Lecture Note 7 — Discrimination and Learning

David Autor
MIT 14.662 Spring 2011

May 11, 2011

1 Distinguishing Statistical from Taste-Based Discrimination: List (2004) [for self-study if interested.]

This study represents an ambitious, multifaceted and a bit complicated attempt to test whether discrimination occurs in a 'well-functioning' marketplace and, if so, to evaluate whether that discrimination is 'taste-based' or statistical (or a combination of both). The setting of the study is a sportscard trading market. Sportscard trading is a popular avocation, and apparently one on which List spends a great deal of time.

In Part I of the experiment, List recruits volunteers at a sportscard show to buy (from dealers) and sell (to dealers) a 1989 *Upper Deck* Ken Griffy Jr. PSA graded "9" baseball card. Apparently, this is a valuable commodity; in the experiment, subjects typically paid over $100 for this card when *buying* from dealers. When *selling* the card to dealers, they typically received about $30 (so, there is quite a large buy-sell spread).

There are many details to the experiment that we will not summarize. The main results of the initial experiment are evident in Table II through Table VI:

1. Initial offers made by dealers to minorities (nonwhites, females and men over age 60) for transacting on the Ken Griffy card are inferior to initial offers to white males. Inferior means *high asking price* when the subject is buying from the dealer and *low offer price* when the subject is selling to the dealer.

2. Discrimination appears much greater in the treatment in which subjects are selling the card to the dealer than in which they are buying (no theory for this).

3. Final offers to minorities are not *as inferior* as initial offers to minorities (relative to white males).

4. But, minorities spend more time bargaining to achieve similar results to white male—suggesting that they have to expend resources to overcome discrimination.

5. Experienced buyers/sellers of minority groups do about as well as white males, but only after having spent considerably more time bargaining.

6. Experienced dealers discriminate *more* than inexperienced dealers.

Thus, discrimination in this market is *amply* evident, and it is greater among more experienced dealers. But what is the nature of this discrimination? List distinguishes three possibilities: animus-based discrimination; statistical discrimination; and differences in bargaining ability.

1.1 The dictator game

The dictator game is a widely-used laboratory experiment. In this particular version, List gives dealers envelopes with five $1 dollar bills and informs the dealer the race/gender of the person to whom he or she is randomly paired (white male age 20-30 WM, white female age 20-30 WF, nonwhite male age 20-30 NMM, white male age 60+ WMM). The dealer anonymously decides how much of the money to keep (take out of the envelope) and how much to leave for the anonymous partner whom s/he will never (knowingly) meet.

Figure I summarizes the results of this experiment. Dictators seem to favor white females. White females are significantly less likely to receive zero dollars and more likely to receive two or three dollars. There is no race differences (nonwhite females were not tested, presumably because there were almost no nonwhite females active in this market). Notice that the pattern of discrimination *favoring* females is opposite to the trading patterns on the floor where females receive worse offers uniformly.

One may object that this dictator game is highly artificial and so the behavior observed here may not match what a dealer would do in a less artificial setting. This particular experiment probably has limited external validity.

1.2 The 'Chamberlain' market

This is a subtle designed to evaluate whether dealers *believe* that minorities are more or less effective at negotiating than nonminorities, and whether this belief explains why dealers make less favorable offers to minorities.

To evaluate, List sets up a market in which dealers and nondealers trade for 'customized' (i.e., defaced) baseball cards (so they have no outside market value) using real money. Dealers and non-dealers are each randomly assigned a reservation value for the card (so, the dealer may be willing to pay no more than \$20 and the seller may be willing to accept no less than \$15). The participants bargain over the price and then each gets to keep his or her surplus (so, in the numerical example, there is \$5 surplus to allocate; if the card sells for \$19, the seller gets one dollar in surplus and the buyer gets \$1 in surplus).

The key manipulation in this study is this: in some experiments, the dealers *are told* that the sellers' reservation values are randomly assigned; in others, they are *not told*.

There are many possible predictions for this experimental setting depending on the underlying model. Let's just focus on two.

1. In the animus-based case, we'd expect minorities to fare worse under both the 'informed' and 'uninformed' cases, since dealers should bargain harder with minorities due to animus.

2. In the case of statistical discrimination, we should expect dealers to bargain harder with minorities if the dealers *do not* know that reservation values are randomly assigned; if they do know that reservation values are randomly assigned, they should treat minority and nonminority sellers similarly. This is particularly true if dealers *knowingly* statistically discriminate, since they should consciously 'shut down' their discriminatory behavior when it is not rational to discriminate (because reservation values are randomly assigned).

Main results:

1. Majority buyers outperform minority buyers when dealers *do not* know that reservation values are determined randomly. This is consistent with either animus-based or statistical discrimination.

2. Majority buyers perform similarly to minority buyers when dealers *do* know that reservation values are determined randomly.

These results suggest that dealers discriminate only when they do not know that minority buyers have randomly assigned reservation values. This suggests that dealers engage in statistical discrimination; dealers believe that minorities are willing to accept less money than nonminorities for similar items.

1.3 Reservation value experiments

Is statistical discrimination rational? That is, should profit-maximizing dealer's rationally discriminate between minority and majority buyers? This will depend on the distribution of reservation values for buying and selling among these groups. Dealers appear to be acting as if these reservation values differ, but we do not know at present if their intuitions are correct. (Statistical discrimination with bad statistics is just discrimination.) The idea here is to directly test whether (a) minorities have a different distribution of reservation values from non-minorities; and (b) dealers believe this to be true.

One can motivate the relevance of the shape of the reservation value distribution by setting up a problem where a buyer has private reservation value $f(w)$ over the object and the seller must make a final take if or leave it price offer to the buyer. The seller's job is to propose a price that maximizes the product of profit conditional on sale times the probability of sale:

$$\max_p \quad (p-c)(1-F(p)),$$

where c is seller's marginal cost of the object. The FOC of this problem is:

$$\begin{aligned} 0 &= (1-F(p)) - (p-c) f(p) \\ p^* &= c + \frac{(1-F(p))}{f(p)}. \end{aligned}$$

As $f(p)$ decreases (representing a less dense distribution), the optimal offer price rises. (Seems like this will only be true for well-behaved distributions—otherwise, other moments will come into play).

The nice set of findings here is that minority reservation values are more *dispersed* than non-minority reservation values, meaning that all else equal, minorities are more likely to have low reservation values (even if on average, their valuations are similar to nonminorities). (Figures II and III are pretty compelling.) This makes it rational for dealers to make them less favorable offers and to bargain harder with them.

A final experiment suggests that dealers know that these reservation value distributions differ. In particular, dealers—especially experienced dealers—are able to guess more accurately than chance would predict which distribution belongs to which race and gender group.

1.4 Conclusions of List study

This study offers is a clever and rigorous effort to evaluate whether discrimination exists in the marketplace and from where it arises.

1. Minorities receive worse offers in the sportscard market
2. There is not strong evidence of dealer animus against minorities
3. When told that reservation values of participants are set randomly, dealers treat minorities/majorities similarly
4. When *not* told that reservation values are set randomly, dealers make worse offers to minorities
5. Minorities have a more dispersed distribution of willingness to pay and willingness to accept
6. Dealers appear to recognize this

The evidence offered here strongly suggests that *statistical discrimination* in the sportscard market is largely responsible for differential bargaining behavior of dealers facing minority versus non-minority buyers. The internal validity of these conclusions looks solid. What about the external validity?

2 Learning models

Statistical models of discrimination are, as a rule, static. They consider setting where employers have a set of beliefs that are rational given the available information. These beliefs represent averages over populations and these averages are applied to individuals. These models do not typically ask what happens as the employer's information set changes. Change may occur for a variety of reasons. In the Autor-Scarborough paper, it occurs because of an improvement in the employer's information set. In the papers by Farber and Gibbons and by Altonji and Pierret, it occurs because employers learn about workers' productivity over the course of the job (and, more generally, over the course of the career). The F&G and A&P papers test rich models at the cost of using representative data that does not link workers to jobs. The A&S paper tests a less sophisticated model in a setting where there is arguably a fair degree of power to evaluate the hypotheses of interest.

2.1 Autor and Scarborough (2008)

- It is widely believed that there is an equity-efficiency tradeoff in job testing. Use of employment tests (such as IQ tests) could raise the productivity of job matches but would reduce opportunities for minority workers, who tend to score lower on these tests. (The Black-White test score gap on IQ tests such as the AFQT is about 1 full standard deviation, which is very large. This observation was the subject of the famous Hernstein-Murray book, *The Bell Curve*.)

- This E-E tradeoff viewpoint is well expressed by the quotation from the Hartigan and Wigdor volume on fairness in job testing given in the paper.

- The A-S paper calls into question the EE tradeoff notion. It argues that the case for a trade-off between equality and efficiency in the use of job testing is not well-established empirically or well-grounded conceptually.

"We start from the presumption that competitive employers face a strong incentive to assess worker productivity accurately, but such assessments are inevitably imperfect. In our discussion and conceptual model, we consider two distinct—and not mutually exclusive—channels by which job testing may affect worker assessment. The first is to raise the precision of screening, which occurs if testing improves the accuracy of firms' assessments of applicant productivity. A large body of research demonstrates the efficacy of job testing for improving precision, so we view this channel as well-established.The second is to 'change beliefs'—that is, to introduce information that systematically deviates from firms' assessments of applicant productivity based on informal interviews. This occurs if either the job test is biased *or* if the informal screen that precedes it is biased—or, potentially, if both are biased, albeit differently.

"To see the relevance of these distinctions, consider a firm that is initially screening informally for worker productivity and which introduces a formal job test that improves the precision of screening. Assuming that minority applicants perform significantly worse than majority applicants on this test, will the gain in screening precision come at a cost of reduced minority hiring? As we show below, the answer will generally be *no* if both the informal screen and the formal test provide *unbiased* measures of applicant productivity. In this case, the main effect of testing will be to raise the precision of screening within each applicant group; shifts in hiring for or against minority applicants are likely to be small and will favor minorities. Notably, this result does not require that both the test and informal screen are unbiased. Our model below suggests that the harm or benefit to minority workers from testing depends primarily on the *relative* biases of the formal and informal screens. So long as the information provided by job tests about minority applicants is not systematically more negative than firms' beliefs derived from informal screens, job testing has the potential to raise productivity without a disparate impact on minority hiring. This result makes it immediately apparent why the presumption that job testing will harm minority workers is suspect: there is little reason to expect that job testing is *more* minority-biased than informal hiring practices."

2.2 Model sketch

- The model in the paper lays out a simple, normal statistical discrimination model of the kind you are now intimately familiar. This is a threshold based hiring model, so decisions are up/down.

- The 'novelty' of the model is to work out what happens when a new source of information is added to the employer's information set.

- Formally, the paper views the job interview as one type of test and the personality test as a second. The paper works out what happens when this second test is added.

- In the normal case, one can decompose the effect of testing on employer's posteriors over worker ability into two effects: a precision effect and a mean shift effect.

- The precision effect simply depends on the added value of the test to raise signal relative to noise variance in the posterior.

- The bias effect depends on the *relative bias* of the test—that is the mean discrepancy between the test and the interview. A key observation is that it is *not* the absolute bias of the test that determines whether the test has an adverse impact on minority hiring. Rather, it's the relative bias of the test (relative to the interview, that is). If the test and interview share the same bias (perhaps they are identically biased or both are unbiased), then testing merely improves precision.

- The main, parametric conclusions of the model are:

  1. If job tests are relatively unbiased, they do not pose an equality-efficiency trade-off;
  2. If job tests are bias-reducing, they pose an equality-efficiency trade-off if and only if interviews are minority-favoring
  3. If job tests are bias-enhancing, they may pose an equality-efficiency trade-off—or they may simply reduce equality and efficiency simultaneously.

- More generally (not relying on normality):

  1. The potential effects of job testing on minority hiring depend primarily on the biases of job tests *relative to* job interviews (and other existing screening methods). Job tests that are unbiased relative to job interviews are unlikely to reduce minority hiring because such tests do not adversely affect firms' average assessments of minority productivity.
  2. Testing is likely to reduce minority hiring when tests are relatively biased against minorities (i.e., relative to interviews). In such cases, testing conveys 'bad news' about the productivity of minority relative to majority applicants and so is likely to adversely affect minority hiring. Nevertheless, if testing mitigates existing biases, it will still be efficiency-enhancing, and so an equality-efficiency trade-off will be present. If instead testing augments bias, it may be efficiency-reducing.
  3. Testing will generally have opposite effects on the hiring and productivity gaps between majority and minority workers; a test that reduces minority hiring will typically differentially raise minority productivity.

- The empirical section of the paper explores what happened when one, large employer (with about ~1,400 stores) introduced job testing. Short answer: minority hiring was unaffected; productivity of Blacks and Whites rose significantly and by the same increment.

2.3 Environment

- There are many firms facing numerous job applicants from two identifiable demographic groups, x_1 and x_2, corresponding to a majority and minority group. For simplicity, assume that each group comprises half of the applicant population (thus, 'minority' refers to historical circumstances rather than population frequency).

- The ability (Y) of job candidates is distributed as

$$Y \sim N\left(\mu_0\left(x\right), 1/h_0\right).$$

  The mean parameter $\mu_0\left(x\right)$ may depend on x. Assume that h_0, equal to the inverse of the population variance σ_0^2, is constant, independent of x.[1]

- Let the ability of each applicant, y, be a random draw from the population distribution for the relevant demographic group (x_1 or x_2). The firm treats the population parameters as known. Thus, the firm's prior distribution for a draw y is the population distribution.

- Firms have a linear, constant returns to scale production technology and are risk neutral.

- Workers produce output, $f\left(y\right) = y$. Hence, ability and productivity are synonymous. Job spell durations are independent of y and wages are fixed,[2] so firms strictly prefer to hire more productive workers.

[1] The assumption that σ_0^2 is independent of x stands in contrast to several models of statistical discrimination in which testing is differentially informative (or uninformative) for minority groups due to their higher (lower) underlying productivity variance, e.g., Aigner and Cain [1977], Lundberg and Startz [1984], and Masters [2006]. We believe that the evidence supports our assumption. Analysis in Hartigan and Wigdor [1989], Wigdor and Green [1991] and Jencks and Philips [1989, chapter 2] all suggest that while tests commonly used for employee selection show marked mean differences by race, the by-race variances are comparable and, moreover, these tests are about equally predictive of job performance for minorities and non-minorities. As shown in Figure II and Table II, *mean* test scores in our sample also differ significantly among White, Black and Hispanic applicant groups but the variances of test scores are nearly identical for all three groups.

[2] As above, the majority of line workers at the establishments we study are paid the minimum wage.

- Job applicants are drawn at random from the pooled distribution of x_1 and x_2 workers. Firms hire applicants using a screening threshold where applicants whose expected productivity exceeds a specified value are hired.

- In a fully elaborated search framework, this screening threshold would depend on technology and labor market conditions. In our reduced form setup, the screening threshold is chosen so that the aggregate hiring rate is held constant at $K \in (0, 0.5)$. This simplification focuses our analysis on the first-order impacts of job testing on the distribution of hiring across demographic groups, holding total employment fixed. We additionally assume that the hiring rate of each demographic group is below 50 percent, so selection is from the right-hand tail of each applicant distribution.

- Initially, applicants are screened using interviews. Each interview generates an *interview signal,* η. When testing is introduced, applicants are screened using both interviews and tests. The test score is denoted by s.

- *Suppose that there is no bias in interviews.* Then the distribution of interview signals will be centered on the true productivity of each applicant. Precisely,

$$\eta \sim N\left(y, 1/h_\eta\right), \tag{1}$$

  where h_η is the inverse of the variance of the interview signal (a measure of accuracy of the interview). Assume h_η does not depend on x.

- Conditional on perceived productivity $\mu_0\left(x\right)$ for group x and the interview signal η, the firm updates its assessment of the expected productivity of the applicant:

$$m\left(x, \eta\right) \equiv y|_{x,\eta} \sim N\left(\mu\left(x, \eta\right), 1/h_I\right), \tag{2}$$

  where the updated degree of precision equals $h_I \equiv h_\eta + h_0$, and the updated mean equals $\mu\left(x, \eta\right) \equiv \left[\eta h_\eta + \mu_0\left(x\right) h_0\right] / h_I$.

- *Suppose that there is no bias in testing.* Then the distribution of test signals will be centered on the true productivity of each applicant. Precisely,

$$s \sim N(y, 1/h_S), \tag{3}$$

where h_S is the inverse of the variance of the interview signal (a measure of accuracy of the interview). Assume h_S does not depend on x.

- This generates a *posterior* for the firm's perception of the applicant's productivity

$$m(x, \eta, s) \equiv y|_{x,\eta,s} \sim N(\mu(x, \eta, s), 1/h_T), \tag{4}$$

where the degree of accuracy for the posterior (based on both testing and interviews) is $h_T \equiv h_S + h_I$; and the updated mean equals $\mu(x, \eta, s) \equiv [sh_S + \mu(x, \eta)h_I]/h_T$. Note that $h_T > h_I$.

2.4 First outcome of interest: Hiring rates

To assess when testing poses an *equality-efficiency trade-off*, we study two outcomes. The first is the *hiring gap*, defined as the hiring rate of majority workers minus the hiring rate of minority workers.

- Denote the hiring decision as $Hire = 0, 1$ for the firm. If there is no testing, the hiring decision will completely depend upon the firm's prior and the results of interviews: $Hire = I\{\mu(x, \eta) > \kappa_I\}$, where κ_I is the screening threshold that yields a total hiring rate of K using interviews and $I\{\cdot\}$ is the indicator function.

- The expected hiring rate of group x applicants who have received the interview is

$$E_\eta[Hire|x] = 1 - \Phi(z_I(x)),$$

where $z_I(x) \equiv [\kappa_I - \mu_0(x)]/\sigma_0\rho_I$ and $\rho_I \equiv Corr[\mu(x, \eta), y] = (1 - h_0/h_I)^{1/2}$.

- Note that we iterate expectations over η to obtain the *unconditional* hiring rate (i.e., not conditional on a specific value of η) for group x applicants based on interviews. Specifically, $E_\eta[Hire|x] = \int E[Hire|x, \eta] f(\eta|x)\, d\eta$.[3]

- If both testing and interviews are used, the hiring decision is $Hire = I\{\mu(x, \eta, s) > \kappa_T\}$, where κ_T is the screening threshold that yields a total hiring rate of K using both

[3] Since η is normally distributed and assessed productivity conditional on η is normally distributed, the unconditional distribution of perceived productivity is also normally distributed. It can be shown that the variance of the unconditional distribution is $V_{\eta,y}(\mu(x, \eta)) = \rho_I^2\sigma_0^2$.

interviews and test scores. The expected hiring rate of group x applicants who have received the interview and the test is:[4]

$$E_{\eta,s}[Hire|x] = 1 - \Phi(z_T(x))$$

where $z_T(x) \equiv [\kappa_T - \mu_0(x)]/\sigma_0\rho_T$ and $\rho_T \equiv Corr[\mu(x,\eta,s), y] = (1 - h_0/h_T)^{1/2}$.

- When hiring is based on interviews, the *hiring gap* between majority and minority workers is

$$\gamma_I = E_\eta[Hire|x_1] - E_\eta[Hire|x_2].$$

- When hiring is based on testing and interviews, this gap is

$$\gamma_T = E_{\eta,s}[Hire|x_1] - E_{\eta,s}[Hire|x_2].$$

- We denote the effect of testing on the hiring gap as $\Delta\gamma \equiv \gamma_T - \gamma_I$.

2.5 Second outcome of interest: Productivity

- A second outcome of interest is the effect of testing on productivity. If only interviews are used, the mean productivity for hired workers of group x is

$$E_\eta[y|Hire = 1, x] = \mu_0(x) + \sigma_0\rho_I\lambda(z_I(x)), \tag{5}$$

where $\lambda(z_I)$ is the inverse Mills ratio $\phi(z_I)/[1 - \Phi(z_I)]$, equal to the density over the distribution function of the standard normal distribution evaluated at z_I.

- If both tests and interviews are used, the mean productivity for hired workers of group x is

$$E_{\eta,s}[y|Hire = 1, x] = \mu_0(x) + \sigma_0\rho_T\lambda(z_T(x)). \tag{6}$$

- A comparison of equations (5) and (6) shows that testing affects the productivity of hired applicants through two channels: selectivity (equal to one minus the hiring rate) and

[4] We iterate expectations over η and s to obtain the unconditional hiring rate for group x applicants based on interviews and tests. It can be shown that the variance of the unconditional distribution is $V_{s,\eta,y}(\mu(x,\eta)) = \rho_T^2\sigma_0^2$.

screening precision. All else equal, a rise in selectivity (i.e., a reduction in hiring) for group x raises the expected productivity of workers hired from group x by truncating the lower-tail of the group x productivity distribution. Screening precision refers to the accuracy of the firm's posterior, and its effect is seen in the terms ρ_I and ρ_T in equations (5) and (6), with $\rho_T > \rho_I$ (more precisely, both ρ_I and ρ_T are increasing functions of screening precision, so $h_T > h_I$ implies that $\rho_T > \rho_I$).

- All else equal, a rise in screening precision improves the accuracy of firms' assessments of worker productivity and so raises the quality of hires from each demographic group.

- In addition to the impact of testing on overall productivity levels, we also study its effect on the *productivity gap*, defined as the mean productivity of majority workers minus the mean productivity of minority workers. This gap proves relevant to our empirical work because our model suggests that testing typically moves the hiring and productivity gaps in opposite directions.

- When hiring is based on interviews, the majority/minority productivity gap is

$$\pi_I = E_\eta\left[y|Hire = 1, x_1\right] - E_\eta[y|Hire = 1, x_2].$$

- When hiring is based on interviews and tests, this gap is

$$\pi_T = E_{\eta,s}\left[y|Hire = 1, x_1\right] - E_{\eta,s}[y|Hire = 1, x_2].$$

- We denote the effect of testing on the productivity gap as $\Delta\pi \equiv \pi_T - \pi_I$.

2.6 The effects of testing when both interviews and tests are unbiased

- The potential for an equality-efficiency trade-off is relevant when one applicant group is less productive than the other. For concreteness, and without loss of generality, suppose that minorities are the less productive applicant group ($\mu_0(x_2) < \mu_0(x_1)$).

- These underlying population productivity differences imply observable differences in the hiring and productivity of minority and majority workers *prior to use of tests*. First, the hiring rate of minority applicants based on interviews will be lower than that of majority applicants ($\gamma_I > 0$). Second, minority workers hired using interviews will be on average less productive than majority workers ($\pi_I > 0$).

- Both inequalities follow from the firm's threshold hiring policy wherein applicants whose assessed productivity (equation (2)) exceeds a reservation value κ_I are hired.[5] This observation is significant for our empirical work because, as shown in Table I, minority workers hired using interviews are *less productive,* as measured by job tenure, than are majority workers hired using interviews.

- To derive the effect of testing on the *hiring gap*, we note that the overall hiring rate in the model is constant at K. Hence, testing must either leave hiring of both groups unaffected or change the hiring rate of each group by equal but opposite amounts. It is straightforward to show by differentiation that: (1) it is not possible for hiring of both groups to be unaffected; and (2) testing *raises* minority hiring or, more generally, raises hiring of the applicant group with lower average productivity (see proof in Appendix):

$$\Delta\gamma < 0.$$

- Intuitively, because the interview signal is error-ridden ($1/h_\eta > 0$) and expected majority applicant productivity exceeds expected minority applicant productivity, firms disproportionately hire applicants from the group favored by their prior—that is, majority applicants. Testing increases minority hiring because the posterior including the test score places more weight on observed signals and less weight on group means. However, simulations show that the effect of testing on the majority/minority hiring gap is typically small under the assumed normality of the productivity distributions. We therefore do not generally expect testing to induce a substantial change in minority hiring.

- We obtain a similar, but stronger, result for the effect of testing on the majority/minority *productivity gap*: although minority workers hired using interviews are less productive

[5] The hiring rule ($Hire = I\{\mu(x,\eta) > \kappa_I\}$) equates the expected productivity of *marginal* hires from each applicant group. Because the *average* majority *applicant* is more productive than the average minority applicant, the average majority *hire* is also more productive than the average minority hire. As a referee pointed out, this result stems from the fact that the normal distribution is thin-tailed.

than majority workers hired using interviews, testing leaves this majority/minority productivity gap essentially unaffected. More precisely, *testing raises productivity of both minority and majority hires approximately equally, with exact equality as selectivity approaches one* (see proof in Appendix). We write:

$$\Delta\pi \approx 0.$$

- The intuition for this result stems from two sources: first, the threshold hiring rule equates the productivity of marginal minority and majority hires both before and after the introduction of testing; second, when selection is from the right-hand tail of the normal distribution, the truncated mean increases near-linearly with the point of truncation with a first derivative that is asymptotically equal to unity.[6] Consequently, a rise in screening precision raises the marginal and average productivity of hires almost identically for minority and majority workers.

- Summarizing, *if both interviews and job tests are unbiased, testing does not pose an equality-efficiency trade-off.* Although job tests unambiguously raise productivity, the gains come exclusively from improved selection within each applicant group, not from hiring shifts against minorities.\

- These results are illustrated in Figure IIIa, which provides a numerical simulation of the impact of testing on hiring and productivity for a benchmark case where majority applicants are on average more productive than minority applicants and job interviews and job tests are both unbiased. The x-axis of the figure corresponds to the correlation between test scores and applicant ability ($\mathrm{Corr}\langle s, y\rangle = 1/\left(1 + h_0/h_s\right)^{1/2}$), which is rising in test precision. The y-axis depicts the hiring rate of majority and minority applicants (left-hand scale) and the expected productivity (equivalently, ability) of majority and minority hires gap (right-hand scale).[7]

[6] Numerical simulations of the normal selection model show that this asymptotic equality is numerically indistinguishable from exact equality at selectivity levels at or above +0.1 standard deviation from the mean (i.e. $z_I, z_T \geq 0.1$). This result is also visible in the numerical simulation in Figure IIIa, where the productivity gap between minority and majority hires is invariant to testing. Recall from Table II that the overall hiring rate at this firm is 8.95 percent, implying that $z_I, z_T \approx 1.34$.

[7] In the simulation, the ability (equivalently productivity) of nonminority applicants is distributed $N(0, 0.29)$,

- Prior to the introduction of testing—equivalently, $\rho = 0$ in the figure—minority applicants are substantially less likely than majority applicants to be hired and are also less productive than majority workers conditional on hire. Job testing slightly reduces the minority/majority hiring gap. But this effect is small relative to the initial gap in hiring rates, even at maximal test precision. By contrast, testing leads to a substantial rise in the productivity of both minority and majority hires, with the degree of improvement increasing in test precision. Consistent with the analytic results, testing has no detectable effect on the majority/minority productivity gap at any level of test precision.

2.7 The effects of testing when interviews and tests are biased: The case of identical biases

Our main result so far is that use of an unbiased test introduced in an unbiased hiring environment raises productivity without posing an equality-efficiency trade-off. We now consider how test and interview biases affect this conclusion.

Suppose there is a *mean bias in interviews.* So, change equation (1) to

$$\eta^* \sim N\left(y + \nu_\eta(x), 1/h\right)$$

where $\nu_\eta(x_1) \neq \nu_\eta(x_2)$. We say that job interviews are *minority favoring* if $\nu_\eta(x_2) > \nu_\eta(x_1)$, and *majority favoring* if $\nu_\eta(x_1) > \nu_\eta(x_2)$). For example, managers may perceive majority applicants as more productive than equally capable minority applicants, or vice versa.[8]

Similarly, suppose there is a *mean bias in job tests.* So, change equation (3) to

$$s^* \sim N\left(y + \nu_s(x), 1/h_S\right)$$

where $\nu_s(x_1) \neq \nu_s(x_2)$, with the definition of minority favoring and majority favoring tests analogous to that for interviews. This might arise if tests are 'culturally biased' so that for given applicant ability, minority applicants score systematically below majority applicants.

the productivity of minority applicants is distributed $N(-0.19, 0.27)$, the precision of the informal ability signal is 1/0.45, and 8.95 percent of applicants are hired. Thus, $h_0 = 1/0.27$, $h_\eta = 1/0.45$ $\mu_0(x_1) = 0$, $\mu_0(x_2) = -0.19$ and $K = 0.0895$. These values are chosen to match estimates from the parametric model simulation in Section **??** of the paper. The precision of the job test ranges from 1/10,000 to 1/0.0001, corresponding to a correlation of $(0.0, 1)$ between test scores and applicant ability (plotted on the x-axis).

[8] Equivalently, $\Delta\nu_\eta$ could be interpreted as taste-discrimination: firms' reservation productivity for minority and majority hires differs by $\Delta\nu_\eta$.

Define the net bias of interviews as $\Delta\nu_\eta = \nu_\eta(x_1) - \nu_\eta(x_2)$ and, similarly, the net bias of tests as $\Delta\nu_s = \nu_s(x_1) - \nu_s(x_2)$. If $\Delta\nu_\eta > 0$, interviews favor majority applicants, and vice versa if $\Delta\nu_\eta < 0$ (and similarly for job tests). We refer to the difference in bias between tests and interviews $(\Delta\nu_s - \Delta\nu_\eta)$ as the 'relative bias' of tests.

Assume that firms' updated assessments of applicant productivity (based on interviews) and posteriors (based on interviews and tests) are still given by equations (2) and (4) except that we now substitute η^* and s^* for η and s. For consistency, suppose that firms' prior for each draw from the applicant distribution is mean-consistent with the information given by interviews, as in the unbiased case: $y|_x \sim N(\mu_0(x) + \nu_\eta(x), 1/h_0)$. Thus, firms do not compensate for biases in interviews or tests and we say that their *perceived* productivity of the applicant distribution is equal to true productivity plus interview bias.

How do these biases affect our prior results for the impact of testing on equality and efficiency? Suppose initially that *interviews and tests are equally biased*—that is, both tests and interviews contain biases but these biases are identical $(\Delta\nu_s = \Delta\nu_\eta \neq 0)$. In this no relative bias case, our prior results require only slight modification:

1. Use of tests that are unbiased *relative to* job interviews does not pose an equality-efficiency trade-off. In particular: (1) testing raises hiring of the applicant group with lower *perceived* productivity, $\Delta\mu + \Delta\nu_\eta$ (the minority group by assumption); and (2) testing raises productivity of both minority and majority hires approximately equally, with exact equality as selectivity approaches one. Thus, expanding on our earlier conclusion: unbiasedness of both interviews and tests $(\Delta\nu_s = \Delta\nu_\eta = 0)$ is a sufficient but not a necessary condition for the no-trade-off result to hold. If both interviews and tests are equally biased $(\Delta\nu_s = \Delta\nu_\eta)$—thus, there is no *relative bias*—testing does not pose an equality-efficiency trade-off.

2. We showed above that if both interviews and tests are unbiased, the applicant group with lower average productivity will have a lower hiring rate and lower productivity conditional on hire than the group with higher average productivity $(\text{Sign}\langle\gamma_I\rangle = \text{Sign}\langle\pi_I\rangle)$. Interview and testing biases can reverse this positive correlation. Because biases reduce selectivity of the favored group and raise selectivity of the non-favored group, it is possible for the

group with a greater hiring rate to have lower productivity conditional on hire.[9] So, if minority hires are observed to be less productive than majority hires, this implies that either minority applicants have lower mean productivity than majority applicants (i.e., $\mu_0(x_2) < \mu_0(x_1)$) or that job interviews are minority-favoring ($\Delta\nu_\eta < 0$) or both.

2.8 The effects of testing when interviews and tests have *non*-identical biases

We finally consider how job testing affects the productivity and hiring gaps when the test is biased *relative to* job interviews (i.e., $\Delta\nu_s \neq \Delta\nu_\eta$). For concreteness, we continue to assume that minority applicants are perceived as less productive than majority applicants: $\mu_0(x_1)+\nu_\eta(x_1) > \mu_0(x_2)+\nu_\eta(x_2)$. It is straightforward to establish the following three results:

1. Use of a job test that is biased relative to interviews: (1) raises the hiring rate of minorities if the test favors minorities (i.e., relative to interviews) but has ambiguous effects on minority hiring otherwise; and (2) reduces the productivity level of the group favored by the test relative to the group that is unfavored. For example, if minority applicants are perceived as less productive than majority applicants, use of a relatively minority-favoring test will raise minority hiring and reduce the productivity of minority relative to majority hires (thus, $\Delta\gamma < 0, \Delta\pi > 0$).

2. If the job test is *bias-reducing*—that is, if the test is less biased than are job interviews (formally, $\Delta\nu_\eta > \Delta\nu_s \geq 0$ or $0 \geq \Delta\nu_s > \Delta\nu_\eta$)—it unambiguously raises productivity.[10] Intuitively, a bias-reducing test improves hiring through two channels: (1) raising screening precision and (2) reducing excess hiring of the group favored by interviews (thus, increasing selectivity for this group). Both effects are productivity-enhancing.

3. By contrast, a *bias-increasing* test ($|\Delta\nu_s| > |\Delta\nu_\eta|$) has ambiguous effects on productivity. Although testing always raises screening precision—which is productivity-enhancing—a bias-increasing test causes excess hiring of the group that is favored by the bias, which

[9] This result requires only that the absolute level of bias is sufficient to offset underlying mean majority/minority productivity differences, which can occur even if there is no relative bias in tests.

[10] However, if the test and interview have biases of *opposite* sign ($\text{Sign}\langle\Delta\nu_s\rangle = \text{Sign}\langle\Delta\nu_\eta\rangle$), testing does not necessarily increase productivity even if job tests are *less* biased than interviews.

is productivity-reducing. The net effect depends on the gains from increased screening precision relative to the losses from increased bias.

Figure IIIb illustrates result (2). Here, we simulate the impact of testing on hiring and productivity for a case where minority applicants are less productive than majority applicants and job interviews are minority-favoring.[11] Prior to testing (equivalently, $\rho = 0$ in the figure), the majority/minority hiring gap is small and the majority/minority productivity gap is large relative to a setting with no biases (Figure IIIa). This contrast with Figure IIIa reflects the fact that a minority-favoring interview raises minority hiring and reduces minority productivity. Job testing counteracts this bias, leading to a marked decline in the hiring of minority applicants and an equally marked decline in the productivity gap between majority and minority hires (with the magnitude depending upon test precision).[12] Thus, an unbiased job test increases efficiency at the expense of equality if job interviews are biased in favor of minorities.

Summarizing our three main conclusions: if job tests are relatively unbiased, they do not pose an equality-efficiency trade-off; if job tests are bias-reducing, they pose an equality-efficiency trade-off if and only if interviews are minority-favoring; if job tests are bias-enhancing, they may pose an equality-efficiency trade-off—or they may simply reduce equality and efficiency simultaneously.

2.9 Empirical implications

Our illustrative model contains many specific—albeit, we believe reasonable—assumptions and so it is unwise to generalize too broadly based on this analysis. In fact, a key purpose of the conceptual framework is to demonstrate that, contrary to an influential line of reasoning, job testing does not pose an *intrinsic* equality-efficiency trade-off, even if minority applicants perform worse than majority applicants on job tests.

Beyond this observation, three general conclusions are warranted. First, the potential effects of job testing on minority hiring depend primarily on the biases of job tests *relative to* job interviews (and other existing screening methods). Job tests that are unbiased relative to job

[11] We use the same parameter values as in Figure IIIa except that we now assume that $\Delta\nu_\eta = \mu_0(x_2) - \mu_0(x_1) = -0.19$.

[12] In the limiting case where job tests are fully informative, the unbiased and biased-interview cases converge to the same hiring rates and productivity levels.

interviews are unlikely to reduce minority hiring because such tests do not adversely affect firms' average assessments of minority productivity.

Second, testing is likely to reduce minority hiring when tests are relatively biased against minorities (i.e., relative to interviews). In such cases, testing conveys 'bad news' about the productivity of minority relative to majority applicants and so is likely to adversely affect minority hiring. Nevertheless, if testing mitigates existing biases, it will still be efficiency-enhancing, and so an equality-efficiency trade-off will be present. If instead testing augments bias, it may be efficiency-reducing.

Finally, testing will generally have opposite effects on the hiring and productivity gaps between majority and minority workers; a test that reduces minority hiring will typically differentially raise minority productivity. This implication proves particularly useful for our empirical analysis.

Below, we use this model to interpret the empirical findings in light of their implications for the relative biases of the job test and the informal screen that preceded it. To make this interpretation rigorous, we parametrically simulate the model in section **??** using observed applicant, hiring and productivity data to calculate a benchmark for the potential impacts of job testing on the majority/minority hiring and productivity gaps under alternative bias scenarios.

- One unusual thing about the paper is its 'parametric simulation' of the model to determine what the results imply about the 'state of the world' prior to the introduction of job testing. The procedure here is a bit complicated, but maybe worth discussing (or maybe not)... Here's how the paper describes it: "Drawing on the applicant, hiring and productivity databases summarized in Tables I and II, we parametrically simulate the model to assess what combinations of interview bias, test bias, and underlying majority/minority productivity differences are most consistent with the findings. One overriding conclusion emerges from this exercise: the data readily accept the hypothesis that both job tests and job interviews are unbiased and that the average productivity of White applicants exceeds that of Black applicants. By contrast, the plausible alternatives that we consider—most significantly, that the job test is relatively biased against minorities—are rejected."

2.9.1 Simulation procedure

Let observed job spell durations, D, be a linear function of applicant ability y , with $D = \alpha + \vartheta y$, where $\vartheta > 0$ is a parameter to be estimated from the data. Suppose that the ability of an applicant drawn at random from the distribution of group x applicants is equal to $y = \mu_0(x) + \varepsilon_0$. Prior to the introduction of job testing, firms have access to an *interview signal*, η, for each applicant that is correlated with ability. When job testing is introduced, it provides a second signal, s, that is also correlated with ability. We assume initially that both interviews and tests are unbiased, with $\eta = y + \varepsilon_\eta$ and $s = y + \varepsilon_s$. In these expressions, ε_0, ε_s and ε_η are mean-zero error terms that are normally and independently distributed, with variances to be estimated from the data.

To estimate the variance parameters, we use the following empirical moments: the mean test score of applicants is normalized to zero and the mean test score of workers hired using the test is 0.707 (Table II); the variance of test scores is normalized at one (hence, $1 = \sigma_0^2 + \sigma_s^2$);[13] the observed hiring rate is equal to 8.95 percent; and the average gain in productivity from testing is 21.8 days (Table IV). We make a further adjustment for the fact that the observed hiring rate is only 22 percent at the 95th percentile of the score distribution (see Table II), implying either that stores are extraordinarily selective or, more plausibly, that a portion of applicants is turned away because there are no vacancies.[14] Since ability is unobserved, we cannot directly estimate the structural relationship between ability and job spell duration, ϑ. Instead, we use the empirical relationship between test scores and productivity from Table V ($\hat{\zeta} = 53.9$ in equation (**??**)) to calculate the implied value of ϑ based on other moments of the model. Putting these pieces together, we calculate that $\hat{\sigma}_s^2 = 0.71$, $\hat{\sigma}_\eta^2 = 0.45$ and $\hat{\sigma}_0^2 = 0.29$. Hence, test scores have approximately 60 percent more measurement error than do interviews.[15]

[13] It is the ratio of variances (σ_η^2/σ_0^2, σ_s^2/σ_0^2), not their levels, that determines the informativeness of the signals. Thus, the normalization that $\sigma_0^2 + \sigma_s^2 = 1$ is innocuous.

[14] To adjust for vacancies, we estimate the hiring rate conditional on a vacancy ('active hiring rate') by calculating what the model implies that the hiring rate *should be* at the 95th percentile of the test score distribution given other estimated parameters. If the observed rate is lower than the calculated rate, we attribute the difference to lack of vacancies and impute the active hiring rate as the ratio of the implied hiring rate to the observed hiring rate. In practice, the active hiring rate is solved simultaneously with the other parameters of the model since they are not independent. We estimate the active hiring rate at 40.4%; that is, 4 in 10 applicants are hired when a vacancy is present.

[15] It would be highly surprising to find that tests are *more informative* than interviews since the item response data gathered by the personality test appear (to us) crude relative to the nuances of attitude and behavior observable during interviews.

Using these parameter estimates in combination with the database of $189,067$ applications summarized in Table II, we implement the following simulation procedure:[16]

1. For each applicant, we draw a simulated ability level, y, as a function of the applicant's observed test score and the estimated error variance of the test. Although this simulated ability level is not observed by employers, it contributes to applicants' interview and test scores and completely determines their job spell durations conditional on hire.

2. Using the ability draws and the estimated variance parameters, we draw an 'interview signal' for each applicant. In contrast to applicant ability levels, these interview signals are observed by firms and are used for hiring.

3. Using applicants' interview signals, their race, and firms' priors, we calculate firms' 'interview-based' posterior productivity assessment for each applicant (see equation (2)).

4. We then simulate hiring under the interview-based regime by calculating a store-specific interview-based hiring threshold such that the count of applicants whose interview-based posterior assessment meets the threshold exactly equals the count of hires observed at the store. Applicants meeting the threshold are labeled 'interview-based hires.'

5. We next use the draws of ability, y, to calculate the job spell durations of interview-based hires (equal to $D = \hat{\alpha} + \hat{\vartheta} y$). In combination, steps (4) and (5) allow us to calculate the race composition and productivity of hires (both overall and by race) under the *interview-based* regime.

6. To obtain analogous outcomes under the *test-based* regime, we repeat steps (3) through (5), making two modifications to the procedure. First, we replace firms' interview-based posterior productivity assessments with their test-based posterior productivity assessments (see equation (4)).[17] Second, when performing the simulated hiring process in step (4), we replace the interview-based hiring threshold with a test-based hiring threshold that generates an identical number of hires at each store.

[16] We sketch the procedure here, with further detals available in an unpublished appendix.

[17] These test-based posteriors differ from the interview-based posteriors *only* in that they incorporate both interview and test signals.

7. In the final step, we compare the race composition and productivity of hires (overall and by race) under the interview-based and test-based regimes. Since the distribution of ability and the hiring rate are identical at each store under each regime, a comparison of (simulated) hiring and productivity outcomes under these two regimes provides an estimate of the pure screening effect of testing on equality and efficiency.

This baseline procedure simulates the case where both interviews and tests are unbiased. It must be slightly extended to explore cases where test or interview biases are present. Table II shows that applicants from the majority group score significantly higher on the job test than applicants from the minority group. We accordingly consider two cases for test bias: in the first case, tests are *unbiased* and, by implication, minority applicants are on average less productive than majority applicants; in the second case, we assume that job tests are *majority-favoring* while minority and majority applicants have the same average productivity.[18]

We allow for the possibility of interview bias in a parallel fashion. Because the data provide no guidance on the possible sign of interview bias, we consider three cases: no bias, minority-favoring bias, and majority-favoring bias. In the unbiased case, the interview signal is equal to $\eta = y + \varepsilon_\eta$, as above. In the minority-favoring case, the interview signal additionally includes an additive bias term that precisely offsets the mean test score differences between minority and majority applicants. Conversely, in the majority-favoring case, the interview signal contains a bias of equal magnitude and opposite sign to the minority-favoring case.

These assumptions give rise to six permutations of the simulation: two cases of testing bias (unbiased and majority-favoring) permuted with three cases of interview bias (unbiased, minority-favoring and majority-favoring). For each scenario, we perform $1,000$ trials of the simulation to obtain mean outcomes and bootstrapped standard errors, equal to the standard deviation of outcomes across trials. Because our focus is on Black-White differences, we discuss and tabulate results for only these two groups. Hispanics are included in the simulation, however.

[18] Since we do not know the *true* mean ability of each applicant group—only the group's mean test score—we make the following ancillary assumptions: if job tests are unbiased, mean ability for each applicant group is equal to the groups's mean test score. If job tests are majority favoring, mean ability for each applicant group is equal to the White mean.

- See Table IX for simulation results. The case of unbiased-interview + unbiased-test is most supported by the data. (This is consonant with List 2004.)

- The main contribution of this paper are: (1) to reframe the equality-efficiency conundrum to show that it corresponds to a strange prior (one where employers do not have rational expectations); (2) to offer a means to test for statistical discrimination by examining how employers react to changes in the information set.

- The specific empirical application has limitations, and certainly more could be done to nail down these findings.

2.10 Farber and Gibbons

Though it may not seem immediately related, the 1996 *QJE* paper by Farber and Gibbons contributes to economic understanding of how information is resolved over time in labor markets. The follow-up paper by Altonji and Pierret, *QJE* 2001, shows how the tools developed by F&G can be used to test for the presence of statistical discrimination (something Farber and Gibbons may have overlooked).

Basic idea:

- When a person enters the labor market, some things about productivity are known to employers, but much is not yet known.

- Given the right data, there may be things about productivity known to the econometrician that cannot be known to employers (like AFQT scores).

- Employers should learn this productivity information as they gather information about workers' productivity.

- Question: What does this learning process imply about wage dynamics?

Two strong initial assumptions:

1. Wages equal expected output at each date (no long term contracts).

2. "The stochastic component of a worker's output has a time invariant distribution, so human capital acquisition is deterministic, and both innate ability and acquired skill have the same value in all jobs." [Translation: error distribution for individual wages is stable at all times, so expected output at any give data is all we need to know for wage determination.]

2.11 Theory: Time invariant worker characteristics

- Let η_i and s_i describe worker's time-invariant ability and (fixed) schooling.
- Assumption: η_i not observed by employers directly (or by econometrician)
- Let X_i be a vector of time-invariant worker attributes (race, gender, date of birth) observable to employers and included in data.
- Let Z_i be a vector of time-invariant worker attributes that are not included in data.
- Let B_i be a vector of time-invariant worker characteristics that are observed in data but not by employers (such as test scores).
- Write the joint distribution of these attributes as $F(\eta_i, s_i, X_i, Z_i, B_i)$.
- Let y_{it} be the output of worker i in the t^{th} period in which she is labor market.
- Assume that outputs $\{y_{it} : t = 1, ..., T\}$ are independent draws from conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$ Note that B_i does not appear in this expression, meaning that it only affects productivity through its relation to other variables (such as η).
- Now assume

  1. All employers know the joint distribution $F(\eta_i, s_i, X_i, Z_i, B_i)$ and conditional distribution $G(y_{it}|\eta_i, s_i, X_i, Z_i)$.
  2. All observe schooling s_i and other worker characteristics X_i, Z_i.
  3. All observe the sequence of outputs $\{y_{i1}, ..., y_{iT}\}$. This last assumption is not mild: **'public learning.'**

- Given these assumptions, wage paid to a worker is expected output given all available information

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

- The rest of the theoretical part of the paper develops implications. These arguments are subtle, but intuition is pretty accessible.

2.12 Three predictions: (1) Effect of schooling on wages

- Consider a cohort of workers entering the labor market simultaneously. For each worker, we observe s_i, X_i and the wage, but *not* output.

- We estimate the following reduced-form equation for the *level* of earnings (one thing that's unusual here about the theory is that it is written in levels not logs, which therefore requires F&G to do the analysis in levels).

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

  Estimated coefficients from this regression are coefficients from linear projection of w_{it} on s_i and X_i. A linear projection is analogous to a conditional expectation function (CEF) except that we have constrained the relationship to linearity. [It could be that the true conditional expectation of structural equation is a non-linear function of independent variables.] Most of us use this terminology loosely, but the F&G usage is more precise.

- *Note that Z_i is not included in the regression–though presumably it does affect wages.* We don't include Z_i because it is not observed in the dataset, though it is observed by the employer.

- Denote linear projection as $E^*(\cdot)$.

$$E^*\left(w_{it}|s_i, X_i\right) = \hat{\alpha}_t + \hat{\beta}_t s_i + X_i \hat{\gamma}_t, \tag{7}$$

  where the $\widehat{hats}$ denote estimated coefficients.

- We now want to deal with the fact that Z is not included in the data. We can handle this by iterating expectations:[19]

$$E^*\left(E^*\left(y|X,Z\right)|X\right)\right) = E^*\left(y|X\right).$$

So, although Z_i, B_i are not included in the data, employers will still be able to form unbiased estimates of expected worker productivity just using the $X's$. Why? Because the estimated coefficients $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ effectively contain the implicit regression of Z, B, on X. So, the omitted variable bias formula says that the coefficients obtained from estimating (7) account for the direct effect of X on w plus the covariance between X and Z, B and the effects of Z, B on w. [You should review/remind yourself of the omitted variable bias formula if you are not already intimately familiar with it. See any intro econometrics text.]

- Some additional iterations gives you

$$E^*\left(w_{it}|s_i, X_i\right) = E^*\left(y_{it}|s_i, X_i\right),$$

meaning that the regression equations gives expected output (equal to the wage by assumption).

- Now, given the assumption that outputs are identically distributed, this arguments says that the effect of schooling on the level of wages is independent of experience.

- Intuition (badly needed...). The assumption that wages equal expected output *and* that outputs are identically distributed implies not only that 1^{st} period wage w_{i1} is expectation of first period output given s_i, X_i but also that no part of 'innovation in wages' between periods 1 and 2, $w_{i2} - w_{i1}$, can be forecast from the information used to determine w_{i1}. Thus:

$$w_{i2} - w_{i1} \perp x_i, z_i, s_i$$

[19] Law of iterated expectations: $E\left(y\right) = E_x\left[E\left(y|x\right)\right]$. In words: although we don't observe x, which is a determinant of y, we can still take unbiased expectations of y by integrating over x. This is so intuitive, it's almost confusing. For example, we know that both schooling and IQ affect earnings. We only observe schooling, not IQ, though we assume they are positively correlated. Can we predict mean earnings by schooling level even though IQ is unobserved? Of course. As a predictive matter, the OLS relationship between schooling and earnings implicitly captures the unobserved relationship between IQ and schooling and earnings as well as the direct relationship between schooling and earnings. Thus, the omitted variables problem does not pose a problem for prediction.

(Though w_{i2} is not independent of w_{i1}.) This does not necessarily mean that wages are not expected to grow between these two periods (although, in this version of the model, they are not – but see below). They key condition is that the 'surprise' element of this growth cannot be forecasted from time invariant characteristics. So this means that education's effect on wages is the same in all periods. (Note, in this version of the model: current age and experience are not included in X_i because they are not time invariant.).

- Thus, w_{i2} is equal to w_{i1} plus a term that depends on y_{i1} but is orthogonal to s_i, X_i, $w_{i2} - w_{i1}$.
- This means that the estimated coefficients for β, γ are the same in all periods.
- Note that this does not necessarily apply to α.

2.13 Three predictions: (2) Effect of unobserved characteristics

- This second prediction is the most important of model (not purely assumption in, result out).
- Recall that B_i is observed by econometrician but not by employers.
- Other observed variables X, Z, s could be correlated with B, however. So let's purge this correlation using the following procedure:

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

  Note that we are including w_{i1} because it contains (by assumption) employers' expectation of productivity at market entry, so by conditioning this out, we are in theory purging everything about B_i that employers may observe.

- Now augment previous regression equation

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

- Question: How will π_t vary with experience? (This is the 'learning' part of the model.) Take B to be scalar. Since B^* is orthogonalized from other regressors, we know that

$$\hat{\pi}_t = \frac{cov\left(B_i^*, w_{it}\right)}{var\left(B_i^*\right)}.$$

Now, note that

$$w_{it} = w_{it-1} + \zeta_{it} = w_{i1} + \sum_{t=2}^{T} \zeta_{it},$$

where ζ_{it} is equal to the 'innovation' in wages in each period following the 1st.

- Since B_i^* is orthogonal to w_{i1} by construction, $\hat{\pi}_0 = 0$. Therefore

$$cov\left(B_i^*, w_{it}\right) = \sum_{t=2}^{T} cov\left(B_i^*, \zeta_{it}\right).$$

In general, this term $cov\left(B_i^*, \zeta_{it}\right)$ will be positive for all $t > 0$. If that regularity condition is satisfied, then $\hat{\pi}_t$ will be rising in t. This is the key result: as experience accumultes, employers will learn about η, and this makes B_i^* increasingly (important) for wage determination.

- Intuition: B_i^* is correlated with ability, and ability affects output. Employers do not observe B_i^* at market entry, but they will learn about output over time. Therefore, the effect of ability – and hence B_i^* – should be rising with market experience. This is quite a nice prediction. It's sensible enough to be intuitive, but sufficiently non-obvious that testing it has power to affirm/reject the plausiblity of the model. Notice: there is no experiment, but still a good test of theory.

2.14 Three predictions: (3) Wage residuals are a martingale

- A martingale is a generalization of a random walk. A random walk has the following two properties:

$$E\left(w_{it+1} | w_{it}\right) = w_{it}, \tag{8}$$

$$var\left(E\left(w_{it+1} | w_{it}\right)\right) = \sigma^2, \tag{9}$$

where σ^2 is not time varying. Note that unconditional variance of w_{it} rises unboundedly with t.

- Like a random walk, a martingale has the the first conditional expectation property. But the second property is more general:

$$var\left(E\left(w_{it+1} | w_{it}\right)\right) = \sigma_t^2.$$

That is, the variance of 'innovations' is not necessarily constant. So, a random walk is a martingale, but a martingale may not be a random walk.

- The martingale property applies readily to the F&G model. Because wages are assumed to incorporate all contemporaneous information on expected productivity, the expected innovation in each period is zero:

$$E\left(\zeta_{it}|w_{it-1}\right)=0,$$

which implies that

$$E\left(w_{it}|w_{it-1}\right)=w_{it-1}.$$

Since successive observations of the wage can be written as

$$w_{it}=w_{i1}+\sum_{t=2}^{T}\zeta_{it},$$

where ζ_{it} are uncorrelated due to the independence of draws assumption. Therefore, $\mathrm{Var}(w_{it})=\mathrm{Var}(w_{i1})+\sum_{t=2}^{T}\sigma_t^2$. This implies that wages (or at least their residuals) are a martingale.

2.15 Productivity that grows with experience

- The problem so far is that they've not really allowed for productivity growth.

- Assume now that productivity grows over time due to experience, human capital, etc.:

$$Y_{it}=y_{it}+h(t),$$

where t is time, and $h(t)$ is the component of output growth due to acquired skill (y_{it} is the part due to fixed characteristics like ability).

- Continue to assume that $\{y_{i1},...,y_{iT}\}$ are *iid* draws from $G\left(y_{it}|\eta_i,s_i,X_i,Z_i\right)$.

- Of course, $\{Y_{i1},...,Y_{iT}\}$ are not *iid*, but if $h(t)$ is deterministic or deterministic plus white noise, then the regression equation becomes something like

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

So, by conditioning on a trend in experience and education, we can get back to a (conditional) *iid* case.

2.16 Farber and Gibbons: Results

- F&G use the NLSY to examine the earnings dynamics of a cohort of workers. The key features of the NLSY for this analysis:

  1. Panel data – can observe wage dynamics
  2. AFQT – correlated with productivity, probably not initially fully observed by employers
  3. Can measure actual experience rather than potential experience.

- First step is to construct B_i^* as

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

  By regressing out other observables *and* the 1^{st} period wage, will have orthogonalized this measure to components that are presumably not known at labor market entry. Though if there is measurement error in wage or if wage does not perfectly reflect expected productivity, this is not going to be quite perfect.

- Table 2 tests their two of three of their main hypotheses:

  1. Estimate effect of education on level of wages does not vary with experience
  2. Estimated effect of variables correlated with ability but not observed by the market increases with experience (this is the most important prediction of the paper).

- Since regressions are in levels, hard to read. But the estimated return to education is roughly 9 percent, which is plausible in this time period, 1981 - 1990.

- There is no evidence that the relationship between education and earnings varies with time, exactly as the learning model predicts. (This is tested by interacting education with experience; experience is almost synonymous with time within a cohort.)

- This finding is sensitive to inclusion/exclusion of education $\times$ year effects; when these are excluded, interaction *is* significant. This is not surprising given the very rapid rise in the return to education during this period; this will look like an 'age' effect if we don't make it into a 'time' effect by adding education $\times$ time.

- Most striking are the 'returns' on the AFQT (and library card) residuals. The interactions between these variables and experience is positive, significant and large in all specifications. This suggests that employers learn about these variables over time, and so they become 'priced' into the wage.

- One concern here is that result could be contaminated by a rise in the return to ability over this period – again, the general identification problem in distinguishing age from time effects. But F&G results are robust to addition of interactions between calendar dummies and the AFQT and library card variables. This suggests that they have found something real.

- The final question that F&G address is whether wages are a martingale. This is of great interest if you do optimal minimum distance (OMD) estimation and/or error components models (drawig on seminal 1984 work by Chamberlain). I will not have time to cover this in class, though it is of independent interest

2.17 Conclusion: Farber and Gibbons

This is an original and important paper that presents a rigorous way to think about a difficult problem – wage dynamics with employer learning. The key result (in my mind) is that unmeasured skill becomes increasingly important to wage setting over time, as employers' learn about ability. The intuition for this result is quite natural, and that makes the finding all the more powerful.

2.17.1 Digression: A note on age, time, and cohort effects

- Write the wage of a person

$$w_{iatm} = \gamma_a + \delta_t + \theta_m + \varepsilon_{it},$$

where a indexes age, t indexes calendar time, and y indexes the year of $i's$ labor market entry. In this notation, γ is an age effect, δ is a time effect, and θ is a cohort effect.

- In theory, all of these 'effects' can be said to exist. That is, there can be a pure wage effect of age (or experience), a pure effect of time (operating through price changes, for example), and a pure effect of being a member of a given cohort (if that cohort has special attributes or, equivalently, if cohorts are imperfect productive substitutes).

- However, there is a fundamental identification problem in measuring these three effects simultaneously. Conditional on year of market entry, age and time are perfectly linearly related – that is, calendar time is simply year of market entry plus age minus age at market entry. So, although we can write this decomposition, we can never estimate all three components simultaneously.

- This identification problem arises repeatedly in labor economics. In particular, if wages are rising for a group of workers, is it because they are getting older (an age effect) or because prices are rising (a time effect)? We'll talk more about this issue in 14.662.

3 Learning Models: Altonji and Pierret (2001)

Altonji and Pierret pick up a thread that Farber and Gibbons left hanging. F&G considered how employer learning affects the evolution of the wage loading of variables that employers do not originally observe. They did not consider what this implies about the evolution of coefficients for variables that: employers do originally observe **and** which are correlated with observables variables that they do not observe. In fact, their estimation strategy precludes investigating this possibility because they purge B_i of correlations with all observables and the 1^{st} period wage.

Here's where A&P come in. Let's say, for example, that race is correlated with AFQT score, and that employers know this. If employers value AFQT score (because of its link to productivity) and use race (or education) as a statistical signal of expected AFQT (that is, they statistically discriminate), this has an immediate implication for the evolution of both the AFQT and race (or education) coefficient. Specifically, AFQT should become more important in wage setting over the workers' career, whereas conditional on AFQT, race (or education) should become *less* important. This is what Altonji and Pierret test.

There are a few important differences between A&P and F&G:

1. A&P do their analysis in logs, which is slightly easier to deal with.

2. Unlike F&G, they do not orthogonalize B_i with respect to X_i, w_{i0}. This is crucial. If B is orthogonal to other covariates, changes in its loading on wages over time *cannot* by construction affect the coefficients on variables in X such as schooling. Hence, their test requires them not to orthogonalize.

3. They look at statistical discrimination on education and race.

I won't develop their model, but the intuition is stated succinctly on page 321, "As employers learn about the productivity of workers, s [which is an observable variable, such as schooling] will get less of the credit for an association with productivity that arises because s is correlated with z [variable like AFQT that is initially unobserved, but is positively correlated with both s and output], provided that z is included in the wage equation with a time-dependent coefficient and can claim the credit."

3.1 Results for education as a signal of ability

- See Tables I, II, III. Results here are quite striking.

- In Table I, first column shows that both education and AFQT have an important effect on wages, and that the effect of education declines insignificantly with time.

- Column (2) adds AFQT $\times$ time/10. Coefficients imply that AFQT has essentially zero effect on wages in year of market entry but by year 10, a one standard deviation higher AFQT raises earnings by 7 log points.

- Most strikingly, addition of this measure dramatically changes the effect of education on earnings. In the first year, the 'return' to education is now 8.3 log points (up from 5.9 in the model excluding AFQT $\times$ time).

- But the education $\times$ time interaction is strongly negative. After 10 years in the market, the effect of education is only 6.0 log points.

- This suggests that employers are 'statistically discriminating' on education – that is, they are initially using education as a proxy of unobserved ability, and they rely on this less as ability becomes known.

- Results using sibling wage and father's education work similarly. All can be thought of correlates of underlying productivity that become know to employers with time.

3.2 Is race used for statistical discrimination?

- A&P report a 1.1 standard deviation black/white mean difference on the AFQT in their sample – large, but consistent with Neal and Johnson.

- Let's say that employers *do not* statistically discriminate on race *but* there is a race difference in productivity on average (owing to the AFQT gap).

- This says that the Black main effect in the wage regression should be negative. But, if we add Black $\times$ experience, the Black main effect should be less negative and the Black $\times$ experience term should be negative. This would be consistent with the idea that employers *don't* use race to proxy productivity at time of hire but do condition subsequent wages on realized productivity.

- Columns 1 and 3 of Table I are consistent with this implication. The Black main effect becomes substantially less negative when a Black $\times$ experience term is added to the model.

- If realized productivity becomes a greater determinant of earnings as experience accumulates, and there is a race gap in AFQT that is not already accounted for in initial wage setting, then inclusion of an AFQT $\times$ time measure should make less negative the cofficient on Black $\times$ time.

- This prediction is supported. Adding AFQT $\times$ time changes the Black $\times$ time coefficient from -0.1315 to -0.0834. This is consistent with the hypothesis that employers are learning over time that Black employees are less productive (i.e., learning about AFQT).

employers had been statistically discriminating, this type of learning should not be going on.

- Taken at face value, these results suggest that employers are not 'accounting' for racial differences in expected productivity at hire (witness: the race intercept is insignificantly positive in year 0 of market entry) and then are paying instead for realized differenes in productivity.

- Employers may be obeying the legal prohibition on statistical discrimination.

- Other explanations?

4 Affirmative action, negative stereotypes, and self-fulfilling prophesies

Affirmative action policies are laws that require employers to give a 'leg up' to minority applicants in hiring (perhaps by enforcing that the minority and majority hiring rate must be equalized). A crucial question is whether, by creating expanded opportunities for minorities in the present, these policies can render themselves unnecessary in the future. By improving employers' perceptions of minorities or improving minorities' skills or both, might affirmative action policies potentially become non-binding? In the best case, AA would eventually cause employers to want to hire minorities even absent AA. Alternatively, by dampening incentives for minorities, one could imagine that affirmative action policies would reduce minority skill investment, thus leading to an equilibrium where employers correctly believe minorities to be less productive than majorities, and so quotas remain binding.

In a seminal paper, Coate and Loury in 1993 provide a theoretical analysis of this question. The starting point of their model is the following three assumptions:

1. The underlying (or perhaps innate) skill distribution of minorities and non-minorities is the same. This skill distribution is modeled as a distribution of *costs* of obtaining a qualification.

2. Employers cannot observe qualifications but do observe noisy signals that are correlated with it.

3. Employers have rational expectations about worker qualifications and workers have rational expectations about employer screening. Thus, in equilibrium, employers beliefs about worker qualifications will be confirmed. And, similarly, workers will make investments consistent with the returns they will receive in the labor market for those investments.

Unlike most static statistical discrimination models, worker skills respond endogenously to employer beliefs (since these effect returns to skill investments). Thus, every equilibrium must be a self-fulfilling prophesy. Once this equilibrium is established, the question is, what happens if we 'shock' the system by imposing affirmative action.

4.1 Structure

- A large number of identical employers, a larger population of workers
- Employers are randomly matched with many workers from this population.
- Workers belong to one of two identifiable groups, B or W.
- Let λ equal the W fraction of the population.
- Employers' sole action is to assign each worker to one of two possible tasks, "zero" and "one."
- All workers are equally good at task "zero," but only qualified workers are successful at task "one."
- Workers get the gross benefit ω if assigned to task one.
- Employers get net return x from assigning a worker to task "one" of the form:

  $$x = \begin{cases} x_q > 0 \text{ if worker is qualified} \\ -x_u < 0 \text{ if worker is unqualified} \end{cases} .$$

  Let $r = x_q/x_u$. Employers get 0 from assigning any worker to task "zero."
- Wages and task productivities are fixed in this model.

- Employers cannot observe worker productivity prior to assignment. They observe worker identity $[B, W]$ and a noisy signal of the worker's qualification level $\theta \in [0,1]$. You can think of θ as emanating from a test.

- The distribution of θ depends on whether or not the worker is qualified, and this relationship does not differ between groups B and W.

- Let $F_q(\theta)$ be the probability that the signal does not exceed θ given that the worker is qualified. And similarly, $F_u(\theta)$ is the probability that the signal does not exceed θ given that the worker is unqualified. Let $f_q(\theta)$, $f_u(\theta)$ be the density functions.

- Write $\varphi(f_u(\theta)/f_q(\theta))$. Assume that $\varphi(\cdot)$ is non-increasing on $\theta \in [0,1]$, which implies that $F_q(\theta) \leq F_u(\theta)$ for all θ. Thus, higher values of the signal are more likely if the worker is qualified, and, for a given prior, the posterior likelihood that a worker is qualified is larger if his signal takes a higher value. So, $\varphi(\cdot)$ has the Monotone Likelihood Ratio (MLR) property. [And $F_q(\theta)$ first order stochastically dominates $F_u(\theta)$.]

- Naturally, employers' assignment policies will depend on a threshold rule where workers with a given signal $\theta \geq \theta^*$ will be assigned to Task 1 and otherwise assigned to Task 0. Since employers observe B and W, the threshold may be group specific θ^*_B, θ^*_W.

- Workers are only *qualified* to perform Task 1 if they have made a costly ex ante investment. The cost distribution varies among workers. Let c be a worker's investment costs and $G(c)$ by the cumulative distribution of investment costs. Suppose for now that $G(c)$ does not differ by group. Thus, there is no ex ante difference in potential qualification levels of B and W workers.

- Equilibrium concept: In equilibrium, it must be the case that workers' investment decisions are consistent with employers' hiring decisions so that workers do not have an incentive to change their investment policies given employer's hiring policies and vice versa. Equilibrium is a pair of employer beliefs about W and B qualification levels that are self-confirming.

- A *discriminatory equilibrium* is one in which employers believe that workers from one group are less likely to be qualified. Notice that because the cost distributions are identical among worker groups, this would imply that employers' beliefs caused one group to invest less than the other.

- See Figure 1 for the sequence of actions

4.2 Employers' decision rule

Consider a worker from group W or B. The fraction of workers qualified in this group is π, and so this is the employer's belief. Conditional on the worker's signal, the employer's posterior probability that the worker is qualified is:

$$\begin{aligned} \xi(\pi,\theta) &= \frac{\pi f_q(\theta)}{\pi f_q(\theta) + (1-\pi) f_u(\theta)} \\ &= 1/\left[1 + ((1-\pi)/\pi)\varphi(\theta)\right]. \end{aligned}$$

The expected benefit of assigning a worker to Task 1 is therefore:

$$\xi(\pi,\theta) x_q - [1 - \xi(\pi,\theta)] x_u.$$

So, an employer will assign a worker to task 1 iff:

$$\begin{aligned} r &= \frac{x_q}{x_u} \geq \frac{1-\xi(\pi,\theta)}{\xi(\pi,\theta)}, \\ r &\geq [(1-\pi)/\pi]\varphi(\theta). \end{aligned}$$

Given the MLR assumption, there will be a threshold standard $s^*(\pi)$ that depends on group membership (through π) so only workers with θ greater than s^* are placed in task 1:

$$s^*(\pi) = \min\{\theta \in [0,1] \quad r \geq [(1-\pi)/\pi]\varphi(\theta)\}.$$

Note that s^* is decreasing in π. Thus, a higher qualification rate of a group will lead to a lower threshold hiring standard s^*. (This should be intuitively clear.)

4.3 Workers' investment decision

The expected gross (*not net*) benefit to obtaining the qualification is:

$$\begin{aligned}\beta(s) &= \omega\left[(1-F_q(s))-(1-F_u(s))\right] \\ &= \omega\left[F_u(s)-F_q(s)\right],\end{aligned}$$

where s is the passing standard.

[Why does π not enter into this equation? Because employers have rational expectations, so all that should matter is the *true* probability that a worker is qualified, not the employer's beliefs about this probability (which of course must match the reality)].

It is critical to observe that $\beta(s)$ is a single-peaked function that satisfies $\beta(1)=\beta(0)=0$ (since there is no point to investing when everyone or no one is assigned to Task 1). Moreover, $\beta(s)$ is rising in s when $\varphi(s)>1$ and falling in s when $\varphi(s)<1$. In other words, the gross (not net) benefit to investing will be rising in the threshold so long as the marginal probability of being assigned to task 1 is increasing in s. To see this, note that:

$$\frac{\partial\beta(s)}{\partial s}=\omega\left(f_u(s)-f_q(s)\right),$$

which is positive iff $f_u(s)>f_q(s)$:

$$\varphi(s)=\frac{f_u(s)}{f_q(s)}>1.$$

Since we know that $\beta(1)=\beta(0)=0$, it follows that this ratio must sometimes be above 1 and sometimes below 1. (If it were always above 1, it would be logical to invest when $\beta(1)=\beta(0)=0$, which cannot be the case. If it were never above 1, no one would invest under any value of s.)

Workers will invest if $\beta(s)\geq c$, so the share of workers investing is $G(\beta(s))$.The paper asserts (and I have no reason to disbelief it) that so long as $G(\cdot)$ is continuos and that $G(0)=0$, it will also be the case that $G(\beta(s))$ is single-peaked and rising (falling) with s as $\varphi(s)>1$ (<1) with $G(\beta(0))=G(\beta(1))=0$. That is, so long as the gross benefit to investing is rising with s, the net benefit should also be rising, and should certainly be falling if the gross benefit is falling. This conclusion comes from the fact that $\varphi(s)$ is non-increasing in s. Thus, the marginal net benefit to investing has to be weakly falling in $\varphi(s)$. So, if the gross benefit is initially positive, it will eventually become negative. Given that $G(0) = 0$, some workers will initially find it optimal to invest if $\varphi(0+\varepsilon) > 1$.

4.4 Equilibrium

An equilibrium is a fixed-point of these investment and hiring policies where beliefs are self-confirming such that:

$$\pi_i = G(\beta(s^*(\pi_i))) \qquad i = W, B.$$

A discriminatory equilibrium (one with $\pi_b < \pi_w$) can occur whenever this equilibrium equation has multiple solutions. In this case, it is possible that employers will believe that $B's$ are less qualified than $W's$ and this set of beliefs will be confirmed by $B's$ investment behavior. Specifically, if employers believe that π_B is lower, they will set a higher threshold for them ($s_B > s_W$) and this *may* reduce B investment. [Again, this is not always the case. Multiple equilibria must be feasible for this outcome to occur. And it infeasible in the model for employers to believe that $B's$ are less productive than $W's$ unless there is an equilibrium that supports this belief.]

Figure 2 illustrates. The y axis is the employers belief, π, and the x axis is the employers screening threshold given π. The locus EE is the threshold s^* that an employer would choose given beliefs π, so $\{(s,\pi) \,|s = s^*(\pi)\}$. The locus WW is the fraction of workers who will obtain the qualification given s^*, so $\{(s,\pi) \,|\pi = G(\beta(s))\}$. Any point where these two loci intersect is a possible equilibrium since at these points, employers beliefs match workers' behavior. There are three points of intersection depicted: s_w, s_b and a middling point that is not labeled.

The point $\tilde{s}$ corresponds to $\varphi(s) = 1$, with $\varphi(s) < 1$ to the right and $\varphi(s) > 1$ to the left.

Proposition 1 proves that under some regularity conditions (spelled out), that *if any solution exists to the equilibrium condition, then at least two solutions will exist.* That is, there are at least two fixed points.

[The figure shows intuitively why this will be true: if $\varphi(s)$ is continuous, and strictly decreasing on $s \in (0,1)$, employers' thresholds will be monotonically falling in the fraction of workers that they believe are qualified (π). Meanwhile, workers actual qualifications will be initially rising and then falling in s. Thus, providing the curves intersect at least once (i.e., on the upward sloping side of WW), they must cross a second time (on the downward sloping side of WW)).

Several key observations:

1. Group identity (B, W) conveys information *only* because employers expect it to.

2. Stereotypes are inefficient. B workers and employers would both be better off if the discriminatory equilibrium were replaced by the non-discriminatory equilibrium.

3. Yet, no single employer could break the discriminatory equilibrium. Employers could understand that there are multiple equilibria and that the alternative equilibria are more desirable. However, without collective action, there is no profitable deviation that a single employer could make.

4. In the discriminatory equilibrium, the employer's expected benefit from hiring a W worker exceeds that of hiring a B worker. This not especially appealing result is typical for statistical discrimination models. The expected benefit from the *marginal* hire is, however equated. [The assumption of the paper is that wages ω in tasks 1 and 0 are equated between the two groups. Otherwise, the lower wages of $B's$ in the discriminatory equilibrium would further dampen investment incentives.]

4.5 Affirmative action

Given that the discriminatory equilibrium is inefficient and there is no underlying difference in the capabilities (cost functions) of B and $W's$, the benevolent social planner can easily rationalize an affirmative action policy in this setting. One such policy would be that the employer apply the rule that $s_W^* = s_B^*$. But this is going to be difficult to enforce since it would require the planner to observe the employer's full information set.

A more realistic policy is one in which the planner mandates that the rate of assignment of B and W workers to Task 1 is equalized. This is the policy that Coate and Loury consider.

Some more notation: Let $\rho(s, \pi)$ be the ex ante probability that a worker is assigned to

Task 1.

$$\rho(s,\pi) = \pi[1 - F_q(s)] + (1-\pi)[1 - F_u(s)].$$

Let $P(s,\pi)$ be the expected payoff from hiring this worker:

$$P(s,\pi) = \pi[1 - F_q(s)]x_q - (1-\pi)[1 - F_u(s)]x_u.$$

Recall that λ is equal to the fraction of W workers in the population.

Under affirmative action, the employers problem is to solve:

$$\begin{aligned} & \max_{s_w,s_b} [1-\lambda] P(s_b,\pi_b) + \lambda P(s_w,\pi_w) \\ \text{s.t. } \rho(s_b,\pi_b) &= \rho(s_w,\pi_w). \end{aligned}$$

Notice that the affirmative action constraint will be non-binding if employers already have homogenous beliefs about the qualifications of B and W workers.

As the paper discusses, it is extremely desirable situation that imposing AA generates a non-discriminatory equilibrium. The paper develops a sufficient condition for this to occur.

You can intuitively see where this comes from in Figure 2. Clearly, raising s reduces the number of workers assigned to Task 1. However, raising s may also raise the fraction of workers investing. In particular, if s is less than $\tilde{s}$ in Figure 2, an increase in s raises the fraction of investors. Conversely, if s is greater than $\tilde{s}$, then lowering s raises investment. Thus, forcing employers to reduce standards may cause a virtuous circle in which worker qualifications improve and so the lower standards are then an equilibrium.

More specifically, if the number of workers the employer expects to assign to task one is always rising with a reduction in the standard s, then an affirmative action policy will necessarily move the equilibrium to one in which employers have the same beliefs about B and W workers. In particular, if any group of workers facing the standard s invests so that the fraction $G(\beta(s))$ is qualified, then all equilibria are self-confirming: $\hat{p}(s) = \rho(s, G(\beta(s)))$. Thus, when employers equate the hiring rate of $W's$ and $B's$, the fraction of qualified $W's$ and $B's$ will also be the same. The AA policy will therefore be self-enforcing after it is in place. (The AA constraint will bind only during the movement between equilibria.)

This is not generally guaranteed to be the case, however. In fact, you can see that it is not true in the example in Figure 2. If at point s_w, the employer lowered the threshold to $s' < s_w$, the fraction of workers investing would fall, and so the employers beliefs about the fraction who are qualified would not be satisfied. Thus, a policy that forced s_w downward would not be self-enforcing.

Coate and Loury define an equilibrium where the affirmative action constraint is permanently binding as a 'patronizing equilibrium,' because employers are compelled, against their unconstrained profit maximizing interests, to lower their hiring standards for $B's$ relative to $W's$ to equate the group hiring rates. Thus, in a patronizing equilibrium, $s_B^* < s_W^*$ and $\pi_B < \pi_W$.

The logic of the patronizing equilibrium is:

- To comply with the equal-assignment mandate, and believing $B's$ to be less productive, employers patronize $B's$ by making it easier for them to achieve the assignment to Task 1.

- Because it is easier for them to succeed, $B's$ find it optimal to invest less, which then confirms employers' negative views.

- Notice one subtlety here: In the *discriminatory equilibrium* without AA, the standard for $W's$ is lower but their rate of investment is higher: $s_B^* > s_W^*$, $\pi_B < \pi_W$. In the *patronizing equilibrium*, the standard for $B's$ is lower and yet their rate of investment is also lower.

- However, we know that $B's$ and $W's$ are identical initially. So why doesn't a lower screening threshold for $B's$ lead them to have higher investment than $W's$? It's all a question of initial conditions. Refer again to Figure 2. Starting from the equilibrium s_W^*, s_B^*, it is clear that even a fairly large movement leftward in s_B will not raise B investment sufficiently to confirm employers' beliefs.

More cogently, imagine that $B's$ are a small proportion of the population so it's near-optimal to employers to accommodate the AA policy by equating the B Task 1 assignment rate to the unconstrained W rate. This will require employers to initially reduce s_B to *below* s_W^* because $B's$ are initially investing less, and so setting $s_B = s_W^*$ would yield a lower Task 1 assignment rate for $B's$ than $W's$. But Figure 2 makes clear that if $s_B < s_W^*$, then $\pi_B < \pi_W^*$ because we are moving leftward in the portion of the WW locus where $\varphi(s) > 1$.Thus, a reduction in s_B

from s_B^* to $s_B' < s_W$ will necessarily produce a patronizing equilibrium and could potentially lower B investment further than in the discriminatory equilibrium such that $\pi_B' < \pi_B^*$.

4.6 A bit more formally...

I will not work through the proofs of the main result, but the figures make it clear how this works. Consult especially Figure 4.

- Imagine we are at a discriminating equilibrium and that AA is imposed.
- Assume that $B's$ are a non-negligible share of the population, so it is optimal to comply with AA by reducing W hiring and raising B hiring.
- This means that there is a negative shadow price on each B hire (a subsidy) and a positive shadow price on each A hire (a tax)
- Panel A of Figure 4 shows the hiring rate of each group as a function of the screening threshold. There is only one line because both groups are the same. However, due to initial conditions, hiring rates differ.
- To lower W hiring and raise B hiring, s_w will have to rise a little bit and s_B will have to fall a lot (from above s_w to below it).
- The new equilibrium depicted has two different EE loci, one for $B's$ and one for $W's$. Notice that both (the new) s_b and s_w are equilibrium points in that these beliefs are self-fulfilling.
- Given that these beliefs are self-confirming, would this equilibrium be stable if the AA policy were removed? The answer is no, because profits would increase if employers could hire more $W's$ and fewer $B's$. In particular, the marginal productivity of B and W hires is not equated. The marginal W generates positive expected profit and the marginal B generates negative expected profit.

The paper proves that such cases do exist under non-extreme assumptions.

4.7 Several small extensions

1. Consult Figure 5. Assume that $B's$ and $W's$ have underlying differences in cost distributions (perhaps because $B's$ attend poor schools). Imagine that we are at a Pareto efficient equilibrium where s_W^* and s_B^* are on the left side of the WW loci for each group (so $\varphi(s_i) > 1$ for each group). If an AA policy is imposed, employers will respond by raising s_W and lowering s_B. Because we are to the left of $\tilde{s}$ for each group, $W's$ will respond with more investment and $B's$ will respond with less investment. Thus, AA would exacerbate the skill deficits that it was presumably enacted to address.

2. In general, subsidizing employers to hire $B's$ is not going to work well. A small subsidy from a discriminating equilibrium will raise B investment, but it will not be self-sustaining. A large subsidy from a discriminating equilibrium could work if the government got it just right. But it could also produce a patronizing equilibrium.

3. By contrast, subsidies to $B's$ to obtain qualifications cannot fail to be beneficial so long as $s_B < 1$ (so some $B's$ are initially hired). A sufficiently large subsidy will always move B investment into the $\varphi(\cdot) > 1$ portion of the curve.

4.8 Conclusion

This is a lovely paper, and also a pleasure to read because it is so well constructed. Fifteen years after its publication, however, we have almost no evidence on what affirmative action does to hiring, employer perceptions and worker investments. Thus, it remains to be established whether the adverse scenarios exposited by this paper are practically relevant. One thing that is clear is that many people in positions of political and business leadership believe that affirmative action is harmful to minorities (and to majorities). (BTW, it's plausible that the 'patronizing equilibrium' is even more salient, and more damaging, in higher education than in the labor market.)

5 Stereotype Threat—Another form of self-fulfilling prophesy

The 'stereotype threat' hypothesis originates with psychologist Claude Steele and coauthors. This hypothesis says that members of groups that are 'stereotypically' believed to have negative attributes may behave in ways that confirm these attributes when the 'stereotype threat' is made salient. For example, female mathematicians may perform badly on math tests when they are reminded that many people believe that women are not as capable as males at mathematics. Or, blacks may perform poorly on IQ tests when the tester subtly suggests that blacks are not as intellectually capable as whites.

This hypothesis may sound far-fetched, but in fact it is easy to think of cases where it *could* be relevant. Steele gives the example of a black male sociologist who feels anxious when he is waiting in line at an ATM if the customer ahead of him happens to be a white woman. Although this sociologist has no criminal intent, he is aware that white female ATM customers may be made anxious by his presence, believing that black males pose a criminal threat. Presumably, the sociologist reacts to this 'stereotype threat' by trying to appear especially non-threatening, perhaps by keeping exaggerated physical distance from other customers. (It's conceivable that, opposite of the intention, this makes other customers more nervous.)

In this example, the stereotype threat does not make the sociologist more likely to engage in a criminal act (which the Steele hypothesis might suggest it should). But it does support the idea that members of discriminated groups may be acutely aware of stereotypes—at least in situations that make these stereotypes salient—and that this awareness could potentially affect behavior and outcomes. (For a fictional example in which stereotype threat experienced by black males *does* directly lead to a criminal act, see the car-jacking scene in the 2005 movie *Crash.*)

Anecdotes are not evidence, but, as we will see, the experimental evidence favoring the stereotype threat hypothesis is somewhat remarkable.

Note that the stereotype threat hypothesis does not fall under the other categories we've studied. It is neither animus-based nor statistical discrimination; it is self-fulfilling prophesy. One could write an economic model about this, but it would be difficult in such a model to motivate the idea that an individual would choose to behave in a way that is individually self-destructive in response to a stereotype. (Note that in the Coate-Loury paper, everyone is playing his or her optimal strategy.)

According to the Steele 1997 article, Stereotype Threat has two testable implications:

1. For domain-identified students, stereotype threat may interfere with their domain-related intellectual performance. Translation: For groups reputed to be worse at a given task, they may perform worse at that task if stereotype threat is activated.

2. "Reducing this threat in the performance setting, by reducing its interfering pressure, should improve the performance of otherwise stereotype-threatened students" [It's not clear to me that this implication is distinct from the prior implication.]

This idea may strike you as far-fetched (it does me). But it is readily confirmed by experimentation. Its economic importance is, however, unknown.

6 'Acting White'

There is a long-standing hypothesis, attributed to Fordham and Ogbu, that part of the deficit in Black academic achievement is due to the cultural stigma that Blacks face for 'acting White,' that is, conforming with the mainstream culture. Under the AW hypothesis, Blacks economic benefits but social costs of achieving, thus reducing skill investment. By contrast, Whites and other groups that do not have an oppositional culture, receive both economic and social benefits from achieving.

The relevance of this hypothesis—which has probably attained a bit of the character of urban myth—has come in for criticism at various points. A widely discussed article by Cook and Ludwig in 1998 (originally published in the *APPAM*) documented using the NLSY that Blacks who are academically successful are as popular as academically successful whites. C&L view this as evidence that Acting White is not in reality stigmatized. However, the 2005 article on your syllabus by Austen-Smith and Fryer counters that C&L confuse the level with the derivative. That is, the AW hypothesis says that the derivative of social status WRT academic success if negative (or less positive) for Blacks than Whites, and this is not at odds with academically successful members of both groups being popular.

Because the AW hypothesis has received so much play, it is potentially valuable to see how it can be framed in economic terms. The paper by Austen-Smith and Fryer offers an example. The set the problem up as one of 'dual audience signaling.' Specifically, agents are attempting to signal their desirability to two groups simultaneously: employers and peers. Employers do not care about peer group membership and peers do not care about skills. But nevertheless, the model identifies a potential tension between the objectives of pleasing both audiences.

A nice insight of the model: In environments in which "acting White" is salient, improved external labor markets have the effect of encouraging more individuals to leave the group, while causing those in the group to invest less in education.