Bridging the Gap: Information, Returns and Choices

Nadav Kunievsky

April 27, 2024

Introduction

- In social systems where individuals' life trajectories are shaped by choices, it is crucial to understand what drives differences in these choices
- Differences in choices can be attributed to:
  - Differences in returns
  - Disparities in information about these returns
- This study employs a structural model to measure how much of the observed choice gap can be attributed to differences in information versus differences in returns

Introduction

- In social systems where individuals' life trajectories are shaped by choices, it is crucial to understand what drives differences in these choices
- Differences in choices can be attributed to:
  - Differences in returns
  - Disparities in information about these returns
- This study employs a structural model to measure how much of the observed choice gap can be attributed to differences in information versus differences in returns

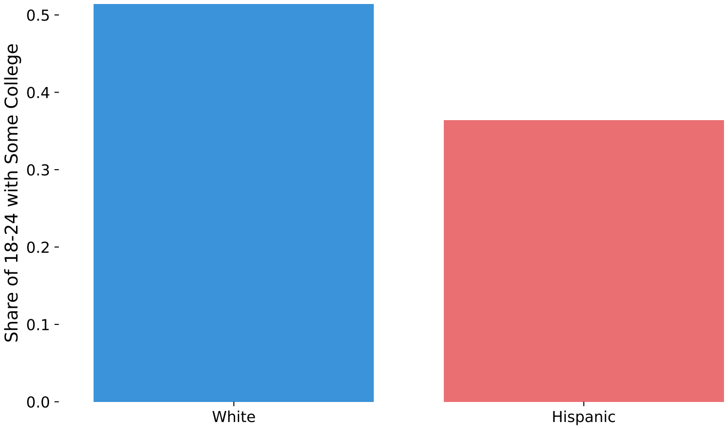
Introduction

- In social systems where individuals' life trajectories are shaped by choices, it is crucial to understand what drives differences in these choices
- Differences in choices can be attributed to:
  - Differences in returns
  - Disparities in information about these returns
- This study employs a structural model to measure how much of the observed choice gap can be attributed to differences in information versus differences in returns

Introduction

- In this paper I propose a Kitagawa-Oaxaca-Blinder decomposition approach to split the gap into two components:
  - **Information Channel:** How much would the gap change if group b had the same information quality as members of group a
  - **Returns Channel:** How much would the gap in choice change if members of a faced the distribution of b's payoffs

The College Enrollment Gap



Source: CPS, 2000-2019 College Attendance over the years and gender

What Drives Differences in Returns

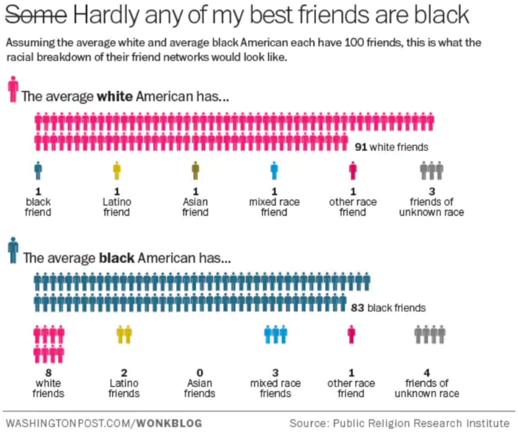
- Factors that influence differences in returns:
  - Self-selection (e.g Roy (1951), Heckman (1979))
  - Investment in human capital (e.g Ben Porath, (1967), Heckman and Cunha, (2005,2007)
  - Unwarranted disparities in the labor market (e.g Bohren et al., (2024), Coate and Loury (1993), Arnold et al. (2022), Aigner and Cain (1997), Becker (1957)

What Drives Differences in Information?

- High school graduates information might include:
  - Social environment
  - Media exposure
  - Academic assessments and school examinations
  - Guidance from educators and counselors
  - Parental insights

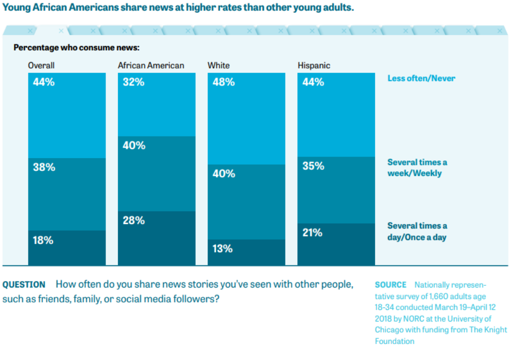
What Drives Differences in Information?

- High school graduates information might include:
  - Social environment
  - Media exposure
  - Academic assessments and school examinations
  - Guidance from educators and counselors
  - Parental insights



What Drives Differences in Information?

- High school graduates information might include:
  - Social environment
  - Media exposure
  - Academic assessments and school examinations
  - Guidance from educators and counselors
  - Parental insights



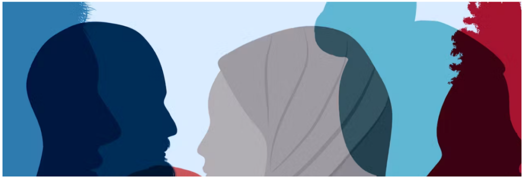
What Drives Differences in Information?

- High school graduates information might include:
  - Social environment
  - Media exposure
  - Academic assessments and school examinations
  - Guidance from educators and counselors
  - Parental insights

DIVERSITY AND EQUITY

School Counselors Have Implicit Bias. Some Are Ready to Address It.

By Emily Tate Sullivan Apr 6, 2021



What do I do in this paper?

- I introduce a method that builds on Marginal Treatment Effect literature to recover the beliefs distribution
- I use the estimates of the beliefs distribution to show how to decompose the choice gap in college attendance between Hispanics and Whites into the information channel and returns channel

What we do in this paper

- I use Texas administrative data to operationalize the decomposition method and measure what drives the gap in the decision to go to college between Whites and Hispanics
- I find that differences in returns, rather than differences in information, drive most of the choice gap.
- I then use the estimates to examine how much information is useful in order to close the choice gap
- I find that to achieve parity in choice requires providing high-quality new information; such high quality is unlikely with the currently available data

Related Literature

- Labour market outcomes and school choices
  Willis and Rosen (1979) Walters (2018)
- Marginal Treatment Effect
  Heckman and Vytlacil (2002,2005), Carneiro, Heckman and Vytlacil (2011)
- Information Counterfactuals
  Bergemann and Morris (2019), Brooks, Bergemann and Morris (2022)

Overview

Decomposition Framework

- Decompose the difference using à la Kitagawa, Blinder, and Oaxaca decomposition

$$P(\text{College}|\text{Group } b) - P(\text{College}|\text{Group } a)$$

Decomposition Framework

- Decompose the difference using à la Kitagawa, Blinder, and Oaxaca decomposition

$$P(\text{College}|\text{Group } b) -$$

$$- P(\text{College}|\text{Group } a)$$

Decomposition Framework

- Decompose the difference using à la Kitagawa, Blinder, and Oaxaca decomposition

$$
\begin{aligned}
& P(\text{College}|\text{Group } b) - P(\text{College}|\text{Group } b \text{ with information of Group } a) \\
+ & P(\text{College}|\text{Group } b \text{ with information of Group } a) - P(\text{College}|\text{Group } a)
\end{aligned}
$$

Decomposition Framework

- Decompose the difference using à la Kitagawa, Blinder, and Oaxaca decomposition

$$\underbrace{P(\text{College}|\text{Group } b) - P(\text{College}|\text{Group } b \text{ with information quality of Group } a)}_{\text{Information Channel}}$$
$$+ \underbrace{P(\text{College}|\text{Group } b \text{ with information quality of Group } a) - P(\text{College}|\text{Group } a)}_{\text{Returns Channel}}$$

- The *Information Channel* captures the change in college enrollment if we gave group b members access to the same quality of information as group a
- The *Returns Channel* captures how college enrollment would change if members of group a faced the same returns distribution as group b

A Gaussian Model of Choice

- Individuals need to decide whether to **attend college** $(D = 1)$ or **not** $(D = 0)$
- Their objective is to maximize earnings Y_d
- Before making a decision they observe vector of signals, $\boldsymbol{S}$
- Earnings and signals are distributed jointly Gaussian

$$\begin{pmatrix} Y_1 \\ Y_0 \\ \boldsymbol{S} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_{Y_1} \\ \mu_{Y_0} \\ \mu_{\boldsymbol{S}} \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_0\rho & \Sigma_{1\boldsymbol{S}} \\ \sigma_1\sigma_0\rho & \sigma_0^2 & \Sigma_{0\boldsymbol{S}} \\ \Sigma_{1\boldsymbol{S}} & \Sigma_{0\boldsymbol{S}} & \Sigma_{\boldsymbol{S}} \end{pmatrix} \right)$$

A Gaussian Model of Choice - Model Uncertainty

- Individuals face model uncertainty
- We assume they have knowledge of the marginal distributions $P(Y_1, \boldsymbol{S})$ and $P(Y_0, \boldsymbol{S})$
- The only information individuals have on ρ is formed by the restriction implied by the model
- We let

$$\Omega = \{\rho \in \mathbb{R} : \text{Covariance matrix is positive-semi-definite}\}$$

  This modeling mimics the fact that individuals, just like econometricians, cannot know the correlation of potential outcomes

- These individuals hold a prior, $h(\rho)$, over Ω

A Gaussian Model of Choice - Beliefs Distribution and Choice

- Let $\mathcal{R} = Y_1 - Y_0$, and $\mu_{\mathcal{R}}$ be the mean returns.
- let $\Sigma_{\boldsymbol{S},\mathcal{R}}$ be the covariance between signals $\boldsymbol{S}$ and the returns
- Beliefs on returns are then given by

$$E[\mathcal{R}|\boldsymbol{S}] = \mu_{\mathcal{R}} + \Sigma_{\boldsymbol{S},\mathcal{R}}^{T} \Sigma_{\boldsymbol{S}}^{-1} (\boldsymbol{S} - \mu_{\boldsymbol{S}}).$$

- Given individual threshold, c, make the decision

$$D = \mathbb{1}\left[E[\mathcal{R}|\boldsymbol{S}] \geq c\right] = \mathbb{1}\left[\mu_{\mathcal{R}} + \Sigma_{\boldsymbol{S},\mathcal{R}}^{T} \Sigma_{\boldsymbol{S}}^{-1} (\boldsymbol{S} - \mu_{\boldsymbol{S}}) \geq c\right].$$

A Gaussian Model of Choice - Beliefs Distribution and Choice

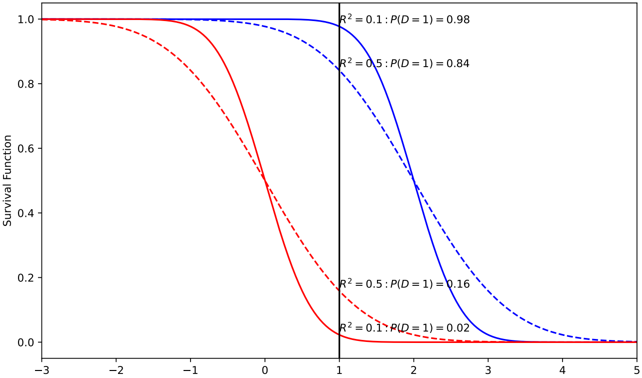
- A key concept is the beliefs distribution, which is given by

$$\mathbb{E}[\mathcal{R}|\boldsymbol{S}] \sim \mathcal{N}\left(\mu_{\mathcal{R}}, \Sigma_{\boldsymbol{S},\mathcal{R}}^{T}\Sigma_{\boldsymbol{S}}^{-1}\Sigma_{\boldsymbol{S},\mathcal{R}}\right),$$

- Given the beliefs distribution and the decision rule, the share of individuals who opt to go to college is given by:

$$P(D=1) = \Phi\left(\frac{\mu_{\mathcal{R}} - c}{\sqrt{\Sigma_{\boldsymbol{S},\mathcal{R}}^{T}\Sigma_{\boldsymbol{S}}^{-1}\Sigma_{\boldsymbol{S},\mathcal{R}}}}\right)$$

Intuition From the Scalar Case



- The impact of information is contingent on the interplay between costs and beliefs

Information Quality

- We define *information Quality* as:

$$R^2 = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}])}{\mathrm{Var}_{total}(\mathcal{R})} = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}])}{\mathrm{E}_h[\mathrm{Var}(\mathcal{R}|\rho)]} = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}])}{\sigma_1^2 + \sigma_0^2 - 2\sigma_1\sigma_0 E_h[\rho]} \tag{1}$$

- For the presentation, we assume a degenerate prior on independent potential earnings $h(\rho = 0) = 1$

$$R^2 = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}])}{\sigma_1^2 + \sigma_0^2}$$

Equating Information Quality

- Assume we have the information quality of each group

$$R_a^2 = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}_a])}{\sigma_{1,a}^2 + \sigma_{0,a}^2} \quad R_b^2 = \frac{\mathrm{Var}(E[\mathcal{R}|\boldsymbol{S}_b])}{\sigma_{1,b}^2 + \sigma_{0,b}^2}$$

- In the model, beliefs dispersion are governing choice - therefore equating information quality would change dispersion
- We define the counterfactual belief dispersion as

$$\mathrm{Var}_{b,a} = R_a^2 \times ({\sigma_{1,b}}^2 + {\sigma_{0,b}}^2).$$

Operationalize the Decomposition

$$P(D = 1|\text{Group b}) - P(D = 1|\text{Group a}) = \underbrace{\Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,b}}}\right) - \Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,a}}}\right)}_{\text{Information Channel}}$$

$$+ \underbrace{\Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,a}}}\right) - \Phi\left(\frac{\mu_{\mathcal{R},a} - c_a}{\sqrt{\mathrm{Var}_{a,a}}}\right)}_{\text{Returns Channel}}$$

Overview

Identification

- We observe the joint distribution $\Pr(\overset{\text{earnings}}{Y}, \overset{\text{decision}}{D}, \overset{\text{instrument}}{Z}, \overset{\text{controls}}{X})$
- To preform the decomposition we need to identify
  - Beliefs Distribution - F
  - Costs and means - $c, \mu_{\mathcal{R}}$
- There are three steps to identification:
  - Step 1: Get the Marginal Treatment Effect curve
  - Step 2: Get the distribution of beliefs
  - step 3: Find the costs and beliefs

Identification - Instrument Assumptions

- Utilizing distance to the nearest college, Z, as a cost shifter
  e.g Card (1995), Carneiro et al. (2011), Nybom (2017), Kapor (2020), Walters (2018), Mountjoy (2022)

$$D = \mathbb{1}[\mathrm{E}[\mathcal{R}|s, x] \geq c_g(z, x)]$$

  In our model, standard instrument assumptions imply:

  1. $Y_1, Y_0 \perp\!\!\!\perp Z|X$
  2. $\boldsymbol{S} \perp\!\!\!\perp Z|X$
  3. Z is continuously distributed on $\mathcal{Z} \subseteq \mathbb{R}$
  4. $\mathrm{E}[\mathcal{R}|\boldsymbol{S}, X]$ is continuously distributed
  5. $c(z, x)$ is differentiable and covers the support of $\mathrm{E}[\mathcal{R}|\boldsymbol{S}, X]$

Identification - Cost and Beliefs

- Step 1: We obtain the Marginal Treatment Effect Curve (Identification Intuition)

$$E[Y_1 - Y_0 | U = u] = E[\mathcal{R} | U = u]$$

where U is the quantile of the choice variable.

- Integrating over u gives us $\mu_{\mathcal{R}}$

Identification - Beliefs Distribution

- Step 2: In our model, the choice variable is $E[Y_1 - Y_0|\boldsymbol{S}]$. Therefore, using Law of Iterated Expectations, and the fact that agents have rational expectations, we have
$$\mathrm{E}[\mathcal{R}|U = u] = \mathrm{E}[\mathcal{R}|\mathrm{E}[\mathcal{R}|s] = F^{-1}(u)] = F^{-1}(u)$$
which allows us to get $u \rightarrow E[\mathcal{R}|\boldsymbol{S}]$ and therefore F.

Identification - Costs

- Step 3: To obtain the costs, we can use the propensity score, $p(Z)$, and noting that
$$p(z) = \Pr\left(E[\mathcal{R}|\boldsymbol{S}] \geq c(z)\right) = 1 - F\left(c(z)\right)$$
Inverting $1 - F(.)$ gives us the cost

Empirical Specification

- We assume a linear cost function

$$c(x,z) = zb_z + \boldsymbol{X} b_x. \tag{2}$$

- Linear outcome equation

$$\begin{aligned} Y_1 &= \boldsymbol{X}\beta_1 + U_1, \\ Y_0 &= \boldsymbol{X}\beta_0 + U_0. \end{aligned} \tag{3}$$

- Assume $Z, \boldsymbol{X} \perp\!\!\!\perp U_1, U_0$

$$\begin{pmatrix} U_1 \\ U_0 \\ E[\mathcal{R}|s,x] \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ \boldsymbol{X}(\beta_1 - \beta_0) \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_{1,0}\sigma_1\sigma_0 & \sigma_{1,E} \\ \rho_{1,0}\sigma_1\sigma_0 & \sigma_0^2 & \sigma_{0,E} \\ \sigma_{1,E} & \sigma_{0,E} & \sigma_E^2 \end{bmatrix} \right) \tag{4}$$

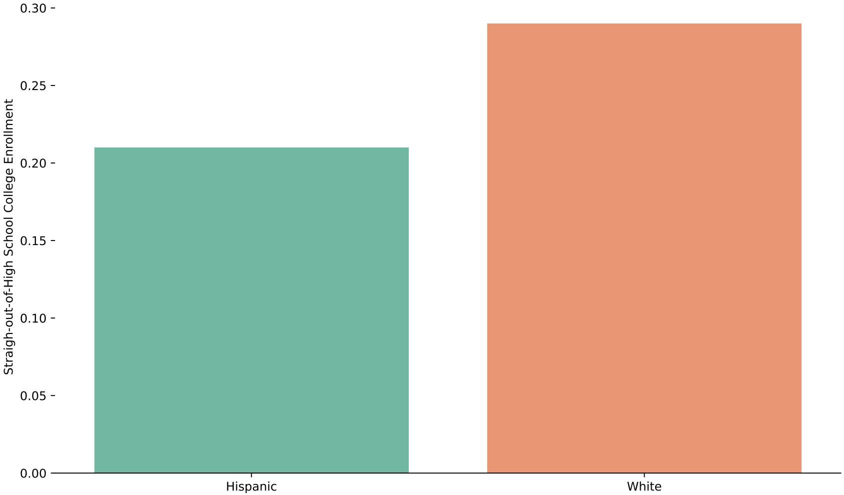
Overview

Data

- Texas administrative data for the population high school graduates from 2003-2020. Focusing the analysis on the 2003-2005 cohorts Descriptives
- The data includes information on
  - (THECB) College enrollment
  - (TEA) High-school Courses, Academic Readiness Test Score, Graduation
  - (TWC) Employed workers wages

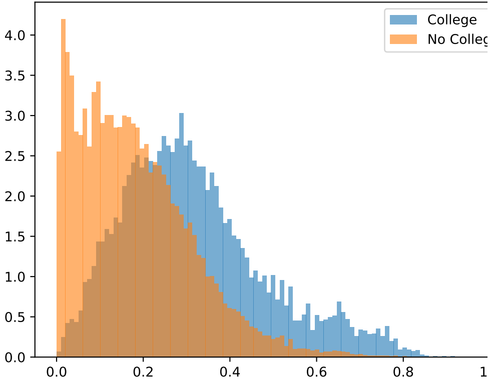
Sample Restrictions Control Variables

College Enrollment Gap

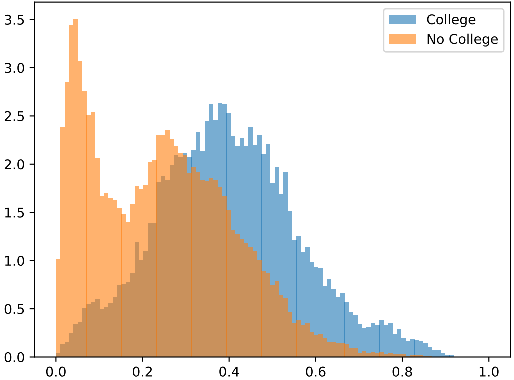


With Controls

College Enrollment Propensity

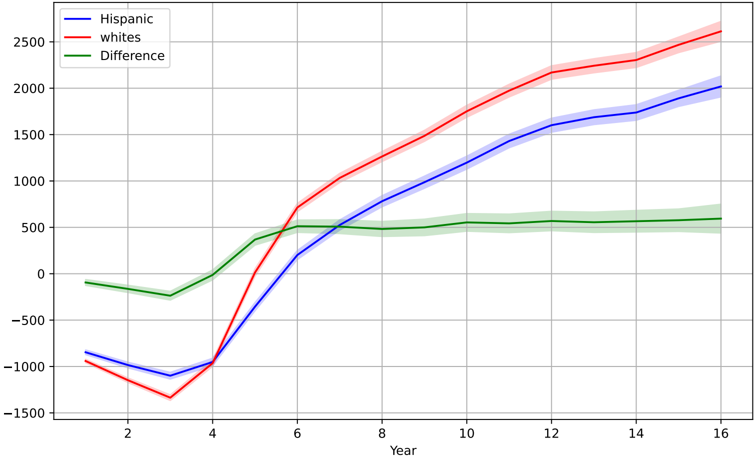


Hispanics



Whites

Returns



with controls

2SLS Specification

- These differences are driven by selection
- We estimate the following 2sls model

$$D = \overbrace{Z}^{\text{Distance to College}} \delta_1 + X\delta_2 + u$$
$$Y_t = \beta_0 + \beta_1 D + X\beta_2 + \epsilon$$

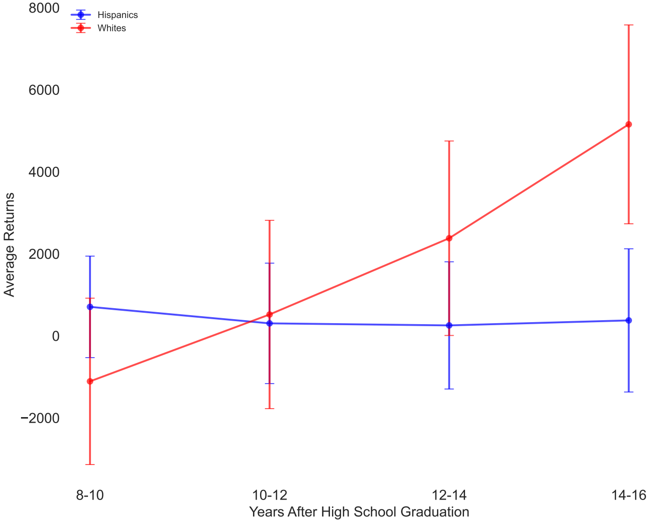
Validity

First Stage - College Enrollment on Distance to College

	All	Hispanic	Whites
No Controls	-0.0008	-0.0007	-0.0013
	(0.0001)	(0.0002)	(0.0001)
	317278	136581	180697
Ind. Controls	-0.0008	-0.0006	-0.0011
	(0.0001)	(0.0001)	(0.0001)
	317278	136581	180697
+ School Char.	-0.0014	-0.001	-0.0019
	(0.0002)	(0.0002)	(0.0002)
	317278	136581	180697
+ Neighborhood Char.	$\boxed{-0.0016}$	$\boxed{-0.0023}$	$\boxed{-0.0012}$
	(0.0002)	(0.0003)	(0.0002)
	317278	136581	180697

Validity

2SLS Specification



Validity

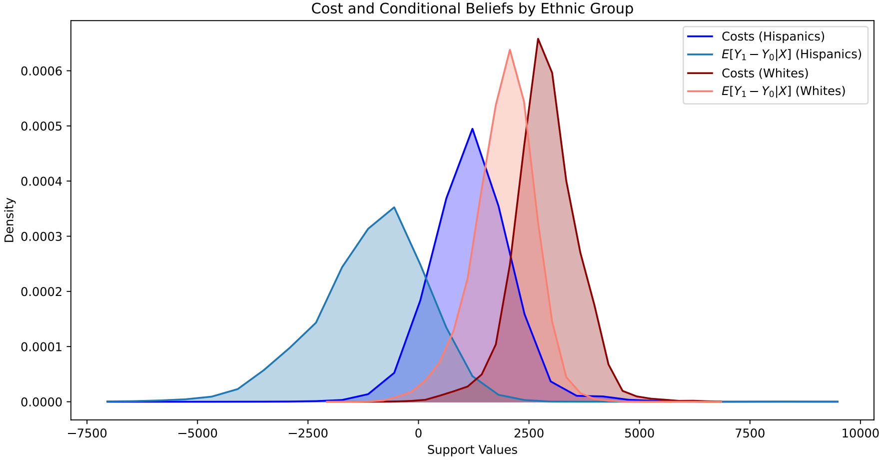
Observable Differences in Information

		Out of Sample R^2
No College	All	0.10
	Hispanic	0.09
	Whites	0.09
College	All	0.10
	Hispanic	0.09
	Whites	0.08

Interaction with School Counselors | Interaction with Parents | Industry Entropy | Wage Dispersion

Overview

Model Results - Mean Returns and Costs



Mean Returns and Costs

Model Parameters

	Share	σ_{E}	σ_1	σ_0	Avg. Cost	$E[\alpha_1 - \alpha_0]$	$\sqrt{\sigma_1^2 + \sigma_0^2}$	R^2
Hispanics	0.21	2381.0	4490.0	6264.0	1200.0	-1057.44	7707.0	0.1
		(657.0)	(125.0)	(818.0)	[890.0]	(3083.0)	(730.0)	(0.01)
Whites	0.29	1414.0	5577.0	6316.0	2880.0	1930.0	8426.0	0.03
		(873.0)	(155.0)	(491.0)	[718.0]	(3083.0)	(399.0)	(0.01)

Decomposition Results

$$\underbrace{P(D=1|\text{Group b}) - P(D=1|\text{Group a})}_{\text{Gap} = 0.8} =$$

$$\underbrace{\Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,b}}}\right) - \Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,a}}}\right)}_{\text{Information Channel} = -0.08\ (0.97\%)} +$$

$$\underbrace{\Phi\left(\frac{\mu_{\mathcal{R},b} - c_b}{\sqrt{\mathrm{Var}_{b,a}}}\right) - \Phi\left(\frac{\mu_{\mathcal{R},a} - c_a}{\sqrt{\mathrm{Var}_{a,a}}}\right)}_{\text{Returns Channel} = 0.16\ (1.97\%)}$$

Uniform Prior | Similar Copula and Degenerate Priors

Overview

How Effective are Information Interventions?

- We consider a policymaker who wants to add to provide better information to Hispanics in order to shrink further the choice gap

- The policymaker provides a new signal s_n on Y_1 and Y_0

- These signals have quality of information $R^2_{1,n}$ and $R^2_{0,n}$

- We assume that this is **new** information

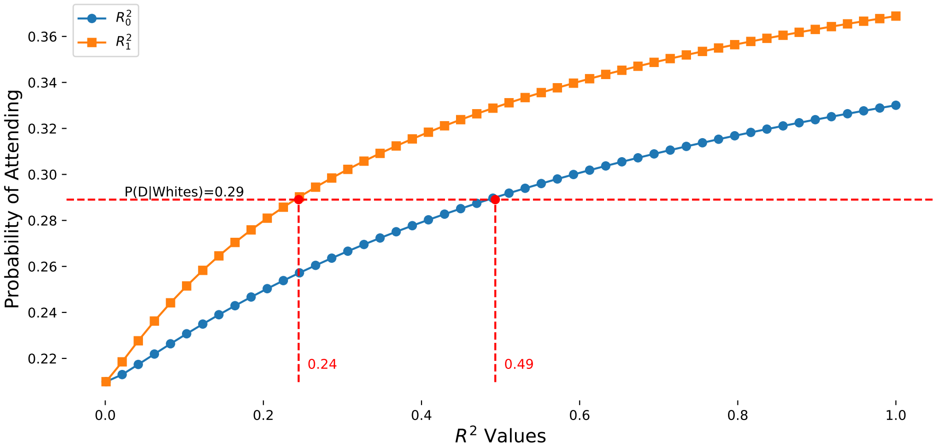
$$\mathrm{s}_n \text{ is independent of } \boldsymbol{S}$$

How Effective are Information Interventions? (Cont.)

- We ask at what precision level this new information need to be in order to induce equality in choice
- Can show that

$$\mathrm{Var}(\mathrm{E}[\mathcal{R}|\boldsymbol{S}, s_n]) = \mathrm{Var}\left(\mathrm{E}[\mathcal{R}|\boldsymbol{S}]\right) + \sigma_1^2 R_{1,n}^2 + \sigma_1^2 R_{0,n}^2 - 2\sqrt{R_{1,n}^2 R_{0,n}^2}\sigma_1\sigma_0$$

How Effective are Information Interventions?



Full Combinations

The Effectiveness of Information Interventions: Is it Reasonable Approach to Close the Gap?

- We showed that using Texas Adminstrative data we can acheive an accuracy of approximately 10% in predicting later life income (Similar results obtained using NLSY97 data)
- Other research has also found limited explanatory power for later life income Salganik et al., 2020; Garip, 2020; Murnane et al., 2000; Watts, 2020; Borghans et al., 2016
- These findings suggest that achieving equality through informational interventions alone may be challenging

Recap

- I introduced a new method to analyze how variations in the information environment impact differences in group choices
- The decomposition showed that differences in information help mitigate the disparity in college education
- It's essential for policymakers to identify the key drivers of inequality
- My counterfactual analysis shows that additional information aids in reducing the gap, but requires better measurement

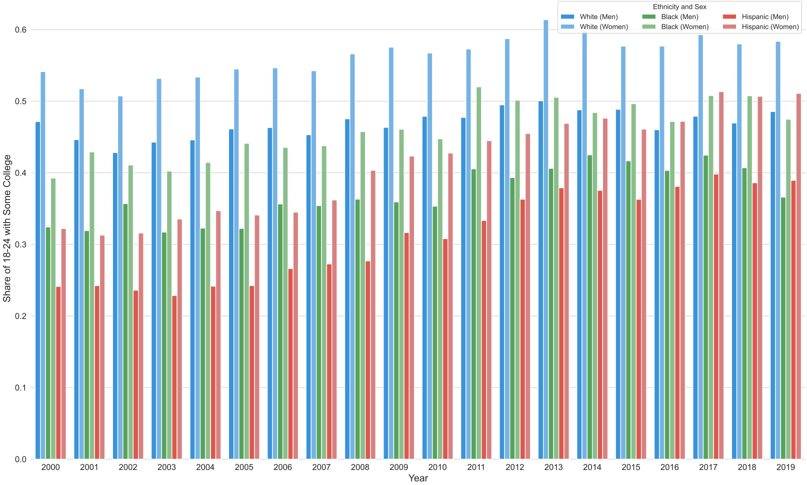
Broader Takeaways

- The decomposition approach can be used with richer data on beliefs and outcomes, for analyzing sources of group disparities and trends over time
  - Analyze the sources of group disparities and trends over time
  - Understand self-perpetuating inequality
  - Examine disparity and discrimination
  - Analyze how policy impacts outcomes

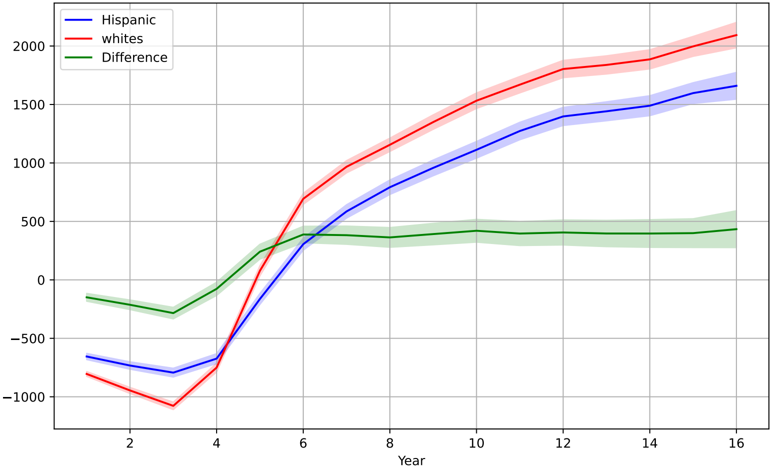
thx.

appendix

Trends in College Attendance Rates Over the Years



Wages



Hispanic
whites
Difference
2000
1500
1000
500
0
−500
−1000
2
4
6
8
10
12
14
16
Year

back

Decomposition - Bayesian Approach

- Let

$$\mathrm{Var}_{SR^2}(E[\alpha_1 - \alpha_0]|s, g, g') = SR^2_{g'} \times (\sigma^2_{1,g} + \sigma^2_{0,g} - 2\sigma_{1,g}\sigma_{0,g} E[\rho; \rho_{min,g'}, \rho_{min,g;}])$$

-

$$P(D = 1|\text{Group b}) - P(D = 1|\text{Group a}) =$$

$$\underbrace{\int_X \Phi\left(\frac{\mu_{\mathcal{R},b,x} - c_b(x)}{\sqrt{\mathrm{Var}_{SR^2}(E[U_1 - U_0|\mathbf{s}]|s, b, b)}}\right) dF_b(x) - \int_X \Phi\left(\frac{\mu_{\mathcal{R},b,x} - c_b(x)}{\sqrt{\mathrm{Var}_{SR^2}(E[U_1 - U_0|\mathbf{s}]|s, b, a)}}\right) dF_b(x)}_{\text{Information Channel}}$$

$$+ \underbrace{\int_X \Phi\left(\frac{\mu_{\mathcal{R},b,x} - c_b(x)}{\sqrt{\mathrm{Var}_{SR^2}(E[U_1 - U_0|\mathbf{s}]|s, b, a)}}\right) dF_b(x) - \int_X \Phi\left(\frac{\mu_{\mathcal{R},a,x} - c_a(x)}{\sqrt{\mathrm{Var}_{SR^2}(E[U_1 - U_0|\mathbf{s}]|s, a, a)}}\right) dF_a(x)}_{\text{Composition Channel}}$$

Decomposition - Bayesian Approach Results

	Information Channel	Returns Channel
Subjective Information Quality		
1) Feasible ρ		
$(\rho_{min} = -0.96, \rho_{max} = 0.9)$	-0.07 (-87.645%)	0.149 (188.0%)
2) All Possible R_1^2		
LB, CF= 0.359,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.069 (-87.549%)	0.149 (188.0%)
UB,CF= 0.352,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.063 (-79.74%)	0.143 (180.0%)
3) All Possible R_1^2, $\rho \geq 0$		
LB, CF= 0.356,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.067 (-84.569%)	0.146 (185.0%)
UB,CF= 0.352,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.063 (-79.74%)	0.143 (180.0%)
4) $R_1^2 \leq 0.3$		
LB, CF= 0.359,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.3$)	-0.069 (-87.55%)	0.149 (188.0%)
UB,CF= 0.357,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.3$)	-0.068 (-86.2%)	0.148 (186.0%)

Information Decomposition - Information Structure

- Decomposing the choice gap to explore how much is driven by differences in *information structure* across groups as follows:

$$P(D=1|\text{Group b}) - P(D=1|\text{Group a}) =$$

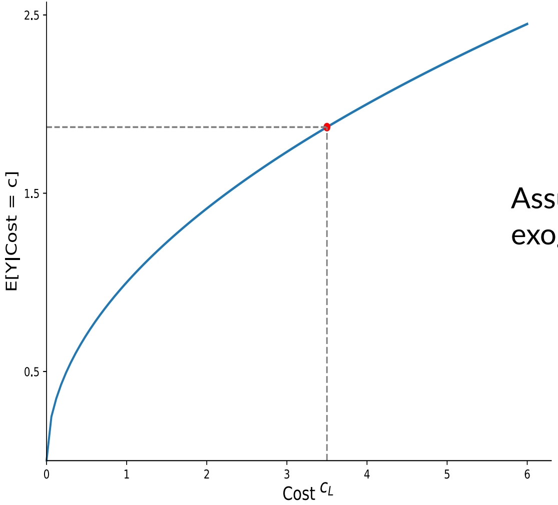
$$\underbrace{\int_{\mathcal{R}\times c} \mathcal{P}(E_{b,b}(s) \geq c|\mathcal{R}, c, b)\pi_b(\mathcal{R}, c)d\mathcal{R}dc - \int_{\mathcal{R}\times c} \mathcal{P}(E_{a,b}(s) \geq c|\mathcal{R}, c, a)\pi_b(\mathcal{R}, c)d\mathcal{R}dc}_{\text{Information Channel}}$$

$$+ \underbrace{\int_{\mathcal{R}\times c} \mathcal{P}(E_{a,b}(s) \geq c|\mathcal{R}, c, a)\pi_b(\mathcal{R}, c) - \int_{\mathcal{R}\times c} \mathcal{P}(E_{a,a}(s) \geq c|\mathcal{R}, a)\pi_a(\mathcal{R}, c)d\mathcal{R}dc}_{\text{Composition Channel}}$$

- where

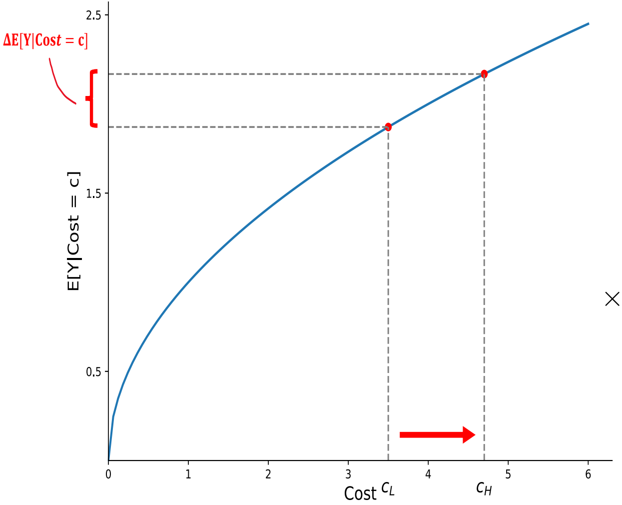
$$E_{a,b}(s) = \int_{\tilde{\mathcal{R}}} \tilde{\mathcal{R}} \frac{\overbrace{P(s|\tilde{\mathcal{R}}, a)}^{\text{Information}} \times \overbrace{\pi_b(\tilde{\mathcal{R}})}^{\text{earnings}}}{\int P(s|\tilde{\mathcal{R}}, a)\pi_b(\tilde{\mathcal{R}})d\tilde{\mathcal{R}}} d\tilde{\mathcal{R}}$$

Marginal Treatment Effect Identification - Intuition



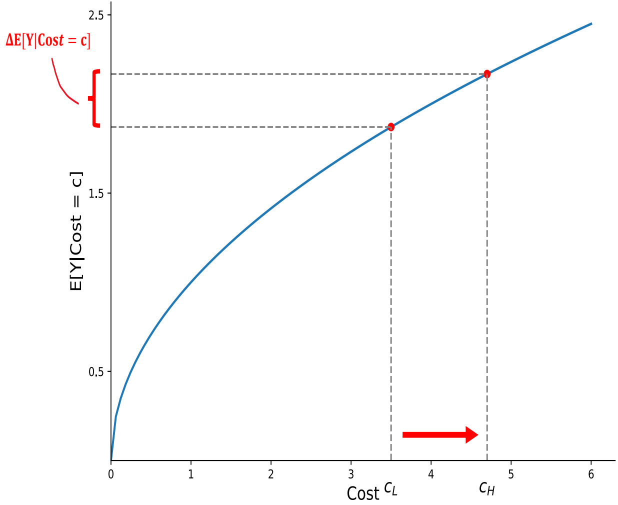
Assume we knew the cost, and we have exogenous cost shifter $c \perp\!\!\!\perp E[R|\boldsymbol{s}]$

Marginal Treatment Effect Identification - Intuition (Cont.)



$$\mathbb{E}[Y|Cost = c_L] - \mathbb{E}[Y|Cost = c_H] = \underbrace{\Big(\mathbb{E}[Y_0|c_L \leq \mathbb{E}[\mathcal{R}|s] \leq c_H] - \mathbb{E}[Y_1|c_L \leq \mathbb{E}[\mathcal{R}|s] \leq c_H]\Big)}_{\text{Change in Earnings of people who opt out}} \times \underbrace{\Big(P(D|Cost = c_L) - P(D = 1|Cost = c_H)\Big)}_{\text{Share of people opting out}}$$

Marginal Treatment Effect Identification - Intuition (Cont.)



$$\mathbb{E}[Y|Cost = c_L] - \mathbb{E}[Y|Cost = c_H] =$$
$$-\mathbb{E}[\mathcal{R}|c_L \leq \mathbb{E}[\mathcal{R}|s] \leq c_H)$$
$$\times \Delta P(D = 1|c)$$

Instrument Validity

	Test Factor	Math	Reading
Hispanic Women	-0.00328	-0.00223	-0.00129
	(0.00265)	(0.00278)	(0.00182)
Hispanic Men	0.00102	0.00432	-0.00291
	(0.00269)	(0.00274)	(0.00193)
White Women	-0.00185	-0.00149	-0.00053
	(0.00234)	(0.00241)	(0.00161)
White Men	0.00033	-0.00139	0.00178
	(0.00234)	(0.00231)	(0.00169)
School Chars	V	V	V
Neighborhood Chars	V	V	V
Individual Chars	V	V	V

Estimation

- Estimate the model parameters using two-step estimation method (Heckman (1979)
- Using parameters estimate $\hat{R}^2$ and the counterfactual component of the decomposition as

$$\underbrace{\hat{P}(D=1|b)}_{\text{Observed}} - \underbrace{\frac{1}{N}\sum_i \Phi\left(\frac{x(\hat{\beta}_1-\hat{\beta}_0)-\hat{c}(x_i,z_i)}{\sqrt{\hat{\sigma}^2_{\mathcal{R}}\hat{R}^2_a}}\right)}_{\text{Counterfactual}},$$

Sample Restrictions

- Sample limited to 2003-2005 cohorts with observed earnings at 12-15 years after high school graduation, excluding:
  - Special education students.
  - Students outside the 17-18 age range in 12th grade.
  - Students not meeting minimum graduation requirements.
  - High-achieving students (above 80th percentile in test scores) due to their higher likelihood of out-of-state college enrollment and missing earnings data.

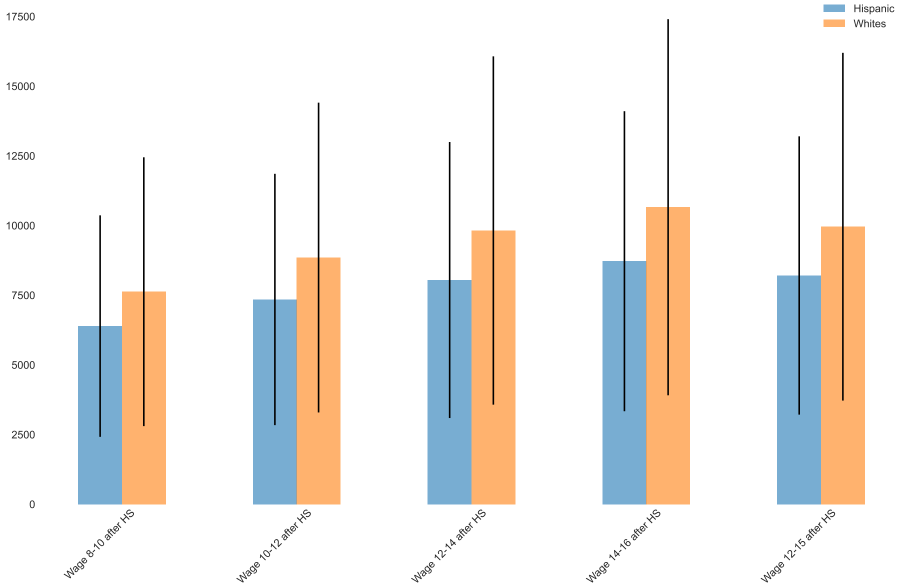
List of Controls

- Control variables:
  - Student-level demographics: gender, economic disadvantage indicator, graduation program type.
  - High school-level controls: geographic locale, proximity to colleges, vocational education indicator, local industry influence (oil and gas employment).
  - Neighborhood characteristics: commuting zones, neighborhood quality index based on median household income, poverty rates, and eligibility for free/reduced-price lunch.

Instrument Validity

	All	Hispanic	Whites
No Controls	-0.0156	-0.0232	-0.0436
	(0.0054)	(0.0053)	(0.0038)
Ind. Controls	-0.0277	-0.0151	-0.0392
	(0.0044)	(0.0066)	(0.0045)
+ School Char.	-0.0061	0.0074	-0.0177
	(0.004)	(0.0057)	(0.004)
+ Neighborhood Char.	-0.0014	0.0009	-0.0036
	(0.0018)	(0.0022)	(0.0021)

Average Wage and Dispersion



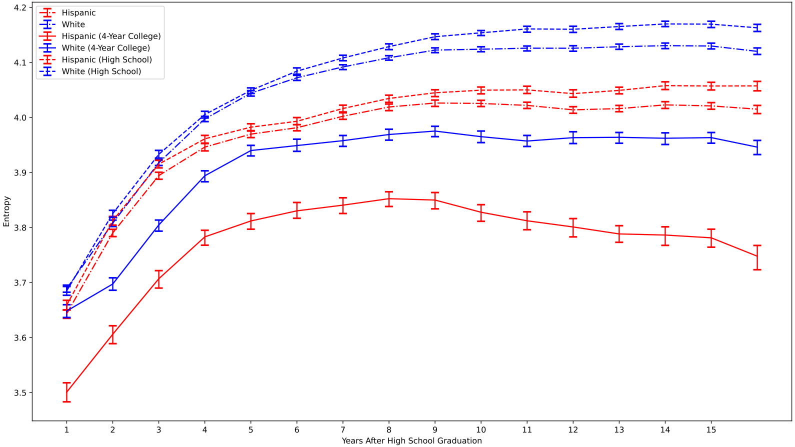
Interaction with Parents

	Education		Important Issues		Job		Relationships		Finance	
	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites
Often	0.42	43.00	0.39	47.00	0.39	37.00	0.36	28.00	0.27	43.00
	(0.50)	(0.36)	(0.49)	(0.37)	(0.49)	(0.31)	(0.48)	(0.22)	(0.45)	(0.34)
Sometimes	0.44	58.00	0.39	61.00	0.43	67.00	0.51	67.00	0.56	54.00
	(0.50)	(0.49)	(0.49)	(0.48)	(0.50)	(0.56)	(0.51)	(0.53)	(0.50)	(0.42)
Never	0.13	18.00	0.22	20.00	0.17	16.00	0.13	32.00	0.18	31.00
	(0.34)	(0.15)	(0.42)	(0.16)	(0.38)	(0.13)	(0.34)	(0.25)	(0.39)	(0.24)
Num Obs	58.00	144	54.00	155	56.00	141	55.00	123	49.00	140

Interaction with Counselors

	course selection		personal problems		school discipline		jobs		educational plans		choosing a college		college applications		letters of rec.		college essays		financial aid		job interviews	
	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites	Hispanic	Whites
Three or more times	0.14	0.10	0.02	0.02	0.01	0.01	0.01	0.01	0.08	0.07	0.10	0.07	0.10	0.09	0.06	0.05	0.03	0.03	0.07	0.04	0.01	0.00
	(0.005)	(0.004)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
Twice	0.29	0.28	0.04	0.03	0.04	0.02	0.04	0.03	0.13	0.12	0.07	0.07	0.07	0.07	0.04	0.05	0.03	0.03	0.05	0.04	0.01	0.01
	(0.006)	(0.007)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.001)	(0.001)
Once	0.28	0.32	0.09	0.08	0.08	0.05	0.13	0.12	0.23	0.26	0.10	0.11	0.11	0.12	0.07	0.09	0.06	0.06	0.11	0.10	0.02	0.01
	(0.006)	(0.007)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)
Never	0.19	0.21	0.82	0.85	0.84	0.89	0.80	0.81	0.50	0.51	0.15	0.19	0.14	0.17	0.25	0.26	0.30	0.33	0.18	0.25	0.38	0.42
	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.007)
No response	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Weighted Obs	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774	139644	212774
Unweighted Obs	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621	11114	12621

Industries Entropy



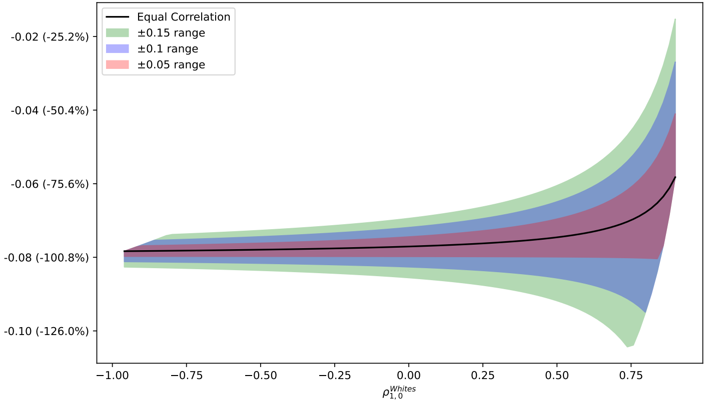
Uniform Prior over Feasible Correlation Values

	Information Channel	Composition Channel
All Possible R_1^2		
LB, CF= 0.359	-0.069 (-87.549%)	0.149 (188.0%)
UB,CF= 0.352	-0.063 (-79.74%)	0.143 (180.0%)

Uniform Prior over Feasible Correlation - Additional Restrictions

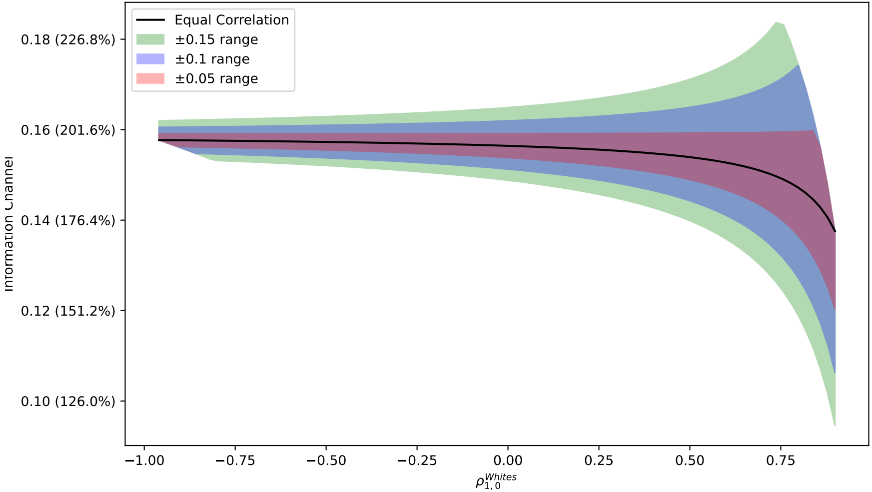
	Information Channel	Composition Channel
1) Feasible ρ		
($\rho_{min} = -0.96$, $\rho_{max} = 0.9$)	-0.07 (-87.645%)	0.149 (188.0%)
2) All Possible R_1^2		
LB, CF= 0.359,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.069 (-87.549%)	0.149 (188.0%)
UB,CF= 0.352,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.063 (-79.74%)	0.143 (180.0%)
3) All Possible R_1^2, $\rho \geq 0$		
LB, CF= 0.356,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.067 (-84.569%)	0.146 (185.0%)
UB,CF= 0.352,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.6$)	-0.063 (-79.74%)	0.143 (180.0%)
4) $R_1^2 \leq 0.3$		
LB, CF= 0.359,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.3$)	-0.069 (-87.55%)	0.149 (188.0%)
UB,CF= 0.357,($R^2_{1,min} = 0.1$, $R^2_{1,max} = 0.3$)	-0.068 (-86.2%)	0.148 (186.0%)

Decomposition - Similar Copulas



- Unrestricted Bounds on Information Channel $-0.17(-216\%) - 0.13(161\%)$
- Unrestricted Bounds on Returns Channel $0.25(316\%) - 0.05(61\%)$

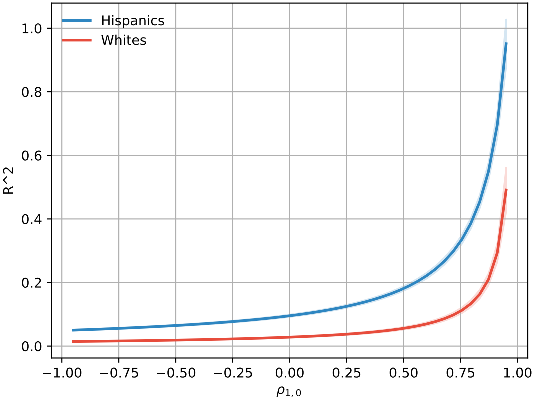
Returns Channel



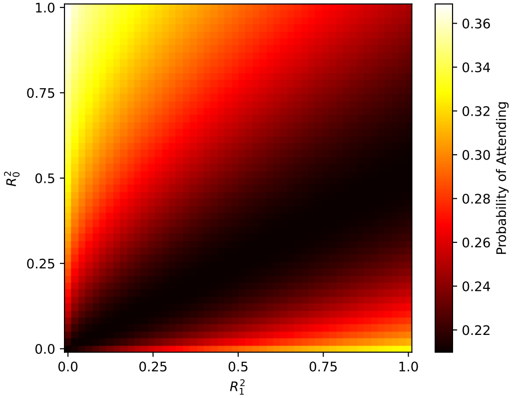
Identification of $\mathrm{Var}_{tota}(\mathcal{R})$ and Ω

- The set of feasible ρs is defined by the covariance matrix of signals and earnings, but we don't observe $\boldsymbol{S}$
- We show that it's enought to know the quality of information on Y_1 to identify the set of feasible ρs
- As we don't know that quality, we explore all values of quality of information to bound the decomposition components

Value of Information



How Effective are Information Interventions?



Back