Lecture Note: The Economics of Discrimination I

David Autor
MIT 14.663 Spring 2009

April 12, 2009

1 Economic models of discrimination

An enormous literature, starting with Becker's 1957 book *The Economics of Discrimination*, explores (you guessed it!) the economics of discrimination. Economic models of discrimination can be divided into two classes: competitive and collective models. Competitive models study individual maximizing behavior that may include discrimination. In collective models, groups acts collectively against each other. Almost all economic analysis has focused on competitive models (and we will do likewise). Competitive models can normally be further divided into taste-based and statistical models of discrimination. Becker's model studied the former case, and it provides an excellent starting point for discussion.

In formal economic terms, discrimination is when members of a minority group are treated differently (less favorably) than members of a majority group with identical productive characteristics. Let the wage Y be the wage equal to

$$Y_i = X_i\beta + \alpha Z_i + e_i, \tag{1}$$

where X_i is a vector of exogenous productivity characteristics and Z_i is an indicator variable for membership in a minority group. Assuming that $X_i\beta$ fully captures the set of productive characteristics and their returns and/or Z_i is uncorrelated with e, then discrimination is a case where $\alpha < 0$.

We already face three difficulties just using this simple definition.

1. 'Productivity' may directly depend on Z—for example, in the entertainment industry or a market in which customers value Z in workers (e.g., discriminating customers). If customers will pay more to see a white actress or a black athlete, is this a legitimate component of productivity?

2. There is also a question of whether β—the production technology—is truly exogenous. For example, operating firefighting equipment requires considerable physical strength and stature. This has historically been used as an argument against the entry of women into this profession. But these physical requirements are engineered attributes and probably could be altered. If humans were 20 percent less physically strong, presumably they could still fight fires.

3. The $X's$ could also be endogenous. Pre-market discrimination—or expectations of future discrimination—could reduce $X's$ for members of the minority group. (Examples: poor schools, or a rational belief among minorities that education will not be rewarded by the market.)

Point (1) is one we may be able to examine directly. Point (2) and (3) are much harder to test. But whether or not these are relevant, it can still be the case that $\alpha < 0$ conditional on both X and β, which would constitute discrimination.

In Becker's 1957 model, employers have a 'taste for discrimination,' meaning that there is a disamenity value to employing minority workers. (Hence, discrimination comes directly out of the utility function.) In this case, minority workers may have to 'compensate' employers by being more productive at a given wage or, equivalently, by accepting a lower wage for identical productivity. The basic insights of this model require almost no formalization. We will formalize only very slightly.

- Let A denote majority group membership and B denote minority group membership.
- Employers will maximize a utility function that is the sum of profits plus the monetary value of utility from employing members of particular groups.
- Let d be the taste parameter of the firm, which Becker called the "coefficient of discrimination."
- Firms will maximize

$$U = pF\left(N_b + N_a\right) - w_a N_a - w_b N_b - dN_b,$$

  where p is the price level, F is the production function, N_x is the number of workers of group $x = \{a, b\}$, and w_x is the wage paid to members of each group.
- Employers who are prejudiced ($d > 0$) will act as if the wage of b group members is $w_b + d$. Hence, they will only hire b group members if

$$w_a - w_b \geq d.$$

- Let $G(d)$ denote the cumulative density function (CDF) of the prejudice parameter d in the population of employers.

- The optimal number of workers hired at each firm is determined by the solutions to

$$\begin{aligned} pF'(N_a + N_b) &= w_a, \\ pF'(N_b + N_b) &= w_b + d. \end{aligned}$$

- Treating p as fixed and aggregating across firms in the economy leads to the market demand functions $N_a^d(w_a, w_b, G(d)), N_b^d(w_a, w_b, G(d))$ for each worker type. Wages are determined by

$$\begin{aligned} N_a^d(w_a, w_b, G(d)) &= N_a^s(w_a), \\ N_b^d(w_a, w_b, G(d)) &= N_b^s(w_b), \end{aligned}$$

  where $N^s(\cdot)$ are the supply functions for the worker types.

- Notice the main point that comes out of this setup is this: A wage differential $w_b < w_a$ will arise if and only if the fraction of discriminating employers (or discriminating jobs) is sufficiently large that the demand for B workers when $w_b = w_a$ is less than the supply.

- In other words, *discrimination on average does not mean discrimination at the margin.* If there are enough non-discriminating employers, then discrimination is competed away. This also implies that minority workers don't work for discriminating employers.

- If, however, the share of prejudiced employers is sufficiently large, then some b group members will work at $d > 0$ employers, and this implies that $w_b < w_a$. In this case, the strength of prejudice at the margin (that is d for the marginal employer of b workers) is what determines the size of the wage gap.

- With free entry or constant returns to scale (CRS), discriminating employers may be competed out of business. In a competitive market, each worker must earn his marginal product. Under CRS, non-discriminating firms would simply expand to arbitrage the wage differential borne by minority workers. In equilibrium, discriminating employers

must fund the cost of their distaste out of their own pockets; they cannot pass the cost onto the minority worker.

So, to summarize:

- In partial equilibrium, minority workers must 'compensate' employers by being more productive at a given wage or, equivalently, accepting a lower wage for equivalent productivity.
- These tastes create incentives for segregation. It is potentially Pareto improving for minority workers to work in their own businesses and similarly for majority workers—then no one bears the cost of the distaste.
- In general equilibrium, it *may* only be possible for employers to indulge these tastes at a positive cost.

Key testable implications of this model are:

1. Wage differentials: Minority workers earn less than majority workers of identical productivity.
2. Preferential hiring: Employers are less likely to hire minority workers of identical productivity.

But these implications may not apply in equilibrium—so it's not clear when we should observe them. Indeed, Keneth Arrow memorably remarked that the employer discrimination model "predicts the absence of the phenomenon it was designed to explain." It is perhaps this perplexing feature of the Becker model (i.e., that it appears to predict that discrimination does not affect labor market outcomes in equilibrium) that has caused economists to primarily *not* seek evidence for taste-based discrimination. Indeed, the paper by Heckman in the 1998 *JEP* symposium on race discrimination makes an impassioned case that the key facts that many sociologists and some economists take as evidence of taste-based discrimination are really nothing of the sort. Heckman's view appears deeply influenced by the Becker model. (To be clear, Heckman is not arguing that there are no employers who harbor a taste for discrimination—only

that this taste doesn't affect the equilibrium wage paid to minorities.) The recent *JPE* paper by Charles and Guryan (2008), however, argues that Arrow critique is misplaced, and that in fact the Becker model applies broadly, both as a matter of theory and of empirical fact.

Though often dismissed by economists, it's instructive to consider the primary evidence for taste-based discrimination

1.1 Audit studies: Pager, Western and Bonikowsi (2006)

Audit studies are a well known tool for measuring discrimination. In the typical audit study, 'matched' testers (who are, in effect, actors) of different races with substantively identical resumes are sent sequentially to employers advertising job vacancies. These studies evaluate whether the minority members of these pairs fare systematically worse as measured by callbacks and job offers.The limitations of such studies are only too obvious: they are not double-blind, and it's natural to suspect that minority testers (who, presumably, are committed to the cause of feretting out discrimination) may unconsciously take actions that reduce their odds of receiving a job offer; sample sizes are invariably small because audit studies are costly; the set of employers audited is likely to be randomly chosen—which, according to the Becker model, does *not* correctly capture the set of employers to whom minorities will apply in equilibrium.[1]

Given these flaws, it is easy to dismiss audit studies out of hand. This is somewhat harder to do, however, if you read them. The working paper by Pager, Western and Bonikowski on your syllabus is a case in point. The study uses matched black, white, and Latino job seekers. To 'calibrate' the magnitude of racial preferences, it compares minority applicants to white applicants just released from prison. White testers in the 'criminal record' group were instructed to reveal that they had recently been released from prison after serving 18 months for a drug felony (possession with intent to distribute cocaine). The white tester's criminal record was additionally signaled on the resume by listing work experience at a state prison, and by listing a parole officer as a reference.

[1] An additional problem is that audit studies essentially attempt to manipulate an economic object that is fixed, which is race. That is, one cannot randomly assign race to a tester. Yet the audit study procedes under the fiction that the minority tester is the counterfactual of the non-minority tester with the difference that the minority is 'treated' with being Black or Hispanic. If you write down the standard causal notation, you'll see that this requires a 'unit homogeneity' assumption (in Holland's terminology) that we rarely apply in the social sciences.

The study uses matched teams of testers who applied for 341 entry-level jobs in New York City over nine months in 2004. All testers were well-spoken, clean-shaven young men ages 22 to 26. Most were college grads between 5 feet 10 inches and 6 feet in height. Testers were given training to standardize/equalize their modes of communication.Testers presented themselves as high school graduates with steady work experience in entry-level jobs.

Employers were sampled from job listings for entry-level positions, defined as jobs requiring no previous experience and no education greater than high school. Job listings were randomly drawn each week from the classified sections of The New York Times, The Daily News, The New York Post, and The Village Voice. From the available population of job listings, the auditors used a simple random sample of advertisements each week. Testers in each team applied to each job within a 24-hour period, randomly varying the order of the applicants.

The statistical approach is probably a bit of overkill. They fit the following model:

$$\ln\left(P_{ij}/\left(1-P_{ij}\right)\right) = \alpha_j + \beta_1 \times \text{Black}_{ij} + \beta_2 \times \text{White}_{ij} + \varepsilon_{ij},$$

where α_j is an employer random effect. This model is estimated using Markov-Chain Monte Carlo (no idea why) but they reassure us that they get the same answers using more conventional procedures. P is an indicator variable equal to one if the applicant received an offer or request for second interview. (I will refer to both as callbacks.)

The first two figures make it clear that there are enormous differences in callback rates between White, Black and Hispanic applicants. White felons do about as well as Black and Hispanic applicants with no (reported) criminal record.

To assess the possibility that the results are driven by the self-fulfilling prophecy effect (what is typically called an experimenter 'demand effect'), they compare callback rates for cases where there was essentially no contact between the employer and the applicant with callback rates where there was in-person contact. The demand effect hypothesis suggests that minority applicants should have fared worse in these in-person interactions. This is opposite to what is found. (Unfortunately, the study does not make it clear how 'no contact' is defined and whether this can be considered an exogenous attribute of the specific application setting or whether it might be an outcome onto itself.)

The paper has many gripping examples of employers' showing favoritism to white applicants.

One objective way to assess this type of favoritism is to measure 'chaneling,' where the applicant is steered to a different job than the one to which he applied. Upward channeling is a positive outcome, whereas downward chaneling is a negative outcome. Table 2 offers fairly compelling evidence of downward chaneling of minorities and upward chaneling of whites.

It is hard to come away from this paper believing that discrimination is not (at a minimum) present among employers of low education workers in New York City.

1.2 A Quasi-Experiment in the Labor Market: Orchestrating Impartiality, Goldin and Rouse (AER, 2001)

This study attempts to isolate the importance of gender preference in a market setting: orchestra auditions. A very simple idea: During the 1970s since 1990s, some orchestras started using screens during solo auditions to hide the identity of performers. Women were historically viewed as unsuitable for orchestras. Did the use of blind screens improve their chances of getting a job?

Statistically, this study uses a differences-in-differences design. The authors evaluate the success rate of women relative to males at orchestras using blind auditions relative to orchestras using non-blind auditions. Using the same group of subjects (individual performers) observed in both venues makes this comparison informative.

Let's develop this logic formally. We would like to know the effect of a candidate's gender on her probability of hire. You may be tempted to think that the causal effect of interest is: $T_i = E\left[Y_{1i} - Y_{0i} | F_i = 1\right]$, where $F = 1$ indicates that the candidate is female ($F = 0$ indicates male), Y_{1i} is the probability of hire if female, and Y_{0i} is the probability of hire if male. But this framing of the question doesn't quite make sense. We cannot actually manipulate the gender of an applicant (without changing many, many other things), so it's perhaps not meaningful to ask how an applicant i would have fared were she instead a male.

But there is a way to frame the question that does not run afoul of this 'reality' constraint. Consider instead the question: how would Male and Female applicants have fared under a blind audition system (i.e., where gender is not known) relative to a non-blind system where gender is revealed. This is a reasonable alternative question to ask because we believe that the *relevant* criterion for selection should be independent of information on gender. Is it interesting to ask

whether masking versus revealing gender affects hiring.

Let Y_1 equal the outcome of a person in a blind audition and Y_0 equal the outcome under a nonblind audition. In this case, the treatment effect of interest is $T = E[Y_1 - Y_0|B = 1]$, that is the difference in outcomes for a women auditioning under a blind screen relative to the outcome she would have experienced under a nonblind screen.

That's a step in the right direction, but there are several issues that we need to consider here before moving to the data:

1. Blind versus nonblind auditions may have a direct effect on hiring odds for both males *and* females. Thus, it may be misleading to only estimate the blind/nonblind contrast for females.

2. The women who choose (or are selected) to audition for orchestras using blind screens may be different from those who choose orchestras with nonblind screens.

3. It's finally possible that the same performers perform *differently* in front of a blind versus under the glare of other musicians. For example, they might be more nervous when performing absent the blind, and this effect could potentially be larger for women than men (given the historical prejudice against women)

The first two of these problems, we can handle. The third we cannot solve using the data available to Goldin and Rouse. We'll discuss why in a minute.

Let's write the expected probability of hire for a performer in a non-blind and blind audition as:

$$\begin{aligned} E[Y_{0i}] &= \alpha_i + \gamma F_i, \\ E[Y_{1i}] &= \alpha_i + \beta. \end{aligned}$$

Thus, the hiring probability in the non-blind condition is a function of individual ability (α_i) and possibly a gender discrimination coefficient (if $\gamma < 0$). The hiring probability in the blind condition is a function of individual ability and a blind condition 'main effect' (β) which may affect hiring odds for both genders. [Note, there is no reason to also introduce a non-blind main effect since the blind main effect is really only defined as a contrast to the non-blind condition].

So, if we compare the outcomes of females ($F = 1$) who audition at both blind and non-blind conditions, we obtain

$$E\left[Y_{1i} - Y_{0i}|F_i = 1\right] = \alpha_i + \beta - \alpha_i - \gamma = \beta - \gamma.$$

Thus, this contrast gives us a combination of the causal effect of interest and a nuisance parameter, which is the main effect of the blind condition. Consider the analogous contrast for males:

$$E\left[Y_{1i} - Y_{0i}|F_i = 0\right] = \alpha_i + \beta - \alpha_i = \beta.$$

Combining these equations gives us the 'difference-in-difference' estimator

$$\hat{T}_{DD} = E\left[Y_{1i} - Y_{0i}|F_i = 1\right] - E\left[Y_{1i} - Y_{0i}|F_i = 0\right] = \gamma.$$

Thus, the difference-in-difference estimator solves *two* problems here: first, contrasting outcomes of females in blind/non-blind relative to males in blind/non-blind eliminates the direct effect of the blind treatment on hiring odds (so we can isolate the pure gender effect). Second, using a sample of candidates who auditioned in *both* venues allows us to eliminate the pure quality effects (α_i) that might otherwise bias our estimates (e.g., if more capable women choose to participate in primarily at blind auditions).

This estimation strategy does not solve the causal inference problem, however, if performers perform differently during blind and nonblind auditions. In that case, α_i is not constant since it will depend on whether or not the audition is blind.

- See Table 1 for a summary on the implementation of screens
- See Figure 3 for long-term trends in female hiring
- Table 4: On average, women do *worse* on blind rounds. But this could be due to composition of the female pool in blind rounds. It's possible that only the very best women compete when the game is lopsided (i.e., in the non-blind rounds).
- Table 5: Models limited to musicians (male and female) who auditioned both blind and non-blind suggest that women did relatively better in blind rounds (diff-females minus diff-males).

- Table 6 gives the main estimates.
- Table 7 estimates models for the 3 orchestras that switched policies; here the authors can include musician and orchestra fixed effects. Results are similar to Table 6, but less precise.
- Conclusion: Fascinating, though the evidence is not as conclusive as one might anticipate.
- It is a virtue of this study that the quasi-experiment takes place in an extant (albeit unusual) market setting. That is, the sample is composed of real, high-stakes employment decisions. We don't have to worry about Hawthorne effects, demand effects, and other distortions potentially induced by a laboratory environment.
- *Does this study provide evidence that orchestras engaged in taste-based discrimination or statistical discrimination?* We cannot tell. It could be that females fared worse in the non-blind condition due to taste discrimination. On the other hand, it could be the case that orchestras statistically discriminated against women in the non-blind condition. The blind condition rules out either taste-based or statistical discrimination—since neither can operate when the orchestra cannot determine the gender of the performer. So, the contrast here is between no discrimination (blind audition) versus unknown form(s) of discrimination (in the non-blind condition).
- These results could also be explained by 'stereotype threat,' i.e., women performed better when auditions were blind because they knew there was no discriminatory expectation? It would be feasible (and fascinating) to assess this hypothesis. If we had recordings of the actual auditions, these could be evaluated by independent experts. If women and/or men performed systematically worse or better under the blind versus non-blind condition, this would invalidate the interpretation of the findings above as a reflection of discrimination. The take-away conclusions from this study rest on the untested (but not untestable) assumption that blind auditions do not directly and differentially affect female versus male performance.

1.3 An audit study without the auditors: Bertrand and Mullainathan (2003)

If you are concerned that 'demand effects' may contaminate the Pager et al study, you may be reassured by the study by Bertrand and Mullainathan (2004). This study involves sending resumes by mail or fax to advertised job positions while manipulating *perceptions* of race by using distinctively ethnic names (otherwise holding constant resume characteristics). The question of the study is whether 'callback' rates lower for applicants with distinctively black names? (B&M are not the first to do this, but they did it on a large scale and their work received considerable attention.)

Briefly consider the statistical framework.We are not manipulating *race*, but we are manipulating perceptions of race by randomly assigning 'black-sounding' or 'white-sounding' names to otherwise identical resumes. Denote each resume by i. We write the probability of callback for resume i given a white-sounding (Y_0) or black-sounding (Y_1) name as:

$$\begin{aligned} E[Y_{0i}] &= \alpha_i, \\ E[Y_{1i}] &= \alpha_i + \gamma N_i. \end{aligned}$$

Hence, the estimated treatment effect in this case is:

$$\hat{T} = E[Y_{1i} - Y_{0i} | N_i = 1] = \gamma.$$

Notice that we do not need a 'second contrast' as we did in the blind versus non-blind audition comparison for males versus males; the only treatment in this study is the 'blind' condition. In the prior study, we were concerned that performers who chose to perform in blind versus non-blind auditions might not be not comparable to one another. We solved this problem by limiting the sample to performers who performed in both audition types. We don't have a similar concern in this case because we have randomized assignment of names to resumes. So, this virtually guarantees comparability in a large sample: $E[\alpha | N = 1] = E[\alpha | N = 0]$.

More formally, think of the basic comparison of hiring odds of black and white-sounding resumes:

$$\hat{T} = E[Y_1 | N = 1] - E[Y_1 | N = 0] = (E[Y_1 | N = 1] - E[Y_0 | N = 1]) + (E[Y_0 | N = 1] - E[Y_0 | N = 0]).$$

In general, the first term following the equal sign is the contrast of interest (i.e., the difference in hiring odds for the same resume assigned a black and white sounding name) and the second term is a bias term (equal to the difference in potential outcomes for resumes with black and white sounding names). Because of the random assignment, however, we can be reasonably confident that this bias is close to zero ($E[Y_0|N=1] - E[Y_0|N=0] = 0$). Hence, the contrast in hiring odds between white-sounding and black-sounding resumes will be due to the random name assignment and no other factor.

Let's examine the results.

- Table 1: Short answer is yes. Callback rates are lower for black sounding names.
- Table 2: In most cases, names receive equal treatment (a criticism that Nobel laureate James Heckman has levied against audit studies more generally). But that's because in most cases, applicants are not called back.
- Table 6: Discrimination based on zip-code characteristics appears quite important and does not systematically differ between white and non-white names. (Authors also view this as evidence against statistical discrimination.)
- Table 8: Considerable overlap in distributions of outcomes between white and black sounding names

Conclusion: Controversial paper with striking findings. It has sparked a great deal of other research in this vein. Questions remaining: How do we translate callbacks into outcomes we care about? Does this provide any information about 'discrimination on average vs. discrimination at the margin'? Does it distinguish taste-based from statistical discrimination?

1.4 Taking the Becker model seriously: Charles and Guryan, 2007

The work described above documents the presence of discrimination in the market, but it tells us almost nothing about the effects of discrimination in general equilibrium (this critique less relevant to G&R's paper), that is, after the mechanisms posited by Becker have worked to segregate minority workers from discriminating employers and to arbitrage (via market competition) employers who harbor discriminatory tastes. Of course, we can conjecture with the some

confidence that the forms of discriminations documented above are *not* fully arbitraged. For example, if job search is costly and workers do not know in advance which employers discriminate, it's very likely that discrimination will lower the reservation wages and employment rates of minorities (see Black, 1995, on your syllabus). Even so, we have no idea by how much. The problem in testing the Becker model is a familiar one. It is a general equilibrium framework, whereas most microeconomic analyses provide only partial equilibrium tests.

Charles and Guryan propose a general equilibrium analysis of the Becker model: does prejudice affects minority wages in market equilibrium? First, let's consider reasons why prejudice *might* survive in equilibrium, despite the forces of arbitrage:

1. Costly search (see Black 1995 or Lang, Manove and Dickens 2005)

2. Nepotism (Goldberg 1982 in the *QJE*). The Goldberg model is a well-known 'tweak' of the Becker model. In the Goldberg model, employers receive positive utility from employing whites, rather than disutility from employing minorities. This is labeled nepotism rather than discrimination. Nepotism makes employers willing to pay from profits for the non-pecuniary gain of indulging their preferences. Thus, they will pay wages in excess of the marginal product to white workers. In the Becker model, discrimination is arbitraged by the market because discriminating employers will presumably be willing to sell their businesses to non-discriminating employers who can earn higher profits with the same capital. But for the nepotistic employer, there is no desire to sell since the non-pecuniary gain compensates for the pecuniary loss. [This is like the sexual harrassment model we discussed earlier. If managers or owners are willing to pay for on-the-job consumption, there's no reason why the market won't accomodate them.] The more general point is that preferences are not generally arbitraged by the market—rather preferences are the underlying force that leads to market equilibrium. A key assumption of the Goldberg model is that there is no cheaper way for nepotistic employers to indulge their preferences outside of the labor market. If there were, they would presumably sell their businesses and indulge these prefernces poolside in Florida.

3. Charles and Guryan propose a third mechanism which is closely related to Goldberg.

They suggest that if a prejudiced employer sells his business, he is likely to become a prejudiced worker. In this case, he may still have to bear the costs of his prejudice by working with minority group members, and so may find it no less psychically expensive to be an employer than an employee. (Though this is unclear. A prejudiced employer may have to pay the full cost of hiring a large number of majority workers to indulge his preference. A prejudiced employee may be able to find a segregated workplace, or his psychic price of working alongside minorities may be less than his monetary cost of running a discriminatory business.)

Let's assume for now that there is a market equilibrium in which prejudiced employers are *not* competed out of business. What should be observed in equilibrium? Figure 2 of C&G suggests some answers. (1) If there are sufficiently few minorities or sufficiently many non-discriminating employers, there may be no discrimination at the margin, meaning no discriminatory wage gap. (2) Holding the distribution of discrimination constant, an increase in the number of minority workers in the market will reduce minority wages if it causes the marginal minority to work at a more discriminating employer. (3) Holding the number of minorities constant, an increase in employer prejudice will lower the wage of minorities if the marginal employer becomes more prejudiced (but not otherwise).

What implications are testable?

1. The marginal level of prejudice matters more than the average level of prejudice for relative wage differences.

2. The racial wage gap increases with the number (or fraction) of blacks in the workforce, holding prejudice constant.

3. Prejudice in the right tail of the employer prejudice distribution should not matter for racial differences, while prejudice in the left tail of the prejudice distribution should affect racial wage gaps

4. The mechanism that should generate these patterns is the tendency of the market to segregate blacks from the most prejudiced whites.

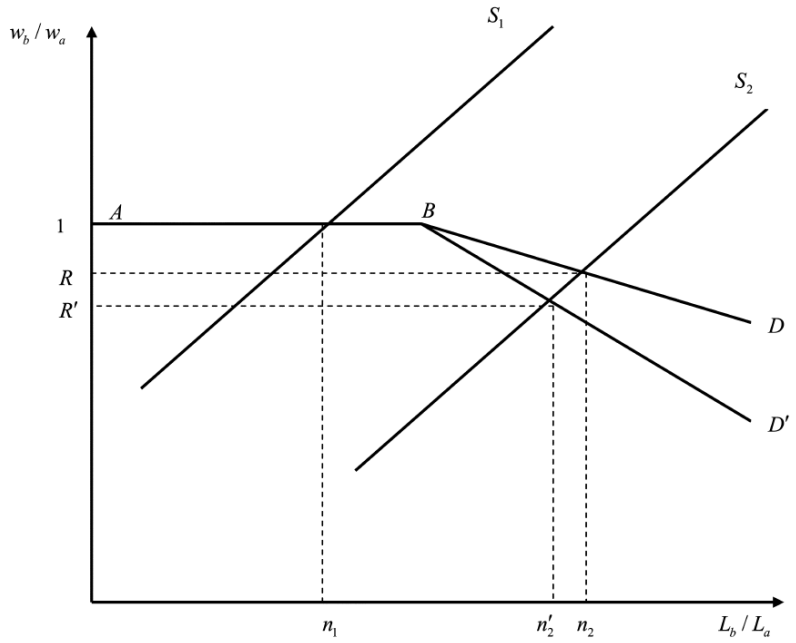

FIG. 2.—Relationship between racial tastes and the relative wages and relative supply of blacks and whites. The figure shows how the equilibrium ratio of black to white wages responds to three sets of market conditions. When the relative supply of black workers is small relative to the number of unprejudiced employers, as is the case when supply is as depicted by S_1, the marginal discriminator is unprejudiced and there is no racial wage gap in equilibrium. When the distribution of racial preferences among employers is held constant, a shift out in the relative supply of black workers (from S_1 to S_2) requires that more prejudiced employers hire blacks, and the ratio of black to white wages falls from one to R. When the relative supply of black workers is held constant, an increase in prejudice among employers likely to be the marginal discriminator (which causes the relative demand curve to rotate from ABD to ABD'), further reduces the equilibrium ratio of black to white wages to R'.

Figure 1: (Charles and Guryan, 2008)

These observations raise a number of tough issues about distinguishing marginal from average prejudice, etc. Charles and Guryan make a bold effort to work through these issues, and the results are highly intriguing.

The first innovation of the paper is to use direct, self-reported measures of (what would typically be interpreted as) prejudicical attititudes from the General Social Survey for years 1972 through 2004 (see paper for examples). The paper uses these measures at the state and regional level. The geographic differences in prejudicial attitudes are remarkably large. The median East South Central region respondent has the same aggregate prejudice as the 81st percentile respondent from New England. The median-prejudiced New England respondent would be at the 26th percentile of the East South Central prejudice distribution. Regions with the largest black population shares are also the most prejudiced.

The primary analysis uses a residualized wage gap as its outcome measure. Specifically, the paper estimates:

$$w_{ijt} = \alpha_j + X_{ijt}\beta + \gamma_t + \theta_j \times \text{Black}_{ij} + \varepsilon_{ijt},$$

where the parameters of interest are the θ_j's, equal to the estimated average residual Black-White wage gap in each state. It's not clear from the text whether the model allows β to vary by year, as should be the case (in which case, β would be denoted as β_t).The θ_j's are used as the dependent variables in the rest of the paper, where weights are inversely proportional to the standard errors of the $\theta's$.

Table 3 contains key initial estimates. As expected, average prejudice is predictive of lower black relative residual wages. More notably, the prejudice of the 'marginal' white is much more strongly predictive of the wage gap than the average prejudice, where the 'marginal' white is defined as the white respondent at the p^{th} distribution of prejudice, where p equals the fraction of workers in a state that is black. (This person is the marginal employer only under very special circumstances, such as all firms having a single employee, and all whites being employers.) The standard deviation of "marginal" prejudice across states is 0.139. The coefficient of -0.213 on the marginal prejudice in the wage model implies that a one-standard deviation increase in prejudice is associated with lower black wages of about 0.028 log points lower—about 23 percent relative to the mean residual wage gap across states. This result is not

quite so surprising after careful inspection. The 'marginal' level of prejudice will mechanically increase with the share of blacks in a state. And we already know that states with a larger share of blacks have larger racial wage gaps.

Columns (5) and (6) are more persuasive, however. Rather amazingly, the 10^{th} percentile of the prejudice distribution (that is, the decile with the least prejudice) is highly predictive of the wage gap, but not the 50^{th} or 90^{th} percentiles. Moreover, this result is robust to controlling for the fraction of the state that is black. It is not obvious what model aside from the Becker model would predict this result. One might have expected that the variance in prejudice at the 10^{th} percentile of the distribution would be mostly noise, and that only the mean or the right tail would have been predictive. Not so.

A significant concern with these estimates is that it might plausibly be the case that blacks are less skilled on average (in unmeasured ways) in regions with greater prejudice. This could, for example, be due to greater racial disparities in school quality in these areas. Table 4 shows, surprisingly, that controlling for mean raical test score differences in reading and math does not greatly reduce the 10^{th} percentile prejudice effect, though the test score gap does have the expected sign for wage gaps. Table 5 shows that limiting the sample to the South or controlling for measures of school quality also does not change the qualitative findings.

Table 6 instruments for fraction black using black population shares in 1920. What problems does this solve, exactly?

Table 7 includes a measure of the fraction of whites' coworkers who are black (measured using the NELS-88 in 2000). This measure is highly predictive of the wage gap—black relative earnings are lower where whites interact more with blacks. Yet this measure is generally not robust to controlling for the prejudice of the marginal discriminator. The final columns of Table 7 are not entirely reassuring, however.

What should you conclude? It's up to you.

2 Statistical Discrimination

The Becker model starts from the presumption that discrimination is essentially non-economic—that is, it stems from employers' tastes rather than from differences in employers' assessments

of workers' productivity. The statistical theory of discrimination, which begins with Phelps (1972) and Arrow (1973), starts from the plausible premise that employers cannot perfectly assess worker productivity (particularly at the time of hire) and instead make educated guesses. This market imperfection (i.e., full information is not available) gives them an incentive to use easily observable characteristics such as race or gender to infer the expected productivity of applicants (assuming these characteristics are correlated with productivity). The Aigner and Cain (1977) article on your syllabus is the standard reference on this topic.

Statistical discrimination is the solution to a signal extraction problem. (There is no statistical discrimination when information is complete.) If an employer observes a noisy signal of applicant productivity and also has prior information about the correlates of productivity (let's say a group-specific mean), then the employer's assessment of applicant productivity should place weight on both the signal and the mean. Two cases are commonly exposited, and we'll also look at a third. We'll use normal distributions for simplicity, but this is not substantively important.

2.1 Case 1: Difference in means

- Assume that when workers apply for jobs, the employer sees the race of the applicant $x = \{a, b\}$ and some error-ridden signal $\tilde{\eta}$ of productivity.

- Assume that employers have learned from experience that

$$\eta_x \sim N(\bar{\eta}_x, \sigma_\eta^2),$$

  with $\bar{\eta}_a > \bar{\eta}_b$, and σ_η^2 identical for a and b. Hence, b group members are less productive on average, but the dispersion of productivity is the same for both groups. We can write $\eta_i = \eta_x + \varepsilon_i$.

- When a worker applies for a job, the employer observes an error-ridden productivity signal of the form:

$$\begin{aligned} \tilde{\eta}_i &= \eta_i + \iota_i \text{ where} \\ \iota &\sim N(0, \sigma_\iota^2), \text{ with } \sigma_\iota^2 > 0. \end{aligned}$$

Hence:

$$\tilde{\eta}_i = \bar{\eta}_x + \varepsilon_i + \iota_i,$$

and $E\left(\tilde{\eta}_i|\eta_i\right) = \eta_i$, meaning that the signal is unbiased.

- What is the expectation of η given $\tilde{\eta}$ and x? This is simply the regression equation,

$$\begin{aligned} E(\eta|\tilde{\eta}, x) &= \bar{\eta}_x\left(1-\gamma\right) + \tilde{\eta}\gamma, \\ &= \bar{\eta}_x + \left(\tilde{\eta} - \bar{\eta}_x\right)\gamma. \end{aligned} \tag{2}$$

  where $\gamma = \sigma_\eta^2 / \left(\sigma_\eta^2 + \sigma_\iota^2\right)$, which is the coefficient from a bivariate regression of η on $\tilde{\eta}$ and a constant (estimated seperately by group). Note that $\gamma_a = \gamma_b$ in this example; all that differs is that $\bar{\eta}_a > \bar{\eta}_b$.

- This equation immediately implies that the for a given $\tilde{\eta}$, the expected productivity of b applicants is below a applicants – even though $\tilde{\eta}$ is an unbiased signal for both workers. In particular

$$E(\eta|\tilde{\eta} = k, x = a) - E(\eta|\tilde{\eta} = k, x = b) = \left(\bar{\eta}_a - \bar{\eta}_b\right) \times \left(1-\gamma\right).$$

  This expression will always be positive provided that $\sigma_i^2 > 0$. (Draw a picture.)

- Notice the main insight/paradox of statistical discrimination:

$$\begin{aligned} E(\tilde{\eta}_i|\eta_i, x) &= \eta_i \\ \text{but } E(\eta_i|\tilde{\eta}_i, x) &\neq \eta_i \text{ (unless } \tilde{\eta}_i = \bar{\eta}_x) \text{ or } \sigma_\iota^2 = 0. \end{aligned}$$

  That is, the expectation of the productivity signal is equal to true productivity. But the expectation of productivity given the signal is not generally equal to actual productivity except in the case where the individual has the average productivity of the group.

- Is there equal pay for equal productivity in this model? No, not in general. There is equal pay for equal *expected* productivity.

- Consider an a and b group worker who both have productivity $\eta = k$. So, $E(\tilde{\eta}_i|\eta_i, x) = k$ for both workers. But, using equation (2), it is clear that $E(\eta|\tilde{\eta}_i = k, x = b) < E(\eta|\tilde{\eta}_i = k, x = a)$.

- Hence, for some workers, statistical discrimination is 'discrimination' in the sense of equation (1).
- But this will not be true on average. Expected productivity equals true productivity *on average* for each group. Be certain that you are clear on this point.

2.2 Case 2: Difference in variances

- Now take a case where $\bar{\eta}_b = \bar{\eta}_a$ and $\sigma^2_{\eta a} = \sigma^2_{\eta b} = \sigma^2_{\eta}$. So, ability is identically distributed in these groups.
- However, the signal $\tilde{\eta}$ may be more informative for one group or another. This would arise if for example a group managers were relatively inaccurate judges of b group ability (or vice versa). This would imply that $\sigma^2_{\iota a} \neq \sigma^2_{\iota b}$.
- It would therefore follow that $\gamma_a \neq \gamma_b$ since

$$\gamma_x = \frac{\sigma^2_{\eta}}{\sigma^2_{\iota x} + \sigma^2_{\eta}}.$$

- For whichever group has lower $\sigma^2_{\iota x}$, $\tilde{\eta}$ will be more informative; employers will put less weight on the mean for this group and more weight on the signal. See Figures 1a and 1b of Cain and Aigner.
- In this case, depending on whether an applicant is above or below the mean of his group, he will be differentially helped or harmed by a steeper γ_x. If i is above the mean, he wants the signal to be as informative as possible. If i is below the mean, he prefers an uninformative signal.
- Contrasting groups a and b as is done in the Cain and Aigner figures, you will see that the expectation of η given $\tilde{\eta}$ will cross at $\bar{\eta}$ for the two groups, and the relative steepness of the a versus b slope will depend positively on $\sigma^2_{\iota b}/\sigma^2_{\iota a}$.

2.3 Case 3: Risk averse employers

- Cain and Aigner also discuss a third plausible case that is rarely examined in the literature: 'Employer risk aversion.' (I'll explain the quotations in a moment.)

- Risk aversion will arise if there are diminishing returns to worker ability. For example, low ability workers may accidentally damage machinery or intentionally harrass patrons. But high ability workers may be only slightly more productive than workers of average ability. (Perhaps there is an upper limit to how well a person can operate a fryolator.)

- If so, uncertainty is harmful to all workers—and the greater the uncertainty, the greater the harm.

- Specifically, the estimation error of equation (2) is

$$e \equiv E(\eta|\tilde{\eta}_i, x_i) - \eta = \frac{\iota\sigma_\eta^2 - \varepsilon\sigma_\iota^2}{\sigma_\iota^2 + \sigma_\eta^2},$$

  which has variance:

$$V(e) \equiv \sigma_e^2 = \frac{\sigma_\eta^2\sigma_\iota^2}{\sigma_\eta^2 + \sigma_\iota^2}.$$

- This variance is increasing in σ_ι^2 since $\partial\sigma_e^2/\partial\sigma_\iota^2 = \sigma_\eta^4/\left(\sigma_\iota^2 + \sigma_\eta^2\right)^2 > 0$. Lower signal precision (a higher value of σ_ι^2) is harmful to workers if employers are risk averse.

- I place the phrase 'risk aversion' is quotations because, although employers in this example act as if they are risk averse, they really are not risk averse in the Von Neumann-Morgenstern sense. Rather, their 'virtual' risk aversion stems from the fact that the production function is concave in worker ability. Accordingly, employers' assessments of expected productivity are below the productivity of expected ability. That's a manifestation of Jensen's inequality.

2.4 Testing statistical discrimination

It is not straightforward to test statistical discrimination since it is difficult to know how employers form expectations. Almost any observed racial/gender difference in pay or hiring can be attributed to statistical discrimination (which is a problem for the theory, not a virtue). Almost all tests of statistical discrimination are therefore indirect. We'll talk about some examples in a moment.

2.5 Statistical discrimination: Efficiency, Legality, Fairness,

2.5.1 Efficiency

It's interesting to speculate on why economists have focused so much more attention on statistical than taste-based discrimination. My guesses:

1. The Becker model employs a modeling trick that many economists consider the last refuge of scoundrels—adding arguments to the utility function. This is a pretty undisciplined technique. By changing the utility maximand, you can pretty much get whatever you want. Arguably, however, it is also the most natural approach here. Casual empiricism suggests that much prejudice takes the form of 'distaste.'

2. Unlike taste-based discrimination, statistical discrimination is not predicted to be 'competed away' in equilibrium. So, we can be reasonably confident that we should be able to find it in a general set of cases.

3. Closely related to (2), statistical discrimination is 'efficient.' That is, statistical discrimination is the optimal solution to an information extraction problem. Economists would generally say that employers 'should' statistically discriminate because it is profit-maximizing, it is not motivated by animus, and it is arguably 'fair' since it treats people with the same expected productivity identically (though not necessarily with the same actual productivity). Many economists might endorse statistical discrimination as good public policy. Indeed, Becker has argued that if discrimination is not based on taste or animus, it's not discrimination.

2.5.2 Legality

Statistical discrimination is generally unlawful in the US. It is illegal in the U.S. to make hiring, pay or promotion decisions based on predicted performance where predictions are based on race, sex, age or disability. Because minorities, women, those over age 40, and the disabled are 'protected groups,' employers are not permitted to hire and fire them 'at will.' (An employer presumably can statistically discriminate among non-disabled, white males under age

40.) Statistical discrimination is probably difficult to detect, however, and so it is plausible that it occurs frequently, despite the law.

2.5.3 Fairness

Leaving aside legality, it is worth asking whether statistical discrimination accords with most commons notions of fairness. Here it's useful to take a loaded example: racial profiling. Say you are a New Jersey State trooper, and there are some number of drug runners who travel on your highways. You have a limited amount of resources to expend on stopping cars, so you want to maximize your productive resources.

- Let's go back to statistical discrimination Case 1: Difference in means. We are going to recast η as a latent index of criminality.

- Assume that when police officers observe cars on the highway, they see the race of the driver $x = \{a, b\}$ and some error-ridden signal $\tilde{\eta}$, corresponding to a latent index that provides information on the probability that the driver is transporting drugs. For simplicity, define this assessment on the real line $\tilde{\eta} \in [-\infty, \infty]$. A lower value of $\tilde{\eta}$ corresponds to a lower likelihood, and a higher value to a higher likelihood. (If you like, you can map these latent index values, $\tilde{\eta}$, into the CDF of the normal to get probabilities.)

- Assume that experience has taught the police that

$$\begin{aligned} \eta_x &\sim N(\bar{\eta}_x, \sigma_\eta^2) \text{ with} \\ \bar{\eta}_a &< \bar{\eta}_b, \text{ and } \sigma_\eta^2 \text{ identical for } a \text{ and } b. \end{aligned}$$

  Thus, b group members are more likely than a group members to be running drugs, though the variance of the latent index is the same for both groups. As above, η_i can be written as $\eta_i = \bar{\eta}_x + \varepsilon_i$.

- The signal for any given car/driver is error ridden.

$$\begin{aligned} \tilde{\eta}_i &= \eta_i + \iota_i \text{ where} \\ \iota &\sim N(0, \sigma_\iota^2), \text{ with } \sigma_\iota^2 > 0, \end{aligned}$$

which is noisy but unbiased.

- How should the police allocate enforcement resources? The optimal decision rule will involve a threshold value of η^*. Police will stop cars with $E(\eta) \geq \eta^*$. One can formalize this rule with a tiny search model, but I will not. The key point is that the police would ideally like to stop only the highest probability cars. But waiting has an opportunity cost, so it would be foolish to wait only for cars with $\tilde{\eta}_i \to \infty$. The optimal decision rule will involve some cutoff η^*: stop any car that meets a critical value η^*. The threshold depends on the distribution of η, the arrival rate of cars, and the opportunity cost of waiting. We'll assume (with justification) that η^* will exist, and (with somewhat less justification) that the New Jersey State Police can solve for it.

- Assume that $\eta^* > \bar{\eta}_a, \bar{\eta}_{b.}$ Hence, only a minority of cars from either group should be stopped.

- Since the police do not observe the true η for any car/driver, they must form an expectation for this value. Using the equations above, the expected value of η given $\tilde{\eta}$ and x is:

$$E(\eta|\tilde{\eta}, x) = \bar{\eta}_x + (\tilde{\eta} - \bar{\eta}_x)\left(\sigma_\eta^2 / \left(\sigma_\eta^2 + \sigma_\iota^2\right)\right),$$

  with estimation error,

$$e = E(\eta|\tilde{\eta}_i, x_i) - \eta = \frac{\iota\sigma_\eta^2 - \varepsilon\sigma_\iota^2}{\sigma_\iota^2 + \sigma_\eta^2},$$

  which has variance

$$V(e) \equiv \sigma_e^2 = \frac{\sigma_\eta^2\sigma_\iota^2}{\sigma_\eta^2 + \sigma_\iota^2}. \tag{3}$$

- The variance of the expectation of η given x is the variance of true productivity minus estimation error. Define $\nu = \varepsilon - e$. We have,

$$V\left(E\left(\eta|\tilde{\eta}, x\right)\right) = V(\nu) = \sigma_\eta^2 + \sigma_e^2 - 2\text{Cov}\left(\varepsilon, e\right) = \sigma_\eta^2 - \sigma_e^2.$$

  This expression underscores a crucial point. So long as $\sigma_\iota^2 > 0$ (the signal is error-ridden), the expectation of η given $\tilde{\eta}, x$ has lower variance than η. Therefore, this estimate 'shrinks' the true range of the underlying variable. (Draw another picture.)

- This is the essence of 'racial profiling.' Since the police know from experience that group b members are more likely to be running drugs than group a members, it is efficient to use this information in determining which cars to stop. You can demonstrate that this is efficient by confirming that the marginal η stopped is the same for both a and b. In fact, this *is* the decision rule.

- Note that for an individual with a $\iota_i = 0, x_i = x$, the police will necessarily under or overestimate her true criminality unless η_i is equal to the group specific mean $\bar{\eta}_x$.

- Two important points follow from this rule.

    1. The share of b cars stopped exceeds the share of a cars stopped. This can be seen as follows:

$$\begin{aligned} \Pr(\text{Stop}|x) &= \Pr\left(E(\eta|x) > \eta^*\right) \\ &= \Pr\left(\nu > \eta^* - \bar{\eta}_x\right) \\ &= \Pr\left(\frac{\nu}{\sigma_\nu} > \frac{\eta^* - \bar{\eta}_x}{\sigma_\nu}\right) \\ &= 1 - \Phi\left(\frac{\eta^* - \bar{\eta}_x}{\sigma_\nu}\right) \end{aligned} \tag{4}$$

    where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution, and the quantity in parentheses $(\eta^* - \bar{\eta}_x)/\sigma_\nu$ is the 'effective screening threshold' for group x. This quantity is the standardized difference between the group's mean and the threshold, scaled by screening precision. The lower is screening precision, the smaller is σ_ν, and the larger is the effective screening threshold.

    - Differentiating (4) with respect to the group mean gives

$$\partial \Pr(\text{Stop}|x)/\partial \bar{\eta}_x = (1/\sigma_\nu)\,\phi\left(\frac{\eta^* - \bar{\eta}_x}{\sigma_\nu}\right),$$

      where $\nu = \varepsilon + \iota$ and $\phi(\cdot)$ is the pdf of the standard normal distribution. Since $\phi(\cdot) > 0$, a higher value of $\bar{\eta}_x$ implies greater odds of being stopped.

    - So, drivers from the b group are stopped more often.

2. The average η (criminality) of b cars stopped exceeds that of a cars stopped. In other words, the level of criminality (or the fraction of criminals) is higher among stopped b cars. You can see this as follows:

$$\begin{aligned} E(\eta|\text{Stop},x) &= \bar{\eta}_x + E\left(\varepsilon_\eta|\nu > \eta^* - \bar{\eta}_x\right) \\ &= \bar{\eta}_x + \sigma_\eta E\left(\frac{\varepsilon_n}{\sigma_\eta}|\frac{\nu}{\sigma_\nu} > \frac{\eta^* - \bar{\eta}_x}{\sigma_\nu}\right) \\ &= \bar{\eta}_x + \rho_{_{\eta\nu}}\sigma_\eta\lambda\left(\frac{\nu}{\sigma_\nu}|\frac{\nu}{\sigma_\nu} > \frac{\eta^* - \bar{\eta}_x}{\sigma}\right) \\ &= \bar{\eta}_x + \frac{E\left[\varepsilon_\eta\left(\varepsilon_\eta - e\right)\right]}{\sigma_\eta\sigma_\nu}\sigma_\eta\lambda\left(Q_x\right) \\ &= \bar{\eta}_x + \frac{\sigma_\nu^2}{\sigma_\eta\sigma_\nu}\sigma_\eta\lambda\left(Q_x\right) \\ &= \bar{\eta}_x + \sigma_\nu\lambda\left(\frac{\eta^* - \bar{\eta}_x}{\sigma_\nu}\right), \end{aligned} \tag{5}$$

where $\lambda(Q) = \phi(Q)/(1-\Phi(Q))$ is the Inverse Mills Ratio (IMR) and $\phi(\cdot)$ is the density function of the standard normal distribution.

   - Differentiating this expression with respect to $\bar{\eta}_x$ gives

$$\partial E(\eta|\text{Stop},x)/\partial\bar{\eta}_x = 1 - \lambda'\left(\cdot\right).$$

   - Since $\lambda'\left(z\right) < 1$ for finite z, the expected criminality of those stopped is increasing in, $\bar{\eta}_x$, the group mean.
   - The reason is that the optimal stopping rule equates the *marginal* return to stopping an a versus b driver. Because there is more mass to the right of η^* for the b group, this means that the *average* criminality of stopped b drivers will be higher than a drivers .

- Most economists would view this stopping rule as 'fair.'

   - The marginal car stopped has the same expected criminality for both groups.
   - There is no animus motivating these choices (i.e., taste-based discrimination).
   - Average criminality is actually *higher* among stopped b group members than stopped a group member, despite the higher frequency of stops of b's.

– Resources are efficiently deployed.

- So, why do civil libertarians complain? And why do b group members get upset about being stopped for "DWB" ('Driving While Black')?

- One possible answer is that well-intended liberals don't understand basic statistical principles.

- Another answer is that this system will seem demonstrably unfair to group b members who are *not* criminals.

- Consider two citizens, one from group a, the other from group b, who have the same $\eta = k$. Assume that $k < \eta^*$, so neither 'should' be stopped. What is the likelihood that each is stopped?

$$\begin{aligned} \Pr\left(\text{Stop}|\eta = k, x\right) &= \Pr\left(E(\eta|x, \eta = k) > \eta^*\right) \\ &= \Pr\left(\tilde{\eta} > \frac{\eta^* - \bar{\eta}_x\left(1-\gamma\right)}{\gamma}|x, \eta = k\right), \end{aligned}$$

where $\gamma = \sigma_\eta^2 / \left(\sigma_\iota^2 + \sigma_\eta^2\right)$, which is the regression coefficient from above.

- The expectation of the left-hand side of this expression is identical for both a and b since $\eta = k$. But the right-hand side is not. Because the effective screening threshold is declining in the group's mean $\bar{\eta}_x$, the b group driver with $\eta = k$ is more likely to be stopped than the a group driver $\eta = k$.

- Another way to see this: For any true level of criminality k, $E\left(\tilde{\eta}|\eta = k, x = a\right) = E\left(\tilde{\eta}|\eta = k, x = b\right)$ is identical for a and b. But $E\left(\eta|\tilde{\eta} = k, x = b\right) > E\left(\eta|\tilde{\eta} = k, x = a\right)$: the police will *not* treat these individuals identically. The b driver is more likely to be viewed as a criminal.

- Substantive point: Although racial profiling is an efficient way to apprehend criminals, it *does* impose a cost on all group b members, including the innocent. There are more Type I errors for group b, i.e., more innocent motorists stopped. If you are a b group member, that may seem patently unfair. In this setting, statistical discrimination appears to pose an equity-efficiency trade-off. (This is reminiscent of the debate about whether airport screeners should differentially search Middle-Eastern airline passengers).

- This observation may explain why economic and lay intuitions on the virtues of statistical discrimination are typically at odds. Many economists would be inclined to view statistical discrimination as socially efficient. Economists may not generally recognize that statistical discrimination is inequitable on average, even if it is efficient by another metric.

- Another point of this example: Many important economic decisions are up/down, yes/no decisions (stop or not, hire or not, promote or not). In these cases, the crucial choice variable is whether some expectation exceeds a critical value, and this can have large distributional consequences. Imagine for example in our racial profiling model that $\bar{\eta}_a < \eta^* < \bar{\eta}_b$ and $\sigma_i^2 \to \infty$. That is, the signal $\tilde{\eta}$ is uninformative and so the decision rule puts full weight on the group mean. In this extreme case, only b group members are stopped and all innocent b group members are inconvenienced.

3 Evidence on statistical discrimination

3.1 Decomposing the beauty premium: Mobius and Rosenblat, AER 2006

There is a robust correlation between 'attractiveness' and earnings (Hamermesh and Biddle, 1994). Why does this correlation arise?

1. They ('Beautiful People' or BP's) are more productive
2. Employers believe that they are more productive (statistical discrimination)
3. Employers or customers may like them better (taste discrimination)
4. BP's may be more self-confident (which may or may not affect or reflect their productivity)

How do we tease these channels apart? This appears daunting...

1. Effect of own appearance on confidence
2. Effect of own appearance on performance
3. Effect of own confidence on other's assessments (statistical discrim)
4. Effect of own appearance on other's assessments (statistical discrim)

5. Effect of own confidence/appearance on other's preferences (taste-based discrimination)

Prior to Mobius and Rosenblat's 2006 paper, it was generally thought that it was not possible to distinguish these channels. Because beauty plausibly affects self-perceptions and other's perceptions, and because self-perceptions also affect other's perceptions, and further because self-perceptions also plausibly affect one's own efficacy/productivity, pinning down the causal channels appeared a bit of a morass.

The Mobius and Rosenblat approach to this problem is reasonably ingenious and subtle. It has six components:

1. Measuring beauty objectively (using external raters)
2. Measuring productivity objectively (mazes)
3. Measuring self-confidence and correlating with attractiveness—Isolates own confidence
4. Controlling employers' knowledge of workers' beauty by limiting information that they are allowed to see before hire:
    - No direct communication
    - Oral communication
    - Oral communication + picture
    - Face to face communication
5. Measuring employers' assessments of expected worker productivity
6. Measuring employers' desire to pay workers more conditional on their expected productivity and beauty

3.1.1 Protocol

Before discussing the statistical model, it's helpful to know the experimental protocol. The first part of the protocol serves to gather information on the relationship between beauty and self-confidence.

1. 'Workers' complete a questionnaire to gather demographic information.

2. Their digital photograph is taken, and their beauty is rated by outsiders.

3. 'Workers' complete a Yahoo maze to practice the job task and to gauge their ability at it.

4. Their time on the practice maze is recorded and placed on their 'resume.'

5. Workers are asked to assess own productivity, equal to the number of mazes they expect to be able to complete in 15 minutes. Workers will be paid according to:

$$100 \times A_j - 40 \times |C_j - A_j| ,$$

where C_j is the workers' predicted maze count and A_j is her is actual count. Observe that this payment scheme provides incentives for workers to report the medians of their posterior distributions (posterior because they have already done a practice maze).

Table 3 shows results of this first exercise:

1. Notice that performance in practice round:

   - Affects own predicted performance—thus, workers learn from the training round.
   - Also predicts actual performance

2. Workers have private information about their own ability:

   - Notice that actual performance 'predicts' estimated performance, even conditional on the 'LNPROJECTED' variable, which extrapolates their expected number of mazes completed during the 15 minute work period based on their time in the practice maze ($LNPROJECTED = \ln((15 \times 60/PRACTICE))$. Thus, workers can meaningfully assess/project their own capabilities.

3. Beautiful people:

   - Believe they will be significantly more productive.

- Are not more productive.

4. Conditional on their projected performance, males are *not* more confident than females. But their realized performance is substantially higher than females.

5. Causality: Are these relationships causal? Is causality relevant here?

Now that we know that beauty affects self-confidence but not productivity in this task, we want to ask how beauty affects employers' willingness to pay (i.e., the wage). There are three primary information channels that may affect employers' willingness to pay:

1. Resume variables:

   - Gender, education, employment
   - Performance during practice round

2. Stereotype and confidence discrimination channels: Statistical

   - Employers may believe that BP's are more productive
   - Employers may believe that confidence predicts productivity—and we know that part of that confidence comes from being beautiful.

3. Stereotype and confidence discrimination channels: Taste-based

   - Independent of beliefs about productivity, employers may prefer to pay BP's more.

Which of these relationships are causal, and which can we experimentally manipulate? Clearly, beauty, confidence, and other $X's$ are fixed, and cannot be manipulated. We can, however, manipulate (1) employers' awareness of beauty, (2) their awareness of confidence, and (3) their incentives to indulge their preferences to reward those traits.

To continue with the protocol:

1. Employers view 5 resumes, each of which includes:

- Standard demographics
- Measured performance on practice round ('Projected performance')

2. There are five information treatment conditions:

   (a) See no other information (Baseline)

   (b) Also see a facial photo (Visual)

   (c) Conduct a 5-minute free-form phone interview (Oral)

   (d) See facial photo and conduct phone interview (VO)

   (e) See photo and conduct 5-minute in-person interview (Face-To-Face)

3. Employers are then asked to choose a wage while facing the following profit function:

$$\pi_i = 4000 - \sum_{j=1}^{5} 40 \times |w_{ij} - A_j| .$$

   This profit function is maximized in expectation when the employer provides his median estimate of the number of mazes that the worker will complete.

4. Prior to employers being asked to choose wages, one of the following two incentive conditions is announced:

   (a) 80% of time: The wage paid to worker will be the one the employer sets. Here, the employer employer can indulge discriminatory taste for the worker *at personal cost* (so this is the Becker setting).

   (b) 20% of time: The wage paid to worker will be set by all other employers (20% of time). In this condition, the employer's taste costs money but does not benefit the worker. So, no gain to indulging tastes.

5. Workers complete mazes for 15 minutes. They are ultimately paid:

$$\pi_j = 100 \times A_j - 40 \times |C_j - A_j| + \sum_{i=1}^{5} w_{ij}.$$

   Thus, worker income is increasing in mazes completed, though workers only make 60 units of income per maze for mazes in excess or shortfall of their projected level. And workers are also paid the wages selected by employers.

3.1.2 Statistical model

Before considering the results, it's potentially useful to write down the statistical model that the authors use to interpret the data.

An employer has to form an estimate about the productivity A of a worker, which is a function of observable resume variables x and an unobservable, normally distributed error component $\eta \sim N\left(0, \sigma_\eta^2\right)$:

$$A = \alpha x + \eta.$$

The worker receives a signal C of his own productivity, where C stands for confidence:

$$C = \eta + \pi B + \varepsilon_c, \text{ where } \varepsilon_c \sim N\left(0, \sigma_c^2\right).$$

The term πB captures any bias in the worker's confidence arising from his physical attractiveness (B stands for beauty). Assume for now that $B = 0$.

Two indicator variables T_O and T_V denote whether this is Oral or Visual communication (both are equal to 1 if there is a face to face conversation).

A key channel through which the model operates is that it is assumed that the employer observes an unbiased signal $\tilde{C}$ of the worker's confidence if there is oral communication:

$$\tilde{C} = C + \varepsilon_{\tilde{c}} = \eta + \pi B + \varepsilon_c + \varepsilon_{\tilde{c}}.$$

Taking conditional expectations, the employer can in theory calculate:

$$w^* = \alpha x + \delta T_0 \times \tilde{C},$$

where

$$\delta = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_c^2 + \sigma_{\tilde{c}}^2}.$$

While w^* is an optimal estimate of productivity (assuming $\pi = 0$), employers might instead make the following estimate:

$$\hat{w} = w^* + \beta_v T_v \times B + \beta_o T_o \times S,$$

where S is self-confidence,

$$S = B + \varepsilon_s, \text{ where } \varepsilon_s \sim N\left(0, \sigma_{\varepsilon_s}^2\right).$$

Thus, there are two additive biases here (added to w^*). There is a visual stereotype channel, refected in β_v (only present if there is visual communication). And there is the oral stereotype channel, which operates through β_0. If worker self-confidence affects employers' expectations of ability ($\beta_o > 0$), and if BP's are more self-confident, then beauty will indirectly affect employers' ability assessment through oral communication.

Now allow $\pi > 0$, so workers' own beliefs about their productivity are affected by their beauty. (This is distinct from the *self-confidence* channel, S, in that self-confidence does not bias a worker's own productivity assessment, C, it's just a separate 'winning' trait that is influenced by beauty.) In this case, the worker's beauty will bias the employer's assessment of w^* during oral interactions because the employer does not observe B but does observe $\tilde{C}$. This bias is $\delta\pi B$.

Notice that if an employer visually and orally interacts with the worker, he can undo this bias if he is aware of it by subtracting πB from $\tilde{C}$.

These components all shape the employer's productivity assessment. The employer may also engage in taste discrimination. So, the final wage is:

$$w = \hat{w} + D(B),$$

where $D(\cdot)$ is the pure Beckarian discrimination function. So, there are four biases operative here:

1. Worker's self-assessment bias, which affects employer assessment in the oral channel (may be undone visually). This bias operates through $\delta\pi B$.

2. Employer's beliefs about the productivity of Beauty. This operates through β_v in Visual interactions.

3. Employer's beliefs about the productivity of Self-confident people. This operates through $\beta_o S$ in Oral interactions.

4. Taste discrimination, which operates through $D(B)$ in Visual interactions.

In practice, the following measures of worker characteristics are key:

- $LNPROJECTED = \ln((15 \times 60)/PRACTICE)$. This is the predicted number of mazes to be completed given the worker's practice time. This is a resume variable
- $LNACTUAL$. This is the actual number of mazes completed
- $BEAUTY$. Standardized raters' assessment of worker's beauty
- $LNESTIMATED$. The worker's own estimate of his expected productivity (meant to measure confidence).

Table 4A

- Employers clearly put weight on past performance ($LNPROJECTED$)
- Beauty raises wages by 13 to 17 percent in all treatments where there is Visual contact.
- Beauty also raises wages by 13 percent in the Oral only treatment (O) where there is no Visual contact. This suggests that the oral stereotype is operative.
- Actual (subsequent) performance does not predict wages, suggesting that there is not another hidden channel by which predictors of actual performance are communicated to employers but not controlled here (for example, if we omitted $LNPROJECTED$, then $LNACTUAL$ would likely be significant).
- There is no evidence that employers prefer to pay higher wages to BP's when they have the opportunity to directly affect their earnings. That is, the Beauty premium in this setting appears to be arising through beliefs about productivity rather than a direct desire to make transfers.

Table 4b assesses the confidence channel

Confidence is $LNESTIMATED$. It is *not* reported to employers. So, the test is whether workers' own beliefs about their ability are somehow communicated to employers and affect employers' beliefs.

- As before Beauty directly affects wages in all cases where there is Visual or Oral contact.

- No compelling evidence of taste discrimination.
- Worker confidence affects wages in cases where there is Oral interaction but not where there is only Visual interaction (column (V)).
- Both the oral stereotype and self-confidence channel appear operative in treatment O. That is, Beauty affects wages directly and Confidence further affects wages (even though confidence is in part a function of Beauty). At some level, this is not entirely surprising since Confidence as constructed here would not be expected to be a sufficient statistic for all information about self-confidence communicated to the employer via Oral interactions.
- In theory Confidence ought to have a steeper slope in VO then O (because employers can increase the precision of the confidence signal by removing the beauty bias in the worker's own confidence estimate). This prediction is not upheld–but it seems like a low powered prediction.

Table 5 attempts to decompose the gross beauty premium into the confidence and residual channel

These channels don't have to be additive and the decomposition does not add up. But it appears that about 20 to 25 percent of the beauty premium can be 'explained' by conditioning on the worker's own beliefs about their productivity, communicated through oral, visual and face-to-face interactions. (Note that own $LNESTIMATED$ is only a proxy for these beliefs—not a latent measure of confidence.)

Table 6 estimates the 'everything' model. No shocking new conclusions here.

3.1.3 Concluding thoughts

1. Why are BP's more confident?
2. Given that beauty is unproductive in this setting, is it irrational for employers to pay a beauty premium?
3. Even though certain traits are immutable, can we learn about their causal effects ('no causation without manipulation' – Rosenbaum)?

4. What are advantages/disadvantages to lab experiments like Mobius/Rosenblat? To field experiments like Bertrand/Mullainathan?