Discrimination and learning

Heidi L. Williams

MIT 14.662

Spring 2015

Discrimination and learning

- Previous models of statistical discrimination were static
- In contrast, Altonji and Pierret (2001) use a dynamic model of employer learning to test for statistical discrimination
  - Theoretical model builds closely on Farber and Gibbons (1996)
- Coate and Loury (1993) use a dynamic model of worker investments to analyze affirmative action

Roadmap for today

- Preliminaries: Farber and Gibbons (1996)
- Testing statistical discrimination: Altonji and Pierret (2001)
- Affirmative action: Coate and Loury (1993)

Farber-Gibbons (1996)

- At the time of labor market entry:
  - Some characteristics (education) observed by employers, but likely convey only partial information about productivity
  - Over time: worker gains experience, more information revealed
- Key insight: the econometrician may observe variables measuring productivity that are *not* observed by employers
  - Example: AFQT scores
  - Can ask how employers learn about these over time
- Farber-Gibbons: implications of employer learning for wages
  - Influential framework
  - Tractable model
  - Empirically testable implications, generally supported by the data

Model: Set-up

- η_i: innate (time-invariant) ability
  - Not observed by employers nor by econometrician
- s_i: (time-invariant) schooling
  - Observed by employers and by econometrician
- X_i: (time-invariant) attributes other than schooling (race)
  - Observed by employers and by econometrician
- Z_i: (time-invariant) attributes (school quality)
  - Observed by employers; not observed by econometrician
- B_i: (time-invariant) attributes (AFQT)
  - Not observed by employers; observed by econometrician

Allow for arbitrary joint distribution $F(\eta_i, s_i, X_i, Z_i, B_i)$

Model: Set-up

- y_{it}: output of i in worker's t^{th} period in labor market
  - $\{y_{it} : t = 1, ..., T\}$: independent draws from conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$
  - Note: B_i does not appear in this conditional distribution (assumes no direct effect on output; can affect output via other variables, like η_i)
- Assume:
  1. Employers know $F(\eta_i, s_i, X_i, Z_i, B_i)$ and $G(y_{it}|\eta_i, s_i, X_i, Z_i)$
  2. Employers observe s_i, X_i, and Z_i
  3. Employers observe outputs $\{y_{i1}, ..., y_{it}\}$ through period t
     - Strong "public learning" assumption

Model: Set-up

Wage paid to a worker in period t is her expected output given all available information available at t about the worker:

$$w_{it} = E\left(y_{it}|s_i, X_i, Z_i, y_{i1}, ..., y_{it-1}\right)$$

Spot-market model of wage determination

- Rules out long-term contracts; strong assumption
- Could be that long-term contracts are not useful, or that they are useful but impossible to enforce

Model: Predictions

Three predictions that can be tested in an earnings regression:

1. Effect of schooling on wages independent of experience
2. Time-invariant worker characteristics correlated with ability but unobserved by employers increasingly correlated with wages as experience increases
3. Wage residuals a martingale

Prediction #1: The effect of schooling on wages

- Consider a panel data set of one cohort of workers
  - Data on s_i and X_i
  - Data on wage in each year $(t = 1, 2, ..., T)$
- Can estimate the following earnings regression:

$$w_{it} = \alpha_t + \beta_t s_i + X_i \gamma_t + \varepsilon_{it}$$

Notes:

- Z_i by construction not included
- Specified in levels, not logs

Prediction #1: The effect of schooling on wages

Notation:

- $E^*(\cdot)$: linear projection
- $E(\cdot)$: conditional expectation

Estimated coefficients ($\hat{\alpha}_t$, $\hat{\beta}_t$, $\hat{\gamma}_t$) are coefficients from linear projection $E^*(w_{it}|s_i, X_i)$ of w_{it} on s_i and X_i:

$$E^*(w_{it}|s_i, X_i) = \hat{\alpha}_t + \hat{\beta}_t s_i + X_i \hat{\gamma}_t$$

Prediction #1: The effect of schooling on wages

Recall:

1. Version of law of iterated expectations: $E^*(E(y|x,z)|x) = E^*(y|x)$ [see notes]
2. $w_{it} = E\left(y_{it}|s_i, X_i, Z_i, y_{i1}, ..., y_{it-1}\right)$

$$\begin{aligned} E^*\left(w_{it}|s_i, X_i\right) &= E^*\left(E\left(y_{it}|s_i, X_i, Z_i, y_{i1}, ..., y_{it-1}\right)|s_i, X_i\right) \\ &= E^*\left(y_{it}|s_i, X_i\right) \end{aligned}$$

Prediction #1: The effect of schooling on wages

- Recall:
  1. s_i, X_i time-invariant
  2. y_{it} independent and identically distributed draws
- $\Rightarrow E^*(y_{it}|s_i, X_i)$ is independent of t
- $\Rightarrow$ effect of schooling on wages is independent of experience

Prediction #1: The effect of schooling on wages

Some intuition:

- Recall:
  1. Wages are assumed to equal expected output
  2. Outputs are independent and identically distributed draws
- $\Rightarrow w_{i1}$ is expectation of first period output given s_i and X_i
- No part of 'innovation' in wages between first, second periods $(w_{i2} - w_{i1})$ can be forecast from information determining w_{i1}
- $\Rightarrow w_{i2} = w_{i1} +$ term depending on y_{i1} but orthogonal to s_i, X_i
- $\Rightarrow$ estimated coefficients on s_i and X_i are the same in the first and second and all subsequent periods

Prediction #2: Unobserved characteristics

- Recall: B_i in data but not observed by employers
  - Note that other variables observable to employers (s_i, X_i, and Z_i) could be correlated with B_i
- Want to create a vector of variables orthogonal to employers' information when worker enters labor market
- B_i^*: residual from a regression of B_i on all the other variables in the data (s_i, X_i) and the worker's initial wage w_{i1}

$$B_i^* = B_i - E^*\left(B_i | s_i, X_i, w_{i1}\right)$$

- Including w_{i1} conditions out employers' information about B_i
  - Caveat: measurement error in initial wage

Prediction #2: Unobserved characteristics

Now add B_i^* as a regressor to our wage equation:

$$w_{it} = \alpha_t + \beta_t s_i + X_i \gamma_t + B_i^* \pi_t + \varepsilon_{it}$$

The question here is how π_t will vary with experience

Prediction #2: Unobserved characteristics

Specialize to case where B is a scalar

- B^* is by construction orthogonal to the other regressors
- $\Rightarrow \hat{\pi}_t = \frac{cov(B_i^*, w_{it})}{var(B_i^*)}$
- We can then write:

$$\begin{aligned} w_{it} &= w_{it-1} + \zeta_{it} \\ &= w_{i1} + \sum_{\tau=2}^{t} \zeta_{i\tau} \end{aligned}$$

where ζ_{it} is the innovation in wages in each period

Prediction #2: Unobserved characteristics

B_i^* is orthogonal to w_{i1} by construction $\Rightarrow \hat{\pi}_1 = 0$ and:

$$\begin{aligned} cov\left(B_i^*, w_{it}\right) &= cov\left(B_i^*, w_{i1} + \sum_{\tau=2}^{t} \zeta_{i\tau}\right) \\ &= cov\left(B_i^*, w_{i1}\right) + cov\left(B_i^*, \sum_{\tau=2}^{t} \zeta_{i\tau}\right) \\ &= \sum_{\tau=2}^{t} cov\left(B_i^*, \zeta_{i\tau}\right) \end{aligned}$$

$cov\left(B_i^*, w_{it}\right)$ will "generally" be positive for every τ

- $\Rightarrow \hat{\pi}_t$ will increase with t
- $\Rightarrow$ if B_i^* is correlated with ability, then the estimated effect of B_i^* on wages should increase with experience

Prediction #2: Unobserved characteristics

Helpful to compare effect of characteristics market cannot observe (B_i^*) with effect of characteristics market can observe (s_i, X_i)

- By construction, former play no role in wage determination, but their estimated effect increases over time as the market learns about ability by observing output
- Latter play a declining role in the market's inference process but have a constant estimated effect

Key prediction of the model

Prediction #3: Wage residuals

- $E\left(\zeta_{it}|w_{it-1}\right) = 0 \Rightarrow$ wages are a martingale: $E\left(w_{it}|w_{it-1}\right) = w_{it-1}$
  - You may be thinking: what is a martingale?
  - Not the focus of Altonji-Pierret test, so see paper for details
- Fact that measured wage growth increases with experience implies wages *not* a martingale; empirics focus on related prediction that wage residuals (not wages) are a martingale

Theory: Time-variant worker characteristics

Model thus far ruled out productivity growth with experience

- Assume i^{th} worker's output in period t is $Y_{it} = y_{it} + h(t)$
  - y_{it}: part of output due to innate ability
  - $h(t)$: part of output due to acquired skill
- Assume output grows with experience by $h(t)$ (OJT)
  - $h(t)$: deterministic, linear
- Write down new wage equation:

$$w_{it} = \alpha_0 + \alpha_1 t + \beta_0 s_i + \beta_1 s_i t + \varepsilon_{it}$$

Empirical analysis

NLSY data

- Panel data: wage dynamics for individuals
- Focus on younger workers
- Detailed experience measures
- AFQT score as a measure of B_i

Construct B_i^*

Many of the determinants of B_i are observable by the market, but we can condition out other observables and the first period wage in order to construct a measure B_i^* that is not observed by employers:

$$B_i^* = B_i - X_i\hat{\gamma} - \hat{\delta} w_{i0}$$

- w_{it} incorporates all information market has on worker's ability
- Caveat: wage may be measured with error $\Rightarrow$ B_i^* term will not be completely purged of attributes observed by the market
- Farber and Gibbons focus on AFQT, library card at age 14 (latter is a proxy for family background)

Table 2

Table 2 tests Farber and Gibbons' first and second predictions:

1. The estimated effect of schooling on the level of wages should be independent of experience
2. Time-invariant worker characteristics correlated with ability but unobserved by employers should be increasingly correlated with wages as experience increases

Table 2

- Column (1): reports the means and standard deviations
- Column (2): basic earnings regression
- Column (3): adds AFQT and library card residuals

Consistent with the model:

1. No evidence that relationship between earnings and education varies with experience
2. Interactions of the AFQT residuals with experience are positive

Farber and Gibbons (1996): Table 2

TABLE II
REGRESSION ANALYSIS OF EARNINGS FUNCTION

Independent variable	(1) Mean [sd]	(2) Wage (level)	(3) Wage (level)	(4) Wage (level)	(5) Wage (Level)	(6) Wage (log)
Constant	1.0	−3.5579	−3.8086	−6.0321	−2.7034	0.0873
		(0.785)	(0.788)	(0.928)	(0.388)	(0.124)
Experience	5.1804	0.4428	0.5054	0.5366	0.2697	0.1012
	[2.502]	(0.102)	(0.103)	(0.100)	(0.069)	(0.013)
Experience squared	33.0953	−0.0178	−0.0185	−0.0178	−0.0198	−0.0027
	[29.947]	(0.003)	(0.003)	(0.003)	(0.003)	(0.000)
Education	13.0450	0.6745	0.6938	0.6719	0.4602	0.0989
	[2.349]	(0.061)	(0.061)	(0.059)	(0.024)	(0.007)
Education × experience	67.5424	−0.0004	−0.0049	−0.0041	0.0172	−0.0026
	[35.014]	(0.008)	(0.008)	(0.007)	(0.005)	(0.001)
AFQT residual/100	0.0024	—	0.6494	0.8734	0.7841	0.1880
	[0.148]		(0.307)	(0.291)	(0.292)	(0.044)
AFQT resid/100 × experience	0.0189	—	0.1938	0.1848	0.1922	0.0187
	[0.856]		(0.064)	(0.060)	(0.060)	(0.008)
Lib card residual/10	−0.0002	—	0.2583	0.2130	−0.0579	0.1440
	[0.043]		(1.035)	(0.988)	(0.989)	(0.146)
Lib card resid × experience/10	−0.00011	—	0.6035	0.6169	0.6448	0.0588
	[0.248]		(0.205)	(0.192)	(0.192)	(0.026)
Year		yes	yes	yes	no	yes
Education × year		yes	yes	yes	no	yes
Other demographic		no	no	yes	yes	yes
R^2		0.215	0.224	0.294	0.289	0.296

The dependent variable is real hourly earnings on the current job (in levels in columns (2)–(5) and in logs in column (6). The mean of the level of earnings is 6.91 (s.d. = 3.30). The mean of the log of earnings is 1.83 (s.d. = 0.448). The numbers in parentheses are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. There are 28,984 wage observations on 4970 individuals. Where included, there are ten year dummies for 1981–1990 and interactions of education with each of the ten year dummies. The base year is 1991. The other demographic characteristics, where included, consist of age at entry, a dummy variable for part-time, the interaction of part-time with education, and dummy variables for collective bargaining coverage, race, sex, marital status, and the interaction of sex and marital status.

Prediction #3

Farber and Gibbons also analyze model's third prediction:
Wage residuals should be a martingale

Not critical to understanding the Altonji-Pierret tests; see paper

Overview of Altonji and Pierret (2001)

Do employers statistically discriminate among young workers on the basis of observable characteristics such as education and race, and as they learn over time do they rely less on such variables?

- Employer learning model as in Farber-Gibbons:
  - Information common across firms
  - Labor market is competitive
- Focus on variables such as race, which employers observe and could be correlated with AFQT scores
- Key idea: statistical discrimination with employer learning should imply coefficient on AFQT will rise with experience whereas (conditional on AFQT) coefficient on race will fall

Differences: Altonji-Pierret and Farber-Gibbons

Very similar models; key differences:

1. Altonji-Pierret model specified in logs rather than levels
2. Whereas Farber-Gibbons orthogonalize B_i with respect to X_i and w_{i0}, Altonji-Pierret do *not* do this - they are interested in how changes in relationship between B_i and wages over time affects coefficients on X_i's such as race and schooling

Altonji-Pierret model in one slide

Proposition 1: Assume schooling s is correlated with the initially unobserved variable z (AFQT score). If we include z in the wage regression with a time-varying coefficient, then as employers learn about the productivity of workers the observable variable s (schooling) will get less of the credit for an association with productivity as z can claim the shifting credit.

Do employers statistically discriminate on education?

- Column (1): education, black, AFQT, educ-exp interaction
- Column (2): adds AFQT-experience interaction
  - Effect of AFQT rises from 0 (exp=0) to 0.0692 (exp=10)
  - Supports that employers learn about productivity over time
  - Coefficient on education-experience interaction declines: supports that employers statistically discriminate on education

Altonji and Pierret (2001): Table 1

TABLE 1
THE EFFECTS OF STANDARDIZED AFQT AND SCHOOLING ON WAGES
Dependent Variable: Log Wage; OLS estimates (standard errors).

Panel 1—Experience measure: potential experience

Model:	(1)	(2)	(3)	(4)
(a) Education	0.0586	0.0829	0.0638	0.0785
	(0.0118)	(0.0150)	(0.0120)	(0.0153)
(b) Black	−0.1565	−0.1553	0.0001	−0.0565
	(0.0256)	(0.0256)	(0.0621)	(0.0723)
(c) Standardized AFQT	0.0834	−0.0060	0.0831	0.0221
	(0.0144)	(0.0360)	(0.0144)	(0.0421)
(d) Education * experience/10	−0.0032	−0.0234	−0.0068	−0.0193
	(0.0094)	(0.0123)	(0.0095)	(0.0127)
(e) Standardized AFQT * experience/10		0.0752		0.0515
		(0.0286)		(0.0343)
(f) Black * experience/10			−0.1315	−0.0834
			(0.0482)	(0.0581)
R^2	0.2861	0.2870	0.2870	0.2873

Panel 2—Experience measure: actual experience instrumented by potential experience

Model:	(1)	(2)	(3)	(4)
(a) Education	0.0836	0.1218	0.0969	0.1170
	(0.0208)	(0.0243)	(0.0206)	(0.0248)
(b) Black	−0.1310	−0.1306	0.0972	0.0178
	(0.0261)	(0.0260)	(0.0851)	(0.1029)
(c) Standardized AFQT	0.0925	−0.0361	0.0881	0.0062
	(0.0143)	(0.0482)	(0.0143)	(0.0572)
(d) Education * experience/10	−0.0539	−0.0952	−0.0665	−0.0889
	(0.0235)	(0.0276)	(0.0234)	(0.0283)
(e) Standardized AFQT * experience/10		0.1407		0.0913
		(0.0514)		(0.0627)
(f) Black * experience/10			−0.2670	−0.1739
			(0.0968)	(0.1184)
R^2	0.3056	0.3063	0.3061	0.3064

Experience is modeled with a cubic polynomial. All equations control for year effects, education interacted with a cubic time trend, Black interacted with a cubic time trend, AFQT interacted with a cubic time trend, two-digit occupation at first job, and urban residence. For these time trends, the base year is 1992. For the model in Panel 1 column (1) the coefficient on AFQT and Black are .0312 and −.1006, respectively, when evaluated for 1983. In Panel 2 the instrumental variables are the corresponding terms involving potential experience and the other variables in the model. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 21,058 observations from 2976 individuals.

Do employers statistically discriminate on the basis of race?

- A statistically discriminating firm might use race along with education to predict the productivity of new workers
- With experience, the productivity of the worker would become more apparent, and compensation would be based on all of the information available rather than just the information at the time of hire
- Hence, if statistical discrimination based on race is important, then adding interactions between t and z variables should make the Black-experience interaction less negative

If firms do not use (or partially use) race as information...

- If race is negatively related to productivity, then the race gap should widen with experience and adding AFQT-experience interaction will reduce the race gap in experience slope

Do employers statistically discriminate on the basis of race?

- Race coefficients in Table 1 not consistent with statistical discrimination
  - Column (3): Black main effect becomes much less negative when Black-experience interaction is added
    - Suggests there is either not much difference in the productivity of black and white men at the time of labor market entry, or that firms do not statistically discriminate much
  - Race gap rises sharply with experience
  - Together, inconsistent with statistical discrimination based on race
- Column (4): Adding AFQT-experience interaction decreases Black-experience interaction
  - Also inconsistent with statistical discrimination based on race
  - One interpretation: employers are obeying the law and not statistically discriminating based on race
  - Paper also discusses alternative explanations

Altonji and Pierret (2001): Table 1

TABLE 1

THE EFFECTS OF STANDARDIZED AFQT AND SCHOOLING ON WAGES

Dependent Variable: Log Wage; OLS estimates (standard errors).

Panel 1—Experience measure: potential experience				
Model:	(1)	(2)	(3)	(4)
(a) Education	0.0586	0.0829	0.0638	0.0785
	(0.0118)	(0.0150)	(0.0120)	(0.0153)
(b) Black	−0.1565	−0.1553	0.0001	−0.0565
	(0.0256)	(0.0256)	(0.0621)	(0.0723)
(c) Standardized AFQT	0.0834	−0.0060	0.0831	0.0221
	(0.0144)	(0.0360)	(0.0144)	(0.0421)
(d) Education * experience/10	−0.0032	−0.0234	−0.0068	−0.0193
	(0.0094)	(0.0123)	(0.0095)	(0.0127)
(e) Standardized AFQT * experience/10		0.0752		0.0515
		(0.0286)		(0.0343)
(f) Black * experience/10			−0.1315	−0.0834
			(0.0482)	(0.0581)
R^2	0.2861	0.2870	0.2870	0.2873

Panel 2—Experience measure: actual experience instrumented by potential experience				
Model:	(1)	(2)	(3)	(4)
(a) Education	0.0836	0.1218	0.0969	0.1170
	(0.0208)	(0.0243)	(0.0206)	(0.0248)
(b) Black	−0.1310	−0.1306	0.0972	0.0178
	(0.0261)	(0.0260)	(0.0851)	(0.1029)
(c) Standardized AFQT	0.0925	−0.0361	0.0881	0.0062
	(0.0143)	(0.0482)	(0.0143)	(0.0572)
(d) Education * experience/10	−0.0539	−0.0952	−0.0665	−0.0889
	(0.0235)	(0.0276)	(0.0234)	(0.0283)
(e) Standardized AFQT * experience/10		0.1407		0.0913
		(0.0514)		(0.0627)
(f) Black * experience/10			−0.2670	−0.1739
			(0.0968)	(0.1184)
R^2	0.3056	0.3063	0.3061	0.3064

Experience is modeled with a cubic polynomial. All equations control for year effects, education interacted with a cubic time trend, Black interacted with a cubic time trend, AFQT interacted with a cubic time trend, two-digit occupation at first job, and urban residence. For these time trends, the base year is 1992. For the model in Panel 1 column (1) the coefficient on AFQT and Black are .0312 and −.1996, respectively, when evaluated for 1983. In Panel 2 the instrumental variables are the corresponding terms involving potential experience and the other variables in the model. Standard errors are White/Huber standard errors computed accounting for the fact that there are multiple observations for each worker. The sample size is 21,058 observations from 2976 individuals.

Overview

- Fryer-Loury (2005): recent overview of affirmative action
  - "Regulations on the allocation of scarce positions in education, employment, or business contracting so as to increase the representation in those positions of people belonging to certain population subgroups"
- Holzer-Neumark (2000)
  - Table 1: key executive orders, regulations, and court decisions regarding affirmative action in the labor market
  - Reviews empirical studies of affirmative action policies

(Selected) Empirics

- Leonard (1984) on Executive Order no. 11246 in 1965
- Chay (1998) on Equal Employment Opportunity Act of 1972
- McCrary (2007) on a series of court-ordered racial hiring quotas in municipal police departments

Affirmative action has been controversial

One key question: can labor market gains from affirmative action policies continue without these policies becoming a permanent fixture in the labor market?

Coate and Loury (1993): how do affirmative action policies impact employers' stereotypes about capabilities of minority workers?

- Break down negative stereotypes: could $\Rightarrow$ permanent gains
- Negative views about minority group are not eroded or are worsened: policy would need to be maintained permanently

Model set-up

- Assume large number of identical employers
- Assume large population of workers, randomly matched
- Workers $\in [B, W]$, where share λ is W

Employers

- Assign workers to job 0 or job 1
- All workers can perform satisfactorily job 0
- Workers differ in qualification for job 1
- Workers earn ω on task 1, 0 on task 0
- Employers earn:
  - $x_q > 0$ for qualified worker on task 1
  - $-x_u < 0$ for unqualified worker on task 1
  - 0 for worker on task 0 (normalization)

Employers

Employers don't observe qualification prior to assignment

- Observe worker identity $\in [B, W]$
- Observe noisy signal $\theta \in [0, 1]$ of worker's qualification (test)
- Distribution of θ depends on qualification, but not group
- $F_q(\theta)$: probability signal does not exceed θ given qualified
- $F_u(\theta)$: probability signal does not exceed θ given unqualified
- $f_q(\theta)$, $f_u(\theta)$: density functions
- $\varphi = \frac{f_u(\theta)}{f_q(\theta)}$: likelihood ratio at θ
- Assume that $\varphi(\cdot)$ is non-increasing on $\theta \in [0, 1]$
  - Implies $F_q(\theta) \leq F_u(\theta)$ for all θ
  - Distribution of the signal for qualified workers first-order stochastically dominates distribution for unqualified workers

Employers

- Employers' assignment policies: thresholds for each group
- Workers are qualified to perform task 1 only if they have made some costly *ex ante* investment
  - c: worker's investment cost
  - $G(c)$: fraction of workers with cost $\leq c$
  - Cost distributions equal across groups

Equilibrium

- Equilibrium is a set of employer beliefs (about workers' qualifications in each group W and B) and workers' investments that are self-confirming
- Discriminatory equilibrium is one in which employers believe that workers from one group are less likely to be qualified

Affirmative action

- Consider an affirmative action policy that mandates that group assignments to task 1 are made at equal rates
- Ask whether introduction of such a constraint is sufficient to induce employers - in the resulting equilibrium - to believe that workers' productivities are uncorrelated with their group identity

Coate and Loury: Conclusions

Coate and Loury conclude by saying their results give “credence to both the hopes of advocates...and the concerns of critics.” There are circumstances under which affirmative action will eliminate negative stereotypes, but equally plausible circumstances under which it fail to do so or even worse stereotypes.

More empirical work on these issues would be useful

1. Preliminaries: Farber and Gibbons (1996)
2. Testing statistical discrimination: Altonji and Pierret (2001)
3. Affirmative action: Coate and Loury (1993)
4. Looking ahead

Looking ahead

Rent-sharing

14.662 Labor Economics II
Spring 2015