INTERNATIONAL ECONOMIC REVIEW
Vol. 44, No. 2, May 2003

THE EFFECT OF STATISTICAL DISCRIMINATION ON BLACK–WHITE WAGE INEQUALITY: ESTIMATING A MODEL WITH MULTIPLE EQUILIBRIA*

By Andrea Moro[1]

Department of Economics, University of Minnesota

This article presents the structural estimation of the parameters of a statistical discrimination model. Although the model is capable of displaying multiple equilibria, an estimation strategy that identifies both the model parameters and the equilibrium selected by the economic agents is developed and empirically implemented. A comparison between the selected equilibria and the other potential equilibria reveals that the decline in wage inequality experienced in the U.S. economy cannot be attributed to changes in the equilibrium selection. Nonetheless, a counterfactual experiment shows that in a color-blind society blacks' wage would have been on average more than 20% higher.

1. INTRODUCTION

In their survey of the empirical literature on racial wage inequality, Donohue and Heckman (1991) argue that measurable factors can account for at most 65% of the reduction in earnings inequality between black and white males.[2] Although better measurement of these factors could enhance their combined explanatory power, an additional explanation is that the unexplained portion of the decline reflects movements between different economy-wide equilibria. In an environment where multiple equilibria coexist under the same fundamentals, a reduction in inequality can be experienced if the economy moves from an equilibrium with high differentials to one with low differentials. Donohue and Heckman conjecture that such a phenomenon might have occurred in the past 30 years, perhaps as a

* Manuscript received June 2001; revised January 2002.

[1] I am deeply grateful to Kenneth Wolpin for his invaluable advice and encouragement. I also received help and useful comments from V. V. Chari, Stephen Coate, Russell Cooper, Zvi Eckstein, Arthur Goldberger, Gautam Gowrisankaran, Tom Holmes, John Kennan, Narayana Kocherlakota, Stephen Morris, Peter Norman, Petra Todd, and one anonymous referee. I also thank Juan Gomez for research assistance. Please address correspondence to: Andrea Moro, Department of Economics, University of Minnesota, 1035 Heller Hall, 271 19th Avenue South, Minneapolis, MN 55455, USA. Tel: +612 625-6353. Fax: +612 624-0209. E-mail: *amoro@atlas.socsci.umn.edu*.

[2] Among the reasons cited for the reduction in the differential are: convergence in years of schooling, convergence in quality of schooling, selective decline of the labor force participation of low-skilled blacks, migration of black workers out of southern regions, affirmative action, and other anti-discrimination legislation. Much of the debate has centered on the relative contributions of each of these explanations. See, for example Card and Krueger (1992), Heckman (1989), Leonard (1986), Smith and Welch (1984, 1989), and Neal and Johnson (1996).

result of the anti-discrimination policies.[3] But they did not provide a structure within which an analysis of this conjecture could be conducted.

This article provides such a structure by proposing a statistical discrimination model capable of displaying multiple outcomes. An estimation strategy that identifies both the fundamentals and the equilibrium chosen by the economic agents is developed and empirically implemented. By estimating the model parameters in different time periods, it is possible to compare the pattern of the equilibria selected in the economy over time with the other equilibria that the model could have generated under the same fundamentals. Results show that during the last 30 years, the economy always selected the equilibrium with the lowest wage differentials. Therefore, the decline in wage inequality cannot be explained by a change in the equilibrium selection.

The model builds on Arrow (1973) and is an extension of the model in Moro and Norman (2003b). Production takes place using two job-tasks of different complexity. Workers belong to two groups (blacks and whites), and face a costly human capital investment decision which, if undertaken, makes a worker productive in the complex task. Workers are heterogeneous in the cost of investment. Therefore, the fraction of workers who invest in human capital depends on how big the benefits from investing are. Firms decide how much to pay workers and how to allocate them between tasks on the basis of a noisy signal of productivity. Equilibria with wage inequality between groups can exist if employers hold asymmetric expectations about aggregate human capital investment of members of different groups. Workers from the group with less productive people will be expected to be less productive than workers from the other group with the same signal. Since, in a competitive environment, wages equal expected marginal products, this generates different incentives to invest. Therefore, one group will have fewer workers willing to undertake the costly human capital investment, fulfilling the employers' asymmetric expectations. It is possible that multiple equilibria coexist with different levels of investment disparity, which implies different levels of wage inequality and different incentives to invest in human capital.

Difficulties in estimating models with multiple equilibria arise because a unique map from the parameters to the likelihood of observable events does not exist (see Jovanovic, 1989). Economists have dealt with the multiplicity problem using different approaches. Dagsvik and Jovanovic (1994) postulate an equilibrium selection mechanism, and show that in some cases the parameters of the mechanism can be estimated together with the parameters of the model. Bresnahan and Reiss (1991) consider multiple equilibria as one event, and show how to predict the joint outcome. Tamer (2003) considers discrete games whose structure provides restrictions containing information about the nonunique outcomes. Brock and Durlauf (2001) present environments where the data contain information not only about the equilibrium, but also about all of the fundamentals, which can be consistently

[3] Theoretical studies have focused on the efficacy of anti-discrimination policies in reducing group disparities. They show that even if the effect of such policies is in general ambiguous, there are cases where such policies can eliminate the equilibria with discrimination and lead the economy towards group equality. See Lundberg (1991), Coate and Loury (1993), and Moro and Norman (2003a).

estimated from the data without imposing the equilibrium restrictions. These procedures assume that the investigator can observe data from different markets potentially generated by different equilibria. Using variability in the observables, it may be possible to recover the common fundamental parameters.

This article assumes, instead, that data are collected from only one market. All observations therefore derive from the same equilibrium. Another difference with respect to the previous literature is that the model in this article is nonstochastic, i.e., the set of equilibria consistent with a given value of the fundamentals does not depend on the realization of a random variable. A procedure is developed to jointly estimate both the fundamental parameters and the equilibrium chosen by the economic agents (which will be referred to as the "selected" equilibrium).

Equilibria are essentially described by a pair of wage distributions. The key feature here is that even if the fundamentals are consistent with multiple equilibria, each equilibrium corresponds to a unique pair of wage distributions. Under some circumstances (i.e., when some *identifying restrictions* apply), a pair of equilibrium wage distributions can be generated by only one value of the fundamentals. Therefore, such fundamentals can simply be computed by backing up which parameters that are consistent with the observed description of the wage distributions.

We show that the problem of determining which fundamental parameters are consistent with a given equilibrium is separable from the problem of determining which equilibrium is more likely to have generated a given set of observations (a pair of empirical wage distributions).[4] The separability implies that it is possible to proceed in two steps.

First, estimate which equilibrium is more likely to have generated the data observed by the investigator. While an equilibrium of the model is a complex object describing the strategy profile of each agent in the economy, a sufficient statistic for such strategy profiles is provided by the description of the wage distributions of the two groups of workers. Hence, the selected equilibrium is identified from the wage distributions of black and white workers that are uniquely characterized by a subset of fundamentals and by the sufficient statistics for workers and firms' equilibrium choices. These variables are treated in the first step as estimable parameters of the equilibrium wage distributions of black and white workers.

In the second step, restrictions are imposed such that a pair of equilibrium wage distributions can be generated by only one value of the fundamentals (despite multiple equilibria still being possible for the same value). In other words, the relation from the space of equilibria to the space of fundamentals is a function, despite its inverse being a correspondence. Applying such function, one can therefore estimate the fundamentals from the equilibrium estimated in the first step. The fundamentals recovered in the second step refer essentially to the distribution of workers over investment costs.

The model parameters have been estimated three times using U.S. male wage data from three 3-year intervals centered at 1965, 1980, and 1995. Results from the first estimation step show that despite its stylized nature, the model can well match the wage distributions. After completing the estimation, it is possible to

[4] This is implied by the nonstochastic nature of the model.

compute whether the model can generate other equilibria under the same set of fundamentals. Results show that whenever the model has multiple equilibria, the selected equilibrium always displays smaller differences in both human capital investment and average wage relative to the other potential equilibria. These findings imply that statistical discrimination and self-fulfilling expectations about groups' average productivity did not exacerbate wage differences in the United States, and that the decline in wage differentials cannot be explained by a change in equilibrium selection. Two possibilities consistent with these results are that prejudice (or "taste discrimination") may have decreased, or that the literature surveyed by Donohue and Heckman could not use good measurements of the relative black–white human capital improvements.

Even if the multiplicity of equilibria cannot account for wage inequality, statistical discrimination has large effects. A counterfactual experiment was performed to compute the equilibrium that the economy would display in a "color-blind society" where employers are not capable of distinguishing workers by their race. In such an environment blacks' average wage would have been 25.2% higher in 1995, and whites' average wage 2.2% lower.

2. THE MODEL

There is a competitive labor market where n firms compete for workers with heterogeneous skill. The firms' objective is to maximize output minus the wage bill.

There is a continuum of workers that belong to one of two identifiable groups, b or w of size λ^b and λ^w, respectively. Each worker must make an ex ante human capital investment decision. The decision is binary: either the worker invests in her human capital and becomes *qualified*, or the worker does not invest remaining *unqualified.* If a worker invests, she incurs cost c, whereas no cost is incurred otherwise. The cost of investment is distributed in each group according to the same distribution $g(c)$, $c \in [\underline{c}, \overline{c}]$. Denote the cumulative distribution of costs by $G(c)$. In addition, workers are heterogeneous because they are exogenously endowed with different skills. A worker with skill e is said to be endowed with "e efficiency-units".[5] The endowment of efficiency units depends deterministically on the investment cost according to $e^j(c)$, $j = b, w$, with $de^j(c)/dc < 0$ (i.e., workers with higher skill have lower investment cost).[6]

All agents are risk neutral, so the utility of a qualified worker earning wage w is $w - c$, whereas the utility of an unqualified worker is w.

To generate output, firms need workers performing two tasks, a *complex task* and a *simple task.*[7] Crucially, all workers employed in the simple task contribute

[5] This is the main difference between this model and the model in Moro and Norman (2003b). Without skill heterogeneity, all workers employed in the simple task would receive the same wage, which would make the estimation problematic.

[6] The assumption that groups have different functions relating skill endowment with investment cost is required for the econometric model to be exactly identified.

[7] The model can be extended to incorporate more tasks (and groups). The simplification adopted in this article is needed because of computational problems and data limitations.

to production, but only qualified workers contribute to production in the complex task. Formally, output is given by

$$y(C, S) = C^{\alpha} S^{1-\alpha} \tag{1}$$

where C and S denote the input of efficiency units of labor in the complex and simple task, respectively. Assume $C = \rho^b C^b + \rho^w C^w$ and $S = S^b + S^w$, where S^j represents the mass of efficiency units from all workers from group j employed in the simple task, and C^j represents the mass of efficiency units from *qualified* workers from group j employed in the complex task. Parameters $\rho^b$and ρ^w represent commonly known exogenous group differences in productivity in the complex task relative to productivity in the simple task.[8]

Employers are unable to observe whether a worker is qualified or not, but observe instead a signal $\theta \in [0, 1]$. Workers from group j have signals distributed according to density f_q^j if the worker is qualified and f_u^j otherwise. Let F_q^j and F_u^j denote the associated cumulative distributions and call this pair of distributions the "testing technology." Assume that both densities are continuously differentiable, and that $f_q^j(\theta)/f_u^j(\theta)$ is strictly increasing in θ. This monotone likelihood ratio property implies that the posterior probability that a worker from group j with signal θ is qualified given prior π^j

$$p^j(\theta, \pi^j) \equiv \frac{\pi^j f_q^j(\theta)}{\pi^j f_q^j(\theta) + (1 - \pi^j) f_u^j(\theta)} \tag{2}$$

is strictly increasing in θ, so qualified workers are more likely to get high signals. Since the testing technology is allowed to be different between groups, signals from workers of different groups can be differently informative even when groups are equal in all other aspects.

We now discuss four modeling assumptions: (1) Assuming that firms observe the outcome of only one test may seem extremely restrictive. However, one can always interpret signal θ as the aggregation of a battery of tests. (2) The assumption of incomplete information about productivity implies an informational asymmetry between tasks: the productivity of one efficiency unit is known with certainty only in the simple task. This is certainly not innocuous, but it seems quite reasonable on empirical grounds. Incomplete information also implies a potential mismatch: some workers without human capital investment will in general end up being employed in the complex task, and some qualified workers will be employed in the simple task. (3) It is important and crucial for the model that qualified and unqualified workers are not interpreted as workers with different quantum of acquired schooling, since this is observable. A more consistent way to interpret human capital in our model is to think of workers with the same level of schooling and interpret human capital as what was learned, which depends on costly effort.

[8] They can also represent a "taste" parameter representing biased preferences of the employers. In this interpretation employers care about both production and the race of their employees, and function (1) is a reduced form representation of the employers' utility function.

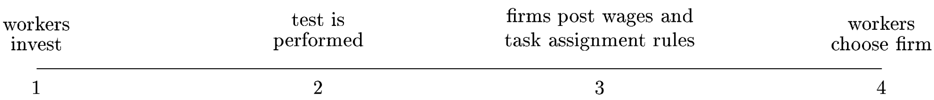

FIGURE 1

TIMELINE

One should think of observable components of human capital being part of the signals that firms observe. (4) It is not necessary to place a restriction guaranteeing that the marginal product of one efficiency unit of labor employed in the complex task is higher than the marginal product of one efficiency unit employed in the simple task. Since in equilibrium each efficiency unit is paid its expected marginal product, this must be true otherwise nobody would find it worthwhile to become qualified.

The timing is described in Figure 1. First, each worker decides whether or not to invest and is then assigned a (publicly observable) signal θ by nature. The firms then simultaneously announce wages and workers' allocation to tasks, which are allowed to depend on the signal and group identity. Formally, an action of firm i is to select some *wage schedule* $w_i^j{:}[0, 1] \rightarrow R_+$ and a *task assignment rule* $t_i^j{:}[0, 1] \rightarrow \{0, 1\}$ for each group j, with the interpretation that if $t_i^j(\theta) = 1(0)$ then firm i places j workers with signal θ in the complex (simple) task. Workers observe the posted wages and task assignment rules, and decide which firm to work for before payoffs are realized (randomizing whenever indifferent).

3. EQUILIBRIUM CHARACTERIZATION

We focus on *Nash Equilibria* of the model. A Nash Equilibrium consists of a pair of wage schedules w_i^j and task assignment rules t_i^j for each firm i, and for each type of worker (its cost c) a human capital investment decision and a choice of which firm to work for (a function of the worker's signal and of the firms' strategy profiles). The Nash equilibria of the game are computed by characterizing first the firms' best responses (in terms of wage schedules and task assignment rules) to workers' investment choices. These responses will determine a unique set of wage schedules given investment behavior by the workers; when optimal worker behavior is imposed, a set of fixed-point equations characterizing the set of equilibrium outcomes is derived.

First, notice that the wage schedule offered by each firm must be identical almost everywhere in any equilibrium. If one firm were to offer a higher wage to a positive mass of workers, then it could slightly lower wages offered to those workers. If the wage cut is small enough, those workers will not choose other firms and profits will increase. Because workers do not care directly what task they are assigned to, they are indifferent about which firm to work for; hence firms equally share workers on the equilibrium path.

Consider for now as given workers' aggregate investment summarized by $(\pi^b, \pi^w) \in [0, 1]^2$, the fractions of b and w workers that invest in human capital (such fractions will be derived in equilibrium at the end of this section). Human capital investment must obey a threshold rule: If a worker with cost c finds it convenient to acquire human capital, so will workers with lower cost. Hence, there must be thresholds B^b, B^w such that every worker from group $j = b, w$ invests in human capital if $c \leq B^j$, and does not invest if $c > B^j$. The thresholds must satisfy $\pi^j = \int_{c\leq B^j} g(c)\,dc$.

Factor inputs C^j, and S^j, $j = b, w$ can be written as follows:

$$C^j = \frac{\lambda^j}{n}\int_\theta t^j(\theta)\int_{c\leq B^j} e^j(c)g(c,\theta)\,dc\,d\theta \tag{3}$$

$$S^j = \frac{\lambda^j}{n}\int_\theta (1-t^j(\theta))\int_c e^j(c)g(c,\theta)\,dc\,d\theta \tag{4}$$

We will now use some equilibrium restrictions to simplify expressions (3) and (4).

Because f_q^j and f_u^j are independent of the cost of investment, the joint distribution of group j workers over costs and signals is

$$g^j(c,\theta) = \begin{cases} g(c)\,f_q^j(\theta) & \text{if the worker invested } (c \leq B^j) \\ g(c)\,f_u^j(\theta) & \text{if the worker did not invest } (c > B^j) \end{cases} \tag{5}$$

It is now possible to use (5) to rewrite (3) and (4) as

$$C^j = \frac{\lambda^j}{n}\int_\theta t^j(\theta)\,f_q^j(\theta)\int_{c\leq B^j} e^j(c)g(c)\,dc\,d\theta \tag{6}$$

$$\begin{aligned} S^j = \frac{\lambda^j}{n}\Bigg(&\int_\theta (1-t^j(\theta))\,f_q^j(\theta)\int_{c\leq B^j} e^j(c)g(c)\,dc\,d\theta \\ &+ \int_\theta (1-t^j(\theta))\,f_u^j(\theta)\int_{c>B^j} e^j(c)g(c)\,dc\,d\theta\Bigg) \end{aligned} \tag{7}$$

To further simplify these expressions, compute $\overline{e_q}^j$ and $\overline{e_u}^j$, respectively, as the average number of efficiency units carried by a qualified and an unqualified worker

$$\overline{e_q}^j = \frac{\int_{c\leq B^j} e^j(c)g(c)\,dc}{\pi^j} \tag{8}$$

$$\overline{e_u}^j = \frac{\int_{c>B^j} e^j(c)g(c)\,dc}{1-\pi^j} \tag{9}$$

Assuming a strong law of large numbers, then in group j there is a total of $\lambda^j\pi^j f_q^j(\theta)$ qualified workers and $\lambda^j(1-\pi^j)f_u^j(\theta)$ of unqualified workers with signal θ. Since (2) is increasing in the signal we focus on task assignment rules $t^j(\cdot)$ with a "threshold property," i.e., workers with signals above the threshold are assigned to the complex task and workers with lower signals are assigned to the simple task. Given threshold $\widetilde{\theta}^j$ the fraction of *qualified* workers with signals above the threshold in group j is $\pi^j(1 - F_q^j(\widetilde{\theta}^j))$ and the fraction of workers (qualified and unqualified) with signals below the threshold is $\pi^j F_q^j(\widetilde{\theta}^j) + (1 - \pi^j)F_u^j(\widetilde{\theta}^j)$. Using (8) and (9) together with the threshold task assignment rules we can finally obtain

$$C^j = \frac{\lambda^j}{n}\pi^j\big(1 - F_q^j(\widetilde{\theta}^j)\big)\overline{e_q}^j \tag{10}$$

$$S^j = \frac{\lambda^j}{n}\big(\pi^j F_q^j(\widetilde{\theta}^j)\overline{e_q}^j + (1 - \pi^j)F_u^j(\widetilde{\theta}^j)\overline{e_u}^j\big) \tag{11}$$

where $\widetilde{\theta}^j$ is the threshold used in the assignment of workers from group j. An interpretation of these equations is that C^j is the total number of *qualified* efficiency units employed by a representative firm in the complex task. This is equal to the proportion of qualified workers (from group j) employed in the complex task $\pi(1 - F_q^j(\widetilde{\theta}^j))$ multiplied by the average number of efficiency units carried by one qualified worker $\overline{e_q}^j$ and by the number of j workers employed by the representative firm λ^j/n. Similarly, S^j is the total number efficiency units from all workers employed in the simple task. This is the sum of the proportions of qualified and unqualified workers employed in the simple task, multiplied by the expected number of efficiency units carried by one qualified and one unqualified worker, respectively.

Since workers do not care about task assignment, the thresholds are the solution to the output maximization problem

$$\max_{(\theta^b,\theta^w)\in[0,1]^2} y(\rho^b C^b + \rho^w C^w, S^b + S^w) \tag{12}$$

Denote the first derivatives of the production function with $y_1(C, S) \equiv \frac{\partial y(C,S)}{\partial C}$, $y_2(C, S) \equiv \frac{\partial y(C,S)}{\partial S}$. The following proposition characterizes firms' best response to some arbitrary investment behavior. It states that wages are given by expected marginal products and job assignments are constrained efficient in any equilibrium.

PROPOSITION 1. *Suppose that fractions* (π^b, π^w) *of b and w workers invest and that* $\widetilde{\theta}^j(\pi^b, \pi^w)$, $j = b, w$ *is a solution to* (12). *Then firms' best response is to offer wages according to*

$$\begin{aligned} &w^j(\theta \mid \pi^b, \pi^w) \\ &= \begin{cases} y_1(C, S)\rho^j\overline{e_q}^j p^j(\theta, \pi^j) & \theta \geq \widetilde{\theta}^j(\pi^b, \pi^w) \\ y_2(C, S)\big(\overline{e_q}^j p^j(\theta, \pi^j) + \overline{e_u}^j(1 - p^j(\theta, \pi^j))\big) & \theta < \widetilde{\theta}^j(\pi^b, \pi^w) \end{cases} \qquad j = b, w \end{aligned} \tag{13}$$

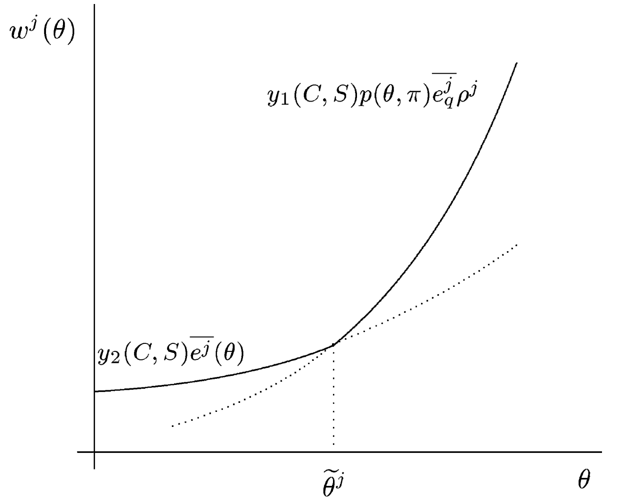

FIGURE 2

THE WAGE FUNCTION

and to assign workers from group $j = b, w$ to the complex task if and only if $\theta \geq \widetilde{\theta}^j(\pi^b, \pi^w)$. Moreover, in any equilibrium where fractions (π^b, π^w) invest, the wage schedules posted by each firm i must agree with (13) for almost all $\theta \in [0, 1]$.

The proof is given in the Appendix. The monotone likelihood ratio property implies that $w^j(\cdot)$ is strictly increasing in θ. The wage function implies that each worker is paid her expected marginal product (see Figure 2). To interpret the expression, note that in the complex task the wage is equal to the marginal product of one efficiency unit of labor, $y_1(C, S)\rho^j$, multiplied by the average number of efficiency units carried by a productive worker $\overline{e_q}^j$, and by the probability of a worker being qualified for the complex task given her group's aggregate investment behavior $p^j(\theta,\pi^j)$. Wages in the simple task are equal to the marginal product y_2 multiplied by the average number of efficiency units carried by a worker with signal θ. This is the weighted average of the expected number of efficiency units of a qualified and unqualified worker, weighted by the probability of being, respectively, qualified and unqualified.

Let η_1^j and η_2^j be the multipliers associated with the boundaries for θ^j. After some manipulations, the Kuhn–Tucker conditions of (12) can be expressed as

$$(14) \qquad -y_1(C, S)\rho^j\overline{e_q}^j\pi^j f_q^j(\theta^j) + y_2(C, S)\big(\overline{e_q}^j\pi^j f_q^j(\theta^j) + \overline{e_u}^j(1-\pi^j) f_u^j(\theta^j)\big) + \big(\eta_1^j - \eta_2^j\big) = 0$$

for $j = b, w$, together with complementary slackness conditions. This condition implies that in an *interior* solution the marginal product of workers with a signal equal to the threshold must be the same in both tasks. However, there may be corner solutions where all workers in one group are assigned to one task only. Because both factors are essential, it cannot be the case that all workers in the economy are assigned to the same task.

We now impose optimal worker behavior to derive the full equilibrium of the model. Denote $B^j(\pi^b, \pi^w)$ to be the gross benefits of investing for a worker belonging to group j when wages are computed in the labor-market equilibrium assuming investment equal to (π^b, π^w). Then,

$$B^j(\pi^b, \pi^w) = \int_\theta w^j(\theta \mid \pi^b, \pi^w)\big(f_q^j(\theta) - f_u^j(\theta)\big)\, d\theta \tag{15}$$

In equilibrium, workers invest if and only if the cost of investment is less than the benefits, so $G(B^j(\pi^b, \pi^w))$ is the best response fraction of investors in group j and equilibria of the model are characterized by the solutions to a set of two fixed point equations.

PROPOSITION 2. *In any equilibrium, a pair of a fraction of investors* (π^b, π^w) *solves*

$$\begin{aligned} \pi^b &= G(B^b(\pi^b, \pi^w)) \\ \pi^w &= G(B^w(\pi^b, \pi^w)) \end{aligned} \tag{16}$$

Moreover, any solution to (16) corresponds to an equilibrium of the model.

The system of equations (16) implies that $B^j(\pi^b, \pi^w) = B^j$. Existence of equilibria is verified by checking continuity of $B^j(\cdot)$. If groups do not possess exogenous differences (i.e., if they have the same testing technology, $\rho^b = \rho^w$, and $e^b(\cdot) = e^w(\cdot)$), then there always exists a symmetric equilibrium. When groups are not identical, then equilibria cannot in general be symmetric: when $\pi^b = \pi^w$, marginal products are different because of the exogenous differences (e.g., a different testing technology); therefore, incentives are different as well.

4. ESTIMATION

This section develops a procedure capable of estimating both the fundamental parameters and the "selected equilibrium" using a cross section of wage data from black and white workers.[9] Since the model is nonstochastic, the only source of noise is sampling error. The data are assumed to have been generated by the same equilibrium. The procedure therefore differs from other techniques that use data potentially generated by different equilibria.

Denote with Φ the parameter space, and with Ξ the space containing all feasible equilibria (for any vector of parameters in Φ). A vector $\xi \in \Xi$ contains sufficient

[9] Data about *observed* human capital investment and job assignment are also available. However, it is not clear how to interpret such information in the context of the model because the investigator does not observe the same signal that employers are assumed to observe, and does not know exactly which jobs correspond to a simple or a complex task. Such information will therefore be ignored.

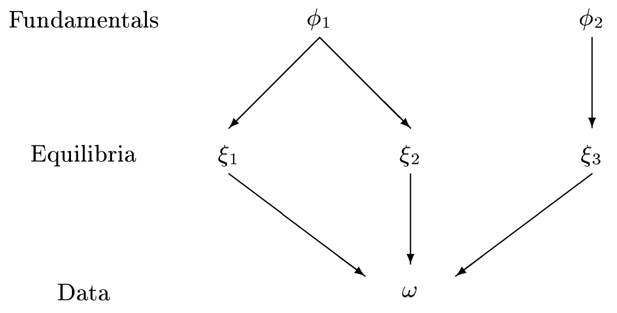

FIGURE 3

A SCHEMATIC REPRESENTATION OF THE ECONOMETRIC MODEL

statistics for the description of an equilibrium. Note that in some models a description of the equilibrium may contain the value of some fundamental parameters, in which cases Ξ and Φ may have some elements in common.[10] Finally, denote with $\Xi_\phi \subseteq \Xi$ the set of equilibria consistent with parameters $\phi \in \Phi$ and let ω be a realization of the data from one equilibrium.

When multiple equilibria are present, the likelihood $L(\omega \mid \phi)$ is a correspondence, but $L(\omega \mid \phi, \xi)$ is a function. Hence, it is possible to estimate ϕ using

$$\max_{\phi \in \Phi, \xi \in \Xi} L(\omega \mid \phi, \xi) = \max_{\phi \in \Phi} \left\{ \max_{\xi \in \Xi_\phi} L(\omega \mid \phi, \xi) \right\} \tag{17}$$

The expression on the right-hand side suggests the following algorithm. Compute the set of all equilibria Ξ_ϕ that can be generated when the vector of parameters is ϕ, compute the likelihood for each of these equilibria, $L(\omega \mid \phi, \xi)$, choose the equilibrium that maximizes the likelihood, and then maximize again by iterating over the set of fundamental parameters.[11]

Identification issues aside, this exercise is computationally expensive and prone to failure given that numerical procedures may sometimes miss the computation of one or more of the equilibria. This is important because in some circumstances it is not known a priori how many equilibria the model displays; moreover, the number of equilibria may change depending on the fundamentals. For example, refer to Figure 3 and assume that the "true" fundamentals are ϕ_1, the "true" equilibrium chosen in the economy is ξ_1, and that the investigator observes some data ω. Adopt the estimation procedure just described and suppose that the numerical procedure fails to compute ξ_1 from parameters ϕ_1. The procedure is, instead, successful in

[10] This is the case in our model: Firms' equilibrium strategy requires a description of the wage schedule that is also a function of the parameters of the testing technology (a subset of the fundamentals). In different models the equilibrium could be described without reference to the fundamentals. For example, a 2×2 game where two players choose between left (L) and right (R), the equilibrium is described by one of the strategy profiles in {LL, LR, RL, RR}, whereas the fundamentals of the game include a list of payoffs for each outcome and player.

[11] This is essentially the strategy adopted by Bisin et al. (2002) to write a method of moments estimator of a cultural transmission model.

computing the other equilibria. Then, it is possible that $L(\omega \mid \phi_2, \xi_3) > L(\omega \mid \phi_1, \xi_2)$, in which case the investigator would conclude that the true parameters are most likely to be ϕ_2.

This article presents an *equivalent* "two-step procedure" that does not require the computation of all the feasible equilibria. The procedure exploits the separability between the equilibrium determination and the data-generating process, and is a direct consequence of the nonstochastic nature of the model. This implies that the fundamental parameters (or, whenever Φ and Ξ have common elements, the fundamentals that are not part of a description of the equilibrium) only affect those that are the feasible equilibria, not how data are generated from such equilibria. Therefore, one can write

$$L(\omega \mid \phi, \xi) = \begin{cases} L(\omega \mid \xi) & \text{if } \xi \in \Xi_\phi \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

and rewrite the likelihood maximization problem as

$$\max_{\Phi, \Xi} L(\omega \mid \phi, \xi) = \max_{\xi} L(\omega \mid \xi) \quad \text{subject to } \xi \in \Xi_\phi \tag{19}$$

Although problem (19) is equivalent to (17), the following proposition shows that under some assumptions it can be solved by a less computationally expensive procedure. The same assumptions guarantee that both the parameters and the equilibrium are identified.

PROPOSITION 3. *If (a) $L(\omega \mid \xi)$ has a unique maximizer for all $\xi \in \Xi$, and (b) the relation $r: \Xi \to \Phi$ where $r(\xi) \equiv \{\phi \in \Phi : \xi \in \Xi_\phi\}$ is single valued and nonempty for all $\xi \in \Xi$, then problem (19) has a unique solution that can be computed using the following two-step procedure:*

First step. *Estimate the equilibrium using* $\widehat{\xi} = \arg\max_{\xi \in \Xi} L(\omega|\xi)$.
Second step. *Compute the fundamentals from* $\widehat{\phi} = r(\widehat{\xi})$.

The first step ignores the equilibrium restrictions $\xi \in \Xi_\phi$ and estimates the equilibrium in an "unrestricted" model. The second step imposes the equilibrium restrictions to recover the fundamental parameters and exploits assumption (b) imposing that the inverse of the equilibrium map is a function (even though the map itself may be a correspondence). The proof follows observing that assumption (a) means that the data uniquely identify the parameters describing the equilibrium. This is a necessary condition for the identification of the equilibrium that must hold even if a different procedure is used and regardless of the existence of multiple equilibria. It is not difficult to find models where this assumption does not hold; for example, all models where equilibria are observationally equivalent.[12]

[12] For example, consider a 2×2 game with identical payoffs equal to θ on the diagonal and 0 payoff off the diagonal. Assume that the investigator observes (with error) several payoffs' realization of the same equilibrium. It is not possible to recover which equilibrium was played.

Condition (b) means that one and only one value of the fundamentals can generate each equilibrium ξ. This not only guarantees identification of the fundamentals, but suggests that once the equilibrium has been estimated in Step 1, it is possible to use function r to obtain the unique value of the parameters that have generated the equilibrium. Clearly, if $\Phi$and Ξ have common elements, the corresponding elements of r contain the identity function. Condition (b) imposes a large amount of structure, requiring the dimensionality of the space of equilibria to be as large as the space of parameters. Examples where the set of all possible equilibria is finite (such as the discrete games analyzed in Tamer, 2003) cannot satisfy (b) if the space of parameters is a continuum.

In our model, an equilibrium is defined by a strategy profile that uniquely determines a pair of wage distributions of black and white workers. Even if the fundamentals may be consistent with multiple equilibria, a vector of parameters ϕ describing a pair of wage distributions is a sufficient statistic of the equilibrium strategy profile. This is because the wage distribution is a function of (1) the firms' strategy profile (summarized by the wage schedule and by the task assignment threshold rule), (2) the agents' aggregate investment decision (summarized by the fraction of investors), and (3) the parameters describing the testing technology. Subsection 4.1 derives the vector of parameters ξ describing the equilibrium and the likelihood $L(\omega \mid \xi)$.

The first step in the estimation consists in finding which values of these parameters best match the empirical wage distribution of each group. This article does not provide a formal proof that condition (a) in Proposition 3 holds, but some intuition can be gained from the simulations in Subsection 4.4.

Subsection 4.2 shows that the restrictions satisfying condition (b) in Proposition 3 are fulfilled when the distribution of investment costs g is uniform, and the functions relating skill endowment with the cost of investment e^j are linear. The results reported in Subsection 5.2 show that such assumptions do not restrict the equilibrium to be unique.

4.1. *First Step: The Wage Distribution.* This subsection shows how the wage distributions of groups b, w can be parameterized as functions of 10 parameters ξ to compute the likelihood $L(\omega \mid \xi)$, where $\omega = (\omega^b, \omega^w)$ is a vector containing wage observations from workers that belong to groups b and w. It is shown below that the likelihood is separable, $L(\omega \mid \xi) = L(\omega^b \mid \xi^b) \cdot L(\omega^w \mid \xi^w)$, so that parameter vectors ξ^b and ξ^w can be separately estimated using wage observations from black and white workers, respectively.

At this stage it is sufficient to express a functional form for the testing technology. A simple one-parameter exponential family is adopted:

$$
\begin{aligned}
f_q^j(\theta) &= \gamma^j \theta^{(\gamma^j - 1)} \\
f_u^j(\theta) &= \gamma^j (1-\theta)^{\gamma^j - 1}
\end{aligned}
\tag{20}
$$

for $j = b, w$. With this functional form the monotone likelihood ratio property is satisfied for any $\gamma^j > 1$.[13] Parameter γ^j can be interpreted as measuring the "informativeness" of the test result: as γ^j becomes larger qualified (unqualified) workers are more likely to receive a test result "close" to one (zero).

Define

$$
\begin{aligned}
y_q^j &\equiv y_2(C, S)\overline{e_q}^j \\
y_u^j &\equiv y_2(C, S)\overline{e_u}^j
\end{aligned}
\tag{21}
$$

The wage distribution of group j will be characterized as a function of $\xi^j \equiv \{\gamma^j, \pi^j, \widetilde{\theta}^j, y_q^j, y_u^j\}$, where $\widetilde{\theta}^j$ is the threshold signal used in the task assignment rule. Among the five parameters, only γ^j is an element of the fundamentals, whereas the others are *equilibrium variables,* treated here as *estimable parameters* of the wage distribution (i.e., we ignore, at this stage, the equilibrium restrictions between such variables and the full set of fundamentals).

It is now possible to write the distribution $f^j(\theta;\pi)$ of test results θ for members of group j as a function of γ^j

$$
\theta \sim f(\theta;\pi^j,\gamma^j) = \pi^j\gamma^j\theta^{(\gamma^j-1)} + (1-\pi^j)\gamma^j(1-\theta)^{\gamma^j-1}
\tag{22}
$$

Consider the wage function (13). Continuity at $\widetilde{\theta}^j$ implies

$$
\begin{aligned}
&y_1(C, S)\rho^j\overline{e_q}^j p(\widetilde{\theta}^j, \pi^j) = y_2(C, S)\overline{e_q}^j p(\widetilde{\theta}^j, \pi^j) + y_2(C, S)\overline{e_u}^j(1 - p(\widetilde{\theta}^j, \pi^j)) \\
&\qquad \Rightarrow y_1(C, S)\rho^j\overline{e_q}^j = y_q^j + y_u^j\frac{(1-\pi^j) f_u(\widetilde{\theta}^j)}{\pi^j f_q(\widetilde{\theta}^j)}.
\end{aligned}
\tag{23}
$$

Using (20), rewrite the wage schedule as a function of ξ^j

$$
w(\theta \mid \xi^j) =
\begin{cases}
\left(y_q^j + y_u^j\dfrac{1-\pi^j}{\pi^j}\left(\dfrac{1-\widetilde{\theta}^j}{\widetilde{\theta}^j}\right)^{\gamma^j-1}\right)\dfrac{\pi^j\theta^{(\gamma^j-1)}}{\pi^j\theta^{(\gamma^j-1)} + (1-\pi^j)(1-\theta)^{\gamma^j-1}} & \text{if } \theta \geq \widetilde{\theta}^j \\[2ex]
y_q^j\dfrac{\pi^j\theta^{(\gamma^j-1)}}{\pi^j\theta^{(\gamma^j-1)} + (1-\pi^j)(1-\theta)^{\gamma^j-1}} + y_u^j\dfrac{(1-\pi^j)(1-\theta)^{\gamma^j-1}}{\pi^j\theta^{(\gamma^j-1)} + (1-\pi^j)(1-\theta)^{\gamma^j-1}} & \text{otherwise}
\end{cases}
\tag{24}
$$

[13] This assumption is necessary for the monotone likelihood ratio property to hold. When $\gamma < 1$ the roles of f_q and f_u are reversed: The lower the signal, the lower the probability that the worker is qualified.

The data ω^j are a vector of wages $(\omega_1, \ldots, \omega_{N^j})$ with corresponding weights $(k_1, \ldots, k_{N^j})$. Assume for simplicity that the vector ω^j contains distinct wage observations sorted in an increasing order; ω_{N^j} is typically a "topcode" wage level. Denote by $w^{-1}(\cdot|\xi^j) : R_+ \to [0, 1]$ the inverse wage function associating with any wage level ω its corresponding signal θ. Given that θ is distributed in the population according to $f(\theta \mid \pi^j, \gamma^j)$, apply a transformation of variables to compute the wage distribution x^j

$$x^j(\omega \mid \xi^j) = f(w^{-1}(\omega \mid \xi^j)|\pi^j, \gamma^j)\frac{dw^{-1}(\omega \mid \xi^j)}{d\omega} \tag{25}$$

where $dw^{-1}(\omega \mid \xi^j)/d\omega$ can be computed analytically using $(dw(\theta)/d\theta)^{-1}$ evaluated at $\theta = w^{-1}(\omega \mid \xi^j)$. At this stage it is therefore possible to derive the likelihood as a function of ξ^j.

To simplify the numerical computation, two restrictions from the model allow us to derive parameters y_q^j and y_u^j and therefore to express the likelihood as a function of γ^j, π^j, and $\widetilde{\theta}^j$ only.

First, note that the model predicts a minimum wage: evaluating (24) at $\theta = 0$ we obtain $w^j(0 \mid \xi^j) = y_u^j$. It is therefore possible to set $\widehat{y_u}^j = \omega_1$.[14] Secondly, the mass of topcoded observations can be used to estimate y_q^j. Given γ^j, π^j, and $\tilde{\theta}^j$ let θ_{N^j} be the signal of the worker that earns the topcode wage level ω_N and compute $\widehat{\theta_{N^j}}$ as the unique root of the equation $\int_{\theta_{N^j}}^1 f(\theta_{N^j} \mid \pi^j, \gamma^j) = k_{N^j} / \sum_{i=1}^{N^j} k_i$. Then, compute $\widehat{y_q}^j$ as the solution to the following equation:[15]

$$\omega_{N^j} = \left(y_q^j + y_u^j \frac{1-\pi^j}{\pi^j}\left(\frac{1-\widetilde{\theta}^j}{\widetilde{\theta}^j}\right)^{(\gamma^j-1)}\right)\frac{\pi^j \widehat{\theta_{N^j}}^{\gamma^j-1}}{\pi^j \widehat{\theta_{N^j}}^{\gamma^j-1} + (1-\pi^j)(1-\widehat{\theta_{N^j}})^{\gamma^j-1}} \tag{26}$$

The right-hand side of (26) is strictly increasing in y_q^j; hence a unique solution is guaranteed. Since the first and last observations are used to get y_u^j and y_q^j, the likelihood of the rest of the sample from group j can be computed using observations from *2* to $N^j - 1$. Substituting (22) in (25) the log likelihood can be written as

$$\begin{aligned} &\log L\big(\omega^j, k^j \mid \widetilde{\theta}^j, \pi^j, \gamma^j; \widehat{y_q}^j, \widehat{y_u}^j\big) \\ &\quad = \sum_{i=2}^{N^j-1}\Bigg(\log\big(\pi^j\gamma^j\big(w^{-1}(\omega_i \mid \xi^j)\big)^{\gamma^j-1} + (1-\pi^j)\gamma^j\big(1-w^{-1}(\omega_i \mid \xi^j)\big)^{\gamma^j-1}\big) \\ &\qquad\qquad + \log\left(\frac{dw^{-1}(\omega \mid \xi^j)}{d\omega}\right)\Bigg)k_i \end{aligned} \tag{27}$$

[14] Denote with carets all estimated parameters.

[15] This is true if we can assume that ω_N has been earned by an individual employed in the complex task. One must always verify this by checking that the maximum likelihood obtained by a model where everybody is employed in the simple task is less than the likelihood obtained with the more general model.

Parameters $\widetilde{\theta}^j, \pi^j, \gamma^j$ can be consistently estimated by maximizing the likelihood function (27). The toughest computational task of calculating the likelihood is to compute signals $w^{-1}(\omega_i \mid \xi^j)$ because this requires numerically inverting the wage function at each wage observation. The derivative $dw^{-1}(\omega \mid \xi^j)/d\omega_i = 1/(dw^j(\theta_i \mid \xi^j)/d\theta_i)$ can be computed analytically.[16]

We are now provided with the relevant information characterizing group j's wage distribution. An estimate of the gross returns to investment in human capital $\widehat{B^j}$ can be computed using equation (15). At this stage nothing guarantees that these estimates are consistent with the full equilibrium of the model. The first step estimates consider only the restrictions imposed by the "labor-market equilibrium" and ignore the full equilibrium restrictions between π^j, $\widetilde{\theta}^j, \gamma^j, y_q^j, y_u^j$, and the parameters of G (the distribution over human capital investment cost), e^j (the skill-cost function), and the production function y. Such restrictions are implied by the system of fixed point equations (16).

4.2. *Second Step: The Production Function, The Cost Distribution, and the Skill Endowment.* The second step consists of using equilibrium moment restrictions derived from the model to obtain the parameters of the production function, of the cost distribution $g(\cdot)$, and skill endowment functions $e^b(\cdot)$, $e^w(\cdot)$.

Several approaches guarantee identification of these functions. The choice adopted in this article is to estimate the model using a cross section of wage observation from the U.S. labor market. The estimations of the two groups' wage distributions obtained in the first step can be combined in order to match some moment restrictions. This requires the identifying assumptions that black and whites have the same, uniform distribution over cost of investment $g(c)$, and that the skill endowment functions $e^b(\cdot)$, $e^w(\cdot)$ are linear.[17]

Whereas this procedure might result in economically meaningless parameters (in a sense that will be described below), when meaningful parameters are obtained the estimated moments will exactly match the theoretical counterpart. This will allow us to make interesting comparisons between the estimated "selected" equilibria and the other equilibria that the model can display.

A. *The production function and moments of skill endowment.* The estimation results obtained from black and white wage distributions are merged to derive parameters of the production functions and moments of the skill endowment. This step does not require any additional parametric assumption. The restrictions that will now be used are: the defining equations (21), continuity of the wage function at $\widetilde{\theta}^j$ (23), and the expression for marginal products derived from the

[16] The algebraic formula is omitted, but is reported in Moro (2002).

[17] Instead of using first-step estimates from the two groups of workers one could alternatively use estimations obtained from two or more time periods, or from different markets. In the first case, as will be explained below, identification requires assuming that cost of investment does not change over time. In the second case the requirement is that the investment cost does not depend on market location.

Cobb–Douglas production function

$$
\begin{aligned}
y_u^j &= y_2(C, S)\overline{e_u}^j && j = b, w \\
y_q^j &= y_2(C, S)\overline{e_q}^j && j = b, w \\
y_1\rho^j(C, S)\overline{e_q}^j &= y_q^j + y_u^j \frac{1-\pi^j}{\pi^j}\left(\frac{1-\widetilde{\theta}^j}{\widetilde{\theta}^j}\right)^{(\gamma^j-1)} && j = b, w \qquad (28)\\
y_1 &= \alpha C^{\alpha-1} S^{1-\alpha} \\
y_2 &= (1-\alpha) C^{\alpha} S^{-\alpha}
\end{aligned}
$$

where $C^j = \lambda^j\pi^j\rho^j(1 - F_q^j(\widetilde{\theta}^j))\overline{e_q}^j$ and $S^j = \lambda^j(\pi^j F_q^j(\widetilde{\theta}^j)\overline{e_q}^j + ((1-\pi^j)\times F_u^j(\widetilde{\theta}^j)\overline{e_u}^j)$. Substituting the estimated values of $\widetilde{\theta}^j, \pi^j$, and γ^j, $j = b, w$, and values of working population sizes λ^b, λ^w obtainable from aggregate employment data[18] such restrictions imply eight equations in nine unknowns (the four moments of the skill endowment $\overline{e_q^w}, \overline{e_q^b}, \overline{e_u^w}, \overline{e_u^b}$, the parameters of the production function ρ^w, ρ^b, α, and marginal products y_1 and y_2). Rescale the model imposing $\rho^w = 1$ (this corresponds to assuming that there is a comparative (dis)advantage in employing productive black workers in a complex task relative to white workers). The set of equations has now a unique solution denoted by $\widehat{\overline{e_q^w}}, \widehat{\overline{e_q^b}}, \widehat{\overline{e_u^w}}, \widehat{\overline{e_u^b}}, \widehat{\rho^b}, \widehat{\alpha}, \widehat{y_1}, \widehat{y_2}$.[19] These estimated parameters inherit the consistency property of the estimation of π^j, θ^j, and γ^j because they satisfy the restrictions in (28) with equality.

B. *The cost of investment distribution.* The theoretical model implies that the equilibrium fraction of investors π^j is equal to the set of workers with cost less than the benefits from investment B^J (see 16):

$$\pi^b = G(B^b) \tag{29}$$

$$\pi^w = G(B^w) \tag{30}$$

The first step delivers estimations of the proportion of people who choose to invest in each group $\widehat{\pi}^b, \widehat{\pi}^w$, and of the benefits from investment $\widehat{B}^b$, $\widehat{B}^w$ computed by substituting the estimated parameters to function (15). It is therefore possible to exactly match the equilibrium moment restrictions (29)–(30) by assuming a linear c.d.f. $G(c) = p_1 + p_2 \times c$. The parameter estimates are computed as follows:

$$
\begin{aligned}
\widehat{p_2} &= \frac{\widehat{\pi}^w - \widehat{\pi}^b}{\widehat{B}^w - \widehat{B}^b} \\
\widehat{p_1} &= \widehat{\pi}^b - \widehat{p_2} \times \widehat{B}^b
\end{aligned} \tag{31}
$$

[18] Constant returns to scale imply that all that matters is the ratio λ^b/λ^w.

[19] Details of the computation of such parameters are available in Moro (2002).

Using $G(\underline{c})=0$ and $G(\overline{c})=1$ compute the estimated bounds of the cost distribution as $\widehat{\underline{c}}=-\widehat{p1}/\widehat{p2}$ and $\widehat{\overline{c}}=(1-\widehat{p1})/\widehat{p2}$. It is important to note that the estimation of p_2 may yield $\widehat{p}_2<0$, which is economically meaningless being G a c.d.f. (see Section 4.3).

C. *The cost-skill function.* In equilibrium, the relationship between ability and cost $e^j(c)$ must satisfy the following moment restrictions:

$$\overline{e_q^w}=\frac{\int_{c<B^w} e^j(c)g(c)\,dc}{\pi^w} \tag{32}$$

$$\overline{e_q^b}=\frac{\int_{c<B^b} e^j(c)g(c)\,dc}{\pi^b} \tag{33}$$

$$\overline{e_u^w}=\frac{\int_{c\geq B^w} e^j(c)g(c)\,dc}{1-\pi^w} \tag{34}$$

$$\overline{e_u^b}=\frac{\int_{c\geq B^b} e^j(c)g(c)\,dc}{1-\pi^b} \tag{35}$$

Assume a linear relationship as follows: $e^j(c)=p_3^j+p_4^j\times c$, $j=b,w$. Then, using the estimated values for B^b, B^w, $\overline{e_q^w}$, $\overline{e_q^b}$, $\overline{e_u^w}$, $\overline{e_u^b}$, and the estimates of the parameters of $g(c)$, these restrictions become four linear equations in four unknowns, with a unique solution. Linearity of g ensures that moments in (32)–(35) can be expressed as simple averages,

$$\begin{aligned}
\overline{e_q}^j &= \frac{p_2p_3^j(B^j-\underline{c})+p_2p_4^j\frac{(B^j)^2-\underline{c}^2}{2}}{\pi^j}\\
&= \frac{p_2(B^j-\underline{c})\left(p_3^j+p_4^j\frac{B^j+\underline{c}}{2}\right)}{\pi^j}=p_3^j+p_4^j\left(\frac{B^j+\underline{c}}{2}\right)\\
\overline{e_u}^j &= \frac{p_2p_3^j(\overline{c}-B^j)+p_2p_4^j\frac{\overline{c}^2-(B^j)^2}{2}}{1-\pi^j}\\
&= \frac{p_2(\overline{c}-B^j)\left(p_3^j+p_4^j\frac{B^j+\overline{c}}{2}\right)}{1-\pi^j}=p_3^j+p_4^j\left(\frac{B^j+\overline{c}}{2}\right)
\end{aligned}\qquad j=b,w. \tag{36}$$

where the last equality in the expressions follows, since $p_2(B^j-\underline{c})=\int_{c\leq B^j} g(c)\,dc=\pi^j$ and $p_2(\overline{c}-B^j)=\int_{c>B^j} g(c)\,dc=1-\pi^j$. The equations say that $\overline{e_q}^j$ is equal to the efficiency units carried by the worker with cost equal to the average of $\underline{c}$ and B^j, whereas $\overline{e_u}^j$ is equal to the number of efficiency units carried by the worker with cost equal to the average between $\overline{c}$ and B^j.

From the pair of sets of two equations in (36) compute

$$
\begin{aligned}
\widehat{p_4^j} &= \frac{\widehat{e_q}^j - \widehat{e_u}^j}{\frac{\widehat{B^j}+\widehat{\underline{c}}}{2} - \frac{\widehat{B^j}+\widehat{\overline{c}}}{2}} = 2\frac{\widehat{e_q}^j - \widehat{e_u}^j}{\widehat{\underline{c}} - \widehat{\overline{c}}} \\
\widehat{p_3^j} &= \widehat{e_q}^j - \widehat{p}_4\left(\frac{\widehat{B^j}+\widehat{\underline{c}}}{2}\right)
\end{aligned}
\tag{37}
$$

The estimated parameters are meaningful only if $\widehat{p_4^j} < 0,\ j = b, w$.

4.3. *Can the Estimation Results Invalidate the Model?* Consider the parameters estimated in the first stage. The theoretical model requires $\widetilde{\theta}^j$ to be strictly between 0 and 1 for at least one group $j = b, w$. When this is the case the wage function displays a kink at the threshold signal, which implies a discontinuity of the wage distribution at $w(\widetilde{\theta}^j)$. A data set without such a feature would generate an estimated value of the threshold equal to zero or one. If this were the case, then the assumption of a two-task production function would be refuted.[20]

The estimation procedure restricts the other first-stage parameter estimates to produce results that satisfy restrictions from the theory. The parameter that defines the precision of the testing technology γ^j is restricted to be greater than one to ensure that high values of the signal θ correspond to more productive workers.[21] Moreover, the structure of the model forces the estimate of π^j to be within the unit interval: $\pi^j < 0$ implies that wages are decreasing in θ; similarly, $\pi^j > 1$ is not consistent with a positive minimum wage.

In two phases of the second stage the estimation may fail to produce economically meaningful results. First, the strategy proposed in this article obtains a meaningful $\widehat{p_2} > 0$ only if the estimated parameters in the first step satisfy

$$
\begin{array}{ccc}
\widehat{\pi}^b < \widehat{\pi}^w & \text{and} & \widehat{B}^b < \widehat{B}^w \\
 & \text{or} & \\
\widehat{\pi}^b > \widehat{\pi}^w & \text{and} & \widehat{B}^b > \widehat{B}^w
\end{array}
\tag{38}
$$

If one of conditions (38) is not satisfied, then the data do not support the assumption that the two groups have an identical cost distribution, and without additional information the parameters of such distributions cannot be separately identified. Because the wage distributions of the two groups are separately estimated in the first stage, nothing guarantees that the estimated parameters satisfy (38).

Secondly, theory also restricts workers with more costly investment to be endowed with fewer efficiency units, which implies that the estimated cost-skill

[20] All of the estimation results performed for this article consistently yield a threshold strictly within the unit interval. I am not aware of other studies focusing on the discontinuity of the wage distribution that these findings imply.

[21] Alternatively, one could estimate a model imposing $\gamma^j < 1$ in which case high values of the signal would correspond to less productive workers.

function has a meaningful interpretation only if $\widehat{p_4^j} < 0$. If $\widehat{p_2} > 0$ then from (37) observe that $\widehat{p_4^j} < 0$ only if $\widehat{e_q}^j > \widehat{e_u}^j$. Again, nothing guarantees that the parameters estimated in the first stage satisfy this restriction. If they do not, the data do not support the assumption that lower-cost workers carry higher ability.

4.4. *How do Parameters Affect the Wage Distribution?* Although local identification in the first step can be proved by checking that the likelihood is not flat at the estimated maximum, global identification should be proved by showing that the likelihood function has a unique maximizer. Given the complexity of the likelihood, which requires the numerical computation of the inverse of the wage function, this is a formidable task. However, some intuition can be gained by showing that the estimated parameters differently affect the wage distribution.

The top-left graph in Figure 4 illustrates the wage distribution using the parameters estimated from white male 1995 data. Each of the other three graphs shows how wage distribution is affected by varying one different parameter. The graphs suggest that parameter π mostly affects the skewness of the wage distribution, whereas γ mostly affects its concavity. Parameter $\widetilde{\theta}$ mainly affects the concavity of the right tail of the distribution, together with the size and the location of the discontinuity of the wage function. A simple inspection of the wage function in (24) reveals that a smaller $\widetilde{\theta}$ also implies a larger *maximum* wage (this is not evident from the graph). Of course, such inspections cannot substitute for a formal proof of global identification, but there is at least evidence that the three parameters play a substantially different role in determining the shape of the distribution.

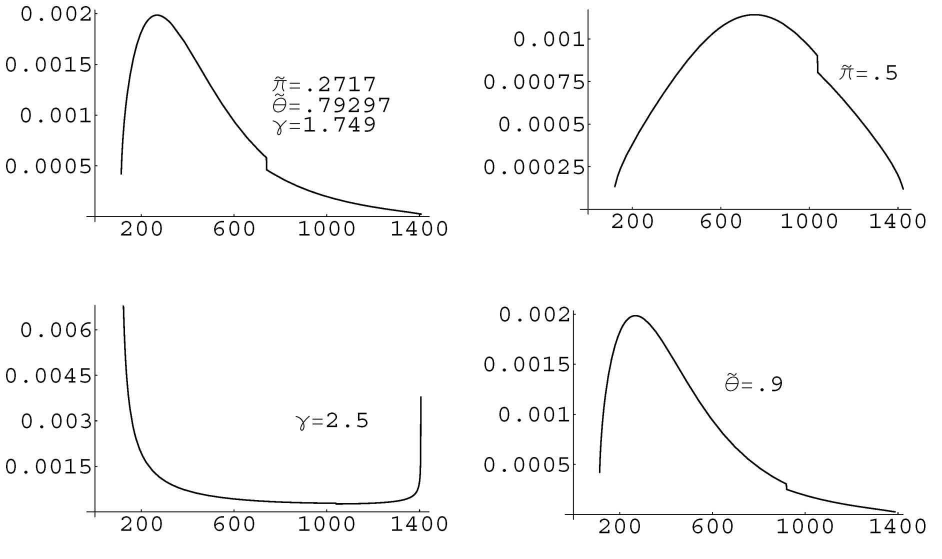

FIGURE 4

WAGE DISTRIBUTION UNDER DIFFERENT VALUES OF $\pi, \widetilde{\theta}, \gamma$

There is no issue of identification in the second step that involves the computation of the unique solution to a set of linear equations.

The parametric assumptions adopted for the production and "testing" technologies are not crucial for identification. It is also not crucial to assume exogenous differences between groups (in terms of skill endowment and testing technology). It is possible to estimate the model assuming that groups are ex ante identical, in which case all the wage differential would be generated by informational asymmetries. Under such hypothesis the first stage of the estimation would require in the first step a joint estimation of the two wage distributions.

Finally, the assumption that groups share the same distribution over investment costs G could also be relaxed. All that is needed to estimate a distribution is a set of observations in the space {*proportion of investors, benefits from investing*}. In this article the choice was to use the two observations from the two groups to exactly match a uniform distribution. Alternatively, one could, for example, use several observations from different markets (in which case the identifying assumption would be that the workers in different markets share the same investment cost distribution).

5. RESULTS

The model is *separately* estimated three times at time intervals centered at years 1965, 1980, and 1995 using weekly wage data from the Current Population Surveys. Wage observations from black and white males working full time are merged over three-year intervals to increase sample size. To perform consistent intertemporal comparisons, wages are standardized to the 1982–1984 level according to the Consumer Price Index. Weekly data are used because information on hours worked was not collected before 1976, which makes it impossible to compute hourly wages. Only workers working full time (full or part year) are considered. The number of observations ranges from 3995 (1965-centered sample) to 5679 (1995) in the blacks' sample, and from 49,087 (1965) to 76,319 (1980) in the whites' sample. In each sample, wages are topcoded at \$1000 to guarantee that at least 1% of the observations is always on the topcoded tail of the wage distribution. A small number of observations below the federal minimum wage were discarded.[22] Information about the size of the employed population was obtained from the Statistical Abstracts of the United States.

5.1. *Parameter Estimates.*

Table 1 reports the parameters directly estimated by maximum likelihood in the first step. The fairly small standard errors (in parenthesis) should convince the reader that the model is locally identified.

The results show that the proportion of blacks investing in human capital has always been smaller than the proportion of whites, on average by 12.3%, but the

[22] Weekly wages were divided by 40 and the result compared with the federal minimum wage. If the federal minimum wage was higher the observation was discarded.

TABLE 1
FIRST-STEP PARAMETER ESTIMATES

Year	Proportion of Investors		Threshold Signal		Testing Technology	
	$\widehat{\pi^b}$	$\widehat{\pi^w}$	$\widehat{\tilde{\theta}^b}$	$\widehat{\tilde{\theta}^w}$	$\widehat{\gamma^b}$	$\widehat{\gamma^w}$
1965	0.1452 (8.0E−3)	0.2408 (3.4E−4)	0.8786 (2.4E−2)	0.7834 (1.1E−3)	1.669 (3.9E−3)	1.591 (3.2E−4)
1980	0.2381 (3.6E−3)	0.2992 (1.5E−4)	0.8883 (1.8E−2)	0.8166 (4.1E−9)	1.744 (2.9E−3)	1.696 (2.7E−4)
1995	0.1938 (5.2E−3)	0.2717 (2.3E−4)	0.9157 (2.5E−2)	0.7297 (1.3E−3)	1.721 (3.2E−3)	1.749 (2.6E−4)

TABLE 2
OTHER PARAMETERS OBTAINED IN THE FIRST STEP

Year	$\frac{\lambda^b}{\lambda^w}$	$\widehat{y_u^b} = \widehat{y_u^w}$	$\widehat{y_q^b}$	$\widehat{y_q^w}$
1965	0.10790	150.6	1336.3 (74.29)	1336.1 (2.03)
1980	0.12108	150.5	1101.7 (18.91)	1240.7 (0.46)
1995	0.12107	111.5	1390.5 (41.51)	1371.0 (1.39)

difference has been declining over time.[23] All six estimations satisfy the model restriction that $\widetilde{\theta}^j$ should be strictly within the unit interval. Groups are estimated to have similar informativeness of the test, measured by γ^j, even though a formal test rejects the hypothesis of equality in any given year. This is interesting because early models of statistical discrimination, such as Phelps (1972), derived wage inequality by assuming that groups carried signals of different quality.

Table 2 reports other parameters obtained in the first step. The employed population ratio λ^b/λ^w and $\widehat{y_u^b}$ does not have standard errors, being obtained directly from employment data and from the minimum wage, respectively. Note that $\widehat{y_u^b} = \widehat{y_u^w}$ every year because the estimations use only observations with wages at or above the federal minimum wage.

Since the first-step parameter estimates are hard to interpret, Figure 5 and Table 3 report information computed from the estimated parameters. Recall that all monetary variables are in dollars per week normalized at the 1982–1984 level.

[23] Since human capital investment in the model is *unobserved*, both by the firms and by the investigator, it is not possible to relate the relationship between the estimated trends in such investment and the convergence in *observable* human capital. A more complete model including an observable investment decision could conceivably be introduced at the cost of much greater complexity (see Fang, 2001 for an application using only one racial group). The main conceptual problem lies in choosing how to model the interaction between the two types of investment because results may be particularly sensitive to the modeling assumptions.

TABLE 3

COMPUTED RESULTS

		Gross Returns to Investment*		Predicted Average Wage*		Workers in Simple Task (%)	
Year	$\widehat{\pi^w} - \widehat{\pi^b}$	Black	White	Black	White	Black	White
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1965	0.0956	166.65	179.07	325.6	443.3	94.6	85.6
1980	0.0611	190.91	219.25	379.1	481.6	93.8	87.4
1995	0.0779	221.17	276.88	360.4	461.9	96.1	81.1

*In weekly dollars normalized at the 1982–1984 level.

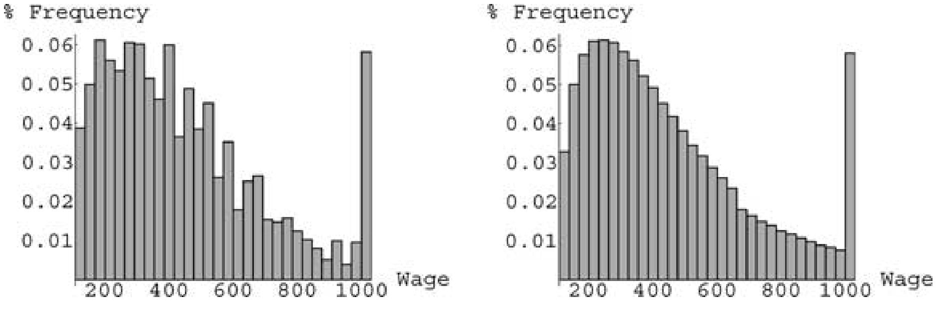

FIGURE 5

EMPIRICAL AND ESTIMATED DISTRIBUTIONS, WHITE MALES, 1995

Figure 5 compares the theoretical and estimated distributions of white male wages in 1995, showing that, in general, the model performs fairly well in describing the empirical wage distribution.[24] The other estimations produce a similar picture.

Column 1 in Table 3 shows that the estimated difference in the proportion of people who choose to invest declines between 1965 and 1980 from 9.56% to 6.11%, and slightly increases afterwards. The increase of incentives to invest (columns 2 and 3) and the decline of average wages after the 1970s (columns 4 and 5) reflect perhaps the increase in wage inequality that occurred after 1970 documented in Juhn et al. (1993) and Levy and Murnane (1992).

Columns 6 and 7 show that on average, 94.8% of blacks and 84.7% of whites were employed in the simple task. This suggests that our estimation is consistent only with a rather narrow definition of "complex task."[25]

[24] To make the objects comparable histograms have been computed from the estimated distribution using the same thresholds.

[25] More troublesome is the observation that the fraction of blacks employed in the complex task is lower in 1995 than in 1965. Recall, however, that the model regards as "complex" a job-task that requires *unobserved* investment in human capital in order to be effectively performed, and the list of jobs that meet this criterion may well change over time without affecting the validity of the intertemporal comparison between equilibria that is the main exercise in this article.

TABLE 4

MOMENTS OF THE SKILL ENDOWMENT AND PARAMETERS OF THE PRODUCTION FUNCTION

Year	$\widehat{e_q^b}$	$\widehat{e_q^w}$	$\widehat{e_u^b} = \widehat{e_u^w}$	$\widehat{\rho^b}$	$\widehat{\alpha}$	$\widehat{y_1}$	$\widehat{y_2}$
1965	2474.1	2473.8	278.9	1.01	0.263	0.630	0.540
	(131.4)	(6.60)	(0.872)	(0.021)	(3.4E−3)	(1.2E−3)	(1.2E−3)
1980	1949.8	2195.8	266.3	0.993	0.237	0.622	0.565
	(28.4)	(5.82)	(0.700)	(0.011)	(2.1E−3)	(1.6E−3)	(1.5E−3)
1995	2758.2	2719.5	221.3	0.960	0.354	0.556	0.504
	(76.8)	(5.88)	(0.537)	(0.012)	(2.6E−3)	(1.8E−3)	(1.7E−3)

TABLE 5

PARAMETERS OF THE COST AND SKILL DISTRIBUTIONS

Year	$\widehat{p_1}$	$\widehat{p_2}$	$\widehat{p_3^b}$	$\widehat{p_3^w}$	$\widehat{p_4^b}$	$\widehat{p_4^w}$
1965	−1.1371	0.007695	7788.4	7994.4	−33.800	−33.781
	(0.258)	(1.4E−3)	(1552.0)	(1114.0)	(8.287)	(6.248)
1980	−0.17325	0.002155	2934.0	3441.6	−7.255	−8.315
	(0.034)	(1.6E−4)	(153.3)	(123.3)	(0.646)	(0.581)
1995	−0.11581	0.001400	3837.4	3977.0	−7.102	−6.994
	(0.038)	(1.3E−4)	(287.2)	(180.2)	(0.906)	(0.666)

Since the estimates are consistent with the restrictions from the model and the empirical wage distributions are well matched by the estimated distributions in each year and race, it is possible to proceed with the second step. Table 4 presents the estimates of the parameters of the production function, marginal products, and moments of skill endowment.[26] As described in Section 4, results in Tables 1 and 4 are obtained without assuming any specific parameterization of the distribution over costs and abilities. When the results obtained from the first step from blacks and whites' data are merged and the parameterization of the cost distribution and skill endowment is adopted, the other parameters of the model can be obtained and welfare analysis can be performed.

Observe first in Table 5 that the model is always capable of providing economically meaningful estimates of p_2, p_4^b, p_4^w. Table 6 presents the computed average cost of investment paid by workers. Average welfare is simply the average wage minus the average cost. The last two columns report the black to white average welfare and wage ratios. Notice that the pattern of reduction in black–white

[26] Another result that may appear counterfactual is the estimation of ρ^b indicating that blacks had a slight advantage or were slightly favored in the complex task in 1965. This may be explained as the result of preferential policies and civil rights laws that were implemented in the mid-1960s, and more generally of the widespread climate toward civil rights. However, even if we believe this interpretation it is not clear why similar effects are not also reflected in the estimation of ρ^b in 1980 and 1995. Since the model ignores the effect of preferential policies on the wage distribution, it is not clear how parameters are affected by such policies, which makes it hard to draw any implication from this result.

TABLE 6
AGGREGATE WELFARE IN THE SELECTED EQUILIBRIUM

	Average Cost of Investment*		Average Welfare*		Black/White Ratio	
Year	Black	White	Black	White	Welfare	Wage
1965	22.8	39.3	302.7	404.0	0.749	0.734
1980	32.3	44.8	346.8	436.8	0.794	0.787
1995	29.4	48.9	330.9	413.0	0.801	0.780

*In weekly dollars normalized at the 1982–1984 level.

TABLE 7
EQUILIBRIA, 1980

Investors		Average Wage		Average Welfare	
π^b	π^w	Black	White	Black	White
0.238	*0.299*	*379.1*	*481.6*	*346.8*	*436.8*
0.121	0.301	377.1	481.8	364.0	436.5

NOTE: Selected equilibrium in italics.

differentials is not substantially different when looking at welfare instead of wages, and confirms that inequality, measured using either welfare or wage, decreased sharply before 1980, but remained approximately stable afterwards. This is consistent with standard evidence reported by the empirical literature on racial wage inequality.

5.2. *Other Equilibria.* A numerical procedure has been used to investigate whether other equilibria were compatible with the estimated parameters. The procedure could not find other equilibria in 1965 and 1995, but another equilibrium has been found using the parameters estimated obtained in 1980. Table 7 reports the main result, showing that the economy selected the equilibrium with the lowest differentials in terms of wage and human capital investment differentials. In both equilibria blacks receive less than whites. In the selected equilibrium differences in welfare are larger because of the higher investment cost paid by blacks.

This result implies that statistical discrimination generated by self-fulfilling expectations about worker's productivity did not exacerbate wage inequality in the United States. Therefore, the observed reduction in wage inequality must be an effect of changes in the fundamental parameters of the model. Two possibilities consistent with these results are that taste discrimination may have decreased, or that we do not have good measurements of the relative black–white human capital improvements reported in the literature surveyed by Donohue and Heckman (1991).

5.3. *A Color-Blind Society.* Table 8 reports a measure of the effects of statistical discrimination computed by calculating what equilibrium the economy would

TABLE 8
A COLOR-BLIND SOCIETY

Year	Investors		Average wage	Wage Gains*		Welfare (Gains*)	
	Black	White		Black	White	Black	White
1965	0.232	0.210	428.6	+31.6%	−3.3%	390.7 (+29.08%)	394.7 (−2.31%)
1980	0.302	0.293	470.5	+24.1%	−2.3%	425.0 (+22.56%)	427.1 (−2.22%)
1995	0.261	0.266	451.4	+25.2%	−2.3%	405.5 (+22.53%)	404.1 (−2.15%)

*Percentage gains relative to the selected equilibrium.

display in a color-blind society, where employers cannot distinguish the race of their employees. The direct effect of this experiment is that employers cannot compute by-group statistics to assess individual productivity; therefore, the average expected productivity of a worker will be a weighted average of a black and a white worker's productivity. This induces an indirect effect of creating similar incentives to invest in human capital. Since the testing technology is different, this does not generate a society without differentials, but differences in terms of wages and welfare are substantially reduced. Incentives are only slightly different, and in 1965 and 1980 they are such that blacks' investment is slightly higher than whites' investment in equilibrium. Despite the burden of a higher investment cost, on average blacks gain 24.5% in terms of welfare with respect to the selected equilibrium, but whites lose only 2.2%, since their wage loss is partially compensated by the smaller cost of investment paid.

It is important to stress that this experiment has been performed under the estimated set of parameters, which are conditional on the existing policy adopted in the United States at the time data were collected. After 1965, the U.S. labor market experienced several waves of enforcement of anti-discrimination policies. Although these policies are not explicitly modeled in this article, they probably affect the estimation of the fundamentals. Since a color-blind society is normally intended as a society without preferential treatment, it would be more appropriate to discount the effect of preferential policies on welfare before performing the experiment. To the extent that preferential policies improve blacks' welfare, the assessment of the gains from a color-blind society performed here overstates the gains for blacks.

6. CONCLUSION

Despite being somewhat stylized, the model performs well in representing the observed wage distributions. The results show that equilibrium selection did not play a role in reducing wage discrimination in the last 30 years. The qualitative results obtained in this article are robust to different modeling assumptions. Not reported in the article are the results of the estimation conducted first using a linear technology, and second assuming a unique job-task where human capital investment increases productivity. In both cases the estimated parameters always imply a unique equilibrium.

The results are particularly interesting since one of the effects that the theoretical literature attribute to antidiscrimination policies such as affirmative action is to change the set of equilibria of the economy. The results of this article imply that if antidiscrimination policy had any role, it was not in effecting the equilibrium selection, but only in changing the exogenous parameters of the model. Two possibilities consistent with these results are that prejudice (or "taste discrimination") may have decreased, or that the reduced-form literature could not use good measurements of the relative black–white human capital improvements.

It is reasonable to ask whether the estimation procedure can be extended to different settings. Providing necessary and sufficient conditions that make this procedure valid in general is outside the scope of this article but my conjecture is that the procedure can be generalized.[27] What is important for the procedure to work is that the relation from the space of equilibria to the space of fundamentals is a function, despite its inverse being a correspondence. Hence, once the equilibrium is given, the fundamentals can be uniquely obtained computing such a function. It is easy to write examples, even in this setup, where this feature would is violated. Subsection 4.3 shows that if the cost distribution were not uniform, multiple values of the fundamentals would be consistent with the same equilibrium. Then, it would not be possible to recover the fundamentals from a description of the equilibrium. More generally, if the model were stochastic, equilibria would depend on the realization of a random variable, and it would not be inconceivable that the same equilibrium could be generated by different parameters (albeit through a different realization of the random variable). Even in this case the procedure developed in this article would not work.

APPENDIX

Proof of Proposition 1. The proof assumes that only two firms exist, and it is easily generalizable to the case of N firms. To save on notation, we define $f^j(\theta;\pi^j) \equiv \pi^j \underline{f_q^j}(\theta) + (1-\pi^j) f_u^j(\theta)$ (the density of workers carrying a given signal θ), and $\overline{e^j}(\theta) \equiv \overline{e_q}^j p^j(\theta,\pi^j) + \overline{e_u}^j(1-p^j(\theta,\pi^j))$ (the average number of efficiency units conditional on carrying signal θ). The proposition considers fraction of investors $(\pi^b,\pi^w) \in [0,1]^2$ as given.

Sufficiency. Given $(\pi^b,\pi^w) \in [0,1]$, let $(\widetilde{\theta}^b(\pi^b,\pi^w), \widetilde{\theta}^w(\pi^b,\pi^w))$ solve the task assignment problem (12). Suppose that each firm posts wage schedules $w^j{:}[0,1] \to R$ given by $w^j(\theta \mid \pi^b,\pi^w)$ defined in (13) and assigns workers according to $t{:}[0,1] \to [0,1]$, the threshold rule with cutoff $\widetilde{\theta}^j(\pi^b,\pi^w)$. C, S denote the factor inputs associated with the candidate equilibrium actions. Suppose that one firm deviates from the proposed equilibrium and plays strategy $(w^{j\prime},t^{j\prime})_{j=b,w} \neq (w^j,t^j)_{j=b,w}$. Define $\Theta^{jh} = \{\theta : w^{j\prime}(\theta) > w^j(\theta \mid \pi^b,\pi^w)\}$ and $\Theta^{jl} = \{\theta : w^{j\prime}(\theta) < w(\theta \mid \pi^b,\pi^w)\}$, respectively, as the set of signals where the deviating firm attracts group j workers

[27] Recently, Antonovics (2002) and Fang (2001) have adapted the procedure developed in this article, respectively, to analyze the persistence of racial discrimination and to quantify the signaling component of the college wage premium.

and where the deviating firm loses group j workers. Finally, $\Theta^{je} = \{\theta : w^{j\prime}(\theta) = w^j(\theta)\}$ is the set of signals such that workers are indifferent between firms (the deviating firm attracts only 1/2 of such workers). Let $C' = \rho^b C^{b\prime} + \rho^w C^{w\prime}$ and $S' = S^{b\prime} + S^{w\prime}$denote the implied factor inputs for the deviator where

$$\text{(A.1)} \quad C^{j\prime} = \lambda^j \left(\int_{\theta\in\Theta^{jh}} t^{j\prime}(\theta)\, \overline{e_q}^j \pi^j f_q^j(\theta)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^{je}} t^{j\prime}(\theta)\, \overline{e_q}^j \pi^j f_q^j(\theta)\, d\theta \right)$$

$$\text{(A.2)} \quad S^{j\prime} = \lambda^j \left(\int_{\theta\in\Theta^{jh}} (1 - t^{j\prime}(\theta)) \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^{je}} (1 - t^{j\prime}(\theta)) \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta \right)$$

Given that the other firms play according to the equilibrium strategies we can express profits for the deviator as

$$\text{(A.3)} \quad \Pi^i_{dev} = y(C', S') - \sum_{j=b,w} \lambda^j \left(\int_{\theta\in\Theta^{jh}} w^{j\prime}(\theta) f^j(\theta, \pi^j)\, d\theta - \frac{1}{2} \int_{\theta\in\Theta^{je}} w^j(\theta) f^j(\theta, \pi^j)\, d\theta \right)$$

Now multiply both sides in (A.1) and (A.2) by $y_1(C, S)\, \rho^j$ and $y_2(C, S)$, respectively, and substitute the identity $\pi^j f_q^j(\theta) = p^j(\theta, \pi^j) f^j(\theta, \pi^j)$

$$\text{(A.4)} \quad y_1(C, S)\, \rho^j C^{j\prime} = \lambda^j \left(\int_{\theta\in\Theta^{jh}} y_1(C, S) \rho^j \overline{e_q}^j p^j(\theta, \pi^j) t'(\theta) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^{je}} y_1(C, S)\, \rho^j \overline{e_q}^j p^j(\theta, \pi^j) t'(\theta)\, d\theta \right)$$

$$\text{(A.5)} \quad y_2(C, S) S^{j\prime} = \lambda^j \left(\int_{\theta\in\Theta^{jh}} y_2(C, S) \overline{e^j}(\theta)(1 - t^{j\prime}(\theta)) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2} \int_{\theta\in\Theta^{je}} y_2(C, S) \overline{e^j}(\theta)(1 - t^{j\prime}(\theta)) f^j(\theta, \pi^j)\, d\theta \right)$$

But $w^j(\theta) = \max\{y_2(C, S)\overline{e^j}(\theta), y_1(C, S)\rho^j \overline{e_q}^j p^j(\theta, \pi^j)\}$; therefore

$$\text{(A.6)} \qquad y_1(C, S)\rho^j C^{j\prime} \le \lambda^j \left(\int_{\theta\in\Theta^{jh}} w^j(\theta) t'(\theta) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2}\int_{\theta\in\Theta^{je}} w^j(\theta) t'(\theta) f^j(\theta, \pi^j)\, d\theta \right)$$

$$\text{(A.7)} \qquad y_2(C, S) S^{j\prime} \le \lambda^j \left(\int_{\theta\in\Theta^{jh}} w^j(\theta)(1 - t^{j\prime}(\theta)) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2}\int_{\theta\in\Theta^{je}} w^j(\theta)(1 - t'(\theta)) f^j(\theta, \pi^j)\, d\theta \right)$$

Summing the inequalities in (A.6) and (A.7) for $j = b, w$ we obtain

$$\text{(A.8)} \quad y_1(C, S)C' + y_2(C, S)S' \le \sum_{j=b,w} \lambda^j \left(\int_{\theta\in\Theta^{jh}} w^{j\prime}(\theta) f^j(\theta, \pi^j)\, d\theta + \frac{1}{2}\int_{\theta\in\Theta^{je}} w^{j\prime}(\theta) f^j(\theta, \pi^j)\, d\theta \right)$$

Furthermore, $y(C', S') \le y_1(C, S)C' + y_2(C, S)S'$ by concavity and constant returns. Use these two facts to substitute into (A.3) to obtain $\Pi^i_{dev} \le 0$. Hence the deviation is not profitable.

Necessity. To simplify the exposition, necessity is proven in four steps.

Step 1. *On the equilibrium path, wage schedule posted by firms are the same almost everywhere.*

Proof. Suppose there is a positive measure set of signals where firms post different wages. Then, this implies that there is a set of signals of positive measure where one firm (say, firm 1) posts wages higher than the wages posted by the other firm. Then, firm 1 can reduce wages on this set, while keeping them above wages posted by the other firm. Firm 1's output does not change, but the deviation is profitable since the total wage bill is smaller. ■

Step 2. *Let $\{t_i^j\}_{j=b,w}$ denote the implied task assignment rules on the equilibrium path for firm i and let t^b, t^w be the cutoff rules with critical point $\widetilde{\theta}^b$, $\widetilde{\theta}^w$ solving (12). Then, $t_i^j(\theta) = t^j$, $j = b, w$, all i for almost all θ.*

Proof. It is not difficult to prove that firms must use a cutoff rule. If not then there are sets $\Theta^h, \Theta^l \subseteq [0, 1]$ with a positive measure such that $\theta^h > \theta^l$ for all θ^h,

$\theta^l \in \Theta^h \times \Theta^l$, $t_i^j(\theta^h) = 0$ for all $\theta^h \in \Theta^h$ and $t_i^j(\theta^l) = 1$ for all $\theta^l \in \Theta^l$, and at least one group $j \in \{b, w\}$. Consider the alternative task assignment rule for group j,

$$t_i'(\theta) = \begin{cases} 1 & \text{if } \theta \in \Theta^h \\ 0 & \text{if } \theta \in \Theta^l \\ t_i^j(\theta) & \text{otherwise} \end{cases} \tag{A.9}$$

(keeping task assignment on the other group unchanged). Let S_i^j, C_i^j and S_i', C_i' be the factor inputs implied by t_i^j, t_i', respectively. Assume w.l.o.g. $\int_{\theta\in\Theta^h} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta = \int_{\theta\in\Theta^l} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta > 0$ (one can appropriately choose sets Θ^h, Θ^l such that this is satisfied). This implies $S'_i = S_i^j$. Since the deviation assigns to the complex task workers who are more likely to be productive, then $C'_i > C_i^j$. To see this, let $l(\theta) = f_q^j(\theta)/f_u^j(\theta)$ denote the likelihood ratio and note that

$$\begin{aligned} 2(C_i' - C_i^j)/\rho^j &= \int_{\theta\in\Theta^h} \overline{e_q}^j(\theta)\pi^j f_q^j(\theta)\, d\theta - \int_{\theta\in\Theta^l} \overline{e_q}^j \pi^j f_q^j(\theta)\, d\theta \\ &= \pi^j \overline{e_q}^j \left(\int_{\theta\in\Theta^h} l(\theta) f_u(\theta)\, d\theta - \int_{\theta\in\Theta^l} l(\theta) f_u(\theta)\, d\theta \right) \end{aligned}$$

By the monotone likelihood ratio there exists θ^* such that $l(\theta) \geq l(\theta^*)$ for all $\theta \in \Theta^h$ and $l(\theta) \leq l(\theta^*)$ for all $\theta \in \Theta^l$, with at least one inequality holding strictly. Hence $C_i' - C_i^j > 0$, so that output, and therefore profits also, are higher under t_i'.

Having established that firms must use a cutoff rule, to complete the proof, we need to show that there is a unique pair of cutoffs. This problem reduces to finding a solution to problem (12) in the main text. It is straightforward but tedious to show that this is a concave problem, and that the Kuhn–Tucker conditions are necessary and sufficient. ■

Step 3. *Suppose $\{w^b, w^w\}$ is a pair of equilibrium wage schedules and let $\{\widetilde{\theta}^b(\pi^b, \pi^w), \widetilde{\theta}^w(\pi^b, \pi^w)\}$ be the solution to (12). Then there are pairs $\{k_s^j, k_c^j\}_{j=b,w}$ such that $w^j(\theta)/\overline{e^j}(\theta)$ is equal to a constant k_s^j for almost all $\theta < \widetilde{\theta}^j(\pi^b, \pi^w)$, $j = b, w$ and $w^j(\theta)/(\rho^j p^j(\theta, \pi^j)\overline{e_q}^j)$ is equal to a constant k_c^j for almost all $\theta > \widetilde{\theta}^j(\pi^b, \pi^w)$, $j = b, w$.*

Proof. First it is shown that $w^j(\theta)/\overline{e^j}(\theta) = k_s^j$ for almost all $\theta < \widetilde{\theta}^j(\pi^b, \pi^w)$. For contradiction assume that for group j there are sets $\Theta^a, \Theta^b \subseteq [0, \widetilde{\theta}^j(\pi^b, \pi^w)]$ with strictly positive measure such that $w^j(\theta)/\overline{e^j}(\theta) < k_1 < k_s^j$ for all $\theta \in \Theta^a$ and $w^j(\theta)/\overline{e^j}(\theta) > k_2 > k_s^j$ for all $\theta \in \Theta^b$, and some k_1, k_2. Choose Θ^a, Θ^b so that the following is satisfied:

$$\int_{\theta\in\Theta^a} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta = \int_{\theta\in\Theta^b} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta > 0 \tag{A.10}$$

Consider a unilateral deviation w_i' by firm i where

$$\text{(A.11)} \qquad w_i'(\theta) = \begin{cases} \left(k_s^j + \epsilon\right)\overline{e^j}(\theta) & \text{for } \theta \in \Theta^a \\ 0 & \text{for } \theta \in \Theta^b \\ w(\theta) & \text{otherwise} \end{cases}$$

(keeping wages in the other group unchanged). Given Step 1, the deviating firm loses to the other firm workers with $\theta \in \Theta^b$, and attracts all workers with $\theta \in \Theta^a$ previously hired by the other firm (their mass is $\lambda^j/2 * f^j(\theta, \pi)$). Task assignment is unchanged; therefore, (A.10) implies that the input of both factors remains constant. Firm i increases profits by the difference in wage payments, i.e.,

(A.12)

$$\begin{aligned} \Delta(\epsilon) &= \lambda^j \left[\int_{\theta\in\Theta^b} \frac{1}{2} w^j(\theta) f^j(\theta, \pi^j)\, d\theta - \int_{\theta\in\Theta^a} \left(\frac{1}{2} w_i'(\theta) + \varepsilon\right) f^j(\theta, \pi^j)\, d\theta\right] \\ &\geq \lambda^j \left[\int_{\theta\in\Theta^b} \frac{1}{2} k_2 \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta - \int_{\theta\in\Theta^a} \left(\frac{1}{2} k_1 + \frac{\varepsilon}{\overline{e^j}(\theta)}\right) \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta\right] \\ &= \lambda^j \frac{1}{2} \int_{\theta\in\Theta^a} (k_2 - k_1) \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta - \int_{\theta\in\Theta^a} \frac{\varepsilon}{\overline{e^j}(\theta)} f^j(\theta, \pi^j)\, d\theta \end{aligned}$$

(in the expression ε is not multiplied by $1/2$ since the deviating firm raises wages by ε also to its current workers in Θ^a). The last equality follows from (A.10). Since $k_2 > k_1$ for ϵ small enough the deviation is profitable.

A symmetric argument proves that $w^j(\theta)/(\rho^j p^j(\theta, \pi^j)\overline{e_q}^j)$ is also equal to a constant k_c^j if $\theta > \widetilde{\theta}^j(\pi^b, \pi^w)$. ■

Step 4. *Proposition 1.*

It remains to be shown that $k_s^j = y_2(C, S)$ and $k_c^j = y_1(C, S)$, $j = b, w$. Recall that homogeneity of degree zero of the partial derivatives of y implies zero profits in equilibrium. Suppose, by contradiction, this is not true for one group and call j such group. It is easy to show that if $k_s^j \leq y_2(C, S)$ and $k_c^j \leq y_1(C, S)$, $j = b, w$, (with at least one of strict inequality), then firms are making positive profits and a deviation where firm i offers $w_i^{j\prime}(\theta) = w_i^j(\theta) + \epsilon$ for all θ and j would be profitable for ϵ small enough. Also, if both inequalities went the other way firms would make negative profits and a deviation to $w_i(\theta) = 0$ for all θ would be profitable. The case where groups have inequalities going in opposite directions $k_s^j \leq y_2(C, S)$, $k_c^j \leq y_1(C, S)$, $k_s^h \geq y_2(C, S)$, and $k_c^h \geq y_1(C, S)$, $j, h \in \{b, w\}$, $j \neq h$ (with at least one strict inequality) can be similarly handled by disposing of h workers and attracting j workers.

The two cases that require some work are when the inequalities work in opposite directions within one group. The arguments are symmetric so consider only the case where there exists a group j with $k_s^j > y_2(C, S)$ and $k_c^j < y_1(C, S)$. In this case, the solution to the task assignment problem (12) must be interior, implying $y_1(C, S)\rho^j \overline{e_q}^j \cdot p^j(\widetilde{\theta}^j(\pi^b, \pi^w), \pi) = y_2(C, S)\overline{e^j}(\widetilde{\theta}^j(\pi^b, \pi^w))$ by (14).

Hence, if $k_s^j > y_2(C, S)$ and $k_c^j < y_1(C, S)$ then we have $k_c^j \rho^j \overline{e_q}^j p^j(\widetilde{\theta}^j(\pi^b, \pi^w), \pi) < k_s^j \overline{e^j}(\widetilde{\theta}^j(\pi^b, \pi^w))$, and by continuity and Step 3 there is an interval (θ', θ^*) around $\widetilde{\theta}^j(\pi^b, \pi^w)$ such that $w^j(\theta^*) < w(\theta')$ which implies

$$k_c^j \rho^j \overline{e_q}^j p^j(\theta^*) < k_s^j \overline{e^j}(\theta') \tag{A.13}$$

The argument below shows that it is better to dispose of some of the workers assigned to the simple task with $\theta \in (\theta', \widetilde{\theta}^j(\pi^b, \pi^w))$ and attract cheaper workers with $\theta \in (\widetilde{\theta}^j(\pi^b, \pi^w), \theta^*)$ from the other firm. To do so, choose $\theta', \theta'', \theta^*$ satisfying $0 < \theta' < \widetilde{\theta}^j(\pi^b, \pi^w) < \theta'' < \theta^* < 1$ so that the following equalities are satisfied:

$$\int_{\theta'}^{\widetilde{\theta}^j(\pi^b,\pi^w)} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta = \int_{\widetilde{\theta}^j(\pi^b,\pi^w)}^{\theta^*} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta \tag{A.14}$$

and

$$\int_{\widetilde{\theta}^j(\pi^b,\pi^w)}^{\theta''} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta = \int_{\theta''}^{\theta^*} \overline{e^j}(\theta) f^j(\theta, \pi^j)\, d\theta \tag{A.15}$$

Consider the following deviation:

$$w_i'(\theta) = \begin{cases} 0 & \text{for } \theta \in [\theta', \widetilde{\theta}^j(\pi^b, \pi^w)) \\ (k_c^j + \epsilon) p^j(\theta, \pi^j) \rho^j \overline{e_q}^j & \text{for } \theta \in [\widetilde{\theta}^j(\pi^b, \pi^w), \theta^*) \\ w(\theta) & \text{otherwise} \end{cases} \tag{A.16}$$

$$t_i'(\theta) = \begin{cases} 0 & \text{for } \theta \in [0, \theta'') \\ 1 & \text{for } \theta \in [\theta'', 1) \end{cases} \tag{A.17}$$

(keeping wages and task assignment on the other group unchanged). The deviating firm is attracting all workers (including those hired by the other firm in the candidate equilibrium) with $\theta \in [\widetilde{\theta}^j(\pi^b, \pi^w), \theta^*)$ and only a fraction (1/2) of workers with $\theta \in [0, \theta') \cup [\theta^*, 1)$ (the other half is hired by the other firm). From (A.14) and (A.15), the input of simple labor is unchanged. The effective units of complex labor from group j strictly increase (equality (A.15) implies that after the deviation the mass of workers in the complex task remains the same but they have higher θ). Thus, output increases. Wages increase by

$$\Delta(\text{wages}) = \lambda^j \left(\int_{\widetilde{\theta}^j(\pi^b,\pi^w)}^{\theta^*} \left(\frac{1}{2} k_c^j + \varepsilon \right) p^j(\theta, \pi^j) \rho^j \overline{e_q}^j f^j(\theta, \pi^j)\, d\theta \right.$$
$$\left. - \int_{\theta'}^{\widetilde{\theta}^j(\pi^b,\pi^w)} \frac{1}{2} k_s^j \overline{e^j}(\theta) f^j(\theta, \pi)\, d\theta \right) \tag{A.18}$$

Take the limit as $\varepsilon \to 0$ and use the fact that $p^j(\theta, \pi)$ and $\overline{e^j}(\theta)$ are both increasing in θ to obtain

$$\text{(A.19)} \qquad \Delta\,(\text{wages}) < \lambda^j \left(\frac{1}{2} k_c^j p^j(\theta^*) \rho^j \overline{e_q}^j \int_{\widetilde{\theta}^j(\pi^b, \pi^w)}^{\theta^*} f^j(\theta, \pi^j)\, d\theta - \frac{1}{2} k_s^j \overline{e^j}(\theta') \int_{\theta'}^{\widetilde{\theta}^j(\pi^b, \pi^w)} f^j(\theta, \pi^j)\, d\theta \right)$$

To prove that this is negative observe that (A.14) and $\overline{e^j}(\theta)$ increasing implies $\int_{\widetilde{\theta}^j(\pi^b, \pi^w)}^{\theta^*} f^j(\theta, \pi^j)\, d\theta < \int_{\theta'}^{\widetilde{\theta}^j(\pi^b, \pi^w)} f^j(\theta, \pi^j)\, d\theta$. Hence

$$\text{(A.20)} \qquad \Delta\,(\text{wages}) < \lambda^j \frac{1}{2} \big(k_c^j p^j(\theta^*) \rho^j \overline{e_q}^j - k_s^j \overline{e^j}(\theta') \big) \int_{\widetilde{\theta}^j(\pi^b, \pi^w)}^{\theta^*} f^j(\theta, \pi^j)\, d\theta$$

From (A.13) the right-hand side is negative. Thus, wages decrease and output increases so the deviation is profitable and Proposition 1 follows.

REFERENCES

ANTONOVICS, K., "Persistent Racial Wage Inequality," University of California, San Diego Discussion paper no. 2002-05, February 2002.

ARROW, K. J., "The Theory of Discrimination" in O. Ashenfelter and A. Rees, eds., *Discrimination in the Labor Markets* (Princeton, NJ: Princeton University Press, 1973), 3–33.

BISIN, A., G. TOPA, AND T. VERDIER, "Religious Intermarriage and Socialization in the U.S.," mimeo, New York University, New York, 2001.

BRESNAHAN, T. F., AND P. C. REISS, "Empirical Models of Discrete Games," *Journal of Econometrics*, 48 (April–May 1991), 57–81.

BROCK, W. A., AND S. N. DURLAUF, "Interaction-Based Models," in J. J. Heckman and E. E. Leamer, eds., *Handbook of Econometrics* (Amsterdam: North-Holland, 2001), Chapter V.

COATE, S., AND G. C. LOURY, "Will Affirmative Action Policies Eliminate Negative Stereotypes?" *American Economic Review*, 83 (December 1993), 1220–40.

CARD, D., AND A. B. KRUEGER, "School Quality and Black–White Relative Earning: A Direct Assessment," *Quarterly Journal of Economics*, 107 (February 1992), 151–200.

DAGSVIK, J., AND B. JOVANOVIC, "Was the Great Depression a Low-Level Equilibrium?" *European Economic Review*, 38 (December 1994), 1711–29.

DONOHUE, J. J., III, AND J. HECKMAN, "Continuous versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, 29 (December 1991), 1603–43.

FANG, H., "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices," Mimeo, Yale University, May 2001.

HECKMAN, J., "The Impact of Government on the Economic Status of Black Americans," in S. Shulman and W. Darity Jr., eds., *The Question of Discrimination: Racial Inequality in the U.S. Labor Market* (Middletown, CT: Welseyan University Press, 1989), 50–80.

JOVANOVIC, B., "Observable Implications of Models with Multiple Equilibria," *Econometrica*, 57 (November 1989), 1431–37.

Juhn, C., K. M. Murphy, and B. Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Econonomy*, 101 (June 1993), 410–42.

Leonard, J. S., "Wage Disparities and Affirmative Action in the 80's," *American Economic Review* 86 (May 1996), 285–89.

Levy, F., and R. J. Murnane, "U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations," *Journal of Economic Literature* 30 (September 1992), 1333–81.

Lundberg, S. J., "The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives," *Quarterly Journal of Economics* 106 (February 1991), 309–26.

Moro, A., "The Effects of Statistical Discrimination on Black–White Wage Inequality: Omitted Computations," mimeo, University of Minnesota, 2002. Available from the author or directly from http://www.econ.umn.edu/∼amoro/research.html.

——, and P. Norman, "Affirmative Action in a Competitive Economy," *Journal of Public Economics*, 87 (March 2003a), 567–94.

——, and ——, "A General Equilibrium Model of Statistical Discrimination," forthcoming, *Journal of Economic Theory*, 2003b.

Neal, D. A., and W. R. Johnson, "The Role of Premarket Factors in Black–White Wage Differences," *Journal of Political Economomy*, 104 (October 1996), 869–95.

Phelps, E. S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62 (September 1972), 659–61.

Smith, J. P., and F. W. Welch, "Affirmative Action and Labor Markets," *Journal of Labor Economics*, 2 (April 1984), 269–301.

——, and ——, "Black Economic Progress after Myrdal," *Journal of Economic Literature*, 27 (June 1989), 519–64.

Tamer, E., "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 70 (January 2003), 147–65.