



Detecting Discrimination

Author(s): James J. Heckman

Source: *The Journal of Economic Perspectives*, Vol. 12, No. 2, (Spring, 1998), pp. 101-116

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/2646964>

Accessed: 15/08/2008 23:53

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=aea>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

---

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

## Detecting Discrimination

James J. Heckman

**I**n the current atmosphere of race relations in America, the authors of the three main papers presented in this symposium are like persons crying “fire” in a crowded theater. They apparently vindicate the point of view that American society is riddled with racism and that discrimination by employers may account for much of the well-documented economic disparity between blacks and whites. In my judgement, this conclusion is not sustained by a careful reading of the evidence.

In this article, I make three major points. First, I want to distinguish market discrimination from the discrimination encountered by a randomly selected person or pair of persons at a randomly selected firm as identified from audit studies.

Second, I consider the evidence presented by the authors in the symposium, focusing for brevity and specificity on labor markets. It is far less decisive on the issue of market discrimination than it is claimed to be. Disparity in market outcomes does not prove discrimination in the market. A careful reading of the entire body of available evidence confirms that most of the disparity in earnings between blacks and whites in the labor market of the 1990s is due to the differences in skills they bring to the market, and not to discrimination within the labor market. This interpretation of the evidence has important consequences for social policy. While undoubtedly there are still employers and employees with discriminatory intentions, labor market discrimination is no longer a first-order quantitative problem in American society. At this time, the goal of achieving black economic progress is better served by policies that promote skill formation, like improving family environments,

■ *James J. Heckman is Henry Schultz Distinguished Service Professor of Economics, University of Chicago, Chicago, Illinois, and Senior Fellow, American Bar Foundation, Chicago, Illinois.*

schools and neighborhoods, not by strengthening the content and enforcement of civil rights laws—the solution to the problem of an earlier era.

Third, I want to examine the logic and limitations of the audit pair method. All of the papers in this symposium use evidence from this version of pair matching. However, the evidence acquired from it is less compelling than is often assumed. Inferences from such studies are quite fragile to alternative assumptions about unobservable variables and the way labor markets work. The audit method can find discrimination when in fact none exists; it can also disguise discrimination when it is present. These findings are especially troubling because the Equal Employment Opportunity Commission has recently authorized the use of audit pair methods to detect discrimination in labor markets (Seelye, 1997).

### **Discrimination: Definition and Measurement**

The authors of these papers focus on the question of whether society is color blind, not on the specific question of whether there is market discrimination in realized transactions. But discrimination at the individual level is different from discrimination at the group level, although these concepts are often confused in the literature on the economics of discrimination.

At the level of a potential worker or credit applicant dealing with a firm, racial discrimination is said to arise if an otherwise identical person is treated differently by virtue of that person's race or gender, and race and gender by themselves have no direct effect on productivity. Discrimination is a causal effect defined by a hypothetical *ceteris paribus* conceptual experiment—varying race but keeping all else constant. Audit studies attempt to identify racial and gender discrimination so defined for the set of firms sampled by the auditors by approximating the *ceteris paribus* condition.

It was Becker's (1957) insight to observe that finding a discriminatory effect of race or gender at a randomly selected firm does not provide an accurate measure of the discrimination that takes place in the market as a whole. At the level of the market, the causal effect of race is defined by the marginal firm or set of firms with which the marginal minority member deals. The impact of market discrimination is not determined by the most discriminatory participants in the market, or even by the average level of discrimination among firms, but rather by the level of discrimination at the firms where ethnic minorities or women actually end up buying, working and borrowing. It is at the margin that economic values are set. This point is largely ignored in the papers in this symposium.

This confusion between individual firm and market discrimination arises in particular in the audit studies. A well-designed audit study could uncover many individual firms that discriminate, while at the same time the marginal effect of discrimination on the wages of employed workers could be zero. This helps to explain the gap between audit-based estimates of discrimination and estimates

based on actual purchase prices that are discussed by Yinger. In fact, the audit evidence reported in all three papers in this symposium is entirely consistent with little or no market discrimination at the margin. Purposive sorting within markets eliminates the worst forms of discrimination. There may be evil lurking in the hearts of firms that is never manifest in consummated market transactions.

Estimating the extent and degree of discrimination, whether at the individual or the market level, is a difficult matter. In the labor market, for example, a worker's productivity is rarely observed directly, so the analyst must instead use available data as a proxy in controlling for the relevant productivity characteristics. The major controversies arise over whether relevant omitted characteristics differ between races, and between genders, and whether certain included characteristics systematically capture productivity differences or instead are a proxy for race or gender.

### **How Substantial is Labor Market Discrimination Against Blacks?**

In their paper in this symposium, Darity and Mason present a bleak picture of the labor market position of African-Americans in which market discrimination is ubiquitous. They present a quantitative estimate of the magnitude of estimated discrimination: 12 to 15 percent in both 1980 and 1990 using standard regressions fit on Current Population Survey and Census data. Similar regressions show that the black/white wage gap has diminished sharply over the last half century. Comparable estimates for 1940 show a black/white wage gap ranging from 30 percentage points, for men age 25–34, to 42 percentage points, men age 55–64. In 1960, the corresponding numbers would have been 21 percent and 32 percent, for the same two age groups; in 1970, 18 and 25 percent (U.S. Commission on Civil Rights, 1986, Table 6.1, p. 191). The progress was greatest in Southern states where a blatantly discriminatory system was successfully challenged by an external legal intervention (Donohue and Heckman, 1991; Heckman, 1990).

How should the residual wage gap be interpreted? As is typical of much of the literature on measuring racial wage gaps, Darity and Mason never precisely define the concept of discrimination they use. As is also typical of this literature, the phrase "human capital variable" is thrown around without a clear operational definition. The implicit definition of these terms varies across the studies they discuss. In practice, human capital in these studies has come to mean education and various combinations of age and education, based on the available Census and Current Population Survey (CPS) data. However, there is a staggering gap between the list of productivity characteristics available to economic analysts in standard data sources and what is available to personnel departments of firms. Regressions based on the Census and/or CPS data can typically explain 20 to 30 percent of the variation in wages. However, regressions based on personnel data can explain a substantially higher share of the variation in wages; 60–80 percent in professional labor markets (for example, see Abowd and Killingsworth, 1983). It is not idle speculation to claim

that the standard data sets used to estimate discrimination omit many relevant characteristics actually used by firms in their hiring and promotion decisions. Nor is it idle speculation to conjecture that disparity in family, neighborhood and schooling environments may account for systematic differences in unmeasured characteristics between race groups.

Consider just one well-documented source of discrepancy between Census variables and the productivity concepts that they proxy: the measurement of high school credentials. The standard Census and CPS data sources equate recipients of a General Equivalence Degree, or GED, with high school graduates. However, black high school certificate holders are much more likely than whites to receive GEDs (Cameron and Heckman, 1993), and a substantial portion of the widely trumpeted “convergence” in measured black educational attainment has come through GED certification. Thus, in 1987 in the NLSY data that Darity and Mason discuss, and Neal and Johnson (1996) analyze, 79 percent of black males age 25 were high school certified, and 14 percent of the credential holders were GED recipients. Among white males, 88 percent were high school certified, and only 8 percent of the white credential holders were GED certified. Given the evidence from Cameron and Heckman that GED recipients earn the same as high school dropouts, it is plausible that standard Census-based studies that use high school credentials to control for “education” will find that the wages of black high school “graduates” are lower than those of whites.

Most of the empirical literature cited by Darity and Mason takes Census variables literally and ignores these issues. The GED factor alone accounts for 1–2 percentage points of the current 12–15 percent black-white hourly wage gap. An enormous body of solid evidence on inferior inner city schools and poor neighborhoods makes the ritual of the measurement of “discrimination” using the unadjusted Census or Current Population Survey data a questionable exercise.

Darity and Mason bolster their case for rampant discrimination by appealing to audit pair evidence. They do not point out that audit pair studies have primarily been conducted for hiring in entry level jobs in certain low skill occupations using overqualified college students during summer vacations. They do not sample subsequent promotion decisions. They fail to point out that the audits undersample the main avenues through which youth get jobs, since only job openings advertised in newspapers are audited, and not jobs found through networks and friends (Heckman and Siegelman, 1993, pp. 213–215). Auditors are sometimes instructed on the “problem of discrimination in American society” prior to sampling firms, so they may have been coached to find what the audit agencies wanted to find. I have already noted that audit evidence does not translate into actual employment experiences and wages obtained by actors who purposively search markets.

Putting these objections to the side, what do the audits actually show for this unrepresentative snapshot of the American labor market? Table 1 presents evidence

*Table 1*  
**Outcomes From Major Audit Studies For Blacks**  
*(outcome: get job or not)*

<i>Number of Audits</i>	<i>Pair</i>	<i>(a) Both Get Job</i>	<i>(b) Neither Gets a Job</i>	<i>Equal Treatment a + b</i>	<i>White Yes, Black No</i>	<i>White No, Black Yes</i>
<b>Chicago*</b>						
35	1	(5) 14.3%	(23) 65.7%	80.0%	(5) 14.3%	(2) 5.7%
40	2	(5) 12.5%	(25) 62.5%	75.0%	(4) 10.0%	(6) 15.0%
44	3	(3) 6.8%	(37) 84.1%	90.9%	(3) 6.8%	(1) 2.3%
36	4	(6) 16.7%	(24) 66.7%	83.4%	(6) 16.7%	(0) 0%
42	5	(3) 7.1%	(38) 90.5%	97.6%	(1) 2.4%	(0) 0%
197	Total	(22) 11.2%	(147) 74.6%	85.8%	(19) 9.6%	(9) 4.5%
<b>Washington*</b>						
46	1	(5) 10.9%	(26) 56.5%	67.4%	(12) 26.1%	(3) 6.5%
54	2	(11) 20.4%	(31) 57.4%	77.8%	(9) 16.7%	(3) 5.6%
62	3	(11) 17.7%	(36) 58.1%	75.8%	(11) 17.7%	(4) 6.5%
37	4	(6) 16.2%	(22) 59.5%	75.7%	(7) 18.9%	(2) 5.4%
42	5	(7) 16.7%	(26) 61.9%	77.6%	(7) 16.7%	(2) 4.8%
241	Total	(40) 16.6%	(141) 58.5%	75.1%	(46) 19.1%	(14) 5.8%
<b>Denver**</b>						
18	1	(2) 11.1%	(11) 61.1%	72.1%	(5) 27.8%	(0) 0.0%
53	2	(2) 3.8%	(41) 77.4%	81.2%	(0) 0.0%	(10) 18.9%
33	3	(7) 21.2%	(25) 75.8%	97.0%	(1) 3.0%	(0) 0.0%
15	4	(9) 60.0%	(3) 20.0%	80.0%	(2) 6.7%	(2) 13.3%
26	9	(3) 11.5%	(23) 88.5%	100.0%	(0) 0.0%	(0) 0.0%
145	Total	(23) 15.8%	(103) 71.1%	86.9%	(7) 4.8%	(12) 8.3%

*Note:* Results are percentages; figures in parentheses are the relevant number of audits.

*Sources:* Heckman and Siegelman (1993).

\* This study was conducted by the Urban Institute.

\*\* Denver pair numbers are for both black and Hispanic audits. For the sake of brevity, I only consider the black audits. The Denver study was not conducted by the Urban Institute but it was conducted to conform to Urban Institute practice.

from three major audits in Washington, D.C., Chicago and Denver. The most remarkable feature of this evidence is the a + b column which records the percentage of audit attempts where black and white auditors were treated symmetrically (both got a job; neither got a job). In Chicago and Denver this happened about 86 percent of the time. The evidence of disparity in hiring presented in the last two columns of the table suggests only a slight preference for whites over minorities; in several pairs, minorities are favored. Only a zealot can see evidence in these data of pervasive discrimination in the U.S. labor market. And, as I will show in the next section, even this evidence on disparity has to be taken with a grain of salt, because it is based on the implicit assumption that the distribution of unobserved productivity is the same in both race groups.

Darity and Mason go on to dismiss the research of Neal and Johnson (1996) who analyze a sample of males who took an achievement or ability test in their early teens—specifically, the Armed Forces Qualifications Test (AFQT)—and ask how much of the gap in black-white wages measured a decade or so after the test was taken can be explained by the differences in the test scores.<sup>1</sup> It is remarkable and important that this early “premarket” measure of ability plays such a strong role in explaining wages measured a decade after the test is taken. This is as true for studies of white outcomes taken in isolation as it is for black-white comparisons. Their findings are important for interpreting the sources of black-white disparity in labor market outcomes.

The goal of Neal and Johnson is not to estimate the racial gap by holding other characteristics constant in the sense of a linear regression. Their estimated race effect is obtained from an equation with essentially only ability and race in it. Ability is allowed to operate both as a direct effect, as in a standard hedonic regression with productivity characteristics included, and as an indirect effect, operating through subsequent schooling, work experience and occupational choices. Therefore, they strip out of their wage regression all of the schooling, occupation and other post-adolescent choice variables used in predecessor studies and commonly employed in “human capital discrimination studies.” In so doing, they bypass the problem of determining the endogeneity of the included control variables.

The Neal-Johnson story is not about genetic determination. They demonstrate that schooling and environment can affect their measured test score. A huge body of evidence, to which the Neal-Johnson study contributes, documents that human abilities and motivations are formed early and have a decisive effect on lifetime outcomes; the evidence is summarized in Heckman (1995) and in Heckman, Lochner, Taber and Smith (1997). Not only is early ability an important predictor of later success for blacks or whites, it can be manipulated. Early interventions are far more effective than late ones because early skills and motivation beget later skills and motivation. As Heckman, Lochner, Taber and Smith document, however, successful early interventions can be quite costly.

The objections raised by Darity and Mason against the Neal-Johnson study are largely specious. For example, Rodgers and Spriggs (1996) miss the point of the Neal-Johnson article by “adjusting” the test score by a later variable, such as schooling. But ability is known to be an important determinant of schooling (Cawley, Heckman and Vtylail, 1998), so it should be no surprise that “adjusting” the score for later schooling eliminates an important component of ability and that adjusted scores play a much weaker role in explaining black-white differentials.<sup>2</sup>

<sup>1</sup> Specifically, Darity and Mason write: “This effort has uncovered one variable in one data set which, if inserted in an earnings regression, produces the outcome that nearly all of the black male-white male wage gap is explained by human capital and none by labor market discrimination.”

<sup>2</sup> The Rodgers and Spriggs comment (1997) on Neal-Johnson raises other red herrings. Their confused discussion of endogeneity of AFQT, and their “solution” to the problem end up with an “adjusted” AFQT measure that is poorly correlated with the measured AFQT, and so is a poor proxy for black ability.

Only one point raised by Darity and Mason concerning Neal and Johnson is potentially valid—and this is a point made by Neal and Johnson in their original article. Black achievement scores may be lower than white scores not because of the inferior environments encountered by many poor blacks, but because of expectations of discrimination in the market. If black children and their parents face a world in which they receive lower rewards for obtaining skills, they will invest less if they face the same tuition costs as whites. Poor performance in schools and low achievement test scores may thus be a proxy for discrimination to be experienced in the future.

There is solid empirical evidence that expectations about rewards in the labor market influence human capital investment decisions; for example, the reward to skills held by black workers increased following the passage of the 1964 Civil Rights Act, and a rapid rise in college enrollment of blacks followed (Donohue and Heckman, 1991). But the difficulty with the argument in this context is that it presumes that black parents and children operate under mistaken expectations about the present labor market. Although it was once true that the returns to college education were lower for blacks than for whites (Becker, 1957; U.S. Civil Rights Commission, 1986), the return to college education for blacks was higher than the return for whites by the mid-1970s, and continues to be higher today. Some parallel evidence presented by Johnson and Neal (1998) shows that the returns to (coefficient on) AFQT scores for black males in an earnings equation are now as high or higher than those for whites, although they used to be lower in the pre-Civil Rights era. Given the greater return for blacks to college education and ability, it seems implausible to argue that a rational fear of lower future returns is currently discouraging black formation of skills.

Ability as it crystallizes at an early age accounts for most of the measured gap in black and white labor market outcomes. Stricter enforcement of civil rights laws is a tenuous way to improve early childhood skills and ability.<sup>3</sup> The weight of the evidence suggests that this ability and early motivation is most easily influenced by enriching family and preschool learning environments and by improving the quality of the early years of schooling.

## **The Implicit Assumptions Behind the Audit Method**

The method of audit pairs operates by controlling for systematic observed differences across pairs. It does this by attempting to create two candidates for jobs or loans who are “essentially” the same in their paper qualifications and personal characteristics, and then comparing their outcomes in their dealings with the same firm. Averaging over the outcomes at all firms for the same audit pair produces an

<sup>3</sup> However, nothing I have said vindicates abolishing these laws. They have important symbolic value and they addressed and solved an important problem of blatant discrimination in the American South.

estimate of the discrimination effect. An average is often taken over audit pairs as well to report an “overall” estimate of discrimination. More sophisticated versions of the method will allow for some heterogeneity in treatment among firms and workers or firms and applicants.

One set of difficulties arise, however, because there are sure to be many unobserved variables. As noted by Heckman and Siegelman (1993), given the current limited state of knowledge of the determinants of productivity within firms, and given the small pools of applicants from which matched pairs are constructed that are characteristic of most audit studies, it is unlikely that all characteristics that might affect productivity will be perfectly matched. Thus, the implicit assumption in the audit pair method is that controlling for some components of productivity and sending people to the same firm will reduce the bias below what it would be if random pairs of, say, whites and blacks were compared using, for example, Census data. The implicit assumption that justifies this method is that the effect of the unobserved characteristics averages out to zero across firms for the same audit pair.

However, the mean of the differences in the unobserved components need not be zero and assuming that it is begs the problem. Nowhere in the published literature on the audit pair method will you find a demonstration that matching one subset of observable variables necessarily implies that the resulting difference in audit-adjusted treatment between blacks and whites is an unbiased measure of discrimination—or indeed, that it is even necessarily a better measure of discrimination than comparing random pairs of whites and blacks applying at the same firm or even applying to different firms. This argument is stated more formally in the Appendix. Here I present an intuitive discussion.

Consider the following example. Suppose that the market productivity of persons is determined by the sum of two productivity components. These two productivity components are distributed independently in the population so their values are not correlated with each other. Both factors affect employer assessments of employee productivity.<sup>4</sup> Suppose further that average productivity of the sum is the same for both whites and blacks; however, blacks are more productive on average on one component while whites are more productive on average on the other. Now consider an audit pair study that equates only the first component of productivity and equates firm effects by sending the audit pair to the same firm. Under these conditions, the audit estimator is biased toward a finding of discrimination, since in this example, only the characteristic which makes black productivity look relatively high is being used to standardize the audit pair. The condition of zero mean of unobservable productivity differences across race groups is not especially compelling and requires a priori knowledge that is typically not available.

Now consider the case in which the observed and unobserved components of

<sup>4</sup> They need not be perfectly observed by employers, but may only be proxied. However, it is easiest to think of both components as fully observed by the employer, but that the observing economist has less information.

productivity are dependent. In this case, making the included components as alike as possible may accentuate the differences in the unobserved components. As a result, it can increase the bias over the case where the measured components are not aligned.

A related crucial assumption behind the audit method is that it typically assumes that the outcome being studied is a linear function of the relevant variables or the outcome can be transformed to be so; that is, more skill leads in a linear way to a greater chance of employment or higher wages. However, in more general models, the decision to offer a job is a nonlinear function of the relevant characteristics. This case arises in many models of the employment decision when productivity must be above a threshold for a job offer to be made. Many models of firm hiring decisions assume that workers must possess minimum qualifications to get a job offer. A similar model is often used to explain the credit applications analyzed by Ladd; only applicants with credentials above a threshold get credit. In this case, making one set of variables more alike for an audit pair may also make the bias in estimating discrimination worse if the distributions of unobserved characteristics are not the same across race/gender groups even if the means are.

To see how this works, consider the intuition of a labor market example, patterned after the analysis of Heckman and Siegelman (1993). Again consider a case where productivity is divided into two components: one which is observable and set by the audit designer as part of deciding on the paper qualifications and personal characteristics that each member of the audit pair should possess; and the other which is unobservable to the audit study, but is at least somewhat visible to the prospective employer and acted on in hiring or credit decisions. Both are determinants of perceived (by the firm) productivity. Let us further assume that: a) the omitted and included components are statistically independent for both race groups; and b) that the mean of the unobserved productivity variable is the same for whites and blacks. Assumption b is a strong one, and it is the key identifying assumption of the audit method as currently practiced. Nothing guarantees that it will be satisfied. Under these assumptions, however, the audit methods would certainly eliminate, or at least not increase, bias in measuring racial discrimination if the measured outcomes are simple sums of productivity characteristics.

However, this conclusion only holds if the firm's treatment of the worker is linear in the attributes of the productivity level. Suppose, instead, that firms hire workers based on a common cutoff value, so that the more stringent the cutoff, the less likely is the auditor to be hired.<sup>5</sup> In the absence of discrimination, the cutoff for blacks will be the same as the cutoff for whites. As long as the distribution of unobserved productivity attributes is the same for each race group, hiring rates should be the same for all levels of observed characteristics set by the audit designer. Evidence of asymmetry of treatment, as manifest in the last two columns of Table

<sup>5</sup> It is straightforward to allow this cutoff value to vary among firms but for simplicity I assume it does not. If these differences in cutoff values are random across firms and pairs, the analysis in the text goes through without any essential modification.

1, would be evidence that the firm cutoff rates are different among race groups, and thus that a different standard is applied to whites than to blacks.

Suppose, however, that whites are more heterogeneous in their observed productivity characteristics; as a result, the distribution of their unobserved productivity characteristics have different tail areas. In this case, the ability of the audit pair method to detect discrimination will depend on the level at which the observed level of productivity was standardized in the audit design.

Here is the intuitive argument. Say that blacks and whites face the same productivity threshold for hiring, and have the same mean characteristics, but that whites are more heterogeneous in their productivity characteristics. For simplicity, assume that these characteristics are symmetrically distributed. Imagine that the audit team members are applying for a job where their observed characteristics, as set by the audit design, are such that they are highly qualified for the job. Then the prospective employer, who has insight into the characteristics unobserved by the designer of the audit study, will tend to hire more blacks than whites. In effect, the longer left-hand tail of unobserved characteristics for whites works against them.<sup>6</sup> Precisely the opposite bias appears if black unobservables are more dispersed than white unobservables. Now consider applying for a job where the observed characteristics, as set by the audit design, are such that the auditors have only a moderate chance of qualifying for the job. Thus most of the applicants of both race groups are below the threshold. Then the prospective employer, who has insight into the characteristics unobserved by the designer of the audit study, will tend to hire more whites than blacks if white traits are more dispersed. In effect, the longer right-hand tail of the distribution of unobserved characteristics for whites works for them.<sup>7</sup>

Another way to state this point is to think of pairing up black and white high jumpers to see if they can clear a bar set at a certain height. There is no discrimination, in the sense that they both use the same equipment and have the bar set at the same level. Suppose now that the chance of a jumper (of any race) clearing the bar depends on two additive factors: the person's height and their jumping technique. We can pair up black and white jumpers so that they have identical heights, but we can't directly observe their technique. Let us make the generous assumption, implicit in the entire audit literature, that the mean jumping technique is equal for the two groups. Then, if the variance of technique is also the same for white and black high-jumpers, we would find that the two racial groups are equally likely to clear the bar. On the other hand, if the variance differs, then whether the black or white pair is more likely to clear the bar will depend on how the bar is set, relative to their common height, and which racial group has a higher variance in jumping technique. If the bar is set at a low level so that most people of the given height are likely to clear the bar, then the group with the lower variance will be

<sup>6</sup> If the black distribution of unobservables is more heterogeneous, whites are favored. This interpretation is consistent with the evidence in Table 1.

<sup>7</sup> This case seems to rationalize the evidence in Table 1.

more likely to clear the bar. If the bar is set at a very high level relative to the given height, then the group with a higher variance in jumping technique will be more likely to clear the bar. A limitation of the audit method is readily apparent from this analogy: there is no discrimination, yet the two groups have different probabilities of clearing the bar.<sup>8</sup> And if there is discrimination—that is, the bar is being set higher for blacks—the differential dispersion in the unobserved component could still cause the minority group to clear the bar more often. The method could fail to detect discrimination when it does exist.

Thus, depending on the distribution of unobserved characteristics for each race group and the audit standardization level, the audit method can show reverse discrimination, or equal treatment, or discrimination, even though blacks and whites in this example are subject to the same cutoff and face no discrimination. The apparent bias depends on whether the level of qualifications set by the audit designer makes it more or less likely that the applicant will receive the job, and the distribution of variables that are unobservable to the audit design. The apparent disparity favoring Washington whites in Table 1 may be a consequence of differences in unobserved characteristics between blacks and whites when there is no discrimination.

Even more disturbing, suppose that there is discrimination against blacks, so the productivity cutoff used by firms is higher for blacks than whites. Depending on the audit designer's choice of what level of qualifications are given to the auditors, the audit study can find no discrimination at all. However, whether the qualifications make it relatively likely or unlikely to get the job is a fact rarely reported in audit studies. This argument is made more precise in the Appendix where two examples are given, where discrimination or reverse discrimination is detected when none exists, and discrimination is disguised when it exists.

Making audit pairs as alike as possible may seem an obviously useful step, but it can greatly bias the inference about average discrimination or discrimination at the margin. Intuitively, by taking out the common components that are most easily measured, differences in hiring rates as monitored by audits arise from the idiosyncratic factors, and not the main factors, that drive actual labor markets. These examples highlight the fragility of the audit method to untested and unverifiable assumptions about the distributions of unobservables. Similar points arise in more general nonlinear models that characterize other employment decision rules.

## The Becker Model

The papers in this symposium make the erroneous claim that in Becker's (1957) model, market discrimination disappears in the long run. It need not. Entrepreneurs

<sup>8</sup> I owe this analogy to Alan Krueger. This analogy also shows how artificial the audit studies are because one would expect to find athletes choosing their sports based on their chances of success, as in the purposive search in the labor market discussed earlier.

can consume their income in any way they see fit. If a bigoted employer prefers whites, the employer can indulge that taste as long as income is received from entrepreneurial activity, just as a person who favors an exotic ice cream can indulge that preference by being willing to pay the price. Only if the supply of entrepreneurship is perfectly elastic in the long run at a zero price, so entrepreneurs have no income to spend to indulge their tastes, or if there are enough nonprejudiced employers to hire all blacks, will discrimination disappear from Becker's model.

However, even if the common misinterpretation of Becker's model is accepted, it is far from clear that the prediction of no or little discrimination in the U.S. labor market in the long run is false. The substantial decline over the past 50 years in wage differentials between blacks and whites may well be a manifestation of the dynamics of the Becker model. It may take decades for the effects of past discrimination in employment and schooling as it affects current endowments of workers to fade out of the labor market. But the evidence from the current U.S. labor market is that discrimination by employers alone does *not* generate large economic disparities between blacks and whites.

## Appendix

### Implicit Identifying Assumptions In The Audit Method

Define the productivity of a person of race  $r \in \{1, 0\}$ , at firm  $f$ , with characteristics  $\bar{X} = (X_1, X_2)$  as  $P(\bar{X}, r, f)$ .  $r = 1$  corresponds to black;  $r = 0$  corresponds to white. Assume that race does not affect productivity so we may write  $P = P(\bar{X}, f)$ . The treatment at the firm  $f$  for a person of race  $r$  and productivity  $P$  is  $T(P(\bar{X}, f), r)$ . Racial discrimination exists at firm  $f$  if

$$T(P(\bar{X}, f), r = 1) \neq T(P(\bar{X}, f), r = 0).$$

As noted in the text, audit methods monitor discrimination at randomly selected firms within the universe designated for sampling, not the firms where blacks are employed.

The most favorable case for auditing assumes that  $T$  (or some transformation of it) is linear in  $f$  and  $\bar{X}$ . Assume for simplicity that  $P = X_1 + X_2 + f$  and  $T(P, r) = P + \gamma r$ . When  $\gamma < 0$  there is discrimination against blacks.  $\gamma$  may vary among firms as in Heckman and Siegelman (1993).<sup>9</sup> For simplicity suppose that all firms are alike. Audit methods pair racially dissimilar workers in the following way: they match

<sup>9</sup> For simplicity, assume that  $\gamma$  is the same across all firms. Alternatively, assume that it is distributed independently of  $\bar{X}$  and  $f$ .

some components of  $X$  and they sample the same firms. Let  $P_1^*$  be the standardized productivity for the black member of the pair;  $P_0^*$  is the standardized productivity for the white member. If  $P_0^* = P_1^*$ ,

$$T(P_1^*, 1) - T(P_0^*, 0) = \gamma.$$

When averaged over firms, the average treatment estimates the average  $\gamma$ .

Suppose that standardization is incomplete. We can align the first coordinate of  $X$  at  $\{X_1 = X_1^*\}$  but not the second coordinate,  $X_2$ , which is unobserved by the auditor but acted on by the firm.  $P_1^* = X_1^* + X_2^1$  where  $X_2^1$  is the value of  $X_2$  for the  $r = 1$  member and  $P_0^* = X_1^* + X_2^0$ . In this case

$$T(P_1^*, 1) - T(P_0^*, 0) = X_2^1 - X_2^0 + \gamma.$$

For averages over pairs to estimate  $\gamma$  without bias, it must be assumed that  $E(X_2^1) = E(X_2^0)$ ; i.e., that the mean of the unobserved productivity traits is the same. This is the crucial identifying assumption in the conventional audit method. Suppose that this is true so  $E(X_2^1) = E(X_2^0) = \mu$ . Then the pair matching as in the audit method does not increase bias and in general reduces it over comparisons of two  $X_1$ -identical persons at two randomly selected firms. Under these conditions, bias is lower than if two randomly chosen auditors are selected at the same firm if  $E(X_2^1) \neq E(X_2^0)$ .

However, the decision rule to offer a job or extend credit often depends on whether or not the perceived productivity  $P$  exceeds a threshold  $c$ :

$$T = 1 \text{ if } P \geq c$$

$$T = 0 \text{ otherwise.}$$

In this case, the audit pair method will still produce bias even when it does not when  $T$  is linear in  $X$  and  $f$ , unless the *distributions* of the omitted characteristics are identical in the two race groups. Suppose that  $P = X_1 + X_2$ .  $X_2$  is uncontrolled. Then assuming no discrimination ( $\gamma = 0$ )

$$T(P_1^*, 1) = 1 \text{ if } X_1^* + X_2^1 + f \geq c$$

$$= 0 \text{ otherwise}$$

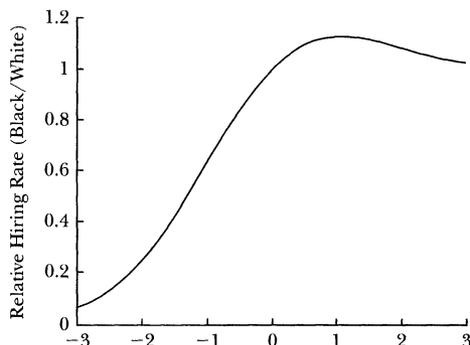
$$T(P_0^*, 0) = 1 \text{ if } X_1^* + X_2^0 + f \geq c$$

$$= 0 \text{ otherwise.}$$

Even if the distributions of  $f$  are identical across pairs, and  $f$  is independent of  $X$ , unless the *distributions* of  $X_2^1$  and  $X_2^0$  are identical,  $\Pr(T(P_1^*, 1) = 1) \neq$

Figure 1

**Relative Hiring Rate as a Function of the Level of Standardization. Blacks Have More Dispersion. Threshold Hiring Rule: No Discrimination Against Blacks Normally Distributed Unobservables**



$X_1^*$  = level of standardization

$X_2^1, X_2^0$  normal

$$E(X_2^1) = E(X_2^0) = 0; \text{Var}(X_2^1) < \text{Var}(X_2^0)$$

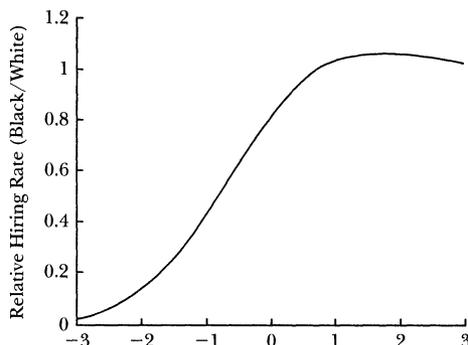
$$\text{Relative Hiring Rate} = \frac{\Pr(T(P_1^*, 1) = 1)}{\Pr(T(P_0^*, 0) = 1)}$$

$$\text{Var}(X_2^0) = 2.25 \text{Var}(X_2^1) = 1$$

$$c_1 = c_0 = 0$$

Figure 2

**Relative Hiring Rate as a Function of the Level of Standardization. Blacks Held to Higher Standard; Blacks Have More Dispersion. Threshold Hiring Rule: No Discrimination against Blacks Normally Distributed Unobservables**



$X_1^*$  = level of standardization

$X_2^1, X_2^0$  normal

$$E(X_2^1) = E(X_2^0) = 0; \text{Var}(X_2^1) < \text{Var}(X_2^0)$$

$$\text{Relative Hiring Rate} = \frac{\Pr(T(P_1^*, 1) = 1)}{\Pr(T(P_0^*, 0) = 1)}$$

$$\text{Var}(X_2^0) = 2.25 \text{Var}(X_2^1) = 1$$

$$c_1 = 0.25, c_0 = 0$$

$\Pr(T(P_0^*, 0) = 1)$  for most values of the standardization level  $X_1^*$ . The right tail area of the distribution governs the behavior of these probabilities. This implies that even if blacks and whites face the same cutoff value, and in this sense are treated without discrimination in the labor market, even if the means of the distributions of unobservables are the same across race group, if the distributions of the unobservables are different, their probabilities of being hired will differ and will depend on the level of standardization used in the audit study—something that is rarely reported. The pattern of racial disparity in Table 1 may simply be a consequence of the choice of the level of standardization in those audits, and not discrimination.

Worse yet, suppose that the cutoff  $c = c_1$  for blacks is larger than the cutoff  $c = c_0$  for whites so that blacks are held to a higher standard. Then depending on the right tail area of  $X_2^1$  and  $X_2^0$ , the values of  $c_1$  and  $c_0$ , and the level of standardization  $X_1^*$ ,

$$\Pr(T(P_1^*, 1) = 1) \cong \Pr(T(P_0^*, 0) = 1).$$

In general, only if the *distributions* of  $X_2^1$  and  $X_2^0$  are the same for each race group,

will the evidence reported in Table 1 be informative on the level of discrimination in the universe of sampled firms.

Figures 1 and 2 illustrate these two cases for  $X_2^1$  and  $X_2^0$  normally distributed (and independent of each other) where  $X_1^*$  is the level of audit standardization and firms are standardized to have  $f = 0$ .<sup>10</sup> In Figure 1 there is no discrimination in the market. Yet the black hire rate falls short of the white rate if the standardization rate is  $X_1^* < 0$ , and the lower the value of  $X_1^*$ , the greater the shortfall. In Figure 2, which is constructed for a hypothetical economy where there is discrimination against blacks, for high standardization rates, audits would appear to reveal discrimination *in favor* of blacks when in fact blacks are being held to a higher standard. The evidence in Table 1 is intrinsically ambiguous about the extent of discrimination in the market. For further discussion, see Heckman and Siegelman (1993).

■ *This research is supported by a grant from the American Bar Foundation. I thank Mark Killingsworth for references and Alan Krueger, Derek Neal and Timothy Taylor for valuable comments.*

<sup>10</sup> Allowing  $f$  to vary but assuming it is normal mean zero and variance  $\sigma_f^2$  does not change the qualitative character of these calculations assuming that  $f$  is distributed independently of the characteristics.

## References

- Abowd, John, and Mark Killingsworth, "Sex, Discrimination, Atrophy, and the Male-Female Wage Differential," *Industrial And Labor Relations Review*, Fall 1983, 22:3, 387-402.
- Becker, Gary, *The Economics of Discrimination*. Chicago: University of Chicago Press, 1957.
- Cameron, Stephen, and James Heckman, "The Nonequivalence of High School Equivalents," *Journal of Labor Economics*, 1993, 11:1, pt1, 1-47.
- Cawley, John, James Heckman, and Edward Vytlačil, "Cognitive Ability and the Rising Return to Education," NBER working paper 6388, January 1998.
- Donohue, John, and James Heckman, "Continuous vs. Episodic Change: The Impact of Affirmative Action and Civil Rights Policy on The Economic Status of Blacks," *Journal of Economic Literature*, December 1991, 29:4, 1603-43.
- Heckman, James, "The Central Role of the South in Accounting For The Economic Progress of Black Americans," Papers and Proceedings Of The American Economic Association, May 1990.
- Heckman, James, "Lessons From the Bell Curve," *Journal of Political Economy*, 1995, 103:5, 1091-1120.

**Heckman, James, and Peter Siegelman**, "The Urban Institute Audit Studies: Their Methods and Findings." In M. Fix and R. Struyk, eds. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Urban Institute, Fall 1993.

**Heckman, James, Lance Lochner, Christopher Taber, and Jeffrey Smith**, "The Effects of Government Policy on Human Capital Investment and Wage Inequality," *Chicago Policy Review*, Spring 1997, 1:2, 1-40.

**Johnson, William R., and Derek Neal**, "Basic Skills and the Black-White Earnings Gaps." In Jencks, Christopher, and Meredith Phillips, eds. *The Black-White Test Score Gap*. Washington, D.C. Brookings, 1998.

**Neal, Derek, and William Johnson**, "The Role of Premarket Factors in Black-White Wage Dif-

ferences," *Journal of Political Economy*, 1996, 104:5, 869-95.

**Rodgers III, William, and William Spriggs**, "What Does AFQT Really Measure: Race, Wages, Schooling and the AFQT Score," *The Review of Black Political Economy*, Spring 1996, 24:4, 13-46.

**Rodgers III, William, William E. Spriggs, and Elizabeth Waaler**, "The Role of Premarked Factors in Black-White Differences: Comment," Unpublished Manuscript, College of William and Mary, May 25, 1997.

**Seelye, Katherine**, "Employment Panel To Send People Undercover to Detect Bias in Hiring," *New York Times*, Sunday, December 7, 1997, p. 22.

**U.S. Commission on Civil Rights**, *The Economic Progress of Black Men in America*, Clearinghouse Publication 91, 1986.