NBER WORKING PAPER SERIES

MEASURING DISPARATE IMPACTS AND EXTENDING DISPARATE IMPACT DOCTRINE TO ORGAN TRANSPLANTATION

Robert Bornholz
James J. Heckman

Working Paper 10946
http://www.nber.org/papers/w10946

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2004

Robert Bornholz is an Associate at the Center for Program Evaluation at the Harris School of Public Policy at the University of Chicago. James J. Heckman is Professor of Economics at the University of Chicago and Senior Fellow of the American Bar Foundation. Discussions with Joseph Gastwirth and Robert Gibbons greatly improved this comment. Remarks by Richard Epstein, Joseph Gelb, and Amy Wax were also helpful. The American Bar Foundation provided generous financial support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Econmic Research.

Measuring Disparate Impacts and Extending Disparate Impact Doctrine to Organ Transplantation
Robert Bornholz and James J. Heckman
NBER Working Paper No. 10945
December 2004
JEL No. K32

ABSTRACT

This paper examines the economic and statistical foundations of proposed tests for discrimination.

We focus on extension of disparate impact doctrine to new domains.

Robert Bornholz
The University of Chicago
klmrob@lily.src.uchicago.edu

James J. Heckman
Department of Economics
The University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
j-heckman@uchicago.edu

There are widespread racial disparities in employment, access to credit, health, and education. Many claim that the original cause is slavery and the subsequent Jim Crow regime of lawful segregation. However disparities persist forty years after the Civil Rights Act of 1964 marked the end of Jim Crow. One segment of opinion holds that continuing racial discrimination prevents progress by African Americans and discrimination law is still a useful instrument to fashion remedies. Because disparate impact liability is based on an effects test rather than proof of discriminatory intent, it appears to be a plausible instrument for redressing disparity itself wherever it may be found.

Ian Ayres (2001) has advocated extending the "discrimination" paradigm beyond the original employment domain; broadening the view of what can constitute "discrimination;" and heightening the sensitivity of tests to set off the alarm that an unjustified act of discrimination has occurred (Ayres, 2001). His advocacy has been influential in public debate and has been persuasive enough to induce defendants to settle some large cases. Without winning a final court showdown, his doctrines are likely to transform distribution practices of the automobile industry. Ayres was successful in using his disparate impact doctrine to persuade the organization that administers the kidney transplantation regulatory regime to change its practices. The importance of these issues has inspired us to take a fresh look at the disparate impact doctrine.

In a recent conference paper, Ayres (2004) sets forth his theory of disparate impact. A view of that paper provides us with the occasion to take a critical look at his theory and tests for disparate impact. His proposed test is actually a test for disparity and is irrelevant for settling the issues raised by disparate impact doctrine as the Supreme Court has stated it. Here we ask whether there is a well-defined economically grounded theory of disparate impact consistent with decided cases. Are there sound ways of measuring disparate impact, and is it wise to extend disparate impact litigation to most areas of economic and social action? We first discuss disparate impact doctrine as it has evolved in the area of employment discrimination.

1 Origin of Disparate Impact Analysis in Employment Discrimination

To understand the distinction between racial disparity in some outcome and the concept

of "disparate impact" of a practice for selecting persons for participation in a productive activity, or rewarding them, it is useful to look at the origin of this concept in employment discrimination law. The Civil Rights Act of 1964 forbids intentional discrimination or "disparate treatment." In the *Griggs* case, the Supreme Court held that liability could be established without a finding of intent to discriminate. But existence of disparity of some outcome like hiring or wages could not alone be the basis for liability. The court focused on the effect of a "practice" used by the defendant more than on the mere existence of a disparity.

> Congress did not intend by Title VII. . . to guarantee a job to every person regardless of qualification. . . Discriminatory preference for any group, minority or majority, is precisely and only what Congress has proscribed. What is required by Congress is the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification. . . . The Act proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation. The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited. (*Griggs v. Duke Power Co.* 401 U.S. 424, 431 [1971])

Not all disparities are the result of intentional discrimination or a discriminatory practice in the meaning of *Griggs*. A practice may have a disparate impact on minorities but be justified as necessary to carry on the employer's business. There may be no alternative practice that produces an equal level of output with lower disparate impact and lower cost. This is the defense of business necessity.

In *McDonnell Douglas v. Green*, 411 U.S. 793 (1973), the Supreme Court took the opportunity to define a framework for disparate treatment in the light of the analysis it had established in *Griggs* for disparate impact. Showing that the employer treats similarly qualified employees or applicants differently establishes a prima facie disparate treatment case. The employer may rebut this claim by showing that there is a nonracial reason for the employer's hiring or firing decision. The plaintiff may then attempt to show that this articulated reason is a mere pretext for discrimination.

The *McDonnell Douglas* court distinguished *Griggs* in a way that points to the accountability of employees as well as the employer.

> *Griggs* was rightly concerned that childhood deficiencies in the education and background of minority citizens, resulting from forces beyond their control, not be allowed to work a cumulative and invidious burden on citizens for the remainder of their lives. (*McDonnell Douglas v. Green* 411 U.S. 793, 806 [1973])

The employee in question, however, had "engaged in a seriously disruptive act" against McDonnell Douglas (*Id*. at 806). This analysis looks at the employment relationship as potentially shaped by the actions of both parties. This might include the negotiation of terms of employment, or the initiative of the employee/applicant in searching for the best opportunity for deploying his or her skills. Decisions, practices, and policies have special meaning in the case of matching people with organizations to engage in productive activity. The consensus position as stated by a unanimous Supreme Court in *McDonnell Douglas* is that employee productivity is the ground for the legitimacy of selection methods.

Subsequent cases articulated a three-step analysis for disparate impact claims. To establish a prima facie case, the plaintiff must show that the facially neutral employment practice had a significantly discriminatory impact. If that showing is made, the employer must then demonstrate that any given requirement has a manifest relationship to the employment in question. Even then the plaintiff may prevail if he can prove that other selection devices with less discriminatory effect would equally serve the employer's legitimate business needs (*Connecticut v. Teal*, 457 U.S. 440, 447, 448 [1982]; *Dothard v. Rawlinso*n, 433 U.S. 321, 329 [1977]).

The statements of the three steps usually made in commentaries obscure similarities of disparate impact and disparate treatment analysis. The actual statement of the third step in *Teal* and some other cases is "however, the plaintiff may prevail, if he shows that the employer was using the practice as a mere pretext for discrimination"—the same third step test as for disparate treatment in *McDonnell Douglas*. If the employer is not using a more profitable practice that has less disparate impact, one may infer a discriminatory purpose.

In his conference paper, and in his book (2001), Ayres extends this body of disparate impact doctrine to new domains. His extensions are not straightforward nor are his criteria. We now turn to his application of this doctrine to organ transplantation.

2 Ayres' Impact on Organ Transplantation Practice

Ayres writes, "Organ transplantation is a natural place to study methods of testing for disparate impacts." (Ayres, 2004). A closer look will show that this "test case" for extending disparate impact doctrine really demonstrates why that extension is unwise. It is far from clear that the reform brought about by his analysis—overriding antigen mismatch using immunosuppressant drugs—is justified in medical terms. His narrow focus on disparate impact in transplantations misses the wider picture of the sources of disparity in health status between blacks and whites.

Ayres' policy intervention in the kidney transplant arena is an example of the myopic search for disparate impacts and "discrimination" that he recommends be undertaken everywhere. Seeing a racial disparity in kidney transplants performed, he assumes this must be due to discrimination. Rejecting intentional bias as implausible, he finds the culprit in the disparate impact of one practice, antigen matching. Antigen matching does indeed select more whites as being good biological matches to receive donated kidneys. One reason is that while a disproportionately (compared to the ratio in general population) greater number of end stage renal failure patients are black, a disproportionately lower number of kidney donors are black. "[W]hile blacks constitute nearly 13 percent of the general population, 34 percent of ESRD (End Stage Renal Disease) patients are black" (Ayres, 2001, 170).

Ayres makes two arguments against the antigen matching regime, one technological, the other normative:

> Advances in the use of drugs that effectively suppress immune responses have dramatically altered the impact of antigen matching: the likelihood of graft survival may now be relatively independent of the degree of antigen matching. . .

> Normatively, we argue that the equitable claims of black dialysis patients for cadaveric kidneys outweigh the marginal improvements in transplant outcomes associated with antigen matching under the old regime. (Ayres, 2001, 171-172)

For Ayres, just because an antigen match to a white recipient may have a biologically more productive outcome—a longer period before rejection and less need for immunosuppressant drugs with side effects—does not mean the less "efficient" transplants should not be performed.

Biological productivity is the standard implicit in the original United Network for Organ Sharing (UNOS) antigen matching criteria. It is not clear what other standard might be applied. In the employment context, the appeal to business necessity implies that the appropriate objective function is the firm's profitability in a competitive marketplace. If organ transplantation were "deregulated" and opened to "free market" enterprise would the profitability or shareholder wealth of "transplant enterprises" be the proper standard? As the framework in Appendix A makes clear, the objective function used to define productivity plays an important role in determining appropriate allocations of treatment to persons.

Ethicists challenge the right of economists to speak about standards in such delicate domains, pointing to the specter of wealth maximization leading to firms dedicated to catering to the whims of rich clients. What criteria should be used to determine productivity once we move from the employment arena where a profit standard is clearly appropriate? If patients could buy kidneys in an open market, some individuals who could not find well-matched organs would buy the best available but still badly matched organ and repeat the process when that kidney failed. If personal preferences can defy nature, one could argue that a social aggregate normative welfare function should be free to defy nature for the sake of racial "equality" as well. Ayres shows his readiness to engage in ad hoc balancing of conflicting objectives when he recognizes six point or perfect antigen matches as legitimately productive, but claims that there should be no preference for partial matches over no matches at all, both being treated by immunosuppressant drugs.

An important feature of the kidney transplant system is that while transplants generally work better for whites (given current donor-recipient imbalances), dialysis works better for blacks. Transplants still may be superior to dialysis for many blacks, but for some blacks, dialysis is better than a transplant. Case by case selection for matches might proceed by evaluating the best medical

"technology" to use for each patient, given available donor matches. Retrospective disparate impact analysis might then proceed by developing models of the productive basis for these matching choices and forecasting their medical outcomes.

Based on a standard of maximizing days of pain-free or useful life, one could determine whether there were "too many" white or black transplants. If blacks were being turned away for transplant matches while a great number of bad matches to whites were being performed, that would be a sign that there was discrimination against blacks. The years of life productivity standard would say that transplants should be performed on members of both groups (black and white), working in sequence from best matches to worst, so that the quality of match (and hence outcome) of the marginal black (the person who is just treated) equals the quality of match of the marginal white. Appendix A develops this analysis more formally.

If whites generally matched more successfully, we would expect to see more white matches than black matches. Ayres (2001), however, makes the claim that this productivity standard is not normatively correct. Even if transplants to all blacks survive five years and transplants to all whites survive ten years, there still should be transplants to blacks—assuming that transplant, while it lasts, does yield a better quality of life than dialysis. If this policy were implemented, the ex post disparate impact analysis would show productive whites being turned away in favor of less productive blacks, indicating discrimination against whites.

The years of life productivity standard was roughly the way the regime worked until Ayres persuaded UNOS to award points for "rare" matches, essentially bonus points to lower the weight given by the antigen matching regime to give African Americans a better chance of receiving transplants. An effect of the policy change has been more black deaths due to transplant rejection. Medicaid provides reimbursement for only two years of drug treatments to suppress rejection. When black Medicaid payments ceased, they went off drug treatment and died.

Attacking one part of the system in the guise of disparate impact doctrine, Ayres ignored the rest of the system. One organization enacted the antigen matching regime and could change it when the disparate impact was called to its attention. But changes in one program were not accompanied by changes in complementary programs. Focusing on the most proximate practice "causing" disparity is likely to leave other causes untouched and possibly to distort allocations

further. Disparity in access to medical care is likely the real culprit and not disparity in organ transplantation. One part of the system may be easily attacked by litigation while other parts are less vulnerable. In the kidney transplant case, it was not litigation, but persuasion brought to bear on one organization that changed the antigen matching regime. Changing Medicaid or access to early health care was more difficult or was not contemplated.

The initial antigen matching method was developed in light of research into how to match kidneys to whites. For biological reasons, matching to blacks is much more complex, with many more heterogeneous factors to consider. The serological agents used for antigen matching were primarily developed in a white population and are not as reliable when used in an African American population (Institute of Medicine, 1999, 42). African Americans exhibit much greater variability in their histocompatibility antigens than whites, making it much more difficult to match organs for them. Nevertheless, the fact that the transplant technology has a disparate racial impact could conceivably be changed by a medical research agenda that took redressing disparity as a priority. Rather than ceasing to use the antigen matching method, one could press for research to take up the problem of how best to match to blacks, or for more research into immunosuppressant drugs.

There are disproportionately more blacks in end stage renal failure than whites. Why? Is this due to discrimination? A proximate cause in the case of kidney transplants is that blacks wait longer before being treated for the diseases that cause kidney failure. One reason is a disparity in access to health care. Lack of health insurance is in turn a major cause of lack of access. There are many other possible causes. Dr. David Satcher, Surgeon General in the Clinton and Bush Administrations 1998-2002, listed some of the "upstream, midstream, and downstream" causes of a racial disparity in health: breakdown of the family, failure of the educational system, crime, the criminal justice system (disproportionately many blacks get infected with AIDS while in prison), bad health habits, lack of exercise, use of drugs, alcohol, and cigarettes, lack of health insurance, and lack of access to health services (Tavis Smiley Presents 2004). The economic point of view asks: what are the relative costs and benefits of investment in policy interventions at different points in the social system which cause the disparity in transplants? That question leads to the broader question: what are the relative costs and benefits of improvements in the transplant domain

compared with preventive health improvements that could be made in other parts of the total health system? When the ethicist says that black/white disparities in kidney transplants should be ended regardless of cost, the economist asks how many more lives (black lives) could be improved by investing that money in subsidizing health insurance or preventive medicine interventions. It is unlikely that systematic discrimination against blacks is at work in the case of kidney transplants. Blacks are not treated systematically differently in the case of liver transplants where compatibility issues are much less of a problem (Institute of Medicine, 1999).

3 The Challenge Ayres Presents

Ayres' claim that normative considerations should prevail over productive considerations is a direct challenge to conventional applications of disparate impact doctrine. His "technological" challenge is based on the claim that immunosuppressant drugs "dramatically altered the impact of antigen matching." In the context of the disparate impact jurisprudence doctrine developed in the employment discrimination context, one way of understanding his proposal is as an alternative practice that has less disparate impact. It is an alternative technology, but not one that produces medical results equivalent to good antigen matching. His alternative trades medically beneficial outcomes for a decrease in disparate impacts. While Ayres borrows the concept of disparate impact from the employment discrimination domain, he has not found an adequate analog of the concept of business necessity that the Supreme Court regarded as essential to it.

His approach assumes the desired outcome is more kidney transplants, not better health and not less need for kidney transplants. The immunosuppressant drug remedial policy counters a deficiency at one stage of the health production process. Transporting the disparate impact frame of analysis from employment discrimination litigation to transplants takes a litigator's view. It ignores alternative policies that might intervene at earlier stages of the process and produce a lower incidence of kidney failure.

What does Ayres' discussion of "Three Tests for Disparity" have to do with his kidney transplantation recommendations? His revision of conventional statistical methods for measuring

disparate impact advocates eliminating variables that might explain why a racial disparity exists. He dismisses a complete model with a sound biological basis as suffering from what he calls the vice of "included variable bias." In analyzing transplant disparity, he proposes to exclude the degree of antigen matching as a control variable. Evidently, it would not fit under whatever analog to "business necessity" Ayres would accept. In Ayres' analysis, disparity itself is the problem. Whatever heroic last stage measures can address it should be applied, regardless of cost. We next turn to Ayres' analysis of three tests for discrimination.

4 Ayres' Three Tests

Ayres presents three tests for discrimination which we now analyze. Appendices B and C present a more formal analysis of the statistical concepts involved.

4.1 Traditional Test

The conventional approach to testing for discrimination analyzes some outcome equation for a person—a wage, an employment rate or the allocation rate of transplant organs. Differences in outcomes between persons of each race are then regressed on measured qualification traits X, a race indicator variable D (=1 if black; =0 otherwise), and an error term to capture unmeasured traits and sources of disparity. If all relevant productivity traits are measured, least squares estimates of disparity (the regression coefficient on D) are unbiased. A negative estimated coefficient on D means that on average blacks do worse on the outcome than the benchmark group, either because of discrimination or because of lower levels of unmeasured productivity traits. Defendants in disparate impact cases attempt to show that there are nonracial variables X that are legitimate for the business decision-maker to consider and that when these variables are included in the regression, the estimated coefficient on D is statistically insignificant. Plaintiffs often attempt to disqualify use of such variables. Appendix B discusses this test in depth.

4.2 Ayres' Omitted Variables Tests

Distributions of some variables that measure productivity traits differ by race, for example test scores and degree of antigen match. That, in Ayres' view, makes them "tainted." Ayres' approach to the tainted variable problem is to throw the tainted variables away. Whenever he finds a productivity variable X correlated with racial difference D, his instinct is to throw X out of the model. The result of this procedure is to leave the cause of the disparate outcome unexplained. Based on his critique of the "traditional test," Ayres presents the "omitted variables" test as an improvement to prevent "included variable bias." He actually has three versions of the omitted variables test.

The first version looks like a test of a selection practice or criterion. "It's inappropriate to control for these nonracial factors in the regression analyzing the impact of a particular set of decisions, because we want to see whether these nonracial factors produce racially disparate outcomes." By nonracial variables, Ayres refers to variables used as selection criteria by the defendant that are not overtly related to race (Ayres 2004). In the second version, he shifts to say that *all* nonracial variables should be omitted. "Excluding nonracial factors is inappropriate in disparate treatment tests, but such exclusion is *necessary* in disparate impact tests…" The radical omission of variables in the second version of his argument reveals that this version of disparate impact analysis is greatly different from disparate treatment analysis.

> If a disparate treatment regression fails to include (or "omits") a non-racial variable upon which the decisionmaker actually based her decision, then the regression can erroneously indicate that the decisionmaker treated minorities differently than whites. For example, if (1) the decisionmaker has a practice of excluding transplant applicants without a high school diploma from the transplant list and if (2) we further assume that the pool of applicants without diplomas is disproportionately comprised of minorities, then omitting from the regression a control for whether applicants graduated from high school might bias the test of disparate treatment.

> Under a disparate impact theory, it is necessary to intentionally omit non-racial variables from a regression to test whether those variables produced a disparate racial impact. . . [T]he idea is to test whether non-racial factors might have caused a racial disparity in the first place. (Ayers 2004)

Ayres tries to draw a neat separation: disparate treatment analysis should include any conceivable nonracial variable that could explain the decision; disparate impact analysis should omit any variable that could explain the decision.

McDonnell Douglas v. Green, 411 U.S. 793 (1973), defined disparate treatment as treating equally qualified or similarly situated persons differently because of their race. *McDonnell Douglas* stated that the "broad, overriding interest, shared by employer, employee, and consumer, is efficient and trustworthy workmanship assured through fair and racially neutral employment and personnel decisions" (411 U.S. 793, 801). Variables by which one determines ex post whether someone is equally qualified or similarly situated as well as the nonracial criteria that the defendant firm used ex ante to make the challenged decision must have some relationship to legitimate business purposes. Where that appears not to be the case, it is because of taking an excessively narrow idea of what constitutes relevance to business goals. Ayres overstates the distinction between disparate treatment and disparate impact analysis.

Variables indicating whether candidates for hiring or promotion are "similarly situated" or "equally qualified" may relate to capabilities of, activities by, or preferences of, the person. These variables may not relate to the employer's selection practice, yet may indicate why someone was not hired or promoted. Actions or preferences on the part of the employee or candidate can have a causal influence on the consummation, terms, or output of the employment relationship. Ayres (2004) states that "More than 30 years after *Griggs* and a dozen years after the purposefully vague Civil Rights Act of 1991, there is still not legal clarity on . . . whether a defendant is liable when both the defendant and plaintiff's actions are but-for causes of the disparate impact." Ayres' omitted variables test is particularly concerned with omitting variables not related to the decision-maker's practice but which could explain the business decisions on the basis of differences between people such as differences in preferences and initiatives in the employment or matching relationship. However, *McDonnell Douglas* makes clear that action on the part of the plaintiff is relevant to disparate impact or disparate treatment analysis.

Ayres gives the example of the high school diploma as a legitimate qualification for a kidney transplant patient in the eyes of a disparate treatment defense, but illegitimate under disparate impact. He evidently believes a high school diploma is not related to the medical

necessity of transplants. But a healthcare provider could make the case that it is. Patients for many procedures have to be able to follow complex post-operative instructions, to take medications on schedule, and be vigilant and responsive to changes in symptoms. It is entirely possible that more educated people can manage their own care better than less educated people.

A high school diploma might not be a criterion the healthcare provider uses. Perhaps physicians made a "subjective judgment" as to how competent a potential patient would be in contributing to successful post-operative care management. An ex post analysis to determine whether there was discrimination might legitimately use variables that are proxies for ability to understand complex instructions.

His "omitted variable test" for disparate impact is really just a test for differences in characteristics across groups. By omitting all of the X variables that might provide a nonracial explanation for disparity, he is left only with race as an explanation for disparity. See Appendix C for more discussion of this point.

In his third version of the omitted variables test, Ayres (2004) claims that the test is a "unified" test for "unjustified disparate impacts" and so may include "legitimate" nonracial variables. The "basic idea is to include in a regression those variables that would reflect a valid justification for the policy in question." The omitted variables test has nothing to say about the degree of medical necessity of antigen matching. "[I]t is essential to have an independent theory of what types of factors might constitute a valid justification."

The "independent theory" will transform the idea of business justification by balancing "efficiency" against "equity." "To ameliorate the disparate impact of a particular policy how much needs to be sacrificed in terms of survivability[?]" He performs the balancing simply by omitting or partially omitting the *normatively* unjustified variables. In the transplant context, "for fuller (5 or 6) antigen matches it would be appropriate to include controls—as the degree of matching is associated with higher survivability. Depending on whether the law or one's private norms require a trade-off or 'accommodation' of equity with efficiency, it might also be necessary to cap the maximum amount that the coefficients on these variables could take" (Ayres 2004). "Valid justification" means normatively as well as productively valid. He reveals his aim at the close of his paper in discussing the third stage of disparate impact analysis: consideration of whether there

is an alternative practice that accomplishes "legitimate business interests while producing a less disparate racial impact." This "might be done as a more of an accomodationist exercise where researchers would investigate how much of a reduction in disparity could be accomplished by demanding that …employers marginally sacrifice some of what would otherwise be their legitimate interests." That is not the law, but by wielding burden of proof requirements, defendants can be intimidated into settling cases or preventing litigation by de facto quotas. Thus Ayres' innovation is immediately to put the burden of proof on the defendant through the uncontrolled version of the omitted variables test, and then to set a narrow definition of what constitutes a legitimate business purpose.

Ayres claims that antigen matching has a disparate impact. He has not proved that it is not business justified nor has he clarified that concept in the transplant setting. The omitted variables "test" only measures disparity in traits between groups. The test does not detect the source of the disparity nor does it determine whether a trait used to employ, pay, promote or assign an organ is a legitimate variable. It does not necessarily identify discrimination or disparate impact not justified by productivity or other legitimate business reasons. We spell these arguments out more formally in Appendix C.

4.3 Becker's Outcomes Test

Ayres looks to Gary Becker's outcomes test to grapple with the question of when a disparity may be "justified by heightened institutional productivity." The essential idea of Becker's outcomes test (Becker, 1993a,b) is captured by the phrase that "a woman, or a black or a Jew has to be better to get a white man's job." If the *marginal* profitability or productivity is higher for a black than an equivalent white, there is productive inefficiency and a profit taking opportunity is foregone. The firm rejects a highly qualified black person to hire a less qualified white. Becker takes this foregone profit as evidence that the firm's managers are indulging a taste for discrimination.

If marginal profits (or productivity) on *equivalent* persons can be measured, then the test

is a strong one. But it is necessary to make sure that equivalents are being compared. This test, like the traditional test, suffers from the same problems of omitted variables, unless special free entry conditions are assumed that arbitrage away marginal profits. In that case, it is not necessary to control for productivity characteristics. Entry guarantees that marginal profits are zero if there is no discrimination.

The outcomes test is a black box method. There is no need to look at what causes any disparity or what method the organization uses to select people. It is unnecessary to control for any characteristics of the persons with whom the organization deals (chooses or rejects as transplant patients). No control variables are necessary because the outcome being observed, the profit of the firm, is determined in a competitive market. The absence of any control variables is perhaps what makes this test attractive to Ayres. "[T]he outcome tests . . . are . . . not susceptible [to] the traditional omitted variable concern."

However, the application of this test to the regulated environments in which transplants are performed, which are far from competitive markets with free entry, is not obvious. A market in transplants would allocate scarce organs not only by the survivability of transplants, but the money value each transplant recipient placed on transplants compared with competing technologies like dialysis. The rational consumer would also regard preventive habits and medical care as substitutes for transplants. The marginal transplant would have a value determined by competing bidders for scarce organs. Each person would be left to judge the degree of match for herself. Wealthy people who are poor matches might choose to have many transplants a year. Because the political process has not permitted such a market, we are in the position of having to define a welfare function to substitute as a mechanism for determining market outcomes.

Ayres simplifies matters by taking survivability as the desired outcome. At this point, he confuses his normative objective. The choice of a welfare function (W in the model in Appendix A) will be highly controversial in the medical domain. For determining whether there has been unjustified discrimination, however, survivability is a reasonable choice to measure outcomes since that has been widely accepted by medical decision-makers.

When *average* black outcomes are inferior to average white outcomes, that is possible evidence of disparate impact. But a basic principle in economics is that efficiency requires

equating at *margins*, which are much harder to measure. If marginal outcomes are equal, any disparate impact is justified by disparities of productivity. The outcomes test requires identifying the marginal white person and marginal black person. Ayres recognizes this issue when he writes of the "infra-marginality problem." The need to identify the "marginal" person from each racial group gets us back to the question of what criteria selected the marginal person and rejected others. This requires that the analyst open up the black box to postulate specific production processes and their relation to matches with persons.

The outcomes test postulates that if the outcome levels of the marginal white and marginal black differ, there exists a person of the disfavored group who would have been more productive than the person chosen. This is unjustified discrimination. However if people are very heterogeneous, especially if people in one of the groups are very heterogeneous, that might not be the case. Suppose there are a large number of "mediocre" whites in terms of productivity (or transplant survivability). Suppose blacks are characterized by two sub-groups, a few extremely productive, but more who are extremely unproductive. Then an equal opportunity practice would first select all the extremely productive blacks, then all the mediocre whites before coming to the extremely unproductive blacks. If the number of available matches (positions, donated kidneys) falls in the mediocre range, the marginal black would have much higher productivity, but there would be no unjustified discrimination because of heterogeneity among people.

Ayres looks to the outcomes test when he must move from finding disparate impact to finding unjustified disparate impact. But his discussion of the outcomes test does not produce any answer to that question that helps his case against antigen matching. Taking survival as the outcome, antigen matching is a good predictor. Comparing two transplant regimes, a regime with antigen matching would show racial disparity but no discrimination, while a regime without antigen matching would show less disparity but discrimination against whites as more productive white transplant opportunities are turned away. Avoiding the factual causes of disparity by ignoring control variables does not solve the problem of identifying unjustified disparity.

5 A Productivity Framework for Analyzing Disparate Impact

Appendix A develops a productivity model for evaluating the "legitimacy" of practices with a disparate impact. The highly abstract model presented there uses the variable X to represent traits. X may be a vector of many productivity traits. Here, we exposit the model.

Taking productivity as the standard for measuring equality of opportunity, the framework is in the tradition of *Griggs* and *McDonnell Douglas*. By providing a framework for elaborating "business necessity" we aim to provide methods for evaluating the extension of disparate impact analysis to new domains and to relate disparate impact litigation to alternative and complementary policy interventions. Developing parallels for the concept of productivity outside of the labor market and employment decisions of firms is a major challenge that Ayres does not adequately meet.

Defining the analog to business necessity for the transplant domain is not so easy. First of all, there is a question of scope. Should we define "medical necessity" as what is necessary for successful transplants? Or is kidney transplantation just one specific technology for producing health, and should the health outcome be defined with respect to the population at large and the multitude of health issues which affect it? "Business necessity" has actually been developed in the light of firms making decisions in product markets, capital markets, and labor markets—a very complex system.

In the employment domain, the competitive market in which the firm operates brings local optimizing decisions into harmony with market-wide optimality in a "global" system. Business necessity lets the firm lawfully use the most profitable practice for participation in the market. What makes profit maximizing decisions a "business necessity" is that in a competitive market, the firm that does not maximize will be replaced by those that do.

The outcome test depends on specifying what the outcome is. Ayres gets trapped in a contradiction. After asserting the priority of his normative objective of decreasing racial disparity, he defines productivity as transplant survivability. There is a tradeoff between lowering disparity and efficiency, but he provides no principled way to determine what the tradeoff should be or even

what the "price" of a certain amount of disparity in terms of efficiency is. Finding no principle in a broader goal of health, he advocates a localized "normative" race-conscious adjustment. Ayres falls back on an arbitrary political balancing of racial "equity" and "efficiency" because he fails to undertake the factual analysis that could inform political decisions that might unite equity and efficiency.

Unlike a firm's profit, transplant survivability is only technological. It is not determined as an equilibrium between supply and demand in markets that incorporates individual preferences. In order to maximize profit, firms need not only possess technological proficiency but must make products valued by customers. These preferences and market decisions include "normative" beliefs of participating individuals.

In the regulated transplant domain, that harmony of local and global efficiency does not obtain. In these non-market situations, it is necessary to develop a "complete" model of productivity in the sense of a model that relates the plausible specification of a local outcome to some broader "global" definition of the outcome appropriate to the domain. In the case of organ transplants, one plausible local definition is transplant survivability. A contender for the global outcome is overall health of the population as a whole. In a market, people would choose to undergo a transplant not as an end in itself but in order to obtain health.

The standard for "completeness" is the economic one of including all substitutes and complements. The economically motivated structural model would relate the mutual selection of persons and organizations to the global productivity of those matches. The scope would expand or contract depending on the time frame. In the longer run, there are more substitutes for any good. Most importantly, preventive medicine and health producing habits tend to be much more effective than heroic last stage measures like transplants.

For purposes of preventing unlawful discrimination, a "firm" (healthcare provider, organ donation administrator) should be able to justify a practice on the basis of a medical necessity based either on local or global efficiency. If the practice is more productive on the basis of either definition, they have not violated the law as defined in *Griggs* and *McDonnell Douglas*.

6 Discrimination and Other Causes of Racial Disparity

We now discuss some additional aspects of Ayres' Three Tests.

6.1 Testing for Disparate Impact

There is discrimination in the sense of Becker if profitable opportunities are turned away. If the least profitable (productive) match (of transplant or employment technology) with a black person is more profitable than the least productive match with a white person, that indicates there is discrimination. The organization is turning away potentially more profitable matches with blacks in order to select less capable whites.

But how do we know the black person turned away would have produced a more profitable transaction? The application of Becker's criterion is based on the assumption of competition and an implicit mathematical assumption of continuity. But if people are very heterogeneous, there may be unique matches. In the extreme, that is equivalent to "monopoly" and to discontinuity in the distribution of traits. With moderate heterogeneity, there may just be very different distributions of traits across populations.

Thus a key assumption of the optimal (most productive) solution in the mathematical model in Appendix A is continuity. Persons differ in productive traits, but there are many close substitutes for any person. If that is not true, Becker's profits test, modified to a general criterion, breaks down. We now discuss further aspects of the three tests.

Should tests for discrimination be conditional on measured characteristics (traits)? The answer is "yes" in both disparate treatment and disparate impact tests. We have shown that Ayres' contrast between tests of disparate impact and disparate treatment is overstated. In both types of test it is legitimate and necessary to include relevant X variables. Contrary to what he claims, there is no formal test for omitted variable bias.

Note that disparate impact tests (as used by Ayres) assume that it is known if X is productive. We need to test the ingredients of the model, and break up the analysis into analyses of technology, preferences, and outcomes. Looking only at outcomes (the average group disparity result) as is traditional in tests of discrimination is not informative.

The framework presented in Appendix A models how productive traits should be utilized.

Optimality implies that the same cutoffs (in "scores" on tests for matching) be used for both race groups, provided that the same technology is appropriate for them. However we may go to a corner solution—all whites and no blacks, or all blacks and no whites are hired (or given organs). This can happen if members of one group have very low endowments of the productive trait (poor antigen matches) compared with the other group. But if there are no corner solutions, so some members of both groups are hired (given organ transplants), optimality requires that marginal returns for both groups are equalized, as long as the same technology is the best for each group. Average returns of those hired may be different across groups unless the productivity traits are equally distributed. If the marginal profitability is the same across race groups, there is no discrimination as measured by the Becker test. If marginal returns are not equalized, the firm (or decision-maker) is acting inefficiently. Disparity can be measured by the foregone profit opportunity.

If there is no discrimination (judged by the productivity standard) marginal returns are equal but average returns may not be. With heterogeneous traits, and selection proceeding from the most to the least productive, there are diminishing returns to the scale of operation. If minorities have a less favorable distribution of productive traits, there would be disparity in enrollment proportion of minorities away from the population proportion of minorities.

This analysis emphasizes the importance of controlling for productivity traits and for looking at marginal persons in order to detect discrimination. In this model, racial disparity in selection for productive relationships (employment, transplants) may not be due to discrimination but may be the consequence of differences in the distributions of productive traits among groups. The right test for discrimination is to see if the optimal selection conditions hold. If the investigator performs the sleight of hand of confusing average with marginal, it becomes a test of disparity, not of unjustified disparity, just like Ayres' omitted variables test.

Technology may not be uniformly effective across racial groups. Disparate impact theory asks that the best practice (business justified practice) be used. We can write the problem as a possible choice of race-specific technology (as in disparate impact cases) where we now allow for different technology and costs. Different antigen matching rates may affect the best choice of technology by race (organ transplant or dialysis therapy). It is known that blacks fare better on

dialysis than whites and black rejection rates exceed those of whites. Optimality and no bias in this case requires using the best mix of technologies but not necessarily the same cut off criterion for each racial group. Becker's test still applies.

Marginal profitability is equalized if hospitals are not indulging their tastes for discrimination and a different cut-off level is used for black and whites. Disparity may increase or decrease when alternative technologies are available. Allowing for differences in efficacy of technology by race produces a technology choice which is race-specific but not racist. Bias is present if there are departures from productive optimality. This can arise if different welfare functions *W* are used to evaluate black and white outcomes. We can separate bias from technology in principle, if we can measure true productivity.

Ayres considers the possibility of racial differences in technology in terms of the "subgroup validity problem." "To put the matter provocatively, when a particular observable characteristic is only a valid proxy of desert for some races then a decisionmaker's *unwillingness* to engage in disparate racial treatment may induce just the racial disparities in outcomes that are generally a concern" (Ayres 2004).

6.2 Tainted Variables

In Ayres' view, the problem with regression models that purport to explain disparities in rewards as caused not by an employer's practices but by other variables (that measure traits of the person) is that these traits are the result of past discrimination, by Jim Crow, and by slavery itself.

Ayres' omitted variables test does not determine if a selection criterion *X* is a legitimate productivity attribute or a smokescreen—the essential question in disparate impact cases. The problem of "tainted variables" is potentially serious. Consider, for example, the preferences by which a patient would evaluate productive outcomes. Market evaluation (profit) is partly determined by preferences of customers. The policy maker's welfare function *W* in the mathematical setup of Appendix A might also be based on patients' preferences. If a person prefers dialysis to a transplant, should that be taken into account? If there are differences in preferences across racial groups, are those preference differences appropriate control variables? Suppose preferences differ with race, and suppose preferences for dialysis are based on fear of risks

of transplant operations. Suppose such fears are influenced by the perception that the medical establishment is hostile to blacks. Suppose past discrimination has caused distrust. Yet if blacks are resistant to a certain treatment for whatever reason, it may be legitimate to respect their wishes whether or not they are well-founded.

6.3 Causal Responsibility of the Plaintiff

Disparate outcomes arise most immediately from a person's interaction with some institution or organization. Proximate causes are to be found in the practices by which people match to organizations and the productivity of those matches. However, failures to create matches may be influenced by many antecedent causes. There are many ways in which the plaintiff (in an employment disparate impact case) has influence over the adverse outcome. In the long run, he can invest in more training. In the short run, he can search for better opportunities and bargain for better terms.

In the health domain, the preferences and initiatives of potential patients are critical to outcomes. Where equity should come in is in devising the best remedy for the fact that some people have bad health outcomes regardless of what their race is. Race neutral policies which tend to improve the health of the American population as a whole will have a disparate impact favoring blacks. When we ask at a program level what would remedy defective outcomes, promoting health-producing habits in the potential patient plays a major role. When we ask how to "reform the system," ways to make people masters of their fate are critical. Focus on "discrimination" is more than a distraction. "Disparate impact" focuses on practices before which the individual is helpless. *Griggs* was concerned about "childhood deficiencies in the education and background of minority citizens, resulting from forces beyond their control" (*McDonnell Douglas v. Green* 401 U.S. 424, 431 [1971]). A major improvement in the health "system" is to increase the awareness of potential patients that there are actions they can take so they will not become patients.

6.4 The Problem of Statistical Discrimination

Ayres is correct that the Becker profits test does not detect statistical discrimination. If race accurately signals the presence of some productive opportunity, the firm could use that race information to make productive matches. In the presence of statistical discrimination, profits are the same for both groups of workers or transplantees. The profits criterion tests for disparate treatment or a selection practice with disparate impact that is productively irrational in the *Griggs s*ense of having no business justification. However, we cannot separate statistical discrimination from no discrimination by this test.

7 The Economic Point of View

Ayres urges that disparate impact analysis be extended to domains beyond employment discrimination. But his analysis does not provide an adequate framework for such an extension. His conception of "disparate impact" is not that defined in *Griggs* and subsequent cases. The Supreme Court in *Griggs* emphasized that the goal of disparate impact analysis is equality of opportunity but not "preference for any group, minority or majority." (401 U.S. 424, 431) Supreme Court decisions do not demand a tradeoff between lowering disparity and efficiency (or profit). They regard inefficient (profit sacrificing) practices as suspect from an equal opportunity point of view.

Ayres writes that "If past government discrimination has caused elevated African American demand for kidney transplantation, could this not justify race-conscious efforts to mitigate the injury?" (Ayres 2001, 221). Advocating deliberate race-conscious policies is not in the spirit of *Griggs* or *McDonnell Douglas*. It risks promoting, not diminishing, the "racial animus" he seeks to counter.

Ayres devises his own interplay of disparate impact and disparate treatment doctrine. Once disparate impact has been discovered, ignoring it becomes disparate treatment, even if the disparate impact might have been justified by "business necessity." Thus he argues that "ignoring the disparate impact of blacks represents selective indifference" (Ayres 2001, 205).

> Even for those who believe that the best allocation should simply try to maximize survival rates, the willingness of the system to respond selectively to other

> equitable claims might argue for considering the claims of blacks as well. In a world where the equitable claims of other discrete groups are heard, UNOS's failure to respond to the equitable claims of black patients becomes suspect. (Ayres 2001, 205-06)

Ending racial disparity at the final stage of the kidney transplant domain might require vast expenditure of resources. Such resources might save more lives if they were devoted to preventive medicine. The economic point of view looks to all of the tradeoffs that are implicit in any policy choice. Ayres' narrowly focused analysis illustrates the limitations of the disparate impact approach to tackling racial disparities. One can sue over any disparate impact of a health insurance reimbursement policy, but not over the fact that the poor (and blacks) tend to have less health insurance. Suits are more effective against end stage selection practices, but not against root causes. Extending disparate impact liability directs attention and resources away from dealing with root causes of disparities.

Appendices

A Formal Statement of the Ingredients Needed to Measure Disparate Impact

A.1 Socially Optimal Allocation Rules and Their Implications for Testing for Disparate Impacts

Let $F_B(X)$ be the distribution of traits X in the black population and $F_W(X)$ be the distribution of traits in the white population. The technology j mapping $X \longrightarrow Y$ is $Y = g_j(X)$. Y is some output.

A trait is productive if $g_j(X)$ is an increasing function of the trait. We initially assume a common technology across all race groups. $C_j(X)$ is the cost of using technology j. The output evaluation of Y is $W(Y)$. In a business setting where Y is output for the market $W(Y) = P_Y Y$, where P_Y is the price of the output. $W(Y)$ may also reflect preferences of the relevant decision-making agents. Assume that $C(X)$ is convex increasing in X; $g_j(X)$ is concave increasing in X; and $W(g_j(X))$ is concave increasing in X.

Net welfare (in utility) of technology j with characteristics X is

$$W(g_j(X)) - C_j(X). \tag{A.1}$$

For a given technology overall (normalizing the size of the population to equal 1) net welfare

in the population is

$$V_j = \max_{R_W, R_B} \left\{ \begin{array}{l} \int_{R_W} \left[W\left(g_j(X)\right) - C_j(X) \right] P_W dF_W(X) \\ + \int_{R_B} \left[W\left(g_j(X)\right) - C_j(X) \right] P_B dF_B(X) \\ + \lambda \left[\mu - P_W \int_{R_W} dF_W(X) - P_B \int_{R_B} dF_B(X) \right] \end{array} \right\}. \tag{A.2}$$

Here R_W and R_B are the regions of X characteristics for whites and blacks, respectively, who are given jobs, organs, credit, etc. P_W is proportion of whites in overall population. P_B is proportion of blacks. $P_B + P_W = 1$. Implicit in this formulation is the definition of a relevant population and the total number of transplants, jobs, etc. available. λ is a multiplier measuring scarcity of jobs, organs, etc. μ is the number of transplants available relative to the total population.

Consider a scalar case, $R_W = r_W, R_B = r_B$ (scalars). This means there is only one scalar attribute X. We can rewrite the problem as

$$V_j = \max_{r_W, r_B} \left\{ \begin{array}{l} \int_{r_W} \left[W\left(g_j(X)\right) - C_j(X) \right] P_W dF_W(X) \\ + \int_{r_B}^{\infty} \left[W\left(g_j(X)\right) - C_j(X) \right] P_B dF_B(X) \\ + \lambda \left[\mu - P_W \int_{r_W}^{\infty} dF_W(X) - P_B \int_{r_B}^{\infty} dF_B(X) \right] \end{array} \right\}. \tag{A.3}$$

Assuming interior solutions to the first order conditions, we obtain optimal cut-off values (r_B and r_W) from the following optimality conditions,

$$\begin{array}{lll} \left[W\left(g_j(r_W)\right) - C_j(r_W) \right] & \cdot P_W f_W(r_W) & = P_W f_W(r_W) \\ \left[W\left(g_j(r_B)\right) - C_j(r_B) \right] & \cdot P_B f_B(r_B) & = P_B f_B(r_B). \end{array} \tag{A.4}$$

(Here f_Q is the density of the random variable Q which is assumed to exist.) Therefore $r_W = r_B = r$ from concavity and interiority. The same cutoffs are used for both race groups. However, we may go to a corner—no blacks or no whites hired (given organs). This can happen if one group has very low endowments (poor antigen matches). Marginal returns for both groups are equalized as long as both are hired. Average returns of those hired may be different across groups unless $F_W = F_B$, so productivity traits are equally distributed. Observe that if $W(Y) = P_Y Y$ so the goal is profit maximization, marginal profitability is the same across race groups. This is Becker's (1993a,b) test for discrimination. If returns are not equalized, then firms (or decision-makers) are acting inefficiently. Assuming interior solutions, racial disparity is

$$P_W \int_r^\infty f_W(z)dz - P_B \int_r^\infty f_B(z)dz. \tag{A.5}$$

In general, unless distributions of characteristics are the same in the two groups, we get a disparity in enrollment of minorities away from the population proportion if we set marginal returns equal across groups. Marginal returns are equalized but average returns are not if there are diminishing returns. This analysis emphasizes the importance of controlling for productivity traits and for looking at marginal persons in making judgements about discrimination.

In this model racial disparity in treatment of employment is a consequence of differences in the distributions of X among groups. The right test is to see if equations (A.4) hold and to confirm if there are productivity (g_j) or cost (C_j) effects of X. If there is no effect on

productivity or cost, there is no basis for using the trait to screen blacks from whites. We next consider the implications of alternative technologies.

A.2 Allowing for Race Specific Technologies and Costs

Technology may not be uniformly effective across racial groups. Disparate impact theory asks that best practice (business justified practice) be used. We can write the problem as a possible choice of race-specific technology (as in disparate impact cases) where we now subscript technology and costs. Different antigen matching rates may affect the best choice of technology by race (organ transplant or therapy). It is known that blacks fare better on dialysis than whites and black rejection rates exceed those of whites. Different workplace technologies may be productive for blacks and whites. we write the problem as

$$\max_{j_W, j_B, R_W, R_B} \left\{ \begin{array}{l} \int_{R_W} \left(W\left(g_{j_W}(X) \right) - C_{j_W}(X) \right) P_W dF_W(X) \\ + \int_{R_B} \left(W\left(g_{j_B}(X) \right) - C_{j_B}(X) \right) P_B dF_B(X) \\ + \lambda \left[\mu - P_W \int_{R_W} dF_W(X) - P_B \int_{R_B} dF_B(X) \right] \end{array} \right\}. \tag{A.6}$$

We allow for the possibility that different technologies will be selected for different groups. Suppose $\hat{j}_B \neq \hat{j}_W$ (different technologies and costs are optimal for different groups). Optimality requires (in the scalar case)

$$\begin{array}{rcl} P_W \left[\left(W\left(g_{j_W}(r_W) \right) - C_{j_W}(r_W) \right) f_W\left(r_W \right) \right] & = & \lambda P_W f_W\left(r_W \right) \\ P_B \left[\left(W\left(g_{j_B}(r_B) \right) - C_{j_B}(r_B) \right) f_B\left(r_B \right) \right] & = & \lambda P_B f_B\left(r_B \right). \end{array} \tag{A.7}$$

Assuming interior solutions, $W\left(g_{j_B}(r_B)\right) - C_{j_B}(r_B) = W\left(g_{j_W}(r_W)\right) - C_{j_W}(r_W)$. In general, no longer does the same cut-off criterion apply at the margin (in general $r_W \neq r_B$). Thus, optimality and no bias would not necessarily produce the same cut-off rule across race groups. However, Becker's test still applies. Marginal profitability is equalized if agents are not indulging their tastes for discrimination even if a different cut-off level is used for blacks and whites.

Disparity may increase or decrease when alternative technologies are available. Allowing for differences in efficacy of technology by race produces a technology choice which is race-specific but not racist. Bias is present if there are departures from (A.7). This can arise if a different W is used for blacks than for whites. Thus we can separate bias from technology in principle, if we can measure true productivity.

A.3 Testing for Disparate Impact

We can use this setup to test for disparate impacts. The first step is to ask:

1. Does an element of X appear in cost $C_j(X)$ or output $g_j(X)$? This is the test that must be done to determine business necessity within a given technology.

2. Is the weighting of the X (the choice of R_B or R_W) different from optimal? If so, there is intentional bias. This arises from unequal weighting of black and white outcomes due to decision-maker preference. This test looks for departures from (A.4) or (A.7).

Ayres' contrast between tests of disparate impact and disparate treatment is overstated. In both types of test it is legitimate and necessary to include relevant X variables. Contrary

to what he claims, there is no formal test for omitted variable bias.[1] Note that disparate impact tests (as used by Ayres) assume it is known if X is productive. We need to test the ingredients of the model, and break up the analysis into analyses of technology, preferences and outcomes. Looking only at outcome equations is not informative.

A.4 Tests of Outcomes

1. Should they be conditional or not? (Should we condition on measured characteristics?) The answer is "yes", in both disparate treatment and disparate impact tests.

2. The use of unconditional tests of the sort advocated by Becker (1993a, b) requires that free entry characterize the industry or activity being studied. Otherwise marginal profitability differences across race groups may be due to a lack of competition.

3. As correctly noted by Ayres, if productivity is the same, under statistical discrimination, profits are the same for both groups. Under animus-based discrimination, profits are higher in transactions with blacks. Therefore, we can test between the two models under free entry. However, we cannot separate statistical discrimination from no discrimination by this test.

A.5 Disparate Impact Alternatives

We can use this framework to test the feasibility of disparate impact alternatives. One version of this policy is to minimize discrepancy in employment (receipt of organs, etc.) by

[1] Gastwirth (1988, 1992, 1996) presents methods based on the Cornfield inequality for determining the sensitivity of estimates to omitted characteristics.

choosing techniques so that a given level of profitability is maintained. Thus we can choose a technology so as to minimize disparity subject to maintaining a given profit level and meeting constraints,

$$\min_{r_W, r_B, j_W, j_B} \left[P_W \int_{r_W}^{\infty} f_W(X)\, dX - P_B \int_{r_B}^{\infty} f_B(X)\, dX \right]$$

subject to

$$\mu = P_W \int_{r_W}^{\infty} f_W(X)\, dX + P_B \int_{r_B}^{\infty} f_B(X)\, dX,$$

and prespecified profit level $\bar{V}$, where

$$\begin{aligned} \bar{V} &= P_B \int_{r_B} \left[W\left(g_{j_B}(X)\right) - C_{j_B}(X) \right] f_B(X)\, dX \\ &\quad + P_W \int_{r_W} \left[W\left(g_{j_W}(X)\right) - C_{j_W}(X) \right] f_W(X)\, dX. \end{aligned}$$

Observe that this policy may not produce the maximum profit result and different technologies may be used for different race groups. The analysis of this appendix illustrates the value of going beyond simple regressions to determine the ingredients of tastes, technology and endowments that produce outcomes.

B Conventional Methods for Measuring Discrimination

The conventional approach to testing for discrimination works with some final outcome equation—a wage, an employment rate or the allocation rate of organs—to detect discrimination. Underlying this approach are two sets of equations: (a) outcome equations and (b) equations determining traits or characteristics denoted X. Let Y_W be the white outcome and Y_B be the black outcome.

The first set of equations is for outcomes:

$$\begin{aligned} Y_W &= \alpha_{0W} + \alpha_{1W} X + \eta_W \\ Y_B &= \alpha_{0B} + \alpha_{1B} X + \eta_B. \end{aligned}$$

The means of η_W and η_B are zero. α_{0W} and α_{0B} capture both unmeasured productivity traits and unmeasured sources of discrimination. Since they are unmeasured, we cannot tell which source of disparity is more important. These equations record how characteristics measured (X) and unmeasured (η_W, η_B) characteristics determine outcomes for whites and blacks. Assume there is only one X. Lack of disparate treatment means that $\alpha_{0W} = \alpha_{0B}$ and $\alpha_{1W} = \alpha_{1B}$. Persons with identical values of traits are treated equally. Disparate treatment arises if $\alpha_{0W} \neq \alpha_{0B}$ or $\alpha_{1W} \neq \alpha_{1B}$ or both. Let $D = 1$ if a person is black; $D = 0$ otherwise.

Then observed outcomes are

$$\begin{aligned} Y &= DY_B + (1-D)\,Y_W \\ &= \alpha_{0W} + (\alpha_{0B} - \alpha_{0W})\,D + \alpha_{1W}X + (\alpha_{1W} - \alpha_{1B})\,DX \\ &\quad + \eta_W\,(1-D) + \eta_B D. \end{aligned}$$

If all relevant productivity traits are measured, least squares estimates are unbiased. A negative estimated coefficient on D means that on average blacks do worse, either because of discrimination or because of lower levels of productivity traits. A negative estimated coefficient on DX means that as productivity traits increase blacks receive a smaller increase in payment (employment, organ transplantation) than whites.

To simplify the argument suppose that $\alpha_{1W} = \alpha_{1B} = \alpha_1$. Then the outcome equation is

$$Y = \alpha_{0W} + \Delta D + \alpha_1 X + \eta, \tag{B.1}$$

where $\Delta = \alpha_{0B} - \alpha_{0W}$ and $\eta = \eta_W\,(1-D) + \eta_B D$. Good (unbiased) estimates of the parameters of this equation can sometimes detect the presence or absence of *disparate treatment*.

The assumptions underlying application of this equation as a measurement framework are: (1) That X accurately captures all relevant productivity factors. Thus D is uncorrelated with η. The estimated $(\alpha_{0B} - \alpha_{0W}) = \Delta$ captures discrimination. Of course, if this is not true, the least squares estimates of Δ reflect both discrimination and unobserved productivity. (2) A second assumption is that X is uncorrelated with (η_W, η_B). A correlation could

arise if η_W and η_B are correlated with unmeasured factors causing X.

Thus we can write

$$\begin{aligned} X_W &= \beta_{0W} + \nu_W \\ X_B &= \beta_{0B} + \nu_B, \end{aligned}$$

where ν_W and ν_B have zero means and are independent of each other. Observed X is

$$\begin{aligned} X &= DX_B + (1 - D) X_W \\ &= \beta_{0W} + (\beta_{0B} - \beta_{0W}) D + \nu, \end{aligned} \tag{B.2}$$

where $\nu = \nu_W (1 - D) + \nu_B D$. Now if ν_B is positively correlated with η_B and ν_W is positively correlated with η_W, say because more productive people (those with higher η) get more training (X), and η_B is uncorrelated with ν_W and η_W is uncorrelated with ν_B, then one can show that if $\beta_{0B} < \beta_{0W}$ (blacks get less training), the least squares estimate of Δ is upward biased as is the least squares estimate of α_1. If $\beta_{0B} = \beta_{0W}$, the OLS estimate of Δ, $\hat{\Delta}$ is unbiased for Δ and if $\beta_{0B} > \beta_{0W}$, $\widehat{\Delta}$ is downward biased for Δ. See the derivations in Appendix C. The reason for the upward bias in the coefficient of α_1 is intuitively obvious. The regression gives too much credit to X and not to ν. The estimated coefficient for α_1 picks up some of the effect of the unmeasured η_W and η_B which positively affect the outcome. The upward bias for Δ is less obvious. Some of the bias arising from the X-η relationship gets shared with D. D is uncorrelated with η and is negatively correlated with X. See the

analysis of Appendix C.

This means that if there is disparity in X, and X is "tainted" (correlated with η), then if higher levels of the η are associated with higher levels of the ν, and there is no disparate treatment ($\Delta = 0$), the estimated Δ is positive, suggesting blacks are *favored*.

One way to undo this bias is to use the method of instrumental variables. If there is a variable Z correlated with X and uncorrelated with η, application of the method produces unbiased estimates of α_1 and Δ.[2]

Ayres' "solution" is different. It is to delete the X, and run the regression. When $\Delta = 0$ (no disparate impact), this is equivalent to running a regression of X on D (See Appendix C). The procedure can detect disparity in X by race. It cannot reveal why there is disparity. Ayres' solution does not answer the question of whether there is disparate impact. It simply tells us there is disparity. Ayres confuses disparity with disparate impact.

Ayres' discussion of omitted variable bias is confusing. There is no formal test for omitted variable bias unless we can measure the omitted variables in which case there need be no bias. There is no mechanical algorithm for picking which variables belong in X when the X are correlated with ν (See, e.g., Heckman and Navarro, 2004).

The real message to take from Ayres' paper is not that X variables should be omitted when testing for disparate treatment and should be included when testing for disparate treatment. The real message is that regressions based on (B.1) cannot separate out bias from endowments. They cannot determine whether X is a genuine productivity attribute. The only way to do that is to find measures of productivity or profitability, instead of

[2]See, e.g., Greene (2003) for a discussion of the method of instrumental variables.

outcomes of an allocation or wage-setting process, against which to measure the effect of X.[3]

[3]Measurement error in X raises a whole set of other problems. Omitting or including variables measured with error may bias α_1 and Δ in any direction if X has many variables.

C A Model of Disparate Impacts with a Tainted Variable

Using equations (B.1) and (B.2) in the previous appendix, and letting "ˆ" denote the OLS estimate,

$$\text{plim} \begin{pmatrix} \hat{\alpha}_1 \\ \widehat{\Delta} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \Delta \end{pmatrix} + \begin{pmatrix} Var(X) & Cov(X,D) \\ Cov(X,D) & Var(D) \end{pmatrix}^{-1} \begin{pmatrix} Cov(X,\eta) \\ Cov(D,\eta) \end{pmatrix}. \quad \text{(C.1)}$$

We have

$$Cov(X,\eta) = E(X\eta) = E(X\eta \mid D=1)P + E(X\eta \mid D=0)(1-P)$$

because $E(\eta) = 0$. Notice that

$$\begin{aligned} E(X\eta \mid D=1) &= E[(\beta_{0B} + \nu_B)\eta_B \mid D=1] \\ &= E(\nu_B \eta_B). \\ E(X\eta \mid D=0) &= E[(\beta_{0W} + \nu_W)\eta_W \mid D=0] \\ &= E(\nu_W \eta_W) \\ E(X\eta) &= PE(\nu_B \eta_B) + (1-P)(\nu_W \eta_W). \end{aligned}$$

$$Cov(X,D) = P(1-P)(\beta_{0B} - \beta_{0W}).$$

$Cov\left(D,\nu\right)=0$ from the definition of the error term ($E\left(\nu_B\right)=0$, $E\left(\nu_W\right)=0$ and $\nu = D\nu_B+(1-D)\,\nu_W$). Define $|\det|$ as the determinant of the regressor matrix which is assumed to be of full rank and is positive,

$$\det\begin{pmatrix} Var\left(X\right) & Cov\left(X,D\right) \\ Cov\left(X,D\right) & Var\left(D\right) \end{pmatrix} = |\det| > 0.$$

Then the probability limit of the least squares estimator with X and D included is

$$\operatorname{plim}\begin{pmatrix} \hat{\alpha}_1 \\ \widehat{\Delta} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \Delta \end{pmatrix} + \frac{1}{|\det|}\begin{pmatrix} Var\left(D\right) & -Cov\left(X,D\right) \\ -Cov\left(X,D\right) & Var\left(X\right) \end{pmatrix}\begin{pmatrix} E\left(X\eta\right) \\ 0 \end{pmatrix}.$$

Assuming that $\beta_{0B}-\beta_{0W}<0$ (disparity in X) and that $E\left(\nu_B\eta_B\right)>0$ and $E(\nu_W\eta_W)>0$ (more X for more productive people), $\widehat{\Delta}$ is upward biased. Assume no disparate treatment so $\Delta=0$, $\widehat{\Delta}>0$. Thus the regression with a tainted X would show *favoritism* for blacks. Note further that $\hat{\alpha}_1$ is upward biased for α_1.

Suppose we omit X as Ayres suggests for his test for disparate impact. This produces the equation

$$Y=\alpha_0+\Delta D+\left\{\alpha_1 X+\eta\right\}.$$

Least squares under the assumption of no disparate impact ($\Delta=0$) is

$$\operatorname{plim}\tilde{\Delta}=\left(\alpha_1\right)\frac{Cov\left(X,D\right)}{Var\left(D\right)}.$$

Summarizing, we obtain in the general case

$$\begin{aligned}
\text{plim } \widetilde{\Delta} &= (\alpha_1)(\beta_{0B} - \beta_{0W}) + \Delta, \\
\text{plim } \widehat{\Delta} &= -\frac{Cov(X, D)}{|\det|}\left[E(\nu_B \eta_B) P + E(\nu_W \eta_W)(1 - P)\right] + \Delta \\
&= -\frac{P(1-P)}{|\det|}(\beta_{0B} - \beta_{0W})\left[E(\nu_B \eta_B) P + E(\nu_W \eta_W)(1 - P)\right] + \Delta.
\end{aligned}$$

When $\Delta = 0$ (no disparate treatment in the outcome equation), we have that if there is disparity in the trait $(\beta_{0B} - \beta_{0W}) < 0$ and $\alpha_1 > 0$, plim $\widetilde{\Delta} < 0$. But plim $\widehat{\Delta} > 0$, which might suggest *favoritism* for blacks. Thus if there is disparity, whatever the source, Ayres will detect it. When $\Delta = 0$, the Ayres method amounts to regressing X on D, circumventing the outcome equation completely. Such a regression cannot decide the sources of the disparity. It also does not test if X is a legitimate productivity attribute or a smokescreen—the essential question in disparate impact cases.

If there is an instrument Z such that $E(Z\nu) = 0$, $E(ZX) \neq 0$, the instrumental variable estimator of Δ is consistent for the parameter. Thus if including X in the regression shows favoritism for blacks (plim $\widehat{\Delta} > 0$) but the IV estimator shows none, we also have evidence of disparity, but not necessarily any form of discrimination.

Notice further that if acquisition of the trait is unrelated to ν ($Cov(\nu, \eta) = 0$ so $E(\nu_B \eta_B) = 0$ and $E(\nu_W \eta_W) = 0$), the standard method (and the instrumental variable method) will show no bias (plim $\widehat{\Delta} = 0$). These tests are for the presence or absence of disparate treatment assuming X is correctly measured.

Note further that if $\beta_{0B} = \beta_{0W}$, plim $\widehat{\Delta} = 0$ and plim $\widetilde{\Delta} = 0$ even if $E(\nu_B \eta_B) \neq$

$E(\nu_W \eta_W)$. Unequal dependence between unobservables in productivity traits and unobservables in outcomes across race groups is another type of disparity in treatment that is off the radar screen of these tests. Methods based on fitting (B.1) or regressing X on D do not provide a way of testing for disparate impact.

References

1) Ayres, I. (2001) *Pervasive Prejudice?* Chicago, University of Chicago Press, 2001.
2) _____ (2003) Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of Included Variable Bias, presented at the Conference on Disparities in Receipt of Treatment, University of Chicago Law School, November, 2003.
3) Becker, G. (1993a) The Evidence Against Banks Does Not Prove Bias, *Business Week*, April 19, 1993.
4) _____ (1993b) The Economic Way of Looking at Behavior, *Journal of Political Economy*, Vol. 101, no. 3 (June 1993) 385--409.
5) *Connecticut v. Teal*, 457 U.S. 440, 447, 448 (1982).
6) *Dothard v. Rawlinson*, 433 U.S. 321, 329 (1977).
7) Gastwirth, J. (1992) Methods for Assessing the Sensitivity of Statistical Comparisons Used in Title VII Cases to Omitted Variables, *Jurimetrics,* Vol. 32, #3, 1992.
8) _____ (1996) Statistical Issues Arising in Equal Employment Litigation, *Jurimetrics,* Vol. 36, #4, 1996.
9) _____ (1988) *Statistical Reasoning in Law and public Policy Vol. 1: Statistical Concepts and Issues of Fairness*, Academic Press, New York and Boston, 1988.
10) *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).
11) Greene, W. (2003) *Econometric Analysis*, 5th edition. Upper Saddle River, New Jersey: Prentice Hall.
12) Heckman, J. and S. Navarro (2004), Using Matching Instrumental Variables and Control Functions to Estimate Economic Choice Models," *The Review of Economics and Statistics*, February 2004, 86(1): 30-57.
13) Institute of Medicine (1999) Organ Procurement and Transplantation: Assessing Current Policies and Potential Impact of the DHHS Final Rule, National Academy Press, Washington, D.C.
14) *McDonnell Douglas Corp. v. Green*, 411 U.S. 793 (1973).
15) Tavis Smiley Presents: The Black Family, C-SPAN February 28, 2004