

Connecticut College

## Digital Commons @ Connecticut College

---

Economics Honors Papers

Economics Department

---

2021

### Discrimination in Modern Labor Markets: A Field Experiment

Taekmin Nam

Follow this and additional works at: <https://digitalcommons.conncoll.edu/econhp>



Part of the [Labor Economics Commons](#)

---

This Honors Paper is brought to you for free and open access by the Economics Department at Digital Commons @ Connecticut College. It has been accepted for inclusion in Economics Honors Papers by an authorized administrator of Digital Commons @ Connecticut College. For more information, please contact [bpancier@conncoll.edu](mailto:bpancier@conncoll.edu).

The views expressed in this paper are solely those of the author.

# Discrimination in Modern Labor Markets: A Field Experiment

By Daniel Taekmin Nam<sup>1</sup>

Thesis Advisor: Professor Mark Stelzner

Honors Thesis

Connecticut College (Department of Economics)

---

<sup>1</sup> I would like thank Professor Mark Stelzner who was my advisor for the thesis. I would like to thank him for the motivation he gave me throughout the project, and the enthusiasm he showed. His guidance was the reason I was able to start, and finish this project. I would also like to thank Nikesh Ghimire who was the research assistant to this project. Nikesh played a pivotal part in the implementation of the study, and without his help, this project would not have taken place. Thanks to Professor Purba Mukerji who is the second reader to the thesis, and who's class in Econometrics 2 helped me build the fundamentals to run this study. Thanks to Andrew Lopez who provided great support with data and literature throughout the whole process. Additionally, I would like to thank Dr. Michael Gaddis, Dr. Patrick Button and Dr. Douglas L. Kruse for their guidance on running correspondence studies.

## **Abstract**

With the growing prominence of big data and algorithms in the hiring process, it seems only apt to systematically identify the effect that these changes are having on labor markets. In this paper, I seek to take the first step in filling this void by systematically analyze the effect of big data and algorithms on discrimination in the hiring process. I conduct a correspondence study adapted to the new job application process to see how big data and algorithms are affecting discrimination in the hiring process.

## Introduction

As we enter what many calls the ‘Fourth Industrial Revolution,’<sup>2</sup> big data and the algorithms that manipulate them have unleashed a wave of dramatic changes on the economy and society. The hiring process is one of the many areas affected by these changes. For example, according to a LinkedIn global recruiting trends report in 2018, big data and artificial intelligence are two of the four major factors that are predicted to shape the future of recruitment (Spar et al. 2018). Furthermore, research done by Jobscan, a recruitment solutions company shows that 98.8 percent of the Fortune 500 companies use applicant tracking systems (ATS) – software that collects data from submitted resumes and automatically filters them through algorithms to make hiring decisions (Qu 2021).

Ardent proponents of the transformed recruitment methods point to saved time and reduced discrimination as the two major reasons why these new methods are superior to the traditional ways of recruitment (Spar et al. 2018). Applicant tracking systems have equipped recruiters with the tool to parse through piles of job applications in a very short period of time – providing employers with large cost-savings (Laumer et al. 2014). This has led the applicant tracking software to become a highly touted novel technological adaptation in the recruitment industry. As a result, applicant tracking software industry is forecasted to grow at over 7.9 percent globally in the next three years and a staggering 9.1 percent in the United States over the same period (Markets and Markets 2020).

Despite the sanguinity of some, it is not clear if the use of big data and algorithms really reduce discriminatory practices (especially racial discrimination) in the labor market. Indeed, there is antidotal evidence of algorithms repeating many of the same discriminating patterns of traditional labor markets – those that existed before big data and AI. However, research on the full effect of hiring discrimination from the introduction of algorithms into the hiring process has been sparse. There is next to no systematic research on whether big data and algorithms used in the labor markets discriminate. To some degree, this is a result of the opacity with which hiring algorithms operate.

---

<sup>2</sup> Charles Schwab, founder of the World Economic Forum defines the different stages of Industrial Revolution as follows: First Industrial Revolution used steam power to mechanize production; Second used electricity to enable mass production; Third used electronics and information technology to automate production and the Fourth is a digital revolution that blurs the boundaries between physical, digital and biological spheres (Schwab 2016).

However, identifying discrimination in the labor market is important because macro data has shown that the difference in wages between like workers is substantial and, for some groups, it has actually been growing. For example, in 2015, after controlling for education, experience, and location, black and white women working fulltime only make 84 and 78 percent of their black and white male counterparts, respectively. And, in the same year, after controlling for education, experience, and location, black males working full-time only made 78 percent of equivalent white males – which represents a five-percentage point decrease in relative wage for black males since 1979 (Daly, Hobijn, and Pedtke, 2017).

With the growing prominence of big data and algorithms in the hiring process, it seems only apt to systematically identify the effect that these changes are having on labor markets. In this paper, I seek to take the first step in filling this void by systematically analyze the effect of big data and algorithms on discrimination in the hiring process.

### **Discrimination in Traditional Labor Markets**

Some argue that macro data, like that presented above, on the relative wage of like workers of different race and gender is misleading. The apparent difference between white and black workers is not a result of discrimination. Instead, it is a result of differences in productivity and the lack of adequate macro data to control for other factors that contribute to the productivity of a worker. For example, Neal and Johnson (1996) show that two thirds of the apparent black-white wage gap in 1990 to 1991 could be attributed to difference in cognitive abilities as proxied through scores on the Armed Forces Qualification Test. A flurry of studies in similar realm argued that employment problems of minority groups were primarily the result of variance in skill or other individual deficiencies rather than a result of discrimination (Farkas and Vicknair 1996; O’Neil 1990). Nobel Prize winning economist James Heckman (1998) states this point succinctly: “most of the disparity in earnings between blacks and whites in the labor market of the 1990s is due to differences in skills they bring to the market and not to discrimination within the labor market.”

Because of the lack of data and the assertion of actual differences in productivity, economists and employers sometimes use group membership as a proxy for missing data that is important for determining a worker’s productivity. This is called statistical discrimination (Phelps 1972; Arrow 1973; Aigner and Cain 1977). As an example, an employer that encounters two applicants with similar noisy but unbiased signals of productivity, might favor the applicant from

the majority group if it is assumed that members from the minority group are on average less productive (Bertrand and Duflo 2016).

There are a number of theoretical frameworks which support the idea that wage differences between workers result from differences in productivity. For example, Becker (1971) argues that some employers could have a distaste for hiring members of the minority group. When there is a sufficiently large enough share of employers discriminating in the economy, a difference in wages can arise between the two group of employees with identical productivity. However, employers engaging in discriminatory practices will face reduced profitability from forgoing the decision to hire workers with the same productivity. As a result, market mechanisms will promote the non-discriminating firms. Still, the model also suggests that, refusing to hire minorities could be efficient if the employer already knows that his firm's productivity will decrease because of his unwillingness, or that of other employees, to work with minorities.

In a direct response to the explanations put forward above, a number of social scientists have attempted to better control for productivity through the use of audit studies to see if discrimination still exists. There are two types of audit studies, in-person and correspondence. In-person audit studies entail creating identities for like workers of different race or gender and training actual individuals to play those roles. Trained testers then apply for jobs and attend interviews to see if there is a difference in treatment for like candidates of a different race or gender. A notable in-person audit study, Neumark, Bank, and Van Nort (1995) investigate the role of gender discrimination in hiring waiters and waitresses. In the study, two male and two female college students were sent to apply for waiter and waitress jobs at 65 restaurants in Philadelphia. The results of the study showed that high-price restaurants that usually paid better, discriminated against female applicants while low-price restaurants favored female applicants.

In another in-person audit study, Pager and Wester (2009) analyze the role of racial discrimination in low-wage labor markets. They trained 340 entry level job testers who subsequently applied to entry level jobs in New York City. The testers were trained to act as candidates with a specified level of experience and productivity. Pager and Wester found that positive response rates for Latino and Black applicants were substantially lower than those for equivalent white applicants. Furthermore, they found that there were double standards where employers sought higher qualifications from black applicants than non-black applicants, or viewed whites as more qualified than minorities holding the same credentials.

The second type of audit studies, correspondence studies, entail the creation of resumes for applicants with identical levels of productivity but of different race or gender groups. Resumes are then sent to employers to detect discrimination in response rates from employers. Correspondence audit studies were created in direct response to criticisms about the inability to perfectly train testers in in-person audit studies. Firstly, some argued that it was not possible to control for all the variables and traits portrayed by testers in the application process (for example, appearance, accent and etc.). These differences across auditors might affect outcomes. Secondly, audit studies are not double-blind, meaning, that the testers know the purpose of the study. Therefore, during the application process, the testers may consciously or unconsciously behave in manners that can bias the results towards their own perspective about the subject (Heckman and Siegelman 1993). Thus, correspondence studies eliminated the human element from job applicant side, but in doing so only evaluates discrimination at the initial employer response stage.

Utilization of resumes as opposed to trained testers allows control over more variables such as personal traits. Since employers can only decide based on the information from the resumes, there are no other factors they can use to assume the difference in productivity between workers. Resumes are created by randomizing inputs which minimizes bias between the candidates. Furthermore, automated correspondence studies allow for a bigger sample of data which, reduces the variance and in turn providing more robust results. At the same time, applications can take place in multiple locations at the same time, which would reduce the geographical bias that may arise in audit studies that look at one specific location (Lahey and Beasley 2018). At the same time, Neumark (2012) shows that if correspondence studies include variation in observable measures of an applicants' quality, like GPA or college or university name, an unbiased estimate of discrimination can be retrieved even when accounting for the variance of the unobservables.

A seminal example of a correspondence audit study, Bertrand and Mullainathan (2004) send out 4,870 fictitious applications to over 1,300 job ads in Boston and Chicago. They code race in the applications through identifying names given to children in Massachusetts between 1974 and 1979 that were common among one racial group while uncommon among the other. In order to address Heckman's critique of unobservable variance, Bertrand and Mullainathan create both high-quality and low-quality resumes. High quality resumes are differentiated from low quality resumes based on level of experience, gaps in employment and skills listed in the resume. Additional signaling of resume quality was through factors such as, in-school work experience,

additional certification degrees, honors and etc. For each job ad, four applications were sent. Applicants had the same gender but differed in race and resume quality. The results showed a 50 percent lower callback rate for black Americans compared to white Americans. Furthermore, whites with higher quality resumes received 30 percent more callbacks than whites with lower quality resumes but such a drastic increase depending on resume quality was not apparent among blacks. Interestingly, living in wealthier and demographically “whiter” neighborhoods increased callback rates for whites but African-Americans did not receive more callbacks for living in a “better” neighborhood.

In another example of a correspondence audit study, Gaddis (2018b) examined whether racial differences in human capital or racial discrimination determines employment opportunities for white and black college graduates from selective versus less selective institutions. In order to examine the effects of college selectivity, Gaddis holds, gender, and social class constant within pairs. Gaddis uses degrees from three highly selective universities (Harvard, Stanford, and Duke) and three from universities with relatively lower selectivity (University of Massachusetts-Amherst, University of California-Riverside, and University of North Carolina-Greensboro). 1008 jobs were applied in three different geographic locations in the United States. The study found that white applicants with a degree from selective universities have a response rate of 17.5 percent, followed by 12.9 percent for black applicants with a degree from selective universities, then 11.4 percent for white applicants with a degree from a less selective university, and 6.5 percent for black candidates with a degree from a less selective university. Thus, the racial discrimination faced by a non-white job candidate cannot be reduced through higher levels of education, even if from the most elite universities. Additionally, Gaddis found that black candidates received responses for jobs with a listed salary of \$3,071 lower than that of job responses for white candidates.

Studies on the effects of ethnic discrimination on callback rates also show similar trends in labor markets outside the United States. Oreopoulos et al. (2016) looked at the effects of immigrant names (Chinese, Indian, Pakistani, and British) on callback rates in Canada. They sent out 12,910 applications to 3,225 job postings. The results showed that English names received a 39 percent higher callback rate than applications with Indian, Pakistani, or Chinese names. Interestingly, there was no significant difference between the callback rate for Canadian sounding names and British sounding names. Furthermore, the majority of employers that were surveyed for follow up



analysis expressed that they treat names and country of education as a signal that an applicant may lack language skills for the job.

In Table 1, I summarize the results from the audit studies mentioned above and a number of others. As we can see, the effects of racial and ethnic discrimination are not limited to the labor market within the United States. Most of the studies that were conducted outside the United States have looked at how ethnicity affects callback rates (Galrza and Yamada 2014; Bartos, Bauer, Chytilova and Matejka 2013; Kaas and Manger 2012; McGinnity, Nelson, Lunn and Quinn 2009; Bursell 2007). All of the studies find employers discriminate between candidates of gender, race, or ethnicity.

**Table 1: Difference in Callback Rates**

<b>Paper</b>	<b>Country</b>	<b>Apps</b>	<b>Vacancies</b>	<b>Call-back ratio difference</b>
Oreopoulos (2016)	Canada	12910	3225	English name-to-Immigrant: ranged from 1.39 to 2.71 (against Indian, Pakistani and Chinese applicants).
Gaddis (2018b)	US	1008	-	White to-African American (high credential): 1.36 White to -African American (low credential): 1.75
Galarza and Yamada (2014)	Peru	4820	1205	White-to-indigenous ratio: 1.8
Nunley, Pugh, Romero, and Seals (2014)	US	9396	-	White-to-black: 1.18 (unconditional)
Bartos, Bauer, Chytilova, and Matejka (2013)	Czech Rep. and Germany	274(Czech) 745(Ger.)	-	Czech-to-Vietnamese: 1.34
Wright, Wallace, Bailey, and Hyde (2013)	US	6400	1600	White-to-Muslim: 1.58
Kaas and Manger (2012)	Germany	1056	528	German-to-Turkish: 1.29
Maurer-Fazio (2012)	China	21,592	10,796	Han-to-Mongolian: 1.36 Han-to-Tibetan: 2.21
Jacquemet and Yannelis (2012)	US	330	990	English-to-foreign names: 1.41 English-to-Black names: 1.46
McGinnity, Nelson, Lunn, and Quinn (2009)	Ireland	480	240	1.8, 2.07, 2.44 in favor of Irish and against Asians, Germans and Africans respectively.
Bursell (2007)	Sweden	3,552	1,776	Swedish-to-foreign names: 1.82
Bertrand and Mullainathan (2004)	US	4,870	1,300+	White-to-African-American: 1.5

**Source:** Bertrand and Duflo (2016) and author's compilation of literature.

As noted, experimental methods in the form of audit studies have been most effective in detecting discrimination in the labor market. These methods proved to be robust and reliable not only in detecting racial discrimination but also discrimination in other forms such as age, gender and disability (Ameri et al. 2017; Neumark et al. 2018; Neumark et al. 2019; Zhang 2008).

All of the audit studies mentioned above were done utilizing the traditional application process wherein auditors directly applied to job vacancies via email or by post. Thus, none of the audit studies to date have analyzed how big data and algorithms have affected discrimination in the hiring process. However, as we have seen above, big data and algorithms are becoming increasingly important in the hiring process.

### **Algorithms as Discriminants**

There are a number of ways that big data and algorithms can replicate discrimination from traditional hiring processes and even magnify it. For example, in the job search process, certain minority groups could be excluded from receiving information about certain jobs due to targeted job advertisement practices (Ajunwa 2019). An investigation by ProPublica showed that Facebook allowed certain advertisers to use ethnicity to determine who viewed their job or housing ad (Angwin 2016). Aptitude and personality tests are sometimes used by employers to filter out people with mental disorders or instability (O’Neil 2016). This case was clearly highlighted by a lawsuit filed by Kyle Behm against Kroger and six other companies. Kyle, a college student with near-perfect SAT score, who happened to be diagnosed with bipolar disorder was screened out of minimum-wage jobs through the personality tests. Additionally, these same tests have been used to filter out people showing signs of strong labor union sympathies (Newman 2017).

Discrimination may also occur when selecting target variables and class labels. A target variable is the outcomes of interest, and class labels are the division of all possible values of the target variable into mutually exclusive categories. For example, in a model that decides which applicant to callback for an interview, the target variable would be ‘acceptable applicants’, and class labels can be variables such as high school GPA or years of job experience. The problem occurs when the target variables bring in past biases through using variable potentially correlated with race or gender. For example, variables that classify the socio-economic background of the candidate may play a role in determining gaps in resumes which affects who would be a successful candidate. Class labels such as neighborhoods can be highly correlated with race. Thus, the model would just be inheriting the traits of human models (Stauffer and Buckley 2005).

Another way discrimination could take place is through biased ‘training data’. The difference between traditional forms of data analysis and data mining is that, data mining looks to discover relationship between variables within the data set while traditional forms of data analysis

simply return summary statistics to certain queries (Fayyad 2001). For algorithms to be able to determine what attributes can be used as relevant proxies for the model, they need to learn from past sample data.<sup>3</sup> Potentially biased training data leads to discriminatory models (Custers 2013). If the model uses biased datasets to learn from, it will reproduce some of the biases that were involved in the training data. Biased training data can result in systemically disadvantaging certain under-represented population (Turner-Lee et al. 2019). For example, in 2015, Amazon scrapped their internal hiring algorithm after they found it had adopted gender-biases from learning from the company's past decade of hiring decisions (Dastin 2018).

Similarly, there also may be problems in the data collection process. Sometimes, datasets might not accurately represent some protected classes. At times, there may be nonrandom and systemic omission of certain groups of people, potentially as a result of poverty, access to internet, geography or lifestyle (Lerman 2013). In most cases, those that are living on big data's 'margins' are those who are underprivileged or under-represented.

Finally, some people may intentionally discriminate using the techniques of data mining. 'Masking' is the technique for purposefully biasing the collection of data or preserving some of the prejudices in the dataset. Institutions could use these techniques to circumvent certain practical and legal barriers that stopped them from excluding individuals from certain racial groups from consideration. For example, it is possible to reliably predict membership in a certain class by data mined about some combinations of 'likes' and network of friends. Employers can then use these traits to discriminate by setting up future models by excluding people holding these traits from considerations (Kosinski 2013).

These potential negative outcomes are more likely to affect low-skilled workers (O'neil 2016). The purpose of using hiring algorithms is to save time in sifting through large piles of job applicants. In low-skilled job postings, there are likely to be larger applicant pools. Therefore, hiring algorithms are especially prominent in low-wage and hourly jobs (Ajunwa 2016). But the possible forms of discrimination via algorithms are not confined to low wage jobs.

Given the many ways that new hiring techniques can discriminate and given the substantial evidence of racial discrimination in traditional hiring markets and labor markets generally, it is import to analyze the effects of big data and algorithms on labor market discrimination. To get a

---

<sup>3</sup> Better known as 'training data'

better understanding of the effect of big data and algorithms on discrimination in the hiring process, I create a correspondence study adapted to the new application process.

## **The Experiment**

As stated above, I conduct a correspondence study adapted to the new job application process to see how big data and algorithms are affecting discrimination in the hiring process. In the experiment, I send 248 fictitious applications to 62 different firms. Applications code race through using distinct racial names, as done in Bertrand and Mullainathan (2004) and Gaddis (2018). To address Heckman's critique, I also create high and low skill applicants by differing the level of experience stated in the resumes. I recorded the callbacks received for applications to determine whether discriminatory practices take place in the hiring process.

## **Research Design**

The two most important factors determining the validity of this experiment are identifying firms that use algorithms (ATS) in the recruitment process and distinguishing between black applicants and white applicants. In determining the firms to apply to, I identify Fortune 500 companies as the most likely firms to use ATS in their recruitment. A study by Gagua (2015) shows that 93 to 96 percent of Fortune 500 companies utilize e-recruitment in 2003, and a more recent study by Jobscan reports that 98.8 percent of Fortune 500 companies use ATS in the hiring process (Qu 2021). In order to reduce the margin for error, I performed a search on all the companies in the study to make sure they use ATS.

To code race, I use the names identified by Gaddis (2018b) as distinct racial names. Gaddis uses data from the New York State Department of Health on births during the early 2000s. He picks names with more than 50 births per year and at least 75 percent born to parents of one particular race (black or white), thereby ensuring that the indication of race is not obscured. He also ensures that the mother's education of those giving the names were similar. I use 8 different names: Jalen and DaQuan (black/male); Ebony and Shanice (black/female); Charlie and Ronny (white/male), and Aubrey and Erica (white/female). In order to make sure that I give a clear indication of race I added one extracurricular activity that can serve as an ethnic indicator (Lahey and Beasley 2018). These extracurricular activities were actual club organizations from the respective universities. Furthermore, it was easier to indicate race in applications in my study

because most of the application portals asked for the candidate's ethnicity. Thereby, I simply had to click on black/African American when asked about race and ethnicity. This addition in the application process eliminates the potential shortcomings pointed out in regard to previous studies not clearly communicating race to employers (Gaddis 2018b; Bertrand and Mullainathan 2003).

I applied to job openings in the San Francisco and New York City area. These two cities had the largest concentration of Fortune 500 companies in the United States which made them favorable locations for the experiment. During the time of application in March 2021, the unemployment rate in the United States was 6.0 percent (U.S Bureau of Labor Statistics). The unemployment rate at San Francisco was 6.3 percent and the unemployment rate at New York was 8.7 percent. Both these cities were subject to more stringent lockdown measures to combat the spread of Covid-19.

### **Building the Resume**

To build the applications that are field specific, I use sample resumes as guides for each job category from Indeed.com.<sup>4</sup> Resume template took on a larger importance than previous correspondence studies because ATS may not be able to parse data from certain resume templates. I referred to career websites to avoid resume templates that are not compatible with ATS. Furthermore, I manually went through all the online applications so that the information from the resumes were being accurately translated onto the job application portals of companies. I used four different resume templates for each of the four applicants for a given job opening in order to decrease the chance of the experiment being exposed.

The types of job openings were largely categorized into: management, sales, administration, finance, IT, customer service, research, retail or a mixture of these. In order to cater specifically to the requirements of each job type, I drafted up different job descriptions and job experiences for each of the job categories. I slightly altered the grammar for each job description for each category but kept the contents of the description largely unchanged. I used real companies in New York and San Francisco to assign to the candidates' experiences. From this, I was able to assign randomly selected job experiences and activities for each job category that would make a good fit for each job opening.

---

<sup>4</sup> I also referred to the sample resumes that were available through the Career Office at Connecticut College.

In order to account for the weaknesses discussed earlier in the literature review, such as the points made by Heckman (1993), I categorized the resumes by altering the quality of resumes. The job candidates identified as a typical high-quality resume had a higher GPA, 1 additional internship experience, 1 additional year of job experience and no gap between jobs. The low-quality resume do not have these qualities and have gaps in the job experience. A typical high-quality resume had a higher GPA, 1 extracurricular activity, 1 internship between academic semesters, 2 years of job experience and 1 more extracurricular activity with racial indicator for black applicants. If the type of job was a mixture of the eight categories mentioned above, the high-quality resume had 1 year of job experience in one industry and 1 year of job experience in the other industry, making a total of 2 years of job experience.

For the study, there was the possibility that the job hiring algorithm could use an applicant's address as a proxy for socioeconomic background or quality of secondary education (Favaretto et al. 2019), so extra care was given into inputting addresses that were apt for the study. There were a couple of factors I considered when determining the addresses. Firstly, I identified the socioeconomic background of addresses by looking at the ethnic break-up within the neighborhood. Secondly, I searched through real-estate websites and housing-price data in order to make sure that the median property prices in the areas being selected were similar. These two steps ensure that the location does not factor deciding by matched candidates sent to a given employer.

To build and randomize the inputs of the resume without any bias, I used the resume randomizing program by Lahey and Beasley (2018). However, there were several adaptations that needed to be made because of the changes firms made in their recruitment process. While a lot of the jobs in the past required applicants to submit job applications by submitting resumes and cover letters via email, almost all of the Fortune 500 companies require applicants to fill out relevant information through their own job application portals. The fact that these firms require applicants to apply by filling out the firms' application portal also indicates that they are using ATS because the firm is optimizing data collection format by customizing their application portal. As a result of the need to enter data through employer specific portals, we had to code ways to enter the inputs in the correct way. So, I compiled all the inputs that would go into the resume in an Excel spreadsheet and built a program that would go through all the application portals and open up relevant inputs for each job.

## **Applying to jobs**

In determining to apply, I identified jobs from a popular job search portal that were entry-level but mostly required bachelor's degree so that I wasn't using overly-qualified resumes for extremely low skilled jobs. I also had to make sure that the jobs didn't require any further aptitude testing such as personality testing or video interview uploads in the initial stages, so that I didn't have to input anything that would bias the results of the study. I also eliminated jobs and companies in industries that required very specific skill sets and capabilities and jobs that required a proof of certification such as a driver's license or a social security number. These unexpected limitations during the application process hindered the study from reaching the goal of applying to 100 companies with a total of 400 job applications.

All the jobs that I applied to were within the 50-mile radius of either San Francisco or New York City, and the applications were sent in within 20 days of the initial job posting. Job postings that expired during the span of the application period were substituted with similar jobs from the same company.

Unlike in traditional correspondence studies where one could simply apply to a job by sending in email or mail in applications with resumes and cover letters, ATS applications require candidates to manually transpose some of the relevant information from the resumes onto their application steps. They also sometimes asked questions about race and work authorization in the United States that requires candidates to individually answer. While most of the ATS parse the data and auto-complete the major sections like education and previous job experiences, candidates still need to fill out the information missed by the ATS when transposed from the resume. This process makes it far more time consuming than previous correspondence studies which did not apply to jobs that use ATS in the recruitment process.

Also, the process of applying could not be entirely automated because of human verifications processes such as Captcha. Most cases also required email verification which needed to be handled by manually verifying the accounts. This made the whole process even more time-consuming than in previous studies, so our application period was for around 2-month between the start of February and end of March in 2021. For all the jobs, we had to apply either directly through the company's website or through the job application portal. None of the application were sent in through the traditional method of emails.

## Results

### *Difference in Callback by Race*

Table 2 shows the callback rates for applicants with different racial sounding names. Applicants with white sounding names had a 17.74 percent chance of receiving an initial callback from employers after the application; applicants with black sounding names had a 18.55 percent chance of receiving a callback. African-American sounding names received callbacks at a 0.81 percentage point higher rate, but as it is apparent from the high p-value, this difference in callback rate is not significant.

Breaking down the results into the two locations of application, New York and San Francisco, the difference in callback rate between white and black sounding names in New York was 3.03 percentage point higher for African-American sounding names. For applications submitted to jobs in San Francisco, the callback rate was 1.73 percentage point higher for white applicants. However, for both the cases, the p-values were high, indicating no real significant difference in callback rates between races.

When we subdivide the callbacks by different gender, we see that male candidates with African-American sounding names received callbacks at a 1.92 percentage point higher than male candidates with white sounding names. For female applicants, there was no difference in callback rate between candidates of different racial groups. In both cases, the p-value for the percentage point difference was high, which once again shows that there is no significant difference in callback rates between races.

From Table 2, we see a large difference in callback rates between New York and San Francisco. There was an 18.5 percentage point higher callback rate for white sounding applicants who applied to jobs in San Francisco compared to those who applied to jobs in New York, and likewise, a 13.74 percentage point higher callback rate for African-American sounding applicants who applied to jobs in San Francisco as opposed to New York. These differences can largely be accounted by the heterogeneity in the jobs applied to in the two different locations and the fact that economy of New York was harder hit by the pandemic than that of San Francisco.



**Table 2: Mean Callback Rates by Race**

	<b>Percent Callback for White Names</b>	<b>Percent Callback for African- American Names</b>	<b>Ratio</b>	<b>Percentage Point Difference (p-value)</b>
Sample: All sent resumes	17.74 (22)	18.55 (23)	0.94	0.81 (0.87)
New York	9.09 (6)	12.12 (8)	0.75	3.03 (0.575)
San Francisco	27.59 (16)	25.86 (15)	1.07	1.73 (0.836)
Males	17.31 (9)	19.23 (10)	0.9	1.92 (0.802)
Females	18.01 (13)	18.01 (13)	1	0 (1.00)

**Note:** Within the parentheses are the total number of callbacks for each corresponding subcategory. The p-values in Column 4 are for a test of proportion testing the null hypothesis signaling that callback rates are equal across different racial sounding names.

Table 3 shows the percent of jobs that required a university degree. 51.52 percent of the job ads I applied to in New York listed a bachelor’s degree as a minimum qualification for the job, while just 41.38 percent of the job ads applied to in San Francisco required a bachelor’s degree. This indicates that in general, the jobs applied to in New York were ones that required higher qualifications and required more skills than jobs applied to in San Francisco. This also stems from the fact that some types of jobs require higher education and skill level even for entry level jobs. The two cities had certain job types that were more prevalent in each of the city, and these differences have translated into a skewed result for the two different locations. Table 4 shows the six different types of jobs that I have applied to. There were three categories where there were clear differences in the number of jobs applied to between the two cities. More administration and finance jobs were applied to in New York while all the IT related jobs were applied to in San Francisco. Administration and finance jobs generally require higher education and skillset than jobs in management, retail and sales. Therefore, we are seeing a skewed result for callbacks between these two locations.

**Table 3: Degree Requirement for a Job by Location**

	<b>Percent Jobs Requiring a University Degrees</b>
New York	51.52 (68)
San Francisco	41.38 (48)

**Note:** A total of 132 applications were applied to jobs in New York and a total of 116 applications were applied to jobs in San Francisco.

**Table 4: Jobs Applied to by Job Type**

	New York	San Francisco
<b>Administration</b>	32	12
<b>Finance</b>	28	16
<b>IT</b>	0	16
<b>Management</b>	32	32
<b>Retail</b>	20	20
<b>Sales</b>	20	20

**Note:** The following categories of job type does not indicate the type of industry the firms applied to were classified under. These categories indicate what type of job the job applied to are.

I run a probit model analysis to further explore the possibilities of racial discrimination in callback rates. For the model, I use callback, which is a dummy variable, as the independent variable and race as the dependent variable, while controlling for gender, race, quality of resume and geography. Table 5 gives a short summary of the coefficients and p-values obtained from the probit model. The coefficient for race is 0.029 which is extremely small. Furthermore, the large p-value of 0.878 indicates that race is not a good explanatory variable for callbacks. The small pseudo r-squared value of 0.006 also confirms that this is a poor model for explaining callback decisions. This result goes hand in hand with the results drawn from Table 2.

**Table 5: Results from Probit Model**

	<b>Coefficient</b>	<b>P-value</b>
<b>Race</b>	0.029	0.878
<b>Gender</b>	-0.079	0.966
<b>Resume Quality</b>	0.215	0.250
<b>Constant</b>	-1.032	0.000

**Note:** For the model, I take callback as the dummy dependent variable. I take race (white or black) as the independent variable and hold gender and resume quality constant for the model. The probability to chi-squared value was 0.715 while the pseudo r-squared value was 0.006, indicating that the model is extremely poor for measuring callback rates.

Table 6 gives a descriptive statistic of callback between different races depending on the quality of resume. For applicants who applied with a high-quality resume, the callbacks were 22.58 percent and 19.35 percent for African-American names and white names respectively. There was 3.23 percentage point difference between the two pool of applicants, which was not statistically significant. For lower quality resumes, the callback rates were 14.52 percent for African-American sounding names and 16.13 percent for white sounding names. In this case, the group of applicants assigned white names had a callback rate 1.61 percentage point higher than that of applicants with African-American names. Once again, the difference in the callback rate was not statistically significant.

However, it is worth noting that there is a greater drop in callback rates for African-American applicants when moving from a higher-quality resume to a lower-quality one. This is a 55.51 percent drop in callback rate. On the other hand, we don't see such a drastic drop in the callback rates with a drop in resume quality for white applicants. This preliminary result draws attention for future research into the effects college credentials or the effects length and quality of experiences may have on callbacks for different races in a labor market setup where AI and big data are playing major roles in the hiring process. Studies similar to those done by Gaddis (2018b) and Oreopoulos (2016) could provide for a more comprehensive picture of the effects AI and big data are bringing onto the labor market.

**Table 6: Callback Rate by Resume Quality**

	<b>Percent Callback for African-American Names</b>	<b>Percent Callback for White Names</b>	<b>Percentage Point Difference (p-value)</b>
High Quality Resume	22.58 (14)	19.35 (12)	3.23 (0.66)
Low Quality Resume	14.52 (9)	16.13 (10)	1.61 (0.81)

***Weaknesses of the Study***

The biggest weakness of this correspondence study is the small sample size. This study was held as a preliminary study to explore the possibilities of applying the traditional methods of a correspondence study to a newer labor market where job applications are parsed by ATS. Therefore, the study was held in a smaller scale than other comparable studies. This smaller scale makes the sample size too small to give a comprehensive view of the labor market.

Secondly, as mentioned earlier, there is a lot of heterogeneity in the type of jobs I applied to. Table 7 shows the callback rates for the six different type of jobs I applied to. Jobs in sales, retail and management had high callback rates, of 45 percent, 25 percent and 17.19 percent respectively. However, jobs in finance and administration had low callback rates of 2.27 percent and 11.36 percent respectively, and jobs in IT had a 0 percent callback rate. Jobs in these three categories have a relatively higher threshold in terms of qualification as opposed to jobs in management retail and sales that have a lower barrier to entry even amongst entry level jobs. This is evident when comparing job ads for degree requirements. Table 8 shows what percent of the job postings for each job type required a college degree. 72.73 percent of the administrative jobs required a college degree while a 100 percent of the jobs in finance and IT required a college degree. Jobs in these industries usually have job specific skills that are usually acquired through college education. This fact is reflected in the low callback rates for administrative, finance and IT jobs. However, just 31.25 percent of the management jobs required a college degree, and an even lower 10 percent of sales jobs required a college degree. None of the retail jobs applied to required a college degree. The large discrepancy in skill required for each type of job led to each company either accepting all 4 candidates applying to the job or rejecting all 4 candidates applying to the job. This also meant that in most cases, the 4 candidates applying to the job were over-qualified or under-qualified. Such a phenomenon could have added more noise in the study, possibly diluting the effects of racial discrimination in this study.

**Table 7: Callback Rate by Job Type**

	Callback Rate	Number of Jobs Applied to
Administrative	11.36 (5)	44
Finance	2.27 (1)	44
IT	0	16
Management	17.19 (11)	64
Retail	25 (10)	40
Sales	45 (18)	40

**Note:** Callback rate is in percent and the numbers in the brackets denote the actual number of callbacks received. The number of callbacks is a total of callbacks received in the two different locations.

**Table 8: College Degree Requirement by Job Type**

	Percent of Jobs Requiring College Degree
Administrative	72.73 (32)
Finance	100 (44)
IT	100 (16)
Management	31.25 (20)
Retail	0
Sales	10 (4)

**Note:** The college degree requirement was counted

Finally, another weakness stemmed from the depressed labor market due to the effects of the Covid-19 pandemic. As mentioned earlier, the unemployment rates at San Francisco and New York were at 6.3 percent and 8.7 percent respectively at the time of application in March 2021. In March 2020, before the pandemic hit, the unemployment rates were 3.4 percent and 4.0 percent respectively. That is an 85 percent rise in unemployment in San Francisco and a 117.5 percent rise in New York. Both these cities were more cautious about reopening, which slowed down the local economy and increased unemployment. This was apparent when finding jobs to apply to. There was significant difficulty finding suitable jobs, let alone matching the jobs for homogeneity across

skill level and across location. Due to firms being unwilling to make additional recruitments, it was hard to make a bigger data set.

## **Discussion**

### ***Takeaways from the result***

While this study does not find evidence of racial discrimination in the initial callback stages, there could be several different interpretations to this result. The use of ATS has led to more firms giving callbacks to all four applicants applying to the job. This could have led to the largely homogeneous result across applicants with different racial names. With ATS evaluating the resume, firms have been able to cut short the time they were previously using in resume evaluation. This has resulted in additional evaluation stages post-ATS. This point will be elaborated on in the next section.

Another, more hopeful interpretation for the result is that we may actually be seeing a decrease in racial discrimination from firms that are using ATS. Starting from the early parts of 2020, we've seen a substantial rise in awareness about systemic racism due to mobilized social movements denouncing discrimination in the United States. While these movements were largely concentrated on police brutality, they served as an alarm call for the prolonged stagnation, or even regression, in the movement towards racial equity. In light of the dialogue surrounding racial equity, firms may have gradually started to appraise their hiring strategies whether voluntarily or out of fear of public criticism. A change in hiring strategy can be relatively easier to implement on a hiring algorithm than on human recruiters. Possible discriminatory practices can be eliminated by changing the internal workings of the algorithm and refining the training data set. However, it is more difficult to eliminate pre-existing biases in human recruiters, even if they make a conscious effort to evaluate candidates equally. While racial discrimination has not been eliminated in the hiring process, there is a possibility that the effects of social changes pressured, or inspired, employers to make changes to their algorithms, leading to a relatively even positive callback distribution amongst the applicants from the two racial groups.

### ***Difficulties in the study and implications for future research***

In this section, I would like to outline some difficulties I met with while conducting the correspondence study, and the implications it has for future research. One of the major difficulties

came from automating the application process. In previous correspondence studies, researchers directly applied to jobs, either by post or by email (Bertrand and Mullainathan 2004; Oreopoulos 2016; Gaddis 2018b). It is relatively easier to automate the application process through these means because the scripts running the application simply need to loop the generation of emails with different resumes and applicant identities. However, with the introduction of ATS, there are far more hurdles to automation due to various authentication processes. Filling out a job application for a firm that uses ATS requires candidates to create their own account, verify the account and answer security questions. In most cases, there is a Captcha process where the candidate needs to click on pictures corresponding to the word displayed on the screen. These processes were incorporated by firms to filter out scripts or bots from applying to jobs. The only way to bypass this initial authentication stage is by manually verifying the accounts for each applicant.

After clearing the verification comes the second problem in automation. Different firms use different ATS systems which requires researchers to create multiple scripts to fill out the application. During the application I was able to identify around 4 large ATS systems that Fortune 500 companies were outsourcing their recruitment process to. Though the application format does not vary by a large amount if the application is managed by the same ATS software, some firms ask additional job specific questions like, “Are you able to lift more than 40lbs?” or “Are you able to work in a remote setting?” These variations that may exist from firm to firm makes it counterproductive to create multiple scripts to automate the application process.

From this study, it was apparent that there needed to be an update in the methods of correspondence studies, and without it, future research would be neglecting an extremely large portion of the labor market that is using ATS. To account for these issues, I chose to make the application process semi-automated and was able to create a script that provides all the information needed for the application process on a terminal. The script would provide information needed for creating and verifying an account, and provide information drawn from each resume that could be used to fill out fields in the application portal. While I had to manually run through all the applications, in essence all I was doing was copy pasting the relevant information into the different fields and providing human verification in places that required it. This way, I could overcome the fastidious process of typing out and recording all the information I used for each candidate. Though this process was more time consuming than the traditional methods of correspondence study, it is

possibly the best way to bypass the hurdles posed by ATS, while keeping applications as homogeneous as possible.

The second difficulty which I previously mentioned is in regards to submitting enough applications to generate an adequate sample size. The first step I had to take in the study was to identify which firms were using ATS. A lot of the private sector market research companies held information about which firms used which ATS, however, these reports were largely inaccessible due to their high price. The second-best alternative was to apply to Fortune 500 companies, because research showed that 98.8 percent of these companies were using ATS (Qu 2021). Additionally, I went through the application portals of all these companies to ensure that they were using ATS. However, this means that the effect recruitment through ATS has on smaller firms are not recorded in the study. There may be different results from smaller companies that use the ATS because they likely receive a far smaller pool of applications than jobs in Fortune 500 companies. This could mean that the initial stages of the recruitment process which is dealt with by ATS in smaller firms may take on larger importance than in the recruitment process for large firms. If so, firms may apply more rigorous filtering through the ATS which would reduce the chances of all 4 applicants being accepted or rejected as in the case for this study. So, identifying a way to find firms of all sizes that use the ATS would provide a foundation for a more robust study.

Furthermore, the introduction of ATS in the modern labor market is a factor that all researchers undertaking a correspondence study must be wary about. With many firms moving towards a highly automated recruitment process, relatively newer correspondence studies may have been recording callbacks from both human recruiters and ATS software. Though now it is common for firms using the ATS to ask its applicants to create an account and apply through their own application portals, in the past, applicants would send in their resumes through traditional means and those resumes would be gathered and parsed by an ATS. As mentioned earlier, around 93 to 96 percent of the Fortune 500 companies were already using e-recruitment as early as 2003 (Gagua 2015). So, the results from correspondence studies showing discrimination in the labor market may not be solely due to human biases but may be a mixture of biases from ATS. In order to get a more robust result through correspondence studies, it would be essential for researchers to identify which firms are using forms of e-recruitment and which are not, and control for this factor.

Another point to note for future research is that the introduction of ATS has actually made it far easier to code for the variable of concern. In the case of this study, the variable being studied



and coded was race. The study by Bertrand and Mullainathan (2004) found a way to code race through names. Most of the correspondence studies have followed this method to code race in the applications. However, it is still unclear whether indication of race through names is being clearly communicated to employers (Bertrand and Mullainathan 2004; Gaddis 2018b). Correspondence studies that looked at other forms of discrimination faced similar problems. A study looking at gender pay gap used similar naming methods to code gender (Zhang 2008), while a study looking at age discrimination coded age by graduation year in the resume (Neumark et al. 2019). However, when applying through ATS, researchers do not need to go through this tedious process. Jobs applied through the ATS explicitly ask for the race, gender and birth date of the applicant (though fields like race and gender are categories that do not require mandatory response). There are far more ways to self-identify in an application through the ATS. For example, one can choose veteran status, citizenship status or disability status in the application portal. This makes it far easier for researchers to code for the discriminatory variable in question, which in turn may open up new dimensions to research on discrimination in the labor market.

Finally, as briefly touched upon earlier, the utilization of ATS and other AI related recruitment software may have reduced the importance of the initial callback stage. The use of ATS has helped employers save time in the initial process of evaluating resumes. In turn, employers seem to have elongated the recruitment process after initial resume evaluations. 56.82 percent of the positive callbacks in my study required applicants to do a follow up interview or assessment through an AI based software. AI based interviews prompt candidates with questions and then evaluate the content, tone, and other factors using video algorithm software. Assessment tests are mostly used to evaluate the personality of a candidate and their fit for the job. They are also conducted online and evaluated by algorithms. Thus, it seems that the use of ATS has come in combination with more non-human evaluation stages in the recruitment process. This also accounts for the fact that resume evaluation through the ATS may be less rigorous than when the process was done by humans. Of the firms that responded positively and required AI based interviews and assessments, 62.5 percent of them gave positive responses to all four applicants applying to the job. The more critical stages of the hiring process, where discrimination might be more prominent, are then pushed down the road as a result of the time savings from ATS.

With this in mind, future research would need a more holistic approach in recording callbacks. Critics of the correspondence study method have argued that callbacks are not a direct

measurement of acceptance to a job (Heckman 1998). The introduction of ATS in the labor market further strengthens the argument against using callbacks from the resume evaluation stage to indicate acceptance to a job. In light of this finding, future research on discrimination should look to implement a combination of both correspondence and in-person audit study to get a deeper understanding into the mechanisms within each stage of the recruitment process. The resume evaluation stage can be done as a correspondence study and the recorded video interview or online assessment can be done by trained testers, similar to those from Pager et al.'s (2009) study. Since the audit study done by testers would be applicable to only the applicants that received the initial callback from resume evaluation, the study can be held at a manageable scale. Using this method would allow for a better indicator than initial callbacks, especially as it gets closer to the actual hiring decision.

## **Conclusion**

In this paper, I conducted a correspondence study to determine whether ATS used in the hiring process racially discriminate. For the study I sent out 248 applications to Fortune 500 companies utilizing ATS. I do not find evidence of racial discrimination in this preliminary study. Potentially, this is a result of employers reacting to increasing conversations about systematic racism in the United States. However, these results could also be from several weaknesses in the study like small sample size, heterogeneity in the jobs applied to, effects of Covid-19. The results could also stem from the elongation of the hiring process as a result of the time saving from adopting ATS and other algorithms in the hiring process. As stated above, 56.82% of positive callbacks outlined the need for candidates to complete video interviews or assessment tests which would be analyzed by AI software, and in many of the cases where an employer responded positively to one candidate, they also responded positively to all the other candidates. Thus, the initial resume evaluation process may be less significant than in the past.

Thus, the results from this paper seem to highlight the growing opacity of the modern hiring process. Future research that wants to test for discriminatory labor market practices where AI and big data are used would most likely have to conduct a double audit study – correspondence for an initial response and in-person audit study to record video interviews or take assessment exams to be evaluated by AI software. And as stated above, the initial stage of this study is more time consuming compared to past audit studies because of the need to both automate and manually input

data into company specific portals. Hence, AI and big data in the job market has made it extremely difficult for people to grasp the internal workings of algorithms.

## Reference:

- Aigner, Dennis J., and Glen G. Cain. 1977 “Statistical Theories of Discrimination in Labor Markets.” *Industrial and Labor Relations Review*, vol. 30, no. 2, p. 175., doi:10.2307/2522871.
- Ajunwa, Ifeoma. 2016. “The Paradox of Automation as Anti-Bias Intervention.” *Cardozo Law Review*, vol. 41, 1671.
- Ameri, Mason, et al. 2017. “The Disability Employment Puzzle: A Field Experiment on Employer Hiring Behavior.” *ILR Review*, vol. 71, no. 2, pp. 329–364., doi:10.1177/0019793917717474.
- Angwin, Julia Terry Parris Jr. “Facebook Lets Advertisers Exclude Users by Race.” *ProPublica*, 28 Oct. 2016, [www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race](http://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race).
- Barocas, Solon, and Andrew D. Selbst. 2016. “Big Data's Disparate Impact.” *SSRN Electronic Journal*, doi:10.2139/ssrn.2477899.
- Vojtech Bartoš, Michal Bauer, Julie Chytilová, and Filip Matejka. 2013. Attention Discrimination: Theory and Field Experiments. CERGE Working Paper, ISSN 1211-3298.
- Becker, Gary S. 1971. *The Economics of Discrimination*. University of Chicago Press.
- Bertrand, Marianne, and Esther Duflo. (2016) “Field Experiments on Discrimination.” *NBER Working Paper Series*, doi:10.3386/w22014.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *SSRN Electronic Journal*, doi:10.2139/ssrn.422902.
- Brannon, Monica M. “Datafied and Divided: Techno-Dimensions of Inequality in American Cities.” *City & Community*, vol. 16, no. 1, 2017, pp. 20–24., doi:10.1111/cico.12220.
- Cappelli, Peter. 2014. *Why Good People Can't Get Jobs the Skills Gap and What Companies Can Do about It*. Recorded Books.
- Custers, Bart. 2013. “Data Dilemmas in the Information Society: Introduction and Overview.” *Studies in Applied Philosophy, Epistemology and Rational Ethics Discrimination and Privacy in the Information Society*, pp. 3–26., doi:10.1007/978-3-642-30487-3\_1.
- D'alessandro, Brian, et al. 2017. “Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification.” *Big Data*, vol. 5, no. 2, pp. 120–134., doi:10.1089/big.2016.0048.
- Dastin, Jeffrey. 2018 “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women.” *Reuters*, Thomson Reuters, 10 Oct. 2018, [www.reuters.com/article/us-amazon-](http://www.reuters.com/article/us-amazon-)

[com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.](https://doi.org/10.1007/978-3-642-30487-3_12)

- Edmund S. Phelps. 1972. The Statistical Theory of Racism and Sexism. *American Economic Review*, pages 659–661.
- Farkas, George, and Keven Vicknair. 1996. “Appropriate Tests of Racial Wage Discrimination Require Controls for Cognitive Skill: Comment on Cancio, Evans, and Maume.” *American Sociological Review*, vol. 61, no. 4, p. 557., doi:10.2307/2096392.
- Favaretto, Maddalena, et al. 2019. “Big Data and Discrimination: Perils, Promises and Solutions. A Systematic Review.” *Journal of Big Data*, vol. 6, no. 1, doi:10.1186/s40537-019-0177-4.
- Gaddis, S. Michael. 2018a. “An Introduction to Audit Studies in the Social Sciences.” *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, pp. 3–44., doi:10.1007/978-3-319-71153-9\_1.
- Gaddis, S. Michael. 2018b. “Discrimination in the Credential Society: An Audit Study of Race and College Selectivity in the Labor Market.” doi:10.31235/osf.io/6qjue.
- Galarza, F. B., & Yamada, G. (2014). Labor Market Discrimination in Lima, Peru: Evidence from a Field Experiment. *World Development*, 58, 83–94. <https://doi.org/10.1016/j.worlddev.2014.01.003>
- Hacker, Philipp, and Bilyana Petkova. 2017. “Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers.” *Northwestern Journal of Technology and Intellectual Property*.
- Hajian, Sara, and Josep Domingo-Ferrer. 2013. “A Methodology for Direct and Indirect Discrimination Prevention in Data Mining.” *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1445–1459., doi:10.1109/tkde.2012.72.
- Heckman, James and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." Pp. 187- 258 in *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by M. Fi and R. Struyk. Lanham, MD: University Press of America
- Heckman, James J. 1998. “Detecting Discrimination.” *Journal of Economic Perspectives*, vol. 12, no. 2, pp. 101–116., doi:10.1257/jep.12.2.101.
- Leo Kaas and Christian Manger. Ethnic Discrimination in Germany’s Labour Market: A Field Experiment. *German Economic Review*, 13(1):1–20, 2012.
- Kamiran, Faisal, et al. 2013. “Techniques for Discrimination-Free Predictive Models.” *Studies in Applied Philosophy, Epistemology and Rational Ethics Discrimination and Privacy in the Information Society*, pp. 223–239., doi:10.1007/978-3-642-30487-3\_12.

- Kenneth J. Arrow. 1973. The Theory of Discrimination. *Discrimination in Labor Markets*, 3(10):3–33.
- Kosinski, M., et al. 2013. “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior.” *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, , pp. 5802–5805., doi:10.1073/pnas.1218772110.
- Lahey, Joanna, and Ryan Beasley. 2018. “Technical Aspects of Correspondence Studies.” *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, pp. 81–101., doi:10.1007/978-3-319-71153-9\_4.
- Lerman, Jonas. 2013. “Big Data and Its Exclusions.” *SSRN Electronic Journal*, doi:10.2139/ssrn.2293765.
- Lutz, Christoph. 2019. “Digital Inequalities in the Age of Artificial Intelligence and Big Data.” *Human Behavior and Emerging Technologies*, vol. 1, no. 2, pp. 141–148., doi:10.1002/hbe2.140.
- Markets and Markets. 2020. "Research and Markets Releases Report : Applicant Tracking System." *Professional Services Close-Up* 16 Mar. 2020: . *Business Insights: Essentials*. Web. 12 Apr. 2021.
- Margaret Maurer-Fazio. Ethnic Discrimination in China’s Internet Job Board Labor Market. *IZA Journal of Migration* 2012, 1(12):1–24, 2012.
- Moa Bursell. What’s in a Name? A Field Experiment Test for the Existence of Ethnic Discrimination in the Hiring Process. *SULCIS WP*, 7, 2007.
- Frances McGinnity, Jacqueline Nelson, Pete Lunn, and Emma Quinn. 2009. *Discrimination in Recruitment*. Equality Research Series.
- Neal, Derek A., and William R. Johnson. 1996. “The Role of Premarket Factors in Black-White Wage Differences.” *Journal of Political Economy*, vol. 104, no. 5, pp. 869–895., doi:10.1086/262045.
- Neumark, David. 2012. “Detecting Discrimination in Audit and Correspondence Studies.” *Journal of Human Resources*, vol. 47, no. 4, pp. 1128–1157., doi:10.1353/jhr.2012.0032.
- Neumark, David, et al. 1995 “Sex Discrimination in Restaurant Hiring: An Audit Study.” doi:10.3386/w5024.
- Neumark, David, et al. 2018. “Do State Laws Protecting Older Workers from Discrimination Reduce Age Discrimination in Hiring? Evidence from a Field Experiment.” *Journal of Political Economy*, vol. 127, no. 2, doi:10.3386/w25369.

- Neumark, David, et al. 2019. "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment." *Journal of Political Economy*, vol. 127, no. 2, pp. 922–970., doi:10.1086/701029.
- Newman, Nathan. 2017. "Reengineering Workplace Bargaining: How Big Data Drives Lower Wages and How Reframing Labor Law Can Restore Information Equality in the Workplace." *University of Cincinnati Law Review*, vol. 85, no. 3, doi:10.2139/ssrn.2819142.
- Nicolas Jacquemet and Constantine Yannelis. Indiscriminate Discrimination: A Correspondence Test for Ethnic Homophily in the Chicago Labor Market. *Labour Economics*, 19(6):824–832, 2012.
- John M. Nunley & Adam Pugh & Nicholas Romero & Richard Alan Seals, Jr., 2014. "An Examination of Racial Discrimination in the Labor Market for Recent College Graduates: Estimates from the Field," Auburn Economics Working Paper Series auwp2014-06, Department of Economics, Auburn University.
- O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin Books.
- O'Neill. 1990. "The Role of Human Capital in Earnings Differences Between Black and White Men." *Journal of Economic Perspectives*, vol. 4, no. 4, pp. 25–45., doi:10.1257/jep.4.4.25.
- Oreopoulos, Philip, et al. 2016. "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes." *AEA Randomized Controlled Trials*, doi:10.1257/rct.1149.
- Pager, Devah, et al. 2009 "Discrimination in a Low-Wage Labor Market." *American Sociological Review*, vol. 74, no. 5, pp. 777–799., doi:10.1177/000312240907400505.
- Qu, Linda. 2021. "Report: 99% of Fortune 500 Companies Use Applicant Tracking Systems." *Jobscan*, Jobscan, 26 Feb. 2021, [www.jobscan.co/blog/99-percent-fortune-500-ats/](http://www.jobscan.co/blog/99-percent-fortune-500-ats/).
- Schwab, Klaus. 2016. "The Fourth Industrial Revolution: What It Means and How to Respond." *World Economic Forum*, [www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/](http://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/).
- Spar, Benjamin, et al. LinkedIn, 2018, pp. 1–62, *Global Recruiting Trends 2018*.
- Stauffer, Joseph M., and M. Ronald Buckley. 2005. "The Existence and Nature of Racial Bias in Supervisory Ratings." *Journal of Applied Psychology*, vol. 90, no. 3, pp. 586–591., doi:10.1037/0021-9010.90.3.586.

- Sweeney, Latanya. 2013. "Discrimination in Online Ad Delivery." *Communications of the ACM*, vol. 56, no. 5, pp. 44–54., doi:10.1145/2447976.2447990.
- Turner-Lee, Nicol, et al (2019). "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms." *Brookings*, Brookings, 25 Oct. 2019, [www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#footnote-5](http://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#footnote-5).
- U.S Bureau of Labor Statistics. "Smoothed Seasonally Adjusted Metropolitan Area Estimates," *U.S Bureau of Labor Statistics [online]*. Available at <https://www.bls.gov/lau/metrossa.htm>.
- Bradley R.E. Wright, Michael Wallace, John Bailey, and Allen Hyde. 2013. Religious Affiliation and Hiring Discrimination in New England: A Field Experiment. *Research in Social Stratification and Mobility*, 34:111–126,.
- Zhang, Liang. 2008. "Gender and Racial Gaps in Earnings among Recent College Graduates." *The Review of Higher Education*, vol. 32, no. 1, pp. 51–72., doi:10.1353/rhe.0.0035.