STATISTICAL DISCRIMINATION AND THE DISTRIBUTION OF WAGES

ONLINE APPENDIX

Prashant Bharadwaj [Q]
Rahul Deb [◊]
Ludovic Renou [‡]

September 18, 2026

[Q] Department of Economics, University of California, San Diego. prbharadwaj@ucsd.edu
[◊] Department of Economics, Boston College. rahul.deb@bc.edu
[‡] W. P. Carey Department of Economics, Arizona State University; School of Economics and Finance, Queen Mary University of London; CEPR. lrenou.econ@gmail.com

1. APPLICATION TO NLSY79

As a complementary empirical analysis to the one we conduct on the Census data, we also implement our test on the NLSY79. An advantage of these data is that they allow us to condition on pre-market ability as measured by AFQT scores; the seminal work of **?** established the importance of this measure in determining racial wage gaps. Controlling for AFQT scores (along with other covariates) allows us to test for the presence of taste-based discrimination under the equal means assumption (Theorem 1).

We restrict the NLSY79 sample to non-Hispanic Black and White prime-aged (30-55) men who worked 50–52 weeks in the year 2000 and report positive wages in 2000. The outcome variable is log annual wages in 2000 (`logwage2000`).

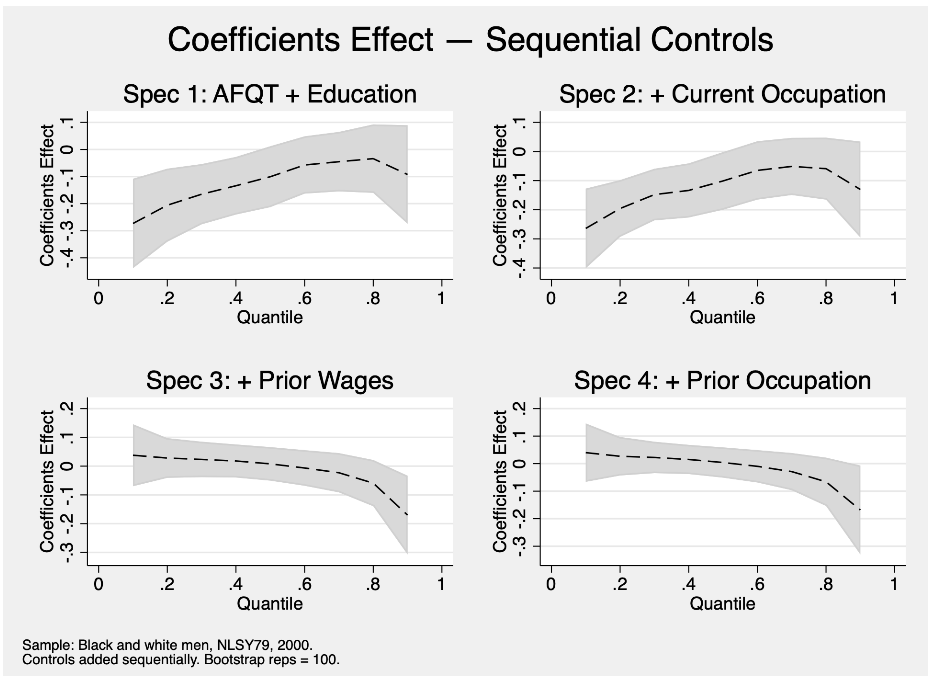
Once again, due to limitations on the number of observations of Black workers, we have to aggregate the data in a few different ways. The NLSY79 records educational attainment as years of completed schooling which ranges from 0 to 20. The returns to education are nonlinear (a completed high school or college degree is disproportionately more valuable than dropping out one year earlier) but we do not have enough observations to include fixed effects for years of schooling. Consequently, we create four groups to match the census `educ` variable used in the ordered means restriction: less than high school ($\leq$ 11 years), high school diploma (12 years), some college or associate's degree (13–15 years), and college or above ($\geq$ 16 years). Because we can control for ability via AFQT scores, we can compare Black and White workers with similar education (as opposed to what we do with the Census data).

The NLSY79 occupation variables are coded on the 1980 CPS occupation classification. We collapse occupations into two categories: white collar (covering Management, Professional, Technical, Sales, and Office/Administrative) and blue collar (covering Service, Farming, Construction, Repair, Production, Laborers, and Transport). This binary classification, while not ideal, is the most occupational heterogeneity we can allow at NLSY79 sample sizes.

?? displays the wage structure effect under increasingly stringent controls; all panels control for educational attainment, age, age-squared and AFQT scores. Both the top left and top right panels show clear evidence of strict first-order stochastic dominance. The top right panel includes the current occupation dummy (blue collar vs. white collar).

Another advantage of the NLSY data (relative to the Census) is that it is a panel. This allows us to control for past wages and occupations. These controls allow us to compare Black and White workers who worked the *same* job in the previous period at the *same* wage. If there was no discrimination in the previous period wages, two employees of each group who are paid the same must have the same expected productivity (since we assume that wages only depend on expected productivities). Conversely, if employers had engaged in discrimination against Black workers in the previous period, we expect Black workers to

Figure 1: Wage Structure Effects from the NLSY



Data is from the NLSY79. Wages recorded in calendar year 2000 for men working full time (50-52 weeks). CDECO Stata command based on **?**. Only the wage structure effect (coefficient plot from CDECO) is displayed. All estimates control for educational attainment, age, age-squared and AFQT scores. In addition, the top right panel controls for the binary occupation dummy in 2000, bottom left controls additionally for wages in 1998, and the bottom right controls additionally for the binary occupation dummy in 1998.

be *more* productive than White workers for them to have been paid the *same* wage. The remaining possibility is that Black workers have a lower expected productivity, but got paid the same as White workers (with higher mean productivity) in the previous period. While unlikely in our view, this is possible if we think there is strong bias in the opposite direction or if forces like affirmative action are at play.

The bottom two panels display the wage structure effect with additional controls for past wages and occupations. In both these panels, we do not find evidence of strict first-order stochastic dominance. This is not entirely unexpected: conditioning on past wages compares workers who have already reached the same wage level, so the test is now asking whether there is additional discrimination in wage growth conditional on past wages and occupations. Detecting such an effect requires substantially more statistical power than the NLSY sample provides.

2. CONVEX AND CONCAVE WAGES

We denote

$$\mathcal{W}_{conv} := \{W \in \mathcal{W} \mid W \text{ is strictly convex}\}$$

to be the set of strictly increasing and strictly convex continuous functions. Likewise, we use $\mathcal{W}_{conc}$ to denote the set of strictly increasing and strictly concave continuous functions.

We say that wage distribution G_i dominates distribution G_j in the *strict increasing concave order*, if

$$\int_0^{\overline{w}} M(w)dG_i(w) > \int_0^{\overline{w}} M(w)dG_j(w)$$

for every strictly increasing, strictly concave function $M : \mathbb{R}_+ \to \mathbb{R}_+$. The *strict increasing convex order* can be analogously defined when the above inequality holds for all strictly increasing, strictly convex functions $M : \mathbb{R}_+ \to \mathbb{R}_+$. For more discussion on these orders, see Section 4.A in **?**.

We derive a version of Theorem 1 assuming wages are convex.

THEOREM 3. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{eq}$, $\mathcal{W}_{conv}$) with equal mean productivities and a strictly convex wage function if, and only if, neither G_1 nor G_2 dominates the other in the strict increasing concave order.*

PROOF.

(Only if.) The proof is by contradiction. Suppose that the wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{eq}$, $\mathcal{W}_{conv}$) and yet G_i dominates G_j in the strict concave order ($i \neq j$). Since the wage distributions are rationalizable, there exists a strictly increasing, strictly convex wage function W, distributions (F_1, F_2) such that $G_k(w) = F_k(W^{-1}(w))$ for all $w \in [0, \overline{w}]$ and priors (H_1, H_2) such that $\mathbb{E}_{H_1}[\theta] = \mathbb{E}_{H_2}[\theta]$ and $F_k \succcurlyeq_2 H_k$ for $k \in \{1, 2\}$. Define $\overline{\theta} = W^{-1}(\overline{w})$.

Since W is strictly increasing and strictly convex, W^{-1} is strictly increasing and strictly concave. To see the latter, choose $(w, w') \in [0, \overline{w}]^2$ with $w \neq w'$. From the strict convexity of W, we have that

$$\alpha W(W^{-1}(w)) + (1-\alpha)W(W^{-1}(w')) > W(\alpha W^{-1}(w) + (1-\alpha)W^{-1}(w')).$$

for all $\alpha \in (0, 1)$. Since W^{-1} is strictly increasing, it follows that

$$W^{-1}\left(\alpha W(W^{-1}(w)) + (1-\alpha)W(W^{-1}(w'))\right) > W^{-1}\left(W(\alpha W^{-1}(w) + (1-\alpha)W^{-1}(w'))\right),$$

which is equivalent to

$$W^{-1}(\alpha w + (1-\alpha)w') > \alpha W^{-1}(w) + (1-\alpha)W^{-1}(w'),$$

as required.

Finally, since we have assumed G_1 and G_2 are rationalizable, we have that

$$\int_0^{\overline{w}} W^{-1}(w)dG_i(w) = \int_0^{\overline{\theta}} \theta dF_i(\theta) = \int_0^{\overline{\theta}} \theta dH_i(\theta) = \int_0^{\overline{\theta}} \theta dH_j(\theta) = \int_0^{\overline{\theta}} \theta dF_j(\theta) = \int_0^{\overline{w}} W^{-1}(w)dG_j(w),$$

which is the required contradiction since G_i dominates G_j in the strict concave order.

(If.) The argument is almost identical to the part of the proof of Theorem 1 showing statement (ii) implies statement (i). Since neither G_1 nor G_2 dominates the other in the strict concave order, the same argument as used in the proof of Theorem 1 implies that there exists a strictly increasing, strictly concave function M such that

$$\int_0^{\overline{w}} M(w)dG_1(w) = \int_0^{\overline{w}} M(w)dG_2(w).$$

Since, affine transformations of M will not affect the above equation, we can, without loss of generality, assume that $M(0) = 0$. Let $\overline{\theta} = M(\overline{w})$.

Take a strictly increasing and strictly convex function $W : \mathbb{R}_+ \to \mathbb{R}_+$ that satisfies $W(\theta) := M^{-1}(\theta)$ for $\theta \in [0, \overline{\theta}]$ and define $F_i(\theta) := G_i(W(\theta))$. Clearly, F_i is a well-defined distribution.

Finally, observe that

$$\int_0^{\overline{\theta}} \theta dF_1(\theta) = \int_0^{\overline{w}} M(w)dG_1(w) = \int_0^{\overline{w}} M(w)dG_2(w) = \int_0^{\overline{\theta}} \theta dF_2(\theta)$$

where the first and last equalities follow from the change of variable $\theta = M(w)$ and the fact that $F_i(M(w)) = G_i(w)$ by construction. Taking $H_i = F_i$ completes the proof. ■

This result can also be used to derive the analogous version of Theorem 2 for the case of ordered means. If we were to instead assume wage functions were concave, the statement would be verbatim with $\mathcal{W}_{conc}$ replacing $\mathcal{W}_{conv}$ and the strict increasing convex order replacing the strict increasing concave order.

If $G_i \succ_1 G_j$, then G_i also dominates distribution G_j in both the strict increasing concave and convex orders. Thus, **??** shows that the testable implications are stronger under the assumption of convex wages but, once again, they take a form that is easy to take to the data. There are well-known tests for higher orders of stochastic dominance (the strict increasing concave order is related to second-order stochastic dominance) developed in the econometrics literature (once again, see **?**).

3. WITHIN-A-NEIGHBORHOOD MEANS

We use

$$\mathcal{H}_{|1-2|\leq d} := \{(H_1, H_2) \mid |\mathbb{E}_{H_1}[\theta_1] - \mathbb{E}_{H_2}[\theta_2]| \leq d\},$$

to denote the set of pairs of productivity distributions whose means differ by at most $d \in \mathbb{R}_+$. Additionally, we assume

$$\mathcal{W}_{L1} := \{W \in \mathcal{W} \mid |W(\theta) - W(\theta')| \leq |\theta - \theta'| \text{ for all } \theta, \theta' \in \Theta\},$$

that is, wage functions are 1-Lipschitz. This technical restriction imposes some discipline on the wage function, that is, it ties changes in wages with changes in productivities. It is weaker than assuming that wage functions are differentiable with slope at most 1.

We now characterize wage distributions that are rationalizable under these assumptions.

THEOREM 4. *Wage distributions G_1 and G_2 are rationalizable (given $\mathcal{H}_{|1-2|\leq d}$, $\mathcal{W}_{L1}$) with mean productivities that are within* d *of each other and a 1-Lipschitz wage function if, and only if, either*

(i) the wage gap is at most d*, that is,* $|\mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w]| \leq d$*, or*

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

PROOF.

(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. Let $\overline{\theta} = W^{-1}(\overline{w})$. 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 $i \in \{1, 2\}$ and $j \in \{1, 2\} \setminus \{i\}$. 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 $G_j(W(\theta)) \geq G_i(W(\theta))$). As required, this shows that two wage distributions (ordered by strict first-order stochastic dominance) are rationalizable (given $\mathcal{H}_{|1-2|\leq d}$, $\mathcal{W}_{L1}$) only if the wage gap is at most d.

(If.) Once again, in light of Theorem 1 (the remark following the proof points out that the wage function can be chosen to be 1-Lipschitz), we only need to consider the case where $G_i \succ_1 G_j$ for $i \in \{1, 2\}$ and $j \in \{1, 2\} \setminus \{i\}$ 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. ■

As Theorem 1 shows, if wages are not ordered by strict first-order stochastic dominance, they can be rationalized with equal mean productivities. Consequently, they will be rationalized (given $\mathcal{H}_{|1-2|\leq d}$, $\mathcal{W}_{L1}$) for any $d \geq 0$; the additional restriction on the wage function $\mathcal{W}_{L1} \subset \mathcal{W}$ does not affect Theorem

1 (the older version of this paper **?** has a formal statement of this result). So the real contribution of **??** is the first condition (i). This shows that the wage gap can be a useful statistic to test for the presence of taste-based discrimination, but *only* when the wage distributions are ordered by strict first-order stochastic dominance.

An issue with applying the result in **??** to data is that the researcher has to choose the appropriate mean productivity difference d to test for. However, the statement of this theorem can be inverted to show that the wage gap is a *tight* lower bound for mean productivity differences required to rationalize the wage distributions. In other words, if we want statistical discrimination alone to rationalize the data, the wage gap is the smallest difference in productivity means required, whenever one wage distribution first-order stochastically dominates the other.

THEOREM 5. *Suppose* $G_2 \succ_1 G_1$. *Then, there exist* (H_1, F_1, H_2, F_2, W) *that jointly induce* G_1, G_2 *such that* $W \in \mathcal{W}_{L1}$ *and* $\mathbb{E}_{H_2}[\theta] - \mathbb{E}_{H_1}[\theta] = \mathbb{E}_{G_2}[w] - \mathbb{E}_{G_1}[w]$.

Moreover, every (H_1, F_1, H_2, F_2, W) *with* $W \in \mathcal{W}_{L1}$, *that jointly induce* G_1, G_2 *satisfy* $\mathbb{E}_{H_2}[\theta] - \mathbb{E}_{H_1}[\theta] \geq \mathbb{E}_{G_2}[w] - \mathbb{E}_{G_1}[w]$.

In words, this result states that every (H_1, F_1, H_2, F_2, W) with $W \in \mathcal{W}_{L1}$ that induce wage distributions $G_2 \succ_1 G_1$ have the feature that the difference in mean productivities is at least the wage gap, and that this bound is tight. Thus, the wage gap is a useful measure of the minimum productivity difference required for statistical discrimination alone to rationalize the wage distributions. The data can be used to contextualize this bound as we briefly mention in the discussion in Section 5 of the paper.

Instead of bounding absolute productivity differences, one could also try to bound percentage productivity differences. Formally, for $\alpha \in (0, 1)$, let

$$\mathcal{H}_{1/2\geq\alpha} := \{(H_1, H_2) \mid \mathbb{E}_{H_1}[\theta_1] \geq \alpha \mathbb{E}_{H_2}[\theta_2]\},$$

denote the set of productivity distribution pairs such that group 1's mean productivity is at least a fraction α of that of group 2. As with **??**, one could first characterize the set of rationalizable wage distributions and then invert the result to derive the bound. Unfortunately, there is too little structure in the model to bound percentage differences.

THEOREM 6. *Suppose* $G_1(0) = G_2(0) = 0$. *Then, for every* $\alpha \in (0, 1)$, *every pair of wage distributions is rationalizable (given* $\mathcal{H}_{1/2\geq\alpha}$, $\mathcal{W}_{L1}$*) with the ratio of mean productivities being at least* α *and a 1-Lipschitz wage function.*

PROOF. We will construct a wage function such that the distribution of posterior expected productivities obtained satisfies $\mathbb{E}_{F_1}[\theta] > \alpha \mathbb{E}_{F_2}[\theta]$. We then take $H_i = F_i$.

Since $G_1(0) = G_2(0) = 0$ (and cumulative distributions are right continuous), there exists a small enough $\tilde{w} > 0$ such that

$$\frac{1 - G_1(w)}{1 - G_2(w)} > \alpha \qquad \text{for all} \quad w \leq \tilde{w}.$$

This in turn implies that

$$\frac{\int_0^{\tilde{w}} [1 - G_1(w)] dw}{\int_0^{\tilde{w}} [1 - G_2(w)] dw} > \alpha.$$

Now take an $\varepsilon \in (0, 1)$ and consider the following wage function

$$W_\varepsilon(\theta) = \begin{cases} \varepsilon\theta & \text{when } \theta \in [0, \tilde{w}/\varepsilon], \\ \tilde{w} + \left(\theta - \frac{\tilde{w}}{\varepsilon}\right) & \text{when } \theta > \tilde{w}/\varepsilon. \end{cases}$$

Note that, by construction, it is 1-Lipschitz. Let $\overline{\theta} = W^{-1}(\overline{w})$.

With this wage function, the ratios of the mean productivities become

$$\begin{aligned}
\frac{\mathbb{E}_{F_1}[\theta]}{\mathbb{E}_{F_2}[\theta]} &= \frac{\int_0^{\overline{\theta}} \theta dF_1(\theta)}{\int_0^{\overline{\theta}} \theta dF_2(\theta)} \\
&= \frac{\int_0^{\overline{\theta}} [1 - F_1(\theta)] d\theta}{\int_0^{\overline{\theta}} [1 - F_2(\theta)] d\theta} \\
&= \frac{\frac{1}{\varepsilon} \int_0^{\tilde{w}} [1 - G_1(w)] dw + \int_{\tilde{w}}^{\overline{w}} [1 - G_1(w)] dw}{\frac{1}{\varepsilon} \int_0^{\tilde{w}} [1 - G_2(w)] dw + \int_{\tilde{w}}^{\overline{w}} [1 - G_2(w)] dw} \\
&= \frac{\int_0^{\tilde{w}} [1 - G_1(w)] dw + \varepsilon \int_{\tilde{w}}^{\overline{w}} [1 - G_1(w)] dw}{\int_0^{\tilde{w}} [1 - G_2(w)] dw + \varepsilon \int_{\tilde{w}}^{\overline{w}} [1 - G_2(w)] dw}.
\end{aligned}$$

Since $\frac{\int_0^{\tilde{w}} [1 - G_1(w)] dw}{\int_0^{\tilde{w}} [1 - G_2(w)] dw} > \alpha$, we can find small enough ε such that

$$\frac{\int_0^{\tilde{w}} [1 - G_1(w)] dw + \varepsilon \int_{\tilde{w}}^{\overline{w}} [1 - G_1(w)] dw}{\int_0^{\tilde{w}} [1 - G_2(w)] dw + \varepsilon \int_{\tilde{w}}^{\overline{w}} [1 - G_2(w)] dw} = \frac{\mathbb{E}_{F_1}[\theta]}{\mathbb{E}_{F_2}[\theta]} > \alpha$$

which completes the proof because W_ε is the requisite wage function. ∎

This result states that, for any fraction α, every pair of wage distributions is rationalizable with productivity distributions whose means are within a fraction α of each other. Loosely speaking, this is because we can consider wage functions that are very flat for low wages but have slope one for higher wages. With such a wage function, absolute differences in higher wages correspond to absolute differences in productivities but because the values of the latter are high, the percentage difference between mean productivities becomes small. The additional assumption $G_1(0) = G_2(0) = 0$ is trivially satisfied in any application since we restrict attention to working adults and no one works for zero wages.

4. AGGREGATION

Our theory assumes that wages are generated by a *single* wage function W. Wage data in a given occupation frequently consists of wages from many different firms. Our test on aggregated data may suffer from either Type-I or Type-II error. We demonstrate this via two examples.

Suppose there are two firms j_1 and j_2 and half the workers from each group work at each firm. Workers from both groups are assumed to have equal mean productivities. Further, suppose $\overline{w} = 1$ and so wage distributions are supported on a subset of $[0, 1]$. We first construct an example to demonstrate that, even though there is taste-based discrimination against group 1 workers, the aggregate wage distributions may not be ordered by strict first-order stochastic dominance. In this case, our test conducted on the aggregated wage distributions will (incorrectly) not find evidence of taste-based discrimination.

Specifically, let group 1's wage distributions

$$G_1^{j_1}(w) = G_1^{j_2}(w) = w$$

be the uniform distribution $\mathbb{U}[0, 1]$ at both firms. Let group 2's wage distributions be

$$G_2^{j_1}(w) = w^2 \quad \text{and} \quad G_2^{j_2}(w) = \begin{cases} 0 & \text{if } 0 \leq w \leq \frac{1}{4}, \\ 2\left(w - \frac{1}{4}\right) & \text{if } \frac{1}{4} < w \leq \frac{3}{4}, \\ 1 & \text{if } \frac{3}{4} < w \leq 1 \end{cases}$$

at firms j_1 and j_2 respectively. Observe that firm j_1 practices taste-based discrimination against group 1 because $G_2^{j_1} \succ_1 G_1^{j_1}$. Firm j_2 does not discriminate and the wage distributions of both worker groups at this firm (uniform distributions $\mathbb{U}[0, 1]$ and $\mathbb{U}[1/4, 3/4]$ respectively) are not ordered by strict first-order stochastic dominance.

The aggregate distribution $\frac{1}{2}G_1^{j_1} + \frac{1}{2}G_1^{j_2}$ for group 1 is the uniform $\mathbb{U}[0, 1]$ distribution. The aggregate distribution for group 2 evaluated at $w = \frac{3}{4}$ satisfies

$$\frac{1}{2}G_2^{j_1}(3/4) + \frac{1}{2}G_2^{j_2}(3/4) = \frac{1}{2} \times \frac{9}{16} + \frac{1}{2} \times 1 = \frac{25}{32} > \frac{3}{4} = \frac{1}{2}G_1^{j_1}(3/4) + \frac{1}{2}G_1^{j_2}(3/4)$$

and therefore the aggregate wage distributions $\frac{1}{2}G_2^{j_1} + \frac{1}{2}G_2^{j_2} \not\succ_1 \frac{1}{2}G_1^{j_1} + \frac{1}{2}G_1^{j_2}$ are not ordered by first-order stochastic dominance. Thus, in this example, our null hypothesis of no taste-based discrimination will incorrectly not be rejected (Type-II error).

Conversely, we now construct an example in which the wage distributions of both groups at both firms are not ordered by strict first-order stochastic dominance but they are in the aggregate. Specifically, let group 1's wage distribution

$$G_1^{j_1}(w) = G_1^{j_2}(w) = w$$

once again be the uniform distribution $\mathbb{U}[0,1]$ at both firms. Group 2's wages are given by

$$G_2^{j_1}(w) = \begin{cases} w + \frac{1}{16} - \left(w - \frac{1}{4}\right)^2 & \text{if } 0 \leq w \leq \frac{1}{2}, \\ w - \frac{1}{32} + \frac{1}{2}\left(w - \frac{3}{4}\right)^2 & \text{if } \frac{1}{2} < w \leq 1, \end{cases} \qquad G_2^{j_2}(w) = \begin{cases} w - \frac{1}{32} + \frac{1}{2}\left(w - \frac{1}{4}\right)^2 & \text{if } 0 \leq w \leq \frac{1}{2}, \\ w + \frac{1}{16} - \left(w - \frac{3}{4}\right)^2 & \text{if } \frac{1}{2} < w \leq 1. \end{cases}$$

Note that for firm j_1, the wage distributions satisfy $G_2^{j_1}(w) > G_1^{j_1}(w)$ when $w \in (0, 1/2)$ and $G_2^{j_1}(w) < G_1^{j_1}(w)$ when $w \in (1/2, 1)$. The opposite inequalities hold for the wage distributions in firm j_2. Thus the wage distributions in both firms are not ordered by first-order stochastic dominance. However, when we aggregate them, we get

$$\frac{1}{2}G_2^{j_1}(w) + \frac{1}{2}G_2^{j_2}(w) = \begin{cases} w + \frac{1}{64} - \frac{1}{4}\left(w - \frac{1}{4}\right)^2 & \text{if } 0 \leq w \leq \frac{1}{2}, \\ w + \frac{1}{64} - \frac{1}{4}\left(w - \frac{3}{4}\right)^2 & \text{if } \frac{1}{2} < w \leq 1, \end{cases}$$

and observe that $\frac{1}{2}G_2^{j_1}(w) + \frac{1}{2}G_2^{j_2}(w) \geq w = \frac{1}{2}G_1^{j_1}(w) + \frac{1}{2}G_1^{j_2}(w)$ with the inequality strict for $w \in (0, 1/2) \cup (1/2, 1)$. Thus, wages $\frac{1}{2}G_1^{j_1} + \frac{1}{2}G_1^{j_2} \succ_1 \frac{1}{2}G_2^{j_1} + \frac{1}{2}G_2^{j_2}$ are ordered by strict first-order stochastic dominance in the aggregate and so our test will incorrectly reject the null and suggest the presence of taste-based discrimination (Type-I error).

5. MULTIDIMENSIONAL PRODUCTIVITY

We use $\theta_i = (\theta_i^1, \ldots, \theta_i^K)$ to denote the K-dimensional productivity of group i. We will assume that employers perfectly observe the workers' productivities. This simplifies the presentation of the result and makes the result stronger because we are not employing the additional generality that signals allow for.

We say that a pair of distributions G_1 and G_2 are rationalizable with K-dimensional productivities if there exist (i) productivity distributions $H_1 \in \Delta\left(\Theta^K\right)$ and $H_2 \in \Delta\left(\Theta^K\right)$ that have the same mean $\mathbb{E}_{H_1}[\theta_1^k] = \mathbb{E}_{H_2}[\theta_2^k]$ in all dimensions $k \in \{1, \ldots, K\}$ and (ii) a continuous wage function $W : \mathbb{R}_+^K \to \mathbb{R}$ that is additively separable and strictly increasing in each productivity dimension such that these jointly induce the observed wage distributions. Observe that we allow the wage function to map to $\mathbb{R}$ instead of $\mathbb{R}_+$; this is purely to simplify the proof and does not affect the result when $G_1(0) = G_2(0) = 0$.

The next result shows that we can rationalize every pair of distributions with four dimensional productivities.

THEOREM 7. *Every pair of wage distributions G_1 and G_2 is rationalizable with 4-dimensional productivities.*

PROOF. Suppose $\mu_1 := \mathbb{E}_{G_1}[w] > \mu_2 =: \mathbb{E}_{G_2}[w]$. The opposite case can be handled analogously. If the means are equal, then the distributions are not ordered by strict first-order stochastic dominance and so can be rationalized with a single dimensional productivity.

We now construct the productivities H_1 and H_2 for each group. For group 1, the distribution H_1 is given by:

- $\theta_1^1 \sim G_1$ is drawn independently and its marginal distribution is the same distribution as the wage distribution of group 1;
- $\theta_1^2 \sim \delta_{\mu_2}$ is the degenerate distribution at the mean wage μ_2 of group 2;
- (θ_1^3, θ_1^4) are independent of the first two dimensions and they take values either $\left(0, \sqrt{2(\mu_1 - \mu_2)}\right)$ or $\left(\sqrt{2(\mu_1 - \mu_2)}, 0\right)$ each with probability $\frac{1}{2}$.

For group 2, the productivity distribution H_2 is given by:

- $\theta_2^1 \sim \delta_{\mu_1}$ is the degenerate distribution at the mean wage μ_1 of group 1;
- $\theta_2^2 \sim G_2$ is drawn independently and its marginal distribution is the same distribution as the wage distribution of group 2;
- $\theta_2^3 \sim \delta_{\frac{\sqrt{2(\mu_1-\mu_2)}}{2}}$ is the degenerate distribution at $\frac{\sqrt{2(\mu_1-\mu_2)}}{2}$;
- $\theta_2^4 \sim \delta_{\frac{\sqrt{2(\mu_1-\mu_2)}}{2}}$ is the degenerate distribution at $\frac{\sqrt{2(\mu_1-\mu_2)}}{2}$.

Notice that each productivity dimension has the same mean under both distributions H_1 and H_2: they are μ_1, μ_2, $\frac{\sqrt{2(\mu_1-\mu_2)}}{2}$ and $\frac{\sqrt{2(\mu_1-\mu_2)}}{2}$ respectively.

Let the wage function be given by

$$W(\theta^1, \theta^2, \theta^3, \theta^4) = \theta^1 + \theta^2 + (\theta^3)^2 + (\theta^4)^2 + \mu_2 - 2\mu_1.$$

and observe that it is continuous, additively separable and strictly increasing in each argument.

Now observe that for group 1, the mean wage is

$$\mathbb{E}_{H_1}[W(\theta^1, \theta^2, \theta^3, \theta^4)] = \mu_1 + \mu_2 + \frac{1}{2} \times 2(\mu_1 - \mu_2) + \frac{1}{2} \times 2(\mu_1 - \mu_2) + \mu_2 - 2\mu_1 = \mu_1.$$

The wage distribution induced by H_1 and W is thus precisely the distribution G_1 as required since the distribution of $\theta_1^2 + (\theta_1^3)^2 + (\theta_1^4)^2 + \mu_2 - 2\mu_1$ is simply the degenerate distribution at 0.

Similarly, observe that for group 2, the mean wage is

$$\mathbb{E}_{H_2}[W(\theta^1, \theta^2, \theta^3, \theta^4)] = \mu_1 + \mu_2 + \frac{2(\mu_1 - \mu_2)}{4} + \frac{2(\mu_1 - \mu_2)}{4} + \mu_2 - 2\mu_1 = \mu_2.$$

The wage distribution induced by H_2 and W is precisely G_2 as required since the distribution of $\theta_2^1 + (\theta_2^3)^2 + (\theta_2^4)^2 + \mu_2 - 2\mu_1$ is simply the degenerate distribution at 0.

This completes the proof. ■

In addition to requiring that the wage function is additively separable, we could assume that each dimension of the productivity $(\theta_i^1, \ldots, \theta_i^K)$ for both groups is drawn independently (with each dimension of the productivity vector having equal means for both groups). Even in this case, Theorems 1 and 2 do not hold; in other words, rationalizable wage distributions are not characterized by the absence of a strict first-order stochastic dominance ordering.

As an example, suppose productivity is two-dimensional and the wage function is given by

$$W(\theta^1, \theta^2) = (\theta^1)^2 + \theta^2.$$

Consider the following productivity distributions H_1 and H_2 and additionally assume that signals are perfectly informative so the distribution of posterior expected productivities satisfy $F_i = H_i$ for both $i \in \{1, 2\}$. H_1 assigns an equal probability to $\left(\frac{1}{3}, 1\right)$ and $(1, 1)$ and H_2 assigns an equal probability to $\left(\frac{2}{3}, \frac{1}{2}\right)$ and $\left(\frac{2}{3}, \frac{3}{2}\right)$. Note that each dimension of the productivity is independently distributed (because one dimension of each of H_1 and H_2 is constant) and the means of each dimension of H_1 and H_2 are the same ($\frac{2}{3}$ and 1 respectively).

These generate wage distributions

$$G_1(w) = \begin{cases} 0 & \text{if } 0 \le w < \left(\frac{1}{3}\right)^2 + 1 = \frac{10}{9}, \\ \frac{1}{2} & \text{if } \frac{10}{9} \le w < 1^2 + 1 = 2, \\ 1 & \text{if } 2 \le w, \end{cases} \quad \text{and} \quad G_2(w) = \begin{cases} 0 & \text{if } 0 \le w < \left(\frac{2}{3}\right)^2 + \frac{1}{2} = \frac{17}{18}, \\ \frac{1}{2} & \text{if } \frac{17}{18} \le w < \left(\frac{2}{3}\right)^2 + \frac{3}{2} = \frac{35}{18}, \\ 1 & \text{if } w \ge \frac{35}{18}. \end{cases}$$

These distributions are rationalizable (with each dimension of the type vector independently drawn and an additively separable wage function) by construction but nonetheless the wage distributions $G_1 \succ_1 G_2$ are ordered by strict first-order stochastic dominance.

6. HIGHER MOMENTS

In the main text, we assumed that the wage is a strictly increasing function of the posterior mean productivity alone. We now show that if wages were to depend on higher moments of the posterior productivity distribution, then all wage distributions are rationalizable.

Once again, we overload notation but the meaning should be clear. Let $W_\alpha(\mu, \sigma^2) = \mu - \alpha\sigma^2$ with $\alpha \geq 0$ be a wage function that assigns a wage based on both the posterior mean productivity μ and the posterior variance σ^2. Note that W_α is a linear function of the posterior mean and variance, is increasing in the former and decreasing in the latter. We define

$$\mathcal{W}_{\mu,\sigma^2} := \{W_\alpha \mid \alpha \geq 0\}$$

to be the set of all such wage functions. Rationalizability (given $\mathcal{H}_{eq}, \mathcal{W}_{\mu,\sigma^2}$) is defined as before, except it is no longer sufficient to simply consider the distribution of posterior mean productivities F_i (since the variance matters too).

We now show that meaningful testable implications of the model disappear even with this limited class of wage functions that are an affine function of only the posterior mean and variance. This shows that a general (non-Gaussian) version of the model of **?** is not refutable. In the following result, a *nontrivial* wage distribution is one that does not have mass 1 on the highest wage $w = \overline{w}$.

THEOREM 8. *Every pair of nontrivial wage distributions G_1 and G_2 is rationalizable (given $\mathcal{H}_{eq}, \mathcal{W}_{\mu,\sigma^2}$) with equal mean productivities and a wage function that is linear in the posterior mean and variance.*

PROOF. If neither G_1 nor G_2 strictly first-order stochastically dominates the other, then the result follows from Theorem 1. So, assume that $G_i \succ_1 G_j$ for $i \neq j$ and hence $\mathbb{E}_{G_i}[w] > \mathbb{E}_{G_j}[w]$.

To ease notation, we normalize the wages to be distributed in $[0, 1]$. (This is without loss of generality since we can consider the re-scaled values $\widehat{w} := w/\overline{w}$ and the re-scaled distribution $\widehat{G}_k$, given by $\widehat{G}_k(\widehat{w}) = G_k(\overline{w} \times \widehat{w})$ for $k \in \{1, 2\}$.)

In what follows, we construct a wage function $W_\alpha \in \mathcal{W}_{\mu,\sigma^2}$ that rationalizes the wage distributions.

Take an $\alpha \geq 0$. For all $w \in [0, 1]$, let $\theta^\alpha(w) \in [0, 1]$ be the greatest solution to

$$w = \theta - \alpha\theta(1 - \theta).$$

Note that the right-hand side of the above equation is continuous in θ, and takes values 0 and 1 at $\theta = 0$ and $\theta = 1$ respectively. Thus, a greatest solution $\theta^\alpha(w) \in [0, 1]$ exists. Moreover, except when $w = 0$, the above equation has a unique solution since the right side is strictly increasing whenever its value is strictly greater than 0. Then, note that $\theta^\alpha(w)$ is strictly increasing in w. Finally, observe that $\theta^\alpha(w) \geq w$ and $\lim_{\alpha\to\infty} \theta^\alpha(w) = 1$.

Let $\Theta = [0, 1]$. Define the prior distribution H_j^α of group j to be the Bernoulli distribution (so supported on $\{0, 1\}$) such that the probability of $[\theta = 1]$ is $\int_0^1 \theta^\alpha(w) dG_j(w)$. Consider the signal (S_j, π_j) where $S_j = [0, 1]$ and π_j is a joint distribution over $\Theta \times S_j$ such that

(i) the marginal distribution of π_j over Θ is the prior H_j^α,

(ii) the marginal distribution over S_j is G_j,

(iii) the posterior distribution upon observing any $s_j \in [0, 1]$ is a Bernoulli distribution with the probability of $[\theta = 1]$ being $\theta^\alpha(s_j)$.

Observe that this is a valid signal because the average of the posterior distributions $\int_0^1 \theta^\alpha(s_j) dG_j(s_j)$ assigns the same probability to $[\theta = 1]$ as the prior. Observe also that, by construction, the prior distribution H_j^α, the signal (S_j, π_j) and the wage function W_α induce the wage distribution G_j.

For group i, we assume that the signal is perfectly informative and that the prior distribution H_i is G_i. Note that since the experiment is perfectly informative, the posterior variance is always zero so the wage at any signal realization is simply the posterior expected productivity. Clearly, the prior H_i, a perfectly informative signal (S_i, π_i) and the wage function W_α induce the wage distribution G_i.

It remains to argue that we can choose α such that

$$\mathbb{E}_{H_j^\alpha}[\theta] = \mathbb{E}_{H_i}[\theta].$$

By construction, we have:

$$\begin{aligned}
&\mathbb{E}_{H_j^\alpha}[\theta] = \int \theta^\alpha(w) dG_j(w) \geq \mathbb{E}_{G_j}[w], \\
&\lim_{\alpha\to+\infty} \mathbb{E}_{H_j^\alpha}[\theta] = 1, \\
&\mathbb{E}_{H_i}[\theta] = \mathbb{E}_{G_i}[w] > \mathbb{E}_{G_j}[w].
\end{aligned}$$

Since $\mathbb{E}_{G_i}[w] < 1$ (as distribution G_i is nontrivial), there exists a requisite $\alpha \geq 0$. This completes the proof. ■

It is possible to show an analogous result if we instead assume $\alpha \leq 0$. Our choice of $\alpha \geq 0$ is motivated by **?** and is natural if employers are risk averse. If we instead allow $\alpha \in \mathbb{R}$ to be either positive or negative, then all distributions (including those that are not nontrivial) can be rationalized with equal mean productivities.

We end this section by studying one additional wage setting process that does not depend on the posterior expected productivity alone. As in the next section, we will assume a finite set of possible wages $\{w_1, \ldots, w_L\}$ where $0 < w_1 < w_2 < \cdots < w_L < 1$ (the last inequality is a normalization). We use $g_{i,\ell}$ to be the probability that workers of group i earn w_ℓ. To ease the exposition, we do not normalize the signals and so each signal realization is not required to be an accurate estimate of the true productivity.

We say that wage distributions G_1 and G_2 are *rationalizable with learning* if there exist (i) productivity distributions $H_1, H_2 \in \Delta(\Theta)$ with equal means, (ii) a signal (S, π_i) for each group $i \in \{1, 2\}$, and (iii) a strictly increasing wage function $W : \mathbb{R}_+ \to \mathbb{R}_+$, such that $\text{Prob}_{\pi_i}(\{s \mid \mathbb{E}_{\pi_i}[W(\theta)|s] = w_\ell\}) = g_{i,\ell}$.

In words, each signal realization s (that arises with positive probability) generates a posterior distribution $\pi_i(\cdot|s)$ over the productivities Θ and the wage corresponding to this signal is $\mathbb{E}_{\pi_i}[W(\theta)|s]$. The rationalization with learning criterion requires that the measure of the set of signal realizations (evaluated using the marginal distribution of π_i over S) that lead to wage w_ℓ is precisely the observed probability $g_{i,\ell}$.

Intuitively, if employers know for sure that the worker's productivity is θ, they will pay them $W(\theta)$. The rationalization with learning criterion captures the fact that employers may eventually perfectly learn the worker's productivity. When the employer is still learning the worker's productivity, the wage is simply the expectation of the future wage.

We now show that there are no testable restrictions subject to this criterion.

THEOREM 9. *Every pair of wage distributions G_1 and G_2 is rationalizable with learning.*

PROOF. Without loss of generality, assume that

$$\mathbb{E}_{G_1}[w] - \mathbb{E}_{G_2}[w] = \sum_{\ell=1}^{L} g_{1,\ell} w_\ell - \sum_{\ell=1}^{L} g_{2,\ell} w_\ell \geq 0$$

or, in words, that group 1 has weakly higher mean wages. We now construct the ingredients of the model that induce the given wage distributions.

The productivity distribution H_1 is supported on two points $\{0, 1\}$. H_1 assigns probability $\sum_{\ell=1}^{L} g_{1,\ell} w_\ell$ to productivity $\theta = 1$ and $1 - \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$ to productivity $\theta = 0$.

For both groups, we assume that the set of signal realizations is $S := \{s_1, \ldots, s_L\}$. The signal (S, π_1) for group 1 is constructed as follows. The marginal distribution over S from π_1 is such that signal realization s_ℓ occurs with probability $g_{1,\ell}$. The posterior distribution $\pi_1(\cdot|s_\ell)$ assigns probability w_ℓ to productivity $\theta = 1$ and probability $1 - w_\ell$ to productivity $\theta = 0$. Observe that the marginal distribution of π_1 over Θ is H_1 as it assigns probability $\sum_{\ell=1}^{L} g_{1,\ell} w_\ell$ to $\theta = 1$ and the remaining probability $1 - \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$ to $\theta = 0$.

Define $W(0) = 0$ and $W(1) = 1$. For any signal realization s_ℓ, the wage offered to workers is given by $\pi_1(0|s_\ell) \times W(0) + \pi_1(1|s_\ell) \times W(1) = w_\ell$ and thus H_1, π_1 along with the defined values of W induce the wage distribution G_1.

If $\sum_{\ell=1}^{L} g_{2,\ell} w_\ell = \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$, then the same construction for H_2 yields an appropriate rationalization. So suppose that $\sum_{\ell=1}^{L} g_{2,\ell} w_\ell < \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$.

Take a value $0 < w^* < w_1$. For each $\ell \in \{1, \ldots, L\}$, we define q_ℓ by the equation

$$q_\ell + (1 - q_\ell) w^* = w_\ell \iff q_\ell = \frac{w_\ell - w^*}{1 - w^*} \in (0, w_\ell).$$

Now observe that

$$\sum_{\ell=1}^{L} g_{2,\ell} q_\ell + \left(1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell\right) \theta = 1 > \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$$

when $\theta = 1$ and

$$\sum_{\ell=1}^{L} g_{2,\ell} q_\ell + \left(1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell\right) \theta = \sum_{\ell=1}^{L} g_{2,\ell} q_\ell \le \sum_{\ell=1}^{L} g_{2,\ell} w_\ell < \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$$

when $\theta = 0$. We take $\theta^* \in (0, 1)$ to be the value that satisfies

$$\sum_{\ell=1}^{L} g_{2,\ell} q_\ell + \left(1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell\right) \theta^* = \sum_{\ell=1}^{L} g_{1,\ell} w_\ell$$

and note that such a θ^* exists by continuity.

The productivity distribution H_2 is supported on two points $\{\theta^*, 1\}$. H_2 assigns probability $\sum_{\ell=1}^{L} g_{2,\ell} q_\ell$ to 1 and the remaining probability $1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell$ to θ^*. Thus, the mean of H_2 is $\sum_{\ell=1}^{L} g_{2,\ell} q_\ell + \left(1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell\right) \theta^*$. Note that, by construction, the means of both productivity distributions satisfy $\mathbb{E}_{H_2}[\theta_2] = \mathbb{E}_{H_1}[\theta_1]$ as required.

Take the wage function W to be any strictly increasing, continuous function that satisfies $W(0) = 0$,

$W(\theta^*) = w^*$ and $W(1) = 1$.

The signal (S, π_2) for group 2 is constructed as follows. The marginal distribution over S from π_2 is such that signal realizations s_ℓ occurs with probability $g_{2,\ell}$. The posterior distribution $\pi_2(\cdot|s_\ell)$ assigns probability q_ℓ to productivity $\theta = 1$ and probability $1 - q_\ell$ to productivity $\theta = \theta^*$. Observe that the marginal distribution of π_2 over Θ is precisely H_2 as it assigns probability $\sum_{\ell=1}^{L} g_{2,\ell} q_\ell$ to productivity $\theta = 1$ and $1 - \sum_{\ell=1}^{L} g_{2,\ell} q_\ell$ to productivity $\theta = \theta^*$.

For any signal realization s_ℓ, the wage offered to workers is given by $\pi_2(\theta^*|s_\ell) \times W(\theta^*) + \pi_2(1|s_\ell) \times W(1) = w_\ell$ and thus H_2, π_2 and W induce the wage distribution G_2.

This completes the proof. ■

7. MULTIPLE GROUPS

In the main text, we have restricted our attention to two groups. The purpose of this section is to extend Theorem 1 to n groups. To ease the arguments in this and the next section, we assume that all wage distributions are finitely supported. Without loss of generality, we assume that the support is included in $\mathbf{W} = \{0, 1, 2, \ldots, L\}$. (In empirical applications, wages are supported on finitely many rationals and we can always redefine them to be supported on finitely many integers.)

Let $N = \{1, \ldots, n\}$ be the set of groups. We use $g_{i,\ell}$ to denote the probability that group i receives wage ℓ and $G_{i,\ell}$ to denote the cumulative probability, that is, $G_{i,\ell}$ is the probability that wages are less than or equal to ℓ.

We once again assume that all groups have the same mean productivity and we denote this using the same notation

$$\mathcal{H}_{eq} := \{(H_1, H_2, \ldots, H_n) \,|\, \mathbb{E}_{H_1}[\theta_1] = \mathbb{E}_{H_2}[\theta_2] = \cdots = \mathbb{E}_{H_n}[\theta_n]\}.$$

As with Theorem 1, we consider unrestricted wage functions in the set $\mathcal{W}$.

Theorem 1 generalizes to $n \geq 2$ groups as follows.

THEOREM 10. *The wage distributions $(G_1, \ldots, G_n)$ supported on a subset of $\mathbf{W}$ are rationalizable (given $\mathcal{H}_{eq}, \mathcal{W}$) if, and only if, there do not exist nonempty subsets $I, I' \subset N$ that satisfy $I \cap I' = \emptyset$, $I \cup I' = N$ and weights $\beta \in \Delta^{|I|-1}$, $\beta' \in \Delta^{|I'|-1}$ such that*

$$\sum_{i \in I} \beta_i G_i \succ_1 \sum_{i' \in I'} \beta'_{i'} G_{i'}.$$

In words, this condition states that we cannot find two subsets of the groups and convex combinations of their *wage distributions that are ordered by strict first-order stochastic dominance.*

PROOF. Since we have assumed wage distributions are discrete, we can rewrite the definition of rationalizability in terms of a system of equalities and inequalities. We then use a version of the Farkas lemma to take the alternative and the latter yields the requisite characterization.

Given any strictly increasing wage function $W \in \mathcal{W}$, there exists a unique θ_ℓ such that $W(\theta_\ell) = \ell$. Thus, to rationalize the wage distributions $(G_i)_{i\in N}$, we need to find a vector $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_\ell, \ldots, \theta_L)$ with $0 \le \theta_\ell < \theta_{\ell+1}$ such that the mean productivities

$$\sum_{\ell=0}^{L} \theta_\ell g_{i,\ell} = \sum_{\ell=0}^{L} \theta_\ell g_{j,\ell}, \tag{1}$$

of all groups $(i, j) \in N \times N$ are equal. In other words, the wage distributions are rationalizable if, and only if, there is a solution to the above equations as then we can choose $W(\theta_\ell) = \ell$, $F_i(\theta_\ell) = G_{i,\ell}$ and $H_i = F_i$ for all $i \in N$ and $0 \le \ell \le L$ to rationalize the wage distributions.

We now rewrite rationalizability conditions as a linear system of equalities and inequalities. To do so, we define two matrices. The matrix $\mathbf{A}$ is

$$\mathbf{A} = \begin{bmatrix} g_{1,0} - g_{2,0} & \cdots & g_{1,\ell} - g_{2,\ell} & \cdots & g_{1,L} - g_{2,L} \\ g_{2,0} - g_{3,0} & \cdots & g_{2,\ell} - g_{3,\ell} & \cdots & g_{2,L} - g_{3,L} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{n-1,0} - g_{n,0} & \cdots & g_{n-1,\ell} - g_{n,\ell} & \cdots & g_{n-1,L} - g_{n,L} \end{bmatrix}$$

and the matrix $\mathbf{C}$ is

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & \ldots & \ldots & 0 \\ 0 & -1 & 1 & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \ldots & -1 & 1 \end{bmatrix}.$$

The matrix $\mathbf{A}$ has $n-1$ rows and $L+1$ columns. The matrix $\mathbf{C}$ has L rows and $L+1$ columns. Wage distributions $(G_i)_{i\in N}$ are rationalizable (given $\mathcal{H}_{eq}, \mathcal{W}$) iff we can find a solution to the system

$$\mathbf{A} \cdot \boldsymbol{\theta} = 0 \text{ and } \mathbf{C} \cdot \boldsymbol{\theta} \gg 0.$$

The first constraint captures the equal mean productivity requirement (**??**) and the second constraint captures the requirement that $\theta_\ell < \theta_{\ell+1}$ for $\ell \in \{0, \ldots, L-1\}$. Note that we do not need to include a constraint that ensures $\boldsymbol{\theta} \ge 0$ since if we find a solution $\boldsymbol{\theta}$, we can also find a positive solution by adding a large enough positive constant to $\boldsymbol{\theta}$.

A general version of the Farkas lemma (see Theorem 1.6.1 in **?**) states that either there exists $\boldsymbol{\theta}$ that satisfies $\mathbf{A} \cdot \boldsymbol{\theta} = 0$ and $\mathbf{C} \cdot \boldsymbol{\theta} \gg 0$ or there exists a solution $(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ to

$$\boldsymbol{\alpha} \cdot \mathbf{A} + \boldsymbol{\gamma} \cdot \mathbf{C} = 0,$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{n-1})$, $\boldsymbol{\gamma} = (\gamma_0, \ldots, \gamma_{L-1})$ are row vectors of dimension $n-1$, L respectively and $\boldsymbol{\gamma} > 0$ (that is, all coordinates are nonnegative with at least one positive).

The vector equation of the alternative $\boldsymbol{\alpha} \cdot \mathbf{A} + \boldsymbol{\gamma} \cdot \mathbf{C} = 0$ can be written in the scalar form

$$\sum_{i=1}^{n-1} \alpha_i (g_{i,\ell} - g_{i+1,\ell}) + (\gamma_{\ell-1} - \gamma_\ell) = 0,$$

for all $0 \leq \ell \leq L-1$. (With the normalization, $\gamma_{-1} = 0$.)

If we sum up these equations from 0 to $0 \leq \bar{\ell} \leq L-1$, we get

$$\sum_{i=1}^{n-1} \alpha_i (G_{i,\bar{\ell}} - G_{i+1,\bar{\ell}}) + \sum_{\ell=0}^{\bar{\ell}} (\gamma_{\ell-1} - \gamma_\ell) = \sum_{i=1}^{n-1} \alpha_i (G_{i,\bar{\ell}} - G_{i+1,\bar{\ell}}) - \gamma_{\bar{\ell}} = 0. \tag{2}$$

Thus, the equation $\boldsymbol{\alpha} \cdot \mathbf{A} + \boldsymbol{\gamma} \cdot \mathbf{C} = 0$ of the alternative is equivalent to (**??**) holding for all $0 \leq \bar{\ell} \leq L-1$.

Further, since $\boldsymbol{\gamma} > 0$, a solution to the alternative exists iff there exists $\boldsymbol{\alpha}$ such that

$$\sum_{i=1}^{n-1} \alpha_i (G_{i,\ell} - G_{i+1,\ell}) \geq 0, \tag{3}$$

for all $0 \leq \ell \leq L-1$ with at least one strict inequality. To see this, observe that if there was an $\boldsymbol{\alpha}$ satisfying the above system of inequalities, we could simply set $\gamma_\ell = \sum_{i=1}^{n-1} \alpha_i (G_{i,\ell} - G_{i+1,\ell})$ and this would satisfy $\boldsymbol{\gamma} > 0$.

For a given ℓ, condition (**??**) can be rewritten as

$$\sum_{i=1}^{n} (\alpha_i - \alpha_{i-1}) G_{i,\ell} \geq 0, \tag{4}$$

with the convention that $\alpha_0 = \alpha_n = 0$. Thus, a necessary and sufficient condition for wage distributions $(G_1, \ldots, G_n)$ to be rationalizable (given $\mathcal{H}_{eq}, \mathcal{W}$) is that there does not exist $\boldsymbol{\alpha}$ that satisfies (**??**) for all $0 \leq \ell \leq L-1$ with at least one strict inequality. We now show that this condition is equivalent to the condition in the statement of the theorem.

First, suppose the wage distributions $(G_1, \dots, G_n)$ are not rationalizable (given $\mathcal{H}_{eq}, \mathcal{W}$). This implies that there exists $\boldsymbol{\alpha}$ that satisfies (**??**) for all $0 \le \ell \le L-1$ with at least one strict inequality. Now since $\sum_{i=1}^{n}(\alpha_i - \alpha_{i-1}) = \alpha_n - \alpha_0 = 0$, the $\alpha_i - \alpha_{i-1}$ terms cannot all be negative or positive and moreover, they cannot all be zero since condition (**??**) has to hold strictly for at least one ℓ.

Without loss, let the groups be labeled such that there is a $1 \le \bar{n} \le n-1$ for which $\alpha_i - \alpha_{i-1} \ge 0$ for $i \le \bar{n}$ and $\alpha_i - \alpha_{i-1} < 0$ for $i > \bar{n}$. We can rewrite inequality (**??**) as

$$\sum_{i=1}^{\bar{n}}(\alpha_i - \alpha_{i-1})G_{i,\ell} \ge \sum_{j=\bar{n}+1}^{n}(\alpha_{j-1} - \alpha_j)G_{j,\ell}. \tag{5}$$

Observe that $0 < \sum_{i=1}^{\bar{n}}(\alpha_i - \alpha_{i-1}) = \alpha_{\bar{n}} = \sum_{j=\bar{n}+1}^{n}(\alpha_{j-1} - \alpha_j)$. Define

$$\beta_i = \begin{cases} (\alpha_i - \alpha_{i-1})/\alpha_{\bar{n}} & \text{if } i \le \bar{n}, \\ (\alpha_{i-1} - \alpha_i)/\alpha_{\bar{n}} & \text{if } i > \bar{n} \end{cases}$$

and observe that $(\beta_1, \dots, \beta_{\bar{n}})$ and $(\beta_{\bar{n}+1}, \dots, \beta_n)$ are $\bar{n}$ and $n - \bar{n}$ dimensional probability vectors. Thus, because (**??**) holds for all $0 \le \ell \le L-1$ with at least one strict inequality, we get that

$$\sum_{j=\bar{n}+1}^{n} \beta_j G_j \succ_1 \sum_{i=1}^{\bar{n}} \beta_i G_i.$$

Conversely, suppose, without loss, that for some $1 \le \bar{n} \le n-1$, we have

$$\sum_{j=\bar{n}+1}^{n} \beta_j G_j \succ_1 \sum_{i=1}^{\bar{n}} \beta_i G_i.$$

for some $(\beta_1, \dots, \beta_{\bar{n}}) \in \Delta^{\bar{n}-1}$, $(\beta_{\bar{n}+1}, \dots, \beta_n) \in \Delta^{n-\bar{n}-1}$. Define

$$\begin{aligned} &\alpha_1 = \beta_1,\ \alpha_2 = \beta_1 + \beta_2,\ \dots,\ \alpha_{\bar{n}} = \beta_1 + \dots + \beta_{\bar{n}} = 1 \quad \text{and} \\ &\alpha_{\bar{n}+1} = 1 - \beta_{\bar{n}+1},\ \alpha_{\bar{n}+2} = 1 - (\beta_{\bar{n}+1} + \beta_{\bar{n}+2}),\ \dots,\ \alpha_{n-1} = 1 - (\beta_{\bar{n}+1} + \dots + \beta_{n-1}) = \beta_n \end{aligned}$$

and observe that this implies that (**??**) holds for all $0 \le \ell \le L-1$ with at least one strict inequality. Thus, these wage distributions cannot be rationalized (given $\mathcal{H}_{eq}, \mathcal{W}$). ■

It is also possible to derive an analogue of Theorem 2. Here, we assume that there are non-empty disjoint subsets $(I_1, \ldots, I_m)$ of N with $\cup_{j=1}^m I_j = N$ and that the set of productivity distributions is given by

$$\{(H_1, H_2, \ldots, H_n) \,|\, \mathbb{E}_{H_j}[\theta_j] \geq \mathbb{E}_{H_{j'}}[\theta_{j'}] \text{ when } j \in I_k,\ j' \in I_{k'},\ k < k' \text{ and } \mathbb{E}_{H_j}[\theta_j] = \mathbb{E}_{H_{j'}}[\theta_{j'}] \text{ when } k = k'\}.$$

We omit the statement for brevity.

8. RATIONALIZATION WHEN THE SLOPE OF THE WAGE FUNCTION IS BOUNDED

In the setting of **??**, we restricted attention to 1-Lipschitz wage functions, which imposes an upper bound on the slope of the wage functions. However, no lower bound was imposed. This made it possible for arbitrarily large changes in expected productivities to induce arbitrarily small changes in wages paid – a fact we exploit in the proof. The purpose of this section is to extend the analysis to wage functions, whose slopes take values in an interval $[a, b] \subset \mathbb{R}_{++}$. The lower and upper bounds may be taken from the literature, for instance from Mincerian wage equation estimates. As in the previous section, we assume (to simplify arguments) that all wage distributions are finitely supported on a subset of $\mathbf{W}$.

Specifically, we assume that the wage functions W are strictly increasing and Lipschitz continuous with a modulus of continuity in $[a, b]$; that is, $a(\theta' - \theta) \leq W(\theta') - W(\theta) \leq b(\theta' - \theta)$ for all $\theta' > \theta$. We denote the set of such wage functions by $\mathcal{W}_{Lab} \subset \mathcal{W}$. Therefore, wherever differentiable, $W'(\theta) \in [a, b]$ for all $\theta \in \mathbb{R}_+$, $W \in \mathcal{W}_{Lab}$. We additionally impose that $W(0) = 0$, so that $a\theta \leq W(\theta) \leq b\theta$: wages are a (possibly variable) fraction between a and b of expected productivity.[1] Throughout this section, we repeatedly use the following two observations.

Observation 1: When $x, x', y, y', \alpha > 0$, the fraction

$$\frac{\alpha x + x'}{\alpha y + y'} = \frac{\frac{x}{y}(\alpha y + y') + y'\left(\frac{x'}{y'} - \frac{x}{y}\right)}{\alpha y + y'} = \frac{x}{y} + \frac{y'\left(\frac{x'}{y'} - \frac{x}{y}\right)}{\alpha y + y'}$$

is increasing in α if $\frac{x}{y} > \frac{x'}{y'}$ and decreasing in α if $\frac{x}{y} < \frac{x'}{y'}$.

Observation 2: When $x, x', y, y' > 0$, then

$$\frac{x}{y} \geq (\text{resp., } \leq) \frac{x + x'}{y + y'} \iff \frac{x}{y} \geq (\text{resp., } \leq) \frac{x}{y} + \frac{y'\left(\frac{x'}{y'} - \frac{x}{y}\right)}{y + y'} \iff \frac{x}{y} \geq (\text{resp., } \leq) \frac{x'}{y'}.$$

We first argue that, because wages are discrete, it is without loss to restrict attention to piecewise linear wage functions.

[1] If instead wages had a floor $W(0) = c > 0$ that is unrelated to productivity (for instance, due to a binding minimum wage), the bounds derived below are obtained by applying the same algorithm to the wage segments above c.

LEMMA 1. *Consider two wage distributions G_1 and G_2 supported on a subset of* $\mathbf{W}$ *and a wage function $W \in \mathcal{W}_{Lab}$. Let F_1 and F_2 be the distributions over posterior expected productivities induced by this wage function, that is, $F_i(\theta) = G_i(W(\theta))$.*

Then, there exists a piecewise-linear wage function $\widehat{W} \in \mathcal{W}_{Lab}$ whose kink points θ satisfy $\widehat{W}(\theta) \in \mathbf{W}$ such that this wage function also induces distributions over posterior expected productivities F_1 and F_2, that is, $F_i(\theta) = G_i(\widehat{W}(\theta))$.

PROOF. Let $\{\theta_0, \ldots, \theta_L\} = \{W^{-1}(0), \ldots, W^{-1}(L)\}$ be the pre-image of the wage function W.

Consider the function

$$\widehat{W}(\theta) = \begin{cases} \frac{\theta-\theta_0}{\theta_1-\theta_0} & \text{if } \theta \in [\theta_0, \theta_1), \\ 1 + \frac{\theta-\theta_1}{\theta_2-\theta_1} & \text{if } \theta \in [\theta_1, \theta_2), \\ \vdots & \vdots \\ L-1+\frac{\theta-\theta_{L-1}}{\theta_L-\theta_{L-1}} & \text{if } \theta \in [\theta_{L-1}, \infty). \end{cases}$$

Note that this function has the feature that $\widehat{W}(\theta_\ell) = \ell$ for all $0 \leq \ell \leq L$.

We now argue that $\widehat{W} \in \mathcal{W}_{Lab}$. It is clearly continuous by construction. Now, observe that for any $0 \leq \ell \leq L-1$ and $(\theta', \theta) \in [\theta_\ell, \theta_{\ell+1}]^2$ with $\theta' > \theta$, we have

$$\frac{\widehat{W}(\theta') - \widehat{W}(\theta)}{\theta' - \theta} = \frac{\widehat{W}(\theta_{\ell+1}) - \widehat{W}(\theta_\ell)}{\theta_{\ell+1} - \theta_\ell} = \frac{W(\theta_{\ell+1}) - W(\theta_\ell)}{\theta_{\ell+1} - \theta_\ell} \in [a, b]$$

where the first equality follows from the piecewise linearity of $\widehat{W}$ and the value of rightmost expression is in $[a, b]$ because $W \in \mathcal{W}_{Lab}$.

Now take any $\theta < \theta'$ such that $\theta \in [\theta_\ell, \theta_{\ell+1}]$ and $\theta' \in [\theta_{\ell'}, \theta_{\ell'+1}]$ where $0 \leq \ell < \ell' \leq L-1$. We then have

$$\begin{aligned} \frac{\widehat{W}(\theta') - \widehat{W}(\theta)}{\theta' - \theta} &= \frac{\widehat{W}(\theta') - \widehat{W}(\theta_{\ell'}) + \widehat{W}(\theta_{\ell'}) - \widehat{W}(\theta_{\ell'-1}) + \cdots + \widehat{W}(\theta_{\ell+1}) - \widehat{W}(\theta)}{\theta' - \theta_{\ell'} + \theta_{\ell'} - \theta_{\ell'-1} + \cdots + \theta_{\ell+1} - \theta} \\ &\in \left[\min_{\ell \leq k \leq \ell'} \left\{ \frac{\widehat{W}(\theta_{k+1}) - \widehat{W}(\theta_k)}{\theta_{k+1} - \theta_k} \right\}, \max_{\ell \leq k \leq \ell'} \left\{ \frac{\widehat{W}(\theta_{k+1}) - \widehat{W}(\theta_k)}{\theta_{k+1} - \theta_k} \right\} \right] \\ &\subseteq [a, b]. \end{aligned}$$

The second line follows from two facts. The first is that $\frac{\widehat{W}(\theta')-\widehat{W}(\theta_{\ell'})}{\theta'-\theta_{\ell'}} = \frac{\widehat{W}(\theta_{\ell'+1})-\widehat{W}(\theta_{\ell'})}{\theta_{\ell'+1}-\theta_{\ell'}}$ and $\frac{\widehat{W}(\theta_{\ell+1})-\widehat{W}(\theta)}{\theta_{\ell+1}-\theta} = \frac{\widehat{W}(\theta_{\ell+1})-\widehat{W}(\theta_\ell)}{\theta_{\ell+1}-\theta_\ell}$ because of the piecewise-linearity of $\widehat{W}$. The second is that the value of the fraction $\frac{x+x'}{y+y'}$

with $x, x', y, y' \geq 0$ lies between $\min\left\{\frac{x}{y}, \frac{x'}{y'}\right\}$ and $\max\left\{\frac{x}{y}, \frac{x'}{y'}\right\}$ (because of Observation 2).

The proof of the lemma is completed by the observation that the distributions of posterior expected productivities F_1 and F_2 obtained by inverting G_1 and G_2 are the same under either W or $\widehat{W}$ since, by construction $W(\theta_\ell) = \widehat{W}(\theta_\ell) = \ell$ for all $0 \leq \ell \leq L$. ■

A direct implication of the lemma is that it is without loss of generality to restrict attention to piecewise linear wage functions. Note that, if W is piecewise linear, so is W^{-1}. Moreover, the set of kink points of W^{-1} is a subset of $\mathbf{W}$ and the slope of W^{-1} between any $\ell, \ell+1 \in \mathbf{W}$ (which is simply $W^{-1}(\ell+1) - W^{-1}(\ell)$) lies in the interval $[1/b, 1/a]$.

We use $\mathcal{M}$ to denote the set of strictly increasing, piecewise-linear continuous functions (the notation captures that M is W upside down) whose finite kink points are a subset of $\mathbf{W}$ and whose slope at every differentiable point lies between $[1/b, 1/a]$. Each $M \in \mathcal{M}$ takes the form

$$M(w) = \begin{cases} \alpha_0 w & \text{if } 0 \leq w \leq 1, \\ \alpha_1(w-1) + M(1) & \text{if } 1 < w \leq 2, \\ \vdots & \vdots \\ \alpha_{L-1}(w-(L-1)) + M(L-1) & \text{if } w > L-1 \end{cases}$$

where all $\alpha_i \in [1/b, 1/a]$.

Our aim is to bound the ratio

$$\frac{\mathbb{E}_{H_1}[\theta]}{\mathbb{E}_{H_2}[\theta]} = \frac{\mathbb{E}_{F_1}[\theta]}{\mathbb{E}_{F_2}[\theta]}.$$

For any $M \in \mathcal{M}$, this ratio equals

$$\frac{\mathbb{E}_{F_1}[\theta]}{\mathbb{E}_{F_2}[\theta]} = \frac{\int_0^{M(L)}[1-F_1(\theta)]d\theta}{\int_0^{M(L)}[1-F_2(\theta)]d\theta} = \frac{\int_0^L[1-G_1(w)]M'(w)dw}{\int_0^L[1-G_2(w)]M'(w)dw} = \frac{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_1(\ell)]}{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_2(\ell)]}.$$

Thus, the bounds for $\frac{\mathbb{E}_{H_1}[\theta]}{\mathbb{E}_{H_2}[\theta]}$ are

$$\left[\min_{\alpha_0,\ldots,\alpha_{L-1}\in\left[\frac{1}{b},\frac{1}{a}\right]} \frac{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_1(\ell)]}{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_2(\ell)]}, \max_{\alpha_0,\ldots,\alpha_{L-1}\in\left[\frac{1}{b},\frac{1}{a}\right]} \frac{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_1(\ell)]}{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_2(\ell)]}\right].$$

In what follows, we describe how to derive the values for these bounds. Specifically, we solve

$$\max_{\alpha_0,\ldots,\alpha_{L-1}\in\left[\frac{1}{b},\frac{1}{a}\right]} \frac{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_1(\ell)]}{\sum_{\ell=0}^{L-1}\alpha_\ell[1-G_2(\ell)]}. \tag{6}$$

The value of the lower bound can be derived analogously. The maximization problem (**??**) is essentially a knapsack problem as the following two lemmata show. We use the convention that $\frac{1-\widehat{G}_1(\ell)}{1-\widehat{G}_2(\ell)} = 0$ whenever both the numerator and denominator are 0.

LEMMA 2. *There is a solution $(\alpha_0^*, \ldots, \alpha_{L-1}^*)$ to* (**??**) *where every $\alpha_\ell^* \in \{\frac{1}{b}, \frac{1}{a}\}$. In words, the slopes of each linear piece take either the maximum or minimum possible value.*

PROOF. Suppose the solution has some $\alpha_\ell^* \in \left(\frac{1}{b}, \frac{1}{a}\right)$. Then, from Observation 1, if

$$\frac{1-G_1(\ell)}{1-G_2(\ell)} \geq \frac{\sum_{\ell' \neq \ell} \alpha_{\ell'}^*[1-G_1(\ell')]}{\sum_{\ell' \neq \ell} \alpha_{\ell'}^*[1-G_2(\ell')]}$$

the value of the objective function in (**??**) will weakly increase if we set $\alpha_\ell^* = \frac{1}{a}$. Conversely if the above inequality is reversed, the value of the objective function in (**??**) will weakly increase if we set $\alpha_\ell^* = \frac{1}{b}$. ■

Now, sort the series

$$\left\{\frac{1-G_1(0)}{1-G_2(0)}, \frac{1-G_1(1)}{1-G_2(1)}, \ldots, \frac{1-G_1(L-1)}{1-G_2(L-1)}\right\}$$

in increasing order and denote this sorted series by

$$\left\{\frac{1-\widetilde{G}_1(0)}{1-\widetilde{G}_2(0)}, \frac{1-\widetilde{G}_1(1)}{1-\widetilde{G}_2(1)}, \ldots, \frac{1-\widetilde{G}_1(L-1)}{1-\widetilde{G}_2(L-1)}\right\}.$$

In other words, $\frac{1-\widetilde{G}_1(\ell)}{1-\widetilde{G}_2(\ell)}$ is the $(\ell+1)$th lowest value in the series $\left\{\frac{1-G_1(0)}{1-G_2(0)}, \frac{1-G_1(1)}{1-G_2(1)}, \ldots, \frac{1-G_1(L-1)}{1-G_2(L-1)}\right\}$. Note that the value of the objective in the solution to the problem

$$\max_{\alpha_0,\ldots,\alpha_{L-1} \in \left[\frac{1}{b}, \frac{1}{a}\right]} \frac{\sum_{\ell=0}^{L-1} \alpha_\ell [1-\widetilde{G}_1(\ell)]}{\sum_{\ell=0}^{L-1} \alpha_\ell [1-\widetilde{G}_2(\ell)]} \tag{7}$$

is identical to (**??**).

LEMMA 3. *There is a solution to* (**??**) *where there is a cutoff index $\hat{\ell}$ such that $\alpha_\ell^* = \frac{1}{b}$ for $\ell \leq \hat{\ell}$ and $\alpha_\ell^* = \frac{1}{a}$ for $\ell > \hat{\ell}$.*

PROOF. From Lemma **??**, there is a solution in which every $\alpha_k^* \in \{\frac{1}{b}, \frac{1}{a}\}$ for $0 \leq k \leq L-1$. So suppose, for contradiction, that there exist $\ell < \ell'$ such that $\alpha_\ell^* = \frac{1}{a}$ and $\alpha_{\ell'}^* = \frac{1}{b}$. Clearly, it must be the case that

$$\frac{1-\widetilde{G}_1(\ell)}{1-\widetilde{G}_2(\ell)} \geq \frac{\sum_{k \neq \ell} \alpha_k^*[1-\widetilde{G}_1(k)]}{\sum_{k \neq \ell} \alpha_k^*[1-\widetilde{G}_2(k)]}$$

as, otherwise, Observation 1 implies that $\alpha^*_\ell = \frac{1}{a}$ would not be a solution to (**??**). It then follows from Observation 2 that:

$$\frac{1-\widetilde{G}_1(\ell)}{1-\widetilde{G}_2(\ell)} \geq \frac{\sum_{k=0}^{L-1}\alpha^*_k[1-\widetilde{G}_1(k)]}{\sum_{k=0}^{L-1}\alpha^*_k[1-\widetilde{G}_2(k)]}.$$

In turn, this implies that

$$\frac{1-\widetilde{G}_1(\ell')}{1-\widetilde{G}_2(\ell')} \geq \frac{1-\widetilde{G}_1(\ell)}{1-\widetilde{G}_2(\ell)} \geq \frac{\sum_{k=0}^{L-1}\alpha^*_k[1-\widetilde{G}_1(k)]}{\sum_{k=0}^{L-1}\alpha^*_k[1-\widetilde{G}_2(k)]} \geq \frac{\sum_{k\neq\ell'}^{L-1}\alpha^*_k[1-\widetilde{G}_1(k)]}{\sum_{k\neq\ell'}^{L-1}\alpha^*_k[1-\widetilde{G}_2(k)]}$$

where the first inequality is a consequence of the fact that $\frac{1-\widetilde{G}_1(\cdot)}{1-\widetilde{G}_2(\cdot)}$ is sorted in increasing order and, the second and third inequalities are a consequence of Observation 2.

Finally, the above inequality implies that the value of the objective function in (**??**) is weakly increasing in $\alpha_{\ell'}$ (fixing all α^*_k with $k \neq \ell'$) so $\alpha^*_{\ell'} = \frac{1}{a}$ must also be a solution to (**??**).

This in turn implies that there is a solution to (**??**) with the requisite cutoff $\hat{\ell}$ in the statement of the lemma. ■

The previous two lemmas yield a simple algorithm to determine the solution to (**??**).

Algorithm:

1. Sort the series $\left\{\frac{1-G_1(0)}{1-G_2(0)}, \frac{1-G_1(1)}{1-G_2(1)}, \ldots, \frac{1-G_1(L-1)}{1-G_2(L-1)}\right\}$ in increasing order and set up optimization problem (**??**).
2. Evaluate the objective in (**??**) for all cutoffs $\hat{\ell}$ described in Lemma **??**. The cutoff that maximizes the value of the objective yields solution to (**??**).

This algorithm is computationally inexpensive: the time complexities of the first and second steps are $O(L\log(L))$ and $O(L)$ respectively.

REFERENCES

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

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

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013): "Inference on counterfactual distributions," *Econometrica*, 81, 2205–2268.

Deb, R. and L. Renou (2022): "Which wage distributions are consistent with statistical discrimination?" *CEPR Discussion Paper No. 17676*.

Neal, D. A. and W. R. Johnson (1996): "The role of premarket factors in black-white wage differences," *Journal of Political Economy*, 104, 869–895.

Shaked, M. and J. G. Shanthikumar (2007): *Stochastic orders*, Springer.

Stoer, J. and C. Witzgall (1970): "Convex Polyhedra," *Convexity and Optimization in Finite Dimensions I*, 31–81.