Amending the Racial Wage Gap,
What Helps and What Does not?

Luis Pinedo Caro∗†

University of Southampton

May 22nd, 2014

Abstract

This paper develops a dynamic model of asymmetric information in the labour market with two racial groups. The objective is twofold. First we show that a racial wage gap cannot persist without pre-market differences between races. Then we argue that statistical discrimination is positive for the earnings prospects of the discriminated group in the presence of performance pay. Thus, the Title VII of the 1964 Civil Rights Act that forbids the usage of race when hiring might have done more harm than good to the blacks’ labour market outcomes.

Keywords: Statistical Discrimination, Performance Pay, Schooling Differences.


∗Correspondence email: pinedo1985@gmail.com
†Acknowledgements: I am indebted to John Knowles for his relentless support and for the uncountable number of discussions that made this paper flourish. I am also thankful to Thomas Gall for his helpful comments on several drafts of the paper.
1 Introduction

The choice of what policies to pursue in the quest for closing the racial wage gap is not simple. In this paper we offer a theoretical analysis to shed some light on the effect of certain policies. First we take on the role of racial historical differences in schooling in shaping the wage gap. Then we analyze whether a triumphal milestone in racial’s american history like the enactment of the 1964 Civil Rights Act was actually helpful for blacks’ interests in the labour market.

To perform the analysis we rely on a model of statistical discrimination and endogenous human capital accumulation. The choice of statistical discrimination as the justification for different treatment of otherwise similar workers is justified based on its rationality and its profit-maximising orientation. Statistical discrimination has been widely used in the literature since Arrow (1973). Firms in Arrow’s setup use a worker’s race during the hiring process insofar it is correlated with its true productivity. Several papers have followed Arrow since then, including Phelps (1974), Borjas & Goldberg (1978), Coate & Loury (1993) and Moro & Norman (2004) among others. These models often feature the existence of either a convergent and a divergent equilibrium. It is particularly worrisome that some of these discriminatory equilibria would unravel were the employers able to observe ex-post the actual ability of the workers. The ex-post observation would allow firms to offer performance pay the convergent equilibrium would be guaranteed due to the linearity of the workers’ utility function.

The aforementioned lack of performance pay may be regarded as a weakness. We see in empirical studies by Altonji & Pierret (2001) or Lang & Siniver (2011) that even though starting salaries do depend on easily available information (signals like years of schooling, name of the school attended), experienced workers’ earnings tend to depend on actual ability, suggesting that firms are capable of learning a worker’s ability over time and use it. My model, thus, combines two strands of the contract theory literature, signalling and adverse selection, applied to the labour market. In this setup firms receive two signals correlated with the workers type; one ex-ante, reflecting a firms’ first impression of a workers, and one ex-post, correlated with how much they have actually produced. It is the existence of new information ex-post incentivizes firms to condition part of the salary until the reception of the ex-post signal.

A recurrent theme in the literature on statistical discrimination is whether an initial negative belief towards one group is self-fulfilling. I show that in presence of performance pay the chances of having a self-fulfilling equilibrium decreases. Even more striking is the model’s suggestion that the use by the firms of statistical discrimination is beneficial for the discriminated group in terms of wage gap closure. Those effects occur due to the firm setting different screening menus for the different groups, in a way it stimulates investment in human capital among the members of the disfavored group.

The extent to which these results hold depends on the firms’ ability to condition payments on the workers’ actual ability. I am aware, though, firms might have a hard time finding out a worker’s actual productivity, and therefore I allow for an imperfect detection of workers’ performance in the model.

In addition, the results of this paper have consequences on how to prove the empirical existence of
statistical discrimination itself. Even though Bertrand & Mullainathan (2004) points that employers might actively use race when hiring Altonji & Pierret (2001) suggest statistical discrimination, if at all, has small prevalence because the race coefficient in the regression does not show correlation with ex-ante wages. This model shows that two opposite effects, self-selection and statistical discrimination might co-exist in a labour market with asymmetric information. Thus explaining wage differentials by conditioning either on schooling, occupation or even ability is not sufficient to get rid of the self-selection bias. Indeed a regression of such kind might potentially yield no difference in wages between races while masking potential usage of race by employers.
2 The model

I develop an overlapping generations model with asymmetric information in the labour market. The asymmetry is generated by an employer who does not know the productivity of his potential employees. This form of asymmetric information gives rise to a problem of adverse selection as the one shown in Akerlof (1970). Akerlof, in his paper, shows that asymmetric information could even cause the cessation of market activities. Typical remedies to the problem of asymmetric information include the use of signals from the education system and ex-post observations of workers’ on-the-job performance, hereinafter referred as ex-ante and ex-post signals, respectively.

The work developed in this paper has the spirit of Coate & Loury’s (1993) signaling model, although wages are endogenous following Moro & Norman’s (2004) general equilibrium model. A crucial innovation that is not present in the two papers just mentioned is the ability of the firms to condition wages on observed performance. When allowing the use of ex-ante and ex-post signals by firms a model of double screening arises. This feature has scarcely been utilised in the literature but is implemented by Gall et al. (2006).

My model also features endogenous human capital accumulation. This part is crucial to show the effects of firms’ current hiring policies on tomorrow’s educational choices. In this model workers have the typical motivations for a worker investing in human capital: they expect stronger signals and to become more productive when handling demanding tasks.

The description of the model starts with the agents (workers and firms) and their strategic interactions. Then a definition of equilibrium is provided and a method for how to solve the equilibrium with one and two groups, respectively.

2.1 Workers

2.1.1 Demographic structure

Every period a unit continuum of workers is born. Workers belong to one of the two existent groups, \( g \in \{W, B\} \). The size of each group is given by \( \lambda^W \) and \( \lambda^B \) with respect to groups W and B. \( \lambda^W + \lambda^B = 1 \). I assume that group identity is easily observable and public knowledge.

Each generation of workers lives for three periods. They represent the workers’ youth, adulthood and retirement. Workers are endowed with one unit of time which is inelastically supplied in their adulthood.

2.1.2 Types and skills

The type of the workers is given by the cost of investing in human capital. This disutility, denoted \( \phi \), is randomly drawn from the distribution \( F^g_\phi \) whose shape might depend on the group.

When young, a worker willing to face the cost \( \phi \) and invests in his human capital becomes high skilled (H), otherwise he becomes low skilled (L). I denote the investment in human capital with an

\[ \text{See Becker (1975) or Spence (1973)} \]
indicator function; $I = 1$ if a worker invests and $I = 0$ if he does not. The cost of investing in human capital is assumed to enter the utility function linearly.\footnote{This is assumed for simplicity. It does not affect qualitatively the results shown.} Finally I denote the share of each group’s high skilled individuals by $\pi^g$.

### 2.1.3 Preferences

Workers derive utility from consuming during their adulthood and their retirement.\footnote{They do not explicitly consume while young. I can assume current adults feed the new generation and that is embedded in their utility.} Their preferences over consumption are summarised by an isoelastic utility function that weights consumption during adulthood $c^a$ and old age $c^o$,

$$u(c^a, c^o) = \frac{c^{1-\epsilon}}{1-\epsilon} + \beta \frac{c^{1-\epsilon}}{1-\epsilon} - \phi,$$

where $\epsilon$ is the coefficient of relative risk aversion and $\beta$ the discount rate. I assume workers face a borrowing constraint. Given wages when they are adults and old, $\{w^a, w^o\}$, the indirect utility function -net of investment costs- is defined as:

$$V(c^a, c^o) = \max_{c^a, c^o} u(w^a, w^o) \quad (2)$$

### 2.1.4 Outside options

High skilled and low skilled workers have outside options denoted by $\{\bar{U}^H, \bar{U}^L\}$ for $g \in \{w, b\}$. Outside options reflect the utility that these workers could obtain if they put their labour back in the market.

### 2.2 Firms

There are two infinitely-lived firms competing à la Bertrand. These firms are myopic \footnote{The myopia is assumed to prevent the model being in steady state all the time. By doing this I obtain dynamics to the steady state.} because they only take into account current profit when deciding their actions. At time 0, I assume that firms hold beliefs on the share of high skilled workers in each group $\{\bar{\pi}^w, \bar{\pi}^b\}$.

#### 2.2.1 Production

Firms produce output, $Y$, that represents a final consumption good. Production of the good requires the performance of both a simple and a complex task. The technology that firms use to produce the consumption good is represented by a CES function mapping complex and simple tasks’ efficiency units of labour $\{Y : \mathbb{R}_+^2 \to \mathbb{R}\}$ into output,

$$Y = A(\alpha C^\gamma + (1 - \alpha)S^\gamma)^{1/\gamma}, \quad (3)$$
where $A$ stands for the total factor productivity, $\alpha$ is the output share associated with the complex task and $\gamma$ determines the degree of substitutability between the efficiency units of each task. The production function satisfies the Inada conditions,

$$\lim_{C \to \infty} \frac{\partial Y}{\partial C} = \lim_{S \to \infty} \frac{\partial Y}{\partial S} = 0, \text{ and}$$

$$\lim_{C \to 0} \frac{\partial Y}{\partial C} = \lim_{S \to 0} \frac{\partial Y}{\partial S} = \infty. \quad (4)$$

The complex task is done more efficiently by high skilled workers (with productivity $A^g_h$) than by low skilled workers (with productivity $A^\ell$), with $A^g_h > A^\ell > 0$ for $g \in \{W, B\}$. On the contrary, both types are equally efficient at doing the simple task, with productivities equal to $1^5$. The efficiency units of labour for each task are made of the mass of high ($H$) and low ($L$) skilled workers performing those tasks times their respective productivities. Note the superscripts $c$ and $s$ reflect the task performed.

$$C = \sum_{g \in \{W, B\}} A^g_h H^c_g + A^\ell L^c \quad (6)$$

$$S = H^s + L^s \quad (7)$$

### 2.2.2 Information

Firms have a belief, $\bar{\pi}^g$, about the proportion of high skilled workers in each group, yet they do not know, within a group, who is high skilled and who is not. This information asymmetry generates a problem of adverse selection. To counter this problem I assume that firms observe two noisy signals that are correlated with the skill level. Adverse selection is, as a result, partially mitigated because firms have the opportunity to use the signals to improve the skill composition of the workers on each task.

**Ex-ante signal:** The first of these signals is denoted $\theta$. This signal is randomly drawn from one of two distributions $F^g_h$ or $F^\ell$. High skilled workers send a signal drawn from $F^g_h$ while low skilled workers draws it from $F^\ell$, with $F^g_h$ first order stochastically dominating $F^\ell$. Employers observe this signal at the moment of hiring workers. The ex-ante signal is best thought as a combination of factors that are observable by a firm when meeting an applicant. These factors might include the CV, the job interview or the neighbourhood where he lives, among others.

**Ex-post signal:** The second signal, denoted $y$, is generated after the completion of a complex task by a worker. The ex-post signal takes on two values, ‘pass’ or ‘fail’ $y \in (p, f)$. High skilled workers

---

5This assumption does not affect the final result. The firms’ objective of matching high skilled workers with complex tasks would remain unaltered had I assumed low skilled workers were more productive than high skilled workers at doing simple tasks.
performing the complex task send out a ‘pass’ signal with probability $\frac{6}{389}$ one. On the contrary, the low skilled workers who perform the complex task send out ‘pass’ signals with probability $\eta \in (0, 1)$. I refer to $\eta$ as the probability of fooling the firm. Low skilled workers are heterogeneous in the probability of obtaining a ‘pass’ signal as $\eta$ is a random variable drawn from the distribution $F_{\eta}$.

### 2.2.3 Hiring game

When workers become adults they randomly choose one of the two existing firms \[7\]. At this point there is an exchange of information between the firm and the worker; the worker sends a signal, $\theta$, to the employer and, at the same time, he gets to know the probability, $\eta$, of obtaining a ‘pass’ signal if he ended up doing complex tasks.

Then, the firm offers a contract that involves either performing simple or complex tasks. Contracts in our setting are defined as follows:

**Definition** A contract between a worker and a firm is defined as a triplet $\{w^{a}, w^{o}, T\}$ that specifies the ex-ante wage $w^{a}$, the ex-post wage $w^{o}$ and the task $T \in \{C, S\}$ that will be performed \[8\].

For the simple task the firm offers a fixed ex-ante wage $w^{a} = w^{s}$ and no ex-post wage $w^{o} = 0$. None of the payments in the simple task contract depend on the signals because firms know with certainty the workers’ productivity when doing simple tasks. The contract for the complex task uses, though, both signals. The form of the ex-ante wage is given by $w^{a} = p(\theta, \pi^{g})w^{c}$, where $p(\theta, \pi^{g})$ is the Bayesian posterior probability of being high skilled given the group and the signal, $p(\theta, \pi^{g}) = \frac{\pi^{g}f_{h}(\theta)}{\pi^{g}f_{h}(\theta) + (1 - \pi^{g})f_{l}(\theta)}$. \[8\]

---

\[6\] This is a simplification that does not affect the results. I could assume a lower probability insofar the signal remains informative.

\[7\] Firms are assumed to be identical. This means all firms receive the same signal $\theta$ from each worker and it is equally easy -or difficult- for a worker to receive a ‘pass’ signal from any of them.

\[8\] In terms of timing $w^{a}$ is given when a worker starts his adulthood and $w^{o}$ as soon as he becomes old and retires.

\[9\] The firm has no incentive to postpone the payment and so I make the calculations simpler by taking away this option.
In turn, the ex-post wage is given by:

\[ w^o = \begin{cases} 
  w^p & \text{if } y = p \\
  0 & \text{otherwise} 
\end{cases} \]

(9)

where \( w^p \in \mathbb{R}_+ \) and firms pay \( w^p \) conditional on the ex-post signal being a ‘pass’. The structure of the ex-ante payment in the complex task contract is taken from Moro & Norman (2003) and I keep it to allow for comparability in the results.

Firms can potentially use the ex-ante and ex-post signals to improve the sorting of high skilled workers into complex tasks; this is done in two ways called Firms’ Selection (FS) and Self-Selection (SS). On the one hand, Firms’ Selection refers to the use of hiring thresholds for each group. Hiring thresholds, \( \{\tilde{\theta}^w, \tilde{\theta}^b\} \), represent the minimum signal a worker needs to be offered a complex task job. Since those who invest in human capital receive, on average, higher signals, setting hiring thresholds may allow the firms to attract more high skilled workers to the complex task.

On the other hand, Self-Selection refers to type revelation by low skilled workers initially assigned to a complex task. We denote as \( F_D \) the proportion of low skilled workers invited to perform the complex task and who accept the invitation, hereinafter called ‘deceivers’. Firms can decrease the size of \( F_D \) with the help of contracts by making deception less attractive\(^10\).

Given hiring thresholds and contracts for simple and complex tasks, I obtain the mass of high and low skilled workers firms expect to hire for each task,

\[
H^c = \sum_{g \in \{W,B\}} [1 - F^g_h(\tilde{\theta}^g)]\tilde{\pi}^g, \quad L^c = \sum_{g \in \{W,B\}} [1 - F^g_l(\tilde{\theta}^g)](1 - \tilde{\pi}^g)F^g_D \quad \text{(10)}
\]

\[
H^s = \sum_{g \in \{W,B\}} F^g_h(\tilde{\theta}^g)\tilde{\pi}^g, \quad L^s = \sum_{g \in \{W,B\}} F^g_l(\tilde{\theta}^g)(1 - \tilde{\pi}^g) + [1 - F^g_l(\tilde{\theta}^g)](1 - \tilde{\pi}^g)[1 - F^g_D]. \quad \text{(11)}
\]

In the above definitions \( H \) and \( L \) stand, respectively, for the mass of high and low skilled workers while the superscripts \( c,s \), stand for the task to which they are allocated. As an example, to obtain the expected mass of high skilled workers performing the complex task I multiply the perceived share of high skilled in society \( \tilde{\pi}^g \) times the mass of high skilled with signals above the hiring threshold \([1 - F^g_h(\tilde{\theta}^g)]\). Calculations for the mass of low skilled on each task follow a similar logic. It should be noted that some low skilled workers initially assigned to a complex task may end up performing simple tasks after revealing their type (the fraction \([1 - F^g_D]\)).

\(^{10}\)Think, for instance, of higher \( w^s \) or \( w^p \).
2.2.4 Profit

The profit function consists of the revenue earned by the firm minus the wages paid to the workers. Note the price of the consumption good is normalised to 1 and profit is

\[ P(\hat{\theta}, w) = Y(\hat{\theta}, w) - \sum_g \omega^g(\hat{\theta}, w). \]  

(12)

Output is defined in equation 1.5; even though the expression in 1.5 does not explicitly depend on either wages or hiring thresholds it does so implicitly through the efficiency units. The expected payroll is calculated assuming that the law of large numbers applies. In particular, the ex-ante payments for workers performing the complex task are calculated by multiplying the expected posterior probability of those who accept the job times the base payment \( w^c \). The ex-post, or performance payments are calculated by multiplying the expected mass of workers with a ‘pass’ signal times the individual payment \( w^p \). Finally, workers doing the simple task are paid \( w^s \). The formula to calculate the payroll is

\[ \omega^g = w^{c,g}(E[p(\theta)]_{I=1}^g H^{c,g} + E[p(\theta)]_{I=0}^g L^{c,g}) + w^{p,g}(H^{c,g} + E[\eta]_{D=1}^g L^{c,g}) + w^{s,g}N^{s,g}, \]  

(13)

where \( E[p(\theta)]_{I=1}^g \) is the expected Bayesian posterior probability of investors doing complex tasks from group \( g \) and \( E[\eta]_{D=1}^g \) is the expected probability of obtaining a pass in the ex-post signal by low skilled workers doing the complex task.

Updating of beliefs: I assume the actual share of high skilled workers can be recovered by the firm after the tasks are done. Firms use this information to update their current beliefs as follows:

\[ \bar{\pi}_{t+1}^g = q \bar{\pi}_{t}^g + (1 - q) \pi_{t}^g \quad \text{with} \quad q \in (0, 1) \]  

(14)

where \( q \) is an exogenous smoothing parameter and \( \pi^g \) is the current share of high skilled workers of the \( g^{th} \) group. This updating rule generalises the one used in Coate & Loury (1993) and I keep it for its simplicity\(^{11}\). Note that the firm possesses private information about former workers but, since the information is learnt after they have left the firms it does not matter for the analysis.

2.3 Equilibrium

The concept of equilibrium of the economy is Bayes-Nash. Since the properties of the equilibrium differ depending on the time span I define the equilibrium in the short-run and in the steady state. The difference lies in the correctness of the firms’ beliefs with respect to the workers’ investment rate. These beliefs can be wrong in short-run but not in the steady state. Since the short-run equilibrium contains the steady state as a sub-case I define it first.

\(^{11}\)See Kim & Loury (2009) for a version of Coate & Loury with forward looking agents.
Equilibrium in the short-run: The equilibrium in the short-run is characterised by the agents maximising their objective functions in addition to the firms updating their beliefs.

Definition Given a set of initial beliefs, \( \{\tilde{\pi}_w, \tilde{\pi}_b\} \), allocations, \( \{\hat{c}_a,i, \hat{c}_o,i\} \), firms’ strategies, \( \{\hat{\theta}_W, \hat{\theta}_B\} \), outside values, \( \{\hat{U}^{H,g}, \hat{U}^{L,g}\} \), and wages \( \{\hat{w}_c,g, \hat{w}_p,g, \hat{w}_s,g\} \), constitute a Bayes-Nash equilibrium if they are such that:

i. Given market outside options \( \{\hat{U}^{H,g}, \hat{U}^{L,g}\} \) and firms’ beliefs, firms’ strategies \( \{\hat{\theta}_W, \hat{\theta}_B\} \) and wages \( \{\hat{w}_c,g, \hat{w}_p,g, \hat{w}_s,g\} \) solve the firms’ problem.

ii. \( \forall i \in (0,1) \), given firms’ strategies, \( \{\hat{\theta}_W, \hat{\theta}_B\} \), and wages, \( \{\hat{w}_c,g, \hat{w}_p,g, \hat{w}_s,g\} \) solve the workers’ investment problem.

iii. \( \forall i \in (0,1) \), given firms’ strategies, \( \{\hat{\theta}_W, \hat{\theta}_B\} \), and wages, \( \{\hat{w}_c,g, \hat{w}_p,g, \hat{w}_s,g\} \) solve the workers’ deceiving problem.

iv. Firms’ expected profit given their beliefs are 0.

v. Beliefs are updated according to equation 1.13.

Steady state: A steady state of this economy is characterised by a sequence of short-run equilibria such that the state variables (firms’ beliefs) remain constant. A formal definition follows.

Definition The steady state of the economy is a Bayes-Nash equilibrium constituted by outside options, \( \{\hat{U}^{H,g}, \hat{U}^{L,g}\} \), and firms’ beliefs, \( \{\pi_w, \pi_b\} \), such that the wages \( \{\hat{w}_c,g, \hat{w}_p,g, \hat{w}_s,g\} \), and the hiring thresholds, \( \{\hat{\theta}_W, \hat{\theta}_B\} \), set by the firms cause the workers’ investment rates to be \( \pi_W = \tilde{\pi}_W, \pi_B = \tilde{\pi}_B \).

2.4 Solving for the equilibrium

This section solves for the equilibrium of the model with double screening. I show how the agents solve their problems in the same order as the decisions are taken for a given generation of workers. First, firms, given their beliefs, decide the wages and the hiring thresholds. Workers then use the firms’ actions to decide whether to invest in their own human capital or not. Lastly, low skilled workers who are invited to perform the complex task decide whether or not to deceive the firm.
2.4.1 The firm’s problem

Since the two firms in the economy are identical I show the behaviour of a representative firm behaving as in perfect competition. Given beliefs \(\{\tilde{\pi}_w, \tilde{\pi}_b\}\) and outside options \(\{\tilde{U}^{H,g}, \tilde{U}^{L,g}\}_{g \in \{W,B\}}\) the firm solves,

\[
\max_{\theta, w} P = Y(\bar{\theta}, w) - \omega(\bar{\theta}, w) \quad (15)
\]

subject to four individual rationality constraints,

\[
\tilde{U}^{H,g} \leq [1 - F_h(\bar{\theta}^g)] V_{h,c}^E + F_h(\bar{\theta}^g) V^s \quad IR.H^g \quad (16)
\]

\[
\tilde{U}^{L,g} \leq [1 - F_l(\bar{\theta}^g)] V_{l,c}^E F_D^g + [1 - F_l(\bar{\theta}^g)] V^s [1 - F_D^g] + F_l(\bar{\theta}^g) V^s \quad IR.L^g \quad (17)
\]

and four incentive compatibility constraints,

\[
V_{h,c}^E \geq V^s \quad \forall \theta > \bar{\theta}^g \quad IC.H^g \quad (18)
\]

\[
V^s \geq V_{l,c}^E(\theta, \eta), \quad \text{for some } (\eta, \theta) \quad IC.L^g \quad (19)
\]

where output is given by equation 1.5 and the expected payroll of the firm is defined in equation 1.13. In addition, four Individual Rationality and four Incentive Compatibility constraints must be satisfied. On the one hand, the four IR constraints reflect that the firm must offer at least the outside option for a worker to choose that firm as his workplace, i.e. it is an ex-ante constraint. The IR constraints are group and skill dependent. Also note that the value functions should have a group superscript since wages may differ between groups; these are omitted for clarity.

The right hand sides of equations 1.16 and 1.17 reflect the expected utility of, respectively, a high and a low skilled worker, before taking the decision of whether to invest in human capital or not. In terms of notation, value functions are assigned superscripts depending on the task \(\{c, s\}\) and the skill of the worker \(\{h, l\}\). More specifically, \(V_{h,c}^E\) is the expected utility that a worker would obtain by investing if he were chosen to perform the complex task. Likewise, \(V_{l,c}^E\) is the expected utility a worker would obtain if he did not invest but was given the opportunity to perform the complex task and accepted a contract to do so -a deceiver-. Lastly \(V^s\) is the utility drawn from performing simple tasks.

On the other hand, I add four IC constraints to the problem of the firm, one per group and skill. First, the IC.H constraints make sure high skilled workers assigned to the complex task do not have an incentive to pretend to be low skilled and thus to switch to the simple task. Then, the IC.L constraints have the opposite purpose; they are intended for low skilled workers assigned to the complex task to

\[\text{Expected because the ex-ante wage depends on the signal and the signal itself is a random variable.}\]

\[\text{There is no expectation as this value is known with certainty by the workers.}\]
leave and perform the simple one instead.

It is important to notice that $IC.H^g$ is, in equilibrium, satisfied and thus no high skilled worker assigned to a complex task would pretend to be low skilled. This is not necessarily true in the case of the $IC.L^g$, and a semi-separating equilibrium is normally achieved. Only in special cases would this constraint be satisfied for all low skilled workers and a pure separating equilibrium would exist.

2.4.2 Workers’ problems.

Workers’ payoffs depend on whether they become high skilled or not, and -only for the low skilled workers who are invited to perform the complex task- on whether they deceive their employer or not. Next we show how these decisions are taken and aggregated.

**Human capital decision:** Young workers make their human capital acquisition decisions based on their investment costs and expected future earnings. I define the benefit of investing as

$$\xi^g = E[U^g|I = 1] - E[U^g|I = 0] \quad g \in (W, B),$$

where $E[U^g|I = 1]$ and $E[U^g|I = 0]$ are the expected utility of a member of group $g$ who, respectively, has and has not invested in human capital. Given information on contracts and hiring thresholds workers compare the payoffs of investing and not investing. This is done by weighting the value at each state by its respective probability.

$$E[U^g|I = 1] = [1 - F_h(\bar{\theta}^g)]V_{E}^{h,c} + F_h(\bar{\theta}^g)V^s \quad g \in (W, B)$$

Every worker whose benefit to investing is higher than his utility cost acquires human capital. The share of each group’s workers that choose to acquire human capital, $\pi^g$, is computed as the mass of workers whose investment cost is below the benefit to invest $\pi^g = F^g_{\xi}(\xi^g)$.

**Deceiving decision:** Workers invited by the firm to perform complex tasks have to decide whether to accept it or to reject the offer. The incentive compatibility constraint $IC.H$ ensures all high skilled workers assigned to the complex task will accept the invitation. On the contrary, low skilled workers invited to perform complex tasks must decide whether they want to deceive the firm or not. Given workers’ knowledge of their ex-ante signals, $\theta$, and their probabilities, $\eta$, of fooling -obtaining a ‘pass’ signal $y = p$- the firm, I define a worker’s expected utility of deceiving ($D=1$) and revealing the type ($D=0$) as,

$$E[U^g|D = 1] = V_{E}^{l,c}(p(\theta)w^{c-g}, w^{p-g}, \eta) \quad g \in (W, B)$$

---

14Suppose $V^s = V^{l,c}$, $\forall (y, \theta)$, then $V^s = V^{h,c}$, $\forall \theta$ and no worker wants to perform complex tasks.

15For instance if $E[\eta] = 0$ and $w^s = w^c$, setting $w^c \leq w^s$ guarantees full separation.
\[ E[U^g|D = 0] = V^s(w^g, 0), \quad g \in (W, B). \] (24)

To calculate the share of deceivers I first define the benefit of deceiving the firm,

\[ \delta^g = E[U^g|D = 1] - E[U^g|D = 0], \quad g \in (W, B). \] (25)

Workers with combinations of \((\theta, \eta)\) such that \(V^l_c > V^s\) will attempt to deceive the firm \((D=1)\) and perform the complex task. The share of these low skilled workers invited to perform the complex task who deceive the firm is given by the joint CDF of the signals and the fooling probabilities denoted \(F^g_D\).

\[ F^g_D = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \int_{\eta_{\text{min}}}^{\eta_{\text{max}}} f_i(\theta, \bar{\theta}, \infty) f_s(\eta, \eta', \eta^h) d\eta d\theta \] (26)

In addition, I need to calculate the expected probability of obtaining a ‘pass’ signal \(y = p\) given deception \(E[\eta|D = 1]\). This is given by:

\[ E[\eta|D = 1] = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \int_{\eta_{\text{min}}}^{\eta_{\text{max}}} f_i(\theta, \delta, \infty) f_s(\eta, \eta_{\text{min}}, \eta^h) d\eta d\theta \] (27)

Note that in equation 2.27 the lower truncation of \(f_s\) depends on \(\eta_{\text{min}}\). Also the lower bound of the inner integral \(\eta_{\text{min}}(\theta)\) depends on the outer variable. The function \(\eta_{\text{min}}(\theta)\) is obtained by first making \(\delta = 0\). This function combines pairs of signals and fooling probabilities such that a worker is indifferent between performing complex or simple tasks. If I solve in terms of the fooling probability\(^{16}\) the result is

\[ \eta_{\text{min}} = \frac{V^s}{V_E^{l,c}(p(\theta)w^c, w^p)^{-1}}. \] (28)

This equation provides the lower bound fooling probabilities that are needed for a worker to deceive given his signal. See Appendix A for extra details on these calculations.

### 2.4.3 Outside options

The outside options, \(\{\bar{U}^H_g, \bar{U}^L_g\}\), of the workers are determined in equilibrium using the zero-profit condition. Throughout the thesis I assume two settings; one where firms’ beliefs are the same for both groups and another one where beliefs may differ. Since the strategy to pin down the outside options is slightly different in each of the cases I start with the simplest one; the case where firms’ beliefs are the same.

**Same beliefs:** If the beliefs are the same firms’ behave as if there were just one group. Therefore, there are two outside options, \(\{\bar{U}^H, \bar{U}^L\}\), one per skill, and I omit the group superscript. To pin them down I choose \(\bar{U}^H\) such that the profit of the representative firm is equal to 0. Then, \(\bar{U}^L\) is given by the expected utility of not acquiring human capital (right hand side of equation 1.20). Note I do not

\(^{16}\)Provided the inverse of \(V\) exists.
Amending the Racial Gap, What Helps and What Does Not?

impose a value for $\bar{U}^L$.

The way to pin down $\bar{U}^L$ as a by-product of $\bar{U}^H$ follows an insurance motive. Given the imperfection of the educational signals, in the unlucky event that a high skilled worker is sent to perform the simple task, a higher salary, $w^*$, helps the firm to reduce the overall payroll. This procedure has implications for the functioning of the labour market. Under perfect information, and assuming workers cannot opt for unemployment benefits and have no other form of wealth, the procedure would lead to low skilled workers being paid the minimum necessary to survive (zero in my model).

**Different beliefs:** In the model with two groups there exist 4 outside options $\{\bar{U}^{H,g}, \bar{U}^{L,g}\}_{g \in \{W,B\}}$, one per group and skill. It is obvious that, for the 2 groups model to be of any interest, there must exist a difference between the firms’ beliefs about the groups W and B. Under this condition the strategy to pin down the outside options takes as a starting point a colour-blind steady state where firms have the same belief about each group. At this stage I shock the economy by shifting down the belief of group, say $b$, and I calibrate the outside options of the high skilled, $\{\bar{U}^{H,w}, \bar{U}^{H,b}\}$, until zero profit holds. The choice of changing $\bar{U}^{H,w}$ or $\bar{U}^{H,b}$ depends on the initial profit after the shock. If the profit is positive I shift up $\bar{U}^{H,w}$, if it is negative I shift $\bar{U}^{H,b}$ down.

**2.4.4 Market clearing**

Two markets must clear in each period, a goods market and two labour markets.

**Goods market:** Equilibrium in the goods market means that the amount of good supplied equals the amount demanded. In the short-run equilibrium the demand of goods might be higher that the production (if the firms’ beliefs are wrong), in that case I assume firms hold a finite stock of the consumption good that is used to keep up their payment promises. The equation for the goods market equilibrium is

$$C_t = Y_t + S_t,$$  \hspace{1cm} (29)

where $C$ is total consumption, $Y$ is output and $S$ is stocked consumption good by the firm.

**Labour markets:** Firms’ demand for high and low skilled workers shape the market. Since workers supply their labour inelastically labour markets always clear.

$$N^d_H = N^s_H \quad ; \quad N^d_L = N^s_L$$  \hspace{1cm} (30)

Firms demand for workers is set when they announce hiring thresholds and wages. If their beliefs are wrong the mass of workers with signals above and below the hiring threshold may not be the expected one. In those cases I allow the firm to change the hiring threshold. If the mass of workers

---

17 By Jensen’s inequality as the utility function is concave.

18 The strategy to pin down the outside values for the colour-blind steady state is the same as that sketched for the one group model.
with signals above the threshold is higher than the expected one, then market clearing occurs by the firm tightening the hiring threshold for the complex task, thus only accepting as many workers as were expected (those with the highest signals). The reverse is partially true and the firm would lower the threshold to accept more workers but only up to $\bar{\theta}_{\text{min},g}$; where $\bar{\theta}_{\text{min},g}$ is the lowest signal such that the $IC.H^g$ constraint is still satisfied, i.e. $V_{h,c} = V_s$.

3 Convergence of earnings under discriminatory beliefs

This part of the chapter can be thought of as a continuation of a literature on statistical discrimination, in particular the models developed by Coate & Loury (1993) and Moro & Norman (2004). An important concern shared with part of the literature on statistical discrimination is whether an initial negative belief towards one of the groups might be self-fulfilling or not. This recurrent theme can be found at Coate & Loury (1993), Moro & Norman (2004) or at the review by Moro & Fang (2010).

I provide simulations of the two groups' model developed in Chapter One to compare the evolution of the firms' beliefs with other papers in the field. Note the economy might end up in either of the two potential steady states. The first of these steady states is a divergent equilibrium, implying that the firms' beliefs with regards the B group diverges to 0. The second steady state is a convergent equilibrium where the beliefs converge to the colour-blind steady state value. Divergence to 0 would put us in a case where due to the low share of high skilled in the black population firms decide not to hire them for the complex task.

Simulations are run in a scenario where firms hold a worse belief towards one of the groups, maybe due to some historical reasons. This seems to capture well the situations of, for instance, blacks and women. For comparability with past papers in the area we are interested in equilibrium where no ex-ante differences exist between the groups. We can, thus, wonder whether an economy where one group is initially disadvantaged is able to reach a race-neutral steady state without help. A group is said to be disadvantaged if firms believe it possess a share of high skilled workers lower than at the other group.

I assume, throughout this section and the remainder of the paper, that B is the disadvantaged group $\tilde{\pi}_w > \tilde{\pi}_b$. The reader should keep in mind that simulations in this section assume that B and W workers obtain, on average, the same advantages and face the same costs from investing in human capital,

$$A_w = A_b \quad F_w = F_b \quad \text{and} \quad F_w = F_b.$$  \hspace{1cm} (31)

Simulations are provided under two economic environments, one where firms have to pay the same wages $\{w^c, w^p, w^s\}$ to both groups (as it is assumed in the literature) and another where I allow complete contract freedom.
3.1 Without contract freedom

The purpose of this numerical experiment is to compare the effect of performance pay in a setup as close as possible to past papers in the area. To accomplish that goal I remove firms’ ability to choose different wages for each group. This assumes $w^c.w = w^{c,b}$, $w^s.w = w^{s,b}$ and $w^{p,w} = w^{p,b}$. Still the firm is able to use group membership in the Bayesian posterior and to use different hiring thresholds as I assume it cannot be stopped from doing that (i.e. legally more difficult to prove).

Figure 1: Convergence. Double screening vs. Firm’s Selection (Signaling)

Figure 1 shows two economies. One where double screening is used by firms (left), and another one where only ex-ante signals exist as a method of screening. The way to interpret the graph is the following, points above the 45 degree line are convergent (i.e. if I let the firm update its beliefs we arrive at a convergent steady state). Points below the 45 degree line are divergent. We can observe that the addition of performance pay greatly reduces the divergent set of points.

I show performance pay helps to reduce the set of initial beliefs where a divergent equilibrium in the sense described in this section occurs. Even though I cannot state the necessary conditions for convergence to be the prevailing situation independent of the initial belief, I can say the existence of convergence is positively associated with the quality of the detection technology. That is, in an economy where performance pay exists, the possibility of being in a divergent situation would be less likely to occur.

3.2 With contract freedom

The second numerical experiment allows the firms to set different wages for different groups. Our starting point assumes a colour-blind steady state where firms’ beliefs are such that $\tilde{\pi}^w = \tilde{\pi}^b$. Then we ask what would happen if the belief with respect to the B-group investment rate deteriorates.

The model implies that no matter how low is the belief for the B group there exists convergence between the beliefs of the two groups. This result does not depend on the existence of performance pay though, rather, it depends on allowing the firm to choose different contracts for different groups. As a result firms’ hiring policies under contract freedom tend to promote the B group investment in human capital.
We can see this result in figure 2. I modify one of the state variable (firms’ belief of blacks) to see how the economy reacts. In particular we show reaction functions of the B and W groups for a range of beliefs for the B group $\tilde{\pi}^b \in (0, \pi_{ss})$. I fix the firms’ belief about the W group equal to the colour-blind steady state investment rate, $\pi_{ss}$. Reactions above the 45 degree line are convergent, below it they are divergent. Convergence means that if I keep iterating the model, the economy will enter the colour-blind steady state. Divergence implies that the belief about the discriminated against group tends to 0 in steady state. A divergent equilibrium is more likely the more the firm rely on W workers than on B workers, thus reducing the benefit to invest of group B.

Next I turn to the screening behaviour of the firm when the belief towards group B becomes discriminatory. The idea we want to show here is that the firm maintains the intensity of each screening mechanism. In order to maintain the strength of each screening method firms have to modify the wage structure. The wage structure is changed so that payments for group B workers on the complex task are higher and the B group simple task wage becomes lower. This modification increases the incentive to invest, which explains the reaction function shown in Figure 2.

The explanation of the screening behaviour can be seen in Figure 3 and 4. Figure 3 shows on the right side of the graph the proportion of low skilled workers that are correctly sent to perform simple tasks, we call this Firm’s Selection. On the left side of the graph I show the proportion of low skilled that is *invited* to perform complex tasks but choose to opt out and perform the simple task; this is called Self-Selection. On the x-axis there is a range of beliefs for group B. I do not set $\tilde{\pi}^b$ to start at 0 because the little knowledge it adds does not compensate the computational demands it requires. I observe that the screening intensity is maintained for both groups with one exception. The firm relaxes the amount of W workers self-selected. This is a technological issue. Since the firm thinks there are less workers who invest in the economy (from group B) the employer decides to re-address this issue.
by using W group workers.

Then in Figure 4 I show the differences between group B and group W’s wages. We can see the reaction of the firm is to make deceiving less appealing for group B (the one perceived to have invested less) by raising the complex task base payments, $w_c$ and $w_p$. Notice the fact that the firm, given total freedom to use the wage mechanism, chooses wages that maintain almost constant the screening intensity for both groups. This is remarkable because in versions of the model where the firm is forced to set $w_b = w_w$ we see different screening behaviour from each group. In particular we observe that the firm relies more on self-selection for group B and on Firm-Selection for the group W.

Two lessons are worth remembering. The first one is that when the firm is allowed to set different wages for each group, the divergent equilibrium does not exist. The model will never have a B group that does not invest in human capital at all. The second lesson is that, even when the firm is not allowed to set different wages, the addition of performance pay shrinks the set of group B beliefs that
lead to a divergent equilibrium. This can be generalised to shocks similar in effect to an improvement in the detection technology. These include positive developments in financial markets that allow workers to better endure performance payments.

4 The case of the US

In section 3 we work with a model of statistical discrimination where both groups are ex-ante equal. I show convergence between two groups does not take longer than 2 to 3 generations. This section starts by wondering why we do not see this in countries like the US, where the racial wage gap has been persistent since the 19th century.

Our aim in the remainder of the paper is to explain the persistence of the US wage gap and why the racial wage gap has not vanished in the US. I do so by offering a computational argument to shed some light on the effect of certain policies. To explain the persistence of the US racial wage gap I now model the role of pre-market factors. Since these are factors differ between races in the model, I can ask not only how policies affect employment outcomes conditional on pre-market factors but also how pre-market factors respond to policies.

Indeed, past models of statistical discrimination can be criticised for lacking an explanation of the observed racial wage gap persistence. This criticism, though, could be judged as unreasonable. These models assume both groups are ex-ante identical and just cover the labour market interactions. Research papers like Heckman (1998) or Neal & Johnson (1996) suggest that different groups do not enjoy the same opportunities before they start working. Their conclusions imply the aforementioned models are just incapable of providing a satisfactory explanation of the race gap's persistence.

In fact, Neil & Johnson argue that most of the wage differential between races could be explained by a measure of actual ability. Heckman (1998) also delves into the same argument when explaining that the use of certain variables in Mincer Equations is misleading when assessing differences in earnings. An example of Heckman’s line of thought relates to the use high school graduation to assess differences in return between blacks and whites. The problem in this case arises because the census data uses as GED (General Equivalence Degree) and graduation from high school as equivalents while the signal sent by each holder is arguably different. Since more blacks than whites hold GED certificates, a standard regression using the generic variable ‘High school graduate’ would wrongly point to black high school graduates being discriminated against. On a more general level Heckman argues that if we could control for what the firm sees (not just what the dataset tells us) the explained portion of the wage would rise from 20%-30% to 60%-80%. To sum up, it can be argued that assuming the residuals from a Mincer Equation can represent discrimination that seems to be far from correct.

My model of double screening allows me to take this argument even further though, by arguing that not only could such variables as years of schooling or occupation be misleading when assessing the

\[19\] See, for instance, Moro & Norman (2004) or Blume (2006)
existence of discrimination, but variables such as the AFQT\textsuperscript{20} could too. This is because the share of black workers performing simple tasks might contain a higher proportion of self-selected individuals in the sense described in this thesis. Thus, finding that both races earn the same (even after controlling for some variables) is not conclusive in terms of rejecting the existence of discrimination.

In light of the results from Neal & Johnson and Heckman it could be argued that pre-market differences might be an important cause of labour market differences between races. They can be classified into two broad groups, schooling opportunities and family background. What I do in this section is extend the analysis of section 3.2 by allowing for pre-market differences between the two groups. We focus on the inter-temporal evolution of the workers pre-market investments in education as an explanation for the race gap persistence.

The last part of this paper is focused on convergence in earnings between blacks and whites in the US. I compare the speed of convergence under two policy regimes, a colour-neutral hiring policy that forbids the usage of race and one that allows firms to use race when hiring. The ban for using race when hiring is associated with the one implied by the title VII of the 1964 Civil Rights Act. My results suggest that allowing firms to discriminate racially in setting wages and in allocating jobs speeds up convergence, and that this result is stronger the more firms rely on performance pay. The result occurs because the incentives to invest for the disfavoured group are stronger under the unconstrained regime than under the one banning the use of race.

4.1 The racial gap in the US

It was not before the 1950’s that the racial wage gap significantly started to decrease. This is shown in Figure 5 where I provide black and white income ratios of each birth cohort. It was for the (1931-1936) cohort that the income wage gap started to decrease as pointed out in Smith (1984). Smith’s idea of showing the data by cohort is to illustrate that something happened after the 1940’s that helped blacks’ income convergence. In particular he supports the idea that convergence in schooling opportunities due to historical reasons were a main driving force.

Indeed, section 3.2’s prediction of a fast catch up of black’s income with whites’ coupled with the actual persistence of the racial income gap seem to point to the existence of certain pre-market differences between groups. These pre-market differences would be the ones preventing full convergence from happening. The importance of these differences in causing the racial wage gap can be found in the work of Heckman (1998) and Neil & Johnson(1996). Their papers mention differences in family background as well as in schooling opportunities\textsuperscript{21} as being able to explain a large part of the wage differential.

According to Heckman (1998) schooling opportunities and family support explain how a worker is perceived by his employer and his actual level of skill which, in turn, is one of the main determinants of a worker’s wage.

\textsuperscript{20} Armed Forces Qualification test.
\textsuperscript{21} See Espenshade & Radford (2009) for a review of how families affect university admissions of their children.
I adapt the model to allow for pre-market factors. In doing so I translate the different distributions of family backgrounds as the cost of investment. I assume workers with worse family background will have a harder time to remain enrolled, find time to study and pay the tuition fees. In addition differences in schooling are translated into the model via the signals and the complex task productivity. On the one hand we assume a worker’s signal depend partially on the school’s name or the networks built while in school as pointed in Borjas (1999). On the other hand, I assume some schools and universities prepare workers better for the labour market than other. In the remainder of the section I work on the assumption that differences between blacks and whites might exist in pre-market factors.

4.2 The effect of schooling differences

Educational opportunities for blacks improved slowly during the decades before 1940 as pointed out in Bowen & Bok (1998). It was not before the 1940’s that the school attendance rates rose as a result of migrational movements of southern blacks to the industrialised north, as is pointed out in Lang (2007) and Bowen & Bok (1998). The movement was due to the higher labour demand that was prevailing at the beginning of the 2nd World War.

The generation that was born during these years enjoyed an improvement of their schooling conditions. Schooling opportunities were better in the north for blacks. Lighter racial bigotry coupled with a higher proportion of blacks now living in cities helped enrolment rates. By 1960 the length of the school term and teachers salaries had almost converged between black and white schools. In addition median years of schooling for blacks aged 25-29 rose from 7 years at the beginning of the 1940’s to 10.5 at the beginning of the 1960’s. Bowen & Bok show data on graduation rates. The proportion of blacks that graduated from high school rose from 12% to 38% while the proportion who graduated from college shifted from 1.6% to 5.4%. It was a result of this new migrational pattern that southern
states started to improve the 1950’s blacks’ schooling conditions in an attempt to stop the flow of workers to the north.

The above facts describe two effects, higher labour demand and better schooling opportunities. Even though arguably the former could have helped blacks’ earnings potential it was the improvements in schooling quality and ex-ante signals that started the virtuous circle as argued in Smith (1988)

A Divergence exercise

In this subsection I simulate a colour-blind steady state where firms hold the same belief about both groups, \( \bar{\pi}_{w_{ss}} = \bar{\pi}_{b_{ss}} \). Then I lower the signal quality of the black investors. The objective is to show that the model can move from a colour blind steady state to a steady state where one group is disadvantaged by creating pre-market differences.

In contrast to the economy shown in section 3.2, the model might have an equilibrium where firms’ belief about blacks is neither zero nor the same as that about whites. A clarification is due: in section 3.2 I comment on the existence of two equilibria. One where the firms’ belief about blacks tends to 0 (the bad, divergent equilibrium), and another one where it converges to the value of the whites (the good, convergent equilibrium). Every equilibrium where the firms’ belief about blacks is neither 0 nor is the same as the firms’ belief about whites belongs to the convergent family and I call it (with some abuse of the language) an interior equilibrium. This interior equilibrium allow us to simulate economies closer to reality.

The strategy for finding the new steady state is to shock the economy with the new, lower signal quality for blacks and let the firm update its belief over time. Figure 6 shows the beliefs of the firm over time, for blacks and whites. The x-axis represents time (each unit of time is thought to represent 25 years) and the y-axis firms’ beliefs. In the graph the beliefs diverge from each other until they enter the new -interior- steady state. This implies that in the presence of pre-market differences between groups, the steady state black/white income ratio can be shifted below 1.

Black workers also suffer in terms of welfare. This can be seen in Figure 7. In this figure the y-axis represents aggregate welfare. The reason why at the beginning blacks are better off is because the firm is being fooled. The employer thinks blacks are more productive than they actually are, therefore paying and hiring more than it would have chosen.

Finally table 3.1 shows the behaviour of the firm in the new interior steady state. The representative firm chooses different screening mechanisms for each race. It uses the hiring threshold to drive out white low skilled workers (F-S, firm selection) and relies on self-selection (S-S) to get rid of the blacks deceivers.

As a summary, I have shown, by means of a simulation, that pre-market differences may play an important role in stopping the convergence of earnings between whites and blacks.

\[22\text{See Appendix C for the calculations.}\]
Amending the Racial Gap, What Helps and What Does Not?

Figure 6: Beliefs converging to a non-colour-blind steady state.

Table 1: Steady state. Discriminatory steady-state.

<table>
<thead>
<tr>
<th>Variables</th>
<th>F-S</th>
<th>S-S</th>
<th>$w^c$</th>
<th>$w^p$</th>
<th>$w^s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solutions W</td>
<td>0.5180</td>
<td>0.1585</td>
<td>5.4384</td>
<td>2.6873</td>
<td>2.8936</td>
</tr>
<tr>
<td>Solutions B</td>
<td>0.1025</td>
<td>0.7694</td>
<td>8.8930</td>
<td>3.0624</td>
<td>3.7174</td>
</tr>
</tbody>
</table>

4.3 The effect of banning race from hiring

Now we move forward to 1964, the year when the first CRA powerful enough to deter discriminatory behaviour was enacted. In particular I focus on the effects that its title VII brought to the labour market. Title VII declares it unlawful ‘to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to her compensation, terms, conditions, or privileges of employment, because of such individual’s race, color, religion, sex, or national origin; or, the usage of the worker’s race when hiring or deciding the compensation’.

This content is translated into the model’s language as a ban for different hiring thresholds, use of race in the Bayesian posterior and use of different base payments for blacks and whites. Because checking firms’ hiring thresholds might be hard, I follow Coate & Loury (1993) by allowing firms to use different ones even after the ban. With regards to the Bayesian posterior, the firm uses the average belief when computing it,

$$p(\theta, \tilde{\pi}^{\text{Ave}}) = \frac{\tilde{\pi}^{\text{Ave}} f_h(\theta)}{\tilde{\pi}^{\text{Ave}} f_h(\theta) + (1 - \tilde{\pi}^{\text{Ave}}) f_l(\theta)}$$

where the average belief is equal to $\tilde{\pi}^{\text{Ave}} = 0.5\lambda w^c \tilde{\pi}^w + 0.5\lambda b \tilde{\pi}^b$. Finally, I assume the ban imposes $w^{c,w} = w^{c,b}$, $w^{p,w} = w^{p,b}$ and $w^{s,w} = w^{s,b}$.

Did the ban help to close the black/white income gap? To answer that question I go back to the
1940’s and simulate two economies. These economies start off from the same discriminatory steady state where blacks are believed to be less productive that blacks doing the complex task \( A_h^b = 0.7, A_h^w = 1 \), to obtain worse signals \( \mu_h^b = 1, \mu_h^w = 2.098 \) and to endure higher investment costs \( \mu_b^\phi = 1.8, \mu_w^\phi = 1.2 \). Then I shock both economies by shutting the gap in signals, \( \mu_h^b = \mu_h^w = 2.098 \), and the complex task productivity, \( A_h^b = A_h^w = 1 \). This is like assuming that all enjoy the same schooling opportunities. In addition, I decrease the gap in investment costs proportionally to the increase in the share of high skilled workers of the last generation. The idea is to associate smarter parents with lower costs of investments to the point that if both groups invest in human capital at the same rate, the investment costs are the same. Table 3.2 shows parameters regarding the preferences and the technology used for the calibration of the initial steady state. The income share of the complex task, \( \alpha \), is calibrated following data from US on high-skilled/low-skilled earnings; \( \beta \) is set so that there is yearly discounting of 4%.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta )</td>
<td>0.6</td>
<td>( \alpha )</td>
<td>0.66</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>0.99</td>
<td>( \gamma )</td>
<td>0.01</td>
</tr>
<tr>
<td>( A )</td>
<td>10</td>
<td>( A_h^b )</td>
<td>0.7</td>
</tr>
<tr>
<td>( A_h^w )</td>
<td>1</td>
<td>( A_l )</td>
<td>0.2</td>
</tr>
</tbody>
</table>

In table 3.3 I show the parameters used for detection technology, signals and investment costs for each race in the initial steady state. The detection technology is assumed to be mild (the average fooling probability is 0.5) as well as the signals distributions. The investment costs’ lower bound is
set to 0 to match the domain of the incentives to invest \( \mathbb{R}_+ \), as allowing for negative investment costs would artificially create investors. In addition I assume the proportion of blacks in the economy is \( \lambda^b = 20\% \).

Table 3: Detection tech., ex-ante signals dist. and investment costs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_{\eta} )</td>
<td>0.5</td>
<td>( \sigma_{\phi} )</td>
<td>2</td>
</tr>
<tr>
<td>( \sigma_{\eta} )</td>
<td>0.4</td>
<td>( \phi^b )</td>
<td>2.6</td>
</tr>
<tr>
<td>( \eta^h )</td>
<td>1</td>
<td>( \phi^l )</td>
<td>0</td>
</tr>
<tr>
<td>( \eta^l )</td>
<td>0</td>
<td>( \mu_{\phi}^b )</td>
<td>2.098</td>
</tr>
<tr>
<td>( \mu_{\phi}^l )</td>
<td>1.2</td>
<td>( \mu_{\phi}^l )</td>
<td>1</td>
</tr>
<tr>
<td>( \mu_{\phi}^h )</td>
<td>1.8</td>
<td>( \sigma_h = \sigma_l )</td>
<td>2</td>
</tr>
</tbody>
</table>

One of the economies is called the unconstrained model, as the firm can use statistical discrimination freely. The other economy is the constrained model that mimics the enactment of the Civil Rights Act with respect to the use of race in the hiring process. The constraint means firms have to use an average belief in the formation of the Bayesian posterior and have to pay the base payments to both groups. Still we assume firms in the constrained model can set different hiring thresholds as their existence is hard to prove.

Results are shown in Figure 8, where time units represent 25 years (a generation). The graph shows faster convergence in earnings when firms statistically discriminate than when they cannot. The simulated steady state implies a high degree of segregation in the job market. Approximately 2\% of the workers doing complex tasks are black. It should be noted blacks represent 20\% of the total population in this economy.

Figure 8: Black/White simulated income ratio
5 Conclusion

This paper proposes a model of statistical discrimination with double screening and endogenous human capital accumulation. The model and the analysis performed can be thought of as the continuation of a literature on statistical discrimination, in particular the models developed by Coate & Loury (1993) and Moro & Norman (2004).

My proposal partly differs from the papers just mentioned because I allow for performance pay and I use a different procedure to find the equilibrium. More importantly I allow firms to pay different wages to workers performing the same task on the basis of which group they belong to.

Section 3 analyzes the predictions of the model assuming that both groups are ex-ante identical and that firms hold different beliefs with regards to the share of high skilled workers at each group. Two lessons emerge from the analysis; first, earnings always converge when I allow the firm to set different wages for each group. Second, had not we allowed firms to set different wages for different groups, the addition of performance pay shrinks the set of B group beliefs that lead to a divergent equilibrium. This can be generalized to shocks similar in effect to an improvement in the detection technology, including positive developments in financial markets that allow workers to better endure performance payments.

Section 4 is a policy oriented one. The section points out the persistence of the differences in black/white income ratio during the last century. I argue that the observed persistence might be due to differences in pre-market factor, as explained by James Heckman and James Smith. In consequence I allow the model to have pre-market differences in schooling and family background.

Two numerical exercises are carried out. The first exercise consist on shocking an economy in a colour blind steady state. All of a sudden firms believe that the share of high skilled workers in one of the groups is lower than it used to be. The objective is to see if my model can generate a situation like the one occurring in the US back in the 1940’s. The model shows a racial wage gap can be created and sustained over time in what we call an interior equilibrium.

Lastly I carry out an exercise to test the effect of two policies, equal education opportunities and an equal pay act. In order to make this computational exercise we calibrate an economy to mimic the black/white income ratio prevailing in the 1960’s. Then I shut down firms’ ability to pay different wages and provide the same education opportunities to both races. The simulation points that the economy where race can be used in contracts shows a faster convergence rate of blacks’ earnings than the economy where the usage of race is banned, thus giving more credit to the role of education opportunities in closing the race gap than we would initially give.
References


Appendix A: Joint CDF and expected value

In section 2 the model developed needs to calculate the joint CDF of two disjoint truncated normal random variables and the expected value of one of the variables. With regards to the former the joint CDF is defined as:

\[
F_D = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \int_{\eta_{\text{min}}(\theta)}^{\eta_{\text{max}}} f_1(\theta, \bar{\theta}, \infty) f_\eta(\eta, \eta', \eta^h) d\eta d\theta
\]  

(33)

but a word must be said about the integration limits. The upper limits of both integrals are given by the high end of both truncated normal distributions i.e. \( \theta_{\text{max}} = \infty \), \( \eta_{\text{max}} = \eta^h \). On the contrary the lower bounds calculation is more subtle. On the one hand the lower bound of the outer integral \( \theta_{\text{min}} \) might depend on the higher bound of the inner integral \( \eta_{\text{max}} \) as follows:

\[
\theta_{\text{min}} = \max\left\{ \frac{V_s}{(V^{l,c}(\eta^\text{max}))^{-1}}, \bar{\theta} \right\}
\]  

(34)

On the other hand the lower bound of the inner integral depends on the outer variable. The idea is that we must ensure the pairs \( \theta, \eta \) satisfy the incentive compatibility constraint for a low skilled to accept a complex task.

\[
\eta_{\text{min}} = \frac{V_s}{(V^{l,c}(\theta))^{-1}}
\]  

(35)

In addition we calculate the expected probability of fooling given a worker deceives \( E[\eta|D=1] \) as:

\[
E[\eta|D=1] = \int_{\theta_{\text{min}}}^{\theta_{\text{max}}} \int_{\eta_{\text{min}}(\theta)}^{\eta_{\text{max}}} \eta f_1(\theta, \theta_{\text{min}}, \infty) f_\eta(\eta, \eta_{\text{min}}, \eta^h) d\eta d\theta
\]  

(36)

where it has to be noted, in addition to what was explained before, that the lower truncation of the distributions follows the lower bounds of the respective integrals. This is because we must ensure the volume adds up to one. In addition the order of integration (first \( \theta \), then \( \eta \)) is not trivial; Since the signal occurs first in time, we need to know what probabilities of fooling are IC and not vice-versa.

Appendix B: Workers’ utility maximization problem.

In the main body of the thesis I state the workers’ utility function and its liquidity constraint. This section specify how the workers maximize their utility for a given contract that pays some money ex-ante and some money ex-post, perhaps with uncertainty.

Utility under certain payments

The utility obtained by high skilled workers in the complex task, with contract \{\( w^a = p(\theta)w^c \), \( w^o = w^p \)\}, and any worker in the simple task with contract \{\( w^a = w^s \), \( w^o = 0 \)\}, is known with certainty. To

\[23\] \( f_1 \) is not truncated from above

\[24\] Provided the inverse of \( V^{l,c} \) exists.
solve for the optimal allocation, workers maximize the following function:

$$\text{Max } U(c^a, c^o) = \frac{c^{a^{1-\epsilon}} - 1}{1 - \epsilon} + \beta \frac{c^{o^{1-\epsilon}} - 1}{1 - \epsilon}$$ (37)

subject to

$$c^a \leq w^a$$ (38)

$$c^a + c^o = w^a + w^o$$ (39)

The utility function evaluated at the optimum takes two values:

$$V = \begin{cases} 
\frac{w^{a^{1-\epsilon}} - 1}{1 - \epsilon} + \beta \frac{w^{o^{1-\epsilon}} - 1}{1 - \epsilon} & \text{if } w^a \leq \frac{w^a + w^o}{1 + \beta^{1/\epsilon}} \\
\frac{w^{a^{1-\epsilon}} - 1}{1 - \epsilon} + \beta \frac{w^{o^{1-\epsilon}} - 1}{1 - \epsilon} & \text{if } w^a > \frac{w^a + w^o}{1 + \beta^{1/\epsilon}}
\end{cases}$$ (40)

Utility under uncertain payments

If the performance pay is uncertain, as it occurs to the low skilled workers when joining the complex task, some precautionary savings will be taken in order to insure themselves against the risk of not having anything when they are old. The maximization program for a given contract \(\{w^a = p(\theta)w^c, w^o = w^p\}\) is,

$$\text{Max } U(c^a, c^{o,h}, c^{o,l}) = \frac{c^{a^{1-\epsilon}} - 1}{1 - \epsilon} + \beta \left[ \frac{c^{o,h^{1-\epsilon}} - 1}{1 - \epsilon} + (1 - \eta) \frac{c^{o,l^{1-\epsilon}} - 1}{1 - \epsilon} \right]$$ (41)

subject to

$$c^{o,h} = w^a + w^o - c^a$$ (42)

$$c^{o,l} = w^a - c^a$$ (43)

The value function analytical expression is not easy to find so I provide, on the one hand, the first order condition that gives the optimum:

$$c^{a^{1-\epsilon}} = \beta \left( \zeta (w^a + w^o - c^a)^{-\epsilon} + (1 - \zeta) (w^a - c^a)^{-\epsilon} \right)$$ (44)

Once \(c^a\) is known I can pin down future consumption using the constraints of the problem. The value function is the utility function evaluated at the optimum \(V = U(c^a^{*, o^h^*, o^l^*})\).

Appendix C: Welfare and income measures

This appendix defines the measures of welfare and earnings used in section 4. First I define welfare by workers' occupation and type (i.e. high skilled in complex task). Then I turn to do the same for the earnings.
Welfare

Using some previously defined objects I define average welfare for high skilled and low skilled workers, at both, simple and complex tasks. This is done for a setting with double screening. For high skilled workers doing complex task welfare (net of investment costs) is:

\[ W^{h,c} = \int_{\bar{\theta}}^{\infty} V^{h,c}(p(\theta)w^c, w^p) f_h(\theta, \bar{\theta}, \infty) d\theta \]  

(45)

For the non-investors who deceive the firm I need to take into account the bias produced by the self-selection\(^{25}\) of workers:

\[ W^{l,c} = \int_{\theta_{\text{min}}}^{\eta_{\text{min}}} \int_{\eta_{\text{min}}}^{\eta_{h}} V^{l,c}(p(\theta)w^c, w^p, \eta) f_l(\theta, \eta, \eta^h, \epsilon) d\eta d\theta \]  

(46)

The average welfare obtained by high skilled doing simple tasks and low skilled doing simple task is given by:

\[ W^{h,s} = V^s \]  

(47)

And total welfare in the economy is defined by the addition of all the partial welfare measures minus the investment costs incurred by the high skilled.

\[ W = H^c W^{h,c} + L^c W^{l,c} + N^s W^s - \int_{\phi}^{\xi} F_\phi d\phi \]  

(48)

Earnings

I do a similar exercise to compute workers’ aggregate earnings. For high skilled workers doing complex task, earnings are given by:

\[ E^{h,c} = \int_{\bar{\theta}}^{\infty} p(\theta)w^c f_h(\theta, \bar{\theta}, \infty) d\theta + w^p \]  

(49)

For the non-investors who deceive the firm I need to take into account the bias produced by the self-selection\(^{26}\) of workers:

\[ E^{l,c} = \int_{\theta_{\text{min}}}^{\eta_{\text{min}}} \int_{\eta_{\text{min}}}^{\eta_{h}} \eta w^p f_l(\theta) f_\eta d\eta d\theta + \int_{\eta_{\text{min}}}^{\eta_{h}} \int_{\theta_{\text{min}}}^{\eta_{h}} w^c p(\theta) f_\eta d\eta d\theta \]  

(50)

The earnings obtained by high skilled and low skilled doing simple tasks is given by:

\[ E^{h,s} = w^s \]  

(51)

\(^{25}\)Refer to Appendix A for information about the bounds of this integral.

\(^{26}\)Refer to Appendix A for information about the bounds of this integral.
And the total earnings in the economy are defined by the addition of all the partial earnings measures weighted by the share of each group in the population,

$$E = H^c E^{h,c} + L^c E^{l,c} + N^s E^s$$

(52)