Wage Gaps and Discrimination

An Overview of Selected Literature

Presentation By Tanya Rajan, George Vojta, and Oscar Volpe

The University of Chicago

Week 4 - Winter Quarter 2021

Presentation Outline

Overview:

Although there were great gains in closing the racial and gender wage gaps in the mid-1900s, the labor-income-discrepancy has proven to be stubborn and by the end of the millennia the gap began to increase again between races and held firm between the white men and women.

Three alarming facts took hold in the end of the 1900s:

- Black males and white women made about $\frac{2}{3}$rds of what white men make in wages (with those other subgroups being worse off).
- This gap is even worse when you look at annual earnings.
- The black unemployment rate doubles the white unemployment rate and while white women unemployment is lower than that of white men, their participation in the labor force is lower (although increasing throughout the century).

Wages by Demographic - Data is 1996 CPS

Table 1
Labor market data by race and gender[a]

	All	White males	Black males	Hispanic males	White females	Black females	Hispanic females
All workers (1995)							
(1) Share of all workers	1.000	0.405	0.037	0.073	0.378	0.049	0.059
(2) Hourly wage	14.88	18.96	12.41	12.20	12.25	10.19	10.94
	(59.48)	(69.11)	(33.21)	(69.33)	(29.34)	(21.89)	(67.72)
(3) Annual earnings	26842	36169	23645	20418	20522	17624	15372
	(1197)	(1346)	(1314)	(926)	(990)	(821)	(917)
(4) Weeks worked	37.0	42.3	34.1	38.6	34.4	31.3	26.3
	(31.11)	(28.8)	(35.0)	(28.2)	(31.8)	(34.2)	(29.9)
(5) Hours worked per week	32.0	38.4	30.3	34.4	27.9	26.3	22.2
	(29.4)	(28.3)	(32.44)	(26.1)	(29.1)	(30.8)	(27)
(6) Share part-time	0.221	0.123	0.153	0.149	0.330	0.254	0.314
(7) Share public sector[b]	0.144	0.120	0.157	0.087	0.165	0.231	0.143
Full-time–full year (1995)							
(8) Hourly wage	14.86	17.97	13.00	11.06	12.51	10.72	9.70
	(24.41)	(27.19)	(25.01)	(18.93)	(19.59)	(17.03)	(18.47)
(9) Annual earnings	34265	42742	29651	24884	27583	22871	20695
	(1236)	(1378)	(1373)	(939)	(963)	(796)	(864)
All persons							
(10) Share ever employed, 1995	0.807	0.892	0.756	0.848	0.769	0.701	0.611
(11) Share ever unemployed, 1995	0.086	0.092	0.119	0.132	0.070	0.091	0.078
(12) Unemployment rate, March 1996	0.044	0.043	0.103	0.080	0.028	0.059	0.057
(13) Employment rate, March 1996	0.731	0.820	0.647	0.768	0.695	0.620	0.532

[a] Source: Current Population Survey, March 1996. Weighted estimates, standard deviations are in parentheses.
[b] Share public sector from March 1996.

Altonji, Joseph G., and Rebecca M. Blank. (1999)

Convergence and subsequent divergence in median weekly earnings

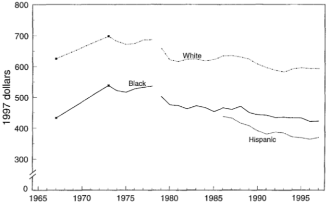


Fig. 1. Median weekly earnings of full-time male workers. Source: Bureau of Labor Statistics.

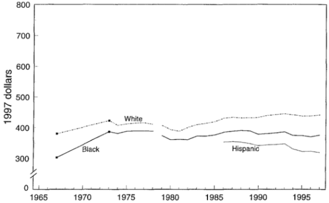
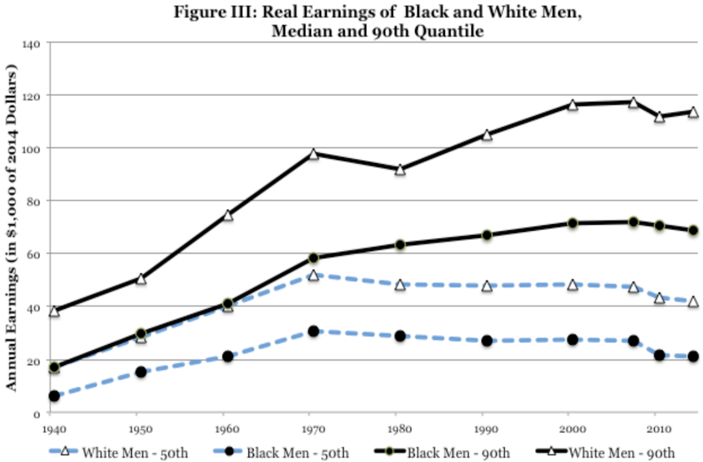


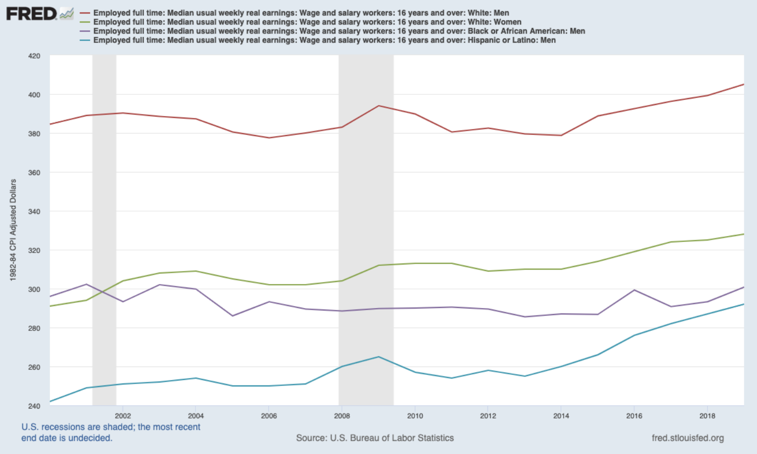
Fig. 2. Median weekly earnings of full-time female workers. Source: Bureau of Labor Statistics.

Median and 90th Percentile Wage-Gaps Today (Neal 2021)



Note: Figure displays earnings of the median and 90th quantile non-Hispanic black and white men measured in the population of all men aged 25-54. Earnings are converted to constant 2014 dollars using the CPI-U price deflator and are measured in thousands of dollars. *Sources:* Census, 1940-2000; American Community Survey, 2005-2014. The sample year labeled '2007' combines ACS samples from 2005-07, '2014' combines those from 2013-14.

Median Earnings Today



(White Women modest gain, Hispanic Men big gain, Black Men loss.)

Comparison of Labor Force Participation

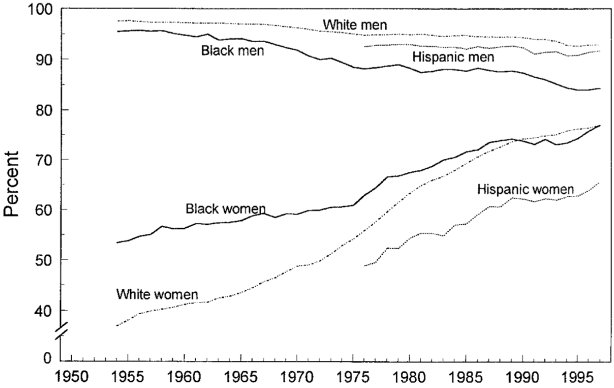
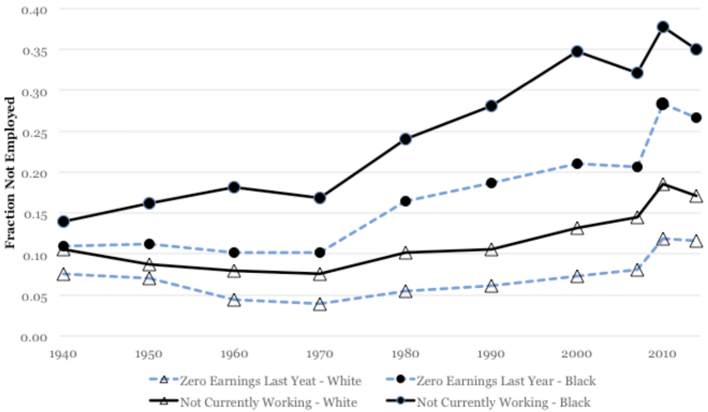


Fig. 5. Labor force participation rates, 25–54-year-olds. Source: Bureau of Labor Statistics.

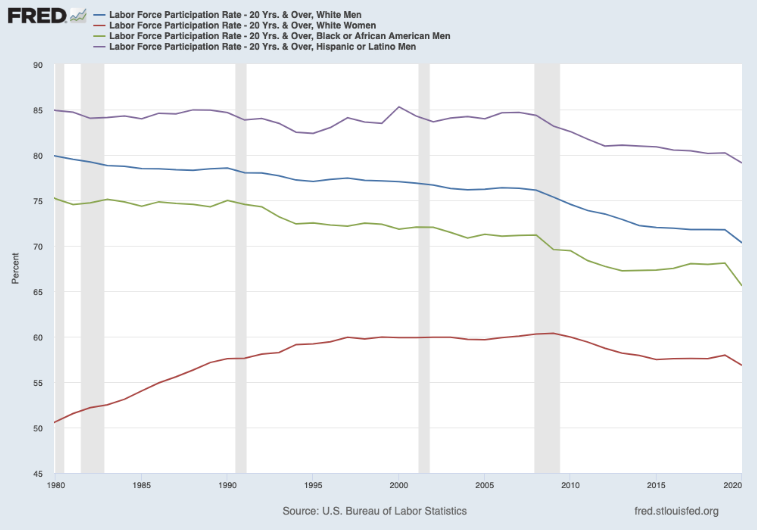
Today (Neal 2021)

Figure II.A: Fraction of Men Not Employed, by Alternative Measure and Race



Note: Figure displays fraction of non-Hispanic black and white men aged 25-54 not working according to two measures: not currently working and zero annual earnings in the previous year. The measure of earnings is labor market earnings plus business and farm income. *Sources:* Census, 1940-2000; American Community Survey, 2005-2014. The sample year labeled '2007' combines ACS samples from 2005-07 and '2014' combines those from 2013-14.

Today - (BLS Most Updated)



Differences in Personal Characteristics

A key difficulty in analyzing discrimination's responsibility of the wage gap is then pulling apart these gaps into two different components, both of which are important: 1) Differences in Personal Characteristics (sometimes called "pre-market discrimination") and 2) Differences due to discrimination

There are large disparities in personal characteristics between these groups:

- Minority workers have substantially lower education levels and work experience when compared to their white compatriots.
- There are large demographic differences between in the industries people work.
- Profound differences in family responsibilities.

How much are these differences due to preferential treatment of white males? How much are due to structural constraints faced by those in less privileged demographic groups? Disentangling these effects is key.

Samples also matter

Heckman et. al (2000) show that using different sampling rules can drastically change wage gap findings

TABLE 1—DECOMPOSITION OF AVERAGE LOG-WAGE GAP UNDER ALTERNATIVE SAMPLING RULES

Sample	1960–1970	1970–1980	1980–1990
SW			
Δ Log-wage gap	13.1	12.3	2.7
Within-cohort component	4.96	3.72	−1.5
Between-cohort component	8.14	8.57	3.65
CK(1), 48 states			
ΔLog-wage gap	10.3	8.49	−1.4
Within-cohort component	0.27	1.16	−5.5
Between-cohort component	9.99	7.33	4.13
CK(2), S–N migrants			
ΔLog-wage gap	5.11	4.03	−0.25
Within-cohort component	−2.0	−3.5	−6.3
Between-cohort component	7.12	7.53	6.09
HLT(1), 48 states			
ΔLog-wage gap	14.8	8.77	2.43
Within-cohort component	3.70	2.06	−1.6
Between-cohort component	11.1	6.71	4.05
HLT(2), 29-states			
Δ Log-wage gap	14.8	9.05	1.52
Within-cohort component	3.58	2.15	−2.6
Between-cohort component	11.3	6.90	4.08

Notes: The table reports the change in the average log-wage gap × 100 (ΔLog-wage gap) for each of several sampling rules defined in the text. The within-cohort component is based on continuing cohorts with base-year weights; the between-cohort component is based on replacement cohorts and changing cohort weights. CK(2) is a subsample of CK(1) containing only South–North migrants. HLT(2) is a subsample of HLT(1), containing only persons born in 29 states with substantial black populations.

So how do we get at these wage gaps?

Start with decomposing wages into "explained," X_{git}, and "unexplained," u_{git} components:

$$W_{git} = \beta_{gt} X_{git} + u_{git}$$

Now take the difference between two groups g, g' over all the individuals (assume $\mathbb{E}[u_{git} | X_{git}] = 0$):

$$W_{gt} - W_{g't} = \beta_{gt}(X_{gt} - X_{g't}) + (\beta_{gt} - \beta_{g't})X_{g't}$$

Showing the differences in mean wages can separate out into two components.

- $\beta_{gt}(X_{gt} - X_{g't})$ is the explained component, due to differences in the aforementioned personal differences.
- $(\beta_{gt} - \beta g' t)X_{g't}$ is the unexplained component, representing differences due to unobserved differences and **discrimination**.

Example: Education

Heckman et. al (2000) show how much the personal characteristic of education explains the black-white wage gap (under various samples)

TABLE 2—TOTAL CONTRIBUTION OF EDUCATION TO CHANGE IN THE RELATIVE WAGE GAP UNDER ALTERNATIVE SAMPLES, HLT MODEL, 1960–1970

Age	Effect	CK(1)	CK(2)	SW	HLT(2)	LFP-corrected[a]
31–40	Main	4.9	4.0	5.0	5.9	4.5
	Race	−3.6	−4.6	−3.9	−4.8	−4.7
	Year	−3.3	−3.0	−4.1	−4.7	−5.2
	Race–Year	7.8	8.4	5.8	13.1	5.7
	Total	5.7	4.8	2.8	9.5	0.4
	Total ΔLWG[b]	8.6	6.5	11.5	12.8	12.8
	Educ.[c]	66.7	73.6	24.2	74.5	2.8
41–50	Main	3.7	2.8	3.7	4.6	3.4
	Race	−6.0	−3.5	−3.6	−6.7	−9.1
	Year	−7.5	10.4	−8.5	−8.0	−6.8
	Race–Year	8.4	−8.6	5.4	13.3	14.0
	Total	−1.3	1.2	−3	3.1	1.5
	Total ΔLWG[b]	5.6	0.0	10.2	10.1	10.1
	Educ.[c]	−23.5	−8661	−29.6	31.3	14.5

Note: "Δ log wage gap" denotes the change in the log wage gap.

Estimating the simple model

Columns (1) and (4) report regression of hourly wages on race and gender dummies, Columns (2) and (5) add in personal characteristics, and Columns (3) and (6) add in industry and occupation controls.

Table 4
Coefficients on race and gender in wage regressions[a]

	1979			1995		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Part (A) all workers						
(1) Black	−0.143	−0.107	−0.061	−0.207	−0.119	−0.089
	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)
(2) Hispanic	−0.152	−0.053	−0.040	−0.379	−0.131	−0.102
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
(3) Female	−0.436	−0.421	−0.348	−0.279	−0.272	−0.221
	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)
Controls						
(4) Education, experience, and region	No	Yes	Yes	No	Yes	Yes
(5) Occupation, industry and job characteristics[b]	No	No	Yes	No	No	Yes
Part (B) full-time-full year workers						
(6) Black	−0.139	−0.115	−0.064	−0.148	−0.102	−0.067
	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)	(0.010)
(7) Hispanic	−0.184	−0.093	−0.076	−0.344	−0.139	−0.101
	(0.012)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)
(8) Female	−0.421	−0.399	−0.360	−0.265	−0.266	−0.241
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Controls						
(9) Education, experience, and region	No	Yes	Yes	No	Yes	Yes
(10) Occupation, industry and job characteristics[b]	No	No	Yes	No	No	Yes

[a] Source: Authors' regressions using the Current Population Survey, March 1980 and March 1996. Standard errors are in parentheses.
[b] Job characteristics include public sector and part-time status.

First pass implies the culprit is personal characteristics

These results imply that the main culprit of the wage gap is our institutions and economic/societal structures.

However, 6-9 % is still a big number! And when looking at the final specifications there are still statistically significant coefficients on minority and female coefficients.

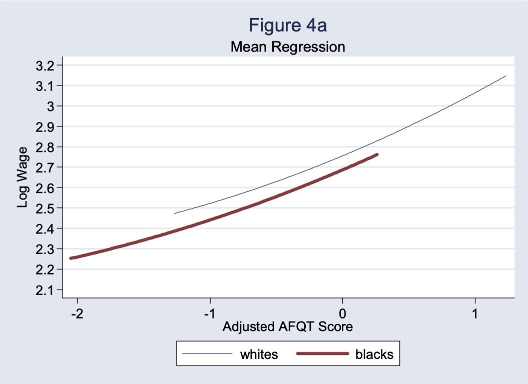
Table 4
Coefficients on race and gender in wage regressions[a]

	1979			1995		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Part (A) all workers						
(1) Black	−0.143	−0.107	−0.061	−0.207	−0.119	−0.089
	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)
(2) Hispanic	−0.152	−0.053	−0.040	−0.379	−0.131	−0.102
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)
(3) Female	−0.436	−0.421	−0.348	−0.279	−0.272	−0.221
	(0.006)	(0.005)	(0.006)	(0.007)	(0.006)	(0.007)
Controls						
(4) Education, experience, and region	No	Yes	Yes	No	Yes	Yes
(5) Occupation, industry and job characteristics[b]	No	No	Yes	No	No	Yes
Part (B) full-time-full year workers						
(6) Black	−0.139	−0.115	−0.064	−0.148	−0.102	−0.067
	(0.012)	(0.011)	(0.011)	(0.012)	(0.011)	(0.010)
(7) Hispanic	−0.184	−0.093	−0.076	−0.344	−0.139	−0.101
	(0.012)	(0.012)	(0.011)	(0.010)	(0.010)	(0.010)
(8) Female	−0.421	−0.399	−0.360	−0.265	−0.266	−0.241
	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)
Controls						
(9) Education, experience, and region	No	Yes	Yes	No	Yes	Yes
(10) Occupation, industry and job characteristics[b]	No	No	Yes	No	No	Yes

[a] Source: Authors' regressions using the Current Population Survey, March 1980 and March 1996. Standard errors are in parentheses.
[b] Job characteristics include public sector and part-time status.

Holding Everything Constant

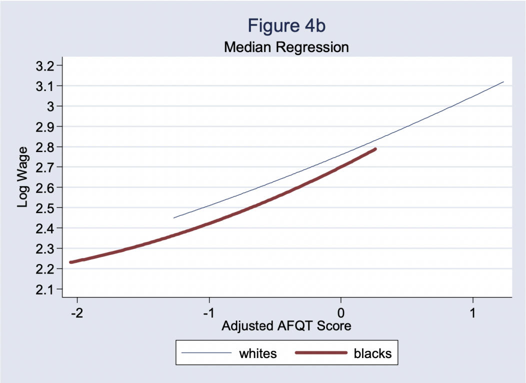
Contingent on schooling, occupation and ability among other controls, wage gap becomes > 95 % explained.



(Neal 2021)

Holding Everything Constant

Contingent on schooling, occupation and ability among other controls, wage gap becomes > 95 % explained.



(Neal 2021)

Patterns

Two main patterns from the previous slides:

- Moving from partial to full specification increases the personal characteristics share of the gap (the "explained").
- However this increase in the share for characteristics decline between the years 1979 to 1995.

However, it should be acknowledged the CPS data has significant limitations (no ability measure, no past labor market experience, etc.)

Exploiting a Richer Data Source

Similar analysis was then conducted using the National Longitudinal Survey of Youth (NLSY... remember the problemset?) Leads to interesting findings

- Wage gap between races become almost entirely due to differences in personal characteristics (e.g. the "explained"/ observed coefficients / pre-market discrimination component.)
- Wage gap between the sexes become almost entirely due to differences in coefficients. (e.g. the "unexplained" / unobserved coefficients / at-hire-discrimination component

However, it should be acknowledged the CPS data has significant limitations (no ability measure, no past labor market experience, etc.)

Exploiting a Richer Data Source

Table 6
Coefficients and decompositions of race and gender wage differentials[a]

	Occupation, industry, job characteristics excluded				Occupation, industry, job characteristics included			
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Controls								
(1) Education, potential experience, and region	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
(2) Add Family background[b] and AFQT	No	Yes	Yes	Yes	No	Yes	Yes	Yes
(3) Use actual experience	No	No	Yes	Yes	No	No	Yes	Yes
(4) Add personal characteristics[c]	No	No	No	Yes	No	No	No	Yes
(5) Occupation, industry and job characteristics[d]	No	No	No	No	Yes	Yes	Yes	Yes
(A) Combined sample with race and gender dummy variables								
(6) Black	−0.154	−0.060	−0.030	−0.029	−0.139	−0.055	−0.028	−0.027
	(0.028)	(0.032)	(0.031)	(0.031)	(0.028)	(0.031)	(0.031)	(0.031)
(7) Female	−0.244	−0.239	−0.211	−0.214	−0.231	−0.225	−0.196	−0.199
	(0.024)	(0.024)	(0.024)	(0.025)	(0.026)	(0.025)	(0.026)	(0.026)
(B) Decompositions based on group specific regressions								
Amount due to (males vs females)								
(8) Coefficients	−0.057	−0.136	−0.225	−0.243	−0.032	−0.111	−0.177	−0.198
(9) Characteristics[e]	−0.171	−0.121	−0.035	−0.011	−0.192	−0.132	−0.069	−0.044
Amount due to (whites vs blacks)								
(10) Coefficients	−0.150	−0.022	−0.005	−0.009	−0.188	−0.044	−0.032	−0.036
(11) Characteristics[f]	0.081	−0.057	−0.057	−0.056	0.036	−0.105	−0.101	−0.099

[a] Source: Authors' regressions using the National Longitudinal Survey of Youth, 1994. Standard errors are in parentheses.
[b] Family background characteristics include mother's and father's education and employment status in 1978.
[c] Personal characteristics include age of youngest child, total number of children, and sex and race when appropriate.
[d] Job characteristics include public sector and part-time status, 1 digit industry and occupation controls.
[e] Coefficients from regression for males.
[f] Coefficients from regression for whites.

What about Labor Force Participation?

Similar patterns as the NLSY study have been found for the white-nonwhite wage gap and the male-female wage gap.

The white-nonwhite wage gap seems to be primarily driven by personal characteristics / “pre-market” discrimination, while the male-female wage gap is due to unexplained forces (including discrimination).

What actually is "Discrimination"

To continue to talk about discrimination it will be best to rigorously define discrimination. Following Cain (1986):

$$Y = X\beta + \alpha Z + e$$

Where:

$$X = \text{Vector of all productivity characteristics}$$
$$Z = \text{Binary variable if person is a minority}$$

This group is discriminated against is $\alpha < 0$.

Huge difficulty in getting all terms to make β, α exogenous.

Breakdown into 2 Versions

1. (Cainian) Discrimination: The idea that hold all other productivity characteristics equal, one is paid less due to their member of a demographic group. (Think taste based)
   - Can be seen as a difference in wages for the same levels of productivity: $\alpha < 0$, or $(\beta_{gt} - \beta_{g't})$ from the basic model.
2. Structural/Statistical/Pre-Market Discrimination: The idea that being member of an out group is correlated to accumulating lesser productivity characteristics.
   - Can be seen as lesser levels of group productivity characteristic gain: $Z \not\perp X$ or $X_{gt} - X_{g't} \neq 0$ from before;

Are These Necessarily Different?

From Altonji and Blank (1999):

"One the one hand, if discrimination is affecting the human capital investments and personal choices that individuals make or if it is affecting job choice, then the 'unexplained gap' will understate discrimination because some of the control variables themselves reflect the impact of discrimination."

Other direction, personal characteristics $\Rightarrow$ discrimination, can be thought of as models of statistical discrimination (discussed later).

Hard to truly disentangle!

Could historical discrimination lead to differences in personal characteristics?

Consider the following example:

- Last Generation: college does not allow members of minority group to enroll
- This Generation: college runs regression on past finding legacy students perform better
- This Generation: college admits far more legacy students than non-legacy students

In this case "legacy" is **an observed personal characteristic** yet is actually the result of previously discriminatory behavior!

Theories of Differential Labor Market Outcomes

Historical there have been three main theoretical explanations for differences in labor market outcomes

1. Driven by differences in skills rather than prejudiced discrimination.
2. Driven by discrimination itself / prejudiced "tastes" by employers, employees or consumers.
3. Driven by occupational exclusion driven from crowding, social norms and other constraints

But what if these explanations were not distinct but endogenous to each other? (Ala statistical, and the previous examples)

What Can Economics Say about Discrimination?

First, we distinguish between two types of models.

- *Competitive Models*: individuals maximize utility by discriminating
  - Taste-based discrimination (Becker, 1957)
  - Statistical discrimination (Phelps, 1972, and Arrow, 1973)
- *Collective Models*: groups act collectively against each other

Most of economic theory considers *competitive models*. So will we.

In this section, our goal will be to:

- Introduce Becker's (1957) taste-based model.
- Discuss the statistical theory of discrimination, which is generally more widely accepted today. To highlight modern work, we will discuss the Coate & Loury (1993) paper on affirmative action.
- Provide motivation for studying systemic discrimination.

Becker's (1957) Model

- Employers have a "taste for discrimination" against minority workers.
  - Two groups: majority (a) and minority (b).
  - When hiring workers, firms maximize profits plus the monetary value of utility from employing members of certain groups.
- Firms set employment levels (N_a and N_b) to maximize:

$$U = pF(N_a + N_b) - w_a N_a - w_b N_b - t_b N_b,$$

  where p is price, F is the production function, w_a (w_b) is the wage in group a (b), and t_b is the "coefficient of discrimination".
- The optimal number of workers hired solves:

$$p\frac{\partial F}{\partial N_a} = w_a \quad \text{and} \quad p\frac{\partial F}{\partial N_b} = w_b + t_b$$

- Intuitively, minorities must be more productive at a given wage. Or, if productivity is fixed, the firm will only hire b if $w_a - w_b \geq t_b$.

Becker's (1957) Model

Now we consider a competitive labor market.

- Let $G(t_b)$ be the distribution of the taste parameter t_b.
- Suppose price p is fixed. For $i \in \{a, b\}$, aggregating across firms gives us labor demand $N_i^D(w_a, w_b, G(t_b))$ and labor supply $N_i^S(w_i)$ functions. In equilibrium, wages are chosen to satisfy:

$$N_i^D(w_a, w_b, G(t_b)) = N_i^S(w_i), \quad \forall i \in \{a, b\}$$

- Wages are unequal (i.e. $w_b < w_a$) *iff* there are sufficiently many discriminating firms such that, when $w_b = w_a$, $N_b^D < N_b^S$.
- Discrimination on average $\not\Rightarrow$ discrimination at the margin. Wage gaps are determined by t_b for the marginal employer of b workers.

Becker's (1957) Model

The consequences of a competitive labor market are:

- Enough unbiased firms $\implies$ discrimination is competed away, and minority workers won't work for discriminating employers.
- Many firms discriminate $\implies$ $w_b < w_a$, and the wage gap is determined by the strength of discrimination at the margin.
- With free entry and constant returns to scale, discriminating firms go out of business, since they inherit the cost of their distaste.

Some testable implications of this model are:

- Wage differentials, controlling for productivity
- Preferential hiring, controlling for productivity
- Segregated industries, e.g. groups only work at their own businesses

Becker explores two other sources of discriminatory tastes:

- *Co-workers*: workers in group a are prejudiced against b workers
- *Customers*: consumers in group a discriminate against b products

Criticisms of the Taste-Based Approach

Arrow's critique of the taste-based framework is twofold:

(1) Discrimination is taken as given, as a feature of preferences.
  - It "risks turning the 'explanation' into a tautology" (Arrow, 1998).

(2) Perfect competition $\implies$ discriminatory firms exit the market.
  - "Becker's model predicts the absence of the phenomenon it was designed to explain" (Arrow, 1972).
  - Alternatively, some variations of Becker's model drop the free entry assumption, e.g. an inelastic supply of entrepreneurs.

In recent decades, economists have moved away from Becker's model and have turned toward models of statistical discrimination.

- Introduced by Phelps (1972) and Arrow (1973). Some seminal models include Aigner & Cain (1977) and Coate & Loury (1993).
- Today, we will give a brief overview of statistical discrimination, and then discuss a more modern signaling model.

Statistical Discrimination: An Overview

Statistical discrimination is the outcome of a signal extraction problem.

- Employer observes a signal (with noise) about worker's productivity.
- Conditional on group-membership, the expectation or variance of productivity may differ, which leads to labor market discrimination.

Basic Setting

- Employer sees: group $x \in \{a, b\}$ and noisy signal $\tilde{\eta}_i$ of productivity.

$$\tilde{\eta}_i = \eta_i + \epsilon_i, \quad \text{where: } \begin{cases} \epsilon_i \sim N(0, \sigma_\epsilon^2) \\ \eta_i \sim N(\bar{\eta}, \sigma_\eta^2) \end{cases}$$

- What happens if $\bar{\eta}$ varies by group, i.e. $\eta_i | x \sim N(\bar{\eta}_x, \sigma_\eta^2)$?
- What happens if σ_ϵ^2 varies by group, i.e. $\epsilon_i | x \sim N(0, \sigma_{\epsilon,x}^2)$?

Signaling Problem: Difference in Means

Assume that $\bar{\eta}_a > \bar{\eta}_b$, i.e. group b is less-productive on average.

- Signal of productivity is: $\tilde{\eta}_i = \bar{\eta}_x + \nu_i + \epsilon_i$, where $\nu_i \sim N(0, \sigma_\eta^2)$.
- Employer forms beliefs about applicant, given group x and signal $\tilde{\eta}_i$.

$$E(\eta_i | \tilde{\eta}_i, x) = \alpha \bar{\eta}_x + (1 - \alpha)\tilde{\eta}_i, \quad \text{where: } \alpha = \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_\eta^2}$$

From this equation, we see that—for a given $\tilde{\eta}_i$—expected productivity of applicants in b is less than it is for applicants in a. We write:

$$E(\eta_i | \tilde{\eta}, a) - E(\eta_i | \tilde{\eta}, b) = \alpha(\bar{\eta}_a - \bar{\eta}_b)$$

As an outcome of this signaling problem, we find that:

$$E(\eta_i | \tilde{\eta}_i) = \eta_i \quad \text{and } \mathbb{E}(\tilde{\eta}_i | \eta_i, x) = \eta_i, \quad \text{but } E(\eta_i | \tilde{\eta}_i, x) \neq \eta_i$$

Equal pay for equal *expected* productivity, but not for equal productivity.

Signaling Problem: Difference in Variance

Suppose the signal $\tilde{\eta}_i$ is *more informative* for group a than for group b.

$$\bar{\eta}_a = \bar{\eta}_b \quad \text{and } \sigma^2_{\eta,a} = \sigma^2_{\eta,b}, \quad \text{but } \sigma^2_{\epsilon,a} < \sigma^2_{\epsilon,b}$$

Just as before, employer forms beliefs about the applicant:

$$E(\eta_i|\tilde{\eta}_i, x) = \alpha_x \bar{\eta} + (1 - \alpha_x)\tilde{\eta}_i, \quad \text{where: } \alpha_x = \frac{\sigma^2_{\epsilon,x}}{\sigma^2_{\epsilon,x} + \sigma^2_{\eta}}$$

- For applicants in group b, employers will put more weight on the mean $\bar{\eta}$ and less weight on the individual signal $\tilde{\eta}_i$.
- Workers above the mean prefer informative signals, while workers below the mean prefer uninformative signals.

Signaling Problem: Difference in Variance

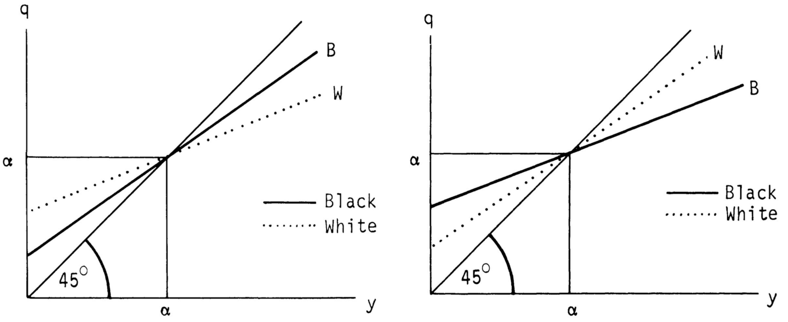


Figure 1A. Predictions of Productivity (q) by Race and Test Score (y), Assuming a Steeper Slope for Blacks.

Figure 1B. Predictions of Productivity (q) by Race and Test Score (y), Assuming a Steeper Slope for Whites.

Source: Cain & Aigner (1977)

- The relative steepness of a versus b increases in $\sigma^2_{\epsilon,b}/\sigma^2_{\epsilon,a}$.
- The expectation of $\eta_i|(\tilde{\eta}_i, x)$ crosses at $\bar{\eta}$ for both groups $x \in \{a, b\}$.

Signaling Problem: Difference in Variance

So far, we assumed that the dispersion in $\eta_i|(\tilde{\eta}_i, x)$ is costless. Instead, let η_i enter the profit function nonlinearly, so the firm is risk averse.

Following Cain & Aigner (1977), we assume labor is the only factor of production, that output is fixed, and that prices & wages are exogenous. Given the number of workers required to maximize profits, the employer need only choose the type of labor in order to maximize:

$$U(\eta_i) = a - b\exp(-c\eta_i), \quad \text{where } b, c > 0$$

The expected utility from $\eta_i|(\tilde{\eta}_i, x)$ is given by:

$$E[U(\eta_i|\tilde{\eta}_i, x)] = a - b\exp\Big(-cE[\eta_i|\tilde{\eta}_i, x] + \frac{c^2}{2}\mathrm{Var}[\eta_i|\tilde{\eta}_i, x]\Big)$$

- The employer now cares about how "risky" the applicant is.
- Noisier signals of productivity now harm *all* workers.

Signaling Problem: Difference in Variance

The error in employer's beliefs is:

$$\begin{aligned} E(\eta_i|\tilde{\eta}_i, x) - \eta_i &= \alpha_x \bar{\eta} + (1-\alpha_x)\tilde{\eta}_i - \eta_i \\ &= \frac{\nu_i \sigma_\eta^2 - \epsilon_i \sigma_{\epsilon,x}^2}{\sigma_{\epsilon,x}^2 + \sigma_\eta^2} \end{aligned}$$

Variance of this term is: $\sigma_{\text{error}}^2 = \text{Var}\big(E(\eta_i|\tilde{\eta}_i, x) - \eta_i\big) = \frac{\sigma_{\epsilon,x}^2 \sigma_\eta^2}{\sigma_{\epsilon,x}^2 + \sigma_\eta^2}$.

- The variance in the error increases in $\sigma_{\epsilon,x}^2$, i.e. $\frac{\partial \sigma_{\text{error}}^2}{\partial \sigma_{\epsilon,x}^2} > 0$.
- Since the signals for group b are noisier, and the employer is risk averse, there is statistical discrimination against *all* b applicants.

Implications of Statistical Discrimination

- Statistical discrimination is often considered to be "efficient".
  - It will exist in equilibrium among profit-maximizing agents, as it is an optimal solution to an employer's signaling problem.
- It is typically illegal, albeit hard to detect.
  - Very difficult to observe how employers form expectations.
- Is statistical discrimination fair?
  - Some economists argue it is fair and rational.
  - Others suggest that statistical discrimination is unfair as it generally reinforces existing socio-economic inequalities.

Affirmative Action: Coate & Loury (1993)

What are the impacts of affirmative action policies in the labor market?

- Perhaps expanding opportunities for minorities will eventually cause employers want to hire minorities regardless of affirmative action.
- Alternatively, it could reduce incentives for minorities to invest in skills, leading to an equilibrium where employers correctly believe that minorities are less productive, and so quotas remain binding.

To address this issue, Coate & Loury (1993) construct a model where:

- The underlying skill distributions are the same for both groups.
- Employers observe noisy signals about worker qualifications.
- In equilibrium, employers form consistent beliefs about worker qualifications, and workers make investments consistent with the returns they will receive in the labor market for their investments.

Affirmative Action: Coate & Loury (1993)

Suppose there are many firms and workers, where workers belong to groups $j \in \{A, B\}$, and λ is the fraction of A's in the population.

- Employer must assign each worker to either Task 0 or Task 1.
  - Only qualified workers are successful at Task 1.
  - Workers earn nothing for doing Task 0 and ω for doing Task 1.
  - For Task 0, firms get nothing. For Task 1, firms receive a return of:

$$x = \begin{cases} x_q > 0 & \text{if worker is qualified} \\ -x_u < 0 & \text{if worker is unqualified} \end{cases}$$

- When making assignments, employers only observe group identity j and a noisy signal $\theta \in [0, 1]$ about the worker's qualification level.
  - $\theta \sim F_q(\theta)$ for qualified workers, $\theta \sim F_u(\theta)$ for unqualified workers.
  - Monotone Likelihood Ratio Property: $\phi(\theta) = \frac{f_u(\theta)}{f_q(\theta)}$ non-increasing.
- To be qualified, workers must make a costly ex ante investment.
  - Worker's investment costs $c \sim G(c)$, where $G(\cdot)$ is continuous.

Affirmative Action: Coate & Loury (1993)

Let $\pi_j \in [0, 1]$ denote the firm's prior belief about worker qualification.

$$\text{Firm's posterior:} \quad \kappa(\pi_j, \theta) = \frac{\pi_j f_q(\theta)}{\pi_j f_q(\theta) + (1 - \pi_j) f_u(\theta)}$$

The firm's expected payoff is $\kappa(\pi_j, \theta) x_q + [1 - \kappa(\pi_j, \theta)] x_u$. It chooses to assign a worker to Task 1 *iff* the signal exceeds the threshold:

$$s^*(\pi_j) = \min \left\{ \theta \in [0, 1] : \frac{x_q}{x_u} > \left(\frac{1 - \pi_j}{\pi_j} \right) \phi(\theta) \right\}$$

The worker's gross benefit to becoming qualified is:

$$\beta(s) = \omega [F_u(s) - F_q(s)],$$

where s is the passing threshold. Since workers only invest if $\beta(s) \geq c$, the share of workers that become qualified is $G(\beta(s))$.

Affirmative Action: Coate & Loury (1993)

In equilibrium, beliefs are self-confirming: $\pi_j = G(\beta(s^*(\pi_j))$.

- Worker skills respond endogenously to employer beliefs.
- Multiple "discriminatory equilibria" (i.e. where $\pi_B \neq \pi_A$) can exist.
  - For example, if employers believe π_b is lower, then they will choose a higher threshold for assigning them to Task 1 (i.e. $s_B > s_A$), which in turn lowers investment for workers of type B.
- Stereotypes are inefficient. Both b workers and employers are better-off without them. Yet, without collective action, no single employer can break the discriminatory equilibrium.

Affirmative Action: Coate & Loury (1993)

Consider an affirmative action policy where a social planner requires both groups have equal assignment rates to Task 1. Firms solve:

$$\max_{s_A, s_B} \left\{(1-\lambda)P(s_B, \pi_B) + \lambda P(s_A, \pi_A)\right\}$$
$$\text{s.t. } \rho(s_B, \pi_B) = \rho(s_A, \pi_A),$$

where $P(s, \pi) = \pi(1 - F_q(s))x_q - (1-\pi)(1 - F_u(s))x_u$ is the expected payoff from hiring a worker, and $\rho(s, \pi) = \pi(1 - F_q(s)) - (1-\pi)(1 - F_u(s))$ is the ex-ante probability that a worker is assigned to Task 1.

- Coate & Loury (1993) provide conditions under which affirmative action will generate a non-discriminatory equilibrium.
- They find that such policies can also generate a "patronizing equilibrium," where the constraint permanently binds.

Affirmative Action: Extensions and Criticism

Coate & Loury (1993) conclude that the impact of affirmative action on a discriminatory equilibrium is theoretically ambiguous.

- If A's and B's have different underlying cost distributions, then affirmative action would only exacerbate skill deficits.
  - For example, suppose B's have worse educational opportunities.
- The subsidies/punishments to get firms to hire more B's has to be "just right" to achieve a sustainable non-discriminatory equilibrium.
  - Some economists argue it is near impossible to achieve in practice.
- Alternatively, subsidies to B's to obtain qualifications will *always* be beneficial, as long as $s_B < 1$ (i.e. if some B's are initially accepted).
  - Job training programs, scholarships, community investment, etc.

Overview of Systemic Discrimination

- Much of economic literature focuses on whether there exists discrimination in a particular domain and point of time.
  - Examples: school admissions, hiring, health care, loan applications.
- Recently, social scientists (e.g. Blank, 2006, and Reskin, 2012) have emphasized the importance of *systemic* discrimination.
  - Given dependence between domains and over time, are we able to measure the *cumulative* effects of one act of discrimination?

Blank (2006) suggests three manifestations of *cumulative* discrimination:

(1) Discriminatory impacts may cumulate within a single domain over time, e.g. in the labor market via hiring, promotions, & job changes.

(2) Discrimination in one social domain can affect outcomes in another social domain, e.g. housing affects education & future earnings.

(3) Discrimination may have inter-generational effects, e.g. education affects lifetime earnings for gen. 1 affects opportunities for gen. 2.

Systemic Discrimination vs. Disadvantage

We should distinguish between systemic discrimination and disadvantage.

- Cumulative disadvantage measures the change in outcome gaps between disadvantaged and advantaged groups over time.
    - It may reflect discrimination, but also other socio-economic factors.
- Cumulative discrimination measures the causal effect of a set of discriminatory activities on particular outcomes over time.
    - Perhaps hiring is unaffected by discrimination, but employment outcomes are discriminatory because of pre-labor market outcomes.
    - *Small* effects of discrimination in one domain can lead to *large* cumulative effects, and the size of these effects may differ by age.

"Disentangling" Discrimination

Some potential mechanisms to consider might be:

- Discrimination in one domain affects outcomes that are important variables to determine future outcomes across various domains.
- Discriminatory events produce feedback effects by influencing the behavior of those who experience it, e.g. two-sided models.

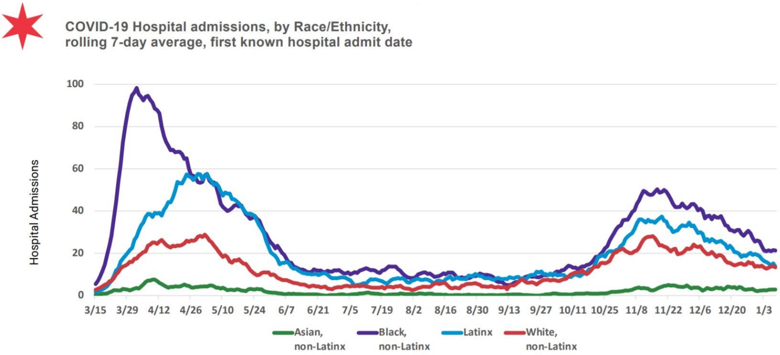
More focus on systemic discrimination in sociology than economics.

- Blank et al. (2004) offers ways to measure systemic discrimination.
- Reskin (2012) argues that race discrimination is a system with emergent properties, which reinforce the effects of their components.

How can we isolate the role of discrimination on inequality today?

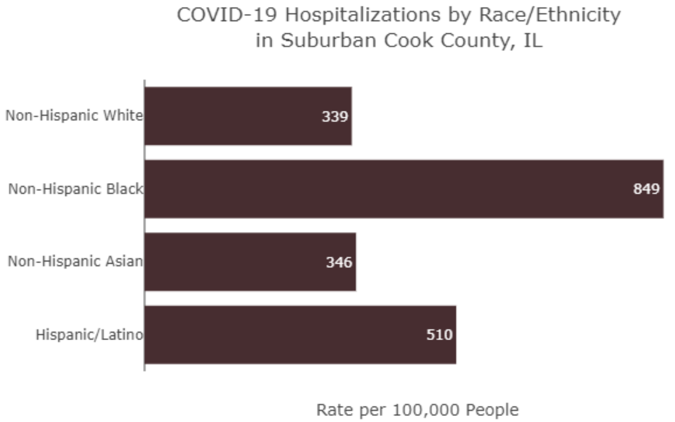
- Will racial inequality "fade away" in absence of discrimination?
  - Statistical discrimination can arise because of *known* inequalities.
- Can exogenous shocks (e.g. pandemics) magnify the impacts of discrimination? How can we quantify these effects?

COVID-19 Hospitalizations in Chicago



Data Source: Cook County Dept. of Public Health

COVID-19 Hospitalizations in the Chicago Suburbs



Data Source: Cook County Dept. of Public Health

Unpacking COVID-19 Disparities

- A lot of ongoing research on the social determinants of health.
  - *Residential segregation*. In Chicago, where you live determines what hospital you are routed to and the quality of care you will receive.
  - Links between health and other disparities (housing, employment, education, neighborhood, etc.) are widely studied.
- Evidence of ongoing discrimination within all of these domains.
  - Can we isolate the impacts of current discrimination on racial disparities, as opposed to existing inequalities?
  - Can we measure the historical effects of past discrimination?

If "small" instances of discrimination in one domain can have large and widespread cumulative effects, then the importance of discrimination may be understated, and perhaps different policies should be considered.

The inherent complexity of these relationships suggests we could model systemic discrimination through networks of individuals/institutions.

Network Theory: Calvó-Armengol & O. Jackson (2004)

The authors examine the role of social networks for obtaining information about job opportunities and implications for the dynamics of employment.

- Each agent is connected to others via a network. Information about jobs comes randomly to agents and spreads within the network.
  - Unemployed agents use information they hear to obtain jobs.
  - Employed agents may use information to obtain a better job, or instead pass the information to other direct contacts.
- Unemployment exhibits duration dependence and persistence.
  - The probability of remaining unemployed increases in the amount of time you are unemployed. (We find real-world evidence for this.)

While this paper does not directly address discrimination, it motivates the importance of social network effects in labor markets.

- What if minorities have weaker networks due to discrimination?
- How can a discriminatory network affect labor market outcomes?

Basic Model: Calvó-Armengol & O. Jackson (2004)

- Agents $i \in \{1, \dots, n\}$, and time $t = 1, 2, \dots$ is discrete.
  - $s_{it} \in \{0, 1\}$ is the employment status of agent i at time t.
  - Let a be the job arrival rate and b be the job breakup rate.
- Any two agents either know each other or do not, and information only flows between agents who know each other
  - A graph g summarizes the links of all agents, where $g_{ij} = 1$ if i & j know each other, and $g_{ij} = 0$ otherwise. Assume $g_{ij} = g_{ji}$.
- If an employed agent hears about a job, she randomly chooses an unemployed contact to give the information.
  - If all contacts are employed, the information is lost.

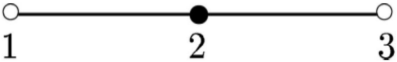
The probability that agent i learns about a job that agent j gets is:

$$p_{ij}(s) = \begin{cases} a & \text{if } s_i = 0 \text{ and } i = j \\ \frac{1}{\sum_{k:s_k=\mathbf{0}} g_{ik}} & \text{if } s_i = 1 \text{ and } g_{ij} = 1 \\ 0 & \text{otherwise} \end{cases}$$

Dynamics: Calvó-Armengol & O. Jackson (2004)

Having employed contacts improves i's prospects for hearing about a job if i is unemployed. There is motivation to reduce competition.

- If friends of my friends are employed, then I'm more likely to be told.



Consider the figure above, a network of 3 agents, and let $s_{t-1} = (0, 1, 0)$.

- *Conditional* on s_{t-1}, the states s_{1t} and s_{3t} are negatively correlated.
  - Agents 1 and 3 are "competing" for information from agent 2.
- However, agent 1 can benefit agent 3 in the long run.
  - When agent 1 is employed, agent 3 is more likely to get information.

Dynamics: Calvó-Armengol & O. Jackson (2004)

Proposition 1. *Under fine enough subdivisions of periods, the unique steady-state long run distribution on employment is such that the employment statuses of any path-connected agents are positively correlated.*

- Despite the short run conditional negative correlation, any interconnected agents' employment will be positively correlated in the long run.
- Intuitively, conditional some set of agents being employed, it is more likely that their neighbors will receive information about jobs, and so on.

Suppose that a person has been unemployed for the last X periods. What is the probability she will be employed next period?

g	1 period	2 periods	10 periods	limit
	0.099	0.099	0.099	0.099
	0.176	0.175	0.170	0.099
	0.305	0.300	0.278	0.099

Setting: $n = 4$, $a = 0.1$, and $b = 0.015$

Dynamics: Calvó-Armengol & O. Jackson (2004)

Proposition 3. *Under fine enough subdivisions of periods and starting under the steady state distribution, the conditional probability that an individual will become employed in a given period is decreasing with the length of their observed (individual) unemployment spell.*

- Longer past unemployment histories lead to worse inferences about the state of one's connections, which leads to worse inferences about the probability of that an agent will hear indirect news about a job.
- In other words, a longer individual unemployment spell makes it more likely that the state of one's social environment is poor, which in turn leads to low forecasts of future employment prospects.

This result reflects a more general persistence in employment dynamics.

Inequality: Calvó-Armengol & O. Jackson (2004)

Suppose there is a cost c_i to staying in the labor market, and agents can decide whether to stay in or to drop out, with no reentry into the network.

- To simplify things, assume all agents make choose whether to drop out of the network simultaneously and only once.
- Agents with better contacts have higher thresholds of dropout costs.
- *Supermodularity.* Having more agents participate is better for any agent, as it improves the long run prospects for future employment.

Proposition 4. *Consider two social groups with identical network structures. If the starting state person-by-person is higher for one group than the other, then the set of agents who drop out of the first group in the maximal equilibrium is a subset of their counterparts in the second group. These differences in drop-out rates generate persistent inequality in probabilities of employment in the steady-state distributions, with the first group having better employment probabilities than their counterparts.*

What Do We Need to Detect?

First, is there discrimination of any kind?

Next, can we distinguish the following?

- Taste-based vs. statistical discrimination
  - Does this distinction matter?
  - Blurred distinction over time: taste-based $\rightarrow$ statistical
  - Accurate vs. inaccurate decision-maker (DM) beliefs
- Implicit biases affecting behavior
  - Defined as: "unconscious mental associations between a target (such as an African-American) and an attitude". (Bertrand et al., 2005)
- Contribution of pre-market gaps vs. market discrimination
  - Disparities due to historical reasons or discrimination?
- Relevance of network-based or structural explanations
  - Jobs through homogeneous networks
  - "In-group" vs. "Out-group" effects

Common Empirical Approaches

A taxonomy from Neumark (2018)

1. Regression decompositions a la Oaxaca (1973)
   + Focus on wages not hiring
   - OVB and effect of imperfectly controlling for differences
2. Production function estimation
   + Direct measurement of productivity
   - Plant level measures, doesn't account for sorting
3. Non-experimental approaches
   - Employer learning: see Foster and Rosenzweig (1993); Altonji and Pierret (2001)
   - Roy Model framework (quasi-experimental)

Common Empirical Approaches (cntd.)

4. Field experiments
   - Largely correspondence and audit studies
5. Lab experiments
   - Vignette studies
   - Identifying taste-based discrimination: add new information
   - Tests for implicit bias

We present critiques of these methods from Heckman (1998), Heckman and Siegelman (1993), and Bertrand and Duflo (2016).

Audit and Correspondence Studies

Audit studies

- Matched individuals apply for jobs in-person
- Better job outcomes (actual offers vs. callbacks)

Correspondence studies

- Fictitious profiles differing by race/gender sent out
- More directly comparable individuals
- Larger samples
- Harder to provide race information on resumes

Both designs have several important implicit assumptions.

Issues with Audit Studies

Data and Design Issues

- Data available to econometrician vs. decision-maker
- Hard to have exactly identical candidates
- Not a double-blinded study

Econometric Issues

- Omitted variable bias
- Pre-market sources of disparity

General Issues with CAS Studies

- Measure average and not marginal discrimination
- Largely apply to entry level positions
  - No information about subsequent promotion decisions
- Ethical issues
  - Wasting scarce employer time
  - Rejecting a job offer might cause employers to update priors

Model Behind CAS studies

The following framework is from the Heckman (1998) critique

$$\text{Performance: } P(X, f, m)$$
$$\text{Treatment (hiring decision): } T(P(X, f, m), m)$$

- m: membership in minority group
- f: firm (usually randomly selected)
- X: vector of characteristics that can be decomposed into:
  - X^E: characteristics observed to econometrician
  - X^U: unobserved characteristics

Assumption $P \perp m$. Discrimination if:

$$T(P(X, f), m = 1) \neq T(P(X, f), m = 0)$$

CAS Model: Linear Case

$$T(P(X,f),m) = \underbrace{\beta X^E + X^U + f}_{\text{performance}} + \gamma m$$

Discrimination captured by γ

- Note that γ can vary among firms

Standardization: picking a specific productivity value

Ideal: $P_0^* = P_1^*$ s.t. $\quad T(P_1^*, 1) - T(P_0^*, 0) = \gamma$

Actual: $X_0^{E*} = X_1^{E*}$ s.t. $\quad T(P_1^*, 1) - T(P_0^*, 0) = X_1^U - X_0^U + \gamma$

Assumption Unbiased estimation requires $\mathbb{E}[X_1^U] = \mathbb{E}[X_0^U]$

CAS Model: Adding Nonlinearities

Threshhold hiring rules add another complication

$$\text{Minorities: } \begin{cases} T(P_1^*, 1) = 1 & \text{if } \beta X^{E^*} + X_1^U + f + \gamma \geq c \\ 0 & \text{otherwise} \end{cases}$$

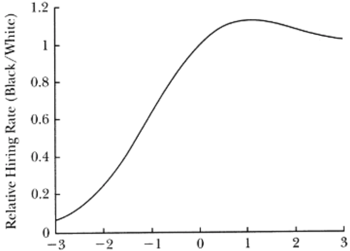
$$\text{Non-minority: } \begin{cases} T(P_0^*, 0) = 1 & \text{if } \beta X^{E^*} + X_0^U + f \geq c \\ 0 & \text{otherwise} \end{cases}$$

Even if distribution of f identical across pairs and $\perp X$, we need:
Assumption $X_1^U \overset{d}{=} X_0^U$ in order for identical hiring probabilities

Simulating Two Hypothetical Worlds

Example: Black workers have lower variance in X_2 unobservables

No Discrimination



X_1^* = level of standardization

X_2^1, X_2^0 normal

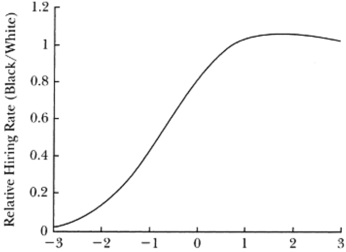
$E(X_2^1) = E(X_2^0) = 0; \ Var(X_2^1) < Var(X_2^0)$

$$\text{Relative Hiring Rate} = \frac{Pr(T(P_1^*, 1) = 1)}{Pr(T(P_0^*, 0) = 1)}$$

$Var(X_2^0) = 2.25 \ Var(X_2^1) = 1$

$c_1 = c_0 = 0$

Discrimination



X_1^* = level of standardization

X_2^1, X_2^0 normal

$E(X_2^1) = E(X_2^0) = 0; \ Var(X_2^1) < Var(X_2^0)$

$$\text{Relative Hiring Rate} = \frac{Pr(T(P_1^*, 1) = 1)}{Pr(T(P_0^*, 0) = 1)}$$

$Var(X_2^0) = 2.25 \ Var(X_2^1) = 1$

$c_1 = 0.25, \ c_0 = 0$

Potential Solution: Neumark (2012)

Framework

- Assume a threshold hiring rule as above
- Parametric assumption on unobservables
- Characteristics affecting perceived productivity do not vary with group membership

Assuming $f = 0$, $X_1^U \sim \mathcal{N}(0, \sigma_1)$ and $X_0^U \sim \mathcal{N}(0, \sigma_0)$, we get callback probabilities:

$$
\begin{aligned}
m = 1: &\qquad \Phi\left[(\beta X^{E^*} + \gamma - c)/\sigma_1\right] \\
m = 0: &\qquad \Phi\left[(\beta X^{E^*} - c)/\sigma_0\right]
\end{aligned}
$$

Unidentified without restrictions on σ_1, σ_0

Potential Solution: Neumark (2012)

1. Define $\sigma^* = \frac{\sigma_1}{\sigma_0}$ and redefine all parameters in relative terms:

$$m = 1: \qquad \Phi\left[(\beta' X^{E^*} + \gamma' - c')/\sigma^*\right]$$
$$m = 0: \qquad \Phi\left[\beta' X^{E^*} - c'\right]$$

2. Identify β' from callback probability for $m = 0$
   - Requires that there is enough variation in X^E to affect the hiring decision
   - Also requires β' constant across groups
3. Back out the value of σ^* from the ratio $\frac{\beta}{\sigma^*}$ from the $m = 1$ callback probability
   - Can test whether $\sigma^* = 1$
4. Estimate this using heteroskedastic probit

Applying the Critique and Solution

Bertrand and Mullainathan (2004) was extremely influential in kicking off a new wave of CAS studies.

- White-sounding and Black-sounding names on both high- and low- quality resumes
- 1300 job ads with 5000 resumes sent
- Black callback rate 33% lower than the 9.65% callback rate for white applicants
- Higher quality resumes did not help Black candidates as much

Neumark (2012) reanalyze the BM data

- Black callback rate 25% lower than for white applicants
- Variance of unobservables contributes to discrimination *in favor* of Black candidates
- Original estimates understated discrimination
- $\sigma^* > 1$ but not statistically significant

Results from Neumark (2012)

Table 2
Heteroskedastic Probit Estimates for Callbacks: Full Specifications

	Males and females		Females	
	(1)	(2)	(3)	(4)
A. Estimates from basic probit (Table 1)				
Black	−0.030	−0.030	−0.030	−0.030
	(0.006)	(0.006)	(0.007)	(0.007)
B. Heteroskedastic probit model				
Black (unbiased estimates)	−0.024	−0.026	−0.026	−0.027
	(0.007)	(0.007)	(0.008)	(0.008)
Marginal effect of race through level	−0.086	−0.070	−0.072	−0.054
	(0.038)	(0.040)	(0.040)	(0.040)
Marginal effect of race through variance	0.062	0.045	0.046	0.028
	(0.042)	(0.043)	(0.045)	(0.044)
Standard deviation of unobservables, black/white	1.37	1.26	1.26	1.15
Wald test statistic, null hypothesis that ratio of standard deviations = 1 (*p*-value)	0.22	0.37	0.37	0.56
Wald test statistic, null hypothesis that ratios of coefficients for whites relative to blacks are equal, fully interactive probit model (*p*-value)	0.62	0.42	0.17	0.35
Test overidentifying restrictions: include in heteroskedastic probit model interactions for variables with |white coefficient| < |black coefficient|, Wald test for joint significance of interactions (*p*-value)	0.83	0.33	0.34	0.56
Number of overidentifying restrictions	3	6	2	6
Other controls				
Individual resume characteristics	X	X	X	X
Neighborhood characteristics		X		X
N	4,784	4,784	3,670	3,670

Note: See notes to Table 1. In the first row of Panel B the marginal effects in Equation 16 are reported, with the decomposition in Equations 16′ and 16″ immediately below; the marginal effects are evaluated at sample means. The standard errors for the two components of the marginal effects are computed using the delta method. Test statistics are based on the variance-covariance matrix clustering on the ad to which the applicants responded. Individual resume characteristics also include the variables listed separately in Table 1.

CAS Studies: Identifying Sources of Discrimination

List (2004) on discrimination by sportscard salesmen

- Discrimination defined as differential treatment of auditors based on race alone
- White males were most successful at bargaining
- More experienced salesmen discriminated more
  - Evidence for statistical discrimination or learning
- Lab evidence against taste-based descrimination
  - Dictator games, auctions, dealer perceptions etc.

Bartoš et al. (2016) designed a correspondence study to detect "attentional discrimination"

- 3 experiments in housing and labor markets in Germany and Czech Republic
- Varying amounts of information provided to DM
  - Information contingent on clicking link
- Found DMs provided less attention to minority candidates

Setup: Examining the Case of Bail Decisions

Thread of literature using data from bail judge decisions

- Data from NYC, Philadelphia, and Miami bail judge decisions
- Quasi-random variation in case assignment
- Easily defined DM goals: minimize pretrial misconduct risk
- Large sample sizes (e.g. Arnold et al. (2020) has 595,186 cases across 268 NYC judges)
- Note: They do note consider the value of the bail set

Modeling Framework

Notation from Canay et al. (2020)

$D = 1$ denotes release by the bail judge.

Y is the pretrial misconduct outcome.

Z indexes judges

$R \in \{w, b\}$ indexes race

V denotes non-race characteristics observed only by judge

$$Y = Y_1 D + Y_0(1 - D)$$
$$D = I\{\underbrace{\Lambda(R, V)}_{\text{cost}} \leq \underbrace{\tau(Z, R, V)}_{\text{benefit}}\}$$

*Note: $Y_0 = 0$ because if a defendant isn't released, they can't have pretrial misconduct

Costs and Benefits of Pretrial Release

$\mathbb{E}$[Cost]: Probability of pretrial misconduct

- $\Lambda(r, v) = \mathbb{E}[Y_1 - Y_0 | R, V] = \mathbb{P}\{Y_1 = 1 \mid R = r, V = v\}$
- Causal effect of release on misconduct
- For simplicity Y is binary but this can generalize
- Doesn't vary by judge due to random assignment

$\mathbb{E}$[Benefit]: Unobserved by analyst

- $\tau(z, r, v)$ can also be interpreted as a threshold for release
- Can differ by race, which is where bias would enter

Assume there are <u>marginal defendants</u> i.e. $\forall z, r, \exists V^*_{z,r} \in \mathcal{V}$ s.t.

$$\begin{aligned} \Lambda(r, v) &< \tau(z, r, v) \quad \text{for all } v < V^*_{z,r} \\ \Lambda(r, v) &> \tau(z, r, v) \quad \text{for all } v > V^*_{z,r} \end{aligned}$$

Formally Defining Bias

There are competing definitions. A judge can be biased if:

1. **Threshold is dependent on race for all $v \in \mathcal{V}$**

$$\tau(z, w, v) > \tau(z, b, v) \text{ for all } v \in \mathcal{V}$$

   - Judges see greater benefit of release for White defendants when defendants have the same characteristics V
   - Doesn't detect bias if only biased for $\mathcal{V}_1 \subset \mathcal{V}$

2. **Threshold is dependent on race "at the margin"**

$$\tau(z, w, V^*_{z,w}) > \tau(z, b, V^*_{z,b})$$

   - Generally $V^*_{z,w}$ isn't equal to $V^*_{z,b}$
   - Can conflate "disparity" and "discrimination"

See Arnold et al. (2018) and Canay et al. (2020) and the corresponding comments for a detailed discussion

Defining and Testing for Bias

We proceed using the definition outlined in Canay et al. (2020)

- **Def:** Judge z is racially unbiased if

$$\tau(z, r, v) = \tau(z, v) \text{ for all } v \in \mathcal{V}.$$

  - Recall this is a highly specific and clear-cut definition of bias

The outcome test proposed by Arnold et al. (2018) is:

- **Test:** There is bias against Black defendants if

$$\Lambda\left(w, V^*_{z,w}\right) > \Lambda\left(b, V^*_{z,b}\right)$$

  - At the release threshold for judge z, White defendants have a higher probability of pretrial misconduct

Critique from Canay et al. (2020)

Canay et al. (2020) argue that this outcome test is logically invalid.

- The test is logically valid if and only if

$$\underbrace{\text{sign}\left(\Lambda\left(w, V^*_{z,w}\right) - \Lambda\left(b, V^*_{z,b}\right)\right)}_{\text{rate of misconduct}} = \underbrace{\text{sign}(\tau(z,w,v) - \tau(z,b,v))}_{\text{bias}}$$

$$\forall v \in \mathcal{V} \text{ and } z \in \mathcal{Z}$$

  Marginal White and Black defendants should have the equal pretrial misconduct if and only if the judge is racially unbiased

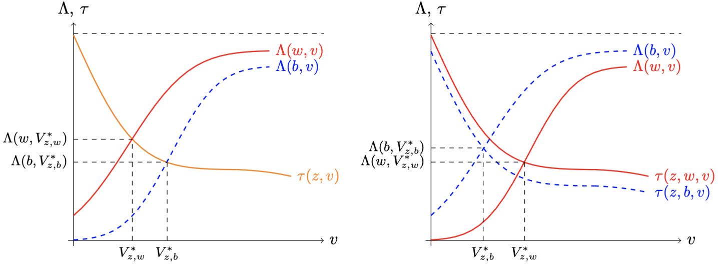
- However Canay et al. (2020) show cases where the test fails:
  1. No bias, costs vary by race, but benefits do not
     - If V correlates with race, then the marginal Black defendant has a lower rate of pretrial misconduct
  2. Bias exists, and both costs and benefits vary by race
     - Can lead to conclusion of bias in the opposite direction

Critique from Canay et al. (2020)

Case 1 (left panel): No bias, but discrimination detected

Note: no discrimination means $\tau(z, r, v) = \tau(z, v)$

Case 2 (right panel): Bias exists, detected discrimination is in the wrong direction



(a) Intuition behind part (i)

(b) Intuition behind part (ii)

Figure 1: Intuition behind Theorem 3.1

Proposed Solutions in Canay et al. (2020)

1. **Assume race doesn't affect pretrial misconduct**
   - Exclude race from cost function $\Lambda(r, v) = \Lambda(v)$ for all $v \in \mathcal{V}$
   - Assume strict monotonicity of cost function

   What does this mean?

     - Assumes judge's information set has all determinants of misconduct correlated with race

2. **Assume threshholds are unaffected by unobserved V**
   - Restrict benefits s.t. $\tau(z, r, v) = \tau(z, r)$ for all $r \in \{w, b\}$
   - The graph of τ from previous figures is flat

   What does this mean?

     - Assumes all bias manifests as racial bias
     - Biased judges are equally biased for all members of each race
     - Measurement error in misconduct uncorrelated with race
     - Judge accuracy at predicting misconduct doesn't vary by race

Other Solutions: Arnold et al. (2020)

Arnold et al. (2020) aim to get rid of OVB from observational data with 2 approaches:

1. Adjusting for pretrial misconduct risk
2. Hierarchical MTEs

Before discussing these, let us set up the framework.

Framework of Arnold et al. (2020)

Following Aigner and Cain (1977), they define a signalling model:

$$\underbrace{\nu_i}_{\text{signal}} = \underbrace{Y_i^*}_{\text{PO}} + \underbrace{\eta_i}_{\text{noise}}$$

$$D_i = 1\left[\pi_{R_i} \geq p\left(\nu_i, R_i\right)\right]$$

where π_{R_i} is the threshold
noise is distributed $\eta_i \mid Y_i^*, (R_i = r) \sim N\left(0, \sigma_r^2\right)$
and $p(\nu_i, R_i) = \Pr\left(Y_i^* = 1 \mid \nu_i, R_i\right)$ is the posterior risk

Definitions

- Racial bias: $\pi_b < \pi_w$
  - Includes animus and inaccurate stereotyping
- Statistical discrimination: Discrimination due to differences in μ_r misconduct risk and signal quality $\tau_r = \frac{1}{\sigma_r}$
- Racial discrimination: Observed discrimination due to both of the above

Defining Discrimination in Arnold et al. (2020)

They define racial discrimination for judge j as:

$$\Delta_j = \mathbb{E}\left[\mathbb{E}\left[D_{ij} \mid Y_i^*, R_i = w\right] - \mathbb{E}\left[D_{ij} \mid Y_i^*, R_i = b\right]\right]$$

where Y_i^* is the unobserved binary state of pretrial misconduct.

This can be rewritten as:

$$\Delta_j = \left(\delta_{jw}^T - \delta_{jb}^T\right)(1 - \bar{\mu}) + \left(\delta_{jw}^F - \delta_{jb}^F\right)\bar{\mu}$$

$$\begin{aligned} \text{where } \delta_{jr}^T &= \Pr\left(D_{ij} = 1 \mid Y_i^* = 0, R_i = r\right) && \text{true neg.} \\ \delta_{jr}^F &= \Pr\left(D_{ij} = 1 \mid Y_i^* = 1, R_i = r\right) && \text{false neg.} \\ \bar{\mu} &= E\left[Y_i^*\right] \end{aligned}$$

Arnold et al. (2020): First approach

1. Characterize OVB as the difference between:

   Judge-Specific Release Rate by Race

   $$\alpha_j = \left(\delta_{jw}^T (1 - \mu_w) + \delta_{jw}^F \mu_w\right) - \left(\delta_{jb}^T (1 - \mu_b) + \delta_{jb}^F \mu_b\right)$$

   Judge-Specific Racial Discrimination

   $$\Delta_j = \left(\delta_{jw}^T (1 - \bar{\mu}) + \delta_{jw}^F \bar{\mu}\right) - \left(\delta_{jb}^T (1 - \bar{\mu}) + \delta_{jb}^F \bar{\mu}\right)$$

   $$OVB = \xi_j \equiv \alpha_j - \Delta_j$$

2. Adjust outcomes by misconduct risk for each group

   $$\Delta_j = \mathbb{E}\left[\Omega_i D_i \mid R_i = w, Z_{ij} = 1\right] - \mathbb{E}\left[\Omega_i D_i \mid R_i = b, Z_{ij} = 1\right]$$

   $$\Omega_i = (1 - Y_i)\,\frac{1 - \bar{\mu}}{1 - \mu_{R_i}} + Y_i \frac{\bar{\mu}}{\mu_{R_i}} > 0$$

Arnold et al. (2020): First approach

3. Identify μ "at infinity" with super-lenient judge who releases everyone
   - Misconduct rate for these released defendants should be close to the population average
4. Else, use non-parametric or model-based extrapolation
   - Extrapolate release rates to a "super-lenient" judge
   - Use local linear regression (or similar methods)

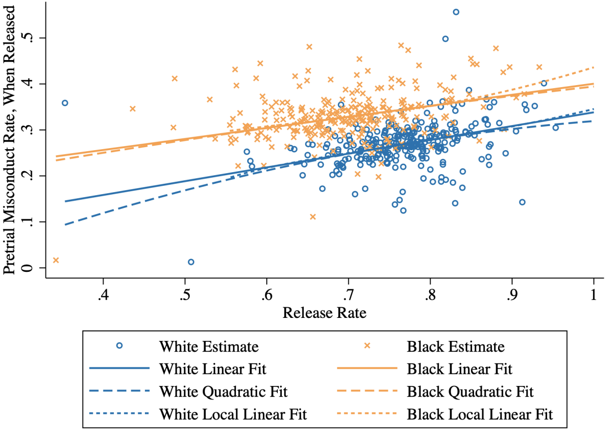
Results

Disparity in release rates: 6.8 percentage points

- Find that 62% can be explained by racial discrimination
- Remaining explained by unobserved diff. in misconduct risk
- 87% of judges discriminate (positive Δ_j)
- Results are more conservative than an uncorrected regression

Estimating Mean Misconduct Risk

Figure 2: Judge-Specific Release Rates and Conditional Misconduct Rates



Notes. This figure plots race-specific release rates for the 268 judges in our sample against rates of pretrial misconduct among the set of released defendants. All estimates adjust for court-by-time fixed effects. The figure also plots race-specific linear, quadratic, and local linear curves of best fit, obtained from judge-level regressions that inverse-weight by the variance of the estimated misconduct rate among released defendants. The local linear regressions use a Gaussian kernel with a race-specific rule-of-thumb bandwidth.

Arnold et al. (2020): Second Approach

First rewrite:

$$D_{ij} = 1\left[\pi_{jR_i} \geq p_j\left(\nu_{ij}; R_i\right)\right] = 1\left[\kappa_{jR_i} \geq Y_i^* + \eta_{ij}\right]$$

where $\kappa_{jr} = p_j^{-1}\left(\pi_{jr}; r\right)$ is the inverse of posterior risk

This function implicitly depends on risk signal quality: $\tau_{jr} = \frac{1}{\sigma_{jr}}$

There is no racial bias if (at the margin):

$$\pi_{jr} = E\left[Y_i^* \mid p_j\left(\nu_{ij}; r\right) = \pi_{jr}\right] = E\left[Y_i^* \mid Y_i^* + \eta_{ij} = \kappa_{jr}\right]$$

The goal is to estimate κ_{jr}, τ_{jr}. Why?

1. Identify statistical discrimination
   - Mean risk = $\mu_{jr}(\kappa) = E\left[Y_i^* \mid Y_i^* + \eta_{ij} = \kappa\right]$
   - Integrate over MTE (i.e. Y_i^*) curve based on risk signal quality
   - Precise signals $\implies$ steep MTEs
2. Allows us to sidestep IV monotonicity assumption ($\eta_{ij} = \eta_i$)

Arnold et al. (2020): Second Approach

Estimation Method

- Assume distributions for κ and τ:

$$\ln \tau_{jr} \sim N\left(\alpha_r, \psi_r^2\right) \text{ and } \kappa_{jr} \sim N\left(\gamma_r, \delta_r^2\right)$$

- Estimate hyperparameters using simulated minimum distance

Results

Rate of pretrial misconduct at the margin (MTE)

- White defendants: 0.651(SE: 0.033)
- Black defendants: 0.576 (SE: 0.021)
- Find higher average risk and less precise signals for Black defendants
- Find significant variation in signal quality

Hierarchical MTE Results

Table 5: Hierarchical MTE Model Estimates

	With Monotonicity			Without Monotonicity		
	White Defendants	Black Defendants	Diff.	White Defendants	Black Defendants	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)
Mean Misconduct Risk	0.346	0.423	-0.077	0.391	0.441	-0.050
	(0.008)	(0.009)	(0.012)	(0.007)	(0.007)	(0.010)
Mean Marginal Released Outcome	0.616	0.511	0.105	0.651	0.576	0.074
	(0.057)	(0.030)	(0.061)	(0.033)	(0.021)	(0.038)
Mean Signal Quality	1.712	0.963	0.749	1.385	0.970	0.416
	(0.219)	(0.141)	(0.271)	(0.104)	(0.073)	(0.128)
Marginal Outcome Std. Dev.	0.211	0.094	0.117	0.080	0.064	0.016
	(0.029)	(0.022)	(0.037)	(0.009)	(0.005)	(0.010)
Signal Quality Std. Dev.				0.196	0.163	0.033
				(0.038)	(0.017)	(0.041)
Covariance of Signal Quality and Marginal Released Outcomes				0.013	0.007	0.006
				(0.005)	(0.002)	(0.005)
Judges	268	268	–	268	268	–

Notes. This table reports simulated minimum distance estimates of moments of the MTE model described in Section 6. See Table A16 for underlying hyperparameter estimates. Columns 4-6 estimate the baseline model, while columns 1-3 impose conventional monotonicity. Robust standard errors, two-way clustered at the individual and the judge level, are obtained by a bootstrapping procedure and appear in parentheses.

Note: Monotonicity assumes judges rank defendants in the same way

Arnold et al. (2020) Policy Simulations

Investigate 2 policy exercises:

1. Attempt to close gap in unwarranted disparity
   - As estimated by their hierarchial MTE approach
2. Attempt to close gap in observed disparity
   - This maps to the more "standard" model that doesn't correct for OVB

Can do this by either:

- Increasing leniency for Black defendants
- Decreasing leniency for White defendants

They find that targeting either unwarranted or observed disparity can reduce discrimination.

Arnold et al. (2020) Policy Simulations

Table 6: Policy Simulations

	Baseline	Target Unwarranted Disparity Posteriors		Target Observational Disparity Posteriors	
		Increase Leniency	Decrease Leniency	Increase Leniency	Decrease Leniency
Panel A: Close All Disparities	(1)	(2)	(3)	(4)	(5)
Mean Unwarranted Disparity	0.047	0.000	0.000	-0.017	-0.019
	[0.037]	[0.020]	[0.026]	[0.020]	[0.026]
Mean Observational Disparity	0.065	0.017	0.019	0.000	-0.000
	[0.038]	[0.020]	[0.026]	[0.019]	[0.026]
Racial Bias	0.074	0.039	0.013	0.025	-0.011
	[0.078]	[0.068]	[0.055]	[0.070]	[0.053]
Panel B: Close Top-Quintile Disparities					
Mean Unwarranted Disparity		0.030	0.030	0.026	0.026
		[0.035]	[0.037]	[0.038]	[0.041]
Mean Observational Disparity		0.047	0.048	0.044	0.043
		[0.035]	[0.037]	[0.039]	[0.040]
Racial Bias		0.062	0.051	0.059	0.045
		[0.075]	[0.076]	[0.076]	[0.080]
Observations	268	268	268	268	268

Notes. This table reports the results from a series of policy simulations. Column 1 reports the mean unwarranted disparity, observational disparity, and racial bias across judges and 250 simulations of the hierarchical MTE model. Average standard deviations across judges are included in brackets. Simulations are based on the estimates from columns 2 and 4 of Appendix Table A16. Column 2 of Panel A recomputes the statistics for a counterfactual in which the lower of the Black or white release rate of each judge is raised to equalize unwarranted disparity posteriors, while column 3 of Panel A does the same by lowering one of the two release rates. Columns 4 and 5 of Panel A instead adjust release rates to equalize observational disparity posteriors. Panel B conducts the counterfactual exercises only on judges ranked in the top quintile of unwarranted (columns 2 and 3) or observational (columns 4 and 5) disparity posteriors. Estimates of the model hyperparameters and empirical Bayes posteriors of all judge-specific parameters are recomputed in each simulation draw via the SMD procedure outlined in the text, using moments simulated according to the estimated distribution of reduced-form estimates in Figure 2.

Arnold et al. (2020) Other Counterfactuals

Appendix Table A19: Unwarranted Disparity Decompositions

	Baseline	No Racial Bias	Equal Signal Quality	Both
Panel A: Change Black Parameters	(1)	(2)	(3)	(4)
Unwarranted Disparity	0.047	-0.042	0.095	0.039
Release Rates (W/B)	0.768 / 0.703	0.768 / 0.795	0.768 / 0.652	0.768 / 0.709
Racial Bias	0.074	0.000	0.074	0.000
Marginal Outcomes (W/B)	0.650 / 0.577	0.650 / 0.650	0.650 / 0.577	0.650 / 0.650
Signal Quality (W/B)	1.386 / 0.970	1.386 / 0.970	1.386 / 1.386	1.386 / 1.386
Panel B: Change White Parameters				
Unwarranted Disparity		-0.006	0.136	0.062
Release Rates (W/B)		0.716 / 0.703	0.853 / 0.703	0.781 / 0.703
Racial Bias		0.000	0.074	0.000
Marginal Outcomes (W/B)		0.577 / 0.577	0.650 / 0.577	0.577 / 0.577
Signal Quality (W/B)		1.386 / 0.970	0.970 / 0.970	0.970 / 0.970
Judges	268	268	268	268

Notes. Column 1 of this table reports average unwarranted disparity and racial bias across judges and 250 simulations of the hierarchical MTE model, along with average release rates, marginal released outcomes, and signal quality of Black and white defendants. Simulations are based on the estimates from columns 2 and 4 of Appendix Table A16. Column 2 recomputes the statistics for a counterfactual in which Black (Panel A) or white (Panel B) release rates are set to eliminate racial bias, while column 3 adjusts Black (Panel A) or white (Panel B) signal quality to equalize signal quality across race. Column 4 applies both counterfactuals simultaneously.

Implicit Association Tests (IATs)

Pioneered by Greenwald, McGhee, and Schwartz (1998) to elicit unconscious associations of groups with "good" or "bad" valence.

Reaction times are a proxy for the magnitude of bias.

Types

- Attitude IAT: "emotional prejudice"
- Stereotype IAT: belief in stereotypes

Pros and Cons

+ Supposedly better than self reports
- Low retest reliability: median of 0.56 (Greenwald et al., 2009)
- Applicability to hiring settings?
- Can be influenced/manipulated with stimuli

IAT Example

Implicit Association Test

Next, you will use the 'E' and 'I' computer keys to categorize items into groups as fast as you can. These are the four groups and the items that belong to each:

Category	Items
Good	Appealing, Excitement, Excellent, Spectacular, Delightful, Attractive, Triumph, Cherish
Bad	Detest, Angry, Abuse, Negative, Pain, Horrible, Horrific, Hurtful
African Americans	
European Americans	

There are seven parts. The instructions change for each part. Pay attention!

Continue

IATs: Validity in Predicting Behavior

Two meta-analyses of interest:

- Greenwald et al. (2009)
  - Generally IAT and self-reports are similarly predictive but self-reports have a larger variance
  - For Black-White interracial behavior, IATs were better
- Oswald et al. (2013)
  - Does not average effects that are dependent on same sample
  - Includes more studies / samples
    - Ignored sterotype-based or attitude-based consistency Greenwald et al. (2014)
  - Fixes errors from Greenwald et al. (2009)
  - Smaller estimates of predictive validity, mostly from brain activity

Oswald et al. (2013) Results by Outcome

Table 1

Meta-Analysis of Implicit-Criterion Correlations (ICCs): Overall and by Subgroups

Criterion	k (s; N_{total})	$\hat{\rho}$ [95% CI]	$\hat{\tau}$	M	SD
All effects: Overall	298 (86; 17,470)	.14 [.10, .19]	.17	.12	.24
Interpersonal behavior	11 (6; 796)	.14 [.03, .26]	.12	.21	.15
Person perception	138 (46; 7,371)	.13 [.07, .18]	.13	.10	.21
Policy preference	21 (9; 4,677)	.13 [.07, .19]	.03	.14	.09
Microbehavior$_a$	96 (21; 3,879)	.07 [–.03, .18]	.19	.10	.24
Response time	6 (5; 300)	.19 [.02, .36]	.27	.31	.28
Brain activity$_a$	26 (8; 447)	.42 [.11, .73]	.68[b]	.26	.40
Black vs. White groups: Overall	206 (63; 9,899)	.15 [.09, .21]	.19	.13	.26
Interpersonal behavior	10 (5; 691)	.14 [.01, .28]	.14	.22	.16
Person perception	75 (30; 3,564)	.13 [.08, .19]	.12	.09	.22
Policy preference$_a$	8 (5; 1,855)	.10 [.02, .19]	.05	.09	.10
Microbehavior$_b$	87 (18; 3,162)	.07 [–.06, .19]	.22	.10	.25
Response time[a]	6 (5; 300)	.19 [.02, .37]	.27	.31	.28
Brain activity$_{a,b}$	20 (8; 327)	.43 [.12, .73]	.67[b]	.30	.42
Ethnic minority vs. majority groups: Overall	92 (24; 7,571)	.12 [.06, .19]	.12	.12	.18
Interpersonal behavior$_a$	1 (1; 105)	.19 [[c]]	—	.19	[c]
Person perception	63 (16; 3,807)	.11 [–.01, .23]	.15	.11	.19
Policy preference	13 (4; 2,822)	.16 [.08, .25]	.00	.17	.07
Microbehavior	9 (3; 717)	.11 [–.09, .31]	.14	.11	.19
Response time	—	—	—	—	—
Brain activity$_a$	6 (1; 120)	.11 [[c]]	—	.11	.27

k = number of effects; s = number of independent samples within each category; $\hat{\rho}$ = meta-analytically estimated population correlation; CI = confidence interval; $\hat{\tau}$ = random-effects standard deviation estimate; M = unweighted mean; SD = unweighted standard deviation

Oswald et al. (2013) Results by Outcome

Table 6
Meta-Analysis of Implicit–Explicit Correlations (IECs) and Explicit-Criterion Correlations (ECCs) by Explicit Measure

Explicit measure	k (s; N_{total})	$\hat{\rho}$ [95% CI]	$\hat{\tau}$	M	SD
		IEC			
Black vs. White groups: Overall	105 (39; 10,739)	.14 [.09, .19]	.13	.13	.15
Thermometer	24 (15; 2,534)	.09 [–.05, .24]	.17	.14	.19
Other existing measure	39 (26; 3,491)	.15 [.09, .21]	.08	.16	.14
Created measure	42 (10; 4,714)	.14 [.05, .24]	.15	.10	.14
Ethnic minority vs. majority groups: Overall	19 (12; 1,339)	.16 [.09, .23]	.00	.13	.14
Thermometer$_a$	3 (2; 511)	.23 [.19, .27]	.08	.31	.11
Other existing measure	6 (5; 402)	.13 [.02, .24]	.00	.13	.09
Created measure$_a$	10 (7; 426)	.07 [–.03, .17]	.00	.07	.12
		ECC			
Black vs. White groups					
Thermometer	29 (18; 2,249)	.11 [–.05, .27]	.16	.07	.24
Other existing measure	112 (31; 6,008)	.11 [.05, .18]	.20	.06	.20
Created measure	57 (14; 4,449)	.06 [.00, .13]	.06	.07	.15
Ethnic minority vs. majority groups					
Thermometer	10 (5; 2,187)	.06 [–.16, .28]	.16	.14	.18
Other existing measure	14 (6; 917)	.09 [–.03, .22]	.04	.05	.10
Created measure	41 (8; 2,413)	.24 [.09, .40]	.18	.15	.21

Note. All effects were coded such that positive correlations are in the direction of promajority group or antiminority group responses or behaviors. The correlation between dependent effects is assumed to be .50. The $\hat{\rho}$ for each category is based on a moderated meta-analysis across categories, where dependent effect sizes (both within and across categories) are accounted for (Hedges et al., 2010), and the overall random-effects variance (tau-squared) weight is applied. $\hat{\tau}$ is also independently estimated within each category in separate analyses. With regard to Heider and Skowronski (2007), these analyses incorporate only the difference score ECCs. Effects sharing subscripts within a category set are statistically significantly different from one another ($p <$ .05). k = number of effects; s = number of independent samples within each category (may not add up to the overall s because of sample overlap across categories); $\hat{\rho}$ = meta-analytically estimated population correlation; CI = confidence interval; $\hat{\tau}$ = random-effects standard deviation estimate; M = unweighted mean; SD = unweighted standard deviation.

Outstanding Issues

- Incorrect beliefs among DMs (see e.g. Bohren et al. (2019)
  - Only 10% of papers in their survey consider this
  - Identify inaccurate beliefs by experimentally eliciting the true distribution of outcomes for a task
  - Provide additional information for DMs to update
- Minority beliefs about discrimination
  - Public health work on how perceived discrimination affects health (see e.g. Lewis et al. (2015))
- Network/structural explanations (see Small and Pager (2020))
  - Institutional discrimination through formal and informal norms
  - Homophilic research networks
  - "Race-neutral" policies applied in context

References I

Altonji, J. G. and Pierret, C. R. (2001). Employer Learning and Statistical Discrimination*. *The Quarterly Journal of Economics*, 116(1):313–350.

Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial Bias in Bail Decisions*. *The Quarterly Journal of Economics*, 133(4):1885–1932.

Arnold, D., Dobbie, W. S., and Hull, P. (2020). Measuring Racial Discrimination in Bail Decisions. NBER Working Papers 26999, National Bureau of Economic Research, Inc.

Bartoš, V., Bauer, M., Chytilová, J., and Matějka, F. (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, 106(6):1437–75.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

References II

Bertrand, M. and Duflo, E. (2016). Field experiments on discrimination. Working Paper 22014, National Bureau of Economic Research.

Bertrand, M. and Mullainathan, S. (2004). Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2019). Inaccurate statistical discrimination: An identification problem. Working Paper 25935, National Bureau of Economic Research.

Canay, I. A., Mogstad, M., and Mountjoy, J. (2020). On the use of outcome tests for detecting bias in decision making. Working Paper 27802, National Bureau of Economic Research.

Foster, A. D. and Rosenzweig, M. (1993). Information, learning, and wage rates in low-income rural areas. *Journal of Human Resources*, 28(4):759–790.

References III

Greenwald, A., Banaji, M., and Nosek, B. (2014). Statistically small effects of the implicit association test can have societally large effects. *Journal of personality and social psychology*, 108.

Greenwald, A., Poehlman, T., Uhlmann, E., and Banaji, M. (2009). Understanding and using the implicit association test: Iii. meta-analysis of predictive validity. *Journal of personality and social psychology*, 97:17–41.

Heckman, J. J. (1998). Detecting discrimination. *Journal of Economic Perspectives*, 12(2):101–116.

Lewis, T., Cogburn, C. D., and Williams, D. R. (2015). Self-reported experiences of discrimination and health: scientific advances, ongoing controversies, and emerging issues. *Annual review of clinical psychology*, 11:407–440.

List, J. A. (2004). The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field*. *The Quarterly Journal of Economics*, 119(1):49–89.

References IV

Neumark, D. (2012). Detecting discrimination in audit and correspondence studies. *Journal of Human Resources*, 47(4):1128–1157.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3):799–866.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709.

Oswald, F., Mitchell, G., Blanton, H., Jaccard, J., and Tetlock, P. (2013). Predicting ethnic and racial discrimination: A meta-analysis of iat criterion studies. *Journal of Personality and Social Psychology*, 105(2):171–192.

Small, M. L. and Pager, D. (2020). Sociological perspectives on racial discrimination. *Journal of Economic Perspectives*, 34(2):49–67.

Altonji, Joseph G., and Rebecca M. Blank. (1999)

Unemployment Rates by Demographic in men and women

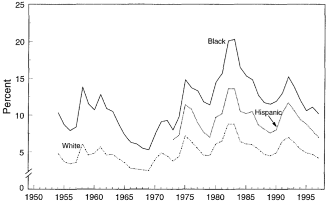


Fig. 3. Male unemployment rates (annual averages). Source: Bureau of Labor Statistics.

decline in their labor force involvement, with the largest declines among black men. Women have shown dramatic increases in labor force participation over these years.

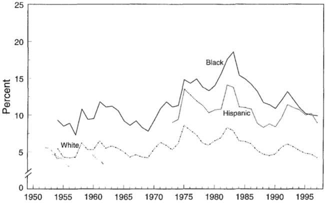


Fig. 4. Female unemployment rates (annual averages). Source: Bureau of Labor Statistics.

Altonji, Joseph G., and Rebecca M. Blank. (1999)

Personal Characteristics

Table 2
Personal characteristics by race and gender, 1996[a]

	All	White males	Black males	Hispanic males	White females	Black females	Hispanic females
(1) Share of all persons	1.000	0.412	0.052	0.055	0.378	0.059	0.039
Education							
(2) Less than high school	0.159	0.118	0.232	0.447	0.105	0.214	0.434
(3) High school	0.331	0.321	0.386	0.275	0.346	0.342	0.275
(4) Some post-HS training	0.281	0.279	0.272	0.192	0.300	0.306	0.215
(5) College degree	0.158	0.184	0.079	0.061	0.177	0.107	0.059
(6) More than college	0.072	0.098	0.030	0.025	0.072	0.031	0.016
(7) Potential experience	23.7	24.1	23.2	22.6	23.8	22.9	22.9
(Age-educ-5)	(23.3)	(23.5)	(25.1)	(21.6)	(23.6)	(23.9)	(21.1)
(8) Share married	0.570	0.605	0.361	0.483	0.624	0.307	0.540
(9) No. children age less than 6	0.24	0.21	0.15	0.29	0.24	0.30	0.41
	(5.07)	(5.02)	(5.03)	(4.99)	(5.07)	(5.69)	(5.19)
(10) Total no. children (age < 18)	0.71	0.63	0.45	0.75	0.73	0.87	1.08
	(7.01)	(6.85)	(7.15)	(6.84)	(6.88)	(7.74)	(6.85)
(11) Share in SMSA[b]	0.489	0.452	0.608	0.655	0.448	0.599	0.658
Region							
(12) New England	0.051	0.060	0.022	0.019	0.059	0.021	0.024
(13) Middle Atlantic	0.145	0.146	0.148	0.125	0.144	0.159	0.146
(14) East-North Central	0.164	0.180	0.149	0.058	0.182	0.149	0.051
(15) West-North Central	0.068	0.082	0.037	0.014	0.081	0.028	0.012
(16) South Atlantic	0.180	0.164	0.324	0.115	0.166	0.332	0.117
(17) East-South Central	0.061	0.060	0.107	0.005	0.062	0.115	0.004
(18) West-South Central	0.109	0.093	0.112	0.213	0.095	0.117	0.212
(19) Mountain	0.060	0.063	0.012	0.095	0.061	0.012	0.093
(20) Pacific	0.161	0.152	0.088	0.357	0.148	0.067	0.340

[a] Source: Current Population Survey, March 1996. Weighted estimates, standard deviations are in parentheses.
[b] Defined as residing in SMSA with at least one million inhabitants.

Altonji, Joseph G., and Rebecca M. Blank. (1999)

Differences in Occupations and Industries

Table 3
Occupation and industry by race and gender, 1996[a]

	All	White males	Black males	Hispanic males	White females	Black females	Hispanic females
Occupation							
(1) Executive, administrative, and managerial	0.107	0.141	0.051	0.050	0.104	0.064	0.047
(2) Professional specialty	0.114	0.121	0.057	0.044	0.139	0.081	0.049
(3) Technicians	0.024	0.024	0.016	0.015	0.028	0.022	0.018
(4) Sales	0.093	0.107	0.050	0.061	0.095	0.073	0.077
(5) Administrative support	0.116	0.048	0.070	0.053	0.187	0.168	0.130
(6) Private household service	0.005	0.000	0.001	0.002	0.005	0.013	0.027
(7) Protective service	0.013	0.022	0.028	0.017	0.003	0.012	0.004
(8) Other service occupation	0.087	0.049	0.111	0.122	0.102	0.152	0.121
(9) Farming, forestry and fishing	0.085	0.167	0.110	0.161	0.014	0.014	0.016
(10) Precision production, craft and repair	0.053	0.059	0.085	0.101	0.032	0.060	0.067
(11) Machine operators, assemblers, etc.	0.033	0.060	0.071	0.062	0.006	0.009	0.004
(12) Transportation and material moving	0.033	0.044	0.094	0.088	0.011	0.017	0.014
(13) Handlers, equipment cleaners, etc.	0.020	0.030	0.013	0.074	0.008	0.001	0.016

Oswald et al. (2013) Criteria I

- **Brain activity:** measures of neurological activity while participants processed information about a member of a majority or minority group
- **Response time:** measures of stimulus response latencies, such as Correll's shooter task
- **Microbehavior:** measures of nonverbal and subtle verbal behavior, such as displays of emotion and body posture during intergroup interactions and assessments of interaction quality based on reports of those interacting with the participant or coding of interactions by observers (this category encompasses behaviors Sue et al., 2007, characterized as "racial microaggressions")
- **Interpersonal behavior:** measures of written or verbal behavior during an intergroup interaction or explicit expressions of preferences in an intergroup interaction, such as a choice in a Prisoner's Dilemma game or choice of a partner for a task

Oswald et al. (2013) Criteria II

- **Person perception:** explicit judgments about others, such as ratings of emotions displayed in the faces of minority or majority targets or ratings of academic ability
- **Policy/political preferences:** expressions of preferences with respect to specific public policies that may affect the welfare of majority and minority groups (e.g., support for or opposition to affirmative action and deportation of illegal immigrants) and particular political candidates (e.g., votes for Obama or McCain in the 2008 presidential election).