DISCUSSION PAPER SERIES

DP17676

WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH STATISTICAL DISCRIMINATION?

Rahul Deb and Ludovic Renou

LABOUR ECONOMICS AND ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH STATISTICAL DISCRIMINATION?

Rahul Deb and Ludovic Renou

Discussion Paper DP17676
Published 16 November 2022
Submitted 18 October 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Labour Economics
- Organizational Economics

Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Rahul Deb and Ludovic Renou

WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH STATISTICAL DISCRIMINATION?

Abstract

We derive a non-parametric test for statistical discrimination that can be applied to cross-sectional wage data. Specifically, we show that the wage distributions for two groups (with identical observable characteristics) are consistent with a general reduced-form model of statistical discrimination if, and only if, neither wage distribution first-order stochastically dominates the other. Our model allows us to interpret a rejection of this condition as evidence of bias.

JEL Classification: N/A

Keywords: N/A

Rahul Deb - rahul.deb@utoronto.ca
University of Toronto and CEPR

Ludovic Renou - lrenou.econ@gmail.com
Queen Mary University of London and CEPR

Acknowledgements
We are very grateful to Paula Onuchic for engaging in a conversation with us that inspired this paper and to her, Prashant Bharadwaj, Aislinn Bohren, Alex Imas, Rob McMillan and Konstantinos Tatsiramos for detailed discussions. We would also like to thank Juan Eberhard, Namrata Kala, Kory Kroft, Aditya Kuvalekar, Marco Manacorda, Mallesh Pai, Maher Said and numerous seminar participants for insightful comments. We would like to thank the gracious hospitality of the Studienzentrum Gerzensee where part of this research was conducted. Deb thanks the SSHRC for their continued and generous financial support without which this research would not be possible.

WHICH WAGE DISTRIBUTIONS ARE CONSISTENT WITH STATISTICAL DISCRIMINATION?

Rahul Deb [ℚ]
Ludovic Renou [†]

November 16, 2022

ABSTRACT

We derive a non-parametric test for statistical discrimination that can be applied to cross-sectional wage data. Specifically, we show that the wage distributions for two groups (with identical observable characteristics) are consistent with a general reduced-form model of statistical discrimination if, and only if, neither wage distribution first-order stochastically dominates the other. Our model allows us to interpret a rejection of this condition as evidence of bias.

[ℚ] Department of Economics, University of Toronto and CEPR, rahul.deb@utoronto.ca
[†] School of Economics and Finance, Queen Mary University of London and CEPR, lrenou.econ@gmail.com
We are very grateful to Paula Onuchic for engaging in a conversation with us that inspired this paper and to her, Prashant Bharadwaj, Aislinn Bohren, Alex Imas, Rob McMillan and Konstantinos Tatsiramos for detailed discussions. We would also like to thank Juan Eberhard, Namrata Kala, Kory Kroft, Aditya Kuvalekar, Marco Manacorda, Mallesh Pai, Maher Said and numerous seminar participants for insightful comments. We would like to thank the gracious hospitality of the Studienzentrum Gerzensee where part of this research was conducted. Deb thanks the SSHRC for their continued and generous financial support without which this research would not be possible.

1. INTRODUCTION

In recent years, considerable progress have been made towards credibly establishing the *presence* of discrimination, that is, the act of treating individuals differently on the sole basis of the groups they belong, or are perceived to belong, to. Examples of such groups include but are not limited to women and men, blacks and whites, juniors and seniors. Field experiments (of both the audit and correspondence variety) have been particularly instrumental in evidencing discrimination, as they allow the researcher to finely control the observables. However, as Bertrand and Duflo (2017) observe in their survey of the literature: "*while field experiments have been overall successful at documenting that discrimination exists, they have (with a few exceptions) struggled with linking the patterns of discrimination to a specific theory.*" The two most prominent theories of discrimination are *statistical discrimination* (outcomes differ because of differences in information) and *taste-based discrimination* (bias or animus towards one group drives outcome differences); establishing which of these is at work is important both for accountability and to devise corrective policies.

In this paper, we propose a general non-parametric model of statistical discrimination in the labor market, and derive a test for statistical discrimination that only requires cross-sectional data on wages. The model is in the spirit of Phelps (1972). There are two groups whose productivity distributions have identical means, but can otherwise be different.[1] The group identity is observable to employers, but productivities are not. Instead, there are group-dependent "statistical experiments" that generate signals about the underlying productivity. As an example, signals could be the information that employers receive from the job screening process that includes interviews, tests, curricula vitae etc. Signals induce posterior productivity distributions (via Bayes' rule) and, in particular, these can be used to compute posterior estimates (the mean of the productivity conditional on the signal) of the unobserved productivity. Therefore, each group's statistical experiment generates a distribution over posterior productivity estimates. Wages are then determined via a strictly increasing, continuous function of the posterior productivity estimate that, importantly, does not depend on the group. The model is reduced form in that we do not microfound the statistical experiments or the wage function (although foundations can easily be provided), but very general in that both are completely unrestricted (as long as the wage function is strictly increasing and continuous). We say that two wage distributions – one for each of the two groups – are consistent with statistical discrimination if they can be rationalized by this model.

We show that two wage distributions are consistent with statistical discrimination if, and only if, neither wage distribution *first-order stochastically dominates* the other (Theorem 1). In addition, we show that a rejection of this test can only be attributed to taste-based discrimination (Theorem 2). In other words, whenever one wage distribution first-order stochastically dominates another, the only explana-

[1] We discuss below how to accommodate different mean productivities.

tion is taste-based discrimination. We stress that our test exploits the information contained in the entire wage distributions, and not just their averages. This is in sharp contrast with the common practice of reporting wage gaps (differences in average wages), which have been, and continue to be, the subject of much public debate.

First-order stochastic dominance might seem like a demanding condition, but there is abundant evidence that the wages of whites are higher than those of blacks and Hispanics not just in average but also at all quantiles; the latter is an equivalent way of stating first-order stochastic dominance. As an example, see Table 1 (taken from the Economic Policy Institute report "State of Working America Wages 2019" by Elise Gould) where the entries are the shares of the hourly wages earned by black and Hispanic workers relative to white workers at the same quantile of the wage distribution; note that every cell is strictly less than a 100%. Bayer and Charles (2018) also document a similar gap in the wages of blacks and whites at the median and 90th percentile of the wage distributions.

	Quantiles									
Year	10th	20th	30th	40th	50th	60th	70th	80th	90th	95th
Wages of blacks as a share of whites										
2000	93.80%	88.40%	84.80%	82.60%	79.20%	79.30%	77.60%	76.90%	74.70%	72.00%
2007	91.30%	86.40%	84.40%	79.70%	77.70%	75.80%	74.90%	72.80%	74.00%	71.70%
2018	90.80%	83.70%	80.00%	77.30%	73.30%	70.80%	69.20%	70.60%	70.70%	68.20%
2019	91.00%	85.00%	83.70%	82.10%	75.60%	74.00%	71.80%	70.10%	69.20%	65.30%
Wages of Hispanics as a share of whites										
2000	90.20%	81.10%	75.20%	72.70%	69.70%	66.90%	66.60%	66.90%	67.40%	69.00%
2007	90.20%	81.00%	76.80%	72.20%	71.90%	68.40%	66.90%	68.40%	67.50%	68.70%
2018	97.20%	87.80%	80.20%	76.50%	73.10%	69.50%	67.60%	67.10%	67.20%	62.90%
2019	94.10%	90.70%	83.50%	79.60%	74.60%	71.60%	67.50%	65.30%	67.10%	63.10%

Table 1: Shares of US hourly wages for the years 2000–2019 taken from the Economic Policy Institute report "State of Working America Wages 2019" by Elise Gould. These are aggregate nationally representative wage shares computed from the Current Population Survey prepared by the U.S. Census Bureau.

The identical pattern has also been documented for the wages of women and men. See, for instance, Table 2 where entries are the difference between log wages of men and women (positive values imply men's wages are higher). This table is taken from Arulampalam, Booth, and Bryan (2007) (Table 4 in their paper). They analyze data from several European countries and, unlike Table 1, they use quantile regressions with a battery of control variables.[2] Recently, Maasoumi and Wang (2019) also find this pattern in most years in US data (1976-2013), even after correcting for selection into employment. Our main result says that such wage distributions are precisely the type of distributions that cannot arise from statistical discrimination on groups with identical mean productivities. If we believe the latter assumption to be reasonable, this provides suggestive evidence of biased employment practices. Conversely, in the absence

[2]Similar evidence of wage gaps across the distribution in Europe is also found by De la Rica, Dolado, and Llorens (2008) and Christofides, Polycarpou, and Vrachimis (2013).

of first-order stochastic dominance, one cannot rule out that wage gaps, however large, are simply the result of rational unbiased employers responding to the information at their disposal from two groups with identical productivity distributions.[3]

		Quantiles						Quantiles				
Country	Mean	10th	25th	50th	75th	90th	Mean	10th	25th	50th	75th	90th
	Public Sector						**Private Sector**					
Austria	0.227	0.191	0.163	0.191	0.221	0.266	0.214	0.182	0.177	0.188	0.21	0.247
Belgium	0.09	0.046	0.05	0.07	0.109	0.169	0.132	0.1	0.121	0.131	0.148	0.185
Britain	0.134	0.091	0.116	0.135	0.144	0.205	0.19	0.155	0.172	0.188	0.213	0.227
Denmark	0.07	0.058	0.051	0.059	0.086	0.136	0.088	0.032	0.065	0.088	0.123	0.161
Finland	0.216	0.115	0.14	0.203	0.269	0.319	0.151	0.068	0.112	0.154	0.188	0.205
France	0.096	0.092	0.077	0.078	0.108	0.167	0.163	0.146	0.126	0.132	0.152	0.19
Germany	0.122	0.111	0.11	0.118	0.14	0.147	0.143	0.088	0.109	0.137	0.166	0.213
Ireland	0.184	0.186	0.169	0.177	0.165	0.181	0.163	0.081	0.143	0.184	0.195	0.206
Italy	0.097	0.041	0.047	0.081	0.138	0.169	0.173	0.148	0.135	0.152	0.179	0.22
Netherlands	0.121	0.039	0.07	0.112	0.16	0.218	0.131	0.059	0.091	0.123	0.168	0.222
Spain	0.083	0.09	0.079	0.095	0.069	0.076	0.181	0.173	0.178	0.184	0.189	0.176

Table 2: Estimated wage gap (in log wages) between men and women using data from 1995-2001 (Arulampalam et al., 2007). All estimates are statistically significant at the 1% level. Models include dummies for whether training was received in the last year, age, education, tenure, marital status, health status, contracts, private sector firm size, any experience of unemployment since 1989, part-time status, fixed term and casual size, region (where possible), year, industry and occupation.

Our test for statistical discrimination has a number of important and attractive features. First, to the best of our knowledge, this is the first non-parametric test of statistical discrimination in wages. In particular, this implies that we do not require productivities or signal noise to be normally distributed or wages to be an affine function of expected productivity.

Second, the condition of our test (first-order stochastic dominance) can be visualized by simply plotting the cumulative wage distributions (and checking if one distribution lies below the other). We view this feature to be important as it makes our test accessible to non-experts (like administrators, journalists and policy makers) who often incorrectly interpret a wage gap (which can be the result of statistical discrimination) to be evidence of bias.

Third, our test can be taken directly to data (without having to separately estimate wage gaps at different quantiles) because there are well known non-parametric statistical tests of stochastic dominance between two distributions. Recent important econometric developments include (but are not limited to) Barrett and Donald (2003), Linton, Maasoumi, and Whang (2005), Linton, Song, and Whang (2010) and Davidson and Duclos (2013).[4] In fact, the aforementioned paper Maasoumi and Wang (2019) conducts

[3]Part iii of Theorem 1 shows that the testable implications of our model do not change if, instead of identical mean productivities, one assumes that both groups have the same productivity distributions.

[4]Our insight also uncovers a connection between the literature on discrimination and other empirical literatures that apply

precisely such an analysis and concludes that "beyond the early 1990s (except for 2010), men's earnings first-order dominate women's in the majority of the cases to a high degree of statistical confidence." Their aim is to credibly establish this empirical fact, whereas our work is complementary in that we provide a theoretical interpretation of this result as evidence of bias.

The key assumption of our model is that the average productivity is the same in both groups. For this assumption to be plausible, at the very least, we should compare wage distributions that are conditioned on rich control variables (as in Table 2). But of course, this may not be sufficient because there might be unobserved mean productivity differences across groups. This is the focus of the literature on statistical discrimination (following Arrow, 1973) which studies how perceived differences in beliefs (or stereotypes) of employers affect *unobserved* investments in productivity (in equilibrium) and, therefore, outcomes. There are some settings where this might be less of an issue: for instance, it seems reasonable to assume that black lawyers, professors or doctors have the same average productivity as observationally equivalent white people doing those jobs. In other settings, the classic work of Coate and Loury (1993) shows how such stereotypes can lead to less unobserved investment by the disadvantaged group and, therefore, lower average productivity (which in turn justifies the stereotype).

We, therefore, generalize our main result to accommodate for different mean productivities (Theorem 3). Here, we assume that the null hypothesis is a *joint* statement that the wage distributions are generated by statistical discrimination alone *and* the difference in mean productivities is less than an exogenously given bound.[5] In this case, a rejection should be interpreted as either evidence of bias or that the difference in mean productivities is larger than the bound (or both). Of course, if we are convinced that the bound is sufficiently generous, then a rejection is just evidence of bias. Once again, the test takes a simple form: the null is rejected if, and only if, one wage distribution first-order stochastically dominates the other *and* the wage gap is greater than the bound. This shows that the wage gap can be a useful statistic to uncover bias but *only* when wage distributions are ordered by first-order stochastic dominance.

While we frame our model in the context of the labor market, it can be applied directly or adapted to analyze other contexts such as housing and financial markets, policing or the criminal justice system. Consequently, we view the general reduced form framework we propose to be one of our main conceptual innovations. We demonstrate this flexibility in two ways. First, in Section 4.1, we examine how our methodology can be applied to settings where, unlike wages, outcomes are binary. Binary outcomes, such as whether or not a job candidate is invited for an interview, are typical in audit and correspondence studies. Echoing Heckman and Siegelman (1993), we demonstrate (in Theorem 4) that it is practically

tests of stochastic dominance (of first and higher orders). Examples are the literature that compares income distributions to infer whether poverty, inequality, or social welfare is greater in one distribution than in another (see Anderson, 1996, Davidson and Duclos, 2000) and the literature on efficient portfolio choice (see Post, 2003, Kuosmanen, 2004).

[5] The bound is expressed in the following normalized units of productivity: a one unit increase in productivity can lead to at most a one dollar increase in wages.

impossible to distinguish between statistical and taste-based discrimination with binary outcomes. Finally, we adapt our framework (Theorem 5) to settings with richer data where outcome tests à la Becker (1957, 1993) are typically employed.

RELATION TO THE LITERATURE

As alluded to above, our test for statistical discrimination leverages the fact that the outcome variable we study—the wage—is *not binary* and we show that a similar test has no empirical bite for binary outcome variables. This result is related to a critique of audit and correspondence studies first made in Heckman and Siegelman (1993) and recently revisited by Neumark (2012). For intuition, consider correspondence studies that send fictitious curricula vitae to employers and measure whether or not the candidate gets invited for an interview; a difference in the call back rates by group status is interpreted as evidence of discrimination. Now suppose that the employer believes that the two groups have the same mean productivity but that the variance of the advantaged group is higher (a feature that our model allows). If employers only call back for interviews those candidates whose productivities they think are above a certain threshold, the differential variance can lead to higher call back rate for the advantaged group. Of course, this could also be the result of taste-based discrimination, but this cannot be differentiated using this binary outcome.

Partly motivated by this difficulty, tests for statistical discrimination in the literature are developed for settings in which the researcher has access to richer data. A classic example is Altonji and Pierret (2001) who test for statistical discrimination in wages on the basis of race. They develop a parametric model which requires panel data and their test is based on the assumption that the researcher has access to information about workers (their AFQT scores) that employers do not. Their insight is that, in a wage regression, the coefficients on variables that employers observe should fall over time but the coefficients on variables that they do not (but the econometrician does) should rise as they learn about the worker. More recently, there is a nascent experimental literature that exploits dynamics (see, for instance Bohren, Imas, and Rosenberg, 2019b) to tease out the sources of discrimination. The key observation is that dynamics help because beliefs respond to information, whereas preferences do not.

We have already briefly mentioned the alternate strand of the literature that develops outcome tests (in the spirit of Becker, 1957, 1993). Papers in this strand consider settings where the researcher has access not just to the decision (whether or not a loan is granted, a driver is searched by a police officer, etc) but also the post-decision result (whether or not the loan is repaid, contraband is found on the driver, etc). Analogous data in our setting would correspond to the researcher observing the productivity of the worker in addition to their wage. The key insight is that even though the rates at which decisions are made may differ due to group differences, the post-decision results of the *marginal* case should be the same if the decision maker is unbiased. This requires devising empirical strategies to identify the post-

decision results of marginal cases or models that provide a systematic relationship between the average and marginal post-decision result.[6] A strength of our framework is that it allows us to derive an outcome test for statistical discrimination which does not require having to identify the marginal case.

We reiterate that, relative to the above mentioned papers, our main insights are that (i) statistical discrimination can be tested on cross-sectional wage data, (ii) our model allows us to to interpret a rejection of our test as evidence of taste-based discrimination and (iii) our test is easy to state and implement so can be used by non-experts.

Before proceeding to our model, it is worth acknowledging that, in addition to the papers already cited, there are large insightful literatures in economics, psychology and sociology studying discrimination and we will not attempt to provide a comprehensive description here. Instead, we refer the reader to several excellent recent surveys in economics—Fang and Moro (2011), Lang and Lehmann (2012), Bertrand and Duflo (2017), Lang and Spitzer (2020), Onuchic (2022)—that cover both the theory and the empirical evidence in a variety of different settings.

2. THE MODEL

To streamline the exposition, we present the model in the context of discrimination in the labor market. However, as mentioned in the introduction, other applications, such as discrimination in policing or in the justice system, also fit our model.

There are two groups—1 and 2—of workers; examples include female and male, black and white, junior and senior, or disabled and able bodied. We do not take a stand on which of these two groups is advantaged/disadvantaged, if any. We observe two wage distributions G_1 and G_2, with $G_i(w) \in [0,1]$ being the fraction of workers in group $i \in \{1,2\}$ who are paid a (hourly) wage of $w \geq 0$ or less.[7] We assume that the wage distributions are bounded, i.e., $G_i(\overline{w}) = 1$ for some $\overline{w} > 0$, $i = 1, 2$.

The question we address is: when are the observed wage distributions rationalized by a reduced-form model of statistical discrimination? The model is simple, non-parametric, and general. In a nutshell, the model assumes that workers' productivity distributions differ, but that there are no significant differences between the two groups; that is, the average productivity is the same in both groups. Employers do not perfectly observe the productivity of workers. Instead, they acquire some information (for instance, through tests, interviews, or referrals), and then pay workers accordingly. We only require that, the higher the expected productivity, the higher the wage.

[6] See, for instance, Knowles, Persico, and Todd (2001), Anwar and Fang (2006), Arnold, Dobbie, and Yang (2018) and Canay, Mogstad, and Mountjoy (2020).
[7] Throughout, all distributions are right-continuous and have limits on the left.

We stress that the only source of discrimination is information. Hiring tools such as personality and aptitude tests or algorithmic resume screeners are all examples of techniques, which may advantage one group over another in signaling their productivity.[8]

We now present the model in detail, starting with the productivity distributions.

Productivity distributions: Workers differ in their productivities, with $\theta_i \in [0, \overline{\theta}] =: \Theta$ denoting the productivity of a worker in group $i \in \{1, 2\}$, and H_i its (cumulative) distribution. We assume that $\int_0^{\overline{\theta}} \theta_1 dH_1(\theta_1) = \int_0^{\overline{\theta}} \theta_2 dH_2(\theta_2)$ or, in words, that the average productivity is the same in the two groups. A case of particular interest is when the two distributions are identical, i.e., $H_1 = H_2$. As we will show, there are no testable differences between a model that assumes identical distributions and another that assumes different distributions, but with identical means. We make no additional restrictions, so that we can accommodate discrete distributions, continuous distributions, or mixtures of the two.

Information: Employers do not directly observe the productivity of workers, but receive informative signals (from curricula vitae, reference letters, interviews, tests etc.). Employers then form an expectation of the productivity of workers and pay them accordingly: wages are strictly increasing in expected productivity. Since wages only depend on the expected productivity, it is without loss to restrict attention to unbiased statistical experiments.

An *unbiased statistical experiment* (S_i, π_i) for group $i \in \{1, 2\}$ consists of a set of *signals* $S_i = \Theta$ and a joint distribution π_i over $\Theta \times S_i$, whose marginal distribution over Θ is H_i. We denote the marginal distribution of π_i over S_i by F_i. Moreover, to reflect the "unbiased" terminology, we require that the *posterior estimate* $\mathbb{E}_{\pi_i}[\theta_i \mid s_i]$ of the productivity satisfy

$$s_i = \mathbb{E}_{\pi_i}[\theta_i \mid s_i],$$

for all s_i in the support of F_i; that is, s_i is an unbiased estimate of the true productivity θ_i. This is without loss of generality, as we can always relabel signals to guarantee that they are unbiased in the above sense. Accordingly, we will write θ_i for the posterior estimate (the signal) in what follows.

It is well known that F_i is a distribution of posterior estimates arising from some statistical experiment if, and only if, the prior distribution H_i is a *mean-preserving spread* of the posterior distribution F_i, which we denote by $H_i \succcurlyeq_2 F_i$. Formally, the mean-preserving spread condition requires that

$$\int_0^{\theta} H_i(\theta_i) d\theta_i \geq \int_0^{\theta} F_i(\theta_i) d\theta_i \text{ for all } \theta \in [0, \overline{\theta}], \text{ with equality at } \theta = \overline{\theta}.$$

[8] In the context of college admission, academic tests, such as SAT and ACT, have been found to discriminate against low-income, minority and female students. See https://www.forbes.com/sites/markkantrowitz/2021/05/21/how-admissions-tests-discriminate-against-low-income-and-minority-student-admissions-at-selective-colleges/.

Note that the requirement of equality at $\theta_i = \overline{\theta}$ is the same as ensuring that H_i and F_i have the same mean.[9] If this inequality is strict for any $\theta \in (0, \overline{\theta})$, we say H_i is a *strict mean-preserving spread* of F_i, which we denote by $H_i \succ_2 F_i$.

We stress that the above formulation subsumes all possible signaling technologies. In particular, this includes the standard way (as in Phelps, 1972; Aigner and Cain, 1977) of modeling (biased) signals as $s_i = \theta_i + \varepsilon_i$, where ε_i is a noise term whose distribution can depend on the productivity θ_i.

Wage function: If an employer estimates the productivity of a worker to be θ, the employer pays the worker $W(\theta)$, where the *wage function* $W : [0, \overline{\theta}] \to \mathbb{R}_+$ is continuous and strictly increasing. Observe that this wage function does not depend on the group identity and, in this sense, there is no bias. In addition, we normalize the wage function to be Lipschitz with constant 1; that is, we assume

$$|W(\theta) - W(\theta')| \leq |\theta - \theta'| \qquad \text{for all } \theta, \theta'.$$

This restriction implies that the slope of the wage function (wherever differentiable) satisfies $W'(\theta) \leq 1$ for all θ.[10] This assumption affects none of our results (relative to a wage function that is just continuous and strictly increasing), but allows us to interpret the productivity parameter in terms of dollars: a one unit increase in productivity has at most a value of one dollar in terms of wages.[11] We maintain the assumptions of strictly increasing and 1-Lipschitz continuous wage functions throughout the paper.

Induced wage distributions: The distribution F_i over posterior estimates induces the *wage distribution* G_i via the wage function W. Formally, for both $i \in \{1, 2\}$, $G_i(w)$ is the measure of the set $\{\theta : W(\theta) \leq w\}$ according to F_i, that is, $G_i(w) = F_i(W^{-1}(w))$ for $w \in [W(0), W(\overline{\theta})]$, $G_i(w) = 0$ for $w < W(0)$ and $G_i(w) = 1$ for $w > W(\overline{\theta})$.[12] Note that, even though the wage function does not depend on group identity, the wage distributions G_1 and G_2 may differ across groups because the distributions of posterior estimates F_1 and F_2 may differ. Moreover, because W is an arbitrary increasing function, G_1 and G_2 may not have the same mean. In other words, the model is consistent with the existence of a wage gap between the two groups.

Consistency with statistical discrimination: We say that the observed wage distributions G_1 and G_2 are *consistent with statistical discrimination* if there exist productivity distributions H_i that have identi-

[9] Integration by parts implies that the mean satisfies $\int_0^{\overline{\theta}} \theta_i dF_i(\theta_i) = \theta_i F_i(\theta_i)|_0^{\overline{\theta}} - \int_0^{\overline{\theta}} F_i(\theta_i) d\theta_i = \overline{\theta} - \int_0^{\overline{\theta}} F_i(\theta_i) d\theta_i$.

[10] Assuming Lipschitz continuity, as opposed to differentiability with a bounded slope, has the advantage that it allows us to model piecewise linear wage functions. The latter is a natural way to model how productivities affect wages across different ranks or professions.

[11] Apart from the interpretation of the productivity parameter, nothing changes if we assume W is Lipschitz with arbitrary constant $k > 0$.

[12] We define W^{-1} as the inverse of W on the domain $[W(0), W(\overline{\theta})]$. None of our results depend on the continuity of W. It would be enough to consider left-continuous and strictly increasing wage functions with generalized inverse $\sup\{\theta : W(\theta) \leq w\}$ at w.

cal means, distributions of posterior estimates F_i that satisfy $H_i \succcurlyeq_2 F_i$ for $i \in \{1, 2\}$, and a strictly increasing, 1-Lipschitz wage function W, such that these jointly induce the observed wage distributions. Put differently, wage distributions G_1 and G_2 are consistent with statistical discrimination if they can be rationalized by our model.

Before stating our main result, we comment on the model. While our model is general in the sense that we allow any statistical experiments and wage functions, we make two key assumptions: (i) the productivity distributions for both groups have identical means and (ii) the wages are a function of the posterior estimate alone.

In the introduction, we discussed the first of these. To reiterate, we are implicitly assuming that the wage distributions G_1 and G_2 are estimated controlling for enough observables (and/or with additional corrections for selection) to make identical mean productivity a reasonable assumption. In Section 3.3, we relax this assumption and devise a test for statistical discrimination that allows mean productivities to differ. We focus on the case of identical mean productivities because this is the simplest demonstration of our theoretical insight and it yields a test with a clear interpretation that is simple enough for non-experts to employ.

We end this section with a brief discussion of the second assumption. Our model is in the spirit of Phelps (1972). Phelps considers two populations, whose productivities are drawn from a normal distribution. Signals are also normally distributed, differ across groups, and the wage function is linear in the posterior mean. If the means of the productivity distributions for both groups are the same, then the Phelps' model implies that the average wage for both groups is the same (because the posterior distribution must have the same mean as the prior, and the wage function is linear). In this case, there is no discrimination at the group level even though the wage distributions differ (so there is individual level discrimination). Aigner and Cain (1977) observe it is possible to generate discrimination at the group level via more general wage functions even when the productivity distributions for both groups are identical. In their model, wages depend both on the mean and the variance of the posterior belief. In the normal learning environment, the variance is the same for all signal realizations so they model the wage as just the difference between the posterior mean and some multiple of the (signal independent) variance of the posterior belief. Hence, different normally distributed signals can generate distinct mean wages.

We could, in principle, further generalize our wage function by allowing it to depend on not only the mean but also higher moments of the posterior productivity distribution. We have chosen not to do so because we felt that our model strikes the right balance between generality and having testable implications that can plausibly be rejected in wage data (as in Tables 1 and 2).

3. TESTING FOR DISCRIMINATION

This section is organized as follows. In Section 3.1, we present our main result (Theorem 1) that characterizes wage distributions that are consistent with statistical discrimination. In Section 3.2, we then show (in Theorem 2) that taste-based discrimination can also be modeled within our framework. The key observation of that sub-section is that every pair of wage distributions are consistent with taste-based discrimination. This observation is important because it allows us to interpret a rejection of the test for consistency with statistical discrimination as evidence of bias. Finally, in Section 3.3, we generalize our test (Theorem 3) to allow for different mean productivities.

3.1. STATISTICAL DISCRIMINATION

Given the generality of our model, the first natural question to ask is: are there *any* wage distributions that are *not* consistent with statistical discrimination? To this point, note that our model allows the posterior estimate distribution of group 1 to be a strict mean-preserving spread of group 2 (or $F_1 \succ_2 F_2$ in our notation), in which case a strictly convex wage function W will generate higher mean wages for group 1. In other words, differences in mean wages (a wage gap) can arise purely via statistical discrimination, even though the productivity distributions have the same mean. So, to find inconsistent distributions, we need to consider higher moments. In fact, as we now argue, we need to consider *all* moments.

The wage distribution G_i *strictly first-order stochastically dominates* the wage distribution G_j, which we denote $G_i \succ_1 G_j$, if $G_i(w) \leq G_j(w)$ for all $w \in \mathbb{R}_+$, with the inequality strict for some w.

Suppose that the wage distribution of group i strictly first-order stochastically dominates that of group j. We now argue that these distributions are *not* consistent with statistical discrimination. For contradiction, assume that these distributions are consistent. This implies that there exist posterior estimate distributions F_i and F_j, and a wage function W, such that

$$F_i(\theta) = G_i(W(\theta)) \leq G_j(W(\theta)) = F_j(\theta) \qquad \text{for all } \theta \in [0, \overline{\theta}],$$

with the inequality strict for some θ, that is, $F_i \succ_1 F_j$. It follows that F_i has a strictly higher mean than F_j, which is a contradiction since F_i and F_j are *mean-preserving* contractions of some productivity distributions H_i and H_j, which both have the same mean.

The above argument shows that a necessary condition for a pair of wage distributions to be consistent with statistical discrimination is that neither strictly first-order stochastically dominates the other. Our main result shows that this condition is also sufficient. In fact, we show a stronger result.

First, we show that if the wage distributions are consistent with statistical discrimination, then they are consistent with statistical discrimination and identical productivity distributions, that is, $H_1 = H_2$.

Second, we show that statistical discrimination and the absence of discrimination – be it statistical or taste-based – have the same testable implications. To state this equivalence, we need to define the latter.

Two wage distributions G_1 and G_2 are *consistent with the absence of discrimination* if there exist productivity distributions H_1 and H_2 that have identical means and a strictly increasing, 1-Lipschtiz wage function W such that these jointly yield the observed wage distributions; i.e., $G_i(w) = H_i(W^{-1}(w))$ for all $w \in [W(0), W(\overline{\theta})]$, $G_i(w) = 0$ for all $w < W(0)$ and $G_i(w) = 1$ for all $w > W(\overline{\theta})$.

We define discrimination to be absent when the employers perfectly observe productivities and determine wages without factoring in group identity. Note that it would be equally natural to assume that employers do not perfectly observe productivities, but learn about them via a single group-independent statistical experiment π. As we now demonstrate, such generality is unnecessary.

THEOREM 1. *The following statements are equivalent.*

(i) The wage distributions G_1 and G_2 are consistent with statistical discrimination.

(ii) Neither G_1 nor G_2 strictly first-order stochastically dominates the other.

(iii) The wage distributions G_1 and G_2 are consistent with statistical discrimination and identical productivity distributions.

(iv) The wage distributions G_1 and G_2 are consistent with the absence of discrimination.

Before presenting a sketch of the proof, it is worth discussing a few implications of this result. It has been argued that, for the distributions of certain traits, men and women have the same mean, but the former have a higher variance. This is sometimes referred to as the "variability hypothesis." The third statement of Theorem 1 implies that any two wage distributions that are not ordered by strict first-order stochastic dominance, no matter how different, could have resulted from statistical discrimination on *identical* populations. In other words, allowing for different variances of the productivity distributions leads to no additional explanatory power of statistical discrimination.

Conversely, the fourth statement says that, when condition (ii) holds, we cannot conclude that discrimination of any form is present. In other words, it is possible that the differences in wages arise from statistical discrimination on identical populations or, simply, from heterogeneous populations (with identical mean productivities) in the absence of discrimination. This result therefore shows that, irrespective of the size of the wage gap, one cannot conclude that there is a cause for concern. However, when condition (ii) is rejected in the data, this not only shows discrimination is present, it says that discrimination cannot be statistical alone! As we will show in the next section (see Theorem 2), we can always interpret this as evidence of taste-based discrimination (possibly, in addition, to statistical discrimination) since every pair

of wage distributions is consistent with taste-based discrimination.

Finally, note that the above result does not require employers to have *accurate* beliefs about either the productivity distributions or about the signal. Employers may believe that the productivity distribution of the advantaged group has higher variance or that the signals for this group are more accurate. The only assumption we require is that these beliefs (whether accurate or inaccurate) satisfy the assumption that they assign the same mean productivity to both groups. Thus, a rejection of the test in Theorem 1 is robust evidence of bias in that it rules out statistical discrimination arising from either accurate or inaccurate information that differs across groups (this feature differentiates our setting from that Bohren, Haggag, Imas, and Pope, 2019a). Of course, employers may have inaccurate beliefs that assign a lower mean productivity to the disadvantaged group. We interpret this to be taste-based discrimination.

We now sketch the proof of the statement $[(ii) \implies (iii)]$, with the help of a simple example. This proof sketch will also provide intuition for why $[(ii) \implies (iv)]$. (Recall that we have already argued that $[(i) \implies (ii)]$, and both $[(iii) \implies (i)]$ and $[(iv) \implies (i)]$ are trivially true.) In Table 3, we have two wage distributions G_1 and G_2, with neither first-order stochastically dominating the other (since $G_1(\$10) > G_2(\$10)$, while $G_1(\$15) < G_2(\$15)$).

Table 3: Sketch of proof

wage/hour	\$10	\$15	\$20
G_1	$1/3$	$5/12$	1
G_2	$1/6$	$1/2$	1

The idea of the proof is to construct a wage function W such that the two distributions F_1 and F_2, defined by $F_i(\theta) := G_i(W(\theta))$ for all θ, $i \in \{1, 2\}$, have the same mean. This also proves $[(ii) \implies (iv)]$ since we can choose the prior productivity distributions H_i to be equal to F_i, thus making the wage distributions consistent with the absence of discrimination.

In this simple example, it suffices to find three points $0 \leq \theta^1 < \theta^2 < \theta^3 \leq 1$ (on which F_1 and F_2 are supported) such that $W(\theta^1) = \$10$, $W(\theta^2) = \$15$, $W(\theta^3) = \$20$, and

$$\theta^1 \underbrace{\left(\frac{1}{3} - \frac{1}{6}\right)}_{>0} + \theta^2 \underbrace{\left(\frac{1}{12} - \frac{1}{3}\right)}_{<0} + \theta^3 \underbrace{\left(\frac{7}{12} - \frac{1}{2}\right)}_{>0} = 0,$$

which, in words, ensures that F_1 and F_2 have the same mean. The wage function W is then the piecewise linear interpolation of these three points on the domain $[0, 1]$. A solution is $\theta^1 = 0$, $\theta^2 = 1/3$ and $\theta^3 = 1$. Notice that a solution exists precisely because neither G_1 nor G_2 strictly first-order dominates the other, which is reflected in the alternating signs in the above expression.

To complete the argument, we need to construct a distribution H such that $H \succcurlyeq_2 F_i, i \in \{1,2\}$. We now argue that the distribution H, defined by $H(\theta) = 7/18$ for all $\theta < 1$ and $H(1) = 1$, is one such distribution. Note that H is supported on $\{0,1\}$ with a probability of $11/18$ on $[\theta = 1]$. The mean of H is $11/18$, as is the mean of F_1 and F_2. Moreover,

$$\int_0^\theta H(x)dx - \int_0^\theta F_1(x)dx = \begin{cases} \left(\frac{7}{18} - \frac{1}{3}\right)\theta & \text{if } 0 \le \theta \le 1/3 \\ \left(\frac{7}{18} - \frac{1}{3}\right)\frac{1}{3} + \left(\frac{7}{18} - \frac{5}{12}\right)\left(\theta - \frac{1}{3}\right) & \text{if } 1/3 < \theta \le 1 \end{cases}.$$

It is easy to check that this is indeed positive for all θ, hence $H \succcurlyeq_2 F_1$. Intuitively, to generate F_1 from H, we construct an experiment with three signals, which generate the posterior beliefs about the event $[\theta = 1]$ of 0, $1/3$ and 1 with probability $1/3$, $1/12$, $7/12$, respectively. Since the expectation of the posterior beliefs is equal to the prior belief ($11/18$), such a construction is possible. A similar argument shows that $H \succcurlyeq_2 F_2$.

Finally, observe that since the wage function W, constructed above, is piece-wise linear with finitely many pieces, it is Lipschitz. If its modulus of continuity is larger than one, we can "stretch" the wage function and rescale the distributions, accordingly, to guarantee a modulus of continuity of one. Formally, let $\overline{\theta} > 1$ and construct productivity distributions and distributions of posterior estimates: $\widetilde{H}_i(\theta) := H_i(\theta/\overline{\theta})$ and $\widetilde{F}_i(\theta) := F_i(\theta/\overline{\theta})$ for all $\theta \in [0, \overline{\theta}]$, for all $i \in \{1,2\}$. Similarly, construct the wage function $\widetilde{W}(\theta) := W(\theta/\overline{\theta})$ for all $\theta \in [0, \overline{\theta}]$. It is routine to verify that these newly constructed model primitives also rationalize the wage distributions G_1 and G_2, but now the slope of the piece-wise linear wage function $\widetilde{W}$ is smaller – we have rescaled the slope of W by a factor $1/\overline{\theta}$. By choosing $\overline{\theta}$ appropriately, we guarantee a modulus of continuity of one.

While the general construction for arbitrary distributions G_1 and G_2 is more elaborate, the same idea works. We first construct a piece-linear function W such that $F_1 := G_1 \circ W$ has the same mean as $F_2 := G_2 \circ W$ by "transporting mass" from the region $\{w : G_1(w) \ge G_2(w)\}$ to the region $\{w : G_1(w) < G_2(w)\}$. We then construct a common distribution H as the right derivative of the convex function

$$\theta \mapsto \max\left(\int_0^\theta F_1(x)dx, \int_0^\theta F_2(x)dx\right).$$

The proof is in Appendix A.

3.2. TASTE-BASED DISCRIMINATION

In this section, we model taste-based discrimination within our framework. We begin with the definition.

Wage distributions G_1 and G_2 are *consistent with taste-based discrimination* if there exist productivity distributions H_1 and H_2 that have identical means and strictly increasing, 1-Lipschitz wage functions

W_1 and W_2 such that these jointly yield the observed wage distributions; i.e., $G_i(w) = H_i(W_i^{-1}(w))$ for all $w \in [W_i(0), W_i(\overline{\theta})]$, $G_i(w) = 0$ for all $w < W_i(0)$ and $G_i(w) = 1$ for all $w > W_i(\overline{\theta})$.

Note the differences of this notion with that of statistical discrimination. We have now removed the noisy signal (the source of statistical discrimination) and instead discrimination is directly introduced via the different wage functions. Discrimination is "taste-based" because two workers from different groups with the same expected productivity can be offered different wages.

THEOREM 2. *Every pair of wage distributions G_1 and G_2 is consistent with taste-based discrimination.*

Theorem 2 shows that, if we allow productivity distributions of both groups to differ (while maintaining the assumption of identical means), then all wage distributions could be the result of taste-based discrimination. This is unsurprising since allowing for different wage functions in addition to distinct productivity distributions introduces a lot of freedom into the model. Importantly, this implies that a rejection of the first-order stochastic dominance condition of Theorem 1 can be in fact interpreted as evidence of bias or animus. This interpretation would not always be correct if there existed wage distributions that were consistent with neither statistical nor taste-based discrimination.

Lastly, Theorem 2 also implies that all wage distributions can be explained with a combination of statistical and taste-based discrimination even if the productivity distributions are *identical*. This is because the statistical experiments can first introduce heterogeneity into the posterior estimate distributions and then we can apply Theorem 2.

3.3. STATISTICAL DISCRIMINATION WITH DIFFERENT MEAN PRODUCTIVITIES

As we highlighted earlier, the key assumption driving our main result (Theorem 1) is that mean productivities of both groups are the same. As we discussed in the introduction, for certain categories of workers, this is an eminently reasonable assumption around which to base a study on wage discrimination. But of course, this may be less true in other settings.

We say that wage distributions G_1 and G_2 are *consistent with statistical discrimination and productivity difference* $d \geq 0$ if there exist productivity distributions H_i whose means differ by less than d (i.e., $|\mathbb{E}_{H_1}[\theta] - \mathbb{E}_{H_2}[\theta]| \leq d$), distributions of posterior estimates F_i that satisfy $H_i \succcurlyeq_2 F_i$ for $i \in \{1, 2\}$, and a wage function W, such that these jointly induce the observed wage distributions.

The next result generalizes Theorem 1 (which is the special case of $d = 0$) and develops a test for statistical discrimination that allows for different mean productivities.

THEOREM 3. *Wage distributions G_1 and G_2 are consistent with statistical discrimination and productiv-*

ity difference d *if, and only if, either*

1. *the wage gap is less than* d*, i.e.* $|\mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w]| \leq d$*, or*

2. *neither* G_1 *nor* G_2 *strictly first-order stochastically dominates the other.*

A few comments about this result are worth making. First, as with Theorem 1, the condition in Theorem 3 is also a necessary and sufficient condition for two distributions to be consistent with the absence of discrimination when mean productivities can differ by at most d.[13] Second, a statistical test of the conditions in the above result can be implemented using standard methods. Third, since this notion of consistency is a joint hypothesis of both the nature of discrimination and the difference in mean productivities, a rejection of the conditions in Theorem 3 must be interpreted accordingly. So, in other words, a rejection is not automatically evidence of bias as it is possible that mean productivities of both groups differ by more than d. But, as with Theorem 1, if we are willing to take a stand on the extent of mean productivity differences, then a rejection indicates bias. To summarize, the wage gap can be a useful statistic to test for the presence of bias, but *only* when the wage distributions are ordered by first-order stochastic dominance. When they are not, a wage gap irrespective of how large is not even conclusive evidence that discrimination of any variety is occurring.

Finally, one can interpret the result in Theorem 3 in a different way. Suppose the researcher is convinced that the setting they are studying is one where only statistical discrimination is at work and bias is absent. Then, the wage gap provides the smallest difference in the mean productivities required to explain two wage distributions ordered by first-order stochastic dominance; otherwise this smallest difference is zero. In other words, the wage gap is a useful statistic to measure differences in productivities, if we are convinced that there is no bias.

We end this section by noting that we do not need to do a similar analysis for taste-based discrimination since we have already shown that every pair of wage distributions are consistent with taste-based discrimination (with the additional restriction to identical mean productivities).

4. DISCUSSION

The purpose of this section is to show that the nonparametric methodology we propose can be adapted to revisit classic approaches to testing for discrimination. First, in Section 4.1, we consider the version of our model where the outcome is binary and we show that statistical discrimination has almost no empirical bite. Then, in Section 4.2, we illustrate the versatility of our approach by revisiting outcome tests à la Becker. Throughout, we will assume that the prior distributions H_i for both groups have the same mean.

[13] To avoid an additional definition, we chose not to include this as a formal statement in the theorem.

4.1. BINARY OUTCOMES

In the introduction, we noted that research studies frequently rely on the differences in *binary* outcomes to document discrimination. For instance, call-back rates from correspondence studies are often used to document discrimination in the labor market. Other instances include mortgage approval rates, credit card approval rates, job promotion and university admission. Dovetailing on the insight in Heckman and Siegelman (1993); Heckman (1998), we stated that statistical discrimination has little bite in such binary settings. We now formalize this statement in the context of our model. Throughout, we use the same notation as in the previous section, but replace the term "wage" with the term "outcome."

There are two outcomes, labeled $w = 0$ and $w = 1$. The distribution G_i is thus a binary distribution, with $G_i(0)$ the probability of outcome $w = 0$. We say that the binary distributions G_1 and G_2 are consistent with statistical discrimination if there exist productivity distributions H_1 and H_2, distributions of posterior estimates F_1 and F_2 that satisfy $H_i \succcurlyeq_2 F_i$, and a *cutoff* $\underline{\theta}$ such that $F_i(\underline{\theta}) = G_i(0)$ for $i \in \{1, 2\}$.

In words, the only difference between this binary outcome setting and the model in Section 2 is that instead of a strictly increasing wage function, there is a group-independent cutoff $\underline{\theta} \in [0, 1]$ such that outcome $w = 1$ occurs only if the posterior estimate is strictly above it. In the context of the labor market, this says that an employer calls back a job candidate only if the candidate's expected productivity is sufficiently high.

The next result characterizes binary outcome distributions that are consistent with statistical discrimination.

THEOREM 4. *Binary outcome distributions G_1 and G_2 are consistent with statistical discrimination if, and only if, it is not the case that either $G_1(0) = 1$ and $G_2(0) = 0$ or $G_1(0) = 0$ and $G_2(0) = 1$.*

In words, any pair of binary distributions is consistent with statistical discrimination unless the outcome for one group is 0 with probability 1, while it is 1 with probability 1 for the other. Again in the context of the labor market, this says that either all job candidates from group 1 are called back and none from group 2 are, or vice versa. Needless to say, such extreme discrimination is never observed and therefore, in practice, statistical discrimination cannot be disentangled from bias or animus in a setting with binary outcomes.

In light of our main result (Theorem 1), it might seem surprising to some readers that Theorem 4 allows two outcome distributions ordered by first-order stochastic dominance to be consistent with statistical discrimination. This difference is driven by the fact that, with binary outcomes, many different posterior estimates are grouped together and assigned a single outcome (depending on whether they are above or

below the cutoff). This is explicitly ruled out in Section 2 because we require the wage function to be strictly increasing and consequently, no two workers with distinct productivity estimates are assigned the same wage.

We end the section by providing a simple argument for this result. By contradiction, suppose that distributions $G_1(0) = 1$ and $G_2(0) = 0$ are consistent with statistical discrimination. Then, the mean of F_1 must be less than or equal to $\underline{\theta}$ (since $F_1(\underline{\theta}) = 1$), whereas the mean of F_2 must be strictly greater (since $F_2(\underline{\theta}) = 0$). This is, of course, not possible since both distributions must have the same mean. A symmetric argument applies when $G_1(0) = 0$ and $G_2(0) = 1$.

Conversely, suppose that $0 < G_1(0) < G_2(0) < 1$. (It is easy to adapt the arguments to treat the other cases.) Let F_1 have binary support $\{0, 1\}$ and assign probability $G_1(0)$ to 0 and, therefore, $1 - G_1(0)$ to 1. Note that F_1 has mean $1 \times (1 - G_1(0)) + 0 \times G(0) = 1 - G_1(0)$. Let F_2 have binary support $\left\{1 - G_1(0) - \frac{\varepsilon}{G_2(0)}, 1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}\right\}$, where $\varepsilon > 0$ is sufficiently small to ensure both points of the support lie in [0,1]. Assign probability $G_2(0)$ to $1 - G_1(0) - \frac{\varepsilon}{G_2(0)}$ and, therefore, $1 - G_2(0)$ to $1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}$. Note that F_2 has mean $G_2(0) \times \left(1 - G_1(0) - \frac{\varepsilon}{G_2(0)}\right) + (1 - G_2(0)) \times \left(1 - G_1(0) + \frac{\varepsilon}{1-G_2(0)}\right) = 1 - G_1(0)$. Thus, the two distributions F_1 and F_2 have the same mean. The second step in the proof of Theorem 1 then shows how to construct a prior H such that $H \succcurlyeq_2 F_i$, as required. The argument is completed by setting the threshold $\underline{\theta} = 1 - G_1(0)$.

4.2. OUTCOME TESTS

The methodology we have introduced is versatile enough to study discrimination in a wide range of settings. As a "proof of concept," we now present one such application — the Becker outcome test — and hope to examine others in future work. To ease the presentation, we frame the application in the context of bail decisions and closely follow Arnold, Dobbie, and Hull (2022). See also Arnold, Dobbie, and Yang (2018), Canay, Mogstad, and Mountjoy (2020), Hull (2021) and Simoiu, Corbett-Davies, and Goel (2017). Several other contexts such as police search decisions and loan decisions also fit the model.

In the context of bail decisions, judges have to decide whether or not to release defendants prior to trials, with $w = 1$ denoting the decision to release a defendant. Upon being released, the defendant may subsequently fail to appear in court or commit another crime, which we model with a binary variable $Y \in \{0, 1\}$; $Y = 1$ indicates a pre-trial misconduct. We–the analyst–observe the fraction $r_i \in (0, 1)$ of defendants from group i released by a judge, and $q_i \in (0, 1)$ the fraction of the released defendants, who committed a pre-trial misconduct.

Unlike the analysis in Section 4.1, where only a single piece of information was observed, we–the analyst–now observe two pieces of information: the release decision along with the post-decision result, that is,

whether the released defendant committed pre-trial misconduct. Now, suppose that the judge releases a defendant if, and only if, the information she possesses signals that the likelihood of pre-trial misconduct is smaller than a given threshold. The Becker outcome test is based on the observation that if the cutoff is the same across groups, the rate of misconduct of the *marginal* defendant of each group should be the same (and equal to the cutoff). Of course, the problem with implementing this test in practice is that the analyst does not observe the identity of the marginal defendant.

So instead, is it possible to detect bias using the averages (r_1, q_1) and (r_2, q_2)? Since the shapes of the signal distributions may differ, the release rates r_1 and r_2 may not be the same even if the judge uses the identical cutoff for both groups. This is the issue with "benchmarking tests" that we discussed in Section 4.1. What if we consider the rates q_1 and q_2 at which misconduct occurs conditional on being released on bail? These too can depend on the shape of the signal distributions and a higher rate of misconduct for either group is possible even if the judge uses a group-neutral cutoff. This is a well-known problem with such "outcome tests" (often referred to as the infra-marginality problem). In the remainder of this section, we show that it is possible to derive a test for statistical discrimination that depends jointly on r_i and q_i, provided we assume that, were all defendants to be released, both groups would commit pre-trial misconduct at the same average rate.

We consider the following model. Assume that Y_i is distributed with (unknown) probability $\theta_i \in [0, 1]$ in group i, with H_i the prior distribution of θ_i. Thus, the ex-ante probability of pre-trial misconduct is $\mathbb{E}_{H_i}[\theta_i]$. Prior to deciding whether to bail a defendant, the judge obtains some information about the likelihood of pre-trial misconduct, and grants the bail only if the perceived probability of misconduct is smaller than the group-independent threshold $\underline{\theta}$. As in previous sections, we assume that the judge receives an unbiased signal about θ_i, with F_i being the distribution of the signal. The distribution F_i is a mean-preserving contraction of H_i. Therefore, the release rate in group i is $F_i(\underline{\theta})$, while the pre-trial misconduct rate conditional on release is $\mathbb{E}_{F_i}[\theta_i | \theta_i \leq \underline{\theta}]$.

Analogous to our previous definitions, we say that the outcomes (r_1, q_1) and (r_2, q_2) are consistent with statistical discrimination if there exist prior distributions H_1 and H_2, posterior distributions F_1 and F_2, and a threshold $\underline{\theta}$ such that (i) $\mathbb{E}_{H_1}[\theta_1] = \mathbb{E}_{H_2}[\theta_2]$, (ii) $H_i \succcurlyeq_2 F_i$, and (iii) $r_i = F_i(\underline{\theta})$ and $q_i = \mathbb{E}_{F_i}[\theta_i | \theta_i \leq \underline{\theta}]$ for $i \in \{1, 2\}$. With this definition in hand, we have the following characterization.

THEOREM 5. *The outcomes (r_1, q_1) and (r_2, q_2) are consistent with statistical discrimination if, and only if, $q_1 < r_2 q_2 + (1 - r_2)1$ and $q_2 < r_1 q_1 + (1 - r_1)1$.*

The above result shows that, under the assumption that both groups commit pre-trial misconduct at the same rate on average, we can precisely identify the conditions under which the outcomes could have arisen from statistical discrimination. As with the rest of this paper, this result requires no further assumptions

on the prior distributions or the signals. It is fully non-parametric and easy to implement. It is worth noting that the aforementioned Simoiu, Corbett-Davies, and Goel (2017) take a different approach. They estimate a parametric model (all distributions lie in certain families parametrized by variables that they estimate), but allow the means of the prior distributions to differ.

5. CONCLUDING REMARKS

In this paper, we introduced a new non-parametric methodology to test whether two distinct wage distributions could have been generated by statistical discrimination alone. Our model is a significantly generalized version of Phelps (1972) and Aigner and Cain (1977): we only require the unobserved productivity distributions to have identical means, the signals from which employers get information about the workers' productivities are unrestricted and wages can be determined by any continuous, strictly increasing function of the posterior productivity estimate. Our test takes a simple form—the wage distributions are consistent with statistical discrimination if neither wage distribution first-order stochastically dominates the other—and the model allows us to interpret a rejection of this condition as evidence of bias. We demonstrate how this test can be generalized to allow both groups to have different mean productivities.

To the best of our knowledge, this paper is the first to analyze the problem of testing for discrimination without resorting to any functional form assumptions and we view this to be one of our main conceptual contributions. In casual discussions of wage gaps, the mean wages of two groups are typically compared and differences are often interpreted (without justification) as evidence of bias. Our test is micro-founded and uses the entire wage distribution but importantly, is just as easy to visualize and implement by non-specialists.

Finally, we hope one consequence of this paper is that the methodology is extended to study discrimination in other settings, possibly with richer data. We have already sketched some possible extensions such as the Becker outcome test, and hope to study more in future work.

A. PROOFS

Recall that G_1 and G_2 have bounded support so $G_i(\overline{w}) = 1$ for some $\overline{w} > 0$, $i = 1, 2$.

PROOF OF THEOREM 1. We have already argued that the first statement implies the second statement. We now argue that the second statement implies the third (the third obviously implies the first).

Proof of [$(ii) \implies (iii)$].

The proof consists of two steps. In the first, we show that there exists a strictly increasing and 1-Lipschtiz piece-wise linear wage function W such that the two distributions defined by $F_i(\theta) = G_i(W(\theta))$ for $i \in \{1, 2\}$ have the same mean. In the second step, we show that for any two distributions with a common mean, there exists a common productivity distribution H such that $H \succcurlyeq_2 F_i$ for both $i \in \{1, 2\}$.

Step 1: There exists an $\overline{\theta} > 0$ and a strictly increasing, 1-Lipschitz function $W : [0, \overline{\theta}] \to [0, \overline{w}]$ such that the two distributions defined by $F_i(\theta) = G_i(W(\theta))$ for $\theta \in [0, \overline{\theta}]$ and $i \in \{1, 2\}$ satisfy $\int_0^{\overline{\theta}} F_1(\theta)d\theta = \int_0^{\overline{\theta}} F_2(\theta)d\theta$.

First, observe that if

$$\int_0^{\overline{w}} G_1(w)dw = \int_0^{\overline{w}} G_2(w)dw,$$

then $W(\theta) = \theta$ along with $\overline{\theta} = \overline{w}$ is trivially the requisite function.

So, without loss, suppose that

$$\int_0^{\overline{w}} G_1(w)dw > \int_0^{\overline{w}} G_2(w)dw.$$

(A symmetric argument applies if we interchange 1 and 2.) Define the function

$$\Delta_G(w) = G_1(w) - G_2(w),$$

and note the above inequality is simply $\int_0^{\overline{w}} \Delta_G(w)dw > 0$.

Since G_2 does not strictly first-order stochastically dominate G_1, there exists a non-empty interval $[\underset{\sim}{w}, \widetilde{w}] \subset (0, \overline{w})$ such that $\int_{\underset{\sim}{w}}^{\widetilde{w}} \Delta_G(w)dw < 0$. (This follows from the right-continuity of G_1 and G_2.) Therefore, there must exist strictly positive constants $\gamma^+ > 0$ and $\gamma^- > 0$ such that

$$\frac{1}{\gamma^+}\int_0^{\underset{\sim}{w}} \Delta_G(w)dw + \frac{1}{\gamma^-}\int_{\underset{\sim}{w}}^{\widetilde{w}} \Delta_G(w)dw + \frac{1}{\gamma^+}\int_{\widetilde{w}}^{\overline{w}} \Delta_G(w)dw = 0.$$

Take any $0 < \kappa < \min\left\{\frac{1}{\gamma^+}, \frac{1}{\gamma^-}\right\}$ and define

$$\begin{aligned}
\underset{\sim}{\theta} &= \frac{\underset{\sim}{w}}{\kappa\gamma^+}, \\
\widetilde{\theta} &= \frac{\widetilde{w} - \underset{\sim}{w} + \kappa\gamma^-\underset{\sim}{\theta}}{\kappa\gamma^-}, \\
\overline{\theta} &= \frac{\overline{w} - \widetilde{w} + \kappa\gamma^+\widetilde{\theta}}{\kappa\gamma^+}.
\end{aligned}$$

Consider the following piece-wise linear wage function

$$W(\theta) = \begin{cases} \kappa\gamma^+\theta & \text{if } 0 \leq \theta < \underset{\sim}{\theta}, \\ \kappa\gamma^-\theta + \underset{\sim}{w} - \kappa\gamma^-\underset{\sim}{\theta} & \text{if } \underset{\sim}{\theta} \leq \theta \leq \widetilde{\theta}, \\ \kappa\gamma^+\theta + \widetilde{w} - \kappa\gamma^+\widetilde{\theta} & \text{if } \widetilde{\theta} < \theta \leq \overline{\theta}. \end{cases}$$

Few observations are worth making. First, W is continuous because

$$\lim_{\theta\uparrow\underset{\sim}{\theta}} W(\theta) = \kappa\gamma^+\underset{\sim}{\theta} = \underset{\sim}{w} = \kappa\gamma^-\underset{\sim}{\theta} + \underset{\sim}{w} - \kappa\gamma^-\underset{\sim}{\theta} = W(\underset{\sim}{\theta})$$

and

$$\lim_{\theta\downarrow\widetilde{\theta}} W(\theta) = \kappa\gamma^+\widetilde{\theta} + \widetilde{w} - \kappa\gamma^+\widetilde{\theta} = \widetilde{w} = \kappa\gamma^-\widetilde{\theta} + \underset{\sim}{w} - \kappa\gamma^-\underset{\sim}{\theta} = W(\widetilde{\theta}),$$

where the equality $\widetilde{w} = \kappa\gamma^-\widetilde{\theta} + \underset{\sim}{w} - \kappa\gamma^-\underset{\sim}{\theta}$ follows from the definition of $\widetilde{\theta}$.

Second, note that W is strictly increasing and 1-Lipschitz since $\kappa\gamma^+ < 1$ and $\kappa\gamma^- < 1$. Third, by construction, observe that $W(\overline{\theta}) = \overline{w}$.

Using the constructed W, define $F_i(\theta) = G_i(W(\theta))$ for $i \in \{1, 2\}$. Let $\Delta_F(\theta) = F_1(\theta) - F_2(\theta)$, and observe that

$$\begin{aligned}
\int_0^{\overline{\theta}} \Delta_F(\theta)d\theta &= \int_0^{\underset{\sim}{\theta}} \Delta_F(\theta)d\theta + \int_{\underset{\sim}{\theta}}^{\widetilde{\theta}} \Delta_F(\theta)d\theta + \int_{\widetilde{\theta}}^{\overline{\theta}} \Delta_F(\theta)d\theta \\
&= \int_0^{\underset{\sim}{\theta}} \Delta_G(W(\theta))d\theta + \int_{\underset{\sim}{\theta}}^{\widetilde{\theta}} \Delta_G(W(\theta))d\theta + \int_{\widetilde{\theta}}^{\overline{\theta}} \Delta_G(W(\theta))d\theta \\
&= \frac{1}{\kappa\gamma^+}\int_0^{\underset{\sim}{w}} \Delta_G(w)dw + \frac{1}{\kappa\gamma^-}\int_{\underset{\sim}{w}}^{\widetilde{w}} \Delta_G(w)dw + \frac{1}{\kappa\gamma^+}\int_{\widetilde{w}}^{\overline{w}} \Delta_G(w)dw \\
&= 0,
\end{aligned}$$

where the second last equality follows from the change of variables from θ to w. Therefore, the constructed distributions F_1 and F_2 have the same mean as required, which completes the proof of this step.

Step 2: Suppose $\int_0^{\overline{\theta}} F_1(\theta)d\theta = \int_0^{\overline{\theta}} F_2(\theta)d\theta$. *Then, there exists a prior distribution* H *such that*

$$\int_0^{\theta} H(x)dx \geq \max\left\{\int_0^{\theta} F_1(x)dx,\ \int_0^{\theta} F_2(x)dx\right\} \textit{ with equality at } \theta = \overline{\theta}.$$

Define the function

$$M(\theta) = \max\left\{\int_0^{\theta} F_1(x)dx,\ \int_0^{\theta} F_2(x)dx\right\}.$$

Observe that M is an increasing, convex function since each $\int_0^{\theta} F_i(x)dx$ is increasing and convex (because F_i is increasing). Also note that

$$M(\overline{\theta}) = \int_0^{\overline{\theta}} F_1(x)dx = \int_0^{\overline{\theta}} F_2(x)dx.$$

Let H be the right derivative of M (the right derivative always exists and, moreover, $M(\theta) = M(0) + \int_0^{\theta} H(x)dx$ since M is convex, hence absolutely continuous). This function is increasing, satisfies $H(0) = 0$, $H(\overline{\theta}) = 1$ and is, therefore, the requisite prior distribution. (Recall the the right derivative of a convex function is right continuous and has limits on the left.)

This completes the proof of $[(ii) \implies (iii)]$.

Step 1 above also shows that the second statement implies the fourth statement (the fourth obviously implies the first). Replacing every instance of F_i by H_i in the statement of Step 1 shows that two wage distributions that are not ordered by first-order stochastic dominance are consistent with the absence of discrimination. This completes the proof of the theorem.

■

PROOF OF THEOREM 2. Without loss, suppose the mean wage of group 1 is the highest; that is, there exists $\alpha \geq 1$ such that $\int_0^{\overline{w}} w dG_1(w) = \alpha \int_0^{\overline{w}} w dG_2(w)$.

Construct the wage functions W_1 and W_2 as:

$$W_1(\theta_1) = \theta_1 \text{ and } W_2(\theta_2) = \theta_2/\alpha,$$

and note that both are clearly strictly increasing and 1-Lipschtiz.

Let $\overline{\theta} = \alpha\overline{w}$ and construct the corresponding prior distributions H_1 and H_2 as follows:

$$H_i(\theta_i) = G_i(W_i(\theta_i)) \text{ for } i \in \{1, 2\}.$$

Since $G_1(\overline{w}) = 1$, $H_1(\overline{w}) = 1$ and, therefore, H_1 is supported on a subset of $[0, \overline{\theta}]$. Similarly, $H_2(\alpha\overline{w}) = G_2(\overline{w}) = 1$ and, therefore, is also supported on a subset of $[0, \overline{\theta}]$.

It remains to verify that the distributions H_1 and H_2 have the same mean. We have:

$$\int_0^{\alpha\overline{w}} \theta_1 dH_1(\theta_1) = \int_0^{\overline{w}} \theta_1 dH_1(\theta_1) = \int_0^{\overline{w}} w_1 dG_1(w_1) = \alpha \int_0^{\overline{w}} w_2 dG_2(w_2) = \int_0^{\alpha\overline{w}} \theta_2 dH_2(\theta_2).$$

This completes the proof. ■

PROOF OF THEOREM 3. (Only if.) Suppose the given wage distributions G_1 and G_2 are induced by model primitives H_1, H_2, F_1, F_2 and W where $|\mathbb{E}_{H_i}[\theta] - \mathbb{E}_{H_j}[\theta]| \leq d$.

Since the wage function is Lipschitz continuous, it is differentiable almost everywhere. Therefore, for each group $i \in \{1, 2\}$, we can write

$$\begin{aligned}\mathbb{E}_{G_i}[w] &= [wG_i(w)]_0^{\overline{w}} - \int_0^{\overline{w}} G_i(w)dw \\ &= \overline{w} - \int_0^{\overline{\theta}} W'(\theta)F_i(\theta)d\theta,\end{aligned}$$

where the second equality follows by a change of variable from w to θ.

In light of Theorem 1, we only need to consider the case where $G_i \succ_1 G_j$ for some $i, j \in \{1, 2\}$. In this case, the above equation implies that

$$\begin{aligned}\mathbb{E}_{G_i}[w] - \mathbb{E}_{G_j}[w] &= \int_0^{\overline{\theta}} W'(\theta)[F_j(\theta) - F_i(\theta)]d\theta \\ &\leq \int_0^{\overline{\theta}} [F_j(\theta) - F_i(\theta)]d\theta \\ &= \mathbb{E}_{F_i}[\theta] - \mathbb{E}_{F_j}[\theta] \\ &= \mathbb{E}_{H_i}[\theta] - \mathbb{E}_{H_j}[\theta] \\ &\leq d,\end{aligned}$$

where the first inequality follows from the fact that W is 1-Lipschitz and $F_j(\theta) \geq F_i(\theta)$ (since $G_j(\theta) \geq G_i(\theta)$). As required, this shows that two wage distributions (ordered by strict first-order stochastic dom-

inance) are consistent with statistical discrimination and productivity difference d only if the wage gap is less than d.

(If.) Once again, in light of Theorem 1, we only need to consider the case where $G_i \succ_1 G_j$ for some $i, j \in \{1, 2\}$ and the wage gap satisfies $|\mathbb{E}_{G_i}[w] - \mathbb{E}_{G_j}[w]| \leq d$.

These wage distributions are induced by $H_i = F_i = G_i$ along with $W(\theta) = \theta$ and these chosen productivity distributions satisfy $|\mathbb{E}_{H_i}[\theta] - \mathbb{E}_{H_j}[\theta]| = |\mathbb{E}_{G_i}[w] - \mathbb{E}_{G_j}[w]| \leq d$ as required. ■

PROOF OF THEOREM 5. (Only if.) Suppose that the outcomes are consistent with statistical discrimination. Since $\mathbb{E}_{F_i}[\theta_i|\theta_i \leq \underline{\theta}] = q_i$, we must have $\underline{\theta} \geq q_i$. Therefore, the mean of F_i must be at least q_i since

$$\mathbb{E}_{F_i}[\theta_i] = F_i(\overline{\theta})\mathbb{E}_{F_i}[\theta_i|\theta_i \leq \underline{\theta}] + (1 - F_i(\underline{\theta}))\mathbb{E}_{F_i}[\theta_i|\theta_i > \underline{\theta}] > r_i q_i + (1 - r_i)\underline{\theta}.$$

Similarly, the mean of F_j, $j \neq i$, is at most $r_j q_j + (1 - r_j)1$. Finally, since F_i and F_j have the same mean, it must be the case that

$$q_i < r_j q_j + (1 - r_j)1,$$

for all (i, j), $j \neq i$.

(If.) The proof is constructive. Without loss of generality, assume that $q_i \geq q_j$. It follows that $r_i q_i + (1 - r_i)1 > q_j$ is automatically satisfied (since $r_i < 1$ and $q_i < 1$). Assume that $r_j q_j + (1 - r_j)1 > q_i$ is also satisfied.

Let $\underline{\theta} = q_i + \delta$ for some $\delta > 0$, and F_i a binary distribution which takes values q_i and $q_i + \varepsilon$ with probability r_i and $1 - r_i$, respectively, where $\varepsilon > \delta$. By construction, the mean of F_i is $q_i + (1 - r_i)\varepsilon$ and $F_i(\underline{\theta}) = r_i$.

We now construct F_j such that its mean is the same as the mean of F_i. The second step in the proof of Theorem 1 then shows how to construct a prior H such that $H \succcurlyeq_2 F_i$ and $H \succcurlyeq_2 F_j$. The distribution F_j is again binary and takes values q_j and $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j}$ with probability r_j and $(1 - r_j)$, respectively. The mean of F_j is:

$$r_j q_j + (1 - r_j)\frac{q_i - r_j q_j + (1 - r_i)\varepsilon}{1 - r_j} = q_i + (1 - r_i)\varepsilon,$$

the same as the mean of F_i.

Finally, we need to choose δ and ε, with $\varepsilon > \delta$, such that (i) $q_i + \varepsilon \leq 1$, (ii) $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j} \leq 1$, and (iii) $\frac{q_i - r_j q_j + (1-r_i)\varepsilon}{1-r_j} > q_i + \delta$. The conditions (i) and (ii) guarantee that F_i and F_j are supported on a subset of $[0, 1]$, while condition (iii) guarantees that $F_j(\underline{\theta}) = r_j$. Since $q_i \geq q_j$ and $\frac{q_i - r_j q_j}{1-r_j} < 1$, it is routine to

verify that we can indeed choose ε and δ as required. ■

REFERENCES

Aigner, D. J. and G. G. Cain (1977): "Statistical theories of discrimination in labor markets," *Industrial and Labor Relations Review*, 30, 175–187.

Altonji, J. G. and C. R. Pierret (2001): "Employer learning and statistical discrimination," *Quarterly Journal of Economics*, 116, 313–350.

Anderson, G. (1996): "Nonparametric tests of stochastic dominance in income distributions," *Econometrica*, 1183–1193.

Anwar, S. and H. Fang (2006): "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence," *American Economic Review*, 96, 127–151.

Arnold, D., W. Dobbie, and P. Hull (2022): "Measuring Racial Discrimination in Bail Decisions," *American Economic Review*, 112, 2992–3038.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial bias in bail decisions," *Quarterly Journal of Economics*, 133, 1885–1932.

Arrow, K. J. (1973): "The theory of discrimination," in *Discrimination in Labor Markets*, Orley Ashenfelter and Albert Rees, eds, 3–33.

Arulampalam, W., A. L. Booth, and M. L. Bryan (2007): "Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution," *ILR Review*, 60, 163–186.

Barrett, G. F. and S. G. Donald (2003): "Consistent tests for stochastic dominance," *Econometrica*, 71, 71–104.

Bayer, P. and K. K. Charles (2018): "Divergent paths: A new perspective on earnings differences between black and white men since 1940," *Quarterly Journal of Economics*, 133, 1459–1501.

Becker, G. S. (1957): *The economics of discrimination*, University of Chicago press.

——— (1993): "Nobel lecture: The economic way of looking at behavior," *Journal of Political Economy*, 101, 385–409.

Bertrand, M. and E. Duflo (2017): "Field experiments on discrimination," *Handbook of Economic Field Experiments*, 1, 309–393.

Bohren, J. A., K. Haggag, A. Imas, and D. G. Pope (2019a): "Inaccurate statistical discrimination: An identification problem," Tech. rep., National Bureau of Economic Research.

Bohren, J. A., A. Imas, and M. Rosenberg (2019b): "The dynamics of discrimination: Theory and evidence," *American Economic Review*, 109, 3395–3436.

Canay, I. A., M. Mogstad, and J. Mountjoy (2020): "On the use of outcome tests for detecting bias in decision making," Tech. rep., National Bureau of Economic Research.

Christofides, L. N., A. Polycarpou, and K. Vrachimis (2013): "Gender wage gaps, 'sticky floors' and 'glass ceilings' in Europe," *Labour Economics*, 21, 86–102.

Coate, S. and G. C. Loury (1993): "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review*, 1220–1240.

Davidson, R. and J.-Y. Duclos (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 68, 1435–1464.

——— (2013): "Testing for restricted stochastic dominance," *Econometric Reviews*, 32, 84–125.

De la Rica, S., J. J. Dolado, and V. Llorens (2008): "Ceilings or floors? Gender wage gaps by education in Spain," *Journal of Population Economics*, 21, 751–776.

Fang, H. and A. Moro (2011): "Theories of statistical discrimination and affirmative action: A survey," *Handbook of Social Economics*, 1, 133–200.

Heckman, J. J. (1998): "Detecting discrimination," *Journal of Economic Perspectives*, 12, 101–116.

Heckman, J. J. and P. Siegelman (1993): "The Urban Institute audit studies: Their methods and findings," *Clear and convincing evidence: Measurement of discrimination in America*, 187–258.

Hull, P. (2021): "What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making," Working Paper 28503, National Bureau of Economic Research.

Knowles, J., N. Persico, and P. Todd (2001): "Racial bias in motor vehicle searches: Theory and evidence," *Journal of Political Economy*, 109, 203–229.

Kuosmanen, T. (2004): "Efficient diversification according to stochastic dominance criteria," *Management Science*, 50, 1390–1406.

Lang, K. and J.-Y. K. Lehmann (2012): "Racial discrimination in the labor market: Theory and empirics," *Journal of Economic Literature*, 50, 959–1006.

Lang, K. and A. K.-L. Spitzer (2020): "Race discrimination: An economic perspective," *Journal of Economic Perspectives*, 34, 68–89.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies*, 72, 735–765.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154, 186–202.

Maasoumi, E. and L. Wang (2019): "The gender gap between earnings distributions," *Journal of*

Political Economy, 127, 2438–2504.

Neumark, D. (2012): "Detecting discrimination in audit and correspondence studies," *Journal of Human Resources*, 47, 1128–1157.

Onuchic, P. (2022): "Recent contributions to theories of discrimination," *arXiv preprint arXiv:2205.05994*.

Phelps, E. S. (1972): "The statistical theory of racism and sexism," *American Economic Review*, 62, 659–661.

Post, T. (2003): "Empirical tests for stochastic dominance efficiency," *Journal of Finance*, 58, 1905–1931.

Simoiu, C., S. Corbett-Davies, and S. Goel (2017): "The problem of infra-marginality in outcome tests for discrimination," *Annals of Applied Statistics*, 11, 1193–1216.