Diversity and Global Policy: Earnings and Wealth Inequality

Germain Gauthier

Bocconi University

Recap of the Last Lecture

There exist large gender gaps worldwide in terms of economic opportunities, education, political empowerment, and health.

Cultural values and stereotypes reflect these inequalities and potentially allow them to persist.

But some of these observed gaps may result from **confounders** that we would like to control for.

e.g., perhaps women earn less because they choose to work in different industries as men

How can we move beyond basic, raw empirical evidence?

Roadmap for Today

In other words, how do we measure inequalities between groups while conditioning on other relevant factors?

We will discuss standard descriptive/non-experimental methods and apply these tools to quantify racial and gender gaps in terms of earnings and wealth inequality.

1. Measurement Tools
2. Gender Earnings Gap
3. Racial Earnings Gap
4. Racial Wealth Inequality

Linear Regression

Standard approach is to measure gaps (*e.g., the gender wage gap*) at the (conditional) mean, using:

$$\log Y_i = \alpha + \beta \text{Group}_i + \gamma' X_i + \varepsilon_i.$$

Y_i is the outcome (*e.g., wage*). Taking the log limits the effect of outliers and makes it easily interpretable.

Group_i is a dummy variable for race, gender, etc.

X_i is a vector of control variables that may explain part of the observed raw gap.

ε_i is the error term.

Quiz

How should we interpret α in this equation without controls?

How should we interpret β in this equation without controls?

... in this equation with controls?

Under which assumption can we interpret the coefficients as causal?

Is this a plausible assumption?

Oaxaca-Blinder Decomposition

A classic tool for measuring discrimination: the Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973).

OB method decomposes a difference between groups into:

- a component explained by observed characteristics
- a component explained by returns to characteristics

Consider individuals in two groups, $G_i \in \{A, B\}$.

Group average outcomes are $\bar{Y}_A$ and $\bar{Y}_B$, where $\bar{Y}_G = \mathbb{E}[Yi \mid Gi = g]$.

We aim to explain group differences with the vector of observed covariates X_i.

Oaxaca-Blinder Decomposition

The quantity to be explained is

$$\Delta = \bar{Y}_A - \bar{Y}_B.$$

In practice, we run a separate regression for each group g:

$$Y_i = X_i'\beta_g + \varepsilon_i.$$

By construction, OLS regression fits each group's average:

$$\bar{Y}_g = \bar{X}_g'\beta_g.$$

A Simple R Simulation

```
> # Simulate a linear regression DGP with two covariates and an intercept
> N = 10000
> beta1 = 2
> beta2 <- 3
> alpha = 1
> x1 <- runif(N)
> x2 <- runif(N)
> epsilon <- rnorm(N,0,10)
> y <- alpha +  beta1*x1 + beta2*x2 + epsilon
>
> # Fit a linear regression model by OLS
> df <- data.frame(y,x1,x2)
> model <- lm(y ~ x1 + x2, data = df)
> summary(model)

Call:
lm(formula = y ~ x1 + x2, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-37.563  -6.639   0.018   6.684  35.070

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.8100     0.2616   3.096  0.00197 **
x1            2.2762     0.3376   6.743 1.64e-11 ***
x2            2.7099     0.3416   7.933 2.36e-15 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.84 on 9997 degrees of freedom
Multiple R-squared:  0.01071,   Adjusted R-squared:  0.01052
F-statistic: 54.13 on 2 and 9997 DF,  p-value: < 2.2e-16

>
> # Mean of the outcome variable in the sample
> mean(y)
[1] 3.325167
>
> # Predicted mean of the outcome based on the model estimates
> unname(model$coefficients['(Intercept)'] + mean(x1)*model$coefficients['x1'] + mean(x2)*model$coefficients['x2'])
[1] 3.325167
```

Oaxaca-Blinder Decomposition

We can therefore write:

$$\Delta = \bar{Y}_A - \bar{Y}_B = \bar{X}_A'\beta_A - \bar{X}_B'\beta_B.$$

Rearranging we get:

$$\Delta = \underbrace{(\bar{X}_A - \bar{X}_B)'\beta_A}_{\text{Explained by } X's} + \underbrace{\bar{X}_B'(\beta_A - \beta_B)}_{\text{Explained by } \beta's}.$$

Oaxaca-Blinder Decomposition

We can therefore write:

$$\Delta = \bar{Y}_A - \bar{Y}_B = \bar{X}_A' \beta_A - \bar{X}_B' \beta_B.$$

Rearranging we get:

$$\Delta = \underbrace{(\bar{X}_A - \bar{X}_B)' \beta_A}_{\text{Explained by } X's} + \underbrace{\bar{X}_B' (\beta_A - \beta_B)}_{\text{Explained by } \beta's}.$$

The first term answers the question: **How much more would A's make than B's if both groups were paid like A's observables?**

e.g., How much more would white workers make than black workers if both racial groups had the same returns to education as white workers?

Oaxaca-Blinder Decomposition

We can therefore write:

$$\Delta = \bar{Y}_A - \bar{Y}_B = \bar{X}_A'\beta_A - \bar{X}_B'\beta_B.$$

Rearranging we get:

$$\Delta = \underbrace{(\bar{X}_A - \bar{X}_B)'\beta_A}_{\text{Explained by } X's} + \underbrace{\bar{X}_B'(\beta_A - \beta_B)}_{\text{Explained by } \beta's}.$$

The second term answers the question: **How much more would A's make than B's if both groups had group B's observables?**

e.g., How much more would white workers make than black workers if both racial groups had the same educational level as black workers?

Alternative Oaxaca-Blinder Decomposition

We wrote:

$$\Delta = \underbrace{(\bar{X}_A - \bar{X}_B)'\beta_A}_{\text{Explained by } X's} + \underbrace{\bar{X}_B'(\beta_A - \beta_B)}_{\text{Explained by } \beta's}.$$

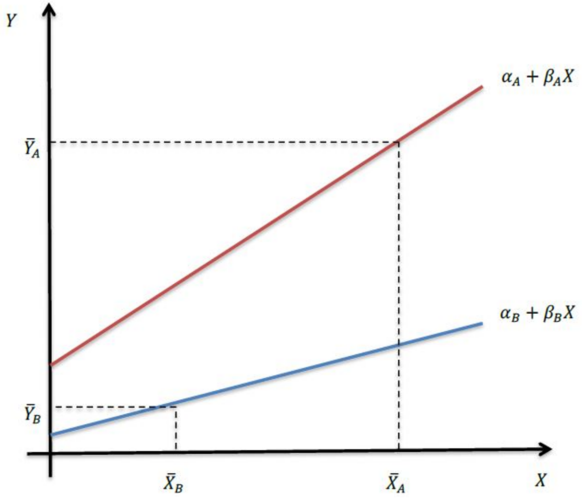
We can also write the alternative decomposition:

$$\Delta = \underbrace{(\bar{X}_A - \bar{X}_B)'\beta_B}_{\text{Explained by } X's} + \underbrace{\bar{X}_A'(\beta_A - \beta_B)}_{\text{Explained by } \beta's}.$$

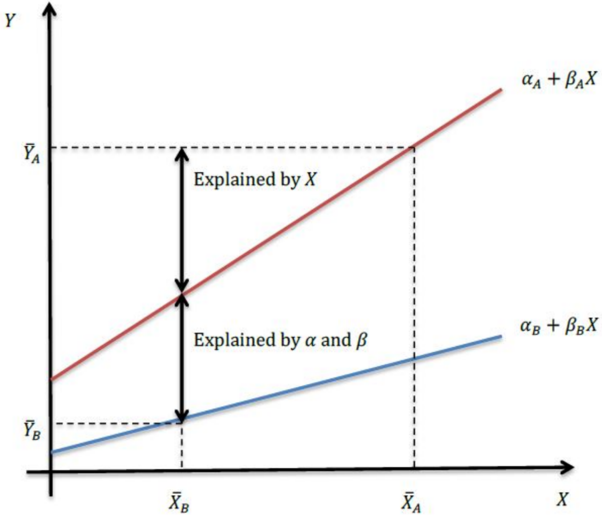
New first term answers the question: How much more would A's make than B's if both groups were paid like B's observables?

New second term answers the question: How much more would A's make than B's if both groups had the A's observables?

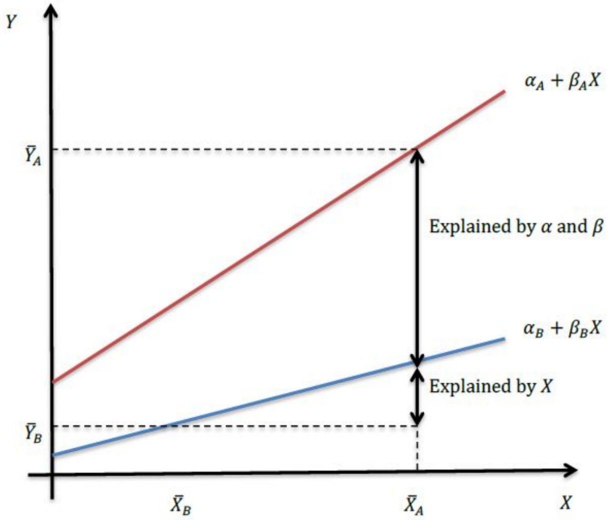
Oaxaca-Blinder Decomposition – Graphical Illustration



Oaxaca-Blinder Decomposition – Decomposition 1



Oaxaca-Blinder Decomposition – Decomposition 2



Quiz

How can we interpret an Oaxaca-Blinder decomposition?

Under which assumptions can we attribute its second term to discrimination?

Is this a plausible assumption?

Quantile Regression

Group differences can also be measured at different percentiles q (e.g., median and 90th pct).

Two common measures of gaps:

- **The level gap**

$$\log Y_i = \alpha(q) + \beta(q)\text{Group}_i + \gamma(q)X_i + \varepsilon_i(q).$$

→ i.e., difference in earnings between groups at the q^{th} percentiles of the outcome distribution.

- **The rank gap**

$$\text{Rank}(Y_i) = \alpha(q) + \beta(q)\text{Group}_i + \gamma(q)X_i + \varepsilon_i(q).$$

→ i.e., difference in rank between where an individual in a group at rank q would be if he were in another group

Quantile Regression Gaps – Graphical Illustration

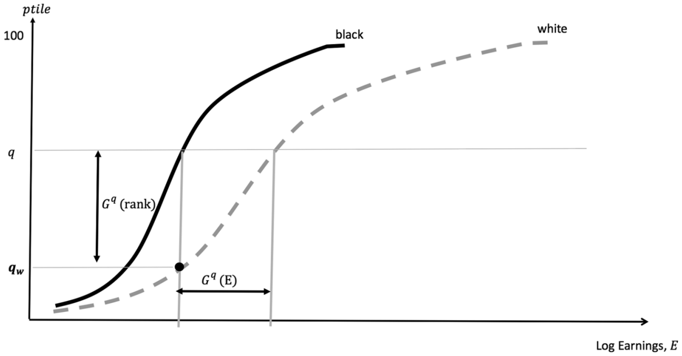


FIGURE I

Racial Earnings Level and Earning Rank Gaps

Quiz

Why would we measure group differences at different percentiles rather than the mean?

Decomposing Quantile Regressions

We can do a similar exercise as the Oaxaca-Blinder decomposition with quantile regressions (Chernozhukov et al., 2013).

Define $F_{Y[A,A]}$ as the **observed distribution** of Y with group A's characteristics and wage function, and $F_{Y[B,B]}$ as the counterpart for group B:

$$F_{Y[A,A]} = \int F_{Y_A|X_A}(y|x)dF_{X_A}(x) \;\;\&\;\; F_{Y[B,B]} = \int F_{Y_B|X_B}(y|x)dF_{X_B}(x)$$

Now define the **hypothetical wage distribution**, $F_{Y[A,B]}$, that group B would have, if they were rewarded with group A's wage function:

$$F_{Y[A,B]} = \int F_{Y_A|X_A}(y|x)dF_{X_B}(x)$$

Decomposing Quantile Regressions

Then one can write:

$$F_{Y[A,A]} - F_{Y[B,B]} = \underbrace{F_{Y[A,A]} - F_{Y[A,B]}}_{\text{Explained by } X's} + \underbrace{F_{Y[A,B]} - F_{Y[B,B]}}_{\text{Explained by } \beta_q' s}$$

Decomposing Quantile Regressions

Then one can write:

$$F_{Y[A,A]} - F_{Y[B,B]} = \underbrace{F_{Y[A,A]} - F_{Y[A,B]}}_{\text{Explained by } X's} + \underbrace{F_{Y[A,B]} - F_{Y[B,B]}}_{\text{Explained by } \beta_q' s}$$

The first term reflects differences in personal characteristics between groups A and B, assuming group A's wage function.

The second term reflects differences in wage functions between groups A and B, assuming group B's distribution of personal characteristics.

Decomposing Quantile Regressions

Then one can write:

$$F_{Y[A,A]} - F_{Y[B,B]} = \underbrace{F_{Y[A,A]} - F_{Y[A,B]}}_{\text{Explained by } X's} + \underbrace{F_{Y[A,B]} - F_{Y[B,B]}}_{\text{Explained by } \beta_q' s}$$

The first term reflects differences in personal characteristics between groups A and B, assuming group A's wage function.

The second term reflects differences in wage functions between groups A and B, assuming group B's distribution of personal characteristics.

Decomposing Quantile Regressions

Then one can write:

$$F_{Y[A,A]} - F_{Y[B,B]} = \underbrace{F_{Y[A,A]} - F_{Y[A,B]}}_{\text{Explained by } X' s} + \underbrace{F_{Y[A,B]} - F_{Y[B,B]}}_{\text{Explained by } \beta_q' s}$$

The first term reflects differences in personal characteristics between groups A and B, assuming group A's wage function.

The second term reflects differences in wage functions between groups A and B, assuming group B's distribution of personal characteristics.

Confidence intervals are obtained via a bootstrap procedure.

Let's put those tools to use!

Gender Wage Gap (Blau and Kahn, 2017)

New empirical evidence of the extent and trends in the gender wage gap in the United States

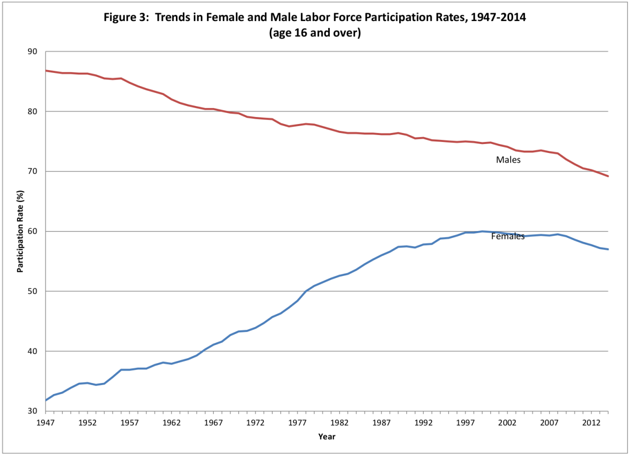
Data:

- Panel Study of Income Dynamics (PSID)
- Current Population Survey (CPS)
- Between 1980 and 2010

Methods:

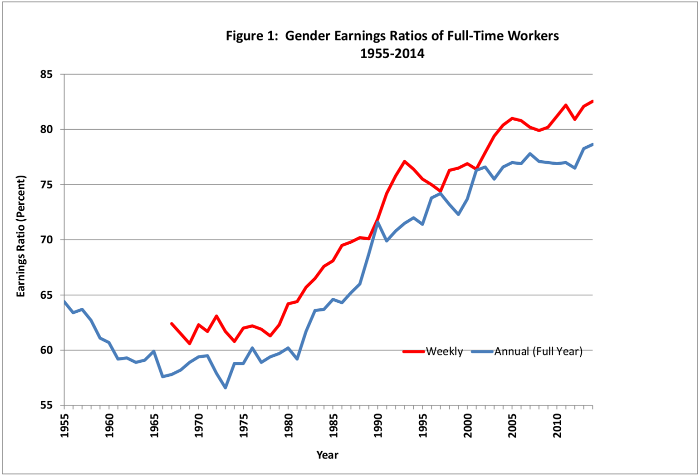
- Oaxaca-Blinder decompositions for differences at the mean
- Quantile regressions to study more precisely trends along different earnings quantiles

Women and Labor Force Participation



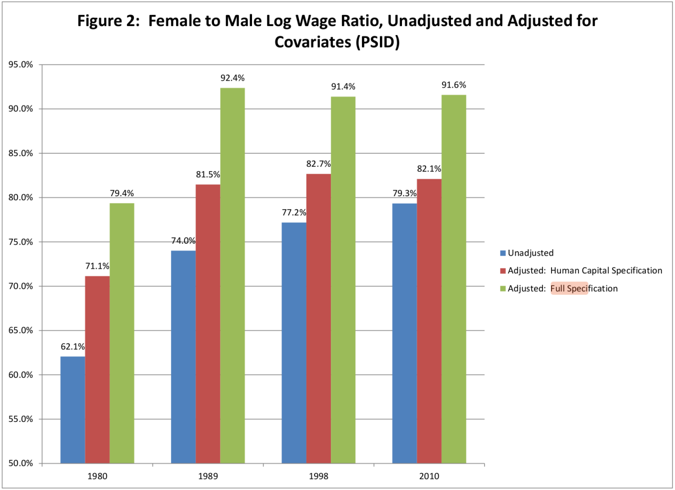
Notes: Updated version of Figure from Blau, Ferber and Winkler (2014) based on data from the Current Population Survey available at www.bls.gov and Employment & Earnings, various issues.

Gender Wage Gap – Raw Means



Notes: Updated version of Figure 7-2 from Blau, Ferber, and Winkler (2014); for additional information on references, see p. 148. Workers aged 16 and over from 1979 onward, and 14 and over prior to 1979.

Oaxaca-Blinder Decomposition



Source: Authors' calculations from Panel Study of Income Dynamics (PSID) data. See text for definitions.

Gender Wage Gap – Raw Quantiles

Table 1: Unadjusted Female/Male Log Hourly Wage Ratios, Full Time Workers

	Sample Size					
Year	Men	Women	Mean	10th Percentile	50th Percentile	90th Percentile
			Panel Study of Income Dynamic (PSID)			
1980	2282	1491	62.1%	64.8%	60.1%	62.4%
1989	2617	2068	74.0%	76.3%	72.4%	74.6%
1998	2391	2146	77.2%	80.3%	79.8%	73.8%
2010	2368	2456	79.3%	81.5%	82.4%	73.9%
			March Current Populations Survey (CPS)			
1980	21428	13484	63.5%	68.7%	61.9%	64.3%
1989	21343	16487	72.4%	78.1%	72.2%	71.4%
1998	17520	14231	77.1%	81.3%	76.2%	76.1%
2010	24229	20718	82.3%	87.6%	82.2%	76.6%

Notes: Sample includes nonfarm wage and salary workers age 25-64 with at least 26 weeks of employment. Entries are exp(D), where D is the female mean log wage, 10th, 50th or 90th percentile log wage minus the corresponding male log wage.

Decomposition with Quantile Regressions

Table 6: Decomposition of the Gender Log Wage Gap by Unconditional Distribution Percentile (PSID)

	1980		2010	
	Specification		Specification	
Percentile	**Human Capital**	**Full**	**Human Capital**	**Full**
A. Effect of Covariates				
10th percentile	0.1767	0.2729	0.0721	0.1648
	(0.0234)	(0.0374)	(0.0249)	(0.0453)
50th percentile	0.1215	0.2381	0.0237	0.1274
	(0.0167)	(0.0279)	(0.0151)	(0.0235)
90th percentile	0.1139	0.2281	0.0265	0.1246
	(0.0188)	(0.0260)	(0.0203)	(0.0329)
B. Effect of Wage Coefficients				
10th percentile	0.2958	0.1886	0.1134	0.0319
	(0.0429)	(0.0487)	(0.0359)	(0.0511)
50th percentile	0.3876	0.2598	0.1836	0.0835
	(0.0220)	(0.0275)	(0.0231)	(0.0255)
90th percentile	0.3316	0.2336	0.2749	0.1790
	(0.0269)	(0.0285)	(0.0341)	(0.0357)
C. Sum of Covariate and Wage Coefficient Effects				
10th percentile	0.4725	0.4615	0.1855	0.1967
	(0.0367)	(0.0353)	(0.0266)	(0.0314)
50th percentile	0.5091	0.4979	0.2073	0.2109
	(0.0226)	(0.0232)	(0.0236)	(0.0211)
90th percentile	0.4455	0.4617	0.3014	0.3036
	(0.0314)	(0.0311)	(0.0346)	(0.0342)

Notes: Sample includes full time nonfarm wage and salary workers age 25-64 with at least 26 weeks of employment. Entries are based on the decomposition of the unconditional gender log wage gap at each indicated percentile, based on methods in Chernozhukov, Fernández-Val and Melly (2013). Standard error are in parentheses and are computed by bootstrapping with 100 repetitions.

Brief Summary

The gender wage gap in the US declined considerably between 1980 and 2010.

Women are much more likely to participate in the labor market over the period.

Conventional human capital variables explain little of the gender wage gap, but differences in occupation and industry remain important factors.

The decline in the gender wage gap is much smaller at the top of the wage distribution.

Racial Wage Gap (Bayer and Charles, 2018)

Bayer and Charles (2018) present new estimates of the black-white differences in earnings in the United States.

Data: US Decennial Census (1940 – 2000) + American Community Survey (2005 – 2014)

Sample: Working and non-working, non-Hispanic black and white men between 25 and 55 years old

Outcome: Annual earnings (not hourly or monthly earnings)

i.e., Holistic measure that subsumes the effect of changes in wage and probability or intensity of working.

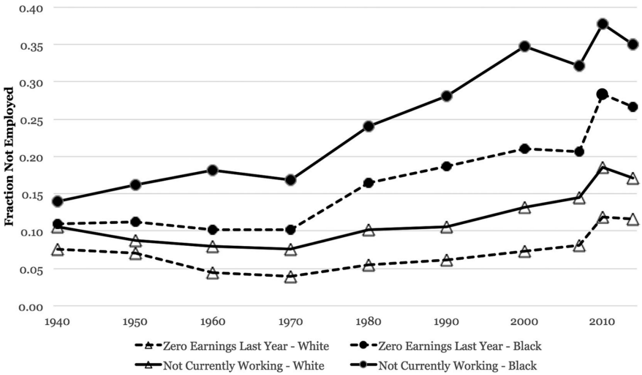


FIGURE II.A

Fraction of Men Not Employed, by Alternative Measure and Race

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.

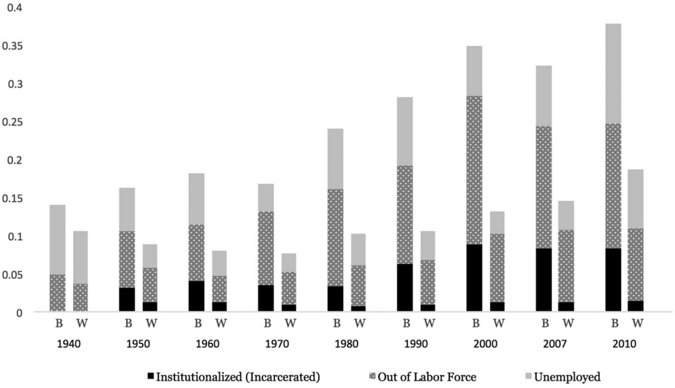


FIGURE II.B

Fraction of Men Not Currently Working, by Explanation and Race

Figure displays fraction of non-Hispanic black and white men aged 25–54 not currently working for three mutually exclusive reasons: institutionalized, not institutionalized but out of the labor force, in the labor force but unemployed. 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.

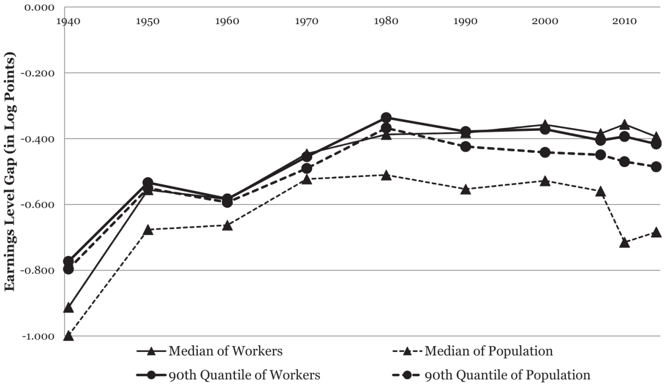


FIGURE IV

Racial Earnings Level Gap, Workers and Population, Median and 90th Quantile

Figure displays earnings level gap, measured in log points, for the median and 90th quantile for non-Hispanic black and white men aged 25–54. Gaps are reported for the sample of workers and the population of all men, including nonworkers. 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.

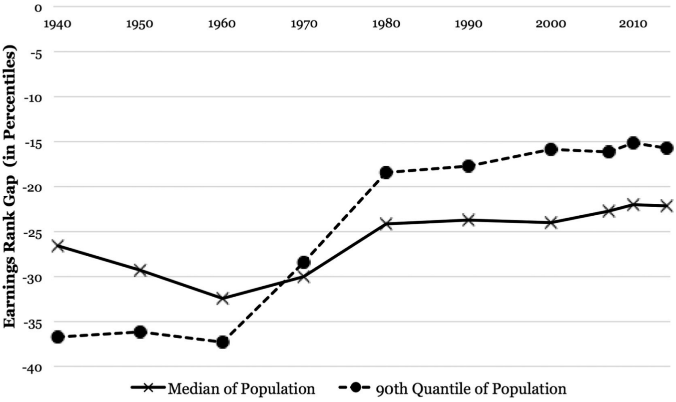


FIGURE V

Racial Earnings Rank Gaps, Median and 90th Quantiles

Figure displays earnings rank gap, measured in percentiles, for the median and 90th quantile in the population of all non-Hispanic black and white men aged 25–54, including nonworkers. 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.

Explaining Changes in Relative Earnings

Distributional Convergence: Forces that stretch out or compress both black and white earnings distributions (“race-neutral”)

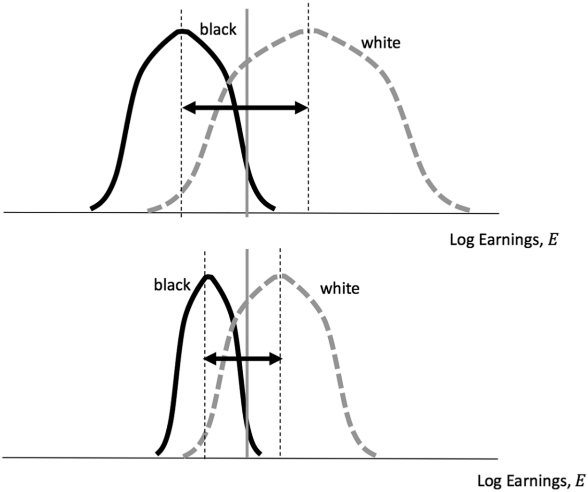
Their effect can vary by race due to the positions Black and white Americans occupy within the distribution.

e.g., “Great compression” in 1940s, or secular increase in inequality since 1980s

Positional Convergence: Forces that create a relative change in actual skill of black man compared with white man (“race-specific”)

e.g., increase in quality of schools attended by black children, decrease in wage discrimination, or occupational exclusion against black people.

(A) Distributional Convergence



(B) Positional Convergence

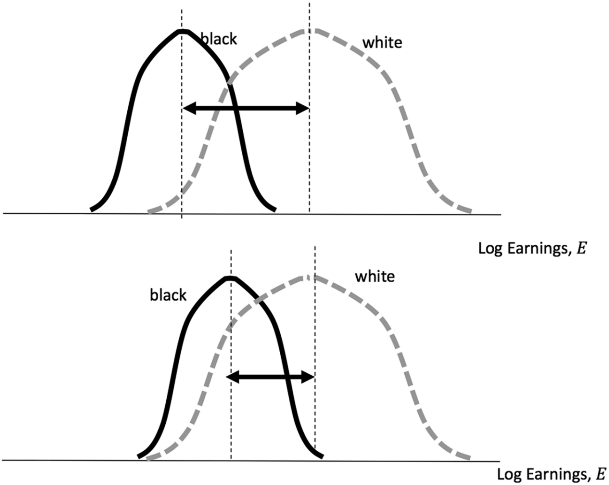


TABLE III

DECOMPOSITION OF CHANGES IN RACIAL EARNINGS AND EMPLOYMENT GAPS: POSITIONAL VERSUS DISTRIBUTIONAL CONVERGENCE

	1940–1970	1970–2014	1940–2014
Panel A: Median earnings level gap			
Total Change	0.476	−0.193	0.283
Distributional convergence	0.643	−0.392	0.251
Positional convergence	−0.167	0.199	0.032
Panel B: 90th quantile earnings level gap			
Total change	0.306	0.019	0.325
Distributional convergence	0.192	−0.177	0.015
Positional convergence	0.114	0.196	0.310
Panel C: Employment gap			
Total change		−0.105	
Distributional convergence		−0.095	
Positional convergence		−0.010	

Notes. The three panels of this table describe a series of decompositions of the change in the earnings gaps at the 50th and 90th quantiles and the racial employment gap, for the time horizon shown in the column heading. All estimates use the sample of all men including those with zero earnings, conditioning on age. The total change in the earnings gap at each quantile and the racial employment gap is decomposed into two components: the portion due to distributional shifts in the overall structure of the earnings distribution and shifts in the relative position of black and white men within the earnings distribution.

Brief Summary

Racial earnings gaps have been flat since 1950, once non-workers are considered.

Thus, despite decades of political change, racial disparities are larger than we thought.

Distributional forces decrease the earnings gap between 1940 and 1970 at the 50th and 90th pct but increase the earnings gap from 1970 to 2014 at the 50th and 90th pct.

Positional forces decrease the earnings gap between 1940 and 2014 at the 50th and 90th pct. They explain c.a. 100% of the gains at the 90th pct, only c.a. 10% at the median.

Wealth of Two Nations (Derenoncourt et al., 2022)

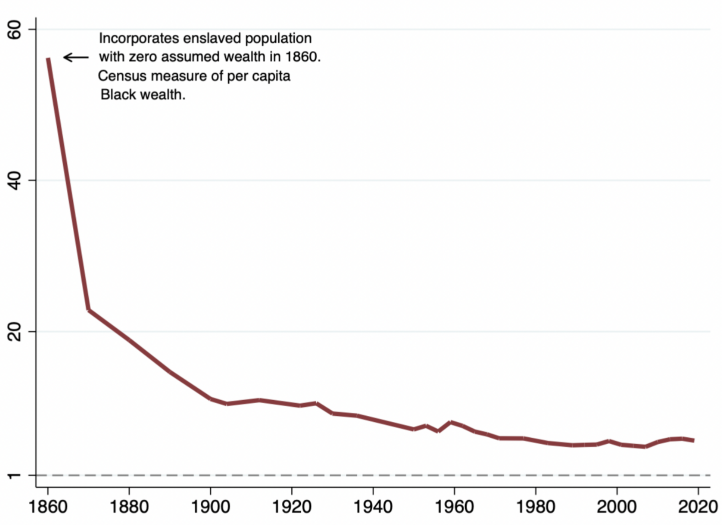
The largest racial economic gap continues to be wealth:

- White to Black wealth ratio in 2019 is 6:1
- Compared to income ratio of 1.5:1

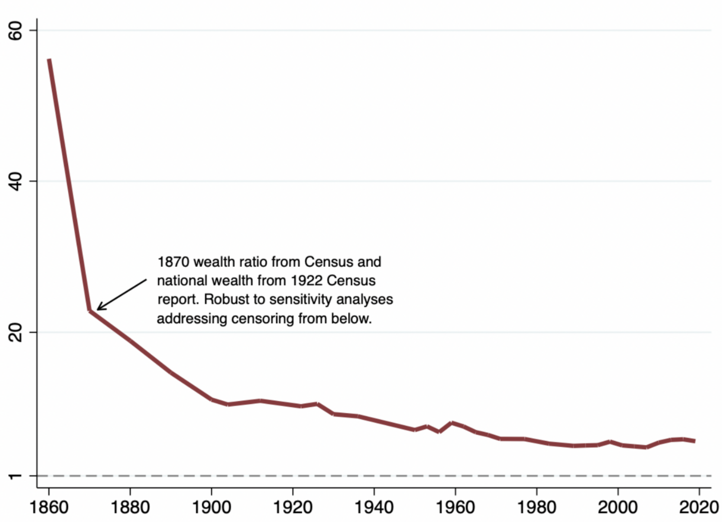
Derenoncourt et al. (2022) compile the first long-run series on the racial wealth gap from the US Civil War to the present, and fill in $\approx$ 100 missing years of data (from the 1880s to the 1980s).

Their goal is to explain the mechanisms behind times of convergence/divergence.

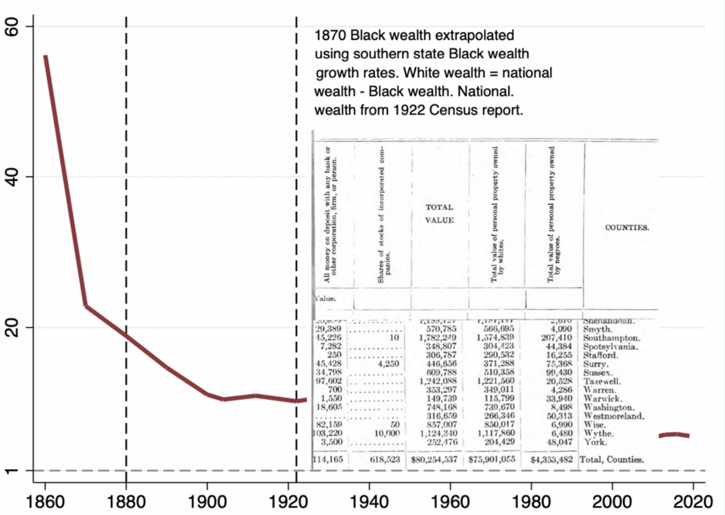
Racial Wealth Gap



Racial Wealth Gap

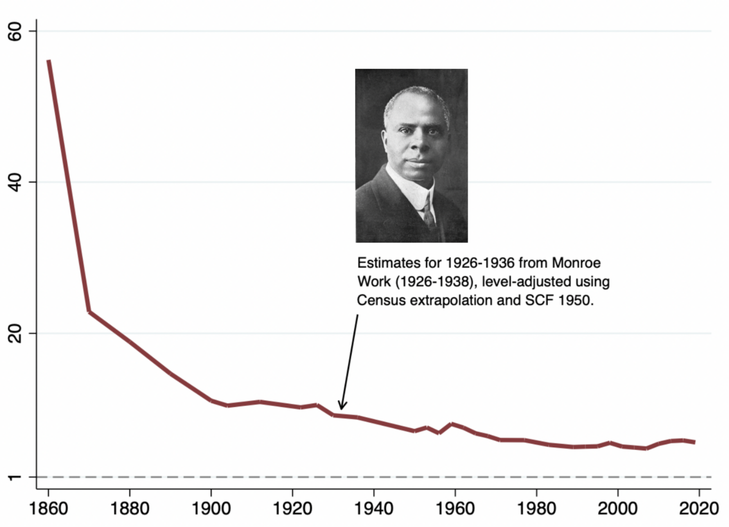


Racial Wealth Gap

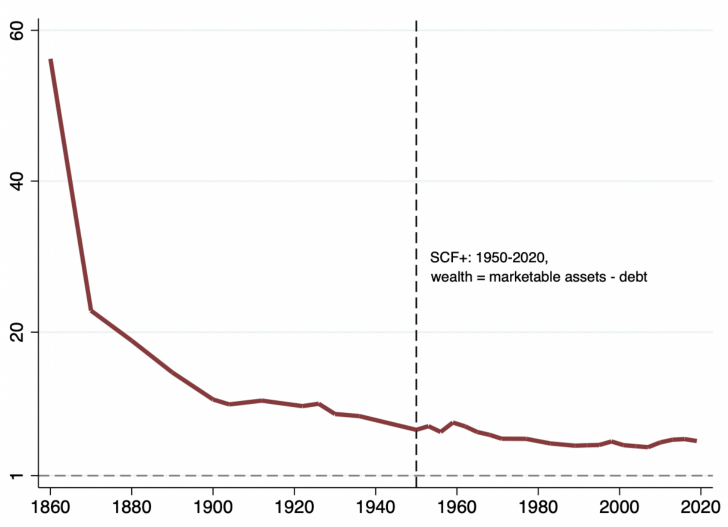


All money on deposit with any bank or other corporation, firm, or person.	Shares of stocks of incorporated companies.	TOTAL VALUE	Total value of personal property owned by whites.	Total value of personal property owned by negroes.	COUNTIES.
Value.					
[illegible]	....	[illegible]	[illegible]	[illegible]	Shenandoah.
29,389	....	570,785	566,695	4,090	Smyth.
45,226	10	1,782,249	1,574,839	207,410	Southampton.
7,282	....	348,807	304,423	44,384	Spotsylvania.
250	....	306,787	290,532	16,255	Stafford.
45,428	4,250	446,656	371,288	75,368	Surry.
34,798	....	609,788	510,358	99,430	Sussex.
97,602	....	1,242,088	1,221,560	20,528	Tazewell.
700	....	353,297	349,011	4,286	Warren.
1,550	....	149,739	115,799	33,940	Warwick.
18,605	....	748,168	739,670	8,498	Washington.
....	....	316,659	266,346	50,313	Westmoreland.
82,159	50	857,007	850,017	6,990	Wise.
103,220	10,000	1,124,340	1,117,860	6,480	Wythe.
3,500	....	252,476	204,429	48,047	York.
114,165	618,523	$80,254,537	$75,901,055	$4,353,482	Total, Counties.

Racial Wealth Gap

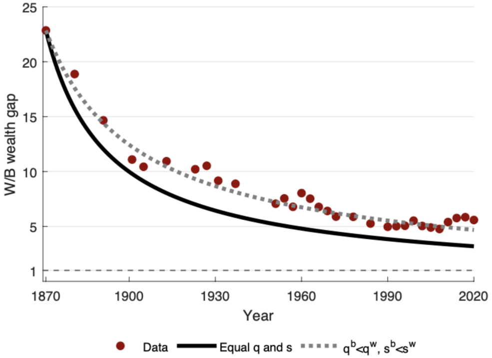


Racial Wealth Gap



Racial Wealth Gap

How would the racial wealth gap have evolved, if Black and white Americans had equal wealth accumulating conditions?



Brief Summary

Rapid convergence after Emancipation.

- In 1860, White to Black wealth ratio is 56 to 1
- By 1920, White to Black wealth ratio is 10 to 1

Convergence slows dramatically by mid 20th century.

- White to Black wealth ratio in 1950s is 7 to 1
- White to Black wealth ratio in 2019 is 6 to 1

Overall, the series exhibits a “hockey-stick” shape.

The wealth gap today is still largely the result of very unequal starting conditions in 1870.

What's next?

The next lecture will study intergenerational mobility in the United States based on income, gender, and race.

We will use the same tools we discussed today. Make sure to master quantile regressions and Oaxaca-Blinder decompositions.

Bayer, P. and Charles, K. K. (2018). Divergent paths: A new perspective on earnings differences between black and white men since 1940. *The Quarterly Journal of Economics*, 133(3):1459–1501.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, 55(3):789–865.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, pages 436–455.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Derenoncourt, E., Kim, C. H., Kuhn, M., and Schularick, M. (2022). Wealth of two nations: The us racial wealth gap, 1860-2020. Technical report, National Bureau of Economic Research.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, pages 693–709.

Tips for the Exam

Understand linear and quantile regressions, as well as their Oaxaca-Blinder-style decompositions.

Understand the interpretation of estimated coefficients (and the underlying assumptions required for such an interpretation).

Have a good knowledge of historical trends in racial and gender earnings and wealth inequality in the United States.