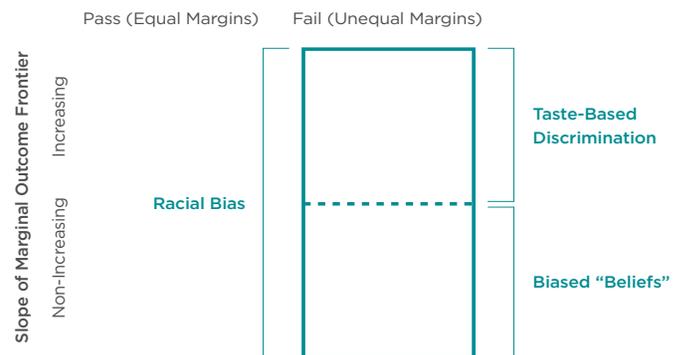
This new work shows how modern empirical tests can detect different forms of bias, from canonical taste-based discrimination to inaccurate beliefs or stereotypes, and offers a new approach to distinguish between the two.

Ever since Gary Becker's path-breaking 1957 work on discrimination, when he introduced the profession to a simple framework for racial bias and its effect on the outcomes of white and Black individuals, economists have built a variety of theoretical models that try to explain the existence of discrimination. In recent years, some researchers have taken a more empirical view of the matter and parsed rich administrative data to find evidence for discrimination in different settings. It is sometimes unclear, however, how this recent empirical literature relates to the classic theoretical framework of Becker and others.

In this new work, Peter Hull of UChicago's Kenneth C. Griffin Dept. of Economics offers a reconciliation of these two literatures, developing a framework for understanding modern tests of decision-making in terms of racial bias. In doing so, Hull shows how modern empirical tests can detect different forms of bias, from canonical taste-based discrimination to inaccurate beliefs or stereotypes, and offers a new approach to distinguish between the two.

Imagine a judge who must decide which defendants to release on bail before trial, with defendants assigned effectively at random to different judges. A recent empirical literature uses such variation to compare the criminal misconduct outcomes of white and Black defendants who a judge is just indifferent to releasing. Inspired by the theory of Gary Becker, racial disparities "at the margin" of treatment may suggest "taste-based discrimination," in which judges hold Black defendants to a different standard than perceivably equal white defendants. But more recent theory may suggest other explanations, such as that the judge is acting on biased beliefs about a defendant's potential for criminal misconduct, or racial stereotypes.

Figure 1 • Result of the "Marginal Outcome Test"



Notes: The change in unemployment is depicted relative to April 2008, when the level of unemployment in the simulation is 5% (as in the data). This figure demonstrates that a counterfactual economy without the observed UI extensions sees a larger rise in unemployment. At its peak, the unemployment rate would have been 0.5 percentage points higher were it not for the benefit extensions. Moreover, at no point is the unemployment rate ever higher with benefit extensions. Thus, the model implies that the benefit extensions were not a cause of unemployment persistence during the Great Recession.

It is theoretically possible, in other words, that a judge with different "marginal outcomes" for white and Black defendants harbors no racial animus, but makes systematic decision-making mistakes that favor white defendants. In practices, judges may base their decisions on inaccurate predictions of defendant misconduct risk after reviewing facts about the defendant's background and prior criminal behavior and other factors. Are these "bad guesses" necessarily evidence of racial bias?

Hull finds that the answer to that question is "No." Differences in decision-making at the margin can reject the possibility that a judge is basing decisions on accurate predictions of misconduct risk in a risk-neutral way. But this does not mean the judge is engaged

in canonical taste-based discrimination. Instead, this finding from the “marginal outcome tests” in the recent empirical literature could be attributed to a judge’s biased beliefs or, more prosaically, their systematic mistakes in predicting whether individual defendants of different races will commit pre-trial crimes.

Hull then offers a new test to disentangle taste-based discrimination from mistaken judgment. This test relies not on the outcomes of white and Black defendants just at the margin of a judge’s decision, but how these marginal outcomes change as a judge becomes more or less lenient. Concretely, imagine that our judge has some sort of internal prediction of pretrial misconduct that she uses to rank white and Black individuals by her desire to release them before trial. If a defendant falls below some potentially race-specific threshold, the defendant is released before trial, while defendants with high misconduct predictions are detained. Currently, researchers look at the outcomes of individuals at these thresholds to determine whether or not a judge is racially biased.

Hull’s insight is to also consider how the misconduct outcomes change as that threshold point moves. In other words, are the judge’s bail decisions resulting in fewer or more crimes at the margin as she releases more or fewer defendants? Hull shows that if marginal outcomes always increase, by race, as more defendants are released then one cannot reject Becker’s classic model of taste-based discrimination. If, however, marginal outcomes do not increase with release rates then it is likely she is just making mistakes.

Importantly, Hull stresses that data can only reveal so much about a person’s intentions. Any conclusions that his test reveals about taste-based discrimination and biased beliefs in a judge’s pre-trial bail decisions, for example, reflect what can be said about the judge’s behavior from her actions, and not necessarily her “true” or intended behavior. The paper nevertheless argues that the results of these empirical tests can be very useful for policymaking.

This work unites classic theoretical framework of racial bias and recent empirical research in several settings, both within and outside of the pretrial setting and criminal justice as a whole. On one hand, Hull shows that existing marginal outcome tests have limits in detecting the canonical taste-based discrimination model of Gary Becker. On the other hand, he shows that a new test which more fully characterizes marginal outcomes can provide a more complete view of racial bias. The paper discusses how both tests can be applied in various settings, and summarizes directions for future empirical work.