Testing the Tests for Racism

Wilfred Reilly

Against the claim of decreased American racism over the past twenty years have come the audit studies. Throughout much of the modern era, a large number of empirically-minded social scientists have pointed out that racism seems by any objective standard to be declining. However, other scholars argue that anonymous tests show considerable modern-era bias against blacks and other racial minorities. How can both of these results co-exist, across dozens of well-designed studies? A review of the audit studies might help.

The fact that racism has declined in the United States in the modern era, following the 1964 passage of the Civil Rights Act, would seem objectively obvious to most people. Eric Kaufmann points out that approval of black/white inter-racial marriage—often presented in the context of a close relative marrying a black person—rose from 4 percent among whites in 1958 to 84 percent in 2013. In 2017, “fewer than 10 percent of whites” surveyed by Pew agreed with a somewhat differently worded question asking whether inter-racial marriage was “a bad thing.” No doubt as a direct consequence of such changing attitudes, the percentage of newlyweds in multi-race marriages rose from 3 percent in the mid-1960s to 17 percent today. Most other anonymous polls focused on key...
issues of race find results similar to those dealing with marriage, with Gallup finding in 2015 that only 8 percent of whites would not vote for a qualified black same-party candidate for President—as vs. 7 percent who would not vote for a Catholic, 9 percent who would not vote for a Hispanic or Jew, and 19 percent who would not vote for a Mormon.5

An especially interesting sub-category of research on the prevalence and effects of modern racism has focused on examining what happens to the performance gaps between groups (in incomes, SAT scores, etc.) that are often attributed to racism—or to genetic factors by the fringe right—when adjustments are made for non-racial characteristics which vary between groups. The economist June O’Neill (1990; 2005), for example, has pointed out that gaps in income between blacks and whites vanish almost entirely following adjustments for such obvious traits as years of education, median age, region of residence, and any aptitude test score. To quote the earlier of her two papers on this subject at some length: “Overall, black men earn 82.9 percent of the white wage. Adjusting for black-white differences in geographic region, schooling, and age raises the ratio to 87.7 percent; adding differences in (standardized) test scores raises the ratio to 95.5 percent, and adding differences in years of work experience raises the ratio to 99.1 percent.”6

More controversially, but with equal empirical support, multiple scholars have noted that adjustment for non-raced variables has the same effect even on “racially disproportionate” rates of black encounters with police.7 At the most basic level, simply adjusting for a black crime rate—and thus black police encounter rate—that is often more than twice the white rate completely eliminates the disparity between the 13 percent representation of blacks within the U.S. population and the roughly 25 percent representation of identified blacks

Testing the Tests for Racism

among those shot and killed by the police.8 At a higher level of analysis, running a fairly comprehensive regression model including multiple victim and officer characteristics reveals that whites are 27.4 percent more likely to be involved in police shootings than are similarly situated black Americans.9 Racism, at least of the contemporary variety, would seem to have essentially no explanatory role regarding findings like this.

However, the positive picture presented so far is complicated by the fact that another field of research—that is, the audit studies conducted primarily by sociologists, seem to virtually always find evidence of rather intense discrimination.10 Perhaps most famously, Devah Pager used teams of “attractive” male college students to test labor market discrimination in Milwaukee in 2001, and discovered that black applicants were a bit less than half as likely as white applicants to be called back for jobs—14 percent versus more than 30 percent.11 Disturbingly, black applicants who did not present themselves as having a criminal record were less likely to be hired than white applicants who did “admit” to having one: 14 percent vs. 17 percent.

Not long after, Bertrand and Mullainathan (2004) conducted a very similar large-N study, using hard copy paper resumes and looking at the effect of “black-sounding” names such as “Lakisha” and “Jamal” on employer interest. Their results were depressingly similar to Pager’s: black-sounding names received only about two-thirds as many callbacks as “white-sounding” names like “Emily.” Literally dozens of similar studies have taken place since these now rather classic pieces were written, with 72 of them (81 percent of the total pool of 89) conducted since 2010, and most prominent published examples of the genre finding discrimination effects at least at the 7-9 percent level.12

So, multiple well-designed regression analyses show that minorities and whites often perform about equally when race-neutral factor variables are equalized, and empiricists have in fact argued for some time that this is the

---

case. However, multiple significant audit studies appear to demonstrate that people of color confront intense bias at every level of the job market—and for that matter the housing market. Simply asked: how can both facts be true at once? While the simple, and greatly under-studied reality of plain human adaptation to moderate levels of discrimination must play some role (“Just apply for two more jobs online, Jamal”), part of the answer almost surely rests with those weaknesses and omissions which do exist within the audit literature.

Setting aside several secondary questions and personal hobby-horses, one of the most obvious issues with audit studies as a class is that most or very many of them seem to look at virtually the same sector of the job market—which could be summed up as the zone of entry-level private-sector jobs, primarily with white-owned employers. For example, Pager’s team of auditors applied entirely to jobs drawn from the “Sunday classifieds section” of the Milwaukee daily newspaper, and from the Jobnet website, and further focused only on “entry level positions requiring no previous experience and no education beyond high school.” Bertrand and Mullainathan’s team responded with fictional resumes “to help-wanted ads in Boston and Chicago newspapers.” In at least the first case, the ten most common categories of job auditors ended up applying for included cashier, delivery driver, laborer, warehouse worker, cook, “waitstaff,” and general service.

Since these two papers have been cited a combined 8,584 times as per a Google search (on 05/12/2021), it is hardly surprising that essentially the same methodology—if in more cyber-friendly form—remains prevalent today, with social scientists being heavily reliant on large general-use portals like Mechanical Turk (Horne et al 2015; Gaddis 2017) and the major jobs databases for large-N survey work in this sector and many others. A fairly typical contemporary audit study design involved the wholly online submission of “9,000 fictitious resumes” to job postings in the specific occupational categories of “administrative assist(ant), customer service, information technology, 

15 One of these “secondary questions” does merit a footnote: what percentage of people in each significant audit study got a job, or would have gotten one had the application process taken place in earnest? This is almost never discussed, but obviously a point of some importance. In Pager’s Milwaukee study, for example, a 14 percent final call-back rate would seem to imply that a black teen would need to apply to only ~7 jobs to find himself hired. It may well be that a fair amount of bias exists in some sectors of the job market, but that this has very little impact on rates of actual hiring for dedicated applicants.
Testing the Tests for Racism

medical assisting (excluding nursing), medical . . . billing, and sales". All applicant resumes submitted during this project included a high school degree, but none included any college achievement beyond “some coursework at a two-year institution.”

There is nothing wrong with focusing research on the entry-level private-sector job market, within which many Americans work. However, a heterodox observer might notice this could well be the only large sector of the U.S. economy where a qualified middle-class minority applicant would face significant discrimination in 2021—and that concentrating studies here could well be one reason for the consistent discovery of anti-POC bias. In contrast to employers like those audited by Pager, colleges and universities—to take one example—engage in very large-scale programs of pro-minority affirmative action, which give potential students and job applicants of color a significant advantage over equally qualified whites.

The same is often true for public-sector employers such as police departments, fire departments, and postal branch offices. While such studies would be more time-consuming than an audit of service jobs posted to Craigslist, it would be fascinating to see the patterns of hiring or admissions bias revealed by well-done audits of these sectors of the economy—or for that matter of jobs advertised specifically as affirmative action or diversity-forward positions, of experience-based union jobs, and the like. I strongly encourage scholars to conduct such research.

For that matter, most targets of audit work to date look awfully pale to me. Often forgotten, during what sometimes seems to be the endless conversation in social science about white discrimination against POC, is the fact that the United States is now just 60.1 percent non-Hispanic white, and minorities seem no less prone to the universal human sin of bias than whites. The researcher Zach Goldberg recently measured levels of in-group preference among various U.S. populations, and found that, while most Americans “like” one another well enough, black Americans preferred their racial group over alternatives by 15.58

---

points, as versus 13.94 points for Asian Americans, 12.83 points for Hispanics, and just 11.62 points for white conservatives.\textsuperscript{19}

Given this context, it is truly surprising that there seem to have been no studies to date focused specifically on auditing the 18.3 percent of businesses owned by members of racial minority groups—or indeed the 38.2 percent of them owned by women and minorities combined.\textsuperscript{20} I strongly encourage social scientists to conduct this work as well, and will be fascinated to see if patterns of pro-minority hiring preference exist, and if so, to what extent. At least some recent empirical work may provide a clue: Kerr and Kerr examine co-ethnic hiring among a heavily-POC subset of new American business ventures, and find that a preference for hiring members of the same ethnic group is “ubiquitous” among immigrants.\textsuperscript{21}

The rate of same-group hiring the authors document—often among fairly small groups providing a limited number of potential employees—averaged 22.5 percent and ranged as high as 40 percent. Whether similar patterns obtain among the owners of the thousands of successful minority businesses listed on directories like Black Business List (https://www.blackbusinesslist.com/), Black Directory (https://blackdirectory.com/), the Virginia Black Business Directory (https://www.vabusinessdirectory.com/), and the Asian American Chamber of Commerce (https://business.asian-americanchamber.org/list/) is a question which remains to be empirically answered.

A second potential explanation for the apparent discrepancy between the results of regression tests attempting to find an accurate residual for discrimination and the results of audit studies is that audit scholars face difficulties when attempting to adjust for social class and other non-raced variables. For at least the past decade and a half, the most common indicator used to signal the “blackness” of an applicant during an audit study has been the use of a name which presumably sounds black. However, a problem with this strategy—noted by several of the authors whose work has been discussed here as well as Roland Fryer and Steven Levitt—is that mean levels of income and of socio-economic status (SES) vary widely among racial groups. Very black names (an example might be “Jamarrian”) are obviously often associated not simply with

minority race but also with “lower income zip codes . . . lower levels of parental education.”

This intertwining of race and class is legitimately difficult to adjust for, and in fact almost never is adjusted for. When skillful adjustments are made to deal with the problem, the results can be striking. Rajeev Darolia and Cory Kodel, in their studies cited above, for example, avoid the risk of class-signaling via parent-assigned first names by using ethnically distinct last names (“Washington,” “Jefferson,” “Garcia,” “Hernandez”) paired with relatively neutral first names like “Chloe” and “Carlos” on resumes. They find almost no effect of race on employer hiring preference.

This paper has faced some attacks based on the potentially confusing nature of the first names used, but the authors stand by their data and point out that roughly 60 percent of all potential employers would have had to mistakenly believe their applicants were white in order for their primary results to be invalid. While Raj Goshal provides a critique of the black names chosen for this project, and argues that essentially this was the case among one sample of respondents, Darolia and Kodel’s conclusions, at least pertaining to discrimination against Hispanics, seem likely to stand for some time—as it is difficult to imagine a sizable number of canny hiring managers mistakenly believing a “Carlos Hernandez” to be of Norwegian descent.

Somewhat similarly, in his 2017 study Michael Gaddis finds that higher-status black names are less likely to be perceived as black at all. To quote: “Respondents are much more likely to (correctly) perceive a black name from mothers with lower education levels like DaShawn, DaQuan, or Lakisha rather than from mothers with higher education levels like Nia, Malcolm, or Malia.”

An entire bloc of common and often middle-class black names analyzed by Gaddis, including “Bria, Sade, Kaylah, Lyric, and Jasmine,” were perceived as black by under 2/3 of study participants engaged using Mechanical Turk—and generally by far fewer respondents than that without an ethnically-signaling last name added. These very common black names would thus, logically, result at the very least in decreased discrimination even from a biased employer. The same, obviously, would be true for almost any of the upper-middle class black

---

names I see daily on the masthead of my own 1776 Unites organization: Bob, Wilfred, Coleman, John, Shelby, Clarence, and so forth.

It would also be interesting to see how “Billy Bob” might fare. While this was by no means his main focus, one of Gaddis’s most interesting points is that the class associations of names exist for whites as well as blacks. He notes at one point that: “Respondents are also much more likely to correctly perceive a white name from mothers with higher education levels like Claire, Jake, or Abigail rather than from mothers with lower education levels like Cheyanne, Irvin, or Jordy.” The names most often perceived as white by Gaddis’s respondents were Katelyn, Hunter, Claire, Jake, and Seth, while those most often perceived as black were DaShawn, Tanisha, Treymayne, Jamal, and Daquan—and at least one or two “poor white” names (“Cheyanne”) were often not perceived as white at all, and could theoretically lead to a holder’s facing discrimination on class and race grounds.

In this now-known context, it would seem critical for auditors attempting to measure race effects, with class adjusted for, to use specifically those names (1) known to be associated with black holders and also (2) verified by respondents as being associated with middle-class status. Further, and in all seriousness, it would be very interesting to see how resumes bearing stereotypically lower-income white first names like “Earline” and “Chevy” do across the New York City, Chicago, and other major urban job markets. Another worthwhile and apparently untired experiment would involve simply submitting resumes bearing a minority-associated last name and a first initial rather than a first name. Something almost never discussed in the extensive literature dealing with contemporary race and class bias is how easy much of it would logically be to avoid: does simply applying for work as “Mr. J. Washington” rather than “Jamarrian Washington” do the trick?

Another tricky and complex issue confronting the audit genre is how to adjust for employer/renter perception of the probable qualifications of applicants—the “unobservables” mentioned by David Neumark and Judith Rich—which often vary significantly at the mean for members of different groups. While it might at first glance seem unbelievably racist for an employer to be (say) one-quarter less likely to hire a black than a white graduate from the same college, or to be as likely to hire a white state college graduate as a black Yale man (Gaddis 2015), an obvious alternative explanation presents itself almost immediately—employers are aware of affirmative action, and of the fact that
a black Ivy Leaguer and white Nittany Lion might well have virtually identical SAT or GRE qualifications.

This is not at all an exaggeration: affirmative action effects have consisted of more than “just a thumb on the scale” for decades now, and are objectively very large. According to 2013-2016 American Association of Medical Colleges data published by the American Enterprise Institute, the medical school acceptance rate for prospective doctors with an MCAT of 23-26 was 6 percent for Asian Americans, 8 percent for whites, 31 percent for Hispanics, and 56 percent for blacks. At the higher level of 30-32 (paired with a GPA of 3.6-3.79 in the data set), the same rates were respectively 58 percent, 63 percent, 83 percent, and 94 percent.25 This sort of thing is not confined to the graduate and professional schools. Peter Arcidiacono, attorney for the plaintiffs during the recent lawsuit brought against Harvard by rejected Asian-American applicants, notes that 2010-15 acceptance rates for students in the top decile of their academic class were a bit over 10 percent for Asians, under 20 percent for whites, around 35 percent for Hispanics, and close to 60 percent for blacks.

A quick glance at any recent year’s SAT data indicates why this is the case on the campus of any research or comprehensive university which desires to “look like America.” As Inside Higher Education noted frankly in the header for a piece titled “New SAT, Old Gaps by Race,” past-decade SAT scores continue to “show large gaps by race and ethnicity.26” In the year in question, 2017, the mean combined math/verbal SAT scores were 941 for blacks, 963 for Native Americans, 987 for Hispanics, 1,118 for whites, and 1,181 for Asians. All of these scores are solid enough, and up from the “800 and perhaps 1,000” which seemed de rigueur for minority and white public high schoolers when I attended East Aurora Senior in the late 1990s. However, the total black-white gap is 177 points, there are only 1,600 points possible on the test, and employers are aware such gaps exist. It would be—to use this language again—very interesting to see what effect adding an entirely objective quantitative metric representing aptitude, such as a GRE score, to resumes would have on employer behavior. I strongly encourage auditors to do so during future studies.


Testing the Tests for Racism

A final important point on the audit literature is not so much whether bias against blacks or Latinos is being measured correctly, but instead how this prejudice measures up to that experienced by members of other populations. As noted early on, an often cited fact is that a well-designed totally anonymous poll—although not an audit study—found an 8 percent rate of voter bias against blacks. However, the same poll found nine percent rates of bias against Hispanics and Jews, and a 19 percent rate of bias against (of all groups) Mormons. The minority of audit studies which have ventured beyond traditional white-black comparisons seem to reveal similar patterns of discrimination. Major projects have found that applicants with Asian last names are 28 percent less likely to receive employer call-backs than white applicants; that stated Islamic faith correlates with 32 percent fewer employer e-mails and 48 percent fewer call-backs; and even that women with “feminine-sounding” names are less likely to be hired as attorneys.27

These esoteric varieties of bigotry, of course, are no more morally acceptable than any other bias. But, to some extent, that’s the point. While they doubtless sometimes encounter prejudice, Americans of East Asian and Middle Eastern descent are arguably the most successful people in the country, and female lawyers don’t do too shabbily. If a similar baseline level of prejudice exists against 89 percent of the country—a rough estimate of the population percentage composed of racial minorities and women, with religious minorities not counted—we may have again bumped up against one of the most important and least discussed realities in social science: the simple existence of bias does not mean that said bias explains all observed discrepancies in performance between groups. As a final suggestion, I strongly encourage auditors to test this “many biases” theory, by including photos with resumes where appropriate, to measure the effect of physical fitness and perceived attractiveness, by looking more at prejudice against gay applicants in the tradition of Drydakis (2014), and by conducting social class-focused studies like the test described above utilizing poor white names.

The audit studies are an interesting sub-genre of research in a discipline methodologically close to my own. Some of their results obviously do indicate

---

that bias remains a reality within significant sectors of the U.S. employment and housing markets. However, these studies rarely if ever examine rates of pro-white (or pro-POC) bias in higher education, the public sector, and the minority business community; very frequently do not include adjustments for social class or perceived competence; and have not extensively compared the bias faced by members of other potentially disadvantaged groups with that faced by blacks. Given this, audit data does not seem to counter the basic observation that citizens of different races with the same background characteristics often perform similarly in life. I certainly do hope audit scholars tackle the points made above going forward, perhaps proving this point wrong—and perhaps proving it right.