Empirical Implications of Statistical Discrimination on the Returns to Measures of Skill

Andrea Moro and Peter Norman

(U. of Minnesota - U. of Wisconsin, Madison)

June 2003

Topic of interest

Quantitatively measure of how different sources of discrimination contribute to wage inequality

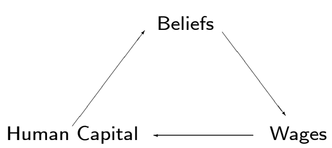
Today

Present a simple model of statistical discrimination.

Estimate the model using NLSY data

How much does statistical discrimination contribute to wage inequality?

Statistical discrimination as self fulfilling stereotypes



- Incomplete information is crucial (we don't observe human capital investment).
- In equilibrium, minority workers have lower incentives to acquire human capital.

A standard argument

- We observe measures of skill (e.g. AFQT, education), and minorities have lower average skill
- Use returns to skill as a proxy for returns to human capital investment.
- Test : look at difference in returns to skill between groups: if they are insignificant, then statistical discrimination is rejected

Examples

- Derek Neal and William Johnson (JPE 1996) on racial differences: returns to AFQT are not significantly different between black and white workers.
- Nicola Persico, Andrew Postlewaite, and Dan Silverman (2002) on height wage differences (a similar test).

Problem with the argument (Moro and Norman, 2003)

- Measures of skill are not perfectly correlated with ability or productivity.
- The econometrician cannot observe the same signals that employers have
- The econometrician's estimate of the returns to his signal of productivity is a biased measure of the return to the firms' signal
- The bias is different across groups

A model of statistical discrimination

Continuous human capital	h
Cost of h	$C(h,i)=\frac{h}{i}$, $\ln(i)=N(\mu_i,\sigma_i)$
Firms' observe signal	$x=\ln(h)+\varepsilon$, $\varepsilon\sim N(0,\sigma_\varepsilon)$
Preferences	$u(w,h)=\ln(w)-c(h,i)$
Technology	production $=h$

Perfectly competitive labor mkts.

Equilibrium

Assume $\ln(h)\sim N(\mu_h,\sigma_h)$ (later verify this is the case)

Firms' signal $x=\ln(h)+\varepsilon$

$$\Longrightarrow f(\ln(h)|x)=N\left(\mu_h\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_h^2}+x\frac{\sigma_h^2}{\sigma_\varepsilon^2+\sigma_h^2}\quad,\quad\frac{\sigma_h\sigma_\varepsilon}{\left(\sigma_\varepsilon^2+\sigma_h^2\right)^{\frac{1}{2}}}\right)$$

$$w(x)=E(h|x)=\exp\left(\mu_h\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_h^2}+x\frac{\sigma_h^2}{\sigma_\varepsilon^2+\sigma_h^2}+\frac{1}{2}\frac{\sigma_h^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_h^2}\right)$$

Log wages are linear in x:

$$\ln(w(x))=\alpha+\beta\cdot x$$

Workers' problem

$u(w,h)=\ln(w)-c(h,i)\Longrightarrow$Expected utility linear in $\ln(h)$:

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

Workers's choice of human capital:

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

With $h\left(i\right)=\beta\cdot i$ human capital is indeed lognormal

$\ln(h)\sim N(\mu_i+\ln(\beta),\sigma_i)$, hence consistency requires:

$$\begin{aligned}\sigma_h &= \sigma_i\\ \mu_h &= \mu_i+\ln\beta=\mu_i+\ln\left(\frac{\sigma_i^2}{\sigma_\varepsilon^2+\sigma_i^2}\right).\end{aligned}$$

Note:

$$\beta=\frac{\sigma_i^2}{\sigma_\varepsilon^2+\sigma_i^2}$$

Summary

- We can compute only one equilibrium (there may be others)
- Our approach: use exogenous differences to rationalize difference in behavior

e.g. $\sigma_\varepsilon^B>\sigma_\varepsilon^W\Rightarrow E^B(h)<E^W(h)$

Econometricians observe a different signal

True d.g.p:

$$\ln\left[w_i^J(x)\right]=\alpha^J+\beta^J x_i$$

$$x_i=\ln(h_i)+\varepsilon_i,\varepsilon_i\sim N(0,\sigma_\varepsilon)$$

But the investigator observes

$$z_i=\ln(h_i)+\delta_i$$

$$\delta_i\sim N(0,\sigma_\delta)$$

Data

NLSY 79, 1990 wages and test scores of young males aged less than 18 when they took the test (1980)

Test: AFQT (verbal, math and arithmetic skills)

	Black	White
Observations	466	825
$\overline{w}^J =$ Average$(log(w))$	6.64	6.89
$\sigma_w^J \equiv$ Stdev$(log(w))$	.46	.43
$\sigma_z^J \equiv$ Stdev$(log(z))$	.82	.93
$\widehat{\beta}_{LS}^J$	0.19 (.02)	0.18 (.02)

Likelihood Function

Hence given a dataset $D = \{w_i, z_i\}_{i=1}^N$ our log likelihood is

$$\begin{aligned} l(\sigma_i, \sigma_\varepsilon, \mu_h, \sigma_\delta | D) &= \sum_{i=1,N} \log \left[f(\ln w_i, z_i) \right] \\ &= \sum_{i=1,N} \log f(\ln w_i | z_i) + \log f(z_i) \end{aligned}$$

Results

		Black	Whites
(cost parameters)	μ_i	4.103871 (.0905601)	5.062817 (.0560569)
	σ_i	13.61844 (.0742783)	12.18787 (.0602092)
(firms' signal)	σ_ε	862.3624 (76.65127)	781.4184 (38.2517)
(econometrician's signal)	σ_δ	˜0 (˜0)	˜0 (˜0)

Simulations

Question: what happens if there were no informational differences, i.e. if the employers had a "race-neutral" test?

Average Wage	Black	Whites	Δ
Data	853.4	1075.6	222.2
Experiment 1 $\sigma_\varepsilon^b = \widehat{\sigma_\varepsilon^w}$	940.2	1075.6	135.4
Experiment 2 $\sigma_\varepsilon^b = \sigma_\varepsilon^w = \frac{\widehat{\sigma_\varepsilon^w} + \widehat{\sigma_\varepsilon^b}}{2}$	894.7	1023.3	128.6

I.e. "Statistical discrimination" accounts for about 40% of the wage differential

The bias of OLS regression

$$\widehat{b} = \frac{Cov_N(z_i, \ln(w_i))}{Var_N(z_i)}$$

$$\begin{aligned} Cov_N(z_i, \ln(w_i)) &= Cov_N\left(\ln h_i + \delta_i, \alpha + \beta(\ln h_i + \varepsilon_i)\right) \\ &= \beta Cov_N(\ln h_i, \ln h_i) + \beta Cov_N(\ln h_i, \varepsilon_i) \\ &\quad + \beta Cov_N(\delta_i, \ln h_i) + \beta Cov_N(\delta_i, \varepsilon_i) \end{aligned}$$

$$\text{p}\lim(\widehat{b}) = \beta \left(\frac{\sigma_i^2}{\sigma_\delta^2 + \sigma_i^2} \right) = \frac{\sigma_i^2}{\sigma_\varepsilon^2 + \sigma_i^2} \cdot \frac{\sigma_i^2}{\sigma_\delta^2 + \sigma_i^2}$$