ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Neumark, David

Working Paper

Detecting discrimination in Audit and correspondence studies

IZA Discussion Papers, No. 5263

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Neumark, David (2010) : Detecting discrimination in Audit and correspondence studies, IZA Discussion Papers, No. 5263, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
https://hdl.handle.net/10419/46132

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DISCUSSION PAPER SERIES

IZA DP No. 5263

Detecting Discrimination in Audit and Correspondence Studies

David Neumark

October 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Detecting Discrimination in Audit and Correspondence Studies

David Neumark
University of California, Irvine,
NBER and IZA

Discussion Paper No. 5263
October 2010

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA Discussion Paper No. 5263
October 2010

ABSTRACT

Detecting Discrimination in Audit and Correspondence Studies*

Audit studies testing for discrimination have been criticized because applicants from different groups may not appear identical to employers. Correspondence studies address this criticism by using fictitious paper applicants whose qualifications can be made identical across groups. However, Heckman and Siegelman (1993) show that group differences in the variance of unobservable determinants of productivity can still generate spurious evidence of discrimination in either direction. This paper shows how to recover an unbiased estimate of discrimination when the correspondence study includes variation in applicant characteristics that affect hiring. The method is applied to actual data and assessed using Monte Carlo methods.

JEL Classification: J7

Keywords: discrimination, audit study, correspondence study

Corresponding author:

David Neumark
Department of Economics
UCI
3151 Social Science Plaza
Irvine, CA 92697
USA
E-mail: dneumark@uci.edu

* I am grateful to the UCI Academic Senate Council on Research, Computing, and Libraries for research support, to Scott Barkowski, Marianne Bitler, Thomas Cornelißen, Judith Hellerstein, James Heckman, and seminar participants at Hebrew University, Tel Aviv University, the Federal Reserve Bank of San Francisco, and the All-California Labor Conference for helpful comments, and to Scott Barkowski, Andrew Chang, Jennifer Graves, and Smith Williams for research assistance. I also thank Marianne Bertrand and Sendhil Mullainathan for supplying their data.

I. Introduction

In audit or correspondence studies, fictitious individuals who are identical except for race, sex, or ethnicity apply for jobs. Evidence of group differences in outcomes – for example, blacks getting fewer job offers than whites – is generally viewed as compelling evidence of discrimination. Across a wide array of countries and demographic groups, audit or correspondence studies conclude that there is evidence of discrimination, including, for example, discrimination against blacks, Hispanics, and women in the United States (Mincy, 1993; Neumark, 1996; Bertrand and Mullainathan [BM], 2004), Moroccans in Belgium and the Netherlands (Smeeters and Nayer, 1998; Bovenkerk et al., 1995), and lower castes in India (Banerjee et al., 2008). These "field experiments" for testing for discrimination are widely viewed as providing the most convincing evidence on discrimination (Pager, 2007; Riach and Rich, 2002), and U.S. courts allow organizations that conduct audit or correspondence studies to file claims of discrimination based on the evidence they collect (U.S. Equal Employment Opportunity Commission, 1996).[1]

Nonetheless, audit or correspondence studies have been subjected to scrutiny and criticism, most notably from Heckman and Siegelman [HS] (HS, 1993; Heckman, 1998). Most important, audit studies – which use "live" job applicants – have been criticized for failing to ensure that applicants from different groups appear identical to employers. Many of these criticisms can be countered by using correspondence studies, which use fictitious applicants on paper, or more recently the internet, whose qualifications can be made identical across groups. However, Heckman and Siegelman (1993) show that even in the ideal correspondence study in which both observed and unobserved group averages are identical, group

[1] EEOC (1996) discusses the case history regarding the use of testers in labor market and housing discrimination, as well as the Commission's conclusions. The EEOC refers to case law from the early civil rights movement when groups' only motivation was to test the law. For example, the EEOC cites *Pierson v. Ray*, 386 U.S. 547 (1967), in which the Supreme Court "held that a group of Black clergymen who were removed from a segregated bus terminal in Jackson, Mississippi had standing to seek redress," ruling that the plaintiffs "had been discriminated against by being ejected from the terminal, despite the fact that the plaintiffs' sole purpose was to test the law rather than to actually use the terminal" (U.S. EEOC, 1996, p. 2). The EEOC cites similar case law regarding plaintiffs testing the legality of segregation laws on buses. The document discusses the well-established standing of testers in housing discrimination cases under Title VIII, based on *Havens Realty Corp. v. Coleman,* 455, U.S. 363, 373 (1982), the parallels to employment discrimination under Title VII, and cases in which damages have been awarded in employment testing. Finally, it discusses *Fair Employment Council of Greater Washington, DC v. BMC Marketing Corporation*, 28 F 3d 1268 (D.C. Circuit 1994), in which standing was limited because individual testers could not show future harm, as well as the implications of Civil Rights Act of 1991, which allows for damages; according to the EEOC, the 1991 act would have given the testers in this case standing had the testing occurred after its passage (p. 4).

differences in the variances of unobservable determinants of productivity can lead to substantial spurious evidence of discrimination or of reverse discrimination; in other words, discrimination is unidentified. This critique is inherently difficult to address because it emphasizes factors outside the control of the researcher. This critique has been ignored in the literature, and researchers have continued to carry out additional field experiments of discrimination without regard to the strong possibility of highly misleading results. In fact, however, the HS critique poses a significant challenge to the audit/correspondence study method.

This paper explicitly addresses the HS "unobservable variance" critique. In particular, it develops and implements a general method of using data from correspondence studies that accomplishes two goals. First, it provides a statistical test of whether this HS critique applies to the data from any particular study. Second, and more important, it develops a statistical estimation procedure that identifies the effect of discrimination.[2] It is a simple matter to collect the requisite data in future correspondence studies, and the method is implemented using data from a correspondence study (BM, 2004) that has the requisite data. Finally, the estimation procedure is assessed via Monte Carlo simulations.

II. Background on Audit and Correspondence Studies

Earlier research on labor market discrimination focused on individual-level employment or earnings regressions, with the estimate of discrimination inferred from the race, sex, or ethnic differential that remains unexplained after including a wide array of proxies for productivity. These analyses suffer from the obvious criticism that the proxies do not adequately capture group differences in productivity, in which case the "unexplained" differences cannot be interpreted as detecting discrimination.

Audit or correspondence studies are a response to this inherent weakness of the regression approach to discrimination.[3] These studies are based on comparisons of outcomes (usually job interviews or job offers) for matched job applicants differing by race, sex, etc. (see, e.g., Turner, et al., 1991; Neumark, 1996; BM, 2004). Audit or correspondence studies directly address the problem of missing data

[2] To clarify, it uncovers an unbiased estimate of discrimination in hiring (or callbacks for interviews) for the applicants in the study. There are other criticisms of audit and correspondence studies related to how generalizable the results are to the population, and what the results tell us about the existence of discrimination at the level of the market. See HS (1993) and Heckman (1998) for more details. This paper does not address those issues.

[3] Another approach is to try to incorporate data in which productivity is observed or can be estimated (e.g., Foster and Rosenzweig, 1993; Hellerstein et al., 1999).

on productivity. Rather than try to control for variables that might be associated with productivity differences between groups, these studies instead create an artificial pool of job applicants, among which there are intended to be no average differences by race, for example. By using either applicants coached to act alike, with identical-quality resumes (an audit study), or simply applicants on paper who have equal qualifications (a correspondence study), the method is largely immune to criticisms of failure to control for important differences between, for example, black and white job applicants. As a consequence, this strategy has come to be widely used in testing for discrimination in labor markets (as well as housing markets). Thorough reviews are contained in Fix and Struyk (1993), Riach and Rich (2002), and Pager (2007).

Despite the widely-held view that audit or correspondence studies are the best way to test for labor market discrimination, critiques of these studies challenge their conclusions (HS, 1993; Heckman, 1998). Some of these criticisms have been readily acknowledged by researchers as potentially valid, and subsequent research has adapted. For example, HS noted that in the prominent audit studies carried out by the Urban Institute (e.g., Mincy, 1993), white and minority testers sent out to apply for jobs were aware of the purpose of the test, and even – in their training – informed about "the pervasive problem of discrimination in the United States," raising the possibility that testers subconsciously took actions in their job interviews that led to the "expected" result (HS, 1993). A constructive response to this criticism has been the move to correspondence studies, which focus on applications on paper and whether they result in job interviews, thus cutting out the influence of the individual job applicants used in the test.

However, a fundamental critique of audit or correspondence studies is one that has not been addressed by researchers. In particular, HS consider what most researchers view as the ideal conditions for an audit or correspondence study – when not only are the observable average differences between groups eliminated, but in addition the observable characteristics used in the applications are sufficiently rich that it is reasonable to assume that potential employers believe there no average differences in *unobservable* characteristics across groups. HS show that, even in this case, audit or correspondence studies can generate evidence of discrimination (in either direction) when there is none, and can also mask evidence of

discrimination when it in fact exists. Given the pervasive use of audit and correspondence studies, the failure of any research to address this critique is a significant gap in the social science and legal literature.

III. The Heckman-Siegelman Critique

Set-up with Continuous "Treatment" Decision

Suppose that productivity depends on two individual characteristics, $X' = (X^I, X^{II})$. Let R be a dummy for race, with R = equal to 1 for minorities and 0 for non-minorities (which I will refer to as "black" and "white" for short). Allow productivity also to depend on a firm-level characteristic F, so that productivity is P(X',F). Let the treatment of a worker depending on P and possibly R (if there is discrimination) be denoted T(P(X',F),R). For now, think of this treatment as continuous, even though that is not the usual outcome for an audit or correspondence study; suppose the treatment is, for example, the wage offered, set equal to a worker's productivity minus a possible discriminatory penalty for blacks as in Becker's (1971) employer taste discrimination model.

Define discrimination as

(1) $T(P(X',F)|R = 1) \neq T(P(X',F)|R = 0)$.

Assume that P(.,.) and T(P(.,.)) are additive, so

(2) $P(X',F) = X^I + X^{II} + F$

(3) $T(P(X',F),R) = P + \gamma'R$.[4]

Thus, discrimination against blacks implies that $\gamma' < 0$, so that blacks are paid less than equally-productive whites at the same firm.

In an audit or correspondence study, two testers (or applications) or multiple pairs of testers (one with R = 1 and one with R = 0 in each pair) are sent to firms to apply for jobs. The researcher attempts to standardize their productivity based on observable productivity-related characteristics. Denote expected productivity for blacks and whites, based on what the firm observes, as P_B^* and P_W^*; note that I have not specified that these are necessarily based on both X^I and X^{II}, as we may want to treat X^{II} as unobserved by

[4] For now, I treat the productivity variables in the abstract, and therefore without loss of generality can assume them to be scaled such that their coefficients equal one. When I turn to the data and relate productivity-related characteristics to expected productivity, I will introduce coefficients multiplying the productivity-related characteristics.

firms. The goal of the audit or correspondence study design is to set $P_B^* = P_W^*$. Given these observables, the outcome T is observed for each tester. So based on equation (3), each test – thought of as the outcome of applications to a firm by one black and one white tester – yields an observation

(4) $T(P_B^*,1) - T(P_W^*,0) = P_B^* + \gamma' - P_W^*$.

If $P_B^* = P_W^*$, then averaging across tests yields an estimate of γ'. Alternatively, γ' can be estimated from the regression

(5) $T(P_{ij}^*) = \alpha' + \gamma' R_i + \mu_j + \varepsilon_{ij}$

where $T(P_{ij}^*)$ is the outcome for worker i at firm j, R_i is a dummy variable for the race of worker i, and μ_j is a vector of firm fixed effects.[5]

Now consider explicitly the two observable components of productivity, X^I and X^{II}. Suppose the audit study controls only one of these; so think of X^I as the schooling level, which is controlled in the resumes/interviews, and X^{II} as another characteristic not controlled in the study and unobserved by employers. Denote by X_B^j and X_W^j the values of X^I and X^{II} for blacks and whites, j = I, II. Suppose that the audit or correspondence study, as is usually done, sets $X_B^I = X_W^I$; this is later denoted as X^{I*}, the level at which X^I is "standardized" across applicants. Then for the test resulting from the application of a pair of black and white testers to a firm, P_B^* and P_W^*, the firm's expected productivity for workers in each group, are

(6) $P_B^* = X_B^I + E(X_B^{II}) + F$

(7) $P_W^* = X_W^I + E(X_W^{II}) + F$.

In this case, each individual test provides an observation equal to

(8) $T(P_B^*,1) - T(P_W^*,0) = P_B^* + \gamma - P_W^* = X_B^I + E(X_B^{II}) + \gamma' - (X_W^I + E(X_W^{II}))$

$= \gamma + E(X_B^{II}) - E(X_W^{II})$.

[5] The inclusion of the firm fixed effects should have no implications for the estimate of γ', since the firm fixed effects are uncorrelated with race. Furthermore, when we introduce productivity-related characteristics X below, as long as testers' characteristics are randomly assigned to employers, which is typically the case in audit and correspondence studies, the estimated differentials associated with X do not depend on the inclusion of firm fixed effects. Thus, they are ignored in what follows. (Their inclusion does, however, affect the standard errors of the estimates, as the firm fixed effects may explain substantial amounts of variation in the outcome, and may also capture non-independence of the outcomes across testers at the same firm.)

Mean Differences

To set the stage, I first discuss issues regarding mean differences across groups in observed and unobserved variables. Clearly observations from a sample of the tests described above provide an unbiased estimate of γ' only if $E(X_B^{II}) = E(X_W^{II})$. Thus, a key assumption in an audit or correspondence study is that all productivity-related factors not controlled for in the test have the same mean for blacks and whites. Heckman (1998) and HS (1993) point out, however, that researchers have limited information about what determines productivity within firms. When there are uncontrolled productivity-related differences between black and white testers, not only will the audit or correspondence study produce a biased estimate of discrimination, but it can produce a *more* biased estimate than using randomly-matched pairs of testers. Suppose that, in addition to the preceding assumptions, X^I and X^{II} are statistically independent, and further assume that mean productivity is the same for blacks and whites,

$$(9) \qquad E(P_B) = E(X_B^I + X_B^{II}) = E(P_W) = E(X_W^I + X_W^{II}),$$

but that the mean of each characteristic differs by race, e.g., $E(X_W^I) \neq E(X_B^I)$. Coupled with equation (9), these assumptions imply that whites are more productive on one characteristic, and blacks on the other.

If an audit study controls for X^I but not X^{II}, the audit study estimates

$$(10) \qquad \gamma' + E(X_B^{II} - X_W^{II}),$$

The bias here can be upward or downward, depending on whether workers are matched on characteristics on which blacks are more or less productive than whites. If the characteristic on which they are matched, X^I, is on average higher for blacks than for whites, then $E(X_B^{II} - X_W^{II}) < 0$ and the audit study overstates discrimination, and conversely (discrimination against blacks implies $\gamma' < 0$).[6] In this case, given the assumption that the expected *sum* of the productivity components is equal for blacks and whites (equation (9)), if we sent *randomly-matched* pairs of testers we would get an unbiased estimate of discrimination, because such an audit study would yield estimates of

[6] The converse case is emphasized by Darity and Mason (1998), who suggest that while whites have higher values of some of the usual productivity controls included in statistical studies of discrimination (such as schooling and AFQT), blacks have higher values of psychological variables such as self-esteem or locus of control, and of effort.

(11) $\gamma' + E(X_B^I - X_W^I) + E(X_B^{II} - X_W^{II}) = \gamma' + E(X_B^I + X_B^{II}) - E(X_W^I + X_W^{II}) = \gamma'$.[7]

The previous case might be viewed as unrealistic, since its starting point is that blacks and whites are equally productive. Certainly much of the discrimination literature is premised on the likelihood that blacks are less productive and have lower values of most productivity-related characteristics (which is why simple comparisons of outcomes between blacks and whites likely overstate discrimination). In this case, under the other assumptions outlined above, the audit study leads to less bias than using randomly-matched pairs, because the two expected differences in the first expression in equation (11) are no longer equal to zero. Controlling for a larger number of productivity-related characteristics in an audit study will reduce the difference in uncontrolled productivity, and the remaining difference might be made sufficiently small that the bias in equation (11) becomes negligible.

However, that reasoning can break down if X^I and X^{II} are not statistically independent. Suppose, as before, that X^I and X^{II} each have different means for blacks and whites, with $E(X_B^{II} - X_W^{II}) < 0$, but that they are *not* independent. The problematic case is when X^I and X^{II} are negatively correlated, in which case

(12) $E(X_B^{II} - X_W^{II}) < E(X_B^{II} - X_W^{II} | X_B^I = X_W^I)$,

so that standardizing on X^I can accentuate the bias in an audit study, possibly leading to more bias from controlling for X^I rather than using randomly-matched testers – and possibly generating evidence of discrimination when γ' is in fact zero.[8] This depends on the relative magnitudes of $\{E(X_B^I - X_W^I) + E(X_B^{II} - X_W^{II})\}$ – the bias with randomly-matched testers – and $E(X_B^{II} - X_W^{II} | X_B^I = X_W^I)$ – the bias with testers matched on X^I. The case where X^I and X^{II} are positively correlated may be more realistic, in which case standardizing on X^I implies that the expected difference in X^{II} between the black and white tester is on

[7] This helps to demonstrate one of the main points of the HS critique with regard to mean differences between testers – that "Nowhere in the published literature on the audit pair method will you find a demonstration that matching one subset of observable variables necessarily implies that the resulting difference in audit-adjusted treatment between blacks and whites is an unbiased measure of discrimination – or indeed, that it is even necessarily a better measure of discrimination than comparing random pairs of whites and blacks applying at the same firm …" (Heckman, 1998, p. 108).

[8] Moreover, in this case the role of factors that are unimportant but still weakly related to productivity can be accentuated. HS discuss some examples in the context of the Urban Institute audit studies using Hispanic-white pairs.

average *smaller* than for a randomly-selected pair.[9] Regardless, the test is still biased.

The discussion to here focuses on the problems that arise because audit or correspondence studies may not control for all characteristics that differ by race. Good audit studies attempt to meet this criterion, but there are limits to how successfully they can do so.[10] Correspondence studies are a response to this potential problem, because, in contrast to audit studies, they do not entail face-to-face interviews that might convey mean differences on uncontrolled variables between blacks and whites.[11]

Even in a correspondence study, though, differences in employer estimates of mean unobserved characteristics for blacks and whites can affect the results, as in equation (10). The difference, in this case, is in part one of legal interpretation. In particular, because employers are not allowed to make assumptions about race (sex, etc.) differences in characteristics unobserved in the job application or interview process,[12] any role of assumed mean differences in characteristics in affecting the outcomes from a correspondence study can be interpreted as statistical discrimination. Consequently, we can interpret the estimate of the expression in equation (10) from a correspondence study as capturing the combined effects of taste discrimination (γ') and statistical discrimination ($E(X_B^{II} - X_W^{II})$).[13] That is, on legal grounds one might argue that correspondence studies, as opposed to audit studies, provide unbiased estimates of discrimination. According to this view, correspondence studies still do not, however, necessarily do any better at isolating taste discrimination – i.e., the discrimination that would remain if the means of all

[9] The argument is analogous to omitted variable bias in OLS regression. Suppose we are estimating the relationship between being black (B) and some outcome Y, and there are two omitted variables W and Z; ignore other covariates. Suppose W and Z are positively correlated with Y conditional on B (parallel to the example in the text, where the variables of interest, X^I and X^{II}, are positively related to productivity), but they have opposite-signed correlations with B. Then omitting both can lead to a less-biased estimate of the coefficient of B than including one but not the other. However, if all else is the same but W and Z have the same-signed correlation with B, then including one of them in the regression will unambiguously reduce the bias.

[10] As such, the statement in Hellerstein and Neumark (2006), that "The audit study approach … creates an artificial pool of labor market participants among whom there are no average differences by race or sex …" (p. 34) should be tempered. An audit study *attempts* to do this.

[11] This is a common argument in favor of correspondence studies, although they pose other disadvantages (e.g., BM, p. 994; Riach and Rich, 2002, p. F485). In addition, an audit study can explicitly compare results based on the application and interview stage, to see whether there is less discrimination at the application stage when face-to-face contact has been avoided (e.g., Neumark, 1996).

[12] See http://www.eeoc.gov/facts/fs-race.pdf (viewed March 23, 2009).

[13] This is noted in the literature. See, for example, Heckman and Siegelman (1993), Riach and Rich (2002), BM, and Lahey (2008). The latter three studies discuss the use of other information to try to distinguish between the two hypotheses.

productivity-related characteristics were the same for blacks and whites. Moreover, economists are inherently interested in whether employers discriminate against groups with equal observed and equal expected unobserved characteristics, and the preceding discussion implies that correspondence studies may not provide an unbiased estimate of this more fundamental type of discrimination. Nonetheless, when a correspondence study includes a rich set of applicant characteristics, it becomes less likely that statistical discrimination plays much of a role in group differences in outcomes.[14]

Distributional Differences in the Context of Hiring

A more troubling result emerges once we take account of the fact that, in audit or correspondence studies, the relevant treatment is not linear in productivity as it might be for a wage offer – like in equation (3) – but instead is non-linear. That is, we think that in the hiring process firms evaluate a job applicant's productivity relative to a standard, and offer the applicant a job (or an interview) if the standard is met. In this case, HS show that, even when there are equal group averages of both observed and unobserved variables, an audit or correspondence study can generate biased estimates, with spurious evidence of discrimination in either direction, or of its absence – or, in other words, discrimination is unidentified. Because this critique applies even to correspondence studies, which meet higher standards of validity – and applies even in the ideal case where there is *no* difference in the means of unobserved productivity measures – the remainder of the discussion refers exclusively to correspondence studies.

The intuitive basis of the HS critique is as follows. Consider the simplest case in which the only difference between blacks and whites is that the *variance* of unobserved productivity is higher for whites than for blacks, for example. That is, we assume $E(X_B^j) = E(X_W^j)$, j = I, II. Imagine a correspondence study that controls for one productivity-related characteristic, X^I, and standardizes on a quite low value of X^I (that is, the study makes the two groups equal on characteristic X^I, but at a low value X^{I*}). The correspondence study does not convey any information on a second, unobservable productivity-related

[14] This is more problematic in correspondence studies of age discrimination, because even with many other qualification on the resumes, if researchers give older applicants the same amount of experience as younger applicants employers are likely to make adverse assumptions about older applicants whose resumes reflect limited work experience. See Lahey (2008) and Riach and Rich (2007) for suggestions for addressing this problem in age discrimination studies.

characteristic, X^{II}. Because employers will offer a job interview only if the expected *sum* of $X^{I} + X^{II}$ is high, when X^{I*} is set at a low level, the employer has to believe that X^{II} is quite high in order to offer an interview. Even though the employer does not observe X^{II}, if the employer knows that the variance of X^{II} is higher for whites, the employer correctly concludes that whites are more likely than blacks to have a sufficiently high sum of $X^{I} + X^{II}$, by virtue of the simple fact that fewer blacks have very high values of X^{II}. Employers will therefore be less likely to offer jobs to blacks than to whites, *even though the observed average of X^{I} is the same for blacks and whites, as is the unobserved average of X^{II}*. The opposite holds if the standardization is at a high value of X^{I}; in the latter case the employer only needs to *avoid* very low values of X^{II}, which will be more common for whites.

It is worth pointing out that the idea that the variances of unobservables differ across groups has a long tradition in research on discrimination, stemming from early models of statistical discrimination. For example, Aigner and Cain (1977) discuss these models and suggest that a higher variance of unobservables for blacks compared to whites is plausible, and Lundberg and Startz (1983) study how such an assumption can lead to an equilibrium with lower investment in human capital by blacks. On the other hand, Neumark (1999) finds no evidence that employers have better labor market information about whites than blacks; if anything, the point estimates go in the opposite direction, although the estimates are imprecise.

To see the bias result formally in our simple framework, suppose that a job offer or interview is given if a worker's expected productivity exceeds a certain threshold c'. As before, suppose that P is determined as a linear sum of X^{I}, X^{II}, and F (equation (2)), with X^{II} (and F) statistically independent of X^{I},[15] and the correspondence study controls for X^{I}. The hiring rules for blacks and whites (with the possibility of discrimination) are

(13) $T(P(X',F)|R = 1) = 1$ if $\beta_I' X_B^{I} + X_B^{II} + \gamma' + F > c'$

(13') $T(P(X',F)|R = 0) = 1$ if $\beta_I' X_W^{I} + X_W^{II} + F > c'$.

Note that now X^{I} is now multiplied by a coefficient β_I', because this discussion pertains to

[15] Again, we can treat F as statistically independent because resume characteristics are, or should be, assigned randomly (e.g., Lahey and Beasley, 2009). And we can always think about X^{II} as the variation in the uncontrolled characteristic that is orthogonal to X^{I}.

estimating models corresponding to equations (13) and (13') using data on an observable productivity variable X^I. But given that X^{II} is unobserved, its coefficient can be standardized to equal one.[16] Discrimination leads employers to "discount" the productivity of a black worker, as captured in γ'.

Assume that the correspondence study controls for X^I, with $X_B{}^I = X_W{}^I = X^{I*}$. Assume further that $X_B{}^{II}$ and $X_W{}^{II}$ are normally distributed, with equal means (set to zero, without loss of generality), and standard deviations $\sigma_B{}^{II}$ and $\sigma_W{}^{II}$. Finally, as long as the firm-specific productivity shifters F are normally distributed and independent of X^{II}, and have the same distribution for blacks and whites, then we can ignore them and focus solely on the variation in X^{II}.[17] Under these assumptions, the probabilities that the left-hand expressions in (13) and (13') equal one (blacks and whites get hired) are

(14) $\Pr[T(P(X^{I*},X_B{}^{II})|R = 1) = 1] = 1 - \Phi[(c' - \beta_I'X^{I*} - \gamma')/\sigma_B{}^{II}] = \Phi[(\beta_I'X^{I*} + \gamma' - c')/\sigma_B{}^{II}]$

(14') $\Pr[T(P(X^{I*},X_W{}^{II})|R = 0) = 1] = 1 - \Phi[(c' - \beta_I'X^{I*})/\sigma_W{}^{II}] = \Phi[(\beta_I'X^{I*} - c')/\sigma_W{}^{II}]$,

where Φ denotes the standard normal distribution function.

The difference between the two expressions in equations (14) and (14') – the success rates for black and white job applicants – is intended to be informative about discrimination. However, even if $\gamma' = 0$, so there is no discrimination, these two expressions need not be equal because $\sigma_B{}^{II}$ and $\sigma_W{}^{II}$, the standard deviations of $X_B{}^{II}$ and $X_W{}^{II}$, can be unequal. It is possible to say something more precise. In particular, consider the earlier case with $\gamma' = 0$, but $\sigma_W{}^{II} > \sigma_B{}^{II}$ – that is, the "uncontrolled" productivity-related variable has a larger variance for whites than for blacks – and X^{I*} is set at a low level – i.e., the standardization level is low. Then the study will generate spurious evidence of discrimination against blacks. In particular, when $\beta_I'X^{I*} < c'$, then $\sigma_W{}^{II} > \sigma_B{}^{II}$ and $\gamma' = 0$ imply that the probability that blacks are hired is lower than the probability that whites are hired, and conversely when $\beta_I'X^{I*} > c'$.[18]

---

[16] Following HS, I assume that the coefficient on X^I is the same for blacks and whites, so the discrimination is reflected only in an intercept difference. I return to this issue later.

[17] That is, one can redefine the random variable in what follows as $X^{II} + F$, and the same reasoning goes through.

[18] Heckman (1998, footnote 7) suggests that the case with a low level of standardization and higher dispersion for whites "seems to rationalize" audit study evidence consistent with discrimination against blacks. It is not clear, however, that we know much about either the level of standardization or the relative dispersion of unobserved productivity. Even though the first issue relates to observables, there is no obvious way to compare the distributions of qualifications of testers in an audit study to the relevant population of job applicants. In fact, there are two

Thus, even if the means of the unobserved productivity-related variables are the same for each group, and firms use the same hiring standard for each group (i.e., $\gamma' = 0$), correspondence studies can generate evidence consistent with discrimination against blacks (or, alternatively, in their favor).[19] Different combinations of the relative magnitudes of σ_W^{II} and σ_B^{II}, and whether the standardization value of X^{I*} is high or low, can generate similar or opposite results. This is the basis for HS's claim that even under ideal conditions correspondence (or audit) studies are uninformative about discrimination.

IV. Detecting Discrimination

With the right data from a correspondence study, the framework from the preceding section can be used to recover an unbiased estimate of discrimination. The intuition is as follows. The HS critique rests on differences between blacks and whites in the variances of unobserved productivity. The fundamental problem, as equations (14) and (14') illustrate, is that we cannot separately identify the effect of race (γ') and a difference in the variance of the unobservables ($\sigma_B^{II}/\sigma_W^{II}$). But a higher variance for one group (say, whites) implies a smaller effect of observed characteristics on employment for that group.[20] Thus, information from a correspondence study on how variation in observable qualifications is related to employment outcomes can be informative about the relative variances of the unobservables, and this, in turn, can identify the effect of discrimination.

More formally, consider a correspondence study with the assumptions from the previous section holding. Equations (14) and (14') imply that the difference in outcomes between blacks and whites is

$$(15) \quad \Phi[(\beta_I'X^{I*} + \gamma' - c')/\sigma_B^{II}] - \Phi[(\beta_I'X^{I*} - c')/\sigma_W^{II}].$$

In a standard probit, we can only identify the coefficients relative to the standard deviation of the unobservable, so we normalize by setting the variance of the unobservable to equal one. In this case,

---

conflicting tendencies in setting standards for audit studies. Setting a low standard implies that call-back rates will be low, reducing the statistical power of the evidence. But setting a standard too high raises concerns about "overqualification" of candidates (e.g., BM, 2004, p. 995), which in an economic context presumably means that the employee will get a better job offer and hence will not take a job at the employer included in the test, in turn deterring the employment from making an offer.

[19] This argument does not depend on normality. It will hold for symmetric distributions (Heckman, 1998).

[20] In the limit, if the variance of unobserved X^{II} were infinite for whites, then X^I, the observed productivity-related variable, would have no effect on whether or not an employer thinks a white applicant meets the standard for hiring, given by equation (13').

impose the normalization for whites only, or $\sigma_W^{II} = 1$. The parameter σ_B^{II} is then the variance of the unobservable for blacks relative to whites. To make this clear, replace σ_B^{II} with $\sigma_{BR}^{II} = \sigma_B^{II}/\sigma_W^{II}$. The normalization $\sigma_W^{II} = 1$ is equivalent to defining all of the coefficients in equation (15) as their ratios relative to σ_W^{II}. Dropping the prime subscripts to indicate that the coefficients are now defined in relative terms, with this normalization equation (15) becomes

(15') $\Phi[(\beta_I X^{I*} + \gamma - c)/\sigma_{BR}^{II}] - \Phi[\beta_I X^{I*} - c]$.

Although this normalization does not solve the identification problem, it helps illustrate the solution. As equation (15') shows, without knowing σ_{BR}^{II} we cannot tell whether the intercepts of the two probits differ because $\gamma \neq 0$ or because $\sigma_{BR}^{II} \neq 1$. However, if there is variation in the level of qualifications used as controls (X^{I*}), and these qualifications affect hiring outcomes, then we can identify β_I/σ_{BR}^{II} and β_I in equation (15'), and the ratio of these two estimates provides an estimate of σ_{BR}^{II}.[21] And if we do inference on this ratio, we can test the hypothesis of equal standard deviations (or variances) of the unobservables. Finally, identification of σ_{BR}^{II} implies identification of γ. Note that without meaningful variation in X^{I*} this is not possible, since in that case all we have in the model are different intercepts with different parameters in both the numerators and the denominators ($(\gamma - c)/\sigma_{BR}^{II}$ and c).

The critical assumption to identify σ_{BR}^{II} and hence γ is that β_I is the same for blacks and whites. Otherwise, the ratio of the two coefficients of X^{I*} for blacks and whites does not identify σ_{BR}^{II}. As HS point out, the constancy of β_I is assumed in the Urban Institute studies that they critique, with discrimination entering through an intercept shift in the evaluation of a worker's productivity, depending on their race. In the real world it is not hard to come up with reasons why the coefficients relating X^I to productivity might differ by race. For example, blacks and whites on average attend different schools, and if white schools are higher quality, a given number of years of schooling may do more to increase white productivity than black productivity. But in a correspondence or audit study, it should be possible to

[21] The discussion here is in terms of probit estimates of callbacks. It could just as well be couched in terms of logit estimation. Although typically (e.g., Maddala, 1983) the logit model is not written with the standard deviation of the error term appearing, it is possible to rewrite it in this way, in which case the difference in coefficients would again be informative about the ratio of the variances of the unobservable (Johnson and Kotz, 1970, p. 5).

control for these kinds of differences; for example, in this case one can control for the area where applicants live, and hence hold school district constant (e.g., BM, 2004).

HS raise other possibilities. One is that there is discrimination in evaluating particular attributes of a group. For example, employers may discriminate against high-education blacks but not low-education blacks. It is not possible to rule out differences in coefficients arising for these reasons. Finally, HS also suggest that differences in coefficients may reflect "statistical information processing," given incomplete information about productivity, as in statistical discrimination models. Of course this is the idea underlying the identification strategy suggested above, as the difference in β_I for blacks and whites is assumed to reflect precisely the accuracy with which X^I signals productivity for each race. However, as discussed below, when there is data on multiple productivity-related characteristics, there is more one can do to test whether there is homogeneity in the coefficients that allows identification of σ_{BR}^{II} and hence γ.

The estimation of β_I/σ_{BR}^{II} and β_I, and inference on their ratio ($\sigma_{BR}^{II} = \sigma_B^{II}/\sigma_W^{II}$), can be done via a heteroskedastic probit model (e.g., Williams, 2009), which allows the variance of the unobservable to vary with race. To do this, pool the data for blacks and whites. Similar to equation (5), there is a latent variable for perceived productivity assumed to be generated by

$$T(P_{ij}^*) = -c + \beta_I X_{ij}^{I*} + \gamma R_i + \varepsilon_{ij}. \quad (16)$$

As is standard, it is assumed that $E(\varepsilon_{ij}) = 0$. But the variance is assumed to follow

$$Var(\varepsilon_{ij}) = [\exp(\mu + \omega R_i)]^2. \quad (17)$$

This model can be estimated via maximum likelihood. The observations should be treated as clustered on firms to obtain a variance-covariance matrix that is robust to the dependence of observations across firms. The normalization $\mu = 0$ can be imposed, given that there is an arbitrary normalization of the scale of the variance of one group (in this case whites, with $R_i = 0$). And the estimate of $\exp(\omega)$ is exactly the estimate of σ_{BR}^{II}.

In this heteroskedastic probit model, maintaining the assumption that β_I is the same for blacks and whites, γ is identified. Observations on whites identify $-c$ and β_I, and observations on blacks identify ($-c$ +

$\gamma)/\exp(\omega)$ and $\beta_I/\exp(\omega)$. Thus, the ratio of $\beta_I/\{\beta_I/\exp(\omega)\}$ identifies $\exp(\omega)$, which, from equation (17), is the ratio of the standard deviation of the unobservable for blacks relative to whites and is, as before, identified from the ratio of the effect of X^{I*} on blacks relative to its effect on whites.[22] With the estimate of $\exp(\omega)$ (or equivalently σ_{BR}^{II}), along with the estimate of c identified from whites, the expression $(-c + \gamma)/\exp(\omega)$ identified from blacks identifies γ as well.[23]

Note that evidence on σ_{BR}^{II} is itself informative. If it equals one, then there is no bias from differences in the distribution of unobservables. This would imply that a correspondence study, at least, is free from bias (under the other maintained assumptions). Alternatively, if σ_{BR}^{II} is not equal to one, but if we had some evidence on how the level of standardization X^{I*} compares to the relevant population of job applicants, we could determine the direction of bias. For example, if the study *detects* discrimination and there is a bias *against* this finding – based on the estimate of the ratio of variances and information about X^{I*}, then the evidence of discrimination is not spurious, because it would be even stronger absent this bias. But because we can identify γ directly under the assumptions above, we can recover an estimate of discrimination that is not biased by the difference in the variances of the unobservables. And we can do this *without* having to make a determination as to whether X^{I*} used in the study is a high or low level of standardization; this is important because it may be impossible to establish the latter.

Thus, given the right data, we can determine whether a race difference in outcomes in a correspondence study in fact reflects discrimination. No doubt those assumptions are restrictive, and there are almost surely ways to relax these assumptions and render the data uninformative about γ. But the approach is structural, and therefore of necessity rests on restrictions regarding parameters and functional forms.

---

[22] The fact that the estimate of σ_{BR}^{II} is inversely related to the relative effect of X^{I*} on the probability that blacks are hired makes sense. For example, when the variance of the unobservable is larger for blacks, a given change in X^{I*} has a smaller effect on the probability that a black is hired than on the probability that a white is hired, because this change in X^{I*} is less informative about black productivity, which is the sum $X^I + X^{II}$. This also clarifies why the availability of data with multiple values of X^{I*} is essential to estimating the relative variance of the unobservables.

[23] Consistent with the earlier discussion of statistical discrimination, we might want to allow for the possibility that $E(X_B^{II} - X_W^{II}) \neq 0$. In this case, we can normalize by assuming $E(X_W^{II}) = 0$ and defining $E(X_B^{II} - X_W^{II}) = \mu_{BW}^{II}$. We can then replace γ in the preceding identification argument with $\gamma + \mu_{BW}^{II}$, and it is this sum of parameters, reflecting the combination of taste discrimination and the expected mean difference in the unobservable, that is identified.

The assumption that β_I is the same for blacks and whites cannot be tested if there is only one productivity control. With more controls, however, there is a testable restriction. In particular, if the coefficients of multiple productivity controls differ between blacks and whites *only* because of the difference in the variance of the unobservable, the ratio of the coefficients between blacks and whites, for each variable, should be the same.[24] Consider the case with two observables X^I and Z^I, modifying equation (15') to be

(15'') $\Phi[(\beta_I^B X^{I*} + \delta_I^B Z^{I*} + \gamma - c)/\sigma_{BR}^{II}] - \Phi[(\beta_I^W X^{I*} + \delta_I^W Z^{I*} - c)]$.

The coefficients on the observables now have B and W superscripts to denote possible differences by race. But if the coefficients differ by race only because $\sigma_{BR}^{II} \neq 1$, then

(18) $\beta_I^B/\beta_I^W = \delta_I^B/\delta_I^W$.

Thus, the restrictions implied by homogeneity of effects but unequal variances of the unobservables can be tested. Of course failure to reject the restrictions does not decisively rule out the possibility that $\sigma_{BR}^{II} = 1$, with the coefficients differing by race for other reasons but equation (18) still holding. With a larger number of control variables, however, it seems unlikely that this alternative scenario would explain failure to reject the restrictions in equation (18). It is also possible to choose – as an identifying assumption – a subset of the observable characteristics for which equation (18) holds, and to identify σ_{BR}^{II} only from the coefficients of this subset of variables, or to test the restriction on the other coefficients as overidentifying restrictions.

A final issue concerns the interpretation of the coefficients from the heteroskedastic probit model. For probits, coefficient estimates are translated into estimates of the marginal effects of a variable using

(19) $\partial P(\text{hire})/\partial Z_k = \beta_k \phi(Z\beta)$,

where, using more general notation, Z_k is the variable of interest with coefficient β_k, and Z is the vector of controls with coefficients β); $\phi(.)$ is the standard normal density and the standard deviation of the unobservable is normalized to one. Typically this is evaluated at the means of Z. When Z_k is a dummy

[24] The possibility of differences in coefficients owing to statistical discrimination could also generate differences in coefficients between blacks and whites, and this could vary for different characteristics to the extent that the signal content of these characteristics varies. The argument below applies to this case as well.

variable such as race, the difference in the cumulative normal distribution functions is often used instead, although the difference is usually trivial.

The marginal effect is more complicated in the case of the heteroskedastic probit model, because if the variances of the unobservable differ by race, then when race "changes" both the variance and the level of the latent variable that determines hiring can shift. As long as we use the continuous version of the partial derivative to compute marginal effects from the heteroskedastic probit model, there is a natural decomposition of the effect of a change in Z_k into these two components. In particular, when the notation of equation (17) is generalized to

(17') $\text{Var}(\varepsilon_{ij}) = [\exp(W\omega)]^2$,[25]

where the vector of variables W includes Z (from equation (19)), with coefficient ω_k, then the overall partial derivative of P(hire) with respect to Z_k is

(20) $\partial P(\text{hire})/\partial Z_k = \phi(Z\beta/\exp(W\omega))\cdot\{(\beta_k - Z\beta\cdot\omega_k)/\exp(W\omega)\}$.[26]

This expression can be broken into two pieces. First, the partial derivative with respect to changes in Z_k affecting only the level of the latent variable – corresponding to the counterfactual of Z_k changing the valuation of the worker without changing the variance of the unobservable – is equal to

(20') $\phi(Z\beta/\exp(W\omega)) \cdot\{\beta_k/\exp(W\omega)\}$.

Second, the partial derivative with respect to changes via the variance of the unobservable is equal to

(20'') $\phi(Z\beta/\exp(W\omega))\cdot\{(-Z\beta\cdot\omega_k)/\exp(W\omega)\}$.

In the analysis below, these two separate effects are reported as well as the overall marginal effect, and standard errors are calculated using the delta method. However, the whole point of the HS critique is that differential treatment of blacks and whites based only on differences in variances of the unobservable should not be interpreted as discrimination.[27] Hence, what we are interested in comparing with the standard marginal effect from the probit is the expression in equation (20') – the effect of race via the latent variable

[25] Recall that μ in equation (17) is normalized to zero.
[26] See Cornelißen (2005).
[27] This issue is discussed in greater detail following the empirical analysis.

– which captures how race affects employers' evaluation of worker productivity.

V. Evidence, Implementation, and Assessment

Existing Evidence

As the preceding discussion shows, we need information on the effects of productivity-related characteristics on hiring or callbacks, estimated separately for blacks and whites (or other groups), to identify discrimination in a correspondence study, or even to assess likely biases. Reporting of such evidence is rare in the literature, because correspondence studies typically create one "type" of applicant. However, BM's well-known correspondence study of race discrimination is unusual in that – for reasons unrelated to the concerns of this paper – it uses two types of applicants.[28]

Part of their analysis studies studying callback differences by race for resumes that they constructed to be low versus high quality.[29] White callback rates are higher for both types of resumes. But although white callback rates increase significantly with resume quality (from 8.5 to 10.8 percent), black callback rates increase only slightly (from 6.2 to 6.7 percent), and the change is not statistically significant, leading them to conclude that, "African-Americans experience much less of an increase in callback rates for similar improvements in their credentials" (pp. 1000-1). Similar qualitative conclusions are reached based on an analysis that measures resume quality for one part of the sample by using an equation for the probability of callbacks estimated from another part of the sample. In this analysis, both groups experience an increase in callback rates from higher-quality resumes, but the effect is larger for whites.

Similarly, BM report probit models estimated for whites and blacks separately (their Table 5). These estimates reveal substantially stronger effects of measured qualifications for whites than for blacks. Among the estimated coefficients that are statistically significant for at least one group, this is true for experience,[30] having an email address, working while in school, academic honors, and other special skills

[28] BM actually study differences in treatment between applicants with black-sounding names and names that do not sound black. For simplicity, I discuss the results as if they capture differences between blacks and whites, which is certainly a plausible interpretation of their findings.

[29] Their analysis is motivated by the question of whether blacks and whites have different incentives to invest in skills, as in the Lundberg and Startz (1983) model.

[30] This variable enters as a quadratic, and the effect of experience is stronger for whites up to about 16 years of experience, more than twice the mean in their sample.

(such as language). The only exception is for computer skills, which inexplicably have a negative effect on callback rates for whites.[31]

As the present paper suggests, an alternative interpretation of smaller estimated probit coefficients or marginal effects for blacks than for whites is a difference in the variance of the unobservables. In particular, the lower coefficients for blacks are consistent with a larger variance for blacks, i.e., $\sigma_{BR}^{II} > 1$. If it is also true that BM standardized applicants at low levels of the control variables, then the HS analysis would imply that there is a bias towards finding discrimination *in favor of* blacks; that is, the evidence of discrimination against blacks would be even stronger absent the bias from differences in the distribution of unobservables. BM explicitly state that they tried to avoid overqualification even of the *higher-quality* resumes (p. 995). But it is very difficult to assess whether the characteristics of applicants were low, since there is no way to identify the population of applicants. Hence, implementation of the estimation procedure proposed in this paper is likely the only way even to *sign* the bias, let alone to recover an unbiased estimate of discrimination.

Implementation Using Bertrand and Mullainathan Data

Because BM's data include applicants with different levels of qualifications, and the qualifications predict callbacks, their data can be used to implement the methods described earlier. Table 1 begins by simply presenting probit estimates for the probability of a callback. Marginal effects are reported for specifications with no controls except a dummy variable for females (in columns (1)-(3)), adding controls for the individual characteristics included on the resumes, and finally adding also neighborhood characteristics for the applicant's zip code; the specific variables are listed in the footnote to the table. Estimates are shown for males and females combined, and for females only; as the sample sizes indicate, the male sample is considerably smaller.[32] Aside from the estimated effects of race, estimates are shown for a few of the resume characteristics capturing applicants' qualifications.

[31] Also, the effect of gaps in employment is inexplicably positive, but not significant for estimates disaggregated by race.

[32] Probits estimated for males only yielded similar results for the effects of race, although the estimated coefficients of some of the productivity-related characteristics were quite imprecise or had unexpected signs. In estimating the heteroskedastic probit model for males, in some cases there were computational problems, likely reflecting these other issues regarding the estimates for males, and perhaps also the much smaller sample for males.

Echoing BM's conclusions, there is a sizable and statistically significant difference between the callback rates for blacks and whites, with the rate for blacks lower by 3-3.3 percentage points (or about 33 percent relative to the white callback rate of 9.65 percent). The estimated race differences are robust to the inclusion of the different sets of control variables, which is what we should expect in a correspondence study in which the resumes are assigned randomly.

Interestingly, in light of the results of other audit and correspondence studies, there is no evidence of lower (or higher) callback rates for females than for males.[33] However, BM's study was to a large extent focused on jobs typically held by females, and was not designed to test for sex discrimination.[34] At the same time, the table also shows that a number of the resume characteristics have statistically significant effects on the callback probability; this, of course, is an essential input for using the methods described above to recover an unbiased estimate of discrimination.

The main analysis is reported beginning in Table 2, for the specifications with the full set of individual resume controls, and then adding as well the full set of neighborhood controls. Panel A simply repeats the estimated race effects from Table 1, for comparison. Panel B begins by reporting the estimated overall marginal effects of race from the heteroskedastic probit model (equation (20)). As the table shows, these estimates are slightly smaller (in absolute value) than the estimates from the simple probits, but trivially so. They remain statistically significant, and indicate callback rates that are lower for blacks by about 2.4-2.5 percentage points (or about 25 percent).

However, these effects represent the effects on both the level of the latent variable (the valuation of the worker's productivity) and the variance of the unobservable. Decomposing the marginal, the effect via the level of the latent variable is larger than the marginal effect from the probit estimation, ranging from −0.054 to −0.086. The effect of race via the variance of the unobservable, in contrast, is positive, ranging from 0.028 to 0.062. (This latter effect is not statistically significant.) The implication is that race discrimination is more severe than indicated by the analysis that ignores the role of differences in the

[33] And although not reported in the tables, this was true if the same methods used below to recover unbiased estimates of race discrimination were applied to the estimation of sex discrimination.
[34] They study sales, administrative support, clerical, and customer service jobs. The male applicants were used almost exclusively for the sales jobs, so the sex difference is identified mainly from the sales jobs.

variances of the unobservables. Note that this evidence that the probit estimates understate discrimination against blacks is consistent with a low level of standardization of X^{I*}, coupled with a higher estimated variance of the unobservable for blacks, as conjectured earlier based on BM's results. And as reported in the next row of the table, the estimated ratio of the standard deviation of the unobservable for blacks to the standard deviation for whites always exceeds one, although the difference is not statistically significant. Note that the positive effect of being black via the variance is what we expect if X^{I*} is low, since then a larger relative variance for blacks increases the relative probability that they are hired (called back).[35]

The next two rows of the table report some diagnostic test statistics. First, the p-values from the test of the restrictions in equation (18) are shown, based on probit specifications interacting all of the controls with race. Recall that this restriction implies that the ratios of coefficients of blacks to whites are equal for all productivity-related characteristics, which is consistent with the differences between the coefficients for blacks and whites reflecting simply a difference in the variances of the unobservables. In all four cases this restriction is not rejected, with p-values ranging from 0.17 to 0.62, implying that the restrictions necessary to identify an unbiased estimate of discrimination from the heteroskedastic probit model are not rejected. Nonetheless, the lower end of this range of p-values suggests that the restrictions sometimes might be fairly inconsistent with the data. As a consequence, below some alternative estimates are discussed that use only a subset of variables for which equation (18) is more consistent with the data.

Finally, the subset of control variables for which the absolute value of the estimated coefficient for whites exceeded that for blacks – consistent with the larger standard deviation of unobservables for blacks – was identified. Then the heteroskedastic probit model was estimated leaving the race interactions of the *other* variables *in* the model – so that the restrictions from equation (18) that were less consistent with the data were not imposed – and the joint significance of these latter variables was tested. Despite this latter subset of variables having estimated coefficients less consistent with the restrictions in equation (18), the p-values indicate that these interactions can also be excluded from the model. This can be viewed as an overidentifying test of the restriction that there are no differences in the effects of any of the control

[35] Note that this mirrors the standard result in the statistical discrimination literature. When a group has a higher variance of the unobservable, then at low levels of qualifications that group will be favored, and vice versa.

variables by race, for the specifications for which the estimates are reported in the first row of Panel B. Although technically it is only necessary to assume that there is a single variable for which the coefficient is the same for blacks and whites, there is no obvious variable to choose for the purposes of identification; here, instead, I let the data select a set of variables more consistent with the identifying restriction.

Table 3 follows up on the last procedure, by instead simply dropping from the analysis the control variables for which the absolute value of the estimated coefficient for whites was less than for blacks. As we would expect, the p-values for the tests of this set of restrictions are now much closer to one, ranging from 0.68 to 0.92, compared with a range of 0.17 to 0.62 in Table 2. However, as the table shows, the estimated effects of race are similar to those in Table 2 and do not point to any different conclusions.

Monte Carlo Assessment

This subsection provides Monte Carlo evidence on how well the estimation procedure proposed in this paper work in terms of removing the bias in estimates of discrimination from correspondence study evidence. First, results similar to those in HS and Heckman (1998) are generated, illustrating the potential bias when using probit analysis. Then the performance of the heteroskedastic probit in eliminating the bias is evaluated.

The upper left-hand panel of Figure 1 replicates the basic result from Heckman (1998), showing that probit analysis of the data from a correspondence study can generate substantial bias in either direction. Paralleling Heckman, this is done for the case in which $c = 0$, $\beta_I = 1$, $Var(X_W^{II})/Var(X_B^{II}) = (\sigma_W^{II})^2/(\sigma_B^{II})^2 = 2.25$,[36] and there is no discrimination ($\gamma = 0$). So for the Monte Carlo simulations, the assumed data generating process is $X^{I*} \sim N(0,1)$, $X_B^{II} \sim N(0,1)$, $X_W^{II} \sim N(0,2.25)$. Paralleling the standardization of correspondence study applicants, the data are generated by sampling X^{I*} from a truncated normal distribution, in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$. The simulation is done 100 times at each value of X^{I*} shown in the graph, with samples of 2,000 blacks and 2,000 whites in each simulation (roughly BM's sample sizes), and a probit model is estimated for each simulated data set.[37] The figures show – for both the estimates of γ

[36] This is the ratio of the variances of the unobservables. Note that a larger value for whites is the opposite of the common assumption in models of statistical discrimination.

[37] Given that the probit estimates were very robust, only a small number of simulations was used.

and the marginal effects – the means of the true values based on the assumed parameters, and the means based on the estimates.[38]

The figure clearly illustrates that, despite the absence of discrimination in the data generating process (the true effect is constant at 0), the evidence can either point to discrimination against blacks or discrimination in favor of blacks, depending on the level of standardization of X^{I*}. The marginal effects show that even though there is no discrimination in the data generating process, quite strong evidence of discrimination in either direction can emerge, with a marginal effect of −0.1 (0.1) for low (high) values of X^{I*}. Finally, as we would expect, only at $X^{I*} = 0$ is the estimate of γ (and the marginal effect) unbiased. The lower two panels of Figure 1 report the same kind of evidence in this case with $\gamma = -0.5$, consistent with discrimination. A similar result is apparent, with substantial bias relative to the true γ or the true marginal effect.

The heteroskedastic probit estimation requires data with multiple levels of the value of X^{I*} at which applicants are standardized. As an intermediate step to separate out the consequences of generating the data this way, and the consequences of implementing the heteroskedastic probit estimator, Figure 2 shows results with such generated data, but continuing to use the probit specification. (Following this, the heteroskedastic probit results with the same type of data are reported.) X^{I*} is now sampled from two truncated normal distributions, one using X^{I*} in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$, as before, and the second using instead $X^{I*} + 0.5$, again in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$. Figure 2 shows qualitatively similar results to Figure 1, so simply using data with variation in productivity-related characteristics does not, in itself, eliminate the bias. Nonetheless, the biases in both the no discrimination and discrimination cases are a bit smaller than in Figure 1 because of the larger range covered by X^{I*}.[39]

---

[38] The true marginals are based on the heteroskedastic probit specification (equation (20'), because the simulated data are heteroskedastic. The true marginal effect is reported as a mean because it is computed once using each simulated data set, holding the parameters fixed, and then averaged.

[39] Note also that in this case the unbiased estimate occurs at the value of −0.25 (for X^{I*}) on the horizontal axis, where the average of the upper and lower standardization levels equals zero. The reduction in bias is little less clear in the discrimination case. To clarify, the bias in Figures 1 and 2 should be contrasted at comparable value of X^{I*}, given that Figure 2 shows the mean estimates at the lower level of standardization of X^{I*}. For example, for the discrimination case, the mean estimate of γ at $X^{I*} = 1$ in Figure 1 should be compared to the mean estimate at $X^{I*} = 0.75$ in Figure 2 (in which case this is the lower standardization level and the average is 1); the latter estimate is in fact closer to zero.

Figure 3 reports results for the heteroskedastic probit estimation, using the same data generating process for simulating data as in Figure 2, although in this case 5,000 simulations are run for each pair of values of X^{I*}, because the heteroskedastic probit estimation is less precise than the simple probit estimation. The top panel covers the no discrimination case ($\gamma = 0$). The left-hand graph shows the means of the true and estimated values of the marginal effects for each value of X^{I*}. These are largely indistinguishable in the figure, indicating no bias. The right-hand panel provides evidence on the distribution of the estimates, showing the distance between the 25th and 75th percentiles of the estimates and between the 2.5th and 97.5th percentile at each value of X^{I*}. The distribution of estimates is quite tight at levels of standardization near the center of the distribution of X^{I*}, but becomes wider at more extreme values, when hiring rates in the generated data move towards zero or one. The discrimination case ($\gamma = -0.5$) similarly demonstrates that the heteroskedastic probit estimation eliminates the bias.

The last analysis, reported in Figure 4, considers the implications of the data generating process violating the identifying assumption that the coefficient(s) on the productivity-related characteristics are equal for blacks and whites. Results are presented for two cases: mild violation in which the coefficient on X^{I*} (β_I) is slightly larger for whites than for blacks (1.1 versus 1); and strong violation in which it is much larger (2 versus 1). As Figure 4 shows, in the case of no discrimination – the left-hand panels – the results are indistinguishable from when the identifying assumption is not violated. In contrast, in the discrimination case the estimated marginal effects become more negative than the true effects over much of the range, only slightly with mild violation of the identifying assumption, but more so when the violation is more pronounced.

The qualitative implications of what happens when the identifying assumption is violated make sense, thinking about how γ is identified. Using estimates of the separate probits in equations (14) and (14'), the ratio of the standardized white probit coefficient (setting $\sigma_W^{II} = 1$) to the black probit coefficient (β_I/σ_B^{II}) identifies σ_{BR}^{II} (which, recall, equals $\sigma_B^{II}/\sigma_W^{II}$). When the true value of β_I is larger for whites than

for blacks, but it is assumed that they are equal, σ_{BR}^{II} is overestimated.[40] Recall from the earlier discussion that the probit for blacks identifies $(-c + \gamma)/\exp(\omega) = (-c + \gamma)/\sigma_{BR}^{II}$. Because c = 0 in the simulations, we identify γ by multiplying the estimate of this expression by the estimate of σ_{BR}^{II}; the upward bias in the estimate of σ_{BR}^{II} therefore implies that the estimate of γ is biased away from zero. In the no discrimination case, when $\gamma = 0$, this is irrelevant; multiplying an estimate that averages zero by the upward-biased estimate of σ_{BR}^{II} has no effect. But when the true γ is non-zero (and negative), this bias leads to an estimate of γ that is more negative. When γ is more negative, we get exactly the "bending" of the estimated marginal effects that the right-hand panels of Figure 4 illustrate. In the standard marginal effect $-\gamma\phi(Z\beta)$ – nearer the center of the distribution the larger estimate of γ dominates the marginal effect, whereas nearer the tails the larger estimate of γ lowers $\phi(Z\beta)$ enough that the product $\gamma\phi(Z\beta)$ is closer to zero. Of course this latter result depends on the direction in which the identifying assumption is violated. Clearly a violation of the assumption in the opposite direction (β_I larger for blacks) would lead to biases in the opposite direction. Nonetheless, it follows from this reasoning that the bias is multiplicative, and hence does not generate the wrong *sign* for the estimate of γ. Moreover, it does not generate spurious evidence of discrimination when there is no discrimination.

VI. The Meaning of Discrimination

The simplest case depicted in the top panels of Figure 1, in which a correspondence study generates evidence of discrimination when $\gamma = 0$, raises the question of whether the evidence reflects a different kind of discrimination. In this case, the productivity of blacks and whites are regarded equally by employers (or equivalently there is no taste discrimination). Moreover, employers are not making any assumption about mean differences in unobservables between blacks and whites. However, they are making assumptions about distributional differences with regard to the variance of unobservables, and it is these assumptions that lead them – given the level of standardization of the study applicants – to prefer one race over the other. This could be thought of as "second-moment" statistical discrimination.

[40] For example, in the case in the top panel of Figure 4, the ratio of coefficients is $(\beta_I \cdot 1.1)/(\beta_I/\sigma_{BR}^{II}) = 1.1 \cdot \sigma_{BR}^{II}$.

To the best of my knowledge, differential treatment based such assumptions (true or not) about variances have never been viewed as discriminatory in either the economics or the legal literature. Nonetheless, the HS critique could be recast as showing that the analysis of data from a standard audit or correspondence study cannot distinguish between discrimination as it usually interpreted, and discrimination based on different variances of unobservables. Similarly, the method proposed in this paper to collect and analyze data from a correspondence study can be viewed precisely as distinguishing between what is typically viewed as discrimination (stemming from tastes or, as noted earlier, from standard statistical discrimination) and discrimination stemming from differences in variances of the unobservable.

Arguably, however, what we are interested in is correspondence study evidence on taste discrimination (or "first-moment" statistical discrimination) – which affects the valuation of the worker's productivity in the latent variable model for expected productivity. In contrast, evidence of second-moment statistical discrimination is largely irrelevant, because it is simply an artifact of how a correspondence study is done – in particular, standardization of applicants to particular, and similar, values of the observables. That is, only because applicants come from a narrow range of the distribution of observables does the variance of the unobservable affect the hiring decision, and as Figure 1 shows, depending on the standardization value, the evidence could point to discrimination against either group. In the real economy, in contrast, applicants – tautologically – would be representative of actual applicants to a job, in which case – as Figure 1 suggests – the differential treatment of blacks and whites depending on their level of qualifications and the variances of the unobservable would likely balance out, leading to no average difference in treatment. In contrast, when a correspondence study detects discrimination that comes through the latent variable – which can be interpreted as the relative valuation of a worker's productivity à la Becker (1971) – that evidence will also carry over to the real economy.

VII. Conclusions and Discussion

Many researchers view audit and correspondence studies as the most compelling way to test for labor market discrimination. And research applying these methods to many different types of groups nearly always finds evidence of discrimination. The use of audit studies to test for labor market

discrimination has been criticized on numerous grounds having to do with whether applicants from different groups appear identical to employers. Many of these criticisms can be countered by using correspondence studies in which fictitious applicants on paper are substituted for fictitious applicants in person.

However, Heckman and Siegelman (1993) show that even in correspondence studies in which group averages are identical conditional on the controls, group differences in the variances of unobservable dimensions of productivity can invalidate the empirical tests, leading to spurious evidence of discrimination in either direction, or spurious evidence of an absence of discrimination. This is an important criticism of correspondence studies, as it implies that evidence regarding discrimination from even the best-designed correspondence study can give misleading evidence about discrimination. Nonetheless, the criticism has been ignored in the literature.

This paper shows that if the correspondence study includes observable measures of variation in applicants' quality that affect hiring outcomes, an unbiased estimate of discrimination can be recovered even when there are group differences in the variances of the unobservable. The method is applied to Bertrand and Mullainathan's (2004) correspondence study, and leads to stronger evidence of race discrimination that adversely affects blacks than is obtained when differences in the variances of the unobservable are ignored. Moreover, this conclusion is bolstered by Monte Carlo simulations suggesting that the estimation procedure performs well, eliminating the problems highlighted by Heckman and Siegelman that could otherwise lead to badly misleading conclusions from the analysis of data from correspondence (or audit) studies.

Finally, it should be recognized that the method proposed here can be easily implemented in any future correspondence study. All that is needed is for the resumes or applications to include some variation in characteristics that affect the probability of being hired. This is different from what is sometimes done in designing correspondence studies, where researchers try to create a bank of resumes of essentially equally-qualified candidates. Clearly researchers sometimes inadvertently create resumes of different quality. All that needs to be done, however, is to intentionally do this. Once a researcher confirms that a set of

productivity-related characteristics on the resumes affected hiring outcomes, it is then possible to test for bias from different variances of the unobservables for the two groups, and more importantly to detect actual discrimination.

References

Aigner, Dennis J., and Glen Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *Industrial and Labor Relations Review*, Vol. 30, No. 2, January, pp. 175-87.

Banerjee, Abhijit, Marianne Bertrand, Saugato Datta, and Sendhil Mullainathan. 2008. "Labor Market Discrimination in Delhi: Evidence from a Field Experiment." *Journal of Comparative Economics*, Vol. 38, No. 1, March, pp. 14-27.

Becker, Gary S. 1971. The Economics of Discrimination, Second Edition. Chicago: University of Chicago Press.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review*, Vol. 94, No. 4, September, pp. 991-1013.

Bovenkerk, F., M. Gras, and D. Ramsoedh. 1995. "Discrimination Against Migrant Workers and Ethnic Minorities in Access to Employment in the Netherlands." *International Migration Papers*, No. 4, International Labour Office, Geneva, Switzerland.

Cornelißen, Thomas. 2005. "Standard Errors of Marginal Effects in the Heteroskedastic Probit Model." Institute of Quantitative Economic Research, Discussion Paper No. 230.

Darity, William A., Jr., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 63-90.

Fix, Michael, and Raymond Struyk. 1993. Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, DC: The Urban Institute Press.

Foster, Andrew, and Mark Rosenzweig. 1993. "Information, Learning, and Wage Rates in Low-Income Rural Areas." *Journal of Human Resources*, Vol. 28, No. 4, Fall, pp. 759-90.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives*, Vol. 12, No. 2, Spring, pp. 101-16.

Heckman, James, and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In Fix and Struyk, eds., Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, D.C.: The Urban Institute Press, pp. 187-258.

Hellerstein, Judith K., and David Neumark. 1999. "Sex, Wages, and Productivity: An Empirical Analysis of Israeli Firm-Level Data." *International Economic Review*, Vol. 40, No. 1, February, pp. 95-123.

Hellerstein, Judith, and David Neumark. 2006. "Using Matched Employer-Employee Data to Study Labor Market Discrimination." In Rodgers, ed., Handbook on the Economics of Discrimination. Great Britain: Edgar Elgar Publishing, pp. 29-60.

Johnson, Norman L., and Samuel Kotz. 1970. Continuous Univariate Distributions – 2. New York: John Wiley and Sons.

Lahey, Joanna N. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources*, Vol. 43, No. 1, Winter, pp. 30-56.

Lahey, Joanna N., and Ryan A. Beasley. 2009. "Computerizing Audit Studies." *Journal of Economic Behavior & Organization*, Vol. 70, No. 3, June, pp. 508-14.

Lundberg, Shelly J., and Richard Startz. 1983. "Private Discrimination and Social Intervention in Competitive Labor Markets." *American Economic Review*, Vol. 73, No. 3, June, pp. 340-7.

Maddala, G.S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge, U.K.: Cambridge University Press.

Mincy, Ronald. 1993. "The Urban Institute Audit Studies: Their Research and Policy Context." In Fix and Struyk, eds., Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, DC: The Urban Institute Press, pp. 165-86.

Neumark, David. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Quarterly Journal of Economics*, Vol. 111, No. 3, August, pp. 915-41.

Neumark, David. 1999. "Wage Differentials by Race and Sex: The Roles of Taste Discrimination and Labor Market Information." *Industrial Relations*, Vol. 38, No. 3, July, pp. 414-45

Pager, Devah. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The Annals of the American Academy of Political and Social Science*, Vol. 609, No. 1, pp. 104-33.

Riach, Peter A., and Judith Rich. 2002. "Field Experiments of Discrimination in the Market Place." *The Economic Journal*, Vol. 112, No. 483, November, pp. F480-518.

Riach, Peter A., and Judith Rich. 2007. "An Experimental Investigation of Age Discrimination in the Spanish Labor Market." IZA Discussion Paper No. 2654.

Smeeters, B., and A. Nayer. 1998. "La Discrimination a l'Acces a l'Emploi en Raison de l'Origine Etrangere: le Cas de le Belgique." *International Migration Papers*, No. 23, International Labour Office, Geneva, Switzerland.

Turner, Margery, Michael Fix, and Raymond Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring*. UI Report 91-9. The Urban Institute, Washington, DC.

U.S. Equal Employment Opportunity Commission. 1996. Notice 915.002, May 22, http://www.eeoc.gov/policy/docs/testers.html (viewed February 28. 2010).

Williams, Richard. 2009. "Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups." Unpublished manuscript, Department of Sociology, Notre Dame University.

Table 1
Probit Estimates for Callbacks: Basic Results

	Males and females			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Black	-.033	-.030	-.030	-.033	-.030	-.030
	(.006)	(.006)	(.006)	(.008)	(.007)	(.007)
Female	.009	-.001	.001	…	…	…
	(.012)	(.011)	(.011)			
Selected individual resume controls						
Bachelor's degree		.009	.009		.019	.019
		(.009)	(.009)		(.010)	(.010)
Experience $\cdot 10^{-1}$		.080	.076		.080	.076
		(.029)	(.028)		(.034)	(.033)
Experience^2 $\cdot 10^{-2}$		-.022	-.021		-.019	-.018
		(.011)	(.010)		(.013)	(.012)
Academic honors		.039	.040		.026	028
		(.015)	(.015)		(.017)	(.017)
Special skills		.056	.055		.060	.059
		(.009)	(.009)		(.010)	(.010)
Other controls:						
Individual resume characteristics		X	X		X	X
Neighborhood characteristics			X			X
Mean callback rate	.080	.080	.080	.082	.082	.082
N	4,784	4,784	4,784	3,670	3,670	3,670

Notes: Marginal effects using equation (19) are reported. Standard errors are computed clustering on the ad to which the applicants responded, and are reported in parentheses; the delta method is used to compute standard errors for the marginal effects. Individual resume characteristics include bachelor's degree, experience and its square, volunteer activities, military service, having an email address, gaps in employment history, work during school, academic honors, computer skills, and other special skills. Neighborhood characteristics include the fraction high school dropout, college graduate, black, and white, as well as log median household income, in the applicant's zip code.

Table 2
Heteroskedastic Probit Estimates for Callbacks: Full Specifications

	Males and females		Females	
	(1)	(2)	(3)	(4)
A. Estimates from basic probit (Table 1)				
Black	-.030	-.030	-.030	-.030
	(.006)	(.006)	(.007)	(.007)
B. Heteroskedastic probit model				
Black (unbiased estimates)	-.024	-.026	-.026	-.027
	(.007)	(.007)	(.008)	(.008)
Marginal effect of race through level	-.086	-.070	-.072	-.054
	(.038)	(.040)	(.040)	(.040)
Marginal effect of race through variance	.062	.045	.046	.028
	(.042)	(.043)	(.045)	(.044)
Standard deviation of unobservables, black/white	1.37	1.26	1.26	1.15
Wald test statistic, null hypothesis that ratio of standard deviations = 1 (p-value)	.22	.37	.37	.56
Wald test statistic, null hypothesis that ratios of coefficients for whites relative to blacks are constant, fully interactive probit model (p-value)	.62	.42	.17	.35
Test overidentifying restrictions: include in heteroskedastic probit model interactions for variables with |white coefficient| < |black coefficient|, Wald test for joint significance of interactions (p-value)	.83	.33	.34	.56
Number of overidentifying restrictions	3	6	2	6
Other controls:				
Individual resume characteristics	X	X	X	X
Neighborhood characteristics		X		X
N	4,784	4,784	3,670	3,670

Notes: See notes to Table 1. In the first row of Panel B the marginal effects in equation (20) are reported, with the decomposition in equations (20') and (20'') immediately below; the marginal effects are evaluated at sample means. The standard errors for the two components of the marginal effects are computed using the delta method. Test statistics are based on the variance-covariance matrix clustering on the ad to which the applicants responded. Individual resume characteristics also include the variables listed separately in Table 1.

Table 3
Heteroskedastic Probit Estimates for Callbacks: Restricted Specifications Using only Controls with Absolute Value of Estimated Effect Larger for Whites than Blacks in Fully Interactive Probit Model

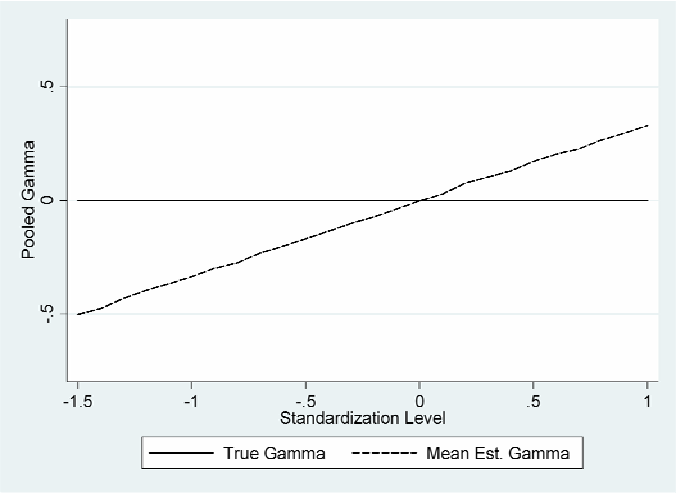
	Males and females		Females	
	(1)	(2)	(3)	(4)
A. Estimates from basic probit				
Black	-.030	-.030	-.030	-.030
	(.006)	(.006)	(.007)	(.006)
B. Heteroskedastic probit model				
Black (unbiased estimates)	-.024	-.025	-.024	-.025
	(.007)	(.007)	(.009)	(.008)
Effect of race through level	-.090	-.080	-.086	-.077
	(.037)	(.036)	(.040)	(.038)
Effect of race through variance	.066	.056	.062	.052
	(.041)	(.039)	(.044)	(.042)
Standard deviation of unobservables, black/white	1.41	1.33	1.37	1.30
Wald test statistic, null hypothesis that ratio of standard deviations = 1 (p-value)	.19	.23	.25	.29
Wald test statistic, null hypothesis that ratios of coefficients for whites relative to blacks are constant, fully interactive probit model (p-value)	.84	.92	.68	.74
Other controls:				
Individual resume characteristics	X	X	X	X
Neighborhood characteristics		X		X
N	4,784	4,784	3,670	3,670

Notes: See notes to Tables 1 and 2.

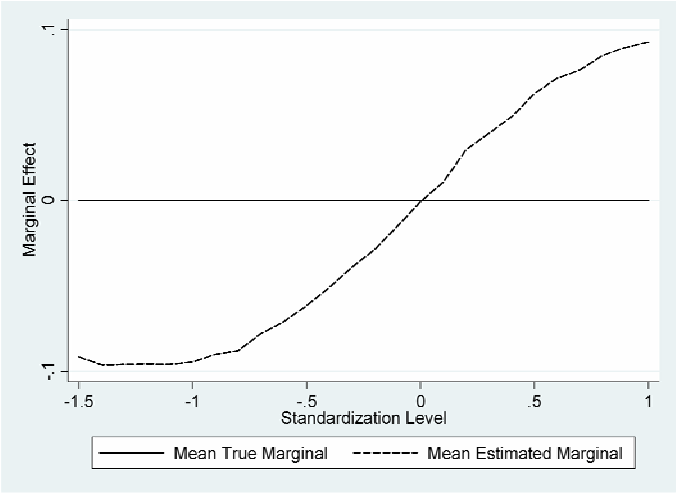
Figure 1
Replication of Heckman (Figure 1, 1998), and Monte Carlo Simulations of Estimates of Marginal Effects from Simple Probit Estimation

No discrimination ($\gamma = 0$)

Probit estimates of γ

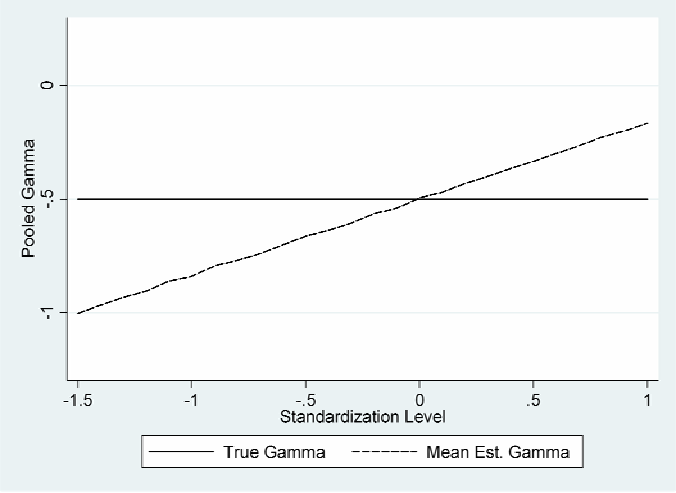


Probit estimates of marginal effect of black

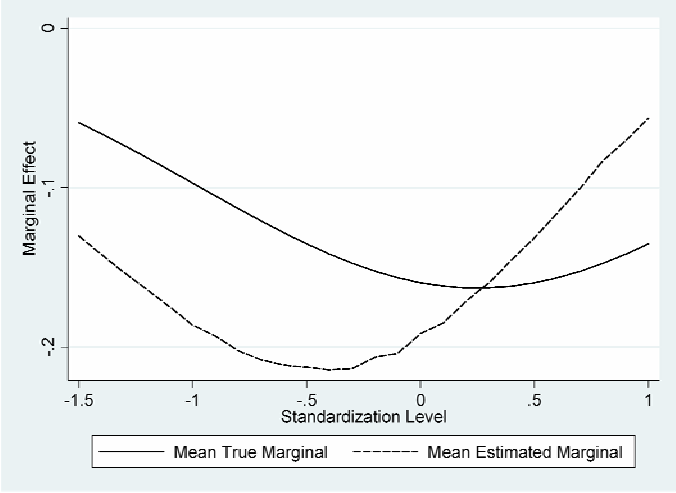


Discrimination ($\gamma = -.5$)

Probit estimates of γ



Probit estimates of marginal effect of black

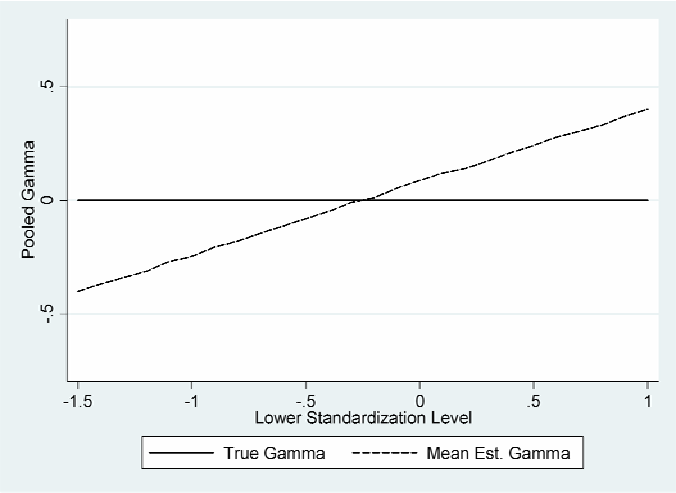


Notes: Left-hand graph shows true γ and mean estimates of γ. Right hand graphs show marginal effects, evaluated at sample means for simulated data. The marginal effects shown correspond to the effect of race on the latent variable, as in equation (20'). In the data generating process, $X^{I*} \sim N(0,1)$, $X_B^{II} \sim N(0,1)$, $X_W^{II} \sim N(0,2.25)$, so $Var(X_W^{II})/Var(X_B^{II}) = 2.25$ (X_W^{II} and X_B^{II} are unobservable); $\beta_I = 1$ and $c = 0$ for both blacks and whites. Estimates are generated by Monte Carlo simulation, drawing 4,000 observations (2,000 white and 2,000 black) from truncated normal distribution at each value of X^{I*} (in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$) and estimating probit model. Simulation is done 100 times at each value of X^{I*}.

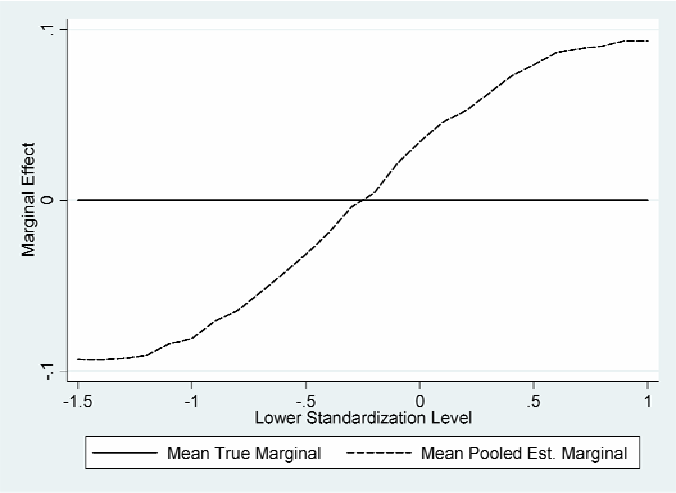
Figure 2
Monte Carlo Simulations of Marginal Effects from Simple Probit Estimation, Bias in Estimate of Discrimination with Two Types of Applicants

No discrimination ($\gamma = 0$)

Probit estimates of γ

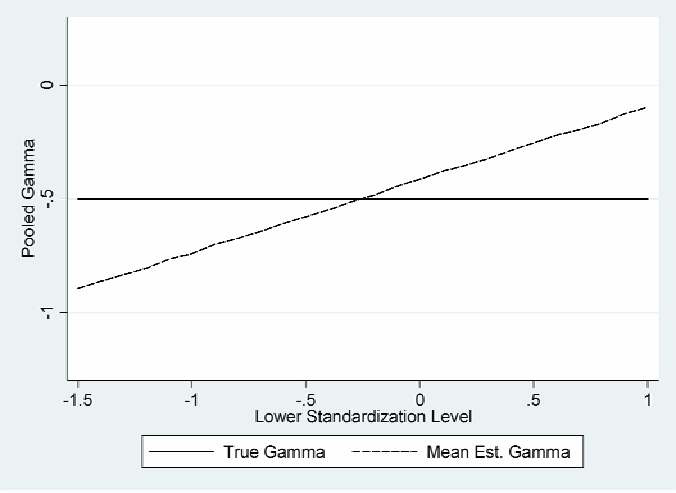


Probit estimates of marginal effect of black

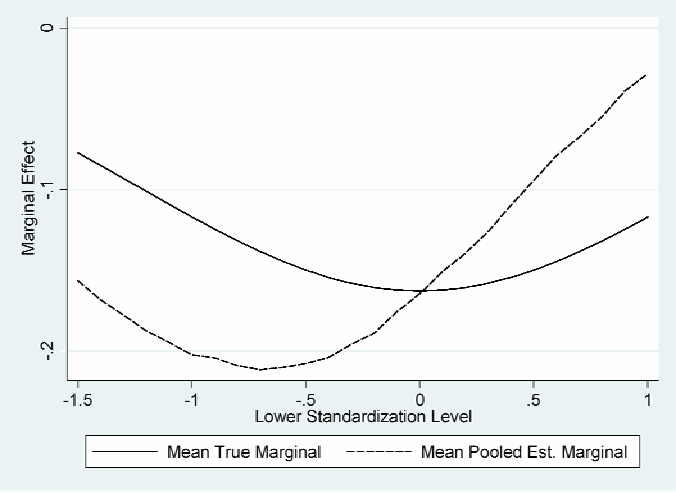


Discrimination ($\gamma = -.5$)

Probit estimates of γ



Probit estimates of marginal effect of black

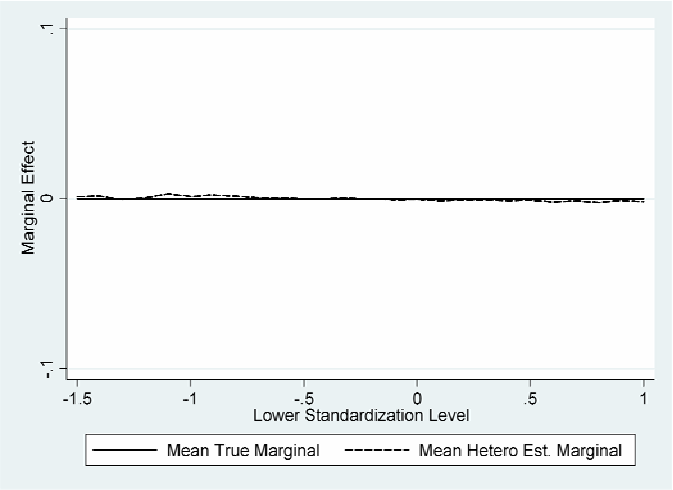


Notes: Left-hand graph shows true γ and mean estimates of γ. Right hand graphs show marginal effects, evaluated at sample means for simulated data. The marginal effects shown correspond to the effect of race on the latent variable, as in equation (20'). In the data generating process, $X^{I*} \sim N(0,1)$, $X_B^{II} \sim N(0,1)$, $X_W^{II} \sim N(0,2.25)$, so $Var(X_W^{II})/Var(X_B^{II}) = 2.25$ (X_W^{II} and X_B^{II} are unobservable); $\beta_I = 1$ and $c = 0$ for both blacks and whites. Estimates are generated by Monte Carlo simulation, drawing 4,000 observations (2,000 white and 2,000 black) observations from two truncated normal distributions (one at each value of X^{I*} (in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$), *and one at each value of* $X^{I*} + .5$ (again in steps of 0.1) $\pm 0.1 \cdot SD(X^{I*})$), and estimating probit model. Simulation is done 100 times at each value of X^{I*}.

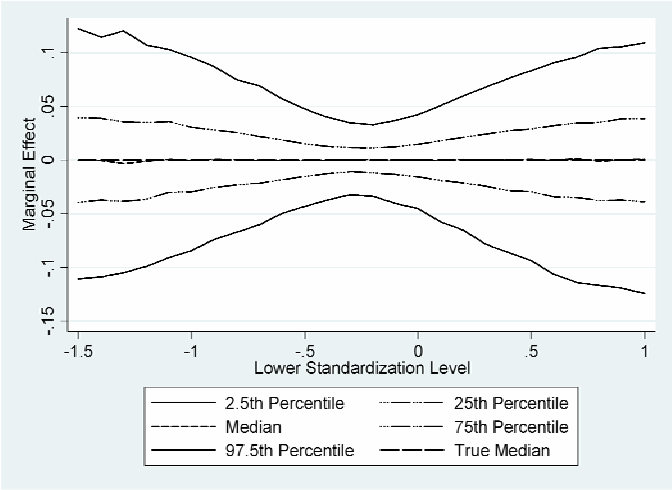
Figure 3
Monte Carlo Simulations of Heteroskedastic Probit Estimation, Estimates of Marginal Effects and Distributions

No discrimination (γ = 0)

Estimates of marginal effects of black

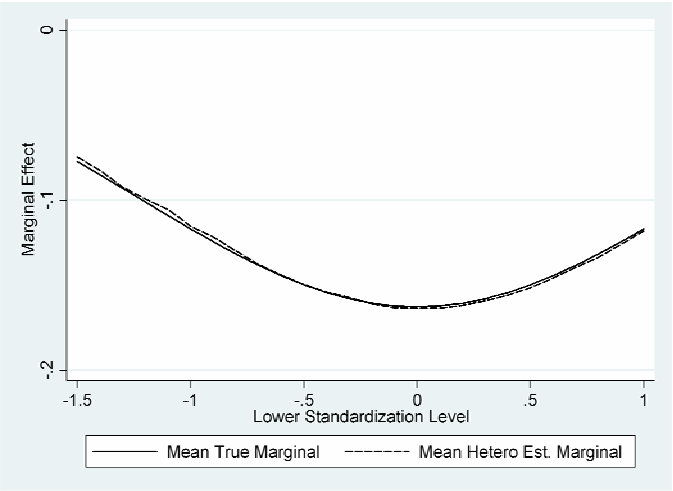


Distribution of estimates

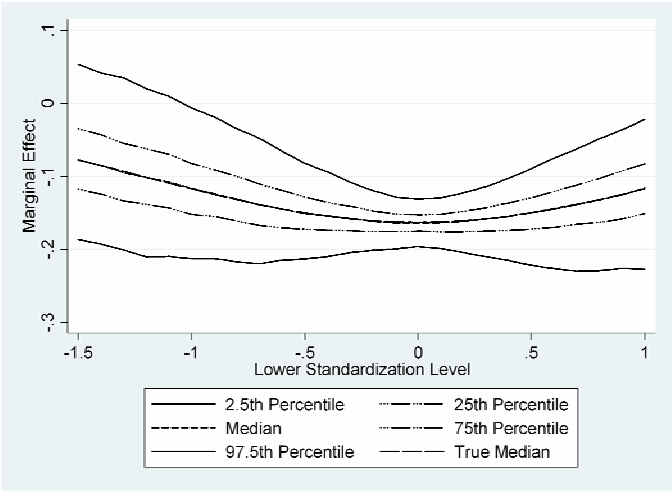


Discrimination (γ = -.5)

Estimates of marginal effects of black



Distribution of estimates

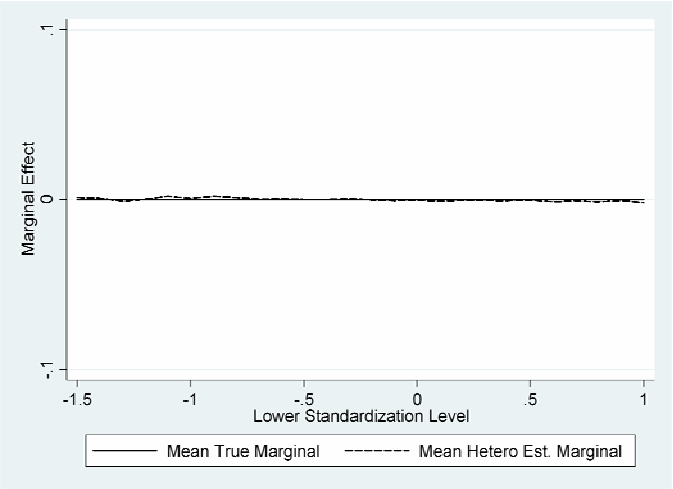


Notes: Left-hand graph shows true marginal effects of being black and marginal effects, evaluate at sample means for simulated data. Right-hand graphs show distributions of estimates. The marginal effects shown correspond to the effect of race on the latent variable, as in equation (20'). Estimates are generated by Monte Carlo simulation, drawing 4,000 observations (2,000 white and 2,000 black) observations from two truncated normal distributions (one at each value of X^{I*} (in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$), and one at each value of $X^{I*} + .5$ (again in steps of $0.1 \pm 0.1 \cdot SD(X^{I*})$), and estimating heteroskedastic probit model. Simulation is done 5,000 times at each value of X^{I*} shown in graph. As in Figure 1, the data generating process has $X^{I*} \sim N(0,1)$, $X_B^{II} \sim N(0,1)$, $X_W^{II} \sim N(0,2.25)$, so $Var(X_W^{II})/Var(X_B^{II}) = 2.25$ (X_W^{II} and X_B^{II} are unobservable); and $\beta_I = 1$ and $c = 0$ for both blacks and whites.

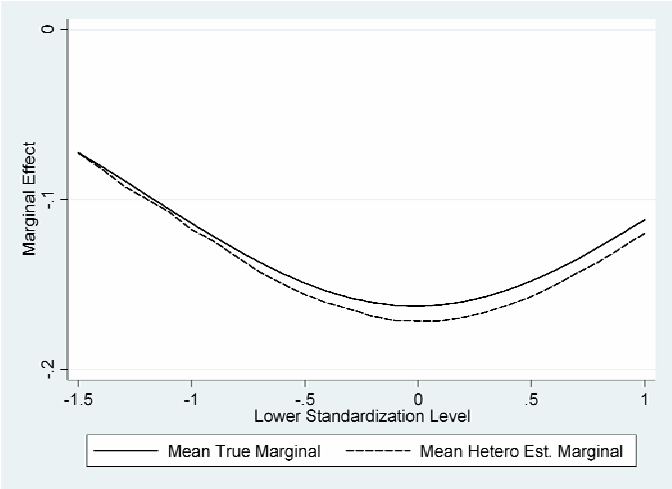
Figure 4
Monte Carlo Simulations of Heteroskedastic Probit Estimation, with Model Misspecification Masking Higher Unobserved Variance for Whites, Estimates of Marginal Effects

Mild violation of identifying assumption in data generating process (β_1 for whites = 1.1)

No discrimination (γ = 0)

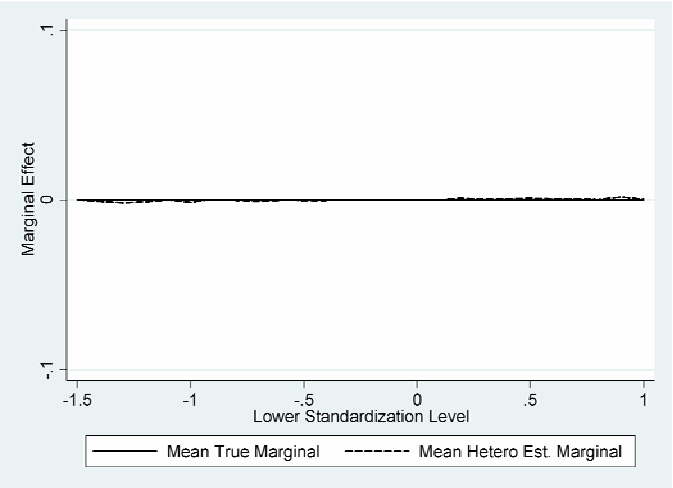


Discrimination (γ = -.5)

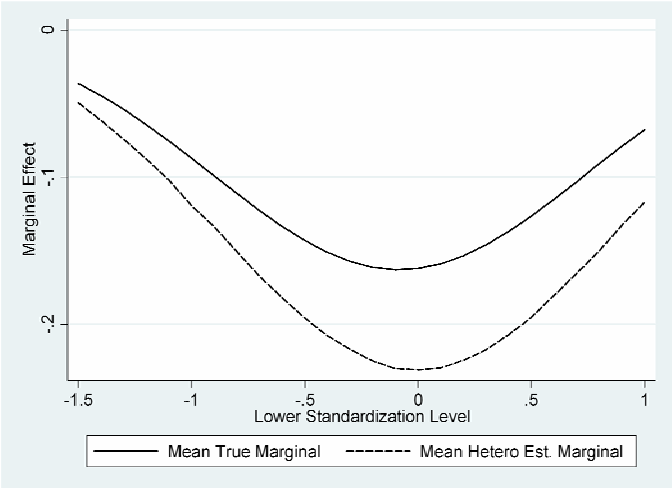


Strong violation of identifying assumption in data generating process (β_1 for whites = 2)

No discrimination (γ = 0)



Discrimination (γ = -.5)



Notes: See notes to Figure 3. The only differences are that β_1 is unequal for blacks and whites; it is always equal to 1 for blacks, and as indicated in the graph headings for whites.