A Simple Theoretical Argument for Affirmative Action

Laurence Kranich

Department of Economics

University at Albany, SUNY

lkranich@albany.edu

September 3, 2012

Abstract

We consider a society jointly committed to ensuring equal opportunity and to increasing aggregate wealth but facing the vestiges of past discrimination manifested in a historically skewed distribution of social resources. Focusing on the problem of allocating the existing quantity of inputs, we contrast two policy instruments: redistributing resources in order to compensate for the effect of the asymmetry on productive abilities, or granting preferential treatment in employment to the disadvantaged group (affirmative action). We show that society is generally better off with affirmative action than without it, and that a socially optimal policy may rely solely on affirmative action.

1 Introduction

Faced with the vestiges of past discrimination, what is the optimal policy of a society which is jointly committed to the principle of equal opportunity and to increasing the aggregate wealth of its members? That is the question we wish to address in this paper.

Unlike other studies which seek to explain why discrimination occurs and ways to prevent it, here we are concerned with social disparities that are due to past discriminatory practices and result in present unfair advantages. Indeed, that is descriptive of the status quo: the prevailing social ethic, as evident in our laws and practices, is that all persons should have equal opportunities, and yet throughout our history discrimination was widespread and even legal. Under such circumstances, simply adopting unbiased statutes or requiring equal treatment does not afford agents equal opportunities. Rather, in order to "level the playing field," the historical asymmetry must be considered. Thus, for example, if members of one group were consigned to an inferior secondary educational system, then simply adopting unbiased college admissions policies would not ensure equal access. Or if one group had been excluded from owning property, then simply repealing such restrictions would not ensure equal ownership opportunities if land is already allocated. Similar initial asymmetries arise in a wide variety of settings, including the employment of women in management and science, access to credit markets, post-colonial trade relations, and climate change negotiations.

A second important distinction between this and previous work concerns the mode of analysis. Clearly, to understand why discrimination occurs and its possible effects, it is necessary to consider the behavior and/or motivation of the agents involved. Thus, whether agents discriminate because they simply prefer one group to another (taste-based) or because it is the rational response to informational asymmetries and group differences (statistical), the focus is on individual choice behavior and incentives. It is implicit in all such studies that discrimination is socially repugnant and means should be

established to prevent it. Nevertheless, the sole criterion by which remedial policies are evaluated is their effect on economic performance, i.e., profits, wage earnings or aggregate wealth. We would argue that this welfare criterion is incomplete: if discrimination is noxious, then its occurrence should adversely affect social welfare.

Since we do not wish to explain discrimination but rather to address its lingering effects, we abstract from incentive issues and individual behavior and, instead, specify the components of social welfare directly.[1] Like existing work, we assume social welfare depends on economic performance. However, we also explicitly include an index of discrimination, which enters negatively.[2] Thus, we consider a society with dual objectives: to promote equal opportunity and to maximize wealth. In such a pure social welfare analysis, the determination of an optimal policy is straightforward. While it is inevitable that some individuals will be helped and some will be hurt when seeking to redress past injustices, the optimal policy is one which maximizes social welfare.

To focus the discussion, we consider the simplest version of the problem which consists of two people (or types of individuals) who compete for the better of two positions – employment, college admissions, etc. The agents' effective abilities are determined by their innate talents and *social factors* (e.g., school quality, nutrition, health care, etc.).[3] We assume the former are drawn randomly from a common distribution while the latter are exogenously

---

[1] Atkinson (1970) notes, "Dalton (1920) argued that we should approach the question [of income inequality measurement] by considering directly the form of the social welfare function to be employed." He adds, "I hope that these conventional measures [of income inequality] will be rejected in favour of direct consideration of the properties that we should like the social welfare function to display."

[2] In light of legislative efforts to eliminate discrimination and fiscal and monetary efforts to stimulate growth, it is obvious that these factors affect social welfare. Moreover, the voluminous economic literature on these topics attests to their significance.

[3] In this paper we abstract from school effort or other voluntary undertakings which might enhance one's ability. Thus, there is no issue of sacrifice or just desserts. Otherwise, our model might apply to the net effect of social inputs after accounting for such private contributions. That is, one could interpret "innate talent" as including the earned portion.

specified and differ for the two types as a result of past discrimination. Once innate talents are drawn and social factors are applied, the positions are assigned on the basis of effective abilities, with the more skilled individual assigned to the better position.

Other things being equal, the social ideal would be that in every instance a more qualified individual is selected for the better position over a less qualified one regardless of their types. It is possible, however, that one individual is more deserving than another on the basis of innate talent, but the order is reversed as a result of differences in social factors. We refer to this as an *inversion,* and we would argue that it is such inversions that are the most objectionable consequence of discrimination.[4] Thus, we assume that social welfare decreases with the likelihood of inversions.[5] The other component of social welfare is aggregate wealth which is determined by the overall value or quality of the match. This is maximized when the individual with the highest effective ability obtains the better of the two positions.

We consider the problem of allocating the existing quantity of social factors, and we compare two policy instruments. The first is simply to reallocate the factors, transferring resources from the advantaged to the disadvantaged type. The second is to grant preferential treatment to members of the disadvantaged group (*affirmative action,* or *AA*). While there are various formulations of AA, here we take it to mean modifying the assignment rule by granting a marginal advantage to the disadvantaged type, that is, to assign the disadvantaged individual the better of the two positions even if its effective ability is marginally less.

After describing the general model, we consider the special case in which innate abilities are distributed uniformly and the effect of social factors is

---

[4] Otherwise, although discrimination might affect the levels of the agents' productivities, it does not affect the order. Hence, the agents are assigned their "correct" positions.

[5] The timing of the model is such that past policies are inherited or new policies are implemented prior to the realization of innate talents. That is, the institutional framework is in place at the start of the next generation.

additive.[6] This enables us to explicitly characterize and contrast the optima under the two policies. In this case we show that unless society is fully committed to either of the extremes (i.e., solely maximizing aggregate wealth or fully equalizing social resources), it is always better to adopt some level of affirmative action than to rely exclusively on resource transfers. Moreover, if we allow the level of AA bias to vary and select the optimal combination of transfers and bias, then at an interior solution (that is, again excluding the extreme commitment to either social objective), the socially optimal policy will rely solely on affirmative action.

1.1   Literature review

Generally, there are two theories of discrimination: *taste-based discrimination* (Becker 1957) and *statistical discrimination* (Phelps 1972 and Arrow 1973). According to the former, agents have different preferences for such ascriptive characteristics as race or gender, and this affects their decision-making. The latter results from informational asymmetries: if relevant characteristics such as productivity are unobservable but known to differ, on average, across groups, then it may be rational to discriminate in seeking to maximize expected profits.

In addition, there have been several theoretical papers on affirmative action. First, Lundberg (1991) compares affirmative action policies (or minority quotas) versus "disparate impact" judgments (restrictions on the type of information employers may use in hiring) as methods to enforce equal opportunity laws. Next, in a model with unobservable investment in skills or training, Foster and Vohra (1992) consider preferential treatment as a means to equalize investment incentives across groups. Otherwise, it is possible in their model that members of one group might fall into an investment trap:

---

[6]We adopt these assumptions for computational convenience. By doing so, we can obtain closed-form solutions to the social planner's problems specified in the sequel. We discuss the restrictiveness of the assumptions in the Conclusion.

they would fail to invest because they would not be hired, and they would not be hired because they fail to invest. In a similar model, Coate and Loury (1993) investigate whether affirmative action tends to dispel or perpetuate employers' negative stereotypes of minority workers.

Chung (2000) addresses a different aspect of affirmative action, namely, in promoting role models who might serve as a source of information about job characteristics, employment opportunities or returns to investment. Next, Austen-Smith and Wallerstein (2006) consider a political economy model in which both affirmative action (quotas) and redistributive taxation are determined endogenously. They wish to determine whether redistribution by race (i.e., AA) tends to crowd out fiscal redistribution.

Fryer and Loury (2007) study the optimal timing and form of intervention in a second-best framework. That is, they concede that AA is welfare reducing – on the basis of economic performance alone – and their objective is to minimize the efficiency loss. Specifically, they consider two policy environments: *sighted*, where group identity is contractible, and *blind*, where it is not. And they invesitgate whether AA policies should be aimed at the skill acquisition stage (e.g., training subsidies) or at the employment stage.[7]

As mentioned earlier, all of the above focus on current discrimination, they consider individual choice behavior and incentives, most take the AA policy to be exogenously specified (generally as a quota), and they evaluate social welfare solely on the basis of economic performance – a necessarily incomplete criterion even within the limited context of these models.

In contrast, ours is a pure social welfare analysis. We abstract from individual incentives[8], we consider the dual social objectives of providing

---

[7]There are several other papers which focus exclusively on the incentive effects of AA rather than on the welfare implications. These include Fu (2006), Casamiglia et. al. (2009) and Franke (2010).

[8]In fact, (microeconometric) estimates of individual labor supply elasticity tend to be quite low. (See Heckman 1993.)

Also, economists are now sufficiently knowledgeable about incentives that if there were adverse incentive effects associated with AA, it might be possible to devise an appropriate

equal opportunities (equity) and maximizing aggregate wealth (efficiency), and we endogenously determine an optimal remedial policy.

1.2   Organization

The paper is organized as follows. In the next section we present the general model and we describe the arguments affecting social welfare. Section 3 focuses on the special case in which innate abilities are distributed uniformly and the effect of social factors is additive. In Section 4 we consider the issue of (re)allocating a fixed quantity of social resources and characterize the optimal level of transfers. In Section 5 we introduce affirmative action as an alternative policy instrument. In order to exactly compute the optima and to explore possible comparative static effects, Section 6 considers an example involving Leontief social welfare functions. In Section 7 we characterize the global optimum when both the affirmative action bias and the level of resource transfers are allowed to vary. Section 8 contains a brief conclusion. Most of the proofs appear in the appendix.

2   Model

In its simplest formulation, the problem consists of two individuals (types), A and B, who must be assigned to two jobs or positions, a high paying job h and a low paying one ℓ. The returns to the latter are denoted w^h and w^ℓ, respectively, where $w^h > w^\ell$. Prior to the assignment, both agents draw randomly from a common distribution of *innate talents*, a°, with (continuous) density f. These are then influenced by *social factors* such as school quality, health care, etc. For simplicity, we consider a single social input, denoted s, and we assume that historically the level of s has differed for the two types.[9]

---

supplemental scheme to address this issue directly.

    [9]While we have in mind such (economic) factors as schooling and health, this might incorporate institutional or cultural impediments as well such as legal strictures or social

Let s_i denote the level of the social input of type i. We will assume that the process by which *effective ability* is determined from innate talent and s (one might think of this as the education technology) is given by $a_i' = \sigma(a_i^\circ, s_i)$, which is increasing in both arguments. Finally, the value or quality of a match between an individual with effective ability a' and a job with return w is given by $a'w$. For convenience, we denote $\mathbf{a}^\circ = (a_A^\circ, a_B^\circ)$ and $\mathbf{s} = (s_A, s_B)$. We refer to $\mathbf{s}$ as an *expenditure policy* or, for now, simply a *policy*.

Proceeding to the assignment of jobs, after innate talents are drawn and are combined with the social input, the agent with the higher effective ability is placed in the high paying position.[10] If $s_i = s_j$, i.e., there is no discrimination in the allocation of s, then the individual with higher innate talent will also have higher effective ability and would be matched to job h. However, if $s_i \neq s_j$, then it is possible that the agent with the lower innate talent would be matched to job h. We refer to such an instance as an *inversion*, and we would argue that it is such inversions that society finds objectionable.

Let $a_H' = \max\{\sigma(a_A^\circ, s_A), \sigma(a_B^\circ, s_B)\}$ and $a_L' = \min\{\sigma(a_A^\circ, s_A), \sigma(a_B^\circ, s_B)\}$. Then the aggregate value of the match, i.e., the aggregate social product or wealth, is given by $V(\mathbf{a}^\circ, \mathbf{s}) = a_H' w^h + a_L' w^\ell$.

Suppose that historically $s_A < s_B$, that is, type A agents have received less of the social input than type B agents. Then while a_A° and a_B° are drawn from the same distribution, the effective abilities a_A' and a_B' are distributed differently. Let f_A and f_B denote the (continuous) densities of a_A' and a_B', respectively, and F_A and F_B the corresponding distribution functions. If there were no discrimination, then *ex ante*, types A and B would be equally likely to be placed in the high paying position, that is, the probabilities are such that $P(a_A' \leq a_B') = P(a_B' \leq a_A')$. Generally, however, the likelihood

---

attitudes. In any case, it is the historical difference in social resources that we refer to as the "vestiges of past discrimination."

[10]It is assumed that abilities, both innate and effective, are observable or easily measured, say, by examination.

that B will obtain job h is given by:

$$P(a'_A \leq a'_B) = \int_{-\infty}^{\infty} F_A(a_B) f_B(a_B) da_B = \int_{-\infty}^{\infty} \int_{-\infty}^{a_B} f_A(a_A) f_B(a_B) da_A da_B.$$

Consequently, the expected value of V under policy $\mathbf{s}$ is given by

$$
\begin{aligned}
EV(\mathbf{s}) &= \left(\int_{-\infty}^{\infty} \int_{-\infty}^{a'_B} \left(a'_B w^h + a'_{At} w^\ell \right) df_A df_B \right) P(a'_A \leq a'_B) + \\
&\quad \left(\int_{-\infty}^{\infty} \int_{-\infty}^{a'_A} \left(a'_A w^h + a'_B w^\ell \right) df_B df_A \right) P(a'_B \leq a'_A).
\end{aligned}
$$

Next, under policy $\mathbf{s}$, the likelihood of inversion can be computed as follows. For given a°_A, $a'_A = \sigma(a^\circ_A, s_A)$ is the effective ability of agent A who receives the level of support s_A; and A will obtain position h over B if $a'_A > a'_B$. However, if A had received the level of support s_B instead of s_A, its ability would be $\sigma(a^\circ_A, s_B)$. Thus, it is those type B agents for whom $a'_B \in (a'_A, \sigma(a^\circ_A, s_B))$, i.e., those whose abilities lie between A's actual ability and what it would have been had it received s_B, who would obtain position h over A under the current scheme (s_A, s_B) but would not if agents received equal funding (at the level s_B). This is depicted in Figure 1.

(Insert Figure 1 here.)

The likelihood of an inversion involving A with innate talent a°_A and effective ability a'_A is[11]

$$I(a^\circ_A, \mathbf{s}) = \int_{a'_A}^{\sigma(a^\circ_A, s_B)} f_B(a_B) da_B. \tag{1}$$

The cumulative likelihood is thus

$$I(\mathbf{s}) = \int_{-\infty}^{\infty} I(a^\circ_A, s) f(a^\circ_A) da^\circ_A. \tag{2}$$

---

[11]Since $s_A \leq s_B$, A could not obtain job h over B undeservedly. That is, inversions always entail a less deserving type B agent obtaining h over a more deserving type A agent.

Next, we assume social welfare, W, depends on two factors: the expected aggregate value of the match (i.e., the "size of the pie" or GDP) and the propensity for inversion. Thus, we take it to be of the form $W(\mathbf{s}) = \Psi(EV(\mathbf{s}), I(\mathbf{s}))$, where Ψ is increasing in its first argument and decreasing in the second.

Before developing a specific case of the model, we note two general influences on aggregate value and thus on social welfare: (1) productivity mismatches between effective abilities and jobs lower aggregate value,[12] and (2) smoothing or averaging effective abilities reduces aggregate product as well. These may have important implications for antidiscrimination policies. Recall, $a'_H = \max\{a'_A, a'_B\}$ and $a'_L = \min\{a'_A, a'_B\}$. Also, let $\Sigma = s_B - s_A$.

Lemma 1 $a'_H w^h + a'_L w^\ell \geq a'_L w^h + a'_H w^\ell$.

Proof. $\left(a'_H w^h + a'_L w^\ell\right) - \left(a'_L w^h + a'_H w^\ell\right) = \left(a'_H - a'_L\right) w^h - \left(a'_H - a'_L\right) w^\ell = \left(a'_H - a'_L\right)\left(w^h - w^\ell\right) \geq 0.\blacksquare$

Hence, the gain from employing the high productivity agent in job h exceeds the loss from their not working in ℓ.

Lemma 2 $a'_H w^h + a'_L w^\ell > (a'_H - \varepsilon)w^h + (a'_L + \varepsilon)w^\ell$, *for small* $\varepsilon > 0$.[13]

Proof. $\left(a'_H w^h + a'_L w^\ell\right) - \left((a'_H - \varepsilon)w^h + (a'_L + \varepsilon)w^\ell\right) = \varepsilon\left(w^h - w^\ell\right) > 0.\blacksquare$

As shown in Lemma 2, smoothing will decrease the aggregate product. However, it will also decrease the likelihood of inversion.

Lemma 3 *The likelihood of inversion, I, is nondecreasing in Σ.*

---

[12] To clarify, we use the term *inversion* to refer to the case in which an individual with lower innate ability has higher effective ability as a result of differences in s. In contrast, a *mismatch* occurs when agents are assigned to the "wrong" jobs, that is, when the individual with lower *effective* ability is assigned to position h, either by mistake or as a result of affirmative action.

[13] That is, for those values of ε which do not reverse the order of the agents' productivities.

Proof. Since σ is increasing in both arguments, it is clear that this is true for each value of a_A° (see (1) and Figure 1). Hence it is true for the cumulative likelihood as well.[14]∎

3 Uniform Innate Talents with Additive Abilities

To demonstrate the trade-off between the two social objectives and to explore the possible role for affirmative action, in the remainder of the paper we focus on a specific case in which innate abilities are distributed uniformly on $[\underline{a}^\circ, \overline{a}^\circ]$ and $\sigma(a^\circ, s_i) = a^\circ + s_i$. In this case, effective abilities, a_A' and a_B', are distributed uniformly on the intervals $[\underline{a}_A', \overline{a}_A'] := [\underline{a}^\circ + s_A, \overline{a}^\circ + s_A]$ and $[\underline{a}_B', \overline{a}_B'] := [\underline{a}^\circ + s_B, \overline{a}^\circ + s_B]$, respectively, where $s_A < s_B$. It is straightforward to evaluate the components of social welfare in this case. First, for notational convenience we define $\Delta = \overline{a}^\circ - \underline{a}^\circ$, $\widehat{a}^\circ = \frac{\overline{a}^\circ + \underline{a}^\circ}{2}$, and $\widehat{a}_i' = \frac{\overline{a}_i' + \underline{a}_i'}{2}$. Then

$$
\begin{aligned}
P(a_A' \;\leq\; a_B') &= \frac{\widehat{a}_B' - \underline{a}_A'}{\Delta} = F_A(\widehat{a}_B') \\
&\text{and} \\
P(a_B' \;\leq\; a_A') &= F_B(\widehat{a}_A').
\end{aligned}
\tag{3}
$$

Hence,

$$
EV(\mathbf{s}) = (\widehat{a}_B' w^h + \widehat{a}_A' w^\ell) F_A(\widehat{a}_B') + (\widehat{a}_A' w^h + \widehat{a}_B' w^\ell) F_B(\widehat{a}_A').
$$

Since here $\Delta = \overline{a}_A' - \underline{a}_A' = \overline{a}_B' - \underline{a}_B'$, if $F_A(\widehat{a}_B') = 1$, then $F_B(\widehat{a}_A') = 0$. Therefore,

---

[14]If there is a finite support of a° (as in the next section), then there may be a level of Σ beyond which B will always be assigned position h and I would be maximal.

$$EV(\mathbf{s}) = \begin{cases} (\widehat{a}'_B w^h + \widehat{a}'_A w^\ell) & if \quad \widehat{a}'_A < \underline{a}'_B \\ (\widehat{a}'_B w^h + \widehat{a}'_A w^\ell)F_A(\widehat{a}'_B) + (\widehat{a}'_A w^h + \widehat{a}'_B w^\ell)F_B(\widehat{a}'_A) & otherwise. \end{cases} \tag{4}$$

Further, for $\widehat{a}'_A \geq \underline{a}'_B$, $F_A(\widehat{a}'_B) + F_B(\widehat{a}'_A) = \frac{\widehat{a}'_B - \underline{a}'_A}{\Delta} + \frac{\widehat{a}'_A - \underline{a}'_B}{\Delta} = \frac{\widehat{a}'_A - \underline{a}'_A}{\Delta} + \frac{\widehat{a}'_B - \underline{a}'_B}{\Delta} = \frac{1}{2} + \frac{1}{2} = 1$. Hence, EV is a weighted average of $(\widehat{a}'_B w^h + \widehat{a}'_A w^\ell)$ and $(\widehat{a}'_A w^h + \widehat{a}'_B w^\ell)$. Since $\widehat{a}'_A \leq \widehat{a}'_B$, the maximum weight on $(\widehat{a}'_A w^h + \widehat{a}'_B w^\ell)$ is $\frac{1}{2}$.

Let $a'_m = \max\{a'_A, \underline{a}'_B\}$. Then

$$\begin{aligned} I(a^\circ_A, \mathbf{s}) &= \int_{a'_m}^{\sigma(a^\circ_A, s_B)} f_B(a_B) da_B \\ &= \frac{\sigma(a^\circ_A, s_B) - a'_m}{\Delta}. \end{aligned}$$

Hence,

$$\begin{aligned} I(\mathbf{s}) &= \int_{\underline{a}^\circ}^{\overline{a}^\circ} \frac{\sigma(a^\circ, s_B) - \max\{\sigma(a^\circ, s_A), \underline{a}'_B\}}{\Delta^2} da^\circ \tag{5} \\ &= \int_{\underline{a}^\circ}^{\overline{a}^\circ} \frac{\min\{a^\circ - \underline{a}^\circ, \Sigma\}}{\Delta^2} da^\circ. \end{aligned}$$

If $\Sigma = 0$, i.e., there is no discrimination, then $I = 0$. Also, for $\Sigma \geq \Delta$, B would always be assigned to h and $I = \frac{1}{2}$. Generally, for $\Sigma \in (0, \Delta)$, I increases monotonically in Σ.

In the absence of discrimination, agent B *should* be selected over A in 50% of the cases, i.e., on the basis of their randomly drawn innate talents. Hence, we take $Q(\mathbf{s}) = \frac{1}{2} - I(\mathbf{s})$ as an index of *equal opportunity*, and we rewrite W as an increasing function:

$$W(\mathbf{s}) = \Phi(EV(\mathbf{s}), Q(\mathbf{s})). \tag{6}$$

We assume Φ is continuous and quasiconcave.

4   Fixed Social Resources

As described above, the levels of the social input s_A and s_B are taken to be exogenous, with $s_A < s_B$, reflecting the historical asymmetry in the treatment of the two types. Now, suppose we consider reallocating units of s from B to A while holding the aggregate quantity fixed. Thus, we abstract from the question of the socially optimal level of s and focus purely on its distribution. We can then reformulate the problem as that of determining an optimal deviation from equal division. Let $\overline{s} = s_A + s_B$ and $\widehat{s} = \frac{\overline{s}}{2}$. Then we describe a distribution of $\overline{s}$ as a pair $(\widehat{s} - \varepsilon, \widehat{s} + \varepsilon)$, where $0 < \varepsilon < \widehat{s}$. For convenience, we abbreviate $(\widehat{s} - \varepsilon, \widehat{s} + \varepsilon)$ by ε. Also, let ε° denote the initial deviation, i.e., $(\widehat{s} - \varepsilon^{\circ}, \widehat{s} + \varepsilon^{\circ}) = (s_A, s_B)$.

By direct calculation from (4) and (5), respectively, we obtain

$$EV(\varepsilon) = \widehat{a}'(w^h + w^{\ell}) + \frac{4\varepsilon^2(w^h - w^{\ell})}{\Delta} \tag{7}$$

and

$$I(\varepsilon) = \frac{2\varepsilon}{\Delta} - \frac{2\varepsilon^2}{\Delta^2}, \tag{8}$$

where $\widehat{a}' := \widehat{a}^{\circ} + \widehat{s}$.

Also, we rewrite (6) as

$$W(\varepsilon) = \Phi(EV(\varepsilon), Q(\varepsilon)). \tag{9}$$

Next, we show that, in general, there is indeed a trade-off between the two social objectives. First, note that in terms of the parameter ε, there is no discrimination when $\varepsilon = 0$ and there is maximal discrimination when $\varepsilon = \frac{\Delta}{2}$. In the case of the latter, $\Sigma = \Delta$ and the effective abilities are such that $\underline{a}'_B > \overline{a}'_A$. Hence, B is always matched to job h. We can then describe the locus of achievable combinations $(EV(\varepsilon), Q(\varepsilon))$ that can be obtained by varying the policy parameter ε from 0 to $\frac{\Delta}{2}$. We refer to this as the *achievable*

set and denote it by $\Lambda(\varepsilon)$. This is depicted in Figure 2.

(Insert Figure 2 here.)

We now establish properties of the achievable set. First, from (7) and (8), respectively, we have

$$EV'(\varepsilon) = \frac{8\varepsilon(w^h - w^\ell)}{\Delta}$$

and

$$I'(\varepsilon) = \frac{2}{\Delta} - \frac{4\varepsilon}{\Delta^2}.$$

Therefore, in the space of (EV, Q), the slope along the achievable set is $-\frac{\frac{\Delta}{2} - \varepsilon}{2\varepsilon\Delta(w^h - w^\ell)}$, which asymptotically approaches $-\infty$ as ε approaches 0 and is 0 at $\varepsilon = \frac{\Delta}{2}$. In addition, it is increasing in ε. Thus the locus is convex.

The optimal trade-off between aggregate wealth and equal opportunity would then be determined by maximizing $W(\varepsilon)$ over the achievable set or, formally, by solving the social planner's problem (PP_1):

$$\max_{\varepsilon \in [0, \frac{\Delta}{2}]} W(\varepsilon) \tag{10}$$

Let ε_1 denote the solution to (10).

Figures 3, 4 and 5 depict three possibilities.[15] In the first, society would choose to fully equalize resources. The second depicts the opposite extreme in which society would choose solely to maximize wealth. And in the third the optimum entails intermediate levels of equality and wealth.[16]

---

[15] The figures include representative social welfare contours.

[16] Notice that if the optimal $\varepsilon > \varepsilon^\circ$, then the economic benefits of *increasing* discrimination (e.g., by creating elite schools or subsidizing private education) would outweigh the social cost of greater inequality.

(Insert Figures 3, 4 and 5 here.)

5   Affirmative Action

In the previous section, we considered the extent to which social inputs or expenditures should be equalized across types when one type has historically received less than the other. We now turn to an alternative policy instrument, namely, *affirmative action* (AA). Here, we take this to mean that in seeking to fill position h, agent A is afforded a marginal advantage. That is, rather than filling h purely on the basis of the comparison between measured productivities a'_A an a'_B, A is awarded the position even if it has somewhat lower productivity. Formally, A is awarded position h if $a'_A + \delta \geq a'_B$ (rather than if $a'_A \geq a'_B$), for some $\delta > 0$. In addition, we require $\delta \leq 2\varepsilon$ to avoid bias in the opposite direction, that is, where A would have an unfair advantage. Increasing δ has the effect of creating more mismatches but decreasing the likelihood of inversions. We wish to determine the effect of AA on social welfare and to contrast the alternative policy instruments of transferring social resources versus AA as means of addressing past injustices.

Under AA, the expressions for EV and I corresponding to (7) and (8) are[17]

$$EV(\varepsilon, \delta) \;=\; \widehat{a}'(w^h + w^\ell) + \frac{(4\varepsilon^2 - 2\delta\varepsilon)(w^h - w^\ell)}{\Delta} \tag{11}$$

$$=\; EV(\varepsilon) - \frac{2\delta\varepsilon(w^h - w^l)}{\Delta} \tag{12}$$

and

$$I(\varepsilon, \delta) = \frac{2\varepsilon - \delta}{\Delta} - \frac{(2\varepsilon - \delta)^2}{2\Delta^2}. \tag{13}$$

---

[17] Again this is a slight abuse of notation. However, the continued use of the notation EV and I should not cause any confusion.

We now establish the effect of this on the achievable set.

Let $\delta > 0$ be given. First, note that for $\varepsilon = \frac{\delta}{2}$, $EV(\varepsilon, \delta) = EV(0)$ and $Q(\varepsilon, \delta) = Q(0)$.[18] Otherwise, for $\varepsilon > \frac{\delta}{2}$, it is clear from (7) and (11) that $EV(\varepsilon, \delta) < EV(\varepsilon)$.

Lemma 4 *For $\delta > 0$, suppose $EV(\varepsilon'', \delta) = EV(\varepsilon')$. Then $Q(\varepsilon'', \delta) > Q(\varepsilon')$.*

Hence, the new achievable set, denoted $\Lambda(\varepsilon, \delta)$, lies above $\Lambda(\varepsilon)$. Also, we have the following:

Lemma 5 *For given $\delta > 0$, $\Lambda(\varepsilon, \delta)$ is convex.*

Summarizing, under the constraint $\delta \leq 2\varepsilon$, the relationship between $\Lambda(\varepsilon, \delta)$ and $\Lambda(\varepsilon)$ is as depicted in Figure 6.

(Insert Figure 6 here.)

As above, in this case we rewrite (6) as

$$W(\varepsilon, \delta) = \Phi(EV(\varepsilon, \delta), Q(\varepsilon, \delta)). \tag{14}$$

Then for given $\delta > 0$, the optimal value of ε would be determined by solving the following variant of (10), which we refer to as PP_2:

$$\max_{\varepsilon \in [0, \frac{\Delta}{2}]} W(\varepsilon, \delta). \tag{15}$$

Analogous to the above, let $\varepsilon_2(\delta)$ denote the solution to (15), for given δ.

Theorem 6 *Social welfare is strictly greater with affirmative action than without it if $\varepsilon_1 \in (0, \frac{\sqrt{\Delta(\Delta - \delta)}}{2})$ or $\varepsilon_2(\delta) \in (\frac{\delta}{2}, \frac{\Delta}{2})$.*

---

[18]If $\varepsilon = \frac{\delta}{2}$, then the affirmative action bias exactly offsets the skewness in s; that is, $a_A' + \delta \geq a_B'$ if and only if $a_A^\circ \geq a_B^\circ$.

Note that in the above result, as δ tends to zero, the range of possible values of both ε_1 and ε_2 for this class of social welfare functions tend to span all feasible values $(0, \frac{\Delta}{2})$. This leads to the following:

Theorem 7 *Unless a society chooses exclusively to maximize aggregate wealth or to fully equalize social inputs, there is a δ sufficiently small such that social welfare is strictly greater with affirmative action than without it.*

The converse of Theorem 7 is suggested by Theorem 6 as well: as $\delta \to \Delta$, the range of ε_2 consistent with a strict improvement in social welfare diminishes. That is, as the magnitude of the AA bias increases, the smaller the domain of social welfare functions for which AA would be welfare improving.

6   Leontief Social Welfare Functions

To demonstrate the above in greater detail, we consider the class of Leontief social welfare functions

$$W(EV, Q) = \min\{\alpha EV, (1 - \alpha)Q\}. \tag{16}$$

(Here, α can be interpreted as the welfare weight assigned to Q in that the greater the value of α, the greater the optimal ratio of Q to EV.)

In this case it is straightforward to determine the optimal level of ε in the absence of AA, ε_1, by computing the intersection of the expansion path with $\Lambda(\varepsilon)$. This is given implicitly by

$$\frac{1}{2} - \frac{2\varepsilon_1}{\Delta} + \frac{2\varepsilon_1^2}{\Delta^2} = \frac{\alpha}{1 - \alpha}\left[\widehat{a}'(w^h + w^\ell) + \frac{4\varepsilon_1^2(w^h - w^\ell)}{\Delta}\right]. \tag{17}$$

The range of optimal ε which might be supported by some member of this class of social welfare functions is delimited by the extreme values associated with $\varepsilon_1 = 0$ and $\varepsilon_1 = \frac{\Delta}{2}$. The former would be optimal for any $\alpha \geq \overline{\alpha} :=$

$\frac{1}{1+2\widehat{a}(w^h+w^l)}$, while the latter would be optimal if $\alpha = 0$. In between, for each $\alpha \in (0, \overline{\alpha})$ there is a unique ε_1 which solves (17). Moreover, α varies inversely with ε_1. Thus, as greater weight is placed on equality, the smaller the optimal deviation from equal division.

Similarly, for given δ, the optimal ε under AA, $\varepsilon_2(\delta)$, is given implicitly by

$$\frac{1}{2} - \frac{2\varepsilon_2 - \delta}{\Delta} + \frac{(2\varepsilon_2 - \delta)^2}{2\Delta^2} = \frac{\alpha}{1-\alpha}\left[\widehat{a}'(w^h + w^\ell) + \frac{(4\varepsilon_2^2 - 2\varepsilon_2\delta)(w^h - w^\ell)}{\Delta}\right]. \tag{18}$$

Again, for any $\alpha \geq \overline{\alpha}$, the optimum would be the minimal deviation from equal division, in this case $\varepsilon_2 = \frac{\delta}{2}$. At the other extreme, for any $\alpha \in [0, \underline{\alpha})$, the optimum would be the maximal deviation $\varepsilon_2 = \frac{\Delta}{2}$, where $\underline{\alpha} := \frac{\delta^2}{\delta^2 + 2\Delta^2(\widehat{a}'(w^h+w^\ell)+(\Delta-\delta)(w^h-w^\ell))}$. For $\alpha \in (\underline{\alpha}, \overline{\alpha})$, there is a unique ε_2 which solves (18).

Using (17) and (18), respectively, we can determine the effects of the parameters on the optimal policy.

Proposition 8 *For the class of social welfare functions given by (16), the comparative static effects of the parameters on ε_1 and $\varepsilon_2(\delta)$ are the following:*

$$\begin{array}{ccc}
\frac{d\varepsilon_1}{d\Delta} & > & 0 \\[4pt]
\frac{d\varepsilon_1}{d\alpha} & < & 0 \\[4pt]
\frac{d\varepsilon_1}{d\widehat{a}} & < & 0 \\[4pt]
\frac{d\varepsilon_1}{d(w^h+w^\ell)} & < & 0 \\[4pt]
\frac{d\varepsilon_1}{d(w^h-w^\ell)} & < & 0
\end{array} \tag{19}$$

$$
\begin{array}{rcl}
\frac{d\varepsilon_2}{d\Delta} & > & 0 \\[2mm]
\frac{d\varepsilon_2}{d\alpha} & < & 0 \\[2mm]
\frac{d\varepsilon_2}{d\widehat{a}} & < & 0 \\[2mm]
\frac{d\varepsilon_2}{d(w^h+w^\ell)} & < & 0 \\[2mm]
\frac{d\varepsilon_2}{d(w^h-w^\ell)} & < & 0 \\[2mm]
\frac{d\varepsilon_2}{d\delta} & > & 0
\end{array}
. \qquad (20)
$$

The last comparative static result, namely, $\frac{d\varepsilon_2}{d\delta} > 0$, suggests that the two corrective measures, transfers and affirmative action, are substitutes; that is, the greater the AA bias, the greater the tolerance for inequality at an optimum. In the next section, we consider the optimal combination of the two measures when δ is endogenous.

7   Optimal Level of AA

We begin by comparing the loci $\Lambda(\varepsilon, \delta)$ and $\Lambda(\varepsilon, \delta')$, where $\delta' > \delta > 0$. First, we have already seen in Lemma 5 and Figure 6 that both $\Lambda(\varepsilon, \delta)$ and $\Lambda(\varepsilon, \delta')$ lie above $\Lambda(\varepsilon)$, are convex, and contain the point $(\widehat{a}'(w^h + w^\ell), \frac{1}{2})$, corresponding to $\varepsilon = \frac{\delta}{2}$ and $\varepsilon = \frac{\delta'}{2}$, respectively. It is straightforward to see that at the right extrema

$$
EV(\frac{\Delta}{2}, \delta') = \widehat{a}'(w^h + w^\ell) + (\Delta - \delta')(w^h - w^\ell) < \widehat{a}'(w^h + w^\ell) + (\Delta - \delta)(w^h - w^\ell) = EV(\frac{\Delta}{2}, \delta)
$$

$$
\text{and}
$$

$$
Q(\frac{\Delta}{2}, \delta') = \frac{\delta'^2}{2\Delta^2} > \frac{\delta^2}{2\Delta^2} = Q(\frac{\Delta}{2}, \delta).
$$

Finally, we can show that at comparable values of ε, the gap between $\Lambda(\varepsilon, \delta)$ and $\Lambda(\varepsilon)$ is increasing in δ. Hence, the relationship between $\Lambda(\varepsilon, \delta)$ and $\Lambda(\varepsilon, \delta')$ is as depicted in Figure 7.

(Insert Figure 7 here.)

18

Letting δ span all values from 0 to Δ, we obtain an envelope, $\boldsymbol{\Lambda}$, of all achievable combinations (EV, Q) under some feasible policy (ε, δ).

(Insert Figure 8 here.)

The global maximum as well as the optimal policy is determined by the social planner's problem PP_3:

$$\max_{\delta\in[0,\Delta],\varepsilon\in[\frac{\delta}{2},\frac{\Delta}{2}]} W(\varepsilon,\delta). \tag{21}$$

Such an optimum is depicted in Figure 9.

(Insert Figure 9 here.)

Turning to the properties of an optimal policy, we have the following:

Theorem 9 *At an interior solution to (21), the optimal policy entails $\varepsilon = \frac{\Delta}{2}$. That is, in allocating a fixed budget $\overline{s}$ between A and B, it is generally optimal to allow maximal discrimination and to rely solely on affirmative action to prevent inversions.*

Some comments are in order. First, the theorem pertains to interior solutions only. It is still possible that a society would choose to focus exclusively on one of the two social objectives, either fully equalizing resources or maximizing aggregate wealth, but these occur only on the boundary. Next, if we were to require that the distribution of social resources be no more skewed than the historical distribution, then rather than the envelope consisting of the values of $\Lambda(\varepsilon, \delta)$ for $\varepsilon = \frac{\Delta}{2}$ as δ varies from 0 to Δ, it would be comprised of the values for $\varepsilon = \varepsilon^\circ$. Finally, these results pertain to the allocation a fixed quantity of social resources. Theorem 9 does not preclude closing the historical gap in social inputs using additional resources.

To see why the theorem holds, consider the two ways to address the historical inequity and thus reduce the likelihood of inversions: transferring

19

social resources from B to A or else adopting an AA policy and granting A a marginal advantage in assigning position h. The former would correspond to moving along the locus $\Lambda(\varepsilon)$, whereas the latter would correspond to moving from $\Lambda(\varepsilon)$ to $\Lambda(\varepsilon, \delta)$, for some $\delta > 0$. If we consider a marginal change that would result in the same reduction in inversions (and hence increase in Q), which of the two would have less of an effect on aggregate income, EV? In the case of transfers, every B type would be adversely affected and every A type would be helped. But as we saw in Lemma 2, smoothing lowers aggregate income. On the other hand, affirmative action would adversely impact aggregate income only at the potentially small margin involving the most qualified disadvantaged individuals and the least qualified advantaged ones. Hence, the impact of AA on aggregate income is likely to be less.

8   Conclusion

In this paper, we have considered the question of how to balance the joint and competing social goals of equal opportunity and growth when society is faced with a history of discrimination. In this case, economic rewards partially accrue to innate skill or talent but are partially due to the unequal distribution of social resources. Accordingly, it is possible that a less innately qualified individual would be selected for a better position over someone who is more innately qualified. While this assignment may be "correct" in terms of effective skills, and thus maximize aggregate income, such inversions are contrary to the ideal of equal opportunity.[19] To balance these dual objectives and to address the historical asymmetry, we have considered two possible policies: transferring social resources and granting preferential treatment in the assignment of positions. In the example presented here, these are substitute measures. But whereas transfers will necessarily reduce (expected)

---

[19]Recall that we abstract from school effort and focus purely on the disparate effect of social inputs.

aggregate income, affirmative action would do so only at the potentially small margin of overlap between the most qualified disadvantaged individuals and the least qualified advantaged ones. Hence, while both measures will reduce the likelihood of inversions, the impact of AA on aggregate income is likely to be less.

In order to render the model tractable, we have made several simplifying assumptions and thus were able to obtain closed-form solutions for the social optima.[20] Nevertheless, the main results are quite intuitive and might be expected to generalize. First, in allocating a fixed quantity of social resources, there is a negative relationship between growth and equal opportunity. Next, providing society places some minimal weight on each of the two objectives (i.e., the optimum is interior), there is some level of bias such that social welfare is greater with affirmative action than without it. And finally, as measured by their impact on aggregate income, AA is less costly than resource transfers as a means to address the vestiges of past discrimination.

The central issue of this paper is quite pervasive. While we have focused on the problem of matching skill level and employment, there are numerous other contexts in which historical asymmetries have resulted in present unfair advantages. As mentioned in the Introduction, these might include access to credit markets, post-colonial trade relations, and climate change negotiations. The results of our study suggest that affirmative action-like measures might be effective in these areas as well.[21]

---

[20]Primarily, these are: (1) labor supply is inelastic, (2) innate talent is uniformly distributed, and (3) the education technology is linear. As mentioned in footnote 8, the assumption that individual labor supply elasticity is low is consistent with the empirical literature. Assumptions (2) and (3) are for analytical convenience. Of these, (2) does not seem to be significant, (Although, the assumption of a finite support allows for a maximal likelihood of inversion. But this is not necessary.) We conjecture that the results would hold for any convex education technology. In particular, this would ensure the convexity of the achievable set. If we were to relax (2) or (3), it may be possible to simulate the model and solve it numerically.

[21]Affirmative action schemes have been proposed and/or adopted for mortgage and commercial lending. Generally, the principle would call for granting a marginal (procedural) advantage to members of a historically disadvantaged group. In the area of climate change,

References

[1] Arrow, Kenneth J. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees. Princeton: Princeton University Press.

[2] Atkinson, Anthony B. 1970. "On the Measurement of Inequality." *Journal of Economic Theory* 2: 244-263.

[3] Austen-Smith, David, and Michael Wallerstein. 2006. "Redistribution and Affirmative Action." *Journal of Public Economics* 90: 1789–1823.

[4] Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

[5] Calsamiglia, Caterina, Jörg Franke, and Pedro Rey-Biel. 2009. "The Incentive Effects of Affirmative Action in a Real-Effort Tournament." Barcelona Economics Working Paper No. 404.

[6] Chung, Kim-Sau. 2000. "Role Models and Arguments for Affirmative Action." *American Economic Review* 99: 640–648.

[7] Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83: 1220–1240.

[8] Dalton, Hugh. 1920. "The Measurement of the Inequality of Incomes." *Economic Journal* 30: 348-361.

[9] Foster, Dean P., and Rakesh V. Vohra. 1992. "An Economic Argument for Affirmative Action. *Rationality and Society* 4(2): 176–188.

---

for example, this could entail affording a marginal advantage to less developed countries under an emissions-trading system or when auctioning pollution rights.

[10] Franke, Jörg. 2010. "Does Affirmative Action Reduce Effort Incentives? – A Contest Game Analysis." *Ruhr Economic Papers* No. 185.

[11] Fryer, Jr., Roland G., and Glenn C. Loury. 2010. "Valuing Identity." NBER Working Paper No. 16568.

[12] Fu, Qiang. 2006. "A Theory of Affirmative Action in College Admissions." *Economic Inquiry* 44(3): 420–428.

[13] Heckman, James J. 1993. "What Has Been Learned About Labor Supply in the Past Twenty Years?" *American Economic Review* 83(2): 116-21.

[14] Lundberg, Shelly J. 1991. "The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives." *Quarterly Journal of Economics* 106: 309-26.

[15] Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62: 659-61.

9   Appendix

This appendix contains proofs of the main results.

Proof of Lemma 4.

Equating $EV(\varepsilon'', \delta) = EV(\varepsilon')$ from (7) and (11), we obtain that the corresponding values of ε are given by

$$\varepsilon'' = \frac{\delta + \sqrt{\delta^2 + 16\varepsilon'^2}}{4}. \tag{22}$$

We then want to show $\frac{1}{2} - I(\varepsilon'', \delta) > \frac{1}{2} - I(\varepsilon')$, or substituting from (13) and (8) and simplifying, that

$$4\varepsilon'\Delta - 4\varepsilon'^2 > 4\varepsilon''\Delta - 2\delta\Delta - (2\varepsilon'' - \delta)^2. \tag{23}$$

23

Substituting from (22) and simplifying considerably, we obtain that (23) is equivalent to

$$2\Delta^2 + \delta\Delta > 4(\Delta + \delta)\varepsilon'. \tag{24}$$

Note that the LHS of (8) is constant and the RHS is linear in ε'. Moreover, LHS = RHS at $\varepsilon' = \frac{\Delta(2\Delta+\delta)}{4\Delta+\delta}$, LHS > RHS for $\varepsilon' < \frac{\Delta(2\Delta+\delta)}{4\Delta+\delta}$, and LHS < RHS for $\varepsilon' > \frac{\Delta(2\Delta+\delta)}{4\Delta+\delta}$. But generally, we require that $\varepsilon \leq \frac{\Delta}{2}$. Hence, in the relevant range, $\varepsilon' \leq \frac{\Delta}{2} < \frac{\Delta(2\Delta+\delta)}{4\Delta+\delta}$. ∎

Proof of Lemma 5.

For given δ, the achievable set $\Lambda(\varepsilon, \delta)$ consists of the locus $(EV(\varepsilon, \delta), Q(\varepsilon, \delta))$ as ε varies. To determine the slope in $EV \times Q - space$, we evaluate $\frac{\frac{\partial Q}{\partial \varepsilon}}{\frac{\partial EV}{\partial \varepsilon}}$. From (11) and (13), this is given by

$$\frac{-2\Delta + 4\varepsilon - 2\delta}{\Delta(8\varepsilon - 2\delta)(w^h - w^\ell)}. \tag{25}$$

Note that under the restrictions $\varepsilon < \frac{\Delta}{2}$ and $\delta < 2\varepsilon$, this is negative. Differentiating this with respect to ε, we obtain

$$\frac{4\Delta(8\varepsilon - 2\delta)(w^h - w^\ell) - 8\Delta(w^h - w^\ell)(-2\Delta + 4\varepsilon - 2\delta)}{\Delta^2(8\varepsilon - 2\delta)^2(w^h - w^\ell)^2}.$$

Simplifying, we see this is positive if and only if $4\varepsilon - \delta > 4\varepsilon - 2\delta - 2\Delta$, which is clearly the case. Hence, the slope increases along the curve in the direction of ε. ∎

Proof of Theorem 6.

Given that $\Phi(EV, Q)$ is increasing in both components and $\Lambda(\varepsilon, \delta)$ dominates $\Lambda(\varepsilon)$, the result follows by directly computing the corresponding value of ε_1 when $\varepsilon_2 = \frac{\Delta}{2}$ using (22). ∎

Proof of Theorem 7.

The result follows immediately from the construction of the envelope Λ in Section 7. ∎

Proof of Proposition 8.

Each of the comparative statics results is obtained by totally differentiating (17) or (18). For example, from (17) we obtain

$$\frac{d\varepsilon_1}{d\Delta} = \frac{-2\varepsilon_1\Delta + 4\varepsilon_1^2 - \frac{\alpha}{1-\alpha}4\varepsilon_1^2\Delta(w^h - w^\ell)}{-2\Delta^2 + 4\varepsilon_1\Delta - \frac{\alpha}{1-\alpha}8\varepsilon_1\Delta^2(w^h - w^\ell)} \tag{26}$$

$$= \frac{2\varepsilon_1^2 - \varepsilon_1\Delta - \frac{\alpha}{1-\alpha}2\varepsilon_1^2\Delta(w^h - w^\ell)}{2\varepsilon_1\Delta - \Delta^2 - \frac{\alpha}{1-\alpha}4\varepsilon_1\Delta^2(w^h - w^\ell)} \tag{27}$$

$$= \frac{\varepsilon_1[(2\varepsilon_1 - \Delta) - \frac{\alpha}{1-\alpha}2\varepsilon_1\Delta(w^h - w^\ell)]}{\Delta[(2\varepsilon_1 - \Delta) - \frac{\alpha}{1-\alpha}4\varepsilon_1\Delta(w^h - w^\ell)]}. \tag{28}$$

Since $\varepsilon_1 < \frac{\Delta}{2}$, $(2\varepsilon_1 - \Delta) < 0$. Also, $\varepsilon_1\Delta(w^h - w^\ell) > 0$. Hence, (28) involves subtracting positive terms in both the numerator and denominator. Therefore, $\frac{d\varepsilon_1}{d\Delta} > 0$.■

Proof of Theorem 9.

Consider the extreme points $(EV(\frac{\Delta}{2}, \delta), Q(\frac{\Delta}{2}, \delta))$, for each $\delta \in (0, \Delta)$. Clearly, each such point is feasible by definition. Moreover, it is undominated. This follows from equations (12) and (13) which show that as δ increases, $EV(\frac{\Delta}{2}, \delta)$ decreases and $Q(\frac{\Delta}{2}, \delta)$ increases monotonically. Finally, note that these points span the frontier since for each value of $EV \in (\widehat{a}'(w^h + w^\ell), \widehat{a}'(w^h + w^\ell) + \Delta(w^h - w^\ell))$, there is a unique $\delta \in (0, \Delta)$ such that $EV = \widehat{a}'(w^h + w^\ell) + \Delta(w^h - w^\ell) - \delta(w^h - w^\ell) = EV(\frac{\Delta}{2}, \delta)$. Therefore, Λ consists entirely of such extreme points, and thus $\varepsilon = \frac{\Delta}{2}$ at an interior solution to (21).■
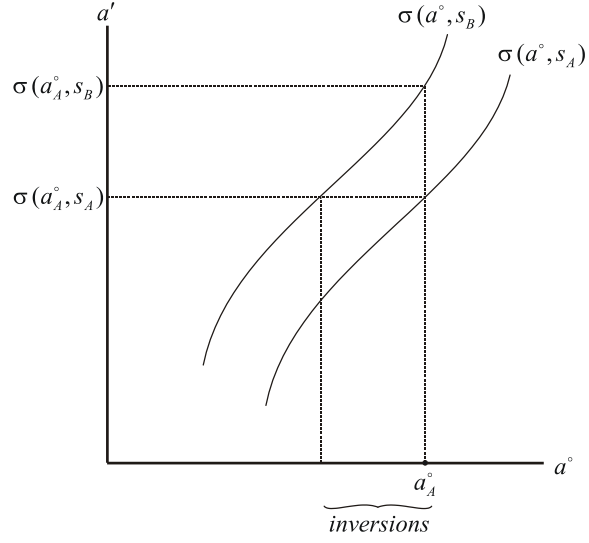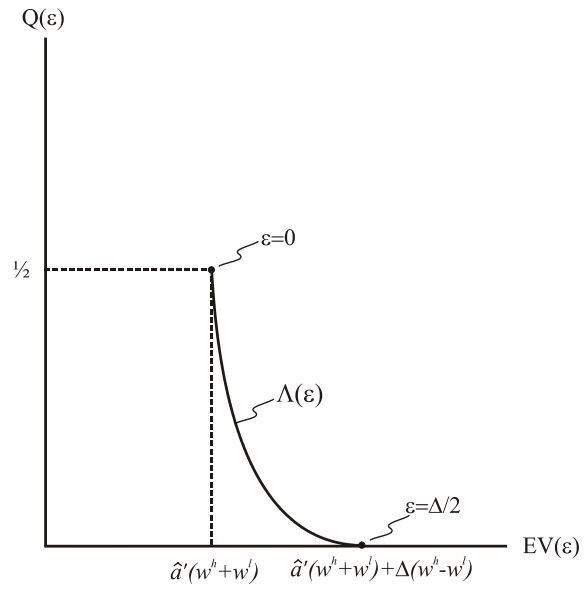
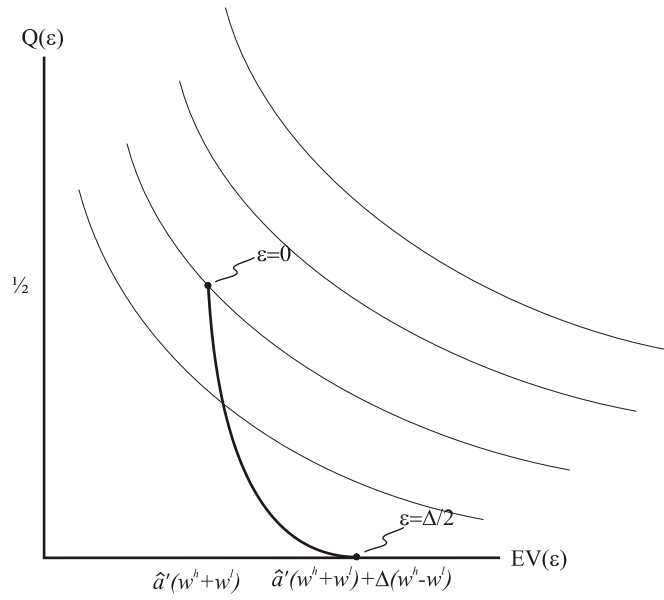Figure 1: Inversions for given a_A°.

Figure 2: Achievable set.

26

Figure 3: Full equality.
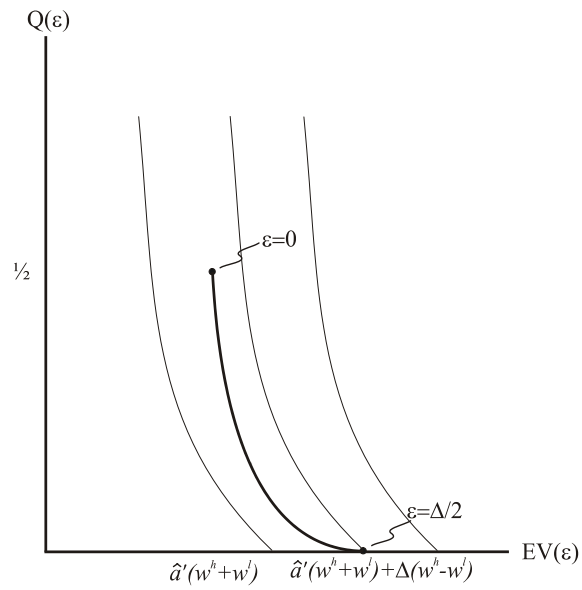
Figure 4: Maximal discrimination.

27

$Q(\varepsilon)$

½

$\varepsilon=0$

$\varepsilon=\Delta/2$

$EV(\varepsilon)$

$\hat{a}'(w^h+w^l)$ $\hat{a}'(w^h+w^l)+\Delta(w^h-w^l)$
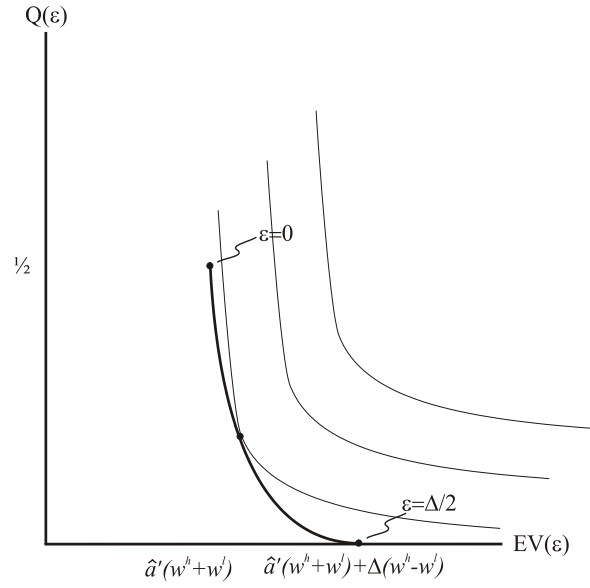
Figure 5: Interior optimum.

$Q(\varepsilon,\delta)$

$\varepsilon=\delta/2$

$\Lambda(\varepsilon,\delta)$

$\varepsilon=\Delta/2$

$\Lambda(\varepsilon)$

$EV(\varepsilon,\delta)$

$\hat{a}'(w^h+w^l)+\Delta(w^h-w^l)-\delta(w^h-w^l)$
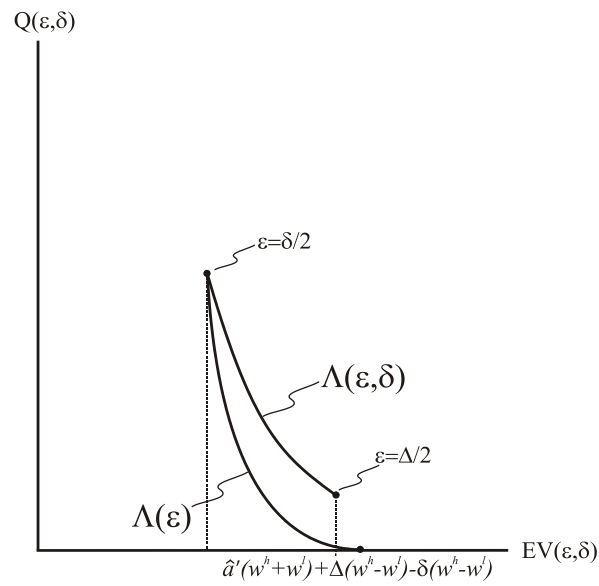
Figure 6: Relationship between $\Lambda(\varepsilon,\delta)$ and $\Lambda(\varepsilon)$.
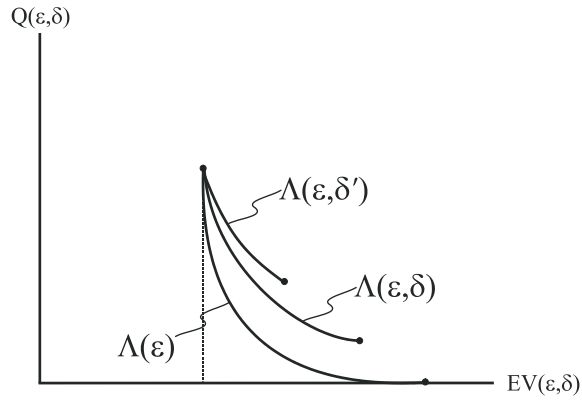
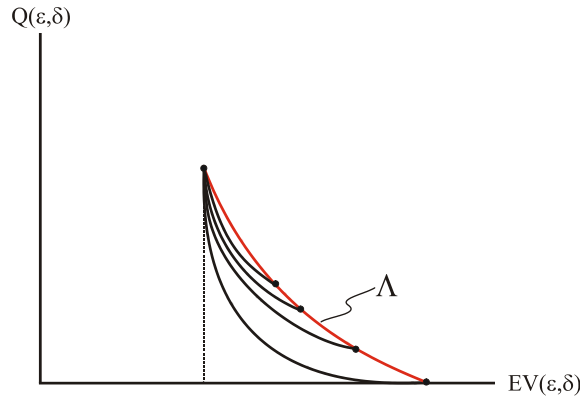Figure 7: $\Lambda(\varepsilon, \delta)$ versus $\Lambda(\varepsilon, \delta')$ for $\delta' > \delta > 0$.
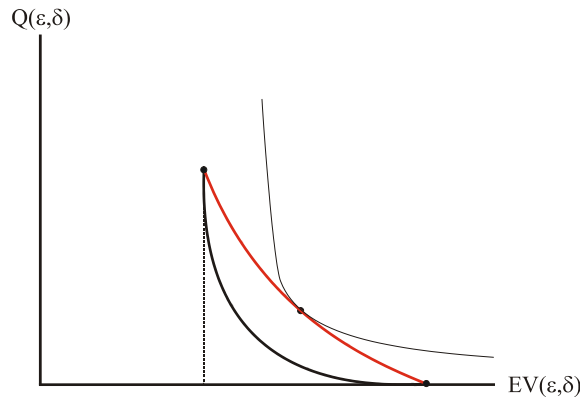

Figure 8: Global achievable set Λ.

Figure 9: Globally optimal policy (ε, δ).

29