Wage Gaps and Test Score Differences: Incentives or Pre-Market Factors?

Andrea Moro (joint with Peter Norman)

Federal Reserve Board, Washington DC, January 27, 2005

Paper and slides available from http://www.econ.umn.edu/ amoro/

The question

- Empirically disentangle two different sources of racial wage inequality

  1. Pre-market factors (Neal and Johnson, JPE 96)
  2. Incentives to acquire human capital

- Horse-race between the two hypotheses using a model of statistical discrimination that nests the two explanations.

The question

- Empirically disentangle two different sources of racial wage inequality

    1. Pre-market factors (Neal and Johnson, JPE 96)
    2. Incentives to acquire human capital

- Horse-race between the two hypotheses using a model of statistical discrimination that nests the two explanations.

- Other explanations: racism (Bowlus & Eckstein 2002), initial conditions/catching up

Facts

- Black/white wage gap has been roughly stable since the 80s (problem for the "catching up" hypothesis)

Facts

- Black/white wage gap has been roughly stable since the 80s (problem for the "catching up" hypothesis)

- Much of the difference is "explained" by differences in "human-capital-like" variables such as schooling and test-scores

Facts

- Black/white wage gap has been roughly stable since the 80s (problem for the "catching up" hypothesis)

- Much of the difference is "explained" by differences in "human-capital-like" variables such as schooling and test-scores

ln($wages$), NLSY males, 1991

Black	−0.244 (0.026)	−0.072 (0.027)
AFQT		0.172 (0.012)

Facts

- Black/white wage gap has been roughly stable since the 80s (problem for the "catching up" hypothesis)

- Much of the difference is "explained" by differences in "human-capital-like" variables such as schooling and test-scores

- "Returns to AFQT" the same for blacks and whites. Neal & Johnson argues that this is evidence against statistical discrimination.

Facts

- Black/white wage gap has been roughly stable since the 80s (problem for the "catching up" hypothesis).

- Much of the difference is "explained" by differences in "human-capital-like" variables such as schooling and test-scores

- "Returns to AFQT" the same for blacks and whites. Neal & Johnson argues that this is evidence against statistical discrimination .

ln($wages$), NLSY males, 1991

	Blacks	Whites
AFQT	0.208 (0.031)	0.183 (0.017)

The standard argument

- We can observe measures of skill (e.g. AFQT)

- If minorities appear to have lower returns to skill, they have less incentives to invest in skills

- Test: look at difference in returns to skill between groups: if they are insignificant, then statistical discrimination is rejected

Examples: Neal and Johnson (JPE 1996), Persico, Postlewaite and Silverman (2004)

Problem with the argument

- Measures of skill are not perfectly correlated with market valued skills.
- Presumably, the econometrician cannot observe the same signals that employers observe
- Using a different signal introduces an "error in variable" bias.

A simple model to illustrate the problem

- Human capital investment $h \in \{0, 1\}$
- Cost of investment, worker $i \sim G$ is $C(i) = i$
- Workers with human capital are called qualified and produce 1; unqualified produce 0
- Competitive firms observe only a noisy signal of productivity $z \in \{good, bad\}$

A simple model to illustrate the problem

- Human capital investment $h \in \{0, 1\}$
- Cost of investment, worker $i \sim G$ is $C(i) = i$
- Workers with human capital are called qualified and produce 1; unqualified produce 0
- Competitive firms observe only a noisy signal of productivity $z \in \{good, bad\}$

Type of worker	Probability of obtaining $z =$*good*
Qualified	p_q
Unqualified	$p_u < p_q$

Incentives to invest in human capital

- π^J = proportion of people who invest in group J

- Firms pay expected productivity, computed using Bayes' rule:

$$w^J(good;\pi^J) = \frac{\pi^J p_q}{\left(1-\pi^J\right) p_u + \pi^J p_q}$$

$$w^J(bad;\pi^J) = \frac{\pi^J (1-p_q)}{\pi^J (1-p_q) + \left(1-\pi^J\right) (1-p_u)}$$

Incentives to invest in human capital

- π^J = proportion of people who invest in group J

- Firms pay expected productivity, computed using Bayes' rule:

$$w^J(good;\pi^J) = \frac{\pi^J p_q}{\left(1-\pi^J\right) p_u + \pi^J p_q}$$

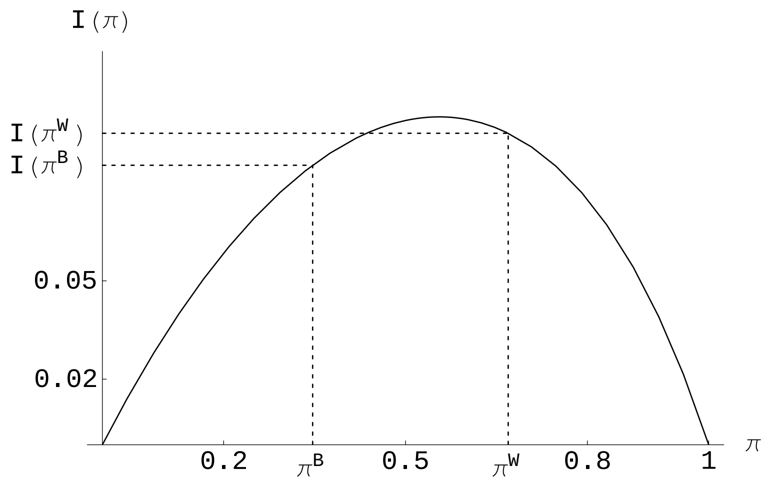
$$w^J(bad;\pi^J) = \frac{\pi^J (1-p_q)}{\pi^J (1-p_q) + \left(1-\pi^J\right) (1-p_u)}$$

- Incentives to invest :

$$\begin{aligned} I(\pi^J) &= E_z\left[w^J(z;\pi^J)|\text{invest}\right] - E\left[w^J(z;\pi^J)|\text{don't}\right] \\ &= (p_q - p_u)\left[w^J(good;\pi^J) - w^J(bad;\pi^J)\right] \end{aligned}$$

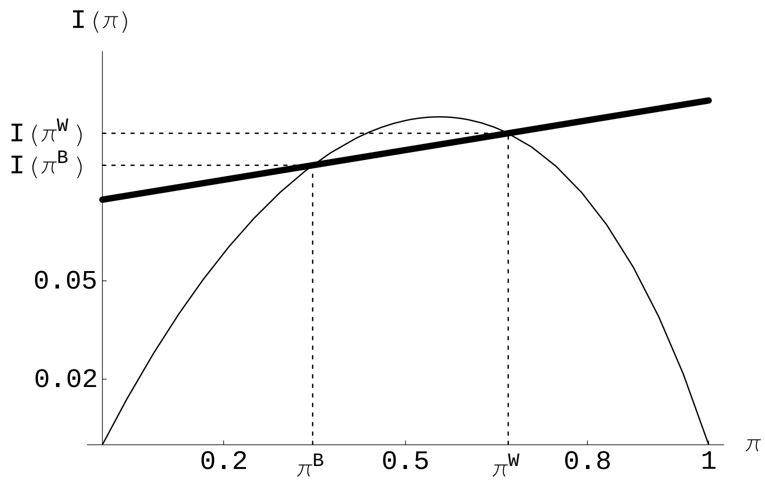
Incentives to invest in human capital

Example drawn using $p_q = .8$ and $p_u = .5$.



Equilibrium

Example drawn using $p_q = .8$ and $p_u = .5$.



Using an appropriate distribution of costs G (thick line) we can support a pair of equilibrium fraction of investors $\pi^B < \pi^W$ so that

$$\pi^J = G(I(\pi^J))$$

The econometric problem

Consider an econometrician observing $x \in \{HIGH, LOW\}$, independent from the firms' signal

Type of worker	Probability of obtaining $x =$*HIGH*
Qualified	r_q
Unqualified	$r_u < r_q$

Returns to the observable measure of skill

The econometrician can measure:

$$R^J(\pi^J) = E[w^J|HIGH] - E[w^J|LOW]$$

Proposition: $R^J(\pi^J) < I^I(\pi^J)$

Returns to the observable measure of skill

The econometrician can measure:

$$R^J(\pi^J) = E[w^J|HIGH] - E[w^J|LOW]$$

Proposition: $R^J(\pi^J) < I^I(\pi^J)$

Intuition:

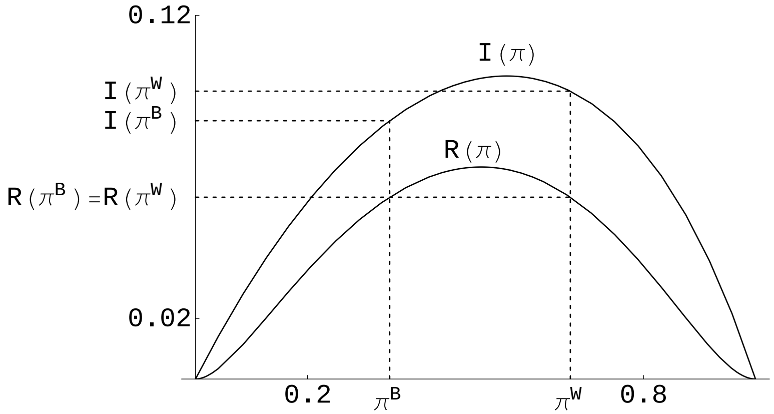
$HIGH$ workers may have "firm's signal" $good$ or bad.
$\Longrightarrow E[w^J|HIGH] < w^J(good)$

LOW workers may have firm's signal $good$, or bad,
$\Longrightarrow E[w^J|LOW] > w^J(bad)$

The possibility of an erroneous conclusion

The bias depends on π.

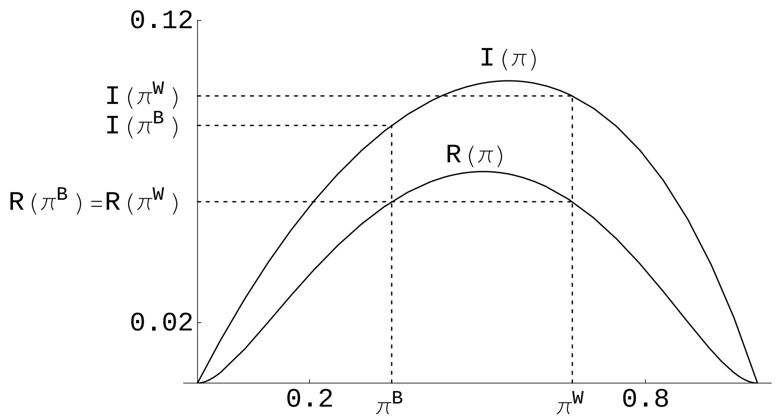
Using $\underbrace{p_q = 0.8, p_u = 0.5}_{\text{Firm's signal}}$, $\underbrace{r_q = 0.8, r_u = 0.1}_{\text{Econometrician's signal}}$



The possibility of an erroneous conclusion

The bias depends on π.

Using $\underbrace{p_q = 0.8, p_u = 0.5}_{\text{Firm's signal}}, \underbrace{r_q = 0.8, r_u = 0.1}_{\text{Econometrician's signal}}$



Note: the econometrician's signal is more informative signal than firms'signal

The Model to be Estimated

- Continuous human capital h . Cost of h is $C(h,k) = h/k$, $\ln(k) = N(\mu_k, \sigma_k)$
- Signal observed by firms: $z = \ln(h) + \varepsilon$, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$
- Preferences $u(w,h) = \ln\, w\, - c(h,k)$
- Competitive firms $w\,(z) = E\,[h|z]$

Looking for a Log-Normal Equilibrium

Assume log normal h (later verify this is the case) with mean/var. μ_h, σ_h^2

$$z = \ln(h) + \varepsilon \implies f(\ln(h)|z) = N\left(\mu_h \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_k^2} + z\frac{\sigma_k^2}{\sigma_\varepsilon^2+\sigma_k^2}, \left(\frac{\sigma_k^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_k^2}\right)^2\right)$$

Wages are log-linear in z:

$$\begin{aligned} w(z) &= E(h|z) = \exp\left(\overbrace{\mu_h \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_k^2}}^{\alpha} + z\underbrace{\frac{\sigma_k^2}{\sigma_\varepsilon^2+\sigma_k^2}}_{\beta} + \overbrace{\frac{1}{2}\frac{\sigma_k^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_k^2}}^{\alpha}\right) \\ \ln(w) &= \alpha + \beta \cdot z \end{aligned}$$

Workers' problem

Expected utility linear in $\ln(h)$:

$$\begin{aligned} E_z\left[\ln\left(w\left(z\right)\right)|h\right] &= E_z\left[\alpha+\beta z|h\right] = \\ E_z\left[\alpha+\beta(\ln(h)+\varepsilon)|h\right] &= \alpha+\beta E_z\left(z|h\right) = \alpha+\beta\ln h \end{aligned}$$

$$\begin{aligned} & \max_{h\geq 0} \alpha+\beta\ln\left(h\right)-\frac{h}{k} \\ \Longrightarrow \quad & h\left(k\right)=\beta k \end{aligned}$$

i.e. human capital is indeed lognormal, $\ln(h) \sim N(\mu_k+\ln(\beta),\sigma_k^2)$

Restriction imposed by the equilibrium

$$\begin{aligned} \text{Stdev of log } h : \sigma_h = &\quad \sigma_k \\ \text{Mean of log } h : \mu_h = &\quad \mu_k + \ln \beta = \mu_k + \ln\left(\frac{\sigma_h^2}{\sigma_\varepsilon^2 + \sigma_h^2}\right) \\ = &\quad \mu_i + \ln\left(\frac{\sigma_k^2}{\sigma_\varepsilon^2 + \sigma_k^2}\right). \end{aligned}$$

Equilibrium

Model has unique log-normal equilibrium (generating log-linear wages).

For any $(\mu_k, \sigma_k, \sigma_\varepsilon)$ there is an equilibrium where

$$
\begin{aligned}
h(k) &= \beta k \\
w(z) &= \exp(\alpha + \beta z),
\end{aligned}
$$

where

$$
\begin{aligned}
\alpha &\equiv \mu_h \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_k^2} + \frac{1}{2} \frac{\sigma_k^2 \sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_k^2} \\
\beta &\equiv \frac{\sigma_k^2}{\sigma_\varepsilon^2 + \sigma_k^2}.
\end{aligned}
$$

In this equilibrium, $\ln h \sim N\left(\mu_k + \ln \beta, \sigma_k^2\right)$.

The econometric problem

- Assume that the econometrician observes a proxy of skill

$$x = \ln h + \delta,$$

  where $\delta \sim N\left(0, \sigma_\delta^2\right)$ is assumed to be independent of ε.

- Since $\ln h = z - \varepsilon$ it follows immediately that

$$x = z - \varepsilon + \delta,$$

  which means that a standard OLS regression of wages on AFQT scores leads to a downwards biased estimate of β in the equilibrium wage function.

Error in variable bias

- $x = \ln(h) + \delta$: econometrician's variable

- $z = \ln(h) + \varepsilon$: firms' signal

- $\ln(w) = \alpha + \beta z = \alpha + \beta x + \beta(-\delta + \varepsilon)$

- The regressor (x) is correlated with the disturbance

$$\Rightarrow p\lim(b_{LS}) = \beta \cdot \frac{\sigma_k^2}{\sigma_\delta^2 + \sigma_k^2} = \frac{\sigma_k^2}{\sigma_\varepsilon^2 + \sigma_k^2} \frac{\sigma_k^2}{\sigma_\delta^2 + \sigma_k^2}$$

Data

NLSY79, 15 to 18 in 1980. Wages observed in 1991.

	< High Sc.		High Sc.		College or more	
	Black	White	Black	White	Black	White
Obs.	(75)	(109)	(323)	(483)	(52)	(219)
E[ln(wage)]	6.46	6.64	6.61	6.84	7.06	7.12
SD[ln(wage)]	0.33	0.41	0.44	0.4	0.39	0.42
E[AFQT]	-1.1	-0.71	-0.61	0.34	0.46	1.3
SD[AFQT]	0.51	0.69	0.73	0.77	0.81	0.54
Corr[wage,AFQT]	0.04	0.4	0.18	0.17	0.41	0.22

(full sample)

Identification strategy

- We observe AFQT, not x, therefore assume for some C, D :

$$C + D \cdot AFQT_i = \ln(h_i) + \delta_i$$

- Assume wages are observed with measurement error $u \sim N(0, \sigma_u^2)$

- Restrict some parameters to be identical across groups: C, D, σ_δ

- Use restrictions implied by the model and its equilibrium

10 parameters to be estimated

- $\mu_k^B, \mu_k^W, \sigma_k^{2B}, \sigma_k^{2W}$: distributions of the investment cost
- $\sigma_\varepsilon^{2B}, \sigma_\varepsilon^{2W}$: the variance in firms' signal
- σ_u^2: measurement error in wage data
- C, D, σ_δ^2: scaling of AFQT and variance of scaled test

10 parameters to be estimated

- $\mu_k^B, \mu_k^W, \sigma_k^{2B}, \sigma_k^{2W}$: distributions of the investment cost
- $\sigma_\varepsilon^{2B}, \sigma_\varepsilon^{2W}$: the variance in firms' signal
- σ_u^2: measurement error in wage data
- C, D, σ_δ^2: scaling of AFQT and variance of scaled test

We can then compute incentives using equilibrium restriction

$$\beta^J = \frac{\sigma_k^{J2}}{\sigma_\varepsilon^{J2} + \sigma_k^{J2}}, J = B, W$$

Identifying Conditions

$$
\begin{aligned}
p\,\text{lim}(b_{LS}^J) &= D\beta^J \frac{\sigma_k^{2J}}{\sigma_k^{2J} + \sigma_\delta^2} \\
\ln(E[w^J]) &= \mu_k^J + \ln \beta^J + \frac{\sigma_k^{2J}}{2} \\
Var[\ln(w^J)] &= \beta^J \sigma_k^{2J} + \sigma_u^2 \\
C + D \cdot E[AFQT^J] &= \mu_k^J + \ln(\beta^J) \\
D^2 VAR[AFQT^J] &= VAR[\ln(h^J)] = \sigma_k^{2J} + \sigma_\delta^2
\end{aligned}
$$

10 conditions in 10 unknowns, but not all parameters are identified

What we can identify

High School Sample	Estimates	Stderr
D	0.199	0.037
$\sigma_k^{2W} - \sigma_k^{2B}$	0.0023	0.0025
$\mu_k^W + \ln(\beta^W) - \left(\mu_k^B + \ln(\beta^B)\right)$	0.189	0.036
$\beta^B \sigma_k^{2B}$	0.0113	0.0045
$\beta^W \sigma_k^{2W}$	0.0106	0.0037
σ_u^{2B}	0.178	0.022
σ_u^{2W}	0.147	0.011

Additional restrictions from the model

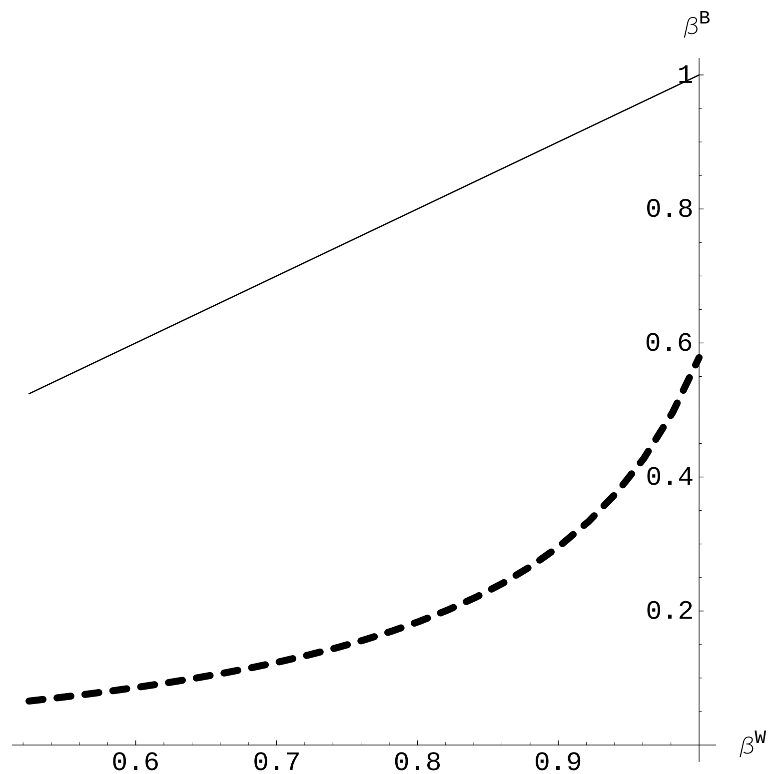
Use $\beta < 1$ and $\sigma_\delta > 0$ to provide an upper bound for σ_k and a lower bound for β

High School	Estimates	Stderr
$\underline{\beta}^B$	0.532	0.210
$\underline{\beta}^W$	0.452	0.171

$<$ High School	Estimates	Stderr
$\underline{\beta}^B$	0.065	0.179
$\underline{\beta}^W$	0.524	0.296

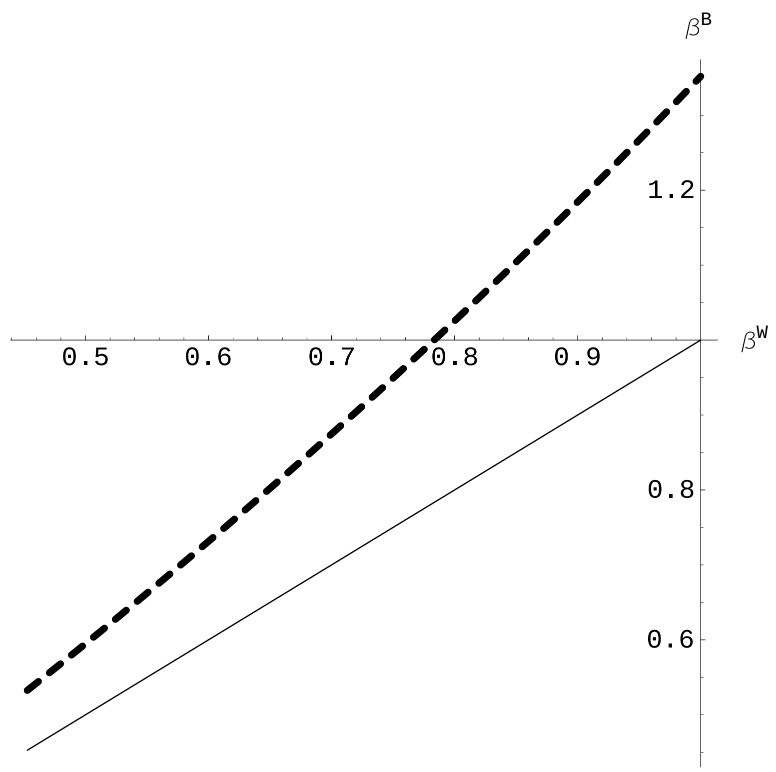
Scenarios, less than high school sample

Solid line: β^W
Dotted line $\beta^B(\beta^W)$



Scenarios, high school sample

Solid line: β^W
Dotted line $\beta^B(\beta^W)$



Conclusion

- A naive look at returns to observable (to the investigator's) may give us biased conclusions about the importance of statistical discrimination

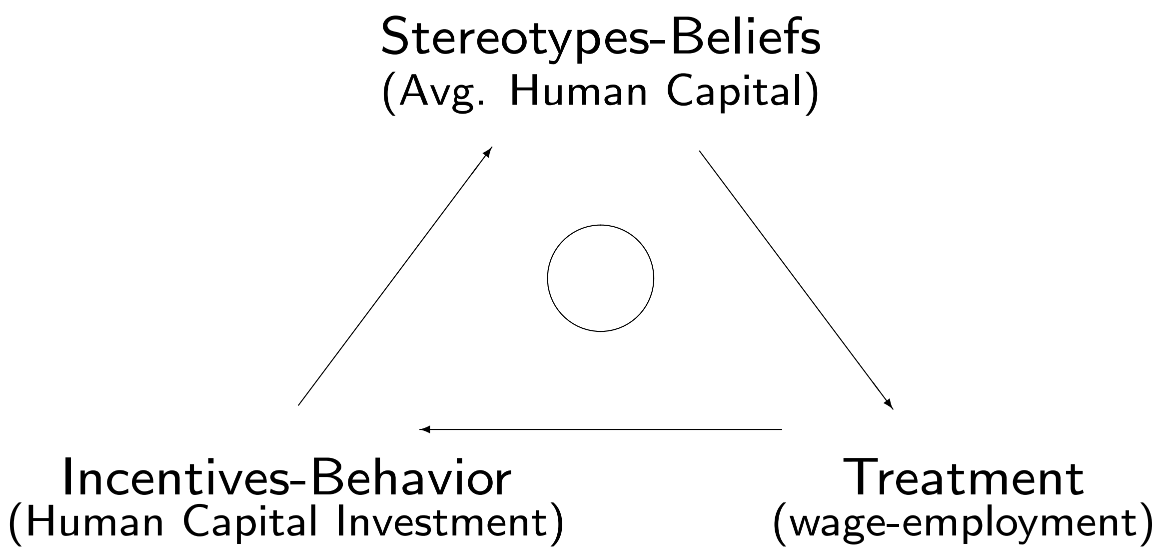
- We look at the data with the guidance of the restrictions imposed by a formal equilibrium model

- Even if we don't achieve full identification, we can provide some clues

- Preliminary results: black high school graduate are statistically discriminated against, but not black high school dropouts

The End

Statistical discrimination: a theory of self fulfilling stereotypes



Incomplete information is crucial.

Data, full sample

Full Sample

	Black	White
N. of obs.	466	825
E[wage]	6.64	6.89
SD[wage]	0.46	0.43
E[AFQT]	-0.57	0.44
SD[AFQT]	0.82	0.93
Corr[wage,AFQT]	0.34	0.38