Replication Research:

Are Emily and Greg More Employable Than Lakisha and Jamal?

A Field Experiment on Labor Market Discrimination

Teresa Nguyen, Jennette Hoang, Latherine Vo

Purpose

Society has long observed persistent racial inequality in the labor market, with worse labor market outcomes for African American workers than the White workers. To examine whether differential treatment contributes to these disparities, the study "*Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination"* by Marianne Bertrand and Sendhil Mullainathan investigates whether employers treat members of different races with equally credentialed backgrounds differently during the hiring process.

Understanding this distinction is important for determining how much it arises from human capital differences versus labor market discrimination. Traditional surveys or interview-based research are subject to measurement bias, and employers rarely admit discriminatory behavior. A field experiment therefore provides an opportunity to isolate discrimination at the earliest stage of hiring, the resume screening process, where employers form first impressions and make early decisions that shape access to employment and upward mobility. The potential policy implications are substantial. If employers discriminate at the resume stage, then improving worker skills alone will not eliminate racial inequality in labor market outcomes. The findings speak to ongoing discussions about Equal Employment Opportunity enforcement, affirmative action, blind hiring practices, and public investments intended to reduce racial gaps in income and employment. The results also inform debates in economics regarding statistical versus taste-based discrimination and the persistence of unequal outcomes even in competitive labor markets.

The primary outcome variable was whether the employer called back the applicant for an interview. The authors estimated differences in callback rates between name groups using statistical tests and linear regression models. Although some of the econometric techniques extend beyond this course,

the core logic is clear: the analysis quantifies whether applicants with African American–sounding names are significantly less likely to receive a callback than equivalent applicants with White-sounding names, holding all resume characteristics constant.

Statistically, the authors compare callback rates between groups and estimate probit models relating callback outcomes to race indicators and resume characteristics. They also explore heterogeneity across cities, industries, occupations, and applicant neighborhoods, as well as interactions between resume quality and racial perception.

The study's central finding is that resumes assigned White-sounding names received dramatically higher callback rates than those with African American - sounding names. In the full sample, applicants with White-sounding names received callbacks 9.65% of the time, compared with only 6.45% for those with African American-sounding names - a 50 percent gap. This difference is both statistically significant and economically meaningful: having a White-sounding name generated the same improvement in employer response as adding approximately eight additional years of work experience to a resume. These results provide direct causal evidence that employers respond differently to observationally identical applicants based solely on perceived race.

Equally striking is the finding that increases in resume quality disproportionately benefit applicants with White-sounding names. High-quality resumes boosted callback rates by roughly 27% for White applicants but by only about 8% for African American applicants, revealing that African American candidates do not receive the same returns to skill investments in the hiring process. Rather than narrowing the racial gap, stronger credentials actually widen it, contradicting the expectation that improved qualifications should alleviate employers' concerns. Together, these results suggest that racial discrimination remains a significant barrier at the very first stage of the

job search process. The authors also find that the racial gap is remarkably consistent across industries, occupations, employer sizes, and whether employers identified as Equal Opportunity Employers. This uniformity suggests that differential treatment is widespread rather than concentrated in specific sectors.

Overall, Bertrand and Mullainathan conclude that racial discrimination remains a prominent feature of the U.S. labor market's initial screening stage. Employers appear to treat African American-named applicants less favorably despite identical qualifications, contradicting claims that discrimination has diminished or reversed. Their evidence highlights persistent barriers to equal employment opportunity and implies that resume-based screening processes may perpetuate systemic inequality.

This paper seeks to reproduce the callback gap identified in the original Bertrand–Mullainathan study by examining how employer responses vary with the racial connotations of applicant names and the quality of submitted resumes. In addition, the analysis will explore how specific resume attributes influence the likelihood of receiving a callback and assess whether African American and White applicants experience different returns to comparable skills and credentials.

Replication Data and Methods

To replicate the callback-rate results in Bertrand and Mullainathan (2004), we use the instructional dataset Names.dta/Names.xlsx released by Marianne Bertrand, which contains 4,870 resume-level observations from the original 2001 field experiment conducted in Boston and Chicago. Each observation corresponds to a fictitious resume submitted to a real job posting and includes binary indicators for race (White-sounding vs. African American-sounding names) and callback

outcomes, along with applicant, resume, and job-posting characteristics. The sample size and one-resume-per-row structure match those reported in the American Economic Review article, allowing direct replication of the study's headline result.

The dataset records key resume features relevant to the experimental design, including years of experience, number of prior jobs, email presence, computer skills, volunteering, and employer-level equal-opportunity statements. These variables allow us to reproduce the main descriptive statistics and estimate the basic linear probability models emphasized in Section III of the original paper. The core experimental structure, randomized name assignment, resume quality variation, and city-level placement, is fully preserved, making the dataset well suited for replicating the primary callback-rate comparisons by race.

However, the replication dataset differs from the original in several important ways. Employer identifiers and detailed job-ad text are not included, which prevents employer-level heterogeneity analyses and job-ad–specific robustness checks. In addition, several occupation-specific variables used in the original study, such as indicators for females in administrative or sales jobs, and detailed job descriptions are missing. While these omissions limit our ability to replicate certain extensions and robustness exercises, the dataset remains sufficiently detailed to replicate the central empirical finding: resumes with African American–sounding names receive significantly fewer callbacks than those with White-sounding names.

Replication Result

The original Bertrand and Mullainathan (2004) study provides the baseline patterns that the replication aims to reproduce. The key finding is that resumes with White-sounding names

receive significantly more callbacks than those with African American - sounding names, even when qualifications are identical. Callback ratios are consistently around 1.5 across subsamples, indicating that White-sounding names receive about 50% more callbacks.

The resume characteristics in the original experiment were carefully randomized, ensuring that applicants assigned different racial-sounding names were otherwise identical. High-quality resumes were designed to include stronger human capital indicators —such such as more experience, volunteering, email availability, honors, and fewer employment gaps -while low-quality resumes intentionally lacked these features.

In the original regression analysis, improvements in resume characteristics generated positive returns overall, but the gains were substantially larger for White-name applicants. African American name resumes received very small or statistically insignificant returns for most resume enhancements, with special skills being one of the few exceptions.

Building these core patterns from the original experiment, the replication analysis evaluates whether the same relationships appear in the provided dataset. By comparing callback rates, resume characteristics, and returns to credentials across racial groups, the replication assesses how closely the reproduced results align with the original study's findings. This allows us to determine whether the underlying evidence of racial disparities persists under the replicated conditions.

The replication results closely match the main patterns of the original study and consistently reproduce the evidence of racial disparities in hiring.

VARIABLES	Percent callback for White names	Percent callback for African American names	Ratio	Percent difference (p-value)
All sent resumes	9.65	6.45	1.49612403	3.2
	(0.00599)	(0.00498)		
Observations	2,435	2,435		
Chicago	8.06	5.4	1.49259259	2.66
	(0.00741)	(0.00615)		
Observations	1,352	1,352		
Boston	11.6	7.76	1.49484536	3.84
	(0.00975)	(0.00813)		
Observations	1,083	1,083		
Females	9.89	6.63	1.49170437	3.26
	(0.00692)	(0.00573)		
Observations	1,860	1,886		
Males	8.87	5.83	1.52144082	3.04
	(0.0119)	(0.0100)		
Observations	575	549		

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Mean Callback Rates by Racial Soundingness of Names

Through ***Table 1***, resumes with White-sounding names receive 9.65% callbacks compared to 6.45% for African American–sounding names, identical to the published figures. Callback ratios and percent differences align closely with the original results, and the overall pattern of racial gaps holds across the full sample, Chicago, Boston, males, and females. The only differences come from subgroups that cannot be reproduced due to missing occupation identifiers, such as female administrative and sales positions.

The dataset also reflects the original study's experimental structure. Resume characteristics-including education, work experience, volunteering, skills, honors, email availability, and employment gaps—are nearly identical across racial groups, showing that the random assignment of names was successful and uncorrelated with resume content. At the same time, ***Table 2*** clearly

retains the intended differences between high- and low-quality resumes, with stronger human capital indicators concentrated in the high-quality group.

	All resume	White names	African American	Higher quality	Lower quality
yearsexp	7.843	7.856	7.830	8.294	7.387
honors	0.0528	0.0542	0.0513	0.0711	0.0342
volunteer	0.411	0.409	0.414	0.792	0.0272
military	0.0971	0.0924	0.102	0.190	0.00330
empholes	0.448	0.450	0.446	0.340	0.557
workinschool	0.560	0.558	0.561	0.718	0.399
email	0.479	0.479	0.480	0.924	0.0305
computerskills	0.821	0.809	0.832	0.910	0.731
specialskills	0.329	0.330	0.327	0.358	0.300
college	0.720	0.716	0.723	0.724	0.715

Table 2: Resume Characteristics – Mean Proportion

These quality differences translate into notable contrasts in callback behavior, as we see in **F*igure 1***. Higher resume quality increases callbacks overall, but the gains are distributed unevenly across racial groups. White-name resumes experience a noticeable rise in callback rates when quality improves, while African American-name resumes see only minimal changes.

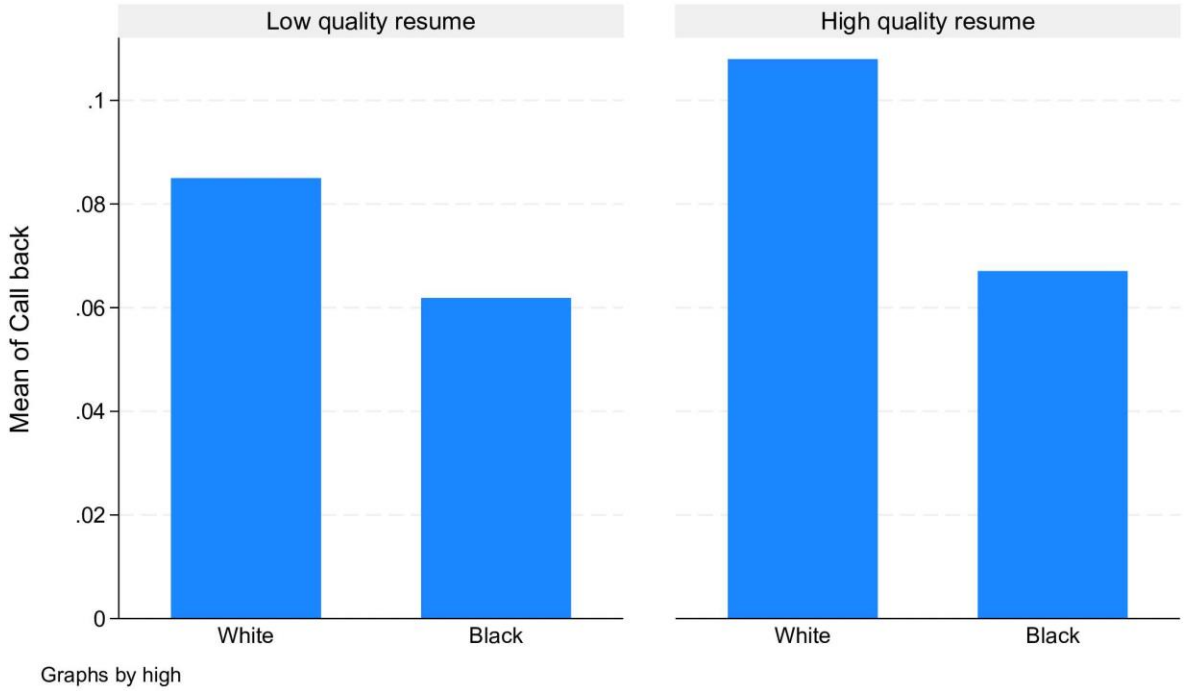

Figure 1. Mean Callback Rates by Race and Resume Quality

This gap reflects broader patterns in how employers respond to specific resume credentials. In the regression results of ***Table 3***, experience has strong positive effects, though the size of the return differs sharply by race. An additional 10 years of experience increases callback probability by roughly 15.4 percentage points for White-sounding names compared to only 2.71 percentage points for African American-sounding names. The negative squared term observed in all groups indicates diminishing returns, consistent with the original study.

Other resume features also show uneven effects. Honors, special skills, and employment holes raise callback probabilities overall, but the boosts are larger and more consistently significant for White-name applicants. African American name resumes receive little to no return from most resume enhancements, with special skills and honors being the only characteristics showing

statistically significant positive effects in the replication. In contrast, several resume features appear to reduce callback rates, with computer skills lowering the likelihood of a callback—most noticeably for White-named applicants (about a 4% decrease) and more modestly for African American–named applicants.

Callback	All resumes	White names	African American names
Years of experience (*10)	0.0818***	0.154***	0.0271
	(0.0269)	(0.0444)	(0.0336)
Years of experience^2 (*100)	-0.0209**	-0.0481***	-0.000713
	(0.0102)	(0.0171)	(0.0130)
Volunteering? (Y =1)	0.000234	-0.0107	0.0133
	(0.0114)	(0.0178)	(0.0140)
Military Experience? (Y=1)	0.00959	0.0390	-0.0119
	(0.0148)	(0.0251)	(0.0162)
E-mail? (Y=1)	0.00987	0.0215	-0.00448
	(0.0113)	(0.0179)	(0.0140)
Employment holes? (Y=1)	0.0278***	0.0384**	0.0176
	(0.00996)	(0.0154)	(0.0124)
Work in school? (Y=1)	0.0150	0.0227	0.00446
	(0.0105)	(0.0159)	(0.0138)
Honors? (Y=1)	0.0583**	0.0673*	0.0459
	(0.0242)	(0.0370)	(0.0302)
Computer skills? (Y=1)	-0.0268**	-0.0429**	-0.00724
	(0.0116)	(0.0169)	(0.0155)
Special skills? (Y=1)	0.0609***	0.0754***	0.0471***
	(0.00942)	(0.0148)	(0.0115)
Constant	0.00670	-0.0157	0.0196
	(0.0174)	(0.0270)	(0.0219)
Observations	4,870	2,435	2,435
R-squared	0.024	0.033	0.017

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, *

Table 3: Return to credentials by race

Because special skills are one of the few characteristics that benefit both groups, though still to different degrees, the comparison in ***Table 4*** highlights the broader asymmetry in employer

responses. While White applicants experience a 7.69% higher rate of callbacks, African American applicants also see an increase in callbacks with 5.15%, but the effect remains smaller than for White applicants, further emphasizing the disparity in employer treatment.

Average marginal effects — Number of obs = 4,870
Model VCE: Robust
Expression: Linear prediction, predict()
dy/dx wrt: 1.specialskills
Over: black

	dy/dx	Delta-method Std.err.	t	P > |t|	[95% conf. interval]	
0.specialsk~s	(base outcome)					
1.specialsk~s						
black						
0	.0768879	.0140539	5.47	0.000	.049336	.1044399
1	.0515027	.0118255	4.36	0.000	.0283193	.074686

Note: dy/dx for factor levels is the discrete change from the base level.

Table 4: Return of special skills on callback rates of different races

Although the replication uses a linear probability model rather than a probit model, the overall relationships remain the same. The effect sizes differ marginally, generally by 1-2 percentage points, but the central conclusions hold: resume improvements produce consistently larger gains for White-name applicants, while African American-name applicants receive weaker or statistically insignificant returns despite having identical credentials.

Potential Extension

A natural extension of Bertrand and Mullainathan's framework would be to investigate how modern hiring technologies influence racial discrimination today. The original study examined human screeners reading paper resumes; however, most firms now use AI-driven applicant tracking systems (ATS), keyword filters, and automated resume parsers. An updated experiment could randomize racially distinctive names *and* resume keyword density to test whether discriminatory gaps persist once algorithms, not people, conduct the first screening. This would clarify whether bias now originates from human judgment, machine learning models, or both.

Another promising extension is to explore intersectionality in discrimination, something the original paper could not fully address due to limited demographic indicators. Race interacting with gender, socioeconomic background (e.g., high school quality, neighborhood characteristics), or immigration status may reveal more complex patterns. For example, do employers treat African American-sounding female names differently from male names? Do ethnic names associated with recent immigrant groups experience similar penalties? Modern labor markets include a wider range of minority identities than those studied in 2001, making this question increasingly relevant.

A third extension concerns changes in employer behavior over time. Labor markets today operate under stronger anti-discrimination enforcement, increased social awareness, and tighter labor shortages in some occupations. Replicating the experiment with 2025-2026 job postings could test whether the magnitude of name-based discrimination has declined, remained stable, or simply shifted to different occupations. The rise of remote work and nationwide hiring may also alter callback patterns, potentially reducing the importance of local racial stereotypes tied to Boston or Chicago.

Additionally, researchers could incorporate richer job-level data, such as the gender composition of occupations, firm diversity policies, recruiter demographics, or industry-specific hiring norms. These factors were largely unobserved in the original dataset but likely shape how discrimination manifests across sectors. Access to detailed job description text, now available through web scraping, would allow for fine-grained analysis of whether certain skills, task requirements, or workplace cultures amplify or mitigate racial bias.

Overall, an updated experimental design leveraging modern data sources, more granular job characteristics, and intersectional identity markers would deepen our understanding of the mechanisms driving callback disparities and reveal whether the patterns documented two decades ago persist in today's hiring landscape.

Bibliography

Bertrand, Marianne, and Sendhil Mullainathan. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." American Economic Review Vol 94, 4 (Sep 2004): 991-1013.