# Testing for Discrimination and the Problem of "Included Variable Bias"

Ian Ayres*

Abstract

Disparate impact and disparate treatment claims have distinct legal elements and require distinct statistical tests. In a disparate treatment test, the primary statistical concern is most often "excluded variable bias" – the worry that the quantitative estimates of disparate treatment are biased because the regression inappropriately excludes necessary non-race variables. But this article shows that in disparate impact testing, the primary statistical concern is most often "included variable bias" – the worry that the statistical estimates of disparate impact are biased because the regression inappropriately includes non-race variables. Somewhat surprisingly, this article will show that it is appropriate to exclude from a regression non-race control variables that even are thought to be causally related to the decision that is being modeled. Appropriate statistical testing for disparate impact that attends to included-variable bias will thus often be less data intensive and particularly well suited for class action adjudication. The article develops regressions specifications testing for four distinct civil rights concerns: disparate treatment, *prima facie* disparate impact, unjustified disparate impact, and subgroup disparate impact.

Words: 20,500

*William K. Townsend Professor & Anne Urowsky Professorial Fellow in Law, Yale Law School. ian.ayres@yale.edu. Rick Brooks, Cynthia Estlund, and David Wilkins provided helpful comments. James Richardson and Wendy Zupac provided excellent research assistance. This article was the basis for my Francis Biddle Lecture at Harvard Law School. While serving as an expert witness in Cason v. Nissan Motor Acceptance Corp (2001) 3-98-0223 (M.D. Tenn.) and a number of parallel pieces of finance litigation, I have offered opinions on the appropriate statistical tests of disparate impact that directly relate to the issues discussed in this article.

## INTRODUCTION

Statistical tests of race discrimination have mistakenly followed a unified strategy of examining whether non-race variables explain away *prima facie* racial disparities in defendant decisionmaking.[1] While the consensus approach makes eminent sense when testing for disparate treatment, this article suggests that testing for disparate impact requires a distinct statistical method. Since disparate treatment and disparate impact claims have different elements, adjudicating these two types of claims should turn on different types of evidence – including different statistical evidence. A disparate treatment claim centrally requires proof of "intentional discrimination."[2] This means that a plaintiff must prove that his or her race was a motivating factor in a defendant's adverse decision.[3]

However, a plaintiff bringing a disparate impact claim need not prove defendant's intentional discrimination.[4] Instead, in disparate impact litigation a violation is "made out when an employer is shown to have used a specific employment practice, neutral on its face but causing a substantial adverse impact on a protected group, and which cannot be justified as serving a legitimate business goal of the employer."[5] In disparate impact litigation, a plaintiff need not show that race played any role in the employer's decision to implement the race-neutral

---

[1] While this article uses discrimination on the basis of race as its motivating example, the methods discussed are generally applicable to discrimination test on other grounds.

[2] *See also* Michael Selmi, *The Price of Discrimination: The Nature of Class Action Employment Discrimination Litigation and its Effects*, 81 TEX. L. R. 1249 (2003).

[3] "[Plaintiff] must prove that [his/her] [protected status] was a motivating factor in [defendant's] decision." MODEL CIVIL JURY INSTRUCTIONS 5.1.1, Third Circuit, *Elements of a Title VII Claim – Disparate Impact – Mixed Motive*, *available at* http://www.ca3.uscourts.gov/civiljuryinstructions/Final-Instructions/Ch5-TitleVII/Ch5-5.1.1.pdf.

[4] EEOC v. Metal Service Co., 892 F.2d 341, 347-348 (3d Cir. 1990) ("No proof of intentional discrimination is necessary.").

[5] *Id.* *See* Civil Rights Act of 1964, 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2006) ("a complaining party [must demonstrate] that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity").

employment practice. Accordingly, a statistical approach that is geared toward testing whether a plaintiff's race caused an employer to behave differently has no necessary relation to the core elements in a disparate impact claim. As the Supreme Court noted in *Watson v. Fort Worth Bank & Trust*, "[t]he factual issues and the character of the evidence are inevitably somewhat different when the plaintiff is exempted from the need to prove intentional discrimination."[6]

This article provides a theory for how to make statistical tests "somewhat different when the plaintiff is exempted from the need to prove intentional discrimination." The key difference turns on the appropriate list of non-race controls. Put simply, it is appropriate to include more non-race control variables in disparate treatment testing than in disparate impact testing. In a disparate treatment test, the central statistical concern is most often "omitted (or excluded) variable bias" – the worry that the statistical estimates of disparate treatment are biased because the regression inappropriately *excludes* necessary non-race variables. If a test fails to control for a non-race factor that may have prompted an employer's adverse decision with regard to a particular plaintiff, then the test may falsely attribute the adverse decision to the applicant's race. But this article shows that in disparate impact testing, the primary statistical concern is most often "included variable bias" – the worry that the statistical estimates of disparate impact are

---

6. *Watson v. Fort Worth Bank & Trust*, 487 U.S. 977, 987 (1988). The need to use a different method of analysis because of this different elements in disparate treatment and disparate impact claims was emphasized in the Supreme Court's most recent disparate impact decisions, *Lewis v. City of Chicago*, No. 08-974, 2010 U.S. LEXIS 4165 (May 24, 2010), where Justice Scalia in analyzing a statute of limitations question found:

> [A] Title VII plaintiff must show a 'present violation' within the limitations. What that requires depends on the claim asserted. For disparate-treatment claims—and others for which discriminatory intent is required—that means the plaintiff must demonstrate deliberate discrimination within the limitations period. But for claims that do not require discriminatory intent, no such demonstration is needed. Our opinions, it is true, describe the harms of which the unsuccessful plaintiffs in those cases complained as 'present effect[s]' of past discrimination. But the reasons they could not be the present effects of present discrimination was that the charged discrimination required proof of discriminatory intent which had not even been alleged. That reasoning has no application when, as here, the charge is disparate impact, which does not require discriminatory intent."

*Id*. at *17 (citations omitted).

biased because the regression inappropriately *includes* non-race variables.

Part I of this article formally lays out and distinguishes the concepts of excluded- and included-variable bias. But the importance of the difference can be gleaned from a stylized analysis of the Supreme Court's very first disparate impact case, *Griggs v. Duke Power Co.*[7] In *Griggs*, the Supreme Court found that Duke Power's requirement of a high school diploma or use of an aptitude test to screen applicants for certain jobs resulted in a disparate impact violation because (1) the requirements caused African-American applicants to be disproportionately rejected, and (2) the requirements were not reasonable measures of job performance. If a researcher today were evaluating the claims in *Griggs v. Duke Power Co.*, one could imagine testing whether the employer was less likely to hire African-American applicants than white applicants. In statistical testing, it would be possible to control for whether particular applicants had received a high school diploma. Under the facts of *Griggs*, such a control would likely have reduced the racial disparity in the hiring rates – for the simple reason that minority applicants at that time were less likely to have a high school diploma. Should a statistical test control for whether or not an applicant had a high school diploma?

In a disparate treatment case, the answer is yes. Under a disparate treatment theory, the trier of fact would be attempting to ascertain whether an applicant's race was the cause of being denied employment. If applicants were rejected because the employer chose not to hire diploma-less applicants, the applicants' race would not be a "motivating factor" in employer's decision

---

[7] *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). Most recently, in Ricci v. DeStefano, 557 U.S. ___; 129 S.Ct. 2658 (2009), the Supreme Court held that absent a strong basis in evidence of an impermissible disparate impact, that a decision maker may not engage disparate treatment to avoid a disparate impact. This strong basis in evidence rule might impact the quantum of evidence that a trier of fact needed to justify a race-conscious remedy but it would not impact the theory of appropriate disparate impact controls set out in this article.

not to hire.[8]

But in testing for disparate impact, it would be inappropriate to control for whether an applicant had a high school diploma. In a disparate impact case, the central question is not whether minority applicants were less likely to be hired after taking account of whether they had a high school diploma. The central question was instead whether the employer's diploma (and aptitude test) requirement had an unjustified disparate impact. The possibility that there would be no statistical difference in an employer's propensity to hire minorities and non-minorities *after* controlling for applicants' diploma status does not speak to whether the employer's decision to condition employment on having a diploma itself had an unjustified disparate impact. In *Griggs*, the Court independently found that the employer had no legitimate business justification to require a high school diploma for the manual labor positions being filled, so the sole statistical question was whether this diploma requirement had a disparate impact on African-American applicants. The possibility that including a diploma variable would reduce the estimated race effect in the regression would in no way be inconsistent with a theory that the employer's diploma requirement disparately excluded African-Americans from employment.

The *Griggs* thought experiment exemplifies the central claims of this article. Excluding non-race factors is inappropriate in disparate treatment tests, but such exclusion is *necessary* in disparate impact tests. Disparate treatment tests, at least in theory, should strive to control for any and all variables that plausibly had a causal impact on a defendant's decisionmaking.[9] But *disparate impact tests should only include controls for attributes that are plausibly business*

---

[8] *See* MODEL CIVIL JURY INSTRUCTIONS, *supra* note 3.

[9] *See infra* text accompanying note 19 for a discussion of how this standard harmonizes with existing judicial approach to "pretextual" controls.

*justified*.[10] If having a high school diploma is not a business justified condition of employment, then it is inappropriate to separately control for diploma status in a disparate impact test. In testing for disparate racial impacts, it is appropriate to exclude non-race control variables, even those that might have causally impacted a defendant's decisionmaking. The diploma status of applicants in *Griggs* may have driven the employer's hiring decisions. But the causal role of an unjustified applicant attribute should not be used to explain away a racial disparity in hiring.

Inappropriately including controls for variables that are not plausibly business justified creates the problem of "included-variable bias." Instead of estimating the disparate racial impact after solely controlling for business justified attributes, the inclusion of inappropriate variables is likely to bias downward the estimates of disparate racial impact. This article will show that included-variable bias has impacted the quality of statistical testing in a variety of civil rights settings.[11]

An appreciation of included-variable bias also has implications for the procedural viability of many civil rights actions. Disparate treatment claims are at times hampered because plaintiffs do not have sufficient data to establish that non-race attributes did not cause the adverse decision. But appropriate statistical testing for disparate impact that attends to included-variable bias is often less data intensive, requiring access to fewer variables. Hence plaintiffs will often have an easier time gathering through discovery the kinds of data needed to conduct the kinds of disparate impact tests described below than they would if the claims being litigated concerned disparate treatment.

---

[10] As applied to non-business decisionmakers, disparate impact test should only include controls for attributes that are "organizationally" justified—that is, that plausibly serve the organizations legitimate interests. Criteria for identifying business-justified factors are discussed *infra* at text accompanying note 45.

[11] See *infra* text accompanying note 33.

Part I develops in greater detail the divergent methods needed to test for disparate treatment and disparate impact. It shows through a series of applications how the hegemony of the disparate treatment approach to discrimination tests has biased estimates in disparate impact litigation. Part II then analyzes outcome tests of discrimination, which assess whether there are racial differences in the success of defendant decisionmaking. This Part will show that outcome tests are not well suited to uncover evidence of disparate treatment, but in appropriate circumstances can provide evidence of what I will call "subgroup disparate impacts." It will also show that the problem of included-variable bias is an even larger concern with outcome testing. The purpose of this article is to provide courts and litigators with a firm basis for adopting a different statistical methodology when confronting a discrimination case with different elements. In all, this article develops regressions specifications testing for four distinct civil rights concerns: disparate treatment, *prima facie* disparate impact, unjustified disparate impact, and subgroup disparate impact. The hope is to provide better theoretical guidance for deciding which controls to include and which to exclude when testing for race effects.

## I. EXCLUDED VS. INCLUDED VARIABLE BIAS

### A. *The "Kitchen Sink" Approach to Disparate Treatment Testing*

In disparate treatment litigation, a crucial question in establishing intentional discrimination is to ask whether the plaintiff's race influenced a defendant's adverse decision. If the plaintiff were a different race, would the defendant's decision have been the same?[12] Thus,

---

[12] In mixed motive cases a plaintiff "is not required to prove that [his/her] [protected status] was the sole motivation or even the primary motivation for [defendant's] decision. [Plaintiff] need only prove that [plaintiff's protected class] played a motivating part in [defendant's] decision even though other factors may also have motivated [defendant]." MODEL CIVIL JURY INSTRUCTIONS 5.1.1, Third Circuit, *Elements of a Title VII Claim – Disparate Impact – Mixed*

for example, in a hiring case where a minority applicant is rejected, the crucial counterfactual is whether a non-minority applicant who was identical to the minority applicant except for being a non-minority would also have been rejected by the defendant. Statisticians testing whether a defendant engaged in a "pattern or practice" of disparate treatment tend to test whether minorities received less favorable defendant treatment than non-minorities after statistically controlling for a host of non-race variables.

Regression analysis of historic decision-making is the central tool by which statisticians test whether the race of the plaintiffs influenced the defendant's decision making.[13] A regression can simultaneously control for a variety of potential influences and estimate the size and statistical significance of the individual influences.[14] In disparate treatment regressions, the central inquiry is, after taking into account the impact of non-race influences, to test whether the race of the plaintiffs had an independent influence on the defendant's decisionmaking. At a sufficient level of generality, there is a well-accepted theoretical approach to specifying the regression equation to accomplish this disparate treatment test. The regression equation takes the form:

$$\text{Defendant Decision} = \alpha + \beta_1 * \text{Minority} + \sum_i \beta_i * (\text{Plausible Non-Race Influences}) + \varepsilon, \qquad (1)$$

where "Defendant Decision" is the defendant decision variable which is the subject to a claim of

---

*Motive*, *available at* http://www.ca3.uscourts.gov civiljuryinstructions/ Final-Instructions/Ch5-TitleVII/Ch5-5.1.1.pdf. This "motivating part" standard suggests that a plaintiff may not need to prove that his or her race was a but-for cause of the adverse decision.

[13] An alternative approach is randomized testing with auditors. *See, e.g.*, Ian Ayres, Fair *Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 Harv. L. Rev. 817 (1991). Randomization and regression are the two central techniques of predictive analytics. IAN AYRES, SUPER CRUNCHERS: WHY THINKING-BY-NUMBERS IS THE NEW WAY TO BE SMART 14 (Bantam 2007).

[14] D. James Greiner, *Causal Inference in Civil Rights Litigation*, 122 HARV. L. REV. 533 (2008). The regression is a statistical procedure that estimates the parameters that produce the "best fit" between a hypothesized model of exogenous influences and some dependent variable. *See* Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299, 1312 (1984).

discrimination, the "Minority" is an indicator (or "dummy") variable taking on the value of 1 if the person subject to the decision is a minority (and 0, otherwise),[15] "Plausible Non-Race Influences" would be a set of variables representing all of the non-race factors that might plausibly have influenced the defendant decision, and $\varepsilon$ is an error term, which is modeled to capture unexplained variation in the defendant's decision. Thus, for example, in a hiring case, a regression might evaluate whether an employer is likely to hire a particular candidate. The left-hand side (or dependent) "Defendant Decision" variable in this regression would equal 1 if a particular applicant was hired, and 0 otherwise; and the right-hand side (independent) control variables would include a potentially long list of factors (such as the applicant's experience or level of schooling) that might influence the employer's willingness to hire that particular applicant.

The regression output returns estimates of the size and statistical significance of $\alpha$, $\beta_1$, and the $\beta_i$ for each of the individual potential influences. In this regression, the sign, size and statistical significance of the $\beta_1$ coefficient are the central tests of disparate treatment. If $\beta_1$ is estimated to be negative and statistically different than zero, the regression indicates that after controlling for other potential non-race influences, that the defendant employer was less likely to hire minority applicants. If properly specified with sufficient controls, this regression specification can provide credible evidence of defendant disparate treatment.[16] In the hiring

---

[15] It is possible to generalize this regression to simultaneously test for multiple racial effects by including more specific racial categories in the specification (for example, African-American, Hispanic, etc.).

[16] Hylton and Rougeau have shown however that the dominant approach is not well-suited for identifying instances of non-mistaken statistical discrimination. For example, imagine an employer who is in fact relying on race as a statistically valid proxy for business-relevant attributes of its applicants. A researcher who subsequently includes controls for those business-relevant attributes is unlikely to uncover a statistically significant estimate of race discrimination – even though the defendant decisionmaker was in fact discriminating. Accordingly, specification (1) is most powerfully tailored to identify instances of disparate treatment that are not based on

example, a negative and statistically significant race coefficient would provide credible evidence that the defendant employer was less likely to hire similarly situated minorities. These kind of regression results would indicate that the applicants' race was impermissibly influencing the defendant's decisionmaking.

A central concern in specifying a disparate treatment regression is deciding on the appropriate set of controls to include in the regression. These controls represent the kinds of non- nondiscriminatory explanations that a defendant would offer (in the *McDonnell Douglas* second stage of proof)[17] for its decisions. In a disparate treatment regression, it is appropriate to include any variable that might provide a non-race basis for a given decision. Thus, even variables that are not related to the profitability of a decisionmaker's business might be properly included and might indicate a non-race basis for superficial racial disparities. For example, consider an employer who mistakenly believes that Pisces are less productive, or simply harbors animus toward Pisces, and refuses to hire applicants who are Pisces. It would be appropriate to include a control for that horoscope category in the regression. Since disparate treatment based on astrological signs is legal, evidence that a plaintiff's adverse employment outcome was driven by her astrological sign would be a defense to a claim of disparate racial treatment. Consistent with *McDonnell-Douglas*,[18] the only variables that should be excluded from a disparate treatment regression are those non-race variables that are merely "pretexts" for masking what

---

statistically valid racial profiling. The authors' analysis also suggests that the control variables included in equation (1) might be limited to those that were known to the decisionmaker at the time of making the decision. *See* Keith N. Hylton & Vincent D. Rougeau, *Lending Discrimination: Economic Theory, Econometric Evidence, and the Community Reinvestment Act*, 85 Geo. L. J. 237 (1996).

[17] *See* McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973).

[18] Under the shifting *McDonnell Douglas* burdens, after the defendant articulates, through admissible evidence, a legitimate, nondiscriminatory reason for its actions, the plaintiff must prove that the employer's stated reason is a pretext to hide discrimination. *McDonnell Douglas*, 411 U.S. at 802-04; Tex. Dep't of Cmty. Affairs v. Burdine, 450 U.S. 248, 252-56 (1981).

otherwise would be race-contingent decisionmaking.[19]

In most disparate treatment disputes, the impulse of most econometricians is toward what I will call the "kitchen sink" approach to disparate impact testing – that is, including in the regression specification any and all possibly plausible controls. The impulse derives from one of the great econometric asymmetries with regard to statistical bias. If a researcher mistakenly includes in equation (1) an irrelevant control (in an otherwise correctly specified equation), then the estimated coefficient of interest, $\beta_1$, will still be an unbiased estimate of the true disparate treatment.[20] In contrast, if a researcher mistakenly excludes from equation (1) a relevant control variable (that is, a variable that in fact influences a decisionmaker's behavior) then the estimate of disparate treatment may be biased away from its true value.[21] The expected bias of estimated coefficients caused by failing to include relevant controls is what statisticians called excluded or omitted bias.

---

[19] A successful refutation of a defendant's asserted reason for an adverse outcome permits, but does not require, a finding of discrimination. St. Mary's Honor Ctr. v. Hicks, 509 U.S. 502, 511 (1993); Anderson v. Baxter Healthcare Corp., 13 F.3d 1120, 1123 (7th Cir. 1994). Moreover, a plaintiff may not survive summary judgment if the plaintiff only refutes as pretextual only one of several reasons for the defendant's decisions. Coco v. Elmwood Care, Inc., 128 F.3d 1177, 1178 (7th Cir. 1997). *But see* Monroe v. Children's Home Ass'n, 128 F.3d 591, 593 (7th Cir. 1997) (plaintiff need not rebut all of defendant's reasons).

[20] An estimated parameter is "unbiased" if the expected value of the estimated parameter is equal to true value of the underlying parameter. *See* WILLIAM H. GREENE, ECONOMETRIC ANALYSIS 117-124 (2d ed. 1993).

[21] For example, consider a variant of equation (1) above in which the true specification for the dependent variable (y) is:

$$y_i = x_i\beta + z_i\delta + u_i, \qquad i = 1, \ldots, n$$

It can be shown that estimating this regression excluding control z will yield estimates for the beta coefficients that in expectation may not be equal to (and therefore biased) their true values:

$$\begin{aligned} E[\hat{\beta}] &= \beta + (X'X)^{-1}X'Z\delta \\ &= \beta + \text{bias}. \end{aligned}$$

*See* WILLIAM H. GREENE, ECONOMETRIC ANALYSIS 245-246 (2d ed. 1993); *Omitted-variable bias*, WIKIPEDIA, http://en.wikipedia.org/wiki/Omitted-variable_bias (last visited Aug. 8, 2010).

If a disparate treatment regression fails to include (or "omits") a non-race 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 only hiring applicants with a high school diploma, and if (2) the pool of applicants without diplomas is disproportionately comprised of minorities, then omitting from the regression whether applicants graduated from high school might bias the test of disparate treatment. The regression's estimate of disparate treatment ($\beta_1$) might superficially indicate that applicant race influences an employer when in fact the decision is solely driven by the presence or absence of a diploma.

The asymmetric impact of expected statistical bias has led many econometricians to "play it safe" when specifying regression equations by being over-inclusive when deciding whether to include marginally plausible controls. If the controls are in fact irrelevant, including them will not bias the estimates of the coefficients of interest. If the controls are in fact relevant, excluding them may bias the coefficients of interest. When in doubt, many statisticians worried about bias are trained to err on the side of inclusion. The cost of this "kitchen sink" approach is traditionally thought to be a loss in the precision of the coefficient estimates. Inappropriately including irrelevant controls will not bias the estimates of the included coefficients, but it will reduce the precision with which these coefficients are estimated. In disparate treatment regressions, inappropriately including irrelevant controls will not bias the estimate of disparate treatment ($\beta_1$), but it will reduce the ability to test whether that coefficient is statistically significant.[22]

---

[22] A standard response to this bias/precision tradeoff is for statisticians to run a series of "nested" disparate regressions in which each successive regression has an increasing number of controls. In this way, a researcher can

The "kitchen sink" approach has become a standard method of testing for disparate treatment discrimination. Under *McDonnell Douglas*, a class of plaintiffs claiming a pattern and practice of disparate treatment may not need to present evidence controlling for a wide variety of non-race attributes. But after a defendant articulates a legitimate, nondiscriminatory reason for the plaintiffs' treatment, the burden shifts back to the plaintiffs to establish that notwithstanding these other considerations, race was still a motivating factor.[23] As a statistical matter, this often means showing that the disparate treatment coefficient continues to be statistically significant even after controlling for additional variables. While the theory behind the "kitchen sink" approach is well settled, the application in actual litigation routinely contested. Defendants often argue that the plaintiffs' regressions suffer from excluded variable bias because the regressions fail to control for all of the non-actionable factors that (might have) influenced the defendant's decision. Plaintiffs in turn routinely respond that these excluded factors are at best pretextual rationales for defendants' behavior.

Often the data for the additional factors is missing, or would be prohibitively expensive to obtain in any credible fashion. For example, consider a class of minority automobile consumers claiming that a dealership discriminated against them in negotiating the price of the cars sold. The plaintiffs might present a disparate treatment regression controlling for a host of factors concerning the type of cars and the timing of the sale, showing that the sale prices for cars sold minorities was significantly higher than the prices of those sold to white customers. But the defendants would likely respond that the regression would also need to control for a long

---

explore whether the size and significance of a disparate treatment estimate is robust to the inclusion of potentially irrelevant variables. For an example of this approach, see IAN AYRES, PERVASIVE PREJUDICE? UNCONVENTIONAL EVIDENCE OF RACE AND GENDER DISCRIMINATION 100-105 (2003) (reporting regressions from Atlanta car dealership data).

[23] *See McDonnell-Douglas*, 411 U.S. at 802-804; *Hicks*, 509 U.S. at 505-507.

list of additional factors concerning the bargaining skills and preferences of the individual consumers and salespeople. If a dealership offers higher prices to customers with less experience negotiating, and if minority customers of this dealership have less experience negotiating, then a discrimination regression which excluded negotiating experience as a control might overstate the estimated amount of disparate treatment because of omitted variable bias. The absence of data – especially the absence of data on individual employees or customers – at times represents a substantial barrier to producing credible evidence of disparate treatment. This is particularly true in class action litigation, where some courts are reluctant to certify a disparate treatment class unless there is a common methodology for establishing the core question of defendant discrimination.[24]

B. *A "Business-Justification" Approach to Disparate-Impact Testing*

In disparate-impact testing, everything changes. Because disparate-impact plaintiffs need not prove that race was a motivating factor in the defendant's decision, disparate-impact tests should not implement the standard disparate-treatment regression (of equation (1) above). The purpose of that regression was to test whether race influenced the defendant's decisions. In disparate impact litigation, a core question is whether defendant's policies have a disparate effect on different racial groups. *Prima-facie* evidence of disparate racial impacts is often captured by mere comparison of averages. For example, if 60% of an employer's white applicants are offered jobs, while job offers are made to only 5% of an employer's Hispanic applicants, then the employer's hiring practices are having a disparate racial impact. Indeed, a stripped down version

---

[24] Carson v. Giant Food, Inc., 187 F. Supp. 2d 462, 471-72 (D. Maryland 2002) ("highly individual nature of Plaintiffs' individual claims" preclude class certification). See also Walsh v. Ford Motor Co., 807 F.2d 1000, 1005 (D.C. Cir. 1986) (Ruth Bader Ginsburg) (plaintiffs must show that "classwide proof applies" commonly to all class members).

of the regression equation (1) can statistically test for such racial disparities in averages:

$$\text{Defendant Decision} = \alpha + \beta_1 * \text{Minority} + \varepsilon. \qquad (2)^{25}$$

Thus, for example, taking a dataset on applicant hiring decisions, this specification (which simply regresses a defendant's decision to hire on just a single variable, applicant race) will return a value ($\alpha$) for the average hiring rate of white applicants as well as a measure ($\beta_1$) of the disparate effect or disparate impact. The estimated minority coefficient ($\beta_1$) specifically measures any difference between the average white and minority hiring rates. If as in the earlier example, an employer hires 60% of white applicants and 5% of minority applicants, then the regression will compute $\alpha$ to be .6 and $\beta_1$ to be -.55. While $\beta_1$ in the disparate-treatment specification (equation 1) was an attempt to measure the causal influence that race had on a defendant's decision, the same coefficient ($\beta_1$) in the disparate-impact specification (equation 2) now has a very different statistical meaning. Now the race coefficient merely measures the average differential impact that all defendant policies – including policies that are not race-contingent – have on minorities relative to whites. These averages could, of course, be calculated without pulling out the heavy machinery of a regression. But an advantage of estimating equation (2) is that the regression results report not only average differential impact on minorities, but also report whether this estimated impact is statistically significant.[26] I will refer to this stripped-down equation (2) as the "*prima-facie* DI" specification, because it is best suited to test whether defendant decisionmaking policies and practices differentially impacted

---

[25] When the defendant's decision is dichotomous, such as a hiring (taking on a value of 1 if applicant is hired, and 0 otherwise), it will often be appropriate to adopt a regression procedure such as logit or probit that more efficiently takes into account the restricted possible outcomes. In contrast, when the defendant's decision takes on continuous values, such as setting the sale price for a car or the APR for an interest rate, it will often be more appropriate to utilize an ordinary least square (OLS) procedure. *See* GREENE, *supra* note 21, at 175.

[26] On-line disparate impact calculators are available which will also automatically estimate the statistical significance of the average differential. *Disparate Impact Analysis,* http://www.hr-software.net/EmploymentStatistics/DisparateImpact.htm (last visited Aug. 11, 2010).

minorities.

A major limitation of the *prima-facie DI* approach is that it fails to take account of any business justifications that a defendant might have for its practices. Under the Civil Rights Act of 1991, an employer is not liable for a practice that produces disparate racial impacts if the employer can show that "the challenged practice is job related for the position in question and consistent with business necessity."[27] A hiring requirement that airline pilots have a pilot's license might have a disparate racial impact (if minority applicants were less likely than white applicants to have the required license), but this hiring requirement would not make out an actionable case for disparate-impact liability because a pilot's license is "job related for the position in question and consistent with business necessity." Analogous business-justification defenses exist in other disparate-impact contexts. For example, consider a lender who has a policy of charging higher interest rates to borrowers with poorer credit scores. If minority borrowers have poorer credit scores, the policy might have a disparate racial impact. The *prima-facie* regression specification (equation 2) would indicate that minorities pay significantly higher APRs than whites. But a defendant/lender might easily be able to establish that the policy is business-justified to cover the heightened cost in defaults from borrowers with poorer credit scores.

The presence of a business-justification defense suggests the need for a specification that tests to see whether disparate racial impacts persist after taking into account legitimate organizational rationales for defendant practices. Luckily, a statistical specification exists that provides just such a test. The specification, which I will call the "unjustified DI" specification, utilizes an intermediate number of controls that are nested between the previous kitchen-sink DT

---

[27] Civil Rights Act of 1991, 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2006).

specification (equation (1)) and the stripped-down "*prima-facie DI*" specification (equation 2):

$$\text{Defendant Decision} = \alpha + \beta_1 * \text{Minority} + \sum_i \beta_i * (\text{Plausible Business-Justified Influences}) + \varepsilon \qquad (3)$$

Equation (3) is identical to equation (1), except that instead of including the larger list of all "plausible non-race influences," equation (3) only includes the smaller list of "plausible business-justified influences." The category of business-justified influences is a subset because (i) any plausible business-justified influence would be a plausible non-race influence, but (ii) some plausible non-race influences might not be business justified.[28]

For example, as discussed above, a high-school diploma might be a plausible non-race-contingent factor influencing hiring at Duke Power, but courts found that it was not a business-justified influence. Accordingly, this influence would be included in a kitchen sink DT specification, but not in equation (3)'s unjustified DI specification. However, it is probably true that being able to stand on your feet and push a broom would have been a business-justified hiring requirement. Accordingly, it would be appropriate to include whether applicants had this physical ability as a control in either the kitchen-sink DT or the unjustified DI specifications.

This shift to the smaller set of business-justified controls in equation (3) dramatically changes how the key racial disparity estimate ($\beta_1$) is interpreted. The DT specification is attempting to establish *causal* relationships.[29] A finding from equation (1) that $\beta_1$ is negative and statistically significant is evidence that defendant decisionmaking was race-contingent. It is a finding that plaintiff's race caused defendant's decision. But by intentionally excluding plausible

---

[28] As a matter of law in the employment context, there are no business-justified race-contingent influences, because race is never a bona fide occupational qualification (BFOQ). *See* 42 U.S.C. § 2000e-2(e) (2006).

[29] Clogg and Haritou identify the precise and daunting assumptions that must hold for what they call the "regression method of causal inference" to be valid. C.C. Clogg & A. Haritou, *The Regression Method of Causal Inference and the Dilemma Confronting This Method*, in Causality in Crisis 83-112 (R. McKim & S.P. Turner, eds. 1997).

causal influences, the unjustified DI specification is no longer attempting to establish the causal relationships that drove the defendant's decisionmaking. A finding from equation (3) that $\beta_1$ is negative and statistically significant no longer is evidence that defendant's decisionmaking was race-contingent. Instead, the estimated $\beta_1$ from equation (3) should be interpreted merely as a *conditional average*. Just as $\beta_1$ from equation (2)'s *prima-facie DI* specification was nothing more than the (unconditional) average racial differential, the $\beta_1$ from equation (3)'s unjustified DI specification is nothing more than the conditional average racial differential – where the average racial differential is estimated conditional on the values of the business-justified influences.

Thus in a stylized *Griggs* equation (3) regression, imagine that we excluded from the specification applicant's high-school diploma status (but included various business-justified factors, including certain physical abilities) and found a negative and statistically significant $\beta_1$. We should not conclude that applicant race was a motivating factor in Duke Power's hiring decision or that defendant engaged in race-contingent decisionmaking. But we should conclude that on average minority applicants received fewer employment offers from the defendant than white applicants even after controlling for business-justified characteristics. The regression would indicate that minority applicants who were similarly situated with regard to business-justified characteristics were adversely impacted by defendant's decisionmaking. Again, this does not mean that defendant's decisionmaking was race-contingent. But it does mean that it was contingent on a factor that was correlated with race, but which was not business justified. It means that Duke Power's decisionmaking produced an unjustified disparate impact.

The upshot of this more appropriate approach is that statisticians in disparate impact testing need to resist the temptation to err on the side of inclusion. Statisticians are trained early

on to include all and any causally-related influence factors in their regression specifications. Leading texts claim that the consequences of including irrelevant variables are "generally less serious than those pertaining to the exclusion of relevant variables."[30] The concern adumbrated above over "omitted variable" bias looms large.[31] Statisticians often reflexively feel that including irrelevant control variables will reduce the precision of the other estimated causal coefficients, but these other coefficients will remain unbiased. The impetus for inclusion is even stronger in specifying controls to test whether defendant "discriminated" because the concept of "discrimination" is still linguistically tied in many non-lawyers' minds to the common-usage idea of race-contingent decisionmaking.

But the foregoing analysis suggests that in disparate-impact disputes, it is necessary to intentionally exclude from the regression controls for certain factors even if those factors are thought to be causally related to the decision being modeled. When testing for unjustified disparate impacts, "included variable bias" – the converse of the "omitted variable bias" – should be a central concern.[32] Including controls for non-race factors that do not represent legitimate organizational justifications can induce included variable bias – in that the estimate of whether a

---

[30] Jack Johnston & John DiNardo, Econometric Methods 110 (McGraw Hill 4th ed.1997).

[31] Ian Ayres, *Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias*, PERSP. BIOLOGY & MED., Winter 2005, at S68 (discussing problem of included variable bias in medical outcomes). *But see* Robert Bornholz & James J. Heckman, *Measuring Disparate Impacts and Extending Disparate Impact Doctrine to Organ Transplantation*, PERSP. BIOLOGY & MED., Winter 2005, at S95 (reply).

32. The term "included variable bias" was first used by Clogg and Haritou, *see supra* note 29. While the text emphasizes the possibility of included variable bias in disparate treatment regressions, including additional variables may increase estimated coefficient bias under limited circumstances in disparate treatment testing as well. In the frequent example, when multiple relevant controls are missing, the inclusion of a subset of the controls has ambiguous effects on the degree of bias on the coefficient of interest. Kevin A. Clarke, *Return of the Phantom Menace: Omitted Variable Bias in Econometric Research*, CONFLICT MGMT. & PEACE SCI. 26:1 (February 2009); Kevin A. Clarke, *The Phantom Menace: Omitted Variable Bias in Econometric Research*, CONFLICT MGMT. & PEACE SCI. 22:4 (Winter 2005).

decision maker's policies produced an unjustified disparate impact may be biased away from their true value.

The methodological confusion over the appropriate set of non-race controls is more than just a theoretical concern. In real world statistical studies, the inappropriate inclusion of factors that do not represent a plausible organization justification have led researchers to misestimate the unjustified racial disparities produced by particular decisionmaking policies. For example, in 2006, the Analysis Group released a report commissioned by the City of Los Angeles analyzing more than 810,000 field data reports (FDRs) collected by the Los Angeles Police Department (LAPD) from July 1, 2003 through June 30, 2004.[33] FDR dataset are completed by LAPD officers whenever an officer stops a vehicle or pedestrian.[34] The FDR data included information on a number of outcomes—including: i) whether a pat-down, frisk or search was conducted; ii) whether contraband was uncovered; and iii) whether an arrest was made or a citation was issued. The Analysis Group analyzed the dataset using multivariate regression to determine whether there was evidence of "racially biased policing."[35] They concluded:

> Although some divisions/bureaus have statistically significant racial disparities for some outcomes and some races, when evaluated across all outcomes, *there is no consistent pattern of race effects across divisions or races*.[36]

But the regression testing approach taken by the Analysis Group failed to distinguish between the core difference between disparate treatment and unjustified disparate impact discussed above.

---

[33] A copy of the Analysis Report is available online. ANALYSIS GROUP, INC., PEDESTRIAN AND MOTOR VEHICLE POST-STOP DATA ANALYSIS REPORT (2006), http://www.analysisgroup.com/AnalysisGroup/uploadedFiles/Publishing/Articles/LAPD_Data_Analysis_Report_07-5-06.pdf (hereafter ANALYSIS REPORT).

[34] Officers did not need to complete an FDR for stops at checkpoints/roadblocks, commercial vehicle safety inspections, stops pursuant to an arrest or search warrant, stops of victims/witnesses, and stops involving calls for service relating to certain particularly dangerous crimes and situations. Ian Ayres & Jon Borowsky, A Study of Racially Disparate Outcomes in the Los Angeles Police Department, prepared for the ACLU of Southern California at 1, *available at* www.aclu-sc.org (October 2008).

[35] ANALYSIS REPORT, *supra* note 33, at 3, 6 n.3.

[36] *Id*. at 4.

The Analysis Group's regressions included a variety of controls — such as the number of complaints that have been levied against the stopping officer — that would not provide a plausible organizational justification for disparate policing behaviors. As Jonathan Borowsky and I wrote in responding to the Analysis Report:

> Controlling for officer complaints might make sense in a test of disparate racial treatment by the officer, because it would be appropriate to control for all non-race factors that might provide alternative (non-pretextual) explanations for a racial disparity in outcomes. But it would be inappropriate to control for officer complaints in a test of disparate racial impacts. Including controls for officer complaints might easily cause a regression to understate the true size of the unjustified racial impact. A policy of assigning officers with multiple complaints to predominantly-minority areas might produce an unjustified impact against minorities who are stopped. Including a control for officer complaints might inappropriately soak up some of the real racial disparity in the data.[37]

Even if individual officers were not engaging in disparate treatment, the department policies might have produced an unjustified disparate impact. This example shows that the failure to attend to the problem of included variable bias is more than just a theoretical concern. When Borowsky and I reran the Analysis Group's regression specification, which included a host of officer attributes that do not offer a plausible organization justification for racial disparities,[38] we still found that that police were about 50% more likely to make a search request to stopped African Americans than stopped whites. But when we ran the same regression excluding these inappropriate controls the racial disparity increased to 76.4%. As in the stylized *Griggs* example, including inappropriate controls can bias downward the estimate of unjustified racial disparities.

---

[37] Ayres & Borowsky ACLU Report, *supra* note 34 at 4.

[38] This specification, following the Analysis Report, inappropriately included controls for: Count of Major Commendations Received by Officer, Officer Age, Officer Gender, Officer Race, Number of Months of Service of Officer, Number of Months of Service of Officer Squared, Officer Assignment (Traffic, Patrol, Other), and Officer Race Interacted with Suspect Race. *See* Ayres & Borowsky ACLU Report, *supra* note 34 at 16-17.

Other scholars have seen the possible biasing effects of including too many control variables in discrimination regressions. However, they fail to make the crucial distinction between the disparate impact and disparate treatment testing. For example, a statistical guide for judges and lawyers uses a stylized version of *Griggs* to emphasize how mistakenly including irrelevant variables can bias a regression's estimate of the racial effect:

> Lastly, and perhaps most important under the heading of legitimacy, is the problem of tainted independent variables. Suppose a regression analysis includes a variable for education that, in a race case, is a key determinant of salary differences between black and white employees in a clearly different job group. Regression analysis indicates a high t-statistic on education and an insignificant t-statistic on the race coefficient. Given that in almost all groups, white employees have received more formal education than black employees, it would appear that education goes a long way towards explaining salary differences between black and white employees. The burden is on the employer, however, to demonstrate separate from the regression, that education was required and affected performance, and hence directly determined salary. To the extent that education is not related to job performance, it is an inappropriate variable to use in a regression. Excluding key variables and including irrelevant variables have the same impact.[39]

Similarly, John Yinger has described how including illegitimate control variables in a discrimination regression can cause included variable bias (what he calls "diverting variable bias"):

> Diverting variable bias arises when a variable that is not a legitimate control variable, but that is correlated with race or ethnicity, is included in the regression. The key issue, of course, is how to define what variables are "legitimate." Under most circumstances, economists are taught to err on the side of including too many variables. In this case, however, illegitimate controls may pick up some of the effect of race or ethnicity and lead one to conclude that there is no discrimination when in fact there is.[40]

---

[39] Thomas R. Ireland *et. al*, EXPERT ECONOMIC TESTIMONY: REFERENCE GUIDE FOR JUDGES & ATTORNEYS (Lawyers & Judges Publishing Co. 1998).

[40] John Yinger, *Evidence on Discrimination in Consumer Markets*, 12 J. ECON. PERSP., 23, 27 (1998).

These claims capture the essence of the included variable bias concern. But neither of these analyses is correct when applied to disparate treatment testing. If an employer relies on applicants' educational attainment in making hiring decisions, there should be no disparate treatment liability even if the educational attainment is shown to be irrelevant to job performance. The possibility for included (or diverting) variable bias arises instead when asking whether an unjustified hiring criterion induces a disparate racial impact. An unjustified disparate impact cannot be explained away by showing that a decisionmaker was influenced by some factor that did not constitute a business justification.

i. *Simpson's Paradox*

One way that both courts and scholars have couched a concern with inadequate statistical controls in discrimination testing is by referring to the possibility that inadequate controls will lead to a "Simpson's Paradox" — which refers to "illusory disparities in improperly aggregated data that disappear when the data are disaggregated."[41] For example, scholars analyzing 1973 admission data from the University of California, Berkeley, uncovered "a clear but misleading pattern of bias against female applicants" because the uncontrolled, aggregate analysis showed that women applicants had a lower overall acceptance rate than men applicants, even though many of the departments admitted women at a higher rate than men.[42]

---

[41] Eng'g Contractors Ass'n of S. Fla. Inc. v. Metro Dade County, 122 F.3d 895, 919 n.4 (11th Cir. 1997). *See also* Dukes v. Wal-Mart Stores, Inc., Nos. 04-16688, 04-16720 (9th Cir. Apr. 26, 2010) (en banc), *available at* http://www.ca9.uscourts.gov/datastore/opinions/2010/04/26/04-16688.pdf; Stagi v. Nat'l R.R. Passenger Corp., No. 03-5702, 2009 WL 2461892 (E.D. Pa. 2009).

[42] P.J. Bickel, E.A. Hammel, & J.W. O'Connell, *Sex Bias in Graduate Admissions: Data from Berkeley*, 187 SCIENCE 398 (1975). Defendants in civil rights actions usually deploy the *possibility* of Simpson's Paradox as a concern without showing the kind of reversal found at Berkeley (or in the stylized university example in the text). *See, e.g.*, *Dukes*, Nos. 04-16688, at 6200 n.30. Actual examples of Simpson's Paradox are fairly rare in real world data. *See also* Marios G. Pavlides & Michael D. Perlman, *How Likely is a Simpson's Paradox?*, 63 AM. STATISTICIAN 226 (2009).

A stylized version of this university example can help us understand how the Simpson's Paradox operates. Imagine there is a university with just two graduate departments (math and English). Of the 1,000 women who apply for graduate admission, imagine that 90 percent apply to the English department and that 10 percent apply to the math department. In contrast, imagine that the 1,000 male applications are evenly divided between the two graduate departments. Finally, imagine that in each department, the admission rate for women is higher than that for men but that the admission rate in the English department for both male and female applicants is markedly lower than in the math department. Specifically, imagine the departments admit men and women at the following rates:

| | Women | Men |
|---|---|---|
| Math | 82% | 80% |
| English | 22% | 20% |

Under these conditions, the overall admission rate of men applicants at the university would be 50 percent, while the overall admission rate for women at the university would be only 28 percent.[43] The paradox in this example is that even though women have a higher admission rate than men in each department, they nonetheless have a lower admission rate for the university as a whole. Failing to control for department effects in a regression would seem to give a false indication of a gender disparity disfavoring women, when in fact women have a statistical advantage of 2 percentage points in each department.

But this concern about the possibility of a Simpson's Paradox again ignores the important differences between disparate treatment and disparate impact testing. In a disparate impact claim,

---

[43] In this example, a total of 280 women would be admitted (82 of 100 would be admitted to the math department and 198 of 900 would be admitted to the English department) and 500 men would be admitted (400 of 500 would be admitted to the math department and 100 of 500 would be admitted to the English department).

where intentional discrimination need not be proven, defendant policies that produce unjustified racial or gender disparities in the aggregate may give rise to liability even if there is no disparity in subsets of the data. Thus in a *Griggs*-like setting, if Duke Power had hired 100% and 99% of blacks and whites (respectively) with diplomas, and only 1% and 0% of blacks and whites (respectively) without diplomas, the unjustified diploma policy might still produce an aggregate disparate racial impact since African-Americans were disproportionately less likely to have diplomas. Similarly in the foregoing university example, the university's policy of admitting a much higher proportion of math applicants than English applicants has an aggregate disparate impact on women applicants because women applicants disproportionately apply to the English department.[44]

In a disparate-impact analysis, controlling for the tendency of the different departments to admit students at different rates would only be appropriate if the university could establish a business justification for its much lower acceptance rate in the department dominantly applied to by women. If the University has a justification for the lower admission rate in the department disproportionately applied to women, then it would be appropriate to control for this factor in a regression testing for unjustified disparate impacts. The Simpson's Paradox anxiety is another example in which disparate-treatment thinking inappropriately infects disparate-impact claims. In disparate treatment analysis of admissions, it would be presumptively appropriate to control for all non-pretextual factors including controls for the individual department effects. Evidence that women applicants were favored in each individual department would be strong evidence against a disparate-treatment claim that women were disfavored by sex-contingent university

[44] In addition, the university may have policies that cause a disparate impact because they tend to induce women to disproportionately apply to the English department.

decisionmaking. In disparate-treatment analysis, the individual department results should trump the reversed finding of disparity in the aggregate data. But the individual department results should not automatically trump the aggregate disparity finding in disparate impact analysis. At the end of the day, a Simpson's Paradox discrimination reversal *only* can occur if some uncontrolled characteristic — like the applicant department or the applicant diploma status — is correlated with the plaintiffs' protected class *and* disfavored by defendant's decisionmaking. From a disparate impact perspective, the key issue is not the possible reversal of the estimated disparity but whether the defendant is justified in disfavoring a category that is disproportionately represented by plaintiff's class.

ii. *What is Business Justified?*

To implement the unjustified DI specification (equation (3)) one needs to identify and control for "plausible business-justified influences." Theories of business justification in disparate impact setting have been hotly contested.[45] In employment, the Civil Rights Act of 1991 defines a defendant's policy as unjustified if, "the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity . . . ."[46] In lending, the test is whether a defendant's policy "meets a legitimate business need."[47] More generally, one could ask whether a challenged policy furthers a legitimate organizational objective.

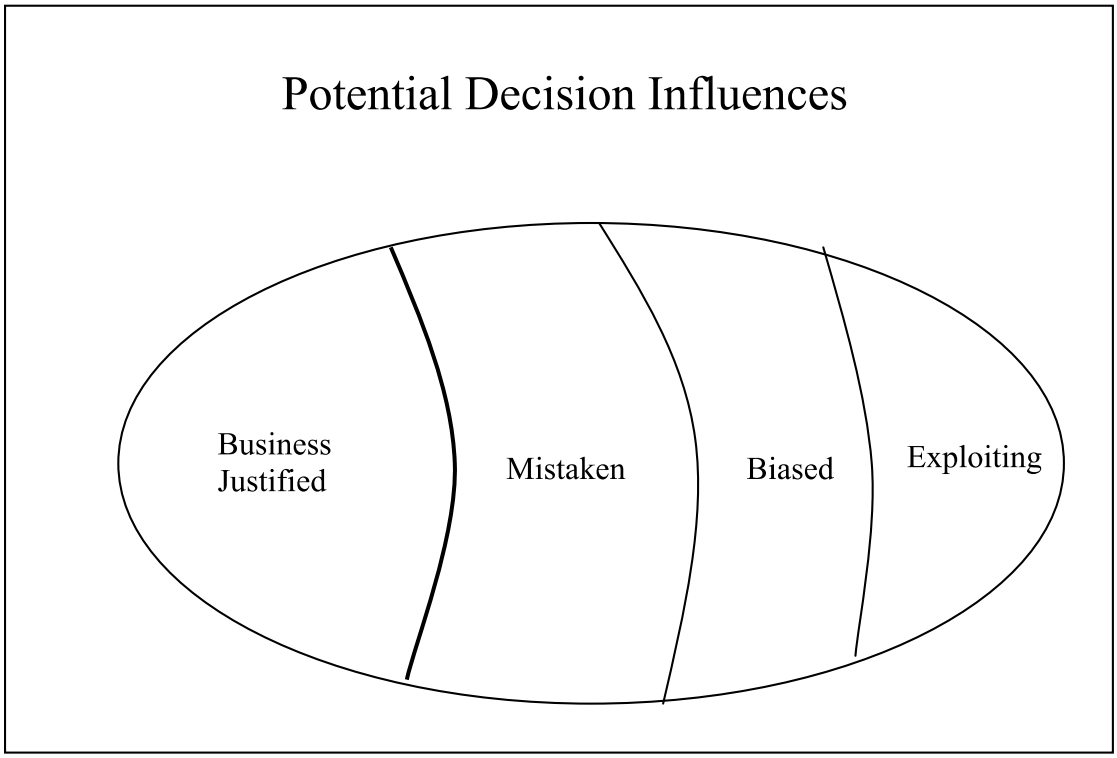
---

[45] President George H.W. Bush felt so strongly that the "business necessity" definition created a "quota bill" – that is, a de facto quota employment requirement – that he vetoed the 1990 legislation. *See, e.g.*, Neil A. Lewis, *President's Veto of Rights Measure Survives By One Vote*, N.Y. Times, Oct. 25, 1990, at A1. The 1990 bill defined the term "required by business necessity" to mean "essential to effective job performance." S. 2104, 101st Cong. § 3(o). *See* Ian Ayres & Peter Siegelman, *The Q-Word As Red Herring: Why Disparate Impact Liability Does Not Induce Hiring Quotas,* 74 TEX. L. REV. 1485 (1996).

[46] 42 U.S.C.A. § 2000e-2(k)(1)(A)(ii) (1994).

[47] The quoted language comes from commentaries on ECOA regulation: "The act and regulation may prohibit a creditor practice that is discriminatory in effect because it has a disproportionately negative impact on a prohibited

If, on the ground, disparate impact testing is going to be different than disparate treatment testing, researchers must have a theoretically and empirically defensible ground for distinguishing between the "plausible non-race influences" which are appropriate controls in the DT specification and the smaller list of "plausible business-justified influences." Figure 1 presents a Venn diagram depicting three categories of decision influences: 1) mistaken, 2) biased, and 3) exploiting factors. These three categories represent non-race factors that actually influenced a decision, but are not business justified.




A "mistaken" influence is a factor is that a decisionmaker relies upon to further the profitability of institution's interests, but which in fact does not further that interest. Thus, if a bail bond dealer mistakenly thought that having a common-law as opposed to a statutory spouse

---

basis, even though the creditor has no intent to discriminate and the practice appears neutral on its face, unless the creditor practice meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact." Official Staff Interpretations, Regulation B (Equal Credit Opportunity), 12 C.F.R. §202.6(a)-2 (2009).

made a defendant a higher flight risk, then charging defendants with common-law spouses higher bail amounts than defendants with statutory spouses might produce an unjustified disparate impact.[48] Similarly, if the human resources at Duke Power mistakenly believed that a high-school diploma requirement would tend to lead to more productive workers, then conditioning hiring on this mistaken belief might have produced an unjustified disparate impact.

In contrast, a "biased" influence is one that is not mistaken but furthers an interest that is not related to the business's or institution's interests. Thus, a human resources director who chooses to favor graduates of her own alma mater—not because they are more productive, but merely to indulge a personal preference for her former school—could by this favoritism induce a disparate racial impact that is not justified as furthering the business's (or institution's) legitimate interests. These biased influences might be seen as a species of agency costs – where the agent's personal preferences cause decisionmaking that diverges from the principal's interest.[49]

The disciplining effects of competition might tend to limit the prevalence of both mistaken and biased decisional influences. Since both of these influences fail to further the institution's own interests, the institution is incentivized to establish procedures to eliminate mistaken or biased decisionmaking. Moreover, courts might lack the institutional competence to accurately judge whether a specific criterion is in fact mistaken. The Supreme Court in *Griggs* was willing to accept the lower court's determination that a high school diploma was not related

---

[48] Ian Ayres & Joel Waldfogel, *A Market Test for Race Discrimination in Bail Setting*, 46 STAN. L. REV. 987 (1994).

[49] At times, it may be difficult to determine the principal's true interest. One can imagine organizing a corporation whose stated objectives include supporting the graduates of a particular college. But at least with respect to for-profit corporations, there is often a default presumption that the objective is to lawfully maximize profitability.JONATHAN R. MACEY, CORPORATE GOVERNANCE 6 (2008).

to the ability to perform the jobs at issue.[50] But one should keep in mind that courts themselves might be mistaken (and inadequately incentivized) to determine which factors are job related.

On the other hand, a defendant business or organization does not internalize all the consequences of its decisionmaking. Indeed, the whole point of disparate impact litigation is that a pattern or practice of decisionmaking can disproportionately disadvantage an external group. Accordingly, an organization deciding how much effort to take to ensure non-mistaken and unbiased decisionmaking may not account for these external costs. If taken to the extreme, this argument would also preclude any claims of non-statistical disparate treatment, with the conclusion being that the organization already has an incentive to eliminate any disparate treatment by its decisionmakers.[51] But in a world where organizations are not able to purge all disparate treatment from their decisionmaking, it is all the more possible that organizations will not be able to eliminate the mistaken and biased decision criteria that produce disparate racial impacts. Indeed, the categories of mistaken and biased influences might be related to pretextual factors that merely are stand-ins for race in the decisionmaking process.

The final category of "exploiting" factors stand on a very different footing with regard to organizational incentives. As I have argued elsewhere,[52] businesses may have an incentive to engage in practices that enhance their profitability merely by exploiting market power in ways that should not constitute a business justification for disparate impact purposes. For example, a

---

[50] *See Griggs*, 401 U.S. at 431 ("On the record before us, neither the high school completion requirement nor the general intelligence test is shown to bear a demonstrable relationship to successful performance of the jobs for which it was used. Both were adopted, as the Court of Appeals noted, without meaningful study of their relationship to job-performance ability.")

[51] The exception might be so called "statistical discrimination," that is, disparate racial treatment that as a statistical matter furthers the institution's interests. *See* Edmund Phelps, *The Statistical Theory Of Racism And Sexism*, 62 AM. ECON. REV. 659-661 (1972); Steven Coate & Glenn Loury, *Will Affirmative Action Eliminate Negative Stereotypes?*, 83 AM. ECON. REV. 1220-40 (1993).

[52] Ian Ayres, *Market Power and Inequality: A Competitive Conduct Standard for Assessing When Disparate Impacts are Justified*, 95 CALIF. L. REV. 669 (2007).

policy of price-gouging a class of customers may enhance a firm's profitability, but it is not "consistent with business necessity." No court has ever accepted price-gouging as a business justification. It is *prima facie* a business justification for a decisionmaker to cover its cost (including earning a risk-adjusted economic return on capital), but it is not a business justification to engage in practices that disproportionately disadvantage a protected class solely to earn a supra-competitive profit. Accordingly, an employer might be justified to pay employees who are primary caregivers less if the employer can demonstrate that these employees are less productive (for example, more likely to be absent from work). But it would not be justified for the employer to pay the workers less simply because they were more necessitous or less able to search for alternative employment. In both cases, the policy might enhance an employer's profitability. But only the former rationale would qualify as "job related for the position in question and consistent with business necessity."[53] While the defendant's profit motive might tend to limit the prevalence of mistaken and biased decisional influences, the same incentive would tend to exacerbate a decisionmaker's impulse to engage in wage or price gouging.

This analysis of business necessity gives a firmer footing for the existing proposals of others. For example, the statistical guide for judges and lawyers previously quoted proposes limiting non-race controls to those that affect job performance in concluding: "To the extent that education is not related to job performance, it is an inappropriate variable to use in a regression."[54] In the credit context, other scholars have similarly applied a performance standard

---

[53] *Id.* at 707-708. *But see* James J. Heckman, Expert Report, Coleman et al. v. GMAC, No. 3-98-0211 (M.D. Tenn. Dec. 15, 2003) (on file with author) (rejecting this example as "clever and polemical.").

[54] Ireland, *supra* note 39.

for determining what characteristics are irrelevant and hence should be excluded from a discrimination regression:

> Discrimination occurs whenever the terms of a transaction are affected by personal characteristics of the participants that are not relevant to the transaction. In credit markets, discrimination on the basis of race and/or gender exist if loan approval rates or interest rates charged differ across groups with equal ability to repay.[55]

Again, it is legitimate to control for factors that relate to a person's probable performance of her contractual commitment – which in the credit context is chiefly whether or not the loan will be repaid:

> Discrimination may be apparent if banks approve loans to equally credit-worthy minority and white-owned firms, but charge the minority-owned firms a higher rate of interest.[56]

Focusing on creditworthiness or the likelihood of repayment is also consistent with a business-justification standard that focuses on a decisionmaker covering its costs of doing business. Borrowers who fail to pay off their loans can impose substantial costs on a lender.[57] Extracting supra-competitive revenues from a class of consumers – not because they impose higher costs on a seller but merely because the seller has the power to do so – is not consistent with business necessity (and thus would constitute an unjustified disparate impact). Thus, John Yinger concludes: "According to the definition of discrimination used here, legitimate controls are those associated with a person's qualifications to rent or buy a house, buy a car or so on – or,

---

55. David G. Blanchflower, Phillip B. Levine, & David J. Zimmerman, *Discrimination in the Small Business Credit Market*, 85 The Rev. of Econ. & Stat., The MIT Press, vol. 85(4), at 930 (2003).

56. *Id*. at 940.

57. See *A.B. & S. Auto Service, Inc. v. South Shore Bank of Chicago*, 962 F. Supp. 1056 (N.D. Ill. 1997) ("[In a disparate impact claim under the ECOA], once the plaintiff has made the *prima facie* case, the defendant-lender must demonstrate that any policy, procedure, or practice has a manifest relationship to the creditworthiness of the applicant….In other words, the onus is on the defendant to show that the particular practice make's defendant's credit evaluation system more predictive than it would be otherwise."). See also *Lewis v. ACB Business Services, Inc.*, 135 F.3d 389, 406 (6th Cir. 1998) ("The Act was only intended to prohibit credit determinations based on 'characteristics unrelated to creditworthiness.'").

to use the legal term, those associated with business necessity."[58] Making a decision contingent on a factor that allows an organization to extract a higher price from costumers or a lower wage from employees merely as a way of extracting a supra-competitive profit is not by this reasoning business-justified.

The question of whether to control for exploiting market power (or realizing supra-competitive profits) is not merely an academic dispute. In lending discrimination cases alleging that lender policies produced unjustified disparate impacts in violation of the Equal Credit Opportunity Act, defendant decisionmakers routinely argue that regressions need to control for borrower sophistication. A prominent defense expert, Marsha Courchane, has published a regression analysis using proprietary mortgage data from a number of mortgage lenders concluding that "up to 90% of the African American APR gap, and 85% of the Hispanic APR gap, is attributable to observable differences in underwriting, costing, and market factors that appropriately explain mortgage pricing differentials."[59] But Courchane's regressions inappropriately include controls for the average level of education attainment in the borrower's neighborhood. Courchane includes this control because she hypothesizes that less sophisticated borrowers, all else equal, are more likely to take out high-APR subprime loans.[60] Somewhat ironically from the perspective of the earlier *Griggs* analysis, Courchane's regressions include a control for the percentage of neighborhood residents without a high-school diploma. Lender

---

[58.] Yinger, *supra* note 40 at 27.

[59] Marsha J. Courchane, *The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain?,* 29 J. REAL EST. RES. 399 (2007).

[60] Courchane says:

> Borrowers are assumed to interact with and seek advice from their neighbors, and that formal education and mortgage market experience both are associated with greater financial knowledge and literacy. As a result, it is hypothesized that the higher the tract population educational attainment and the greater the percentage of homeowners in the tract, the lower the probability of taking out a subprime mortgage.

*Id.* at 413.

policies that market high-cost subprime loans to neighborhoods with low-educational attainment can potentially work a disparate impact on minorities that is unrelated to the lender cost of underwriting. Contrary to Courchane's claim, the lack of a high-school diploma is not a factor which "appropriately explain[s] mortgage pricing differentials." Like in the earlier LAPD example, where researchers inappropriately included controls for officer complaints,[61] the choice to include controls for borrower sophistication runs the risk of understating the true unjustified disparate impact.

Regardless of what substantive standard is adopted for determining what qualifies as a business justification (and hence what should be included in the unjustified DI specification), the application of the standard will turn on facts and or reasoning that are external to the regression itself. Specifically, one cannot determine the appropriateness of a particular factor merely by looking to see whether it is shown in the regression to exert a statistically significant influence on the defendant's decisionmaking. On this dimension, disparate impact testing is different than disparate treatment testing. As a general tendency, when statisticians attempt to infer causal influences from regression testing, they tend to interpret a statistically-significant coefficient estimate for a particular control variable as evidence that the variable was appropriately included in the regression. But in disparate-impact regressions, where the goal is not to test whether race causally influenced a defendant's decisionmaking, the statistical significance of a control coefficient does not indicate that it was appropriate to include the variable in the regression. Again, the stylized *Griggs* thought experiment is a case in point. The mere fact that a variable indicating whether an applicant has a high school diploma is estimated to be statistically significant may tell us that this factor had a causal influence on Duke Power's decisionmaking

[61] *See supra* text accompanying note 33.

but it does not tell us anything about whether that influence was business justified or not. In fact, the statistical significance of estimated coefficient creates a larger likelihood that the disparate impact estimate ($\beta_1$) will suffer from included-variable bias.

An analogous argument applies to goodness-of-fit measures. Regression output includes measures, such as the R-squared or the adjusted R-squared estimates, for how well the regression was able to explain the dependent variable.[62] For example, an R-squared of 25% in a disparate impact regression would indicate that a regression was able to explain one-quarter of the variance in the defendant's decisionmaking. Omitting a control that in fact exerted a causal influence on defendant decisionmaking (but was not business justified) is likely to reduce the estimated R-squared measure of goodness-of-fit. In *Griggs,* one could imagine that a regression that included applicant diploma status would perfectly fit (i.e., explain) when applicants would or would not be hired. Excluding the diploma control would predictably reduce the goodness of fit measures – so the unjustified DI specification (equation (3)) would explain less of the variance in defendant's decisionmaking than the DT specification. But this reduction in the goodness of fit induced by the omission of the diploma control should not be taken as a weakness of the specification. Since the disparate impact regression is not attempting to suss out all the causal influences on the defendant's decisionmaking, the inability of the regression to capture all the nuances of the decisionmaking is not a weakness of the regression specification.[63] The statistical significance of an adverse disparate impact estimate ($\beta_1$) should be the touchstone of concern.

iii. *Capped Coefficients*

---

[62] GREENE, *supra* note 21 at 74.

[63] Even in disparate treatment regressions, the goodness-of-fit measure might be substantially below 100% if some aspects of the decisionmaking were random, arbitrary or capricious. For example, one would not expect a well-specified regression of major league pitching to perfectly explain when a pitcher would throw a curve or a fastball, because pitchers have strong strategic reasons to randomize some of their pitch selection.

The inclusion/exclusion choice can have dramatic impact on the disparate impact estimate ($\beta_1$). Including additional control variables – whether or not the included variables are appropriate business justifications – can explain away what would otherwise be a statistically significant disparity. Including additional controls can reduce the size of the estimated coefficient itself as well as the estimated level of statistical significance.[64] The choice whether to include or exclude seems to be a high-stakes, all-or-nothing decision.

Fortunately, however, there is a straightforward way to modify the unjustified DI specification (equation (3)) to limit the impact of an included control so that it does not explain away more of racial disparity than is justified by legitimate organizational interests. For example, imagine that a particular class of borrowers on average exposes a lender to $100 of higher costs. Imagine also that minority borrowers are more likely than non-minority borrowers to fall into this higher-cost class. In testing for unjustified disparate impacts, it would be appropriate to include in the regression a control for this cost-related attribute. However, what should be done if the regression specification (3) reveals that the lender was charging customers with this attribute a price that was $1100 higher than customers without this attribute? A lender trying to cover its cost would be justified charging borrowers with this attribute $100 more (even if that higher $100 charge disproportionately fell on minority borrowers). But, under the prior section's analysis of market-power exploitation, the supra-competitive charge of an additional $1,000 is not justified. In econometric terms, including an unrestricted control for this cost-related attribute would allow the regression control to explain away too much of the disparate

---

[64] Although in both theory and practice, including additional variables may reveal heightened racial disparities that are not found in less controlled regressions. For example, one might observe a community development bank where minority and non-minority borrowers had the same average loan approval rate, but where a statistically significant racial disparity was exposed after controlling for applicant credit history.

racial impact.

It is possible to modify regression specification (3) to cap the coefficient on cost-related variables so that only the justifiable amount of a possible racial disparity. In the foregoing hypothetical example, where specification (3)'s initial coefficient on the cost-related attribute was $1,100, it is possible to re-estimate the disparate impact coefficient ($\beta_1$) after capping the coefficient to be the cost-justified $100.[65] The re-estimated measure of racial disparity then indicates how much a racial disparity remains after allowing for the business justified enhanced charges for particular borrower classes.

Instead of making an all-or-nothing decision on whether to include or exclude a particular variable, the capped-coefficient approach effectively allows researchers to *partially* include the effects of a control in a regression where the degree of inclusion is naturally related to the degree of the business justification. Indeed, where there is uncertainty about the exact amount of a price enhancement that would be cost-justified, it is possible for researchers to vary the cap to see how alternative business justifications would impact the estimate of unjustified racial disparity.

iv. *Causation*

While plaintiffs in disparate impact litigation need not prove that race was a motivating factor or causal influence in plaintiffs' adverse treatment, they do need to establish that defendant's race-neutral practices caused a disparate racial impact. In employment, the Civil Rights Act of 1991, which codified the elements of a disparate impact claim, requires "a complaining party [to] demonstrate[] that a respondent uses a particular employment practice

---

[65] Capping a control coefficient can be accomplished by subtracting the capped effect from the decision variable and running the adjusted decision variable on the remaining right-hand side variable (excluding the capped coefficient variable). In the foregoing lending example, one would create an adjusted finance charge by subtracting $100*(higher-cost indicator) from the actual finance charge and then regressing this adjusted finance charge on the pre-existing right-hand- controls except without the higher-cost indicator variable.

that causes a disparate on the basis of race, color, religion, sex, or national origin."[66] Section (k)(1)(B)(i) of the act adumbrates further this causation element:

> With respect to demonstrating that a particular employment practice causes a disparate impact as described in subparagraph (A)(i), the complaining party shall demonstrate that each particular challenged employment practice causes a disparate impact, except that if the complaining party can demonstrate to the court that the elements of a respondent's decisionmaking process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice.[67]

Under this provision, a plaintiff must either show that particular practices cause a disparate impact or, if an analysis of individual processes is not possible, that the practices taken together have a disparate impact.[68] I will refer to these two different approaches as the "individual practice" and the "unitary practice" approach. The need for the unitary practice approach, where the defendant's decisionmaking process is "analyzed as one employment practice," can be seen, for example, in the context where an employer promotes employees on the basis of several "tightly woven and overlapping criteria."[69] If the promotion system combines objective criteria – such as attendance history and seniority – with subjective and arbitrary application of these objective criteria on the part of managers, it is not possible to separate the promotion criteria for review.[70] In such cases (particularly where there is subjective or discretionary decisionmaking), the statute allows an analysis testing for the disparate impact of the decisionmaking as "one employment practice."

---

[66] Civil Rights Act of 1991, 42 U.S.C. § 2000e-2 (k)(1)(A)(i) (2006).

[67] *Id.* at 42 U.S.C. § 2000e-2 (k)(1)(B)(i).

[68] *See* Meacham v. Knolls Atomic Power Lab, 185 F. Supp. 193, 207-08 (N.D.N.Y. 2002) (guidelines for determining which employees were exempt for an involuntary reduction in force consisted of four criteria, with several factors considered for each criterion; these elements not capable of separation for analysis); Graffam v. Scott Paper Co., 870 F. Supp. 389, 395 (D. Me. 1994) (assessment process analyzed as one employment practice); Stender v. Lucky Stores, 803 F. Supp. 259, 335 (N.D. Ca. 1992) (elements of company's "subjective and ambiguous decision making process" not separable for the purposes of analysis).

[69] *See* McClain v. Lufkin Indus., 519 F.3d 264, 277-78 (5th Cir. 2008).

[70] *Id.* at 278-279.

When there is not sufficient data to estimate the marginal impact of individual policies, the unjustified DI specification (equation (3)) can by itself provide evidence that an amalgam of discretionary practices caused an unjustified disparate impact. The inclusion of plausible business justification factors in equation (3) in fact suggests the alternative decisionmaking policies that could have been used without causing an unjustified disparate impact. Thus, for example, if the results of an equation (3) regression suggest that a lender was justified in charging borrowers with poorer credit scores higher interest rates, then the lender would not have an estimated unjustified disparate impact if it had implemented a policy of uniformly charging all borrowers with the poorer credit score the higher rates recommended by the regression. Or if an equation (3) regression indicated that a particular business-justified applicant attribute increased the probability of hiring, then a policy of enhancing the likelihood of hiring all applicants with this attribute could have been implemented without producing a statistically-significant disparate impact estimate. The estimated regression coefficients thus point to a set of alternative business-justified policies that a decisionmaker might have used which would not have produced evidence of an unjustified disparate impact.[71] It is relative to this alternative policy set that the defendant's actual policies caused an unjustified disparate impact. Thus, when there is not sufficient data to estimate the impact by caused individual practices, a statistically significant estimate of $\beta_1$ is by itself evidence that defendant's practices caused an unjustified racial disparity.

To establish that an "individual practice" caused a disparate impact, it is natural to compare the disparate impact estimate ($\beta_1$) in a regression that excluded a control for the practice

---

[71] This alternative set of policies would implement the decisions produced by the estimated regression equation – except ignoring any estimated $\beta_1$ race effect.

in question with a regression that included the control. For example, in our recurring stylized version of *Griggs*, a researcher would compare the coefficient value estimated for $\beta_1$ in a regression that first included a control for applicants' diploma status and then in a regression that properly excluded the applicants' diploma status variable. If the disparate impact estimate ($\beta_1$) becomes more negative when the policy control is excluded from the regression, that adverse movement indicates that the defendant's policy of hiring based on diploma status caused a disparate racial impact. When you exclude a causal variable from the specification, the regression in effect attempts to reattribute the causal influence of the excluded variable to other control variables that are correlated with the excluded variable. An adverse movement in the disparate impact estimate ($\beta_1$) when the diploma variable is excluded is evidence both that the diploma status is correlated with race and that the diploma hiring criterion is causing an adverse impact on minority relative to non-minority applicants.

As a conceptual matter, individual practice causation might be shown by starting with the DT specification (equation (1)) and then eliminating a single practice to assess the impact on the disparate impact estimate ($\beta_1$). In the more inclusive regression, a statistically significant coefficient on the diploma indicator variable would indicate that this attribute causally influenced the defendant's decisionmaking.[72] Additionally, a statistically significant change in the $\beta_1$ coefficient when the diploma variable was removed from the specification would indicate that

---

[72] An ancillary regression of the form:

$$\text{Minority} = \alpha + \beta_1 * \text{Unjustified Influence} + \varepsilon,$$

where "Unjustified Influence" represents the specific unjustified practice that is claimed to influence the defendant decision, would assess whether minorities were more likely to be subjected to the practice. For example, in the foregoing *Griggs* hypothetical, regressing the minority indicator variable on the high school diploma factor would have established that African-American applicants were less likely to have high school diplomas (and that this short fall was statistically significant).

the diploma policy caused a racial disparity.[73]

Indeed, even if both of the $\beta_1$ coefficients — that is, in both the more inclusive DT regression and the less inclusive unjustified DI regression — indicate that minority applicants were systematically favored by a defendant decisionmaker, a statistically significant adverse change in estimate when the unjustified policy variable was eliminated from the regression specification would indicate that the particular policy on the margin caused an unjustified disparate racial impact.[74] This is particularly relevant to *Connecticut v. Teal*,[75] which found that unjustified disparate impacts in any part of the decision-making process can be actionable even if the decision-making process overall does not produce unjustified racial disparities.[76] Under *Teal*, it is the adverse change in the racial disparity that is crucial and not the level of the change before or after excluding the controls for an unjustified policy.[77]

---

[73] While statistical packages by default report whether a coefficient in a regression is statistically different from 0, it is possible to test where a coefficient is statistical different from any constant. Thus, if the more inclusive regression estimated a $\beta_1$ coefficient of "x," it would be possible to test whether the estimated $\beta_1$ coefficient in less inclusive regression was statistically worse than "x."

[74] I made an analogous point in describing the possible results of a disparate impact analysis regarding the impact of antigen matching in allocating kidneys for transplantation:

> For example, in the foregoing hypothetical probit regressions estimating transplantation probabilities, one could imagine that the coefficient on the minority indicator variable was estimated to be positive in both the regressions including and excluding controls for partial antigen matching, thus indicating that minority applicants had a heightened probability of qualifying for transplantation. Nonetheless, if the regressions indicated that excluding the partial antigen control caused a statistically significant drop in the minority coefficient (but still left the coefficient positive), this would be evidence that the partial antigen matching preference had an unjustified disparate impact on minority applicants.

Ayres, *Three Tests supra* note 31 at S77.

[75] 457 U.S. 440, 445-51 (1982).

[76] I have been critical of the implications of *Teal*. *See* Ayres & Siegelman, *supra* note 45. But there are still signs of its vitality. *See, e.g.*, *Lewis v. City of Chicago*, 2010 U.S. LEXIS 4165 at *16 (May 24, 2010) (Scalia, J.) (citing to *Teal* in support of proposition that Chicago's "decision to adopt the cutoff score (and to create a list of the applicants above it) gave rise to a freestanding disparate impact claim"). *See also* Robert B. Fitzpatrick, *Disparate-Impact Claims Get a Boost in Unanimous Supreme Court Opinion Written by Justice Scalia,* FITZPATRICK ON EMPLOYMENT LAW (May 25, 2010), http://robertfitzpatrick.blogspot.com/2010/05/disparate-impact-claims-get-boost-in.html.

[77] This analysis takes plaintiffs' behavior and the behavior of the plaintiffs' non-minority counterparts as given. In some settings, the decisions and behaviors of plaintiffs may also be a but-for cause of an estimated disparate racial impact. Thus in a stylized *Griggs* example, the failure of plaintiffs to earn a diploma could just as much cause a disparate racial impact as an employer diploma requirement. The civil rights act of 1991 however on its face

v. *Qualified Pools*

The Civil Rights Act of 1991, in overturning the result of the Supreme Court's 1989 *Wards Cove* decision, made clear that after a plaintiff establishes a *prima facie* case of disparate impact the defendant has the burden of persuasion to prove the validity of an asserted business justification.[78] But an analogous question of justification has been imported into the adjudication of even the question of whether there is initial *prima facie* disparate impact as courts have limited their attention to racial disparities within "qualified" pools. For example, in *Wards Cove*, the Supreme Court found that "a comparison – between the racial composition of the qualified persons in the labor market and persons holding at-issue jobs – that generally forms the proper basis for the initial [*prima facie*] inquiry in a disparate-impact case.."[79] Thus, in the earlier airline example,[80] the need for pilots to have a license might be raised, on the back end, as an affirmative defense to rebut a plaintiff's *prima facie* case, or on the front end, as an effort to limit the *prima facie* analysis to a qualified pool of licensed applicants.

Determining whether a business justification issue will be adjudicated on the front end or the back end impacts both procedural burdens and the appropriate statistical approach. The

---

contains no contributory negligence defense. It would be perverse to require plaintiffs to acquire an unjustified attribute (such as a diploma in *Griggs)* or to thereby fail to establish causation. *See* Peter Siegelman, Contributory *Disparate Impacts in Employment Discrimination Law*, 49 WM. & MARY L. REV. 515 (2007). When the policies of both plaintiffs and defendants both cause (in a sense of formal logic) a disparate impact, the law tends to ask whether, holding precontractual attributes of other market actors constant, defendant's policies caused a disparate impact relative to other policies that they might have used.

[78] Wards Cove Packing v. Atonio, 490 U.S. 642 (1989) *overruled by the* Civil Rights Act of 1991, 42 U.S.C. § 2000e-2(k), *see* EI v. SEPTA, 479 F.3d 232, 241 (3d Cir. 2006) (recognizing [*Wards Cove*] as a departure from *Griggs*, Congress responded with the Civil Rights Act of 1991 (the "Act"), which placed back on the employer the burden of proof.)

[79] *Id. at* 650-51 (1989). *See also* New York City Transit Auth. v. Beazer, 440 U.S. 568, 585 (1979) (restricting analysis to "otherwise-qualified applicants"); and Ian Ayres, *Outcome Tests of Racial Disparities in Police Practices*, 4 JUSTICE RES. & POL'Y 131 (2002). Ayres and Siegelman argued that it was easier for plaintiffs to bring disparate impact firing cases than hiring cases because courts in firing cases would more likely accept that the class of existing workers as the qualified pool with regard to the firing decision. *See* Ayres & Siegelman, *supra* note 45.

[80] *See* text accompanying note 27..

plaintiff has the burden on the front end of proving a *prima facie* case which would include a showing that it had sufficiently limited its analysis to qualified persons. Take, for example, a disparate impact case against a law school employer. In order to show that a employer practice had a disparate racial impact, plaintiffs would be required to show as part of a *prima facie* case that the proportion of minorities hired was lower than the proportion of minorities within the set of people with the minimal teaching qualifications. In contrast, the defendant has the burden on the back end of showing that the employment policy that causes an adverse *prima facie* impact is nonetheless justified by promoting a legitimate organizational interest. Thus, an employer policy of being less willing to hire, say, aspiring political candidates might be justified (notwithstanding a disparate racial impact) if the employer could show on the back end that aspiring political candidates were more likely to be absent from teaching.

As a matter of statistical testing, "qualified pool" attributes should be used to restrict the dataset on which a regression runs, while "justification" attributes should be included as controls in a regression specification and not used to limit the number of observations. Thus, for example, the qualified pool of pilots only includes those people with a valid pilot's license, then the employer's treatment of applicants without a license should be eliminated from the analysis. However, if the employer is justified in giving a preference to military pilots (because such pilots tend to be more productive) then this attribute should be included as a control in the regression. At times, a defendant's own actions will constitute a qualified pool. For example, in promotion or firing cases alleging disparate impact, the pool of qualified employees is usually taken to be

the group of relevant employees that the employer hired in the past.[81] In Equal Credit Opportunity Act cases claiming disparate impact with regard to APR setting,[82] the defendant's own choice to extend credit to a class of customers naturally establishes a qualified pool of borrowers for purposes of APR analysis.[83] As in other areas of statistical analysis, it is necessary to determine when it is appropriate to control by excluding class of observations from a regression and when it is appropriate to control by adding right-hand-side variables. Included-variable bias can be created by inappropriately using either method of control. The larger thesis of this article is that in many circumstances neither type of control is appropriate. But when controlling is appropriate, the law's front end/back end approach to burdens can also guide a statistician on whether to control by excluding observations or by adding an additional control variable to the regression.[84]

vi. *Class Feasibility*

In disparate impact testing, the need to exclude (from regression equation (3)) all control variables that are not plausibly related to legitimate organizational goals makes disparate impact analysis particularly well suited for class-wide analysis. Under Rule 23, class certification is only

---

[81] Of course, not all existing employees are equally qualified for promotion or layoff. But for purposes of establishing a *prima facie* case, the existing employees are normally deemed to be minimally qualified. See Ayres & Siegelman, *supra* note 45, at 1489.

[82] The Equal Credit Opportunity Act, 15 U.S.C. § 1691 *et seq.* (2006); *see, e.g.*, Bayard v. Behlmann Auto. Servs., 292 F. Supp. 2d 1181 (E.D. Mo. 2003).

[83] If a suit alleged disparate impact in the class of customer who received loans, it would be necessary to look at the larger set of minimally qualified borrowers who applied or might have applied in the absence of defendant policies. *See* Jonah B. Gelback, Jonathan Klick and Lesley Wexler *Passive Discrimination: When Does It Make Sense to Pay Too Little?*, 76 U. CHICAGO L. REV. 797, 799-800 (2009) (race-neutral policies can disproportionate chilling effect on minority applications).

[84] As a general matter, econometric tests tend to eschew controlling by dropping variables. *See, e.g.*, DAMODAR N. GUJARATI, BASIC ECONOMETRICS 522-525 (McGraw-Hill, 3d ed. 1995). *See also* H.D. VINOD & AMAN ULLAH, RECENT ADVANCES IN REGRESSION MODELS 248 (Marcel Dekker, 1981): "When dealing with cross-section and time series data, where each individual cross-section sample is small so that sharp inferences about the coefficients are not possible, it is a common practice in applied work to pool all data together, and estimate a common regression. The basic motivation for pooling time series and cross-section data is that if the model is properly specified, pooling provides more efficient estimation, inference, and possibly prediction."

appropriate when there are "questions of law or fact common to the class"[85] and in some circumstances may only be appropriate "when questions of law or fact common to the members of the class predominate over any questions affecting only individual members."[86] Disparate treatment testing may force researchers to control for more idiosyncratic influences that provide non-race explanations for a decisionmaker's behavior. Information on these idiosyncratic variables may be difficult to obtain in a form that can readily be included in aggregate regression analysis.[87] As a result, resolving class claims of disparate treatment may depend on proof of individual acts of disparate treatment.[88]

But in contrast disparate impact litigation is virtually always based on aggregate statistical analysis that controls for a smaller universe of factors.[89] The substantive goal of avoiding "included variable bias" renders disparate impact claims more amenable to class-wide analysis. The *prima facie* issue of whether the defendant policies had an adverse impact on the plaintiffs' class presents a "common question of fact" that can be answered with a single aggregate estimation. And the second-order statistical question of whether the disparate impact persists after controlling for plausible business justifications likewise is a "common question of fact" which often can be analyzed with retained defendant data on the costs of doing business.[90]

---

[85] FED. R. CIV. P. 23(a)(2).

[86] FED. R. CIV. P. 23(b)(3).

[87] *See also* Brown, *et al*. v. Nucor Corp., No. 08-1247 (4th Cir. Aug. 7, 2009), *available at* http://pacer.ca4.uscourts.gov/opinion.pdf/081247.P.pdf (overturning denial of class certification by the District Court).

[88] *See* International Brotherhood of Teamsters v. United States, 431 U.S. 324, 338 (1977) ("The Government bolstered its statistical evidence with the testimony of individuals who recounted over 40 specific instances of discrimination.").

[89] Richard Nagareda, *Class Certification in the Age of Aggregate Proof*, 84 N.Y.U. L. REV. 97 (2009); James T. Tsai, *23(B)(2) Class Certification: Choosing an Approach for Certifying Civil Rights Discrimination Class Action Suits* (Berkeley Electronic Press, Working Paper No. 1984, 2007), *available at* http://law.bepress.com/expresso/eps/1984.

[90] For example, as discussed previously, *supra* at text accompanying note 83, modern lenders routinely retain underwriting information on borrowers (including, e.g., credit scores, loan-to-value ratios) to assess the cost of

## II. Outcome Tests As Measures of Disparate Impact And The Recurring Problem of Included Variable Bias

Up until now, the regression specifications have attempted to model defendant decisionmaking — the left-hand side of the regression equations have been dichotomous (hire/don't hire, lend/don't lend) or continuous (APR, wage) decision variables. But Gary Becker suggested a very different approach to test for discrimination by statistically modeling, not defendant decisions, but the economic outcomes of those decisions. Becker, first in a *Business Week* article and later in his Nobel Prize lecture,[91] suggested that lending discrimination could be inferred by analyzing the profitability of loans made to different races: "If banks discriminate against minority applicants, they should earn greater profits on the loans actually made to them than on those to whites."[92] The idea behind Becker's outcome test is that if banks are only willing to lend to relatively overqualified minority borrowers, then these minority borrowers should default less and impose fewer costs on the banks — thus producing higher profits. A similar argument could be applied to college admissions. If a college is only willing to accept overqualified female applicants (relative to male applicants), then one would expect female graduates of the college to outperform male graduates on blind examinations.

Statistically, an outcome test can be implemented with the following regression equation:

$$\text{Defendant Success} = \alpha + \beta_1 * \text{Minority} + \varepsilon. \qquad (4)$$

---

lending to particular borrower types. It is inappropriate to include controls for borrower sophistication and other factors that relate solely to a lender's ability to borrower weakness.

[91] Gary S. Becker, *The Evidence Against Banks Doesn't Prove Bias*, Business Week, Apr. 19, 1993, *available at* http://www.businessweek.com/archives/1993/b331513.arc.htm; Gary S. Becker, *Nobel Lecture: The Economic Way of Looking at Behavior*, J. Pol. Econ., June 1993, at 385, 389.

[92] *Id.*

where the "Defendant Success" variable measures the outcome or success of a defendant decision. In some circumstances, the success variable will be dichotomous (for example, whether a police search uncovered contraband). In other circumstances, the success variable will be continuous (for example, the profitability to a lender from a loan). Being able to quantify success in terms of a defendant's own preferences is a sine qua non of being able to run an outcome regression.[93]

Becker extolled the outcome testing as a "direct" approach to measuring discrimination. But Becker failed to distinguish between disparate outcomes that are caused by disparate racial treatment – that is race-contingent decisionmaking – and disparate outcomes that are caused by unjustified disparate impacts. An outcome test is at most an alternative way of testing for unjustified disparate impacts and not a "direct" test of disparate treatment.[94] A bank that gives a preference to college graduates because of a mistaken belief that such borrowers are lower risk or because of a lender's bias in favor of college borrowers may induce an outcome test to indicate racial bias if minority applicants are less likely than non-minority applicants to be college graduates. Favoring a category that is correlated with non-minority status, but not with

---

[93] My employer, Yale Law School, would have a difficult time finding a metric for quantifying the professional success of graduates.

[94] As measure of disparate treatment, an outcome test is both over and under inclusive. It is over-inclusive because outcome disparities might capture instances of unjustified disparate impact. It might be under-inclusive because it might not capture instances where decisionmaker was engaging in statistical discrimination, that is, race contingent decisionmaking based on statistically valid inferences. *See* Hylton and Rougeau, *supra* note 16 (discussing the difficulty of statistically testing for disparate treatment caused by statistically valid racial inferences); and *see infra* at text accompanying note 110 (discussing possibility that racial disparities in success rates may be induced by a decision maker's unwillingness to engage in statistical discrimination).

the decisionmaker's legitimate goal, will tend to induce racial disparity in outcomes — even when there is no disparate treatment.[95]

The right-hand side of equation (4) is identical to the right-hand side of equation (2) in that both specifications exclude all non-race controls. But while equation (2) merely implements a *prima-facie* test for either justified or unjustified disparate impacts, the outcome specification in equation (4) can provide evidence of whether defendant's decisionmaking produced unjustified disparate impacts. A *prima-facie* test of disparate impact in interest rate setting (based on equation (2)) might indicate that minority borrowers on average were charged higher APRs than white borrowers. Yet without adding controls for credit scores and other for plausible business justifications, this crude test could not tell us whether the racial disparity was justified. It would not be able to distinguish between justified and unjustified racial disparities. In contrast, the outcome test which uncovers racial disparities in the *success* of defendant decisionmaking provides evidence of unjustified racial disparities—even though it fails to control for plausible justification factors. The central idea of Becker is that a maximizing decisionmaker should have already taken into account all justified factors that systematically influence success, with the result that an analysis of success itself should already control for those factors. For example, if a judge's goal is to assure that all criminal defendants will appear at trial with some minimum expected probability, then bail should be raised to the level that (conditional on all the factors) induces this minimum probability of appearance.[96] The

---

[95] Becker's outcome test is also not well suited to capture instances of statistical discrimination – that is disparate treatment which is motivated by a statistically valid difference between minority and non-minority customers or employees.

[96] Ayres and Waldfogel, *supra* note 48, used the ex ante prices that bond dealers charged as a proxy for the ex post probability of success. An analysis of bond dealer pricing as a proxy for success similarly need not control for plausible justifications:

researcher need not observe and control for in specification (4) all the justified factors that might plausibly impact the decisionmaker's own criterion for success under the maintained assumption that the decisionmaker was already taking these factors into account. A racial difference in the average success rates can provide evidence that the decisionmaker must have been taking into account factors unrelated to success (or failing to take into account factors related to success) that are correlated with the plaintiffs' race.[97]

The problem of included-variable bias is even more pronounced in outcome testing. In the earlier unjustified DI specification (equation (3)), where the defendant's decision was the left-hand side variables, it was only necessary to exclude controls that did not represent plausible decision justifications. But in an outcome test (equation (4)), it is necessary to avoid included-variable bias to exclude even controls that represent a plausible decision justification. To see why this is so, consider again an outcome test of judicial bail setting. Imagine that it is plausible for judge to consider defendants a greater flight risk if in the past they have "failed to appear" when released on bail. Judges would be plausibly justified in setting higher bail amounts for defendants with this characteristic. Nonetheless, this characteristic should *not* be included as a control in an outcome regression. If judges are setting bail so as to assure the same minimum probability of appearance, then we should expect that this control variable would have no

---

> [O]ur analysis avoids the problem of omitted variable bias. Bond rates provide a market-disciplined assessment of a defendant's probability of flight, given her bail amount. As a result, bond rates obviate the need to observe and measure defendant characteristics which, in traditional discrimination studies, serve as indirect proxies for the defendant's flight probability. Knowledge about bond prices substitutes for the traditional requirement that the researcher control for everything that might have affected courts' decisions.

*Id.* at 993.

[97] As a theoretical matter, either an outcome test for unjustified disparate impacts (equation 4) or the earlier decision test for unjustified disparate impacts (equation 3) can uncover evidence of statistically significant racial disparities adverse to minorities—even though a *prima facie* test (equation 2) indicated no disparity or a statistically significant disparity favoring minorities. Thus, even if the uncontrolled equation 2 analysis shows that minorities were more likely to receive a favorable decision (being hired, receiving a loan, avoid search) than non-minorities, an analysis of either equations 3 or 4 might indicate that a consideration of plausible controls suggests that, after controlling for the decisionmaker's own objectives, they should have been favored even more.

statistically significant impact on the probability that defendants appear. Optimizing judges would already have taken this risk factor into account in setting bail — so that defendants with this factor (and the higher bail amounts) should have the same probability of appearance as defendants without this factor (and lower bail amounts). In econometric terms, we should expect that the estimated coefficient on the "failure to appear" (FTA indicator) variable would not be statistically different than zero.

A regression that improperly included this factor runs the risk of inducing included-variable bias. A regression finding that bailees with this factor are more likely to appear for trial suggests that judges are not setting bail to achieve uniform minimum probability of appearance. Even though judges are justified in setting higher bail amounts for bailees who in the past had failed to appear, such a regression finding would suggest that the judges were overdeterring bailees with this FTA factor relative to bailees without this FTA factor. Including an FTA control would induce included-variable bias if minorities were disproportionately likely to have failed to appear in the past. The unjustified overdeterrence of bailees with the FTA characteristic would work an unjustified disparate impact on minority bailees. But the outcome regression would fail to capture the true unjustified disparate impact in a misspecified regression, because some of the unjustified racial disparity would be improperly attributed to the FTA coefficient. So with regard to output tests, it is not just that researchers do not have to be worried about excluded variable bias, they need to worry that erroneously including even plausible business justified variables will bias their estimate of any unjustified disparate impact.

In outcome regressions, the only appropriate controls to add to specification (4) should concern factors that legitimately alter the decisionmaker's goal. For example, in the preceding bail example, one could imagine that it is justified for a judge to demand a higher probability of

appearance for more serious crimes — where the social costs of a defendant's failure to appear might be expected to be larger. Thus, when Joel Waldfogel and I conducted an outcome test of judicial bail setting in Connecticut state courts, we included controls for offense severity and certain offense categories.[98] We found that judges effectively demanded higher probabilities of appearance for series felonies (relative to misdemeanors) and for drug and gun charges (relative to other charges).[99] Yet as before,[100] the choice to include or exclude does not have to be all or nothing. While a judge might be justified to demand a higher probability of appearance on drug possessions, there might some point at which elevating the probability of appearance would not be justified. It would be possible to include controls for such "goal shifters" but to cap the coefficients so that the controls are not allowed to improperly soak up more than the justified disparate racial impact.

The outcome test specification has theoretical similarities to randomized experiment testing. In a randomized experiment, the test of whether a "treatment" randomly assigned to a portion of a sample produced different outcome than the outcomes for those assigned to a "control" group uses a similar stripped-down specification like equation (4), where a "treatment" indicator substitutes for a "race" indicator.[101] In a randomized experiment, it is the process of randomization that obviates the need of a researcher to control for other potential outcome

---

[98] Ayres &Waldfogel, *supra* note 48 at 1010 tbl.3.

[99] *Id.*

[100] *See supra* text accompanying note 65 (discussing capped coefficient specifications).

[101] *See, e.g.*, Ian Ayres, Sophie Raseman, & Alice Shih, *Evidence From Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage* (July 16, 2009), *available at* http://ssrn.com/abstract=1434950. *See also* James Heckman, *Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy*, 48 J. ECON. LITERATURE, no. 2, 2010, at 356.

influences. In an outcome experiment, it is the process of decisionmaker maximization that obviates the need of a researcher to control for other potential outcome influences.[102]

To be valid indicators of unjustified disparate impacts, outcome tests need to overcome two difficulties, which I will refer to as the "infra-marginal problem" and the "subgroup validity problem." The possibility of these problems suggest that statistically significant racial disparities uncovered in an outcome test should not be taken as conclusive evidence of unjustified disparate impacts. Such evidence should, I propose, shift the burden to defendants to explain why their decisionmaking is systematically less successful when applied to the plaintiff's class.

*A. Infra-Marginal Problem*

The outcome test has been criticized as a test of discrimination because researchers at times are only able to measure the average outcome and not the outcomes associated with the marginal decision.[103] For example, in the mortgage context, a test of disparate treatment would want to assess whether the least qualified whites to whom banks were willing to lend had a higher default rate than the least qualified minorities to whom banks were willing to lend. If lenders dislike lending to minorities, then the least qualified minority to whom they would be willing to lend (the marginal minority borrower) should have a lower expected default rate than the least qualified nonminority to whom they are willing to lend (the marginal nonminority

---

[102] In both stripped down specifications, the researcher should be unconcerned about the degree of goodness of fit measured for example by R-squared—since the purpose of neither the outcome test nor the randomized test is to explain all the variation in outcomes. But an important difference between the two specifications concerns the causal interpretation of the results. In a randomized test, a statistical significant coefficient on the treatment indicator is an indication that the treatment relative to the control caused a different outcome. Whereas a statistically significant coefficient on a race indicator in an outcome test is at most an indicator that the decisionmaker relied on unjustified factors that were correlated with the race indicator.

[103] James H. Carr & Isaac F. Megbolugbe, *The Federal Reserve Bank Of Boston Study On Mortgage Lending Revisited*, 4 J. HOUSING RES., no. 2, 1993 at 277, 309; *see also* George C. Galster, *The Facts of Lending Discrimination Cannot Be Argued Away By Examining Default Rates*, 4 HOUSING POL'Y DEBATE, no. 2, 1993, at 141. This subsection is generalizes an argument regarding police practices made first in Ayres, *Outcome Tests, supra* note 79.

borrower). Unfortunately marginal default rates are unobservable and researchers are often only able to estimate the average default rates conditional on being above this marginal lending threshold. Lenders might still discriminate against minority borrowers—in the straightforward sense that the lending threshold for minorities might be more stringent than for nonminorities—but we might still see that the average rate of minority default is higher than the average rate of nonminority default. As long as infra-marginal nonminority borrowers have lower expected default rates (than infra-marginal minority borrowers), a comparison of average defaults may mask disparate treatment by lenders in setting the minimum thresholds for granting loans. For example, Figures 2 depicts a scenario in which a lender engages in disparate treatment — requiring a higher credit score from minority borrowers than non-minority borrowers, but in which the average default rate of non-minority borrowers is nonetheless likely to be lower because infra-marginal nonminority borrowers have lower expected default rates.

**Figure 2: Example in which Minority Borrowers have lower marginal but higher average default rates than Non-Minority Borrowers**


The figure shows that even though the lender engages in express disparate treatment in requiring higher credit scores from minority applicants before lending, the average default rate of minority borrowers is likely to be higher than the average default rate of non-minority (because

non-minority borrowers are disproportionately infra-marginal, with credit scores well above both the minimum lending thresholds). The figure shows that the outcome test as a measure of disparate treatment can be underinclusive. But more generally, the infra-marginality problem can render outcome tests under or over-inclusive as tests of disparate treatment.

One response to the infra-marginality problem is to focus on outcome tests where researchers are better able to directly estimate the marginal impacts of the decision. For example, in the bail bond setting, judges can set bail along a spectrum to assure a minimum probability of appearance for each defendant. The continuous nature of the decision places each defendant on the margin — so that average racial disparity should not include infra marginal effects.[104]

---

[104] The ability to infer marginal effects from average effects is dependent on the absence of "selection effects" (i.e., that only some defendants accept judicial bail offers) and the ability of higher bails to deter flight. *See* Ayres & Waldfogel, *supra* note 48 at 1032-1035 (discussing both of these issues). Figure 1 depicts a typology of outcome tests distinguished by whether the decision or the outcome is a dichotomous or continuous choice:

Table 1: Typology of Outcome Tests
with characteristic examples of decisions (and outcomes in parentheses)

| | | Decision | |
|---|---|---|---|
| | | Dichotomous | Continuous |
| Success | Dichotomous | Police Search (Discover Contraband) | Judicial Bail (Appear at Trial) |
| | Continuous | Journal Acceptance (Citations) | Analyst Earnings Prediction (Actual Future Earnings) |

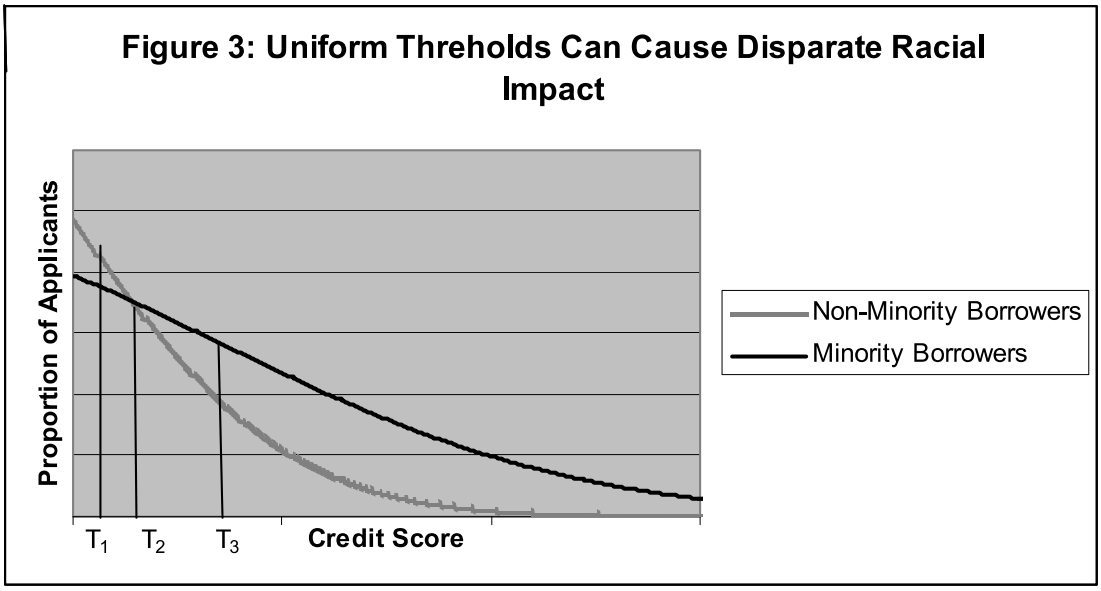
But a more fundamental response to the infra-marginal concerns again focuses on the difference between disparate treatment and disparate impact testing. As already discussed, outcome tests are better suited as tests of disparate impacts than tests of disparate treatment. A showing that minorities have higher success rates than non-minorities does not mean that a decisionmaker is discriminating on the basis of race — but it does suggest that decisionmaker is relying on a criterion that is correlated with race but does not in fact predict success. Notwithstanding the possibility of infra-marginal differences between the average and the marginal racial disparities, outcome tests still provide evidence that the decisionmaker policies caused disparate impacts. For example, imagine that equation (4) shows that the average white loan default rate was 15%, while the average minority default rate was 10% of the time. The lender police could raise infra-marginality as a defense to the claim that this finding proves disparate treatment: for example, they might argue that they lend to all applicants with less than a 20% default probability (and of this group, it just so happens that 15% of whites default while only 10% of minorities do). In essence, the lender would be arguing that they apply a uniform (20%) threshold to all applications regardless of race—so that at the margin there is no disparate treatment.

But this infra-marginal argument would not be a defense to the claim that the lender's choice of a 20% threshold itself produced a disparate impact on minorities. Raising or lowering the decision threshold will frequently mitigate the size of average racial disparity. For example,

---

Examples of the four types can be found in Ayres & Borowsky, *supra* note 34 (police searches); Ian Ayres & Fredrick E. Vars, *Determinants of Citations to Articles in Elite Law Review*, 29 J. LEGAL STUD. 427 (2000). (law review citations); Ayres & Waldfogel, *supra* note 48 (bail); Alok Kumar & Justin Wolfers, *Under-Estimating Female CEOs*, ALEA 2008, NBER Law and Economics Summer Institute 2008, 2008 Conference on Empirical Legal Studies (analyst earnings predictions)' Knowles, Persico and Todd suggest that the average success rate of searches for different races will also tend toward equality because of the strategic reaction of the individuals subject to searches. *See also* John Knowles, Nicola Persico, & Petra Todd, *Racial Bias in Motor Vehicle Searches: Theory and Evidence*, 109 J. POL. ECON. 203, 203-29 (2001); Ayres, *Outcome Tests*, *supra* note 79 at 131.

Figure 3 depicts the racial disparities that would be created by three different lending thresholds ($T_1$, $T_2$ and $T_3$):




All three lending thresholds likely to produce racial outcome disparities with minority borrowers exceeding the threshold having higher average credit scores than whites exceeding a given threshold. But thresholds $T_1$ and $T_3$ are likely to produce greater outcome disparities than $T_2$.[105]

While a finding of a racial disparity in the average success rates would still often be evidence of a disparate impact, this finding might—taking into account the infra-marginality—no longer imply evidence of an *unjustified* disparate impact. In the previous lending hypothetical, one would need a separate analysis to assess whether a particular threshold was business justified. The decisionmaker could still argue that the uniform credit score threshold produced a justified disparate impact relative to other credit score thresholds. When a decisionmaker is making dichotomous decisions that create the possibility of infra-marginal

[105] Thresholds 3 at the margin disproportionately excludes minority borrowers — so that a lowering of the minimum credit score would disproportionately include minorities with relatively poor credit scores, thus tending to mitigate the expected default disparity. An analogous argument explains why increasing the threshold requirement from $T_1$ would be likely to disproportionately exclude non-minority borrowers with relatively poor credit scores.

reversals, the ultimate question of whether the threshold criteria are justified or not will turn on external evidence concerning the organizational impact of potentially less restrictive alternatives. It might be appropriate for courts to resolve such issues by shifting the burden to the decisionmaker to offer plausible evidence why the particular threshold which produces a disparate impact is nonetheless "consistent with business necessity."

*B. Subgroup-Validity Problem*

A second limitation on the use of outcome tests concerns what I term the subgroup-validity problem.[106] Put simply, when a particular characteristic is valid for some races but not for others, it is possible that a decisionmaker conditioning her decisions on this characteristic generally might induce racially disparate outcomes. The non-race characteristic may be a valid predictor of success for the population as a whole, but relying on this characteristic may nonetheless produce racially disparate outcomes if the characteristic is not a valid predictor of success for a minority subgroup. To put the matter more provocatively, when a particular non-race characteristic is only a valid decisionmaking criterion for some races, then a decisionmaker's unwillingness to engage in disparate racial treatment (only deploying the criterion for those races) may induce racial disparities in outcomes.

To see the possibility of perverse results from outcome testing when a characteristic of generally valid, but not valid for a minority subgroup, consider the following "baseball cap" hypothetical:

> [I]magine that wearing a particular type of baseball cap is strong evidence of drug possession when done by whites but not when done by minorities. In the extreme, imagine that 100% of whites wearing this cap possess drugs, and 0% of minorities wearing this cap possess drugs. And finally imagine that if the police stopped all people

---

[106] *See* Ayres, *Outcome Tests, supra* note 79.

> wearing such a baseball cap, that 75% of those stopped would be white (possessing illicit drugs) and 25% would be minorities (not possessing illicit drugs).[107]

These stylized statistics suggest that the baseball cap is a valid indicator of illicit activity for whites but it is not valid for the minority subgroup. Moreover, the cap is a fairly valid predictor for the overall population—since there is a 75% chance that a cap search will uncover illicit drugs.

A police department that chose to adopt a policy of stopping and searching all those who wear the cap (minorities and nonminorities alike) would produce disparate racial outcomes. In this extreme hypothetical, the success rate for minorities searched would be 0% and the success rate for whites searched would be 100%. The racial disparity in outcomes is caused because the cap search criterion (i) is correlated with minority status and (ii) is not correlated with drug contraband for the minority subgroup.

The cap hypothetical has two implications for civil rights testing. First, it shows that a decisionmaker's unwillingness to engage in "statistical discrimination" can produce outcome disparities.[108] When race (possibly in combination with other factors) is a valid predictor of success, then failing to condition decisions on race can produce evidence of disparate outcomes. In this hypothetical the systematically lower minority search success rate is caused by the police department's unwillingness to engage in disparate racial treatment—its unwillingness to engage in racial profiling. By the internal logic of the outcome testing — which assesses defendant behavior by the sole criterion of whether it produces equivalent success rates, the decisionmaker

---

[107] This example is adapted from Ayres, *Three Tests*, *supra* note 31 at S82-S83.

[108] Analytically, this baseball cap hypothetical is somewhat the opposite of the facts in *Ricci v. DeStefano*, 557 U.S. ___; 129 S.Ct. 2658 (2009), where the city of New Haven engaged in disparate treatment in order to avoid imposing what it believed was an unjustified disparate impact on African-American applicants.

policy of *not* engaging in disparate racial treatment produces evidence of an unjustified disparate racial impact. When statistical discrimination is (statistically) valid, the failure to use it will produce outcome disparities. The cap hypothetical shows the limited normative scope of outcome evidence because the law might reject disparate treatment even if it is statistically valid.[109]

Second, the hypothetical shows that outcome tests are limited to testing whether decisionmaking criteria are less effective as applied to particular subgroups. Especially if we rule out the possibility of decision making policies — like racial profiling — which are explicitly race contingent, then outcome tests at most are evidence that defendant decisionmaking policies were unjustified as applied to particular subgroups.[110] But the hypothetical suggests that criteria that are invalid with regard to a subgroup may nonetheless be valid as applied to the larger group of all races.[111] Thus, the subgroup validity problem means that a finding of racial disparate

---

[109] *See* Becker, *Evidence Against Banks*, *supra* note 91. *See also* Stewart Schwab, *Is Statistical Discrimination Efficient?*, 76 AM. ECON. REV., no. 1, 1986, at 228. See also R. Richard Banks, Race-Based Suspect Selection and Colorblind Equal Protection Doctrine and Discourse, 48 UCLA Law Review 1075-1124 (2001).

[110] Ordinarily courts refuse in disparate impact cases to consider whether race-contingent policies could mitigate the disparate impact created by some preexisting practice. But the Supreme Court in Ricci v. DeStefano, 557 U.S. ___; 129 S.Ct. 2658 (2009), held that disparate treatment can be justified by a strong basis in evidence of an impermissible disparate impact. As Christine Jolls has recently noted (see Christine M. Jolls, *Accomodation Mandates and Antidiscrimination Law*, 53 STAN. L. REV. 223 (2000)), a disparate racial impact decision invalidating an employer's "no beard" policy (as having an unjustified disparate impact on African Americans) has expressly endorsed race-contingent remedies. *See* Bradley v. Pizzaco of Nebraska, 7 F.3d 795, 799 (8th Cir. 1993) ("injunction shall be carefully tailored to place Domino's under the minimal burden of recognizing a limited exception to its no-beard policy for African American males who suffer from PFB and as a result of this medical condition are unable to shave"). *See also* Cason v. Nissan Motor Acceptance Corporation, Settlement Agreement, Class Action No. 3-98-0223 (M.D. Tenn. Feb. 18, 2003). Such decisions suggest that decisionmakers may have a duty to remedy racial disparate impacts by resorting to express racial disparate treatment. However, such a duty may run afoul of the 1991 Civil Rights Act's ban on race-norming. *See* 42 U.S.C. § 2000e-2(l) (Supp. IV 1992).

[111] The 1966 and 1970 EEOC guidelines required evidence of subgroup racial validity (so called "differential validation") requiring employers to conduct separate validation studies for different racial groups. *See* United States v. City of Chicago, 549 F.2d 415, 433 (7th Cir. 1997) (requiring differential validation). The Supreme Court even endorsed it in *Albermarle Paper Co. v. Moody*, 422 U.S. 405, 435 (1975). But the Uniform Guidelines eliminated the requirement for differential validation and replaced it with something called "unfairness studies." *See* 29 C.F.R § 1607.14B(8). Subgroup validation is still required with language; *see* Mark Kelman, *Concepts of Discrimination in 'General Ability' Job Testing*, 104 HARV, L. REV. 1157 (1991).

outcomes may not indicate that the decisionmaking was unjustified with regard to the decisionmaker's legitimate organizational interests. For this reason, I will refer to the outcome test using equation (4) as the *subgroup disparate impact* test. As before, the subgroup validity problem is best handled by burden shifting. When a plaintiff shows systematic racial difference in the success of defendant decisionmaking that burdens the plaintiff class, the defendant should be given the opportunity to explain why a generally valid criterion is not accurately predicting with regard to this particular subgroup.

For example, Alok Kumar and Justin Wolfers have analyzed more than 93,000 quarterly earning announcements to assess whether stock analysts are equally successful at predicting corporate earnings when the analyzed firms have male and female CEOs.[112] Their statistical analysis is a version of the equation (4) specification in that they control for no other variables and just see whether quarterly earnings predictions are equally accurate when the analyzed firm has a female as opposed to a male CEO. They find that analysts are 3.7% more likely to underpredict earnings when a CEO is female than when a CEO is male, and analysts are 4.6% more likely to overpredict earnings when a CEO is male than when a CEO is female.[113] To put it simply, the analysts' forecasts display systematic bias in their earning predictions against female CEOs relative to male CEOs. This gender disparity in forecast success rates suggests that analyst decisionmaking has a disparate impact adverse to women CEOs. The results are not qualified by infra-marginal concerns, because, as argued above, the continuous nature of earnings estimates make each forecast marginal. However, there is a possibility of a subgroup

---

[112] Kumar & Wolfers, *supra* note 104.

[113] Specifically, they find that "positive earning surprises" – where actual earnings exceed forecasts — 62.5% and 58.7% of the quarters for female and male CEOs respectively; and "negative earning surprises" — where actual earnings are below forecasts — 22.5% and 27.1% of the quarter for female and male CEOs respectively. *Id.*

validity problem. Analysts might base forecast decisions on criteria that are (i) predictive for corporations generally, but (ii) not predictive for firms led by female CEOs. While Kumar and Wolfers have strong evidence that analyst forecasts have an adverse impact on the companies led by female CEOs, there is still the possibility that it is a justified disparate impact. If the question of justification somehow arose in litigation, my approach would be to allow analyst defendants to bring forward evidence showing that traded firms led by women CEOs were affected by a criterion that while generally valid was not valid as applied to them.

## Conclusion

The word "discrimination" has two distinct legal meanings related to disparate treatment and disparate impact claims. It is important that statistical evidence reflect the different elements in these two distinct claims. This article has shown that imprecise thinking about disparate treatment and disparate impact has led a host of analysts to draw erroneous conclusions. Nobel Prize economist, Gary Becker, describes an outcome test as a "direct" test of discrimination when it is at most evidence of subgroup disparate impact.[114] Statistical regressions commissioned by the LAPD purport to show no consistent evidence of racially biased policing — but treat as justified the elevated propensity of officers with numerous civilian complaints to search and frisk stopped citizens in minority neighborhoods. [115] Mortgage disparity regressions inappropriately control for lenders' attempts to prey on less sophisticated borrowers. [116] In all these cases, attention to the risk of "included variable bias" could lead to improved statistical estimation and inference.

---

[114] *See supra* text accompanying note 91.
[115] *See supra* text accompanying note 33.
[116] *See supra* text accompanying note 59.

It is time that courts and policy analysts move beyond the disparate treatment regression model when trying to assess and estimate whether a decisionmaker's policy produced unjustified disparate impacts. This article has tried to offer a reasoned blueprint for crafting regression specifications to provide evidence that fits the precise elements of the case. In disparate impact litigation (and other policy contexts where assessing whether policies produce unjustified racial disparities), inappropriately including variables that do not offer plausible justifications for defendant decisionmaking can bias the statistical estimates of racial bias. The possibility that a particular control variable in fact influenced a decision or that the variable is shown in regression analysis to be statistically significant does not, by itself, justify the variables inclusion in a regression testing for unjustified disparate impacts. Analysts ignore the problem of "included-variable bias" at their peril.