Empirical Topics in Domestic Public Finance

Discrimination of Race in the Labor Market

Şebnem Sera Uysal
Matriculation number: 12629559
Supervisor: Prof. Francis Wong
Winter Semester, 2024/25

Ludwig Maximilian University of Munich
Department of Economics

November 18, 2024[1]

[1] 47921 characters

1. Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

1.1. Introduction

The purpose of Bertrand and Mullainathan's study (2004) is to explore whether racial discrimination continues to play a significant role in the US labor market. Specifically, they aim to determine whether employers treat job applicants differently based solely on the perceived race indicated by their names, even when applicants have identical qualifications. The research question they seek to answer is: *To what extent does the racial perception of a name impact callback rates for job applicants with equivalent qualifications?*

To address this question, they conducted a field experiment between 2001 and 2002 by sending nearly identical fictitious resumes to real job ads in Boston and Chicago. Names perceived as White or African-American were randomly assigned to these resumes to isolate the effect of perceived race on hiring decisions. They specifically examined whether applicants with African-American-sounding names received fewer callbacks for interview than those with White-sounding names, as a lower callback rate would indicate racial discrimination in hiring practices, suggesting that biases may persist even when qualifications are equal.

During the experiment, they sent around 5,000 resumes in response to 1,300 job ads across various job categories, including sales, administrative support, clerical and customer service. They also investigated whether African-American applicants experience diminished returns to quality indicators on their resumes, such as additional work experience or specific skills, compared to their White counterparts. This aspect of the study aims to determine if higher credentials help African-American applicants overcome discrimination or if racial biases persist regardless of improved qualifications. Their findings reveal significant racial disparities in the job application process, which suggests that African-American applicants face barriers that are not alleviated by improving their credentials alone. Overall, the study provides evidence of structural racial bias in hiring practices and highlights the limitations of relying solely on improved qualifications to overcome these biases, emphasizing the need for systemic changes to address discrimination effectively.

1.2. Data

The data used in their study consists of nearly identical resumes, crafted from actual job seekers' resumes but substantially modified to anonymize content. The resumes were identical in qualifications but differed only in the names assigned, which signal perceived race, i.e., either African-American or White. The authors selected names using birth certificate data to clearly indicate race without directly stating it, choosing those most commonly associated with African-Americans or Whites. They validated these names through a survey in Chicago to ensure respondents identified the expected racial group. Names were then randomly assigned to the resumes, allowing the study to isolate the effect of perceived race on hiring decisions. This setup enabled a controlled comparison to determine how racial identity influences the likelihood of employers' callbacks.

Compared to other studies on labor market discrimination, which often rely on quasi-experimental methods using observational data or conducting in-person field experiments (known as audit studies), their field experiment offers several advantages. Unlike observational data, which can be confounded by unobserved variables, this dataset allows for a clearer isolation of race as the influencing factor on hiring outcomes. Since resumes were randomly assigned African-American or White-sounding names, any difference in callback rates can be attributed to the perceived race of the applicant, which offers strong causal evidence of racial bias. Using real job advertisements enhances external validity by replicating a realistic hiring scenario and reflecting how employers behave in actual recruitment processes.

By using fictitious resumes, the study also avoids the issue of participants consciously or unconsciously adjusting their behavior. Unlike audit studies, where participants are aware that they are part of an experiment and thus might unconsciously alter their behavior, employers in their experiment were completely unaware that they were being studied. This "double-blindness" enhances the reliability of their results. Furthermore, the relatively low cost of sending resumes, compared to the high costs associated with audit studies, allowed the researchers to generate a larger sample. Overall, these lead to a more controlled and cost-effective approach to studying racial discrimination, addressing key weaknesses in previous research.

However, there are notable disadvantages to using fictitious resumes. The standardized resumes, while effective for isolating the effect of race, limit the ability to capture the diversity seen in real-world applicants. Real job seekers present varied experiences, skills, educational backgrounds and unique resume details, such as the name of educational institutions or specific job histories, that may influence employer perceptions beyond race. This uniformity can oversimplify the complexity of how discrimination operates in real hiring practices and restrict the generalizability of their findings. Furthermore, since the study focuses solely on Boston and Chicago, the findings may not represent hiring discrimination practices across the broader US labor market, where racial dynamics, labor market conditions and hiring practices can vary significantly from region to region. These geographic limitations affect the generalizability of the conclusions to the labor market nationwide. This gap was later addressed by Kline et al. (2022), who broadened their study to include more firms and achieve greater geographical coverage, allowing for a more comprehensive analysis of racial discrimination in the US labor market.

Additionally, the dataset only measures the initial stage of the hiring process, i.e., callbacks for interviews, without extending to subsequent stages such as job offers, wage negotiations or career advancements. As a result, it cannot fully assess the extent of racial discrimination throughout the employment cycle. The study also relies solely on names to signal race, assuming that all employers interpret these names as clear racial indicators. In reality, individual perceptions can vary and some employers may associate names with socioeconomic status, potentially leading to inconsistencies in racial signaling.

On the other hand, they exclusively used newspaper ads for job applications, and this does not account for other common job-seeking methods, such as networking or online platforms. If African-Americans rely more heavily on social networks to find jobs, this approach might miss a significant aspect of their job-search experience. Finally, the callback rate in their study is relatively low. Kline et al. (2022) report a callback rate that is three times higher than that observed in Bertrand and Mullainathan's (2004) study. This discrepancy may suggest that the fictitious resumes used in their study were not viewed as well-suited or competitive candidates by employers, which could have influenced the study's findings.

1.3. Empirical Strategy

The empirical strategy in this study utilized a correspondence testing methodology, which involves sending fictitious resumes, each with a randomly assigned name, to job advertisements. This ensures that any systematic differences in callback rates can be attributed solely to the perceived race signaled by the names. Their primary analysis involved comparing average callback rates between resumes with White-sounding names and those with African-American-sounding names. This comparison was conducted across the entire sample and within various subsamples such as city (Boston and Chicago) and applicant gender (male and female).

The study further explored how the racial gap in callback rates varies with resume quality. The idea is to see whether improvements in resume quality resulted in the same increase in callback rates for African-American applicants as they did for White applicants[2]. For this analysis, the key identifying assumption is that assigning high and low-quality resumes to names is random. Initially, they compared callback rates using a subjective classification of resume quality based on predefined criteria, such as work experience and additional skills, for both racial groups. To provide a more objective assessment, they used a predicted measure of quality by estimating separate probit regressions for White and African-American-sounding names. These regressions used a random subsample of one-third of the resumes, where the callback dummy was regressed on resume characteristics and control variables such as gender, city, occupation dummies and job requirements (e.g., education, experience and communication skills). The coefficients obtained from these probit models were then applied to the remaining two-thirds of the sample to predict

[2] As previously mentioned, the authors previously classified resumes into two quality levels: higher-quality and lower-quality. Higher-quality resumes had more favorable characteristics like additional work experience, fewer employment gaps, volunteering or military experience, email addresses, and special skills (e.g., computer proficiency and certifications).

callback likelihoods for each racial group separately. In this context, resumes classified as "high-quality" are those predicted to rank above the median callback likelihood, while "low-quality" resumes fall below this threshold.

In a probit model, the probability that resume *j* for job *i* receives a callback is represented as:

$$P(\text{ Callback}_{ij} = 1) = \Phi(\beta_0 + \beta_1 \text{Resume Characteristics}_{ij} + \beta_2 \text{Sex Dummy}_{ij} + \beta_3 \text{City Dummy}_{ij} + \sum_{k=4}^{9} \beta_k \text{Occupation Dummies}_{ijk} + \beta_{10} \text{Job Requirements}_{ij})$$

where:

- ϕ is the cumulative distribution function of the standard normal distribution, which transforms the linear combination of predictors into a probability between 0 and 1.
- $P(Callback_{ij} = 1)$ is the probability that resume *j* for job *i* receives a callback.
- $Resume\ Characteristics_{ij}$ is a vector of variables that describe the attributes of the resume (e.g., years of experience, education level, special skills).
- $Sex\ Dummy_{ij}$ is a binary variable indicating the gender of the applicant.
- $City\ Dummy_{ij}$ is a binary variable representing the city where job *i* is located (e.g., 1 for Boston, 0 for Chicago).
- $\sum_{k=4}^{9} \beta_k \text{Occupation Dummies}_{ijk}$ is a set of binary variables indicating the type of job or occupation (e.g., administrative, sales), each with its own coefficient.
- $Job\ Requirements_{ij}$ is a vector that includes specific job requirements mentioned in the ad (e.g., required communication skills, minimum education).

The main coefficient of interest, β_1, reflects the effect of specific resume characteristics on the probability of receiving a callback within each racial group. A positive and statistically significant β_1 suggests that these characteristics enhance the likelihood of a callback for that group.

Random assignment of names and resume quality ensures comparability across groups, aligning with standard experimental practices and strengthening the credibility of causal interpretations. This randomization helps isolate the impact of perceived race and resume quality on callback rates, supporting the validity of their findings. This design also assumes that observed differences in callback rates are mainly driven by racial cues in names and resume quality, an expectation that seems credible given the study's carefully controlled methodology and experimental setup.

1.4. Results

The main findings of the study, as shown in Table 1, reveal significant racial discrimination in the hiring process, where resumes with White-sounding names receive 50% more callbacks than those with African-American-sounding names, even when all other resume characteristics are equal. This difference suggests that White applicants would receive a callback after applying to about 10 jobs, while African-American applicants would need to apply to approximately 15 to get one callback. The racial gap in callback rates persists across various settings, including cities, job types and gender, though the degree of disparity differs slightly across these dimensions.

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (p-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Notes: The table reports, for the entire sample and different subsamples of sent resumes, the callback rates for applicants with a White-sounding name (column 1) an an African-American-sounding name (column 2), as well as the ratio (column 3) and difference (column 4) of these callback rates. In brackets in each cell is the number of resumes sent in that cell. Column 4 also reports the p-value for a test of proportion testing the null hypothesis that the callback rates are equal across racial groups.

Table 1. Mean callback rates by racial soundingness of names

Source: Bertrand and Mullainathan (2004)

In terms of resume quality, the analysis finds that higher-quality resumes increase the likelihood of receiving a callback for both groups; however, the improvement is less pronounced for resumes with African-American-sounding names, as seen in Table 2. For example, high-quality resumes with White-sounding names see a callback increase of 6.42 percentage points (pp, 1.89 times higher) compared to low-quality resumes, whereas resumes with African-American-sounding names show a smaller increase of 3.23 pp (1.60 times higher). This shows that improving qualifications or acquiring additional skills benefits African-American applicants less and fails to bridge the racial gap in callbacks.

	Panel A: Subjective Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (p-value)
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)

	Panel B: Predicted Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (p- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Notes: Panel A reports the mean callback percents for applicant with a White name (row 1) and African-American name (row 2) depending on whether the resume was subjectively qualified as a lower quality or higher quality. In brackets is the number of resumes sent for each race/quality group. The last column reports the p-value of a test of proportion testing the null hypothesis that the callback rates are equal across quality groups within each racial group. For Panel B, we use a third of the sample to estimate a probit regression of the callback dummy on the set of resume characteristics as displayed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). We then use the estimated coefficients on the set of resume characteristics to estimate a predicted callback for the remaining resumes (two-thirds of the sample). We call "high-quality" resumes the resumes that rank above the median predicted callback and "low-quality" resumes the resumes that rank below the median predicted callback. In brackets is the number of resumes sent for each race/quality group. The last column reports the p-value of a test of proportion testing the null hypothesis that the callback percents are equal across quality groups within each racial group.

Table 2. Average callback rates by racial soundingness of names and quality

Source: Bertrand and Mullainathan (2004)

Moreover, the study demonstrates that racial disparities remain consistent across a range of industries, job categories and employer sizes. Even when accounting for neighborhood characteristics, the racial gap remains; while living in a better neighborhood affects callback rates, it does not provide more advantage to African-Americans than to Whites.

1.5. Contribution and Implications

Racial discrimination in hiring processes has been a persistent area of research, with early studies primarily relying on regression analyses of survey data to detect disparities between racial groups (Altonji and Blank, 1999). These studies faced challenges due to unobservable variables and potential biases that made it difficult to isolate discrimination as the cause of hiring disparities. Experimental methods were introduced to overcome these challenges, with Bertrand and Mullainathan (2004) pioneering the use of correspondence testing methodology, where they revealed significant discrimination in the early hiring stages based on racial signals conveyed by names. By introducing the correspondence testing method in the US labor market, they provided a clearer, more controlled measure of discrimination, offering robust, experimental evidence that has since shaped the design of similar studies globally. Therefore, their approach laid the groundwork for many subsequent studies that analyze racial and ethnic discrimination in diverse settings and countries[3].

Discrimination in hiring is difficult to measure, especially as it can be subtle or systemic, embedded in company practices. Correspondence testing has proven highly effective in detecting and quantifying discrimination, allowing researchers to identify potential biases and produce evidence that informs anti-discrimination policies. Over the past decade, these studies conducted worldwide have consistently shown racial and ethnic discrimination in hiring.

In the US, numerous studies have built on Bertrand and Mullainathan's findings, investigating variations in discrimination across different settings and applicant characteristics. For instance, Nunley et al. (2015) conducted a large-scale correspondence experiment in seven US cities, targeting recent college graduates. Their results indicated that Black-sounding names received 14% fewer callbacks than White-sounding names, especially for customer-facing roles. Their findings suggest that recruiters might not themselves have racial biases but anticipate such biases from customers.

Jacquemet and Yannelis (2012) investigated discrimination patterns in Chicago, focusing on the concept of 'ethnic homophily', where employers tend to favor candidates from the majority ethnic group over those perceived as ethnically distinct. Using Anglo-Saxon, African-American, and ethnically ambiguous names, they found discrimination in predominantly White suburban areas. This trend was partly attributed to the 'white flight' phenomenon, where White residents historically moved away to suburban areas as minority populations grew. Their findings suggest that employers were more likely to discriminate against non-majority names, emphasizing how geographic and historical factors can shape hiring biases.

Using large-scale correspondence testing, Kline et al. (2022) recently identified racial discrimination patterns across US companies and industries. Their study revealed significant variation in callback rates across firms, with some employers exhibiting a marked "contact gap" that favored White applicants over Black applicants, regardless of location. The findings suggest that racial biases are often systemic within certain firms, embedded in company culture or hiring practices. This form of systemic discrimination limits opportunities for Black applicants, even when equally qualified and contributes to racial disparities in employment.

While all of these studies reveal clear empirical evidence of systemic discrimination in hiring practices, Darolia et al. (2016) offered a different perspective by examining discrimination against African-American, Hispanic and White applicants across seven US states. Deviating from Bertrand and Mullainathan's approach, they used racially ambiguous names rather than stereotypically African-American-sounding names. Their findings revealed minimal differences

[3] For further details on correspondence experiments, Baert (2018) offers an extensive review of nearly all correspondence studies since 2005, categorizing them by methodology, findings and context, providing an invaluable resource.

in employer responses across racial groups. This indicated little evidence of systematic discrimination based on race or gender, suggesting that racial biases may have diminished over time or that the lack of strong socioeconomic signals reduced perceived differences among applicants.

Other studies have found similar patterns of discrimination in different labor markets. For instance, Oreopoulos (2011) extended Bertrand and Mullainathan's study to the Canadian labor market, examining biases against applicants with foreign-sounding names. His findings revealed significant name-based discrimination that persisted despite Canada actively promoting immigration to attract skilled workers. Banerjee et al. (2018) further analyzed Oreopoulos's dataset to explore the role of firm size, finding that larger organizations exhibited slightly lower discrimination, possibly due to more structured hiring practices, though biases remained across all firm sizes. Similarly, Carlsson and Rooth (2007) used a similar approach in Sweden to compare callbacks for applicants with Swedish and Middle Eastern names and their results demonstrated a clear bias favoring Swedish names, with callbacks notably higher in lower-skilled positions and at smaller firms.

Kaas and Manger (2012) investigated similar patterns in Germany by examining job advertisements for internship positions, comparing callback rates for applicants with German and Turkish names. Their findings showed that applicants with German names were significantly more likely to receive callbacks, with discrimination especially pronounced in smaller firms. Notably, callback rates for Turkish applicants improved when references were included on resumes, suggesting a form of statistical discrimination where employers may perceive Turkish applicants as less productive without additional verification of qualifications[4]. Similarly, Drydakis and Vlassis (2010) used a correspondence test to examine ethnic discrimination in Greece, finding that low-skilled Albanian applicants faced significantly lower callback rates than Greek applicants. Moreover, their study uncovered evidence of wage discrimination, with Albanian candidates receiving lower wage offers and reduced insurance coverage, highlighting both hiring and post-hiring discrimination against ethnic minorities.

Collectively, the literature provides strong evidence of racial discrimination across various labor markets, suggesting that this issue is deeply rooted and persistent. Despite policy efforts aimed at addressing racial disparities through skill development programs, the findings suggest that these traditional approaches alone may fall short. Bertrand and Mullainathan's (2004) study is particularly instructive, as it reveals that racial bias affects hiring decisions even when the qualifications of applicants are identical, indicating that discrimination is structural rather than merely reflective of differences in skills. This reduces the returns on resume quality enhancements for African-American applicants, discouraging further investment in skill

[4] Taste-based discrimination occurs when employers, coworkers or customers have a preference or "taste" against working with or hiring individuals from certain groups, even if it reduces efficiency or profits. Statistical discrimination, on the other hand, arises when employers use group-level characteristics, such as race or gender, as proxies for individual productivity due to incomplete information about applicants.

development. These findings call for a reassessment of policy interventions, emphasizing the need for targeted and structural measures that address the implicit biases within hiring processes, i.e., hiring practices themselves are fair and free from hidden prejudices that disadvantage minority applicants.

Further reinforcing the need for such interventions, Bertrand and Mullainathan's study introduces the lexicographic search model as a potential mechanism for discrimination. This model suggests that employers might make immediate judgments based on an applicant's name, stopping their review process without considering the applicant's full qualifications. Such a screening method could explain the lower returns to credentials observed for African-American applicants, as employers may not assess candidates' skills comprehensively. This finding underscores the importance of encouraging hiring practices that promote a thorough review of all applicant qualifications rather than allowing implicit biases to impact decisions.

1.6. Critique

The study by Bertrand and Mullainathan (2004), while pioneering in its approach, has several limitations, some of which the authors themselves acknowledged. One key limitation is the assumption that employers' callback decisions directly reflect their hiring intentions and accurately represent overall interest in candidates. This implies that receiving a callback signals real hiring intent; however, while callbacks are an initial and necessary step in the hiring process, they do not fully capture employer preferences throughout the entire process and fail to reveal who ultimately receives the job and salaries offered. This aspect is important because, even if equal treatment is observed in callback rates, it does not necessarily indicate an absence of discrimination in the later stages. Discrimination might still occur during interviews, salary negotiations or when assessing candidates with varied qualifications, which were not part of their study. Although the authors argue that fewer callbacks likely lead to fewer job offers, this link remains untested.

Moreover, discrimination in hiring processes is only one aspect of discriminatory practices in the labor market. For example, their study does not address on-the-job discrimination, such as disparities in wage rises, promotions or job retention, which would provide a more comprehensive view of racial discrimination. Indeed, these forms of discrimination may be more severe and pervasive than what is evident at the hiring stage alone. Consequently, the study's focus on callbacks may lead to an underestimation of the full extent of workplace discrimination over the course of an employee's career.

Additionally, the study relies on racially distinct names to suggest race rather than explicitly stating the applicant's racial identity. This approach introduces certain limitations. Not all employers may recognize their racial content correctly. Some might overlook the names or fail to associate them with a specific racial group, potentially undermining the intended signal.

Furthermore, the results may not be generalizable to all African-American applicants, as not all have names strongly associated with their racial group. Another limitation is that some employers might associate certain names with socioeconomic background rather than race.

Moreover, the study's scope is limited to the private sector, focusing only on callbacks from privately held and non-profit organizations. This approach overlooks public sector employment, which may operate under different regulations and hiring practices that could influence discriminatory behavior. By not including public sector jobs, the study does not provide insights into whether discrimination dynamics differ between public and private sectors, potentially missing an important part of understanding labor market discrimination.

Lastly, it does not fully align with the current digital nature of job applications. Today, most firms publish job vacancies online and applicants submit their resumes through digital platforms. While the method employed by Bertrand and Mullainathan was novel at the time, its relevance may have diminished with the rise of online application systems. More recent studies, such as Kline et al. (2022), utilized online job portals to distribute resumes, reflecting the current state of the job market more accurately.

Future research could address these weaknesses through various approaches. First, to overcome the limitation of focusing solely on callback rates, future studies could track subsequent stages of the hiring process, such as job offers and salary negotiations. This would provide a more comprehensive understanding of discrimination across the entire hiring procedure. Second, instead of relying exclusively on racially distinct names, future studies could explicitly signal racial identity, for example, through photos or self-reported race, to investigate whether clearer racial identification influences employer behavior. Third, expanding the scope to include public sector jobs and a more geographically diverse sample beyond Boston and Chicago would improve the generalizability of findings.

Additionally, data collection should go beyond hiring outcomes to capture disparities in wage progression, promotions and job retention, enabling a more comprehensive assessment of discrimination. Finally, future research could evaluate the effectiveness of policies aimed at reducing discrimination in the labor market, such as name-blind recruitment or pay transparency laws recently implemented in various US states and other countries. Empirical assessments of these interventions could offer valuable insights for policymakers, where we will explore the potential of pay transparency policies to reduce racial wage gaps in the next section.

2. The effect of pay transparency policies in reducing racial discrimination in the US labor market

2.1. Research Question and Existing Literature

Black workers in the US have consistently faced significant labor market disparities, with unemployment rates roughly double those of White workers across all education levels and demographics, a trend that has persisted since the 1970s. Even during severe economic downturns, the Black unemployment rate frequently exceeds 10%, a threshold never reached by White workers (Maye, 2023). Strikingly, higher education offers little relief, as Black workers with advanced degrees often experience higher unemployment rates than less-educated White workers. Similarly, wage disparities persist, with White households earning nearly twice the median income of Black households. Despite controlling for factors such as education, experience and location, much of the Black-White wage gap remains unexplained, pointing to systemic discrimination and unequal bargaining power. Recent data from the US Department of Labor reveal that Black workers earn only $0.76[5] for every dollar earned by White workers, a disparity that has widened over time with no signs of narrowing (Gudell, 2023). These enduring inequities underscore the systemic barriers shaping unequal labor market outcomes and emphasize the need for targeted policy interventions.

One promising approach to addressing these disparities is the adoption of pay transparency laws. These policies aim to reduce structural barriers and promote fairness by requiring employers to disclose salary information to job candidates and current employees, fostering greater workplace equity. Some measures go further by prohibiting retaliation against employees who discuss their pay, requiring employers to provide pay data to government agencies or banning salary history inquiries from candidates. By enhancing visibility into pay structures, these laws empower workers to negotiate more effectively and hold employers accountable. Through these mechanisms, pay transparency laws hold the potential to reduce wage disparities and address systemic inequities in the US labor market.

Several US states have recently enacted pay transparency laws to address wage gaps and promote equity in the labor market. Maryland was among the first to implement its law in October 2020, followed by Nevada and Colorado in early 2021, with more recent adoptions in New York (2023) and Hawaii (2024). States such as Minnesota, Illinois, Massachusetts and Vermont have also passed laws that will take effect between 2025 and 2026, reflecting the continuing momentum of this policy trend. By increasing transparency, these laws aim to combat systemic inequities and reduce wage gaps in the US labor market. Despite their increasing adoption at the state level, there is no comprehensive federal pay transparency legislation, resulting in significant variation in implementation across the US.

[5] Available at US Department of Labor, Office of Federal Contract Compliance Programs.

This study takes a step forward by moving beyond the hiring stage to examine the impact of pay transparency policies on current workers in the US labor market. Unlike correspondence studies or field experiments, which are often used to detect discrimination but cannot directly measure wage and employment outcomes, this study leverages observational data to evaluate whether these policies reduce racial disparities. Specifically, the research seeks to answer this key question: *Have state-level pay transparency laws successfully narrowed the wage gap between Black and White workers?* By examining the effects of these laws, such as mandating salary disclosures, encouraging pay negotiations and addressing structural barriers, this study aims to evaluate their role in fostering equity and addressing the long-standing racial wage gap.

The key hypothesis is that pay transparency policies, by mandating the disclosure of salary information for existing employees and in publicly advertised job postings, can reduce racial wage gaps and promote greater equity in the labor market. Rooted in the principle of equal pay for equal work, these laws offer significant potential to address systemic discrimination by increasing the visibility of wage disparities, thereby shifting the bargaining power of historically underpaid groups, such as Black workers, in their favor.

The literature on pay transparency policies is relatively recent, reflecting the increasing adoption of these measures in the US and globally. Although evidence of their direct impacts is still emerging, early studies provide meaningful observations. For example, Arnold et al. (2022) analyzed Colorado's 2021 pay transparency law through a dynamic difference-in-differences (DiD) framework and found a 3.6% increase in posted salaries. Moreover, the study identified general equilibrium effects, indicating that even firms already compliant with pay transparency practices raised their wages, highlighting the broader economic impact of these policies.

Other studies emphasize the potential of pay transparency to address gender wage disparities. Chen et al. (2023) demonstrated that salary range disclosures, especially narrow ones, encourage women to apply for jobs and negotiate higher salaries, suggesting that these policies could reduce gender wage gaps. Similarly, Frimmel et al. (2024) examined mandatory wage postings in Austria and observed a slight reduction in gender wage gaps, particularly for women in lower wage brackets, without negatively impacting overall wages. Cullen (2024) extended the literature by exploring different forms of pay transparency, i.e., horizontal (within firms), vertical (across seniority levels) and cross-firm. The study found that horizontal transparency empowers underpaid employees to negotiate fairer wages, while vertical transparency enhances motivation by clarifying career progression opportunities. Together, these studies demonstrate the promise of pay transparency policies in promoting equity. However, further research is needed to assess their broader impacts, particularly on different demographic groups and economic contexts.

2.2. Empirical Strategy

This study evaluates the impact of state-level pay transparency policies on racial wage disparities in the US using a stacked DiD approach. Initially proposed by Cengiz et al. (2019) and refined by Wing et al. (2024), this methodology is particularly well-suited for analyzing staggered policy adoptions, such as pay transparency laws introduced at different times across US states. The stacked DiD approach mitigates potential biases in traditional DiD designs by addressing challenges posed by time-varying treatment effects and heterogeneous treatment timings. Traditional DiD assumes that all treated units experience the policy at the same time, which is not the case for pay transparency laws introduced in different years across states. This mismatch can lead to biased estimates, especially if the treatment effects vary over time or earlier and later adopters differ significantly in ways that affect outcomes. By structuring the analysis into separate sub-experiments and focusing on short-term effects within consistent time windows, the stacked DiD design accounts for these variations, ensuring more reliable estimates of the policy's impact.

In the stacked DiD, the analysis is divided into multiple sub-experiments, each centered on a specific state's policy adoption. For each sub-experiment, treated states are compared to "clean controls," which remain untreated during the relevant time window. These sub-experiments are then concatenated into a single stacked dataset, allowing for causal validity by isolating the policy's short-term effects.

Additionally, this study applies corrective weights, an improvement introduced by Wing et al. (2024). While Cengiz et al.’s stacked DiD framework effectively isolates treatment effects by creating sub-experiments, it does not account for imbalances caused by differences in sample sizes or temporal proximity among sub-experiments. Wing et al. enhanced this approach by introducing corrective weights to ensure that each sub-experiment contributes proportionally to the overall analysis. Without these weights, states with larger sample sizes or states whose policy adoption occurred closer in time to others might disproportionately influence the results. Corrective weights address this imbalance.

The key identifying assumption of this framework is the parallel trends assumption, which assumes that in the absence of pay transparency policies, the wage of workers in treated states would have evolved similarly to the wage of workers in control states. To validate this assumption, pre-treatment event windows are included in the analysis, and statistical insignificance of the pre-treatment coefficients ($\beta_{\tau<0}$) confirms its validity. These results are reported in Table 4 and can also be seen in Figure 1 at the end of the paper.

The estimating equation is as follows:

$$\log(\text{ Wage diff}_{ijt}) = \sum_{\tau=-\kappa_{\text{pre}}}^{\kappa_{\text{post}}} \beta_{\tau} \cdot I_{\tau,ijt} + \mu_{ij} + \rho_{jt} + X_{ijt}\gamma + \varepsilon_{ijt}$$

where

- dependent variable is the logarithm of the wage differential between year t and year $t-1$ for individual i in state j, capturing changes in wage growth due to pay transparency policies.
- $I_{\tau, ijt}$ is an indicator variable for event time τ relative to policy adoption, where τ = 0 is the treatment year, τ < 0 are pre-treatment years, and τ > 0 are post-treatment years.
- μ_{ij} is the individual-by-state fixed effects, controlling for unobservable time-invariant characteristics of individuals and states. For example, if some states historically have higher wages regardless of the treatment, these fixed effects absorb that variation.
- ρ_{jt} is the state-by-year fixed effects, capturing state-specific trends over time
- X_{ijt} are the vector of covariates that control for gender, marital status, tenure, education level, job category, whether individual i is located in the South region and hours worked per week.
- ε_{ijt} is the error term.

The coefficient of interest, $\beta_{\tau > 0}$, measures the average treatment effect of pay transparency policies on wage growth disparities after policy adoption (ATT). A positive and significant $\beta_{\tau > 0}$ indicates that the policies improved wage growth for the treated group. Conversely, an insignificant or negative $\beta_{\tau > 0}$ would suggest limited or no impact.

To ensure compositional balance and avoid misleading conclusions, trimming is applied by defining fixed pre and post-treatment windows (κ_{pre}, κ_{post}). Only states with sufficient pre and post-treatment data are included, ensuring clean comparisons and causal validity.

To further explore the policies' impacts across different demographics, the analysis examines subgroups defined by race (e.g., Black vs. White workers). By comparing $\beta_{\tau > 0}$ coefficients for Black workers to those for White workers, the analysis evaluates whether pay transparency laws disproportionately benefit historically underpaid groups.

2.3. Data

This study employs observational data covering 2010–2024, focusing on US states that implemented pay transparency laws during this period. The treated states consist of ones that adopted such laws, including early adopters like Maryland (2020), Nevada and Colorado (2021), as well as more recent adopters such as California and Washington (2023). States that enacted pay transparency policies recently, such as New York and Hawaii, are excluded due to insufficient post-treatment observations necessary for the analysis.

The primary data source for this study is microdata from the Economic Policy Institute (EPI, 2024), derived from the Current Population Survey (CPS). This dataset offers comprehensive individual-level information on employment, wages and demographic characteristics, enabling a detailed examination of factors contributing to racial wage disparities. It also provides sufficient coverage and consistency to analyze trends across multiple states and over time.

To focus the analysis on the direct impact of pay transparency laws, the sample is restricted to currently employed workers in the private sector, whereas individuals employed in the public sector, state or federal government are excluded. Furthermore, the analysis is limited to workers who remained with the same employer as in the previous month to better capture the effects of pay transparency policies on ongoing employment relationships. The resulting dataset provides a robust foundation for examining the causal effects of pay transparency laws on racial wage gaps between Black and White workers in the US labor market.

2.4. Hypothetical Results

Our analysis reveals that the implementation of state-level pay transparency policies has led to an average wage increase of 3.5%, with Black workers experiencing a larger benefit than White workers, thereby significantly reducing the racial wage gap in the US. More specifically, we observe a 6.7% average wage rise among Black workers, highlighting the significant positive effect of these policies on historically underpaid groups. In contrast, White workers experienced a 1.3% wage increase, substantially smaller than the benefit observed for Black workers. This disparity suggests that the policies have effectively contributed to narrowing the racial wage gap.

Moreover, our event-study analysis shows a pronounced immediate effect in the year of policy implementation, with a 9.6% wage increase for Black workers and a 2.4% increase for White workers. These effects persist over time but diminish slightly in magnitude in subsequent years.

Furthermore, our analysis suggests that the policies are most effective for Black women. These findings emphasize the potential of pay transparency policies to benefit vulnerable groups in the labor market. Factors such as marital status and geographic location (urban/rural) appear to have limited influence on the policy's effects. However, education and tenure positively correlate with wage increases, indicating that workers with these characteristics are better positioned to leverage the opportunities created by pay transparency laws.

2.5. Discussion

The results of this study provide meaningful contributions to academic research and policy discussions on labor market discrimination. First, the results highlight the effectiveness of pay transparency policies in addressing systemic disparities, particularly in reducing the racial wage gap. The findings underscore that economic behavior can be influenced through greater transparency as workers gain better information for negotiations. This expands our understanding of how policy interventions can reshape labor market dynamics. Additionally, this study addresses a key limitation in Bertrand and Mullainathan (2004), which primarily identified discrimination at the hiring stage but did not evaluate its later effects on wage disparities. By focusing on post-hiring wage gaps, we extend the analysis to uncover how policy interventions can alter existing compensation structures.

From a policy perspective, the findings emphasize the importance of implementing and expanding pay transparency measures to enhance wage equity. The significant benefits observed for Black workers, particularly in narrowing the racial wage gap, suggest that these laws are a valuable tool for promoting fairness in compensation structures. Moreover, the persistence of wage increases over time for historically underpaid groups indicates the potential for long-term benefits of such policies. However, there is also a need for complementary interventions to address limitations, such as occupational segregation and broader systemic barriers.

Policy implications extend to encouraging federal-level adoption of pay transparency laws to standardize their application across states, ensuring consistency in outcomes and closing disparities in regions without such legislation. Additionally, the findings point to the utility of broader data reporting requirements, such as pay data disclosures linked to demographic characteristics, which could further enhance accountability and equity in wage practices[6]. Overall, this study supports the notion that targeted transparency initiatives can create fairer labor market conditions and reduce racial inequities.

Finally, while the immediate benefits are clear, further research is necessary to evaluate these policies' long-term effects and applicability across diverse labor market contexts. This study provides a compelling argument for pay transparency as a critical tool to advance racial equity and economic efficiency.

[6] One type of progress would be the possible reinstatement of EEO-1 Component 2, which mandates employers to submit employee compensation and demographic data to the US Equal Employment Opportunity Commission (EEOC) (Madden & Wiggins, 2024).

2.6. Tables and Figures

Variable Name	Description
Age	Age of respondent
Female	A dummy variable for gender (0 = male, 1 = female)
Married	A dummy variable for marital status (0 = Not married, 1 = Married)
Race	Indicator for race (1 = white, 2 = black)
Education	Education level (1 = Less than high school, 2 = High school, 3 = Some college, 4 = College, 5 = Advanced)
Metropolitan	Whether or not the household resides in a metropolitan area (urban dummy)
State-FIPS	FIPS code for the state of the household
South	Whether or not the household resides in the South region of U.S.
Hours worked	Number of hours worked last week at the respondent's primary job
Person ID	A person identifier, unique within year, month and household identifier
Year	Calendar year of the interview, generated from the survey year
Wage	Hourly wage in dollars (adjusted)
Occupation category	10-category recode of the major occupation of the job of the respondent
Tenure	Number of years worked for the current employer

Table 3: List of variables
Source: Economic Policy Institute (2024)

Dependent Variable:	Wage difference (log)		
	Full sample (1)	Black population (2)	White population (3)
A. Average Treatment Effect on Treated (ATT)			
Treated (=1) × Post (=1)	0.035***	0.067***	0.013***
	(0.005)	(0.012)	(0.007)
B. Event-studies			
Treated (=1) × Event-time, -5 (=1)	0.0050	0.0081	0.0025
	(0.0138)	(0.0270)	(0.0091)
Treated (=1) × Event-time, -4 (=1)	0.0027	-0.0191	0.0178*
	(0.0080)	(0.0147)	(0.0082)
Treated (=1) × Event-time, -3 (=1)	0.0129	-0.0040	0.0254
	(0.0094)	(0.0141)	(0.0143)
Treated (=1) × Event-time, -2 (=1)	-0.0190	-0.0005	-0.0317**
	(0.0131)	(0.0243)	(0.0107)
Treated (=1) × Event-time, 0 (=1)	0.0531***	0.0958***	0.0244*
	(0.0114)	(0.0208)	(0.0115)
Treated (=1) × Event-time, 1 (=1)	0.0412**	0.0647**	0.0253**
	(0.0137)	(0.0221)	(0.0112)
Treated (=1) × Event-time, 2 (=1)	0.0445**	0.1082**	0.0013
	(0.0168)	(0.0350)	(0.0209)
Age	0.0054**	0.0035**	0.0075**
	(0.0039)	(0.0068)	(0.0073)
Metropolitan	-0.0011	0.0085	-0.0075
	(0.0039)	(0.0068)	(0.0073)
Tenure	0.004**	0.003*	0.005**
	(0.0004)	(0.0003)	(0.0006)
Education	0.005**	0.002**	0.006**
	(0.0029)	(0.0055)	(0.0034)
Married	0.0150	0.0209	0.0115
	(0.0048)	(0.0074)	(0.0044)
Female	0.079**	0.084**	0.077**
	(0.019)	(0.018)	(0.011)
South	-0.003	-0.008	-0.001
	(0.0019)	(0.0018)	(0.0011)
Occupation category	-0.0003	0.0009	-0.0012
	(0.0018)	(0.0021)	(0.0020)
Fixed-effects			
State-by-year FE	Yes	Yes	Yes
Individual-by-state FE	Yes	Yes	Yes
Fit statistics			
Observations	228,726	92,228	136,498
R^2	0.00046	0.00117	0.00061
Within R^2	0.00037	0.00087	0.00049

Standard errors are clustered at the state level and reported in parentheses.
*Sign. levels: ***: 0.01, **: 0.05, *: 0.1*

Table 4: Hypothetical results
Source: own study

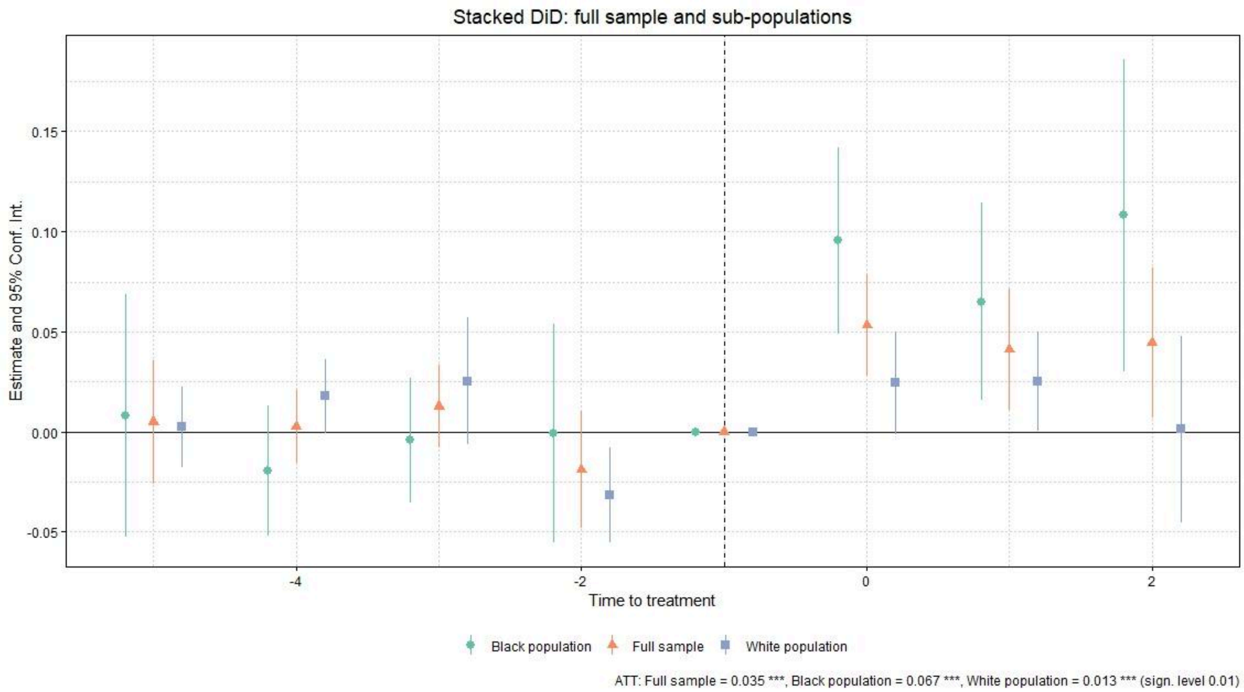

Figure 1: Hypothetical results
Source: own study

References

Altonji, Joseph G., and Rebecca M. Blank. "Race and gender in the labor market." Handbook of labor economics 3 (1999): 3143-3259.

Arnold, David, Simon Quach, and Bledi Taska. "The impact of pay transparency in job postings on the labor market." Available at SSRN 4186234 (2022).

Baert, Stijn. Hiring discrimination: "An overview of (almost) all correspondence experiments since 2005". Springer International Publishing, 2018.

Banerjee, Rupa, Jeffry Reitz, and Phil Oreopoulos. "Do large employers treat racial minorities more fairly." A new analysis of Canadian field experiment data. Retrieved from University of Toronto, Munk School of Global Affairs website: http://www. hireimmigrants. ca/wpcontent/uploads/Final-Report-Which-employers-discriminate-Banerjee-Reitz-Oreopoulos-January-25-2017. pdf (2017).

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 94.4 (2004): 991-1013.

Carlsson, Magnus, and Dan-Olof Rooth. "Evidence of ethnic discrimination in the Swedish labor market using experimental data." Labour economics 14.4 (2007): 716-729.

Cengiz, Doruk, et al. "The effect of minimum wages on low-wage jobs." The Quarterly Journal of Economics 134.3 (2019): 1405-1454.

Chen, Clara Xiaoling, Victoria Fung, and Lisa LaViers. "Labor Market Participants' Reactions to Salary Range Disclosures." Available at SSRN (2023).

Cullen, Zoë. "Is pay transparency good?." Journal of Economic Perspectives 38.1 (2024): 153-180.

Darolia, Rajeev, et al. "Race and gender effects on employer interest in job applicants: new evidence from a resume field experiment." Applied Economics Letters 23.12 (2016): 853-856.

Drydakis, Nick, and Minas Vlassis. "Ethnic discrimination in the Greek labour market: occupational access, insurance coverage and wage offers." The Manchester School 78.3 (2010): 201-218.

Economic Policy Institute. Current Population Survey Extracts, Version 1.0.57, 2024, https://microdata.epi.org.

Frimmel, Wolfgang, et al. External pay transparency and the gender wage gap. No. 07/24. RF Berlin-CReAM Discussion Paper Series, 2024.

Gudell, Skylar. "How Pay Transparency Can Help Close Wage Gaps in the Workplace." World Economic Forum, 25 Apr. 2023, www.weforum.org/stories/2023/04/how-pay-transparency-help-close-wage-gaps-in-workplace/.

Kaas, Leo, and Christian Manger. "Ethnic discrimination in Germany's labour market: A field experiment." German Economic Review 13.1 (2012): 1-20.

Kline, Patrick, Evan K. Rose, and Christopher R. Walters. "Systemic discrimination among large US employers." The Quarterly Journal of Economics 137.4 (2022): 1963-2036.

Madden, M., and Wiggins, L. "The Momentum and Future of Pay Transparency in the US" WTW, 4 Mar. 2024, www.wtwco.com/en-mu/insights/2024/02/the-momentum-and-future-of-pay-transparency-in-the-us.

Maye, Adewale A. "Chasing the dream of equity: how policy has shaped racial economic disparities." (2023).

Nunley, John M., et al. "Racial discrimination in the labor market for recent college graduates: Evidence from a field experiment." The BE Journal of Economic Analysis & Policy 15.3 (2015): 1093-1125.

Jacquemet, Nicolas, and Constantine Yannelis. "Indiscriminate discrimination: A correspondence test for ethnic homophily in the Chicago labor market." Labour Economics 19.6 (2012): 824-832.

Office of Federal Contract Compliance Programs. Earnings Disparities by Race and Ethnicity. US Department of Labor, n.d., www.dol.gov/agencies/ofccp/about/data/earnings/race-and-ethnicity.

Oreopoulos, Philip. "Why do skilled immigrants struggle in the labor market? A field experiment with thirteen thousand resumes." American Economic Journal: Economic Policy 3.4 (2011): 148-171.

Wing, Coady, Seth M. Freedman, and Alex Hollingsworth. Stacked difference-in-differences. No. w32054. National Bureau of Economic Research, 2024.

Overview of tools used

Product name: ChatGPT by OpenAI.

How it was used: Assisted in revising the language in this paper, including the data generation script (written in R).

Extent: Provided suggestions for the formal, stylistic and grammatical optimization of the text and the R script.

Purpose: To check and improve the clarity and quality of the language used.

I hereby declare that I have prepared this paper independently and without the use of aids other than those specified, that I have not yet submitted it to another examination authority and that it has not yet been published. In the case of the use of generative models for the creation of texts, illustrations, calculations and other services, I am fully responsible for the selection, adoption and all results of the generated output used by me. In the list "Overview of tools used" I have named all generative models used with their product names and indicated how, to what extent and for what purpose they were used.

18.11.2024