Affirmative Action Policies for Top-k Candidates Selection

With an Application to the Design of Policies for University Admissions

Michael Mathioudakis
University of Helsinki

Carlos Castillo
Universitat Pompeu Fabra

Giorgio Barnabo
Sapienza University of Rome

Sergio Celis
Universidad de Chile

ABSTRACT

We consider the problem of designing affirmative action policies for selecting the top-k candidates from a pool of applicants. We assume that for each candidate we have socio-demographic attributes and a series of variables that serve as indicators of future performance (e.g., results on standardized tests). We further assume that we have access to historical data including the actual performance of previously selected candidates. Critically, performance information is only available for candidates who were selected under some previous selection policy.

In this work we assume that due to legal requirements or voluntary commitments, an organization wants to increase the presence of people from disadvantaged socio-demographic groups among the selected candidates. Hence, we seek to design an *affirmative action* or *positive action* policy. This policy has two concurrent objectives: (i) to select candidates who, given what can be learnt from historical data, are more likely to perform well, and (ii) to select candidates in a way that increases the representation of disadvantaged socio-demographic groups.

Our motivating application is the design of university admission policies to bachelor's degrees. We use a causal model as a framework to describe several families of policies (changing component weights, giving bonuses, and enacting quotas), and compare them both theoretically and through extensive experimentation on a large real-world dataset containing thousands of university applicants. Our paper is the first to place the problem of affirmative-action policy design within the framework of *algorithmic fairness*. Our empirical results indicate that simple policies could favor the admission of disadvantaged groups without significantly compromising on the quality of accepted candidates.

1 INTRODUCTION

In recent years, researchers across several areas of computing, including data mining, machine learning, and information retrieval, have sought to develop methods that are non-discriminatory, i.e., that do not generate an unjustified disadvantage for members of a socially salient group.

Predictive policies for top-k selection. We consider a general setting in which we have a large pool of applicants for, e.g., an educational program, a scholarship, or a job, and we want to select the k most promising ones. This selection is based on the candidates' predicted future performance. For instance, in the case of university admissions, this prediction stems from previous grades obtained in different subjects (transcripts), scores in relevant standardized tests, ratings from interviews with the candidate, and so on.

The predictor, which induces a *selection policy*, is created using historical data that tracks the performance of previously selected candidates through outcome variables such as academic achievement or productivity measurements. A key constraint in our setting is that we assume such outcomes can only be observed for candidates that have been selected in the past.

Affirmative action in top-k selection. One of the most common ways of operationalizing algorithmic fairness is through a notion of *demographic parity* or *statistical parity*, which requires that a decision such as accepting or rejecting a candidate is independent of the protected attribute (for a criticism of this notion see [8]). In our setting, for example, we want to admit candidates at similar rates throughout all groups that make up the whole population. Given that disparities exist, we seek an *affirmative action* policy: a temporary intervention aimed at increasing the representation of an underrepresented group.

University admissions. Our motivating example and the problem domain we use throughout this paper are university admission policies. In several countries, standardized tests for university admission are administered at the end of the last year of high school or community college. They are often framed as a measurement of aptitude for university studies, i.e., a prediction of how a student would eventually perform if admitted. For most students, admission decisions are based on an *admission score* computed from test scores and/or high school grades (e.g., through a linear combination).

However, an admission score is not the only mechanism for university admission. Universities may want to have a diverse body of students, or may have a social role in addition to their educational one, such as promoting gender equality and social mobility. For this reason, university admissions often include a predictor of students' future performance plus a patchwork of additional affirmative action programs addressed at increasing the number of accepted candidates from disadvantaged backgrounds.

Designing affirmative action policies. The goals of an affirmative action top-k selection policy are two fold. First, the policy should select candidates who have a high expected performance. Second, the policy should ensure a sufficient number of candidates from disadvantaged backgrounds are selected. Both properties can be seen as expressing notions of fairness: the former towards good performance; the latter against disadvantages that are not related with performance.

Effectively, to design such a policy one needs to be able to predict its effect before it is enacted. The goal of this paper is to formalize the problem of designing data-driven affirmative action policies and to develop algorithms to build such policies starting from historical data. Given the experimentation we perform, our contributions have a direct application to university admissions, and they are directly extensible to other applications.

Our approach. To formalize the questions we consider, we build a probabilistic causal model that captures the selection process and makes our assumptions about the underlying data explicit. The causal model takes into account each applicant's input attributes (e.g., test scores), as well as a number of *sensitive* socio-demographic attributes, to compute an admission score, which is the main basis for their selection. Given the causal model, we consider three types of policies: (1) COEFFICIENTS: policies that adjust the weight of different components of the admission score to emphasize those that favor disadvantaged groups; (2) BONUS: policies that give extra points in the admission score to applicants from disadvantaged groups; and (3) QUOTA: policies that enforce a certain selection rate for applicants from disadvantaged groups. Policies are considered from the perspective of a single institution and optimized to increase the expected performance of the selected applicants while accepting a suitable proportion of disadvantaged applicants.

In our analysis, we build instances of causal models from the available data, find suitable parameters for a series of affirmative action policies, and present the results of an empirical study over university admissions considering thousands of applicants.

2 RELATED WORK

2.1 Fairness-aware top-k selection

This paper belongs to the emergent research field of *algorithmic fairness*, specifically to fairness in predictive decision making (for a survey, see [20]). Historically, data mining and machine learning applications have been the focus of most works, with comparatively less research dedicated to fairness in ranking or top-k selection.

Ranking with fairness constraints, in batch processing and online processing scenarios respectively, is the focus of two recent works by Celis et al. [5] and Stoyanovich et al. [26]. In these works, every element to be ranked has a certain utility and belongs to a certain group, and for every ranking position $r = 1, 2, \ldots, k$ there are two constraints per group indicating the minimum and maximum number of candidates from that group that must be among the top-r positions. The goal is to rank candidates by decreasing utility without violating the constraints.

Zehlike et al. [28] define a *fair representation condition* based on a binomial test that seeks to determine if a ranking of k elements is compatible with a process in which candidates have been drawn at random from protected and non-protected populations with a known probability. The fair representation condition together with a process for adjusting for multiple hypothesis testing can be used to obtain the bounds used in ranking with fairness constraints. Kearns et al. [16] study the case of selecting k candidates drawn from d different groups in which candidates across groups are not comparable. Hence, what matters is the relative position of candidates within their groups.

2.2 Affirmative action

"Downstream" effects of affirmative action. Several recent papers study "downstream" or "cascade" effects of affirmative action policies, mostly through a theoretical approach. Hu and Chen [14] consider a labor market in which firms are either temporary or permanent employers. Temporary employers are forced to use hiring policies that favor disadvantaged groups, while permanent employers can hire based on expected utility alone. The goal is to increase the employability of candidates from disadvantaged groups through their employment in the temporary market. Kannan et al. [15] analyze the potential effects of affirmative action in education on the employability of graduates from disadvantaged groups. The goal is to achieve parity of employability or create an incentive for employers to adopt a color-blind hiring policy. Mouzannar et al. [21] analyze the effect of affirmative action policies on the qualifications of different groups in society.

Data-driven analysis. Over the last four decades prestigious universities, many of which are public institutions, have become even more selective, having a negative effect on students from low socio-economic backgrounds and underrepresented minorities [24]. This effect is mainly explained by the lower performance of disadvantaged students on standardized tests and other admission requirements that favor applicants from well-off families [3]. However, admission scores might underestimate university performance for disadvantaged students. For instance, a study by Wightman [27] on nearly 25,000 law students in the US found that while students of color had significantly poorer results on admission tests than white students, their difference in terms of performance once they were admitted (their probability of passing the bar exam) did not justify the large gap between the admission rates of both groups.

As a result, countries and universities have promoted affirmative action policies. For instance, Brazil has promoted bonus [9] and racial quotas [10]. The evidence suggests that the implementation of these policies increased both black students and students coming from low socioeconomic backgrounds in large public universities [10]. In Chile, where significant gaps are observed in standardized math tests between students of public versus private schools [6] and between women and men [1], coefficient-based policies have also shown some effectiveness [19]. Institutional policies seem to have an effect on admission systems that include socio-demographic information in the selection process, as done by several universities in the US. For instance, Bastedo et al. [4] found that when admission officers in the US consider contextual variables (economic hardship, family background) into the admission decision, the chances of admitting disadvantaged students increases.

2.3 Causal reasoning and fairness

In a recent paper, Hardt et al. [13] explain how causal reasoning, as developed by Pearl [22] and others, offers a formal and rich theoretical framework to express notions of algorithmic fairness that goes beyond statistical / observational measures. Indeed, the work of Kusner et al. [18] and Kilbertus et al. [17] builds upon that intuition by studying notions of *counterfactual fairness,* i.e., fairness notions that take into account the causal structure of the data's generative model. In our work, we use the framework of causal reasoning to define the model for the admission mechanism (Section 3), and confirm that the available data allow us to make unbiased estimates of the performance of alternative admission policies (Section 6).
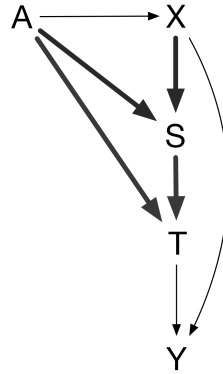
3 SETTING

In the setting we consider, selection of candidates is performed by *one institution*. We begin by defining a probabilistic causal model that captures this process, expresses our assumptions about the data, and allows us to quantify the effect of different selection policies (Section 3.1). For a given policy, we then define measures of *utility*, capturing the performance of selected candidates, and *fairness*, capturing the similarity between selection rates (Section 3.2). Next, we describe three types of policies (Section 3.3), and present a formal definition of the technical problem we address as a trade-off between utility and fairness (Section 3.4). Finally, we discuss various modeling choices we make (Section 3.5).

3.1 Causal model

Our design of selection policies is guided by historical data that are the result of a particular enacted selection policy. To predict what data we would observe under alternative policies, it is necessary to express formally our knowledge and assumptions about the decision mechanisms we study. We do this by defining a *causal diagram* [22] capturing our background knowledge about causal relationships between quantities of interest, as shown in Figure 1. The diagram describes the selection mechanism from the point of view of *one candidate* to *one institution*, and assumes that each candidate is associated with socio-economic attributes A, which are considered sensitive, and a series of input scores X (e.g., results of qualification tests), which are considered non-sensitive. These are combined into a selection score S which is the basis for the institution's decision T. In particular, given input scores $X = x$ and socio-economic attributes $A = a$, the candidate obtains a selection score $S = s$ according to $S = f(X, A)$. Here, we allow the selection score S to depend on A to allow for affirmative action through the selection score function. Based on this score S a decision T is made in favor or against the candidate. Again, we allow T to be determined by A so as to allow also for affirmative action through policies that operate at the stage of decision.

For the purposes of affirmative action policy design, it is also crucial that we model the possibility that sensitive socio-economic attributes A influence the input scores X. This is done through the dependency $A \rightarrow X$ in the diagram. This is the case if, for instance, university applicants from families of higher income have access to training or resources that are not affordable for families of lower income. In that case, an affirmative action policy tries to bring *equality of opportunity* [25].

One could also consider the possibility that sensitive socio-economic attributes A influence the performance of the selected candidates Y, as well. This would be particularly useful to model special cases where low socio-economic status may be associated to *low* input scores but *high* performance of selected candidates, for instance due to the persistence required to overcome their initial disadvantage (see, e.g., observations by Glynn [12] on community colleges in the US); in this case an affirmative action policy can lead to the admission of applicants of higher Y by *reducing discrimination*. Nevertheless, in this first effort to formalize the design of selection policies as an algorithmic problem, we have opted not to include this dependency into the analysis. We have made this choice because (1) it simplifies the analysis, (2) it turns out not to

Figure 1: Diagram of causal model. Selection policies are defined via the causal relationships $(A, X) \rightarrow S$ **and** $(A, S) \rightarrow T$.

Symbol	Description
A	Socio-economic attributes
X	Input scores (e.g., results of qualification tests)
S	Selection score: combination of input scores
T	Decision of the institution: select or not select
Y	Performance of selected candidate

be crucial for our empirical analysis case on university admissions (Section 6), (3) it is reasonable to assume that, in real-life systems that have been used in practice for many years (as are, typically, university admission systems) the input scores X are adaptively designed over time to be good predictors of performance Y. Therefore, in what follows, we assume that input scores X provide all available information to predict performance Y (hence the dependency $X \rightarrow Y$ in the diagram), while they might be influenced by sensitive attributes A (hence the dependency $A \rightarrow X$). In the context of university admissions, this would mean that socio-economic attributes A influence the grades X obtained by university applicants – and that, among all available information for applicants, the grades X are enough to predict as accurately as possible an applicant's performance Y at the university.

The institution under consideration is interested in the acceptance of only a certain number or fraction of top candidates. Formally, we require every selection policy to be *calibrated* – i.e., that it lead, in expectation, to the selection of a certain fraction θ of candidates, considered known and fixed for the institution.

$$\mathbb{E}[T] = P(T = 1) = \theta \qquad (1)$$

If the decision for a candidate is positive ($T = 1$), we observe their performance Y. If a candidate is rejected the performance of that candidate is not observed, and we write $Y = n/a$.

3.2 Measures

We aim to define selection policies that lead to the selection of the candidates with the highest expected performance. At the same time, we aim to mitigate the disadvantage of certain demographic groups. Towards this end, we define the following measures.

Utility of Selection (UoS). A policy is useful if it leads to the selection of the candidates that actually perform best. This is captured by the following measure, defined in terms of the causal model.

$$\mathbf{UoS} = \mathbb{E}[Y|T = 1]$$

Demographic Disparity (DmD). Our objective is to produce selection policies that minimize the difference between the selection rates of candidates of different backgrounds, according to their socio-economic features $\mathbf{A}$. The following measure is defined for a group of candidates with sensitive properties $\mathbf{A} = \mathbf{a}$; it compares the selection rate for this group to the candidates outside the group.

$$\mathbf{DmD_A} = P(T = 1|A = a) - P(T = 1|A \neq a)$$

The demographic disparity is equivalent to the *risk difference* measure from the literature [7], where risk is the probability of not being selected $1 - P(T = 1)$, and the difference is taken between the sensitive group and its complement.

3.3 Types of selection policies

To define alternative selection policies, we essentially have to define instances of the relationships $(A, X) \rightarrow S$ and $(A, S) \rightarrow T$, marked with thick arrows in Figure 1. In other words, for a given causal model, we aim to define the scoring function $S = f(A, X)$ and selection function $T = t(A, X)$ so that the policy exhibits desired values of the utility and fairness measures as described earlier. We say that every pair of functions $f(A, X)$ and $t(A, X)$ defines a *selection policy* within the *selection mechanism* defined by the model. We consider three types of policies: Coefficients, Bonus, and Quota.

Coefficients-**based policies**. This type of policy is completely determined by a series of coefficients $\mathbf{w}$ that are used to compute the selection score S as a linear combination of the input scores X. No sensitive attributes are used. The score is computed as

$$S = f(X = x) = w \cdot x, \quad |w|_1 = 1. \tag{2}$$

Note that $\mathbf{w}$ is normalized. Each component of $\mathbf{w}$ represents the degree to which an input score counts towards the selection score. Candidates are admitted if their score exceeds a threshold τ, in which case we write $T = 1$ — otherwise, we have $T = 0$.

$$T = t(S) = \begin{cases} 1 \text{ if } S \geq \tau \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

Bonus **policies**. This type of policy differs from Coefficients policies in that sensitive attributes $\mathbf{A}$ contribute to the selection score of candidates. Specifically, the group of candidates with $\mathbf{A} = \mathbf{a}$ receive an additive bonus b in their score.

$$S = f(X = x, A) = \begin{cases} w \cdot x + b, & \text{if } A = a \\ w \cdot x, & \text{if } A \neq a \end{cases}, \quad |w|_1 = 1. \tag{4}$$

Note that the selection decision T is determined in exactly the same way as for Coefficients policies, i.e., via a threshold τ.

Quota **policies**. This type of policy determines the selection score in the same way as Coefficients, i.e., via a linear function with normalized weights $\mathbf{w}$. However, unlike Coefficients policies, sensitive attributes $\mathbf{A}$ contribute to the decision T. Quota policies

constraint the selected candidates from group $\mathbf{A} = \mathbf{a}$ to be (in expectation) a fraction $q \in [0, 1]$ of accepted candidates.

$$P(T = 1, A = a) = q \cdot \theta \tag{5}$$

Specifically, the selection decision T under a Quota policy is determined based on separate thresholds $\tau_{(A \neq a)}$ and $\tau_{(A = a)}$ for the two groups, $(A = a)$ and $(A \neq a)$, respectively.

$$T = t(S, A) = \begin{cases} 1 \text{ if } (A \neq a) \text{ and } S \geq \tau_{(A \neq a)} \\ 1 \text{ if } (A = a) \text{ and } S \geq \tau_{(A = a)} \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

The two thresholds are chosen so that Equation 5 is satisfied.

3.4 Problem definition

Having defined the set of policies that we consider, we can now define our objective formally. If we were to ignore the goal of favoring disadvantaged group $(A = a)$ via an affirmative action policy, it would be natural to aim for the policy that optimizes the utility $\mathbf{UoS}$ of the selection mechanism. On the other extreme, if we were to target only the demographic disparity of those selected, it would be natural to aim for a policy that eliminates $\mathbf{DmD}$ (i.e., leads to $\mathbf{DmD} = 0$). In general, we consider cases where we are willing to trade λ units of utility to decrease disparity by one unit. We thus define our technical objective as follows.

Problem 1. *For sensitive property A, and two associated groups of candidates $(A = a)$ and $(A \neq a)$, define $f(A, X)$ and $t(A, S)$ so as to obtain a calibrated policy that maximizes*

$$o(f, t) = \mathbf{UoS}(f, t) - \lambda |\mathbf{DmD_A}(f, t)|, \quad \lambda \geq 0.$$

Notice that the objective function considers the absolute value of $\mathbf{DmD}$, as we wish to eliminate $\mathbf{DmD}$ rather than maximize it in favor of the currently disadvantaged group.

3.5 Discussion on modeling choices

The causal model we described above might not match exactly every real selection mechanism, as it is based on certain assumptions, some made for the purposes of simplification. We now briefly discuss those in the context of university admissions, which is the domain in which we perform our real data analysis (Section 6).

Candidate decisions might not be independent. In our setting, the decision is made independently for each applicant based on their individual attributes. We have made this modeling choice because treating applicants independently simplifies analysis considerably, while it still allows us to understand the properties of the different policies. However, we should keep in mind that the independence assumption does not hold in some university admission systems. For example, for many universities there is a hard constraint on the absolute number of students to be admitted – and hence admission decisions are *not independent*, as one candidate's admission can be another one's rejection.

Selection does not imply enrollment. We have also made the simplifying assumption that an applicant who is admitted by a university is bound to attend it. However, in reality applicants do not apply for admission to one institution only and might be admitted

by several. In some admission systems, applicants are forced to prioritize their preferences for institutions and attend the first to admit them. In other admission systems, applicants have the opportunity to choose which one to attend. The model does not capture explicitly such admission systems. However, its assumptions work well for top universities that students are almost certain to attend if admitted. We find the model to be a reasonable approximation of reality for our empirical analysis, which concerns a top university in its country and to which most selected applicants enroll (about 95%). We also acknowledge that highly selective universities with affirmative action policies may create a mismatch issue (i.e., accepted students may be better off in another type of institution) as reported in countries such as the US and India [11].

4 POLICY SEARCH FOR A SINGLE SENSITIVE ATTRIBUTE

For ease of exposition, this section describes algorithms to determine affirmative action policies for a single sensitive attribute $\mathbf{A}$. Section 5 extends them to multiple sensitive attributes.

4.1 COEFFICIENTS policy

Let us consider the case of $\lambda = 0$, in which the objective function simplifies to the measure of utility **UoS**. Let $m(\mathbf{x}) = \mathbb{E}[\mathbf{Y}|\mathbf{X} = \mathbf{x}, \mathbf{T} = 1]$ be the expected performance $\mathbf{Y}$ of a selected applicant with input scores $\mathbf{x}$. By straightforward calculations, we find that

$$\mathbf{UoS} = \mathbb{E}[\mathbf{Y}|\mathbf{T} = 1] = \frac{1}{\theta} \int m(\mathbf{x})\, \delta(\mathbf{f}(\mathbf{x}) \geq \tau)\, d\mathbf{P}(\mathbf{x}). \quad (7)$$

Intuitively, Eq. 7 expresses that **UoS** is the average expected performance of applicants that exceed the threshold τ. Using $\mathbf{f}(\mathbf{x}) = m(\mathbf{x})$ maximizes **UoS** over all possible scoring functions. In other words, if we could use the expected performance of an applicant as the scoring function, the resulting COEFFICIENTS policy would select those applicants that are expected to have the best performance $\mathbf{Y}$ and it would be optimal for $\lambda = 0$. However, we are constrained to use linear functions as scoring function $\mathbf{f}(\mathbf{X})$, as defined in Equation 2.

In the case where $m(\mathbf{x})$ is indeed a linear function, then using $\mathbf{f}(\mathbf{x}) = m(\mathbf{x})$ leads to an optimal COEFFICIENTS policy for $\lambda = 0$. To state the claim formally, let $F(\mathbf{S} = \mathbf{s})$ be the cumulative probability distribution for selection $\mathbf{S}$ and $F^{-1}(\beta)$ the inverse cumulative distribution, i.e., the function that returns the score below which lies a fraction β of the population of applicants.

LEMMA 1. *Let $m(\mathbf{x}) = \alpha_0 + \alpha\, \mathbf{c} \cdot \mathbf{x}$ with normalized coefficients, $|\mathbf{c}| = 1$. Then, the following* COEFFICIENTS *policy is optimal for $\lambda = 0$:*

$$\mathbf{S} = \mathbf{f}(\mathbf{X}, \mathbf{A}) = \mathbf{c} \cdot \mathbf{X}$$

$$\mathbf{T} = \mathbf{t}(\mathbf{S}, \mathbf{A}) = \begin{cases} 1 \ if\, \mathbf{S} \geq \tau \\ 0 \ otherwise \end{cases} \quad (8)$$

$$\tau = F^{-1}(1 - \theta)$$

Conceptually, the optimal COEFFICIENTS policy for $\lambda = 0$ corresponds to ordering applicants in decreasing order of $\mathbf{c} \cdot \mathbf{X}$ and keeping the top θ-fraction of them by setting the threshold τ appropriately. By construction, this policy is calibrated.

For general $\lambda \geq 0$, $\mathbf{c}$ is not necessarily optimal anymore – instead, the optimal COEFFICIENTS policy would require different score
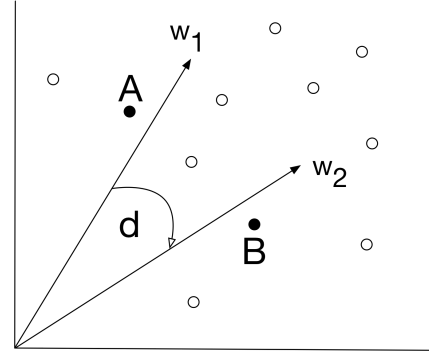

Figure 2: For a given weight vector w, the selection score of each candidate with input scores x is defined as the inner product $(\mathbf{w} \cdot \mathbf{x}) \propto \|\mathbf{x}\| \cos(\mathbf{w}, \mathbf{x})$. In the figure, each circle corresponds to the input scores of a candidate. The two candidates A and B have input scores with same l_2-norm. Candidate A is assigned with a higher selection score than B under score weight vector w_1. However, the situation is reversed under score weight vector w_2, obtained by a rotation of angle d from w_1. Under w_2, candidate B is assigned a higher selection score than B.

weights. To see why, notice that, for two data points $\mathbf{x}_1, \mathbf{x}_2$, we have

$$\mathbf{w} \cdot \mathbf{x}_1 > \mathbf{w} \cdot \mathbf{x}_2 \ \text{iff} \ \|\mathbf{x}_1\| \cos(\mathbf{w}, \mathbf{x}_1) > \|\mathbf{x}_2\| \cos(\mathbf{w}, \mathbf{x}_2).$$

Therefore, the selection score depends on the angle between the vectors of input scores $\mathbf{x}$ and score weights $\mathbf{w}$. In general, a rotation of score weights $\mathbf{c}$ by some angle $\mathbf{d}$ might lead to score weights $\mathbf{w}$ that change the ordering of datapoints $\{\mathbf{x}\}$ by score value $\{\mathbf{f}(\mathbf{x})\}$ – and thus possibly lead to smaller **DmD** and better objective value (see Figure 2).

To identify the optimal COEFFICIENTS policy we could use a sample of data $\{\mathbf{x}\}$ and follow an approach similar to Asudeh et al. [2]. For a given set of applicants with input scores $\{\mathbf{x}\}$, it is possible to partition the vector space of $\mathbf{w}$ into regions such that any vector of score weights $\mathbf{w}$ within the same region leads to the same ordering of applicants, where the ordering is performed by $\{\mathbf{w} \cdot \mathbf{x}\}$. However, we would have to perform an exhaustive search of such regions, which is inefficient for high-dimensional data. Instead, we experiment with a simple grid-like search strategy: starting with vector $\mathbf{w} = \mathbf{c}$, we rotate the score weight vector $\mathbf{w}$ by small increments towards different directions $\mathbf{d}$, to obtain new weight vectors $\mathbf{w}' = \rho(\mathbf{w}; \mathbf{d})$ and evaluate the objective function in every step. The approach is described in Algorithm 1.

4.2 BONUS policy

Let $\mathbf{M}$ be the optimal COEFFICIENTS policy, having parameters $\mathbf{c}$ and τ as defined in Equations 8. We now consider a BONUS policy $\mathbf{B}$ that favors with b points the group $\mathbf{A} = \mathbf{a}$ that is *disadvantaged* under $\mathbf{M}$, assuming without loss of generality that

$$\mathbf{P}(\mathbf{T} = 1|\mathbf{A} = \mathbf{a}; \mathbf{M}) \leq \mathbf{P}(\mathbf{T} = 1|\mathbf{A} \neq \mathbf{a}; \mathbf{M}).$$

For the moment, let us restrict $\mathbf{B}$ to use the same score weights $\mathbf{c}$ as $\mathbf{M}$, but score threshold τ_b that is potentially different than the

threshold τ used by **M**. We lift the restriction on score weights at the end of Section 4.2.

Calibrating τ_b. As discussed in Section 3, all policies are required to be calibrated, i.e., adhere to condition 1. Towards this end, let us consider a fixed bonus b $\geq$ 0. The lemma below guarantees that there is an algorithm to choose its threshold τ_b so that **B** becomes calibrated.

LEMMA 2. *The score threshold τ_b that makes **B** calibrated can be found via binary search over the interval $[\tau, \tau + b]$*

Optimizing b. Given τ_b obtained with the procedure described above, and **w** = **c**, how do we determine b = b_{OPT} that optimizes the objective of Problem 1? Our main observation is that we can pre-determine a narrow range of values within which the optimal b_{OPT} must fall. To be precise, let $F(S = s|A = a; M)$ and $F(S = s|A \neq a; M)$ be the cumulative probability functions for the score S of the two groups under COEFFICIENTS policy **M**. Let $g(\beta) = F^{-1}(\beta|A = a; M)$ and $h(\beta) = F^{-1}(\beta|A \neq a; M)$ be the corresponding inverse functions: given a cumulative probability value β, they return the score S of the β-th applicant of the respective group. The search for optimal b_{OPT} can be focused in the interval stated in the following lemma.

LEMMA 3. $b_{OPT} \in I = [0, g(1 - \theta) - h(1 - \theta)]$

PROOF. *First*, notice that the following condition holds for **M**.

$$P(T = 1|A = a) \leq \theta \leq P(T = 1|A \neq a) \qquad (9)$$

This follows from the assumption that **A** = **a** is the disadvantaged group and the fact that the overall proportion θ of accepted applicants must be between the proportions of accepted applicants of the two groups **A** = **a** and **A** $\neq$ **a**. It is easy to see that the proportion $P(T = 1|A = a)$ of admitted candidates of the disadvantaged group is non-decreasing for increasing b $\geq$ 0, while the corresponding proportion $P(T = 1|A \neq a)$ for the advantaged group is non-increasing for increasing b $\geq$ 0. This implies that, for increasing b $\geq$ 0, there is a bonus value b = b_{DmD} for which we have

---

Algorithm 1: Search for COEFFICIENTS policy
Parameters: {**d** }: set of rotation directions, *k*: number of rotations in each direction

1 /* initialize weights */
2 **w**:= **c**, $\mathbf{c}_{opt}$:= **c**
3 **for** *each direction* **d do**
4   **for** *i = 1..k* **do**
5   /* rotate the vector of score weights */
6   **w**:= $\rho(\mathbf{w}; \mathbf{d})$
7   /* normalize the vector */
8   **w** := $\mathbf{w}/|\mathbf{w}|_1$
9   /* set the threshold τ to make the policy calibrated */
10         $\tau = F^{-1}(1 - \theta)$
11   /* update best policy */
12   **if** $o(\mathbf{w}) > o(\mathbf{c}_{opt})$ **then**
13             $\mathbf{c}_{opt}$:= **w**
14 **return** $\mathbf{c}_{opt}$

Algorithm 2: Grid-based search for b_{OPT}
Parameters: *k*: granularity

1 $\epsilon = b_{DmD}/k$
2 b_{OPT}:= 0
3 **for** *i = 1..k* **do**
4   b = $i \cdot \epsilon$
5   τ_i := binarysearch($\tau_{i-1}, \tau + b$)
6      $v := o(b, \tau_b)$
7   **if** $v > v_{OPT}$ **then**
8   b_{OPT}:= b; $v_{OPT} := v$
9 **return** b_{OPT}

---

equality between the quantities of Condition 9 and **DmD** = 0. This is achieved for a bonus value that makes the (1 - θ)-th applicant of group **A** = **a** have a score equal to the score of the (1 - θ)-th applicant of group **A** $\neq$ **a**. Formally, we have $b_{DmD} = g(1-\theta)-h(1-\theta)$ and the BONUS policy with b = b_{DmD} is calibrated for score threshold $\tau_{b_{DmD}} = F^{-1}(1 - \theta|A \neq a; M)$. The above implies that the absolute value |**DmD**| of disparity decreases over the interval b $\in [0, b_{DmD}]$ and increases over b $\in [b_{DmD}, \infty]$. *Second*, it is easy to see that **UoS** is decreasing for increasing b. The two observations imply that the objective function (Problem 1) is optimized in $[0, b_{DmD}]$. □

One simple algorithm to search for b_{OPT} is grid-based search, in which we evaluate the performance of BONUS (b) policies for uniformly distributed b in the range **I**. The calibration for each policy explored in this grid-based search can be done somewhat more efficiently based on the following generalization of Lemma 2, which allows us to narrow the range of binary search for the threshold τ that makes a policy calibrated. The search algorithm for b_{OPT} is shown in Algorithm 2 (proof omitted).

LEMMA 4. *For $i \in \{1, 2\}$, let τ_i be the threshold that makes the BONUS policy with b_i bonus points calibrated, and let τ be the threshold for **M**. If $b_1 \leq b_2$, then $\tau_1 \leq \tau_2 \leq \tau + b_2$.*

Solving Problem 1. So far in Section 4.2, we have constrained ourselves to BONUS policies that adopt fixed weights **w** = **c**. This raises the question: how do we choose **w** jointly with other parameters to find the optimal bonus policy for Problem 1? The answer is that we can simply use **w** = **c**, the optimal score weights for a COEFFICIENTS policy with $\lambda = 0$, as stated in following lemma.

LEMMA 5. *There is an optimal bonus policy with **w** = **c**.*

PROOF. *First*, consider two BONUS policies $\mathbf{B}_1$ and $\mathbf{B}_2$ with the following properties: (i) $\mathbf{B}_1$ and $\mathbf{B}_2$ are associated with the same value of |**DmD**|; (ii) $\mathbf{B}_1$ uses score weights **w** = **c**, while $\mathbf{B}_2$ does not. Then, $\mathbf{B}_1$ is associated with higher **UoS** and therefore higher objective value than $\mathbf{B}_2$. *Second*, notice that for well-behaved score distributions (specifically, for cumulative score distributions $F(S = s|A = a; M)$ and $F(S = s|A \neq a; M)$ that are continuous), it is possible to choose an appropriate bonus value b $\in (-\infty, +\infty)$ to achieve *any* value of **DmD** $\in (-1, 1)$, irrespective of score weights **w**. Combined, the above two arguments imply that for any BONUS policy $\mathbf{B}_2$ with **w** $\neq$ **c**, we can build a policy with same **DmD** and **w** =**c**, which is associated with higher objective value. □

4.3 Quota policy

We find that there is a relationship of equivalence between Bonus and Quota policies, as pointed out in the following lemma.

Lemma 6. *For a given* Bonus *policy, there is a* Quota *policy that leads to the selection of exactly the same candidates, and vice versa.*

Proof. Let $P(T = 1, A = a)$ be the probability admitting applicants of group $A = a$ under a Bonus policy with bonus value b. Let also q_b be such that

$$P(T = 1, A = a) = q_b\, \theta$$
$$P(T = 1, A \neq a) = (1 - q_b)\, \theta.$$

and define the probabilities

$$\sigma_{A=a} = P(T = 1 \mid A = a) = \frac{q_b\, \theta}{P(A = a)}$$
$$\sigma_{A\neq a} = P(T = 1 \mid A \neq a) = \frac{(1 - q_b)\, \theta}{P(A \neq a)}.$$

We can construct a Quota policy with a quota $q = q_b$ for the disadvantaged group $(A = a)$, by setting

$$\tau_{(A=a)} = F^{-1}(1 - \sigma_{A=a}|; A = a)$$
$$\tau_{(A\neq a)} = F^{-1}(1 - \sigma_{A\neq a}|; A \neq a),$$

as per the definition of Quota policies in Section 3.3. By construction, we have a Quota policy that accepts exactly the same applicants as the given Bonus policy. A similar argument can be made for the opposite direction. □

This lemma allows us to focus our empirical study (Section 6) mainly on Coefficients and Bonus policies.

5 POLICY SEARCH FOR MULTIPLE SENSITIVE ATTRIBUTES

Our discussion so far has focused on policies that target one sensitive attribute. In this section, we introduce the problem of policy design for multiple sensitive attributes $\{A_i\}$.

Problem 2. *For sensitive properties* $A_1, A_2, \ldots, A_m$, *each partitioned into two associated groups of applicants* $(A_i = a_i)$ *and* $(A_i \neq a_i)$, *define* $f(A_i, X)$ *and* $t(A, S)$ *so as to maximize*

$$o(\mathbf{f}, \mathbf{t}) = \mathbf{UoS}(\mathbf{f}, \mathbf{t}) - \sum \lambda_i\, |\mathbf{DmD}_A(\mathbf{f}, \mathbf{t})|, \quad \lambda_i \geq 0.$$

Algorithms. For Coefficients policies, Algorithm 1 is directly extensible to the setting of multiple sensitive attributes.

For Bonus policies, our goal is to assign bonuses $b_1, b_2, \ldots, b_m$ to the respective disadvantaged groups $(A_i = a_i)$. One approach is, similarly to Algorithm 2, to perform a grid search over possible bonus values for each sensitive attribute. However: (i) Lemma 2 does not extend directly to multiple dimensions, and so does not binary search; (ii) grid search at fine granularity and multiple dimensions is inefficient. Instead, we employ an incremental algorithm that works as follows: (i) at every step, the algorithm maintains bonus values $b_1, b_2, \ldots, b_m$ for each sensitive attribute, (ii) at every step, it greedily increments by an increment δb the bonus value b_i of the attribute A_i which leads to the best objective value.

For Quota policies, the equivalence with Bonus extends to multiple sensitive attributes (proof omitted), and so we do not explore them further here.

6 APPLICATION

We now evaluate the performance of affirmative action policies on a large real-world dataset of university admissions.

6.1 Data

We analyze anonymized data about university admissions from an OECD country (details withheld for double-blind review). In this country, the admission to undergraduate programmes in all major universities is based on the students' performance in a series of standardized tests, administered by a public entity, on language, mathematics, natural sciences, and human sciences, plus their grades in high school. The latter are converted to the same range of values as the standardized tests using a simple formula. Together, test results and grades form the input scores.

Each university announces the degrees it offers (such as engineering, medicine, or law), and coefficients to convert the different input scores into an admission score. For instance, the engineering school we analyze below uses 30% grades, 10% language, 45% mathematics, 15% natural sciences, and 0% human sciences. At the same university, the law school instead uses 40% grades, 25% language, 10% mathematics, 0% natural sciences, and 25% human sciences. After learning their results in the tests, candidates rank degrees and universities by decreasing order of preference. Then, candidates applying to each degree are ranked by admission score and the top ones are admitted.

We have access to two datasets from this admission process. The first corresponds to *national-level data* about students taking the standardized tests; the second corresponds to *university-level data* of students admitted to one of the most sought-after engineering degrees in one of the largest universities (95% of the admitted applicants enroll for this degree). The national-level data is provided by country authorities in anonymized form under a research agreement. . This dataset includes the gender, high school type (public or private), income decile, grades in high school, and results in standardized tests of about 3, 500 applicants who gave the test in 2017 and included the engineering degree of our university-level data among their preferences. The university-level data is similarly provided by university authorities in anonymized form specifically for research on discrimination and bias. For the university-level data, we have access to the same variables, except for income decile, for about 3, 000 admitted applicants from 2010 to 2014. Most importantly, the university-level data contains the grades obtained by admitted students during their first year at the university. We note that the period of the two datasets do not overlap, but due to standardization of grades and test scores, the distributions of input scores do not differ significantly across years.

6.2 Building a causal model

The causal model (see Figure 1) is built from the two datasets we have at our disposal. In the causal model, there are two conditional relationships that need to be set to implement an admission policy,
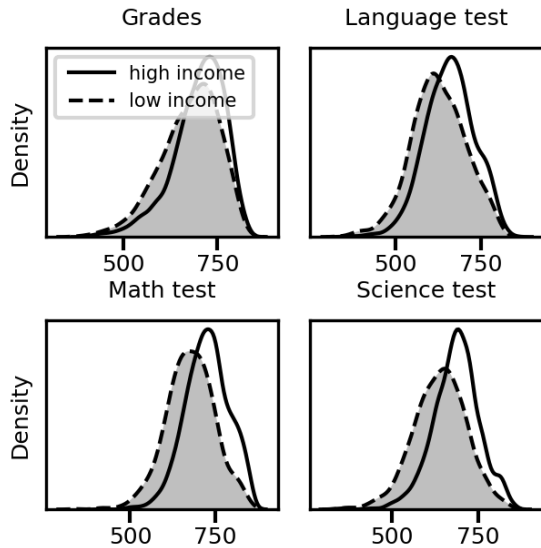
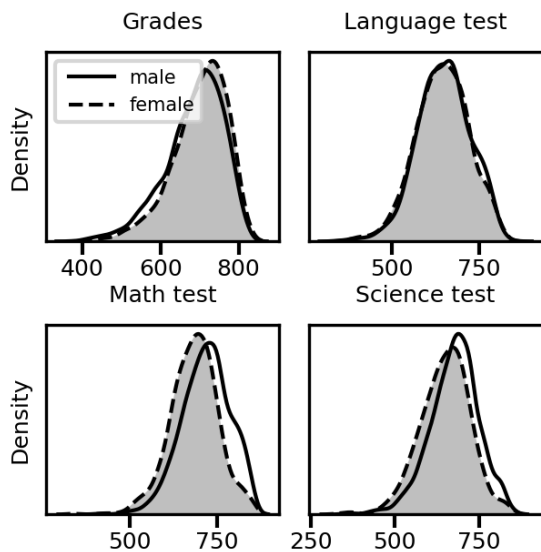Figure 3: Differences in input scores by income.

Figure 4: Differences in input scores by gender.

Figure 5: First year performance vs admission score of the optimal Coefficients **policy with** $\lambda = 0$. **For clarity, the plot shows only a 20% sample of the data points, stratified by** Y.

namely $(A, X) \to S$ and $(S, A) \to T$, and two conditional relationships that need to be learnt from data, namely,

$$A \to X : P(X \mid A) \tag{10}$$

$$X, T = 1 \to Y : P(Y \mid T = 1, X). \tag{11}$$

Both probabilistic quantities above can be extracted directly from the data (e.g., by fitting a statistical model on the related attributes). Here is a subtle point that needs to be addressed for $P(Y \mid T = 1, X)$ (Expr. 11): if we learn a model for it from the data of students who were admitted ($T = 1$) according to a specific admission policy, can we use it to make predictions for students who *would be admitted* under a different policy – or are the observed data somehow *biased* by the decisions of the admission policy that generated the data?
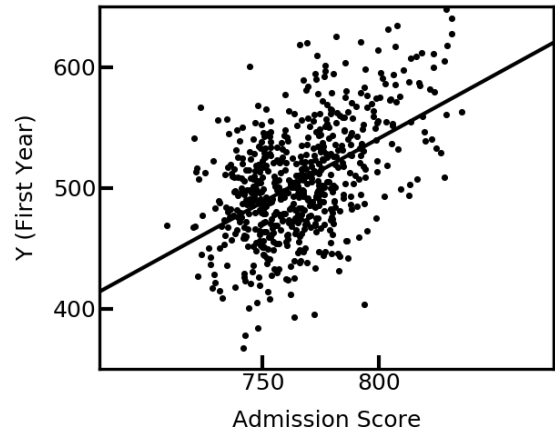
The answer is provided by causality theory [23], from which we deduce that we can, indeed, use the model we learn for from the data to make predictions under different admission policies. The reason is that attribute X is an appropriate 'control' variable for Expr. 11 (more formally, X blocks all back-door paths from T to Y; for an intuitive explanation see [23], Chapter 7).

In what follows, we use the attributes of gender (male or female), high school type (public or private), income (above or below the median) as sensitive attributes A, with the population split in the two groups in parentheses; the values of grades, language test, math test, and natural sciences test as input scores X; and the sum of grades during first year at university as Y (weighted by the credits of each subject; all first-year students take the same subjects).

For Expr. 10, we use directly the distributions extracted from the data, as shown in Figures 3-4 (the plots for school type are omitted here and whenever they are similar to those for income). As we see, the two groups for income (and similarly for school type) exhibit visibly different distributions (Fig. 3), while the distributions differ less for gender (Fig. 4). This suggests that, for the admission system that we study here, there is a stronger case for affirmative action policies that target financial attributes than gender, and the enactment of such policies has larger potential for impact.

For Expr. 11, we learn the best linear weights **c** for a Coefficients policy that targets only UoS (i.e., considers $\lambda = 0$) – they are 47.6% grades, 1.69% language, 23.8% mathematics, and 26.9% natural sciences. As discussed in Section 4.1, these weights also correspond to the linear function that expresses the expected performance Y of a selected applicant with given input scores. Figure 5 shows the actually observed values of admission scores S and performance Y of admitted students during their first year of studies. There is a correlation between the two, which supports our choice of a linear function to model the expectation $m(\mathbf{x}) = \mathbb{E}[Y|X = \mathbf{x}, T = 1]$ (see also beginning of Section 4.1).

We should also note, however, that there is variance in the observed values of Y for a given score value. Such variance does not directly affect the task of policy design in the setting we described in
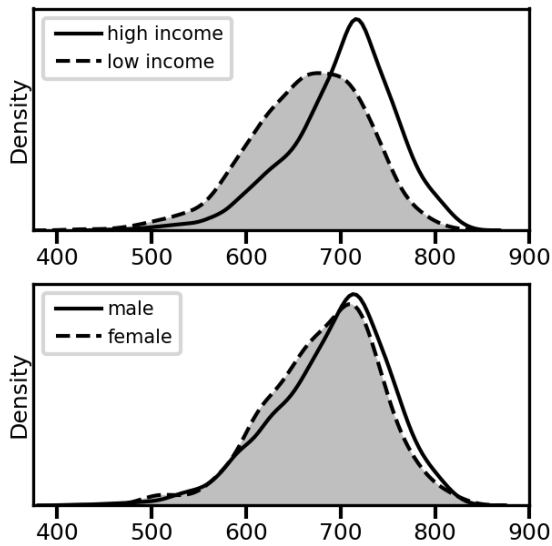
Figure 6: Distribution of admission scores for the optimal Coefficients policy with $\lambda = 0$.

Section 3, as utility UoS is defined in terms of expected performance Y. Generally speaking, however, we should note that large variance in Y could trigger political arguments in favor or against affirmative action policies in practice. On one hand, large variance might indicate that, instead of affirmative action, policy design should focus on the design of tests, the values X of which ought to be better predictors of future performance. On the other hand, values of variance that are too large compared to the expected performance of applicants might also support arguments in favor of affirmative action policies: if we cannot predict accurately who is going to be a good university student, then we should not rely too much on test scores, and instead we should provide more opportunities to applicants from disadvantaged backgrounds.

Going back to our data, the distributions for admission score with weights **c** are shown in Figure 6. It becomes obvious that, for the Coefficients policy that targets only **UoS**, high-income applicants have an advantage over low-income applicants (similarly for private-school applicants against public-school applicants), in the sense that they are admitted by larger proportions. Male applicants have an advantage over female applicants, but the advantage is smaller than in the case of income or school type, consistently with our earlier observations.

6.3 Affirmative action policies

Figure 7 shows the empirical performance of different affirmative action policies, each policy concerning a single sensitive attribute (school type is omitted due to similarity to the plot for income). To produce the plot, we employed Algorithm 1 to search for alternative Coefficients policies and Algorithm 2 to search for different Bonus policies. The plots also show a few indicative quota percentages for the corresponding Quota policy of the same attribute. The plots also depict the optimal Coefficients policy with $\lambda = 0$, the best Coefficients and Bonus (and equivalent Quota) policy with
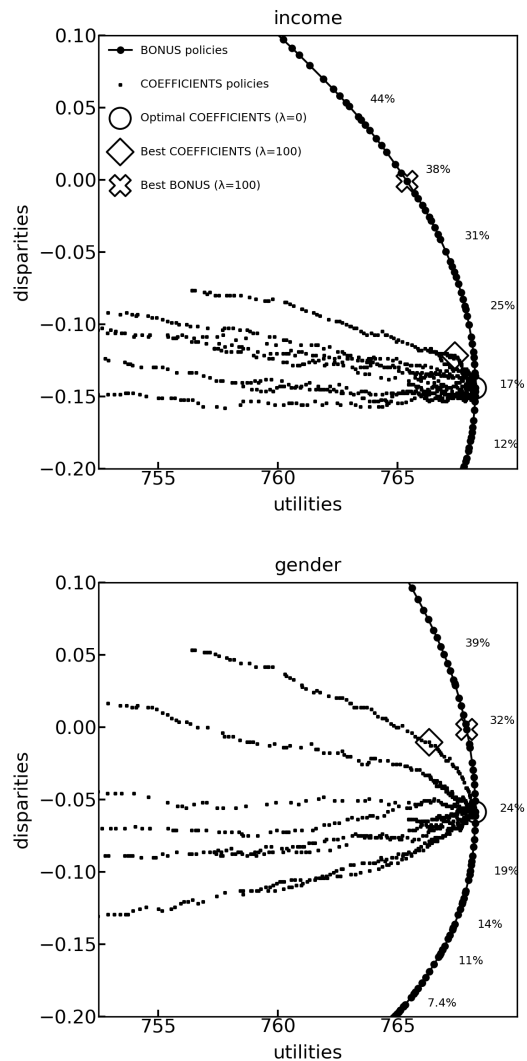

Figure 7: Comparison of Coefficients and Bonus policies. Percentages on the right are corresponding quotas for the disadvantaged groups under Quota policies matching sample Bonus policies.

$\lambda = 100$ discovered by the policy search algorithms 1-2, where λ is the parameter that dictates the trade-off between **UoS** and **DmD** in the optimization function (see Problem 1).

We make a few observations. (1) Based on its performance for income and school type, the Bonus and Quota policies provide better exploration of the trade-off between **UoS** and **DmD**: they lead to a better value of the optimization function (not shown in the plot) and policy instances with practically zero disparity **DmD**, with little loss in **UoS**. (2) The difference is less pronounced when gender is used as the sensitive attribute. This is explained by the fact that student grades show smaller discrepancy across genders than across income and school types (see Figure 3-4) (3) Consistently with the previous observation, the optimal Bonus policy uses larger
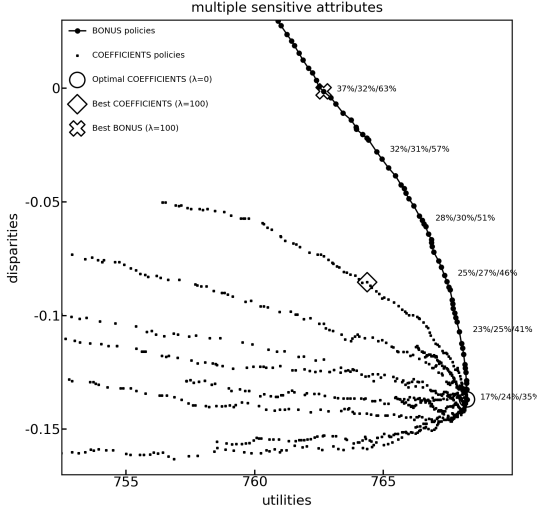
Figure 8: Performance of admission policies for multiple sensitive attributes. Percentages on the right are corresponding quotas for the disadvantaged groups (by income, gender, and school type) under QUOTA **policies matching sample** BONUS **policies.**

bonus value for income and school type (b = 30 and 35, respectively) than the one for gender (b = 9.4). To put these numbers in context, consider that the admission scores of the top applicant is about 833 points and of the last admitted applicant is in the range 722 - 733, depending on the attribute **A**. Hence, the gender bonus is negligible. (4) As expected, BONUS policies that lead to higher disparity (i.e., more favorable for the disadvantaged group) correspond to higher quota values for the corresponding QUOTA policy.

Figure 8 shows the results when all three attributes (income, school type and gender) are considered. Again, we mark the optimal COEFFICIENTS policy for $\{\lambda_i = 0\}$, as well as the optimal COEFFICIENTS and BONUS (and equivalent QUOTA) policy with $\{\lambda_i = 100\}$, where $\{\lambda_i\}$ are the parameters that dictate the trade-off between **UoS** and **DmD** in the optimization function (see Problem 2). Consistently with the case of single sensitive attributes, the optimal bonus values assigned to income and school type (b = 16 and 26, respectively) are larger than the one for gender (b = 7).

7 CONCLUSIONS

In this paper, we demonstrated how the problem of designing affirmative action policies can be addressed in an algorithmic framework. We explored three types of policies, i.e., COEFFICIENTS, BONUS, and QUOTA. We showed the equivalence of BONUS and QUOTA policies, and provided efficient search algorithms for the optimal parameters of COEFFICIENTS and BONUS policies. We employed these algorithms on real data for university admissions and obtained instances of admission policies that achieved good utility (performance of admitted applicants) while virtually eliminating disparity (i.e., reducing the difference between the proportions of admitted

applicants in advantaged and disadvantaged groups). Future steps include the application of the proposed algorithms on a wider range of real settings and the treatment of other measures of fairness.

REFERENCES

[1] Oscar Arias, Alejandra S Mizala, and Francisco Meneses. 2016. *Brecha de género en matemáticas: el sesgo de las pruebas competitivas (evidencia para Chile)*. Technical Report.
[2] Abolfazl Asudeh, HV Jagadish, Julia Stoyanovich, and Gautam Das. 2019. Designing fair ranking schemes. In *Proc. of SIGMOD*.
[3] Michael N Bastedo and Nicholas A Bowman. 2017. Improving admission of low-SES students at selective colleges. *Educational Researcher* 46, 2 (2017), 67–77.
[4] Michael N Bastedo, Nicholas A Bowman, Kristen M Glasener, and Jandi L Kelly. 2018. What are we talking about when we talk about holistic review? Selective college admissions and its effects on low-SES students. *Journal of Higher Education* 89, 5 (2018), 782–805.
[5] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2018. Ranking with fairness constraints. In *Proc. of ICALP*.
[6] Rodrigo Cornejo Chávez. 2006. El experimento educativo chileno 20 años después: una mirada crítica a los logros y falencias del sistema escolar. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio enEducación* 4, 1 (2006).
[7] Michele Donini, Luca Oneto, Shai Ben-David, John S Shawe-Taylor, and Massimiliano Pontil. 2018. Empirical risk minimization under fairness constraints. In *Advances in Neural Information Processing Systems*. 2791–2801.
[8] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. ACM, 214–226.
[9] Fernanda Estevan, Thomas Gall, and Louis-Philippe Morin. 2018. Redistribution without Distortion: Evidence from An Affirmative Action Programme At a Large Brazilian University. *The Economic Journal* 129, 619 (2018), 1182–1220.
[10] Andrew M Francis and Maria Tannuri-Pianto. 2015. Using Brazilı£¡s racial continuum to examine the short-term effects of affirmative action in higher education. *Journal of Human Resources* 47, 3 (2015), 754–784.
[11] Veronica Frisancho and Kala Krishna. 2016. Affirmative action in higher education in India: targeting, catch up, and mismatch. *Higher Education* 71, 5 (2016).
[12] Jennifer Glynn. 2019. *Persistence: The Success of Students Who Transfer from Community Colleges to Selective Four-Year Institutions*. Technical Report.
[13] Moritz Hardt, Eric Price, Nati Srebro, et al. 2016. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*.
[14] Lily Hu and Yiling Chen. 2018. A short-term intervention for long-term fairness in the labor market. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. 1389–1398.
[15] Sampath Kannan, Aaron Roth, and Juba Ziani. 2019. Downstream effects of affirmative action. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, 240–248.
[16] Michael Kearns, Aaron Roth, and Zhiwei Steven Wu. 2017. Meritocratic fairness for cross-population selection. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1828–1836.
[17] Niki Kilbertus, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. 2017. Avoiding discrimination through causal reasoning. In *Advances in Neural Information Processing Systems*. 656–666.
[18] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *Advances in Neural Information Processing Systems*. 4066–4076.
[19] Tomas Larroucau, Ignacio Rios, and Alejandra S Mizala. 2015. Efecto de la incorporación del ranking de notas en el proceso de admisión a las universidades chilenas. *Pensamiento Educativo* 52, 2 (2015), 95–118.
[20] Shira Mitchell, Eric Potash, and Solon Barocas. 2018. Prediction-based decisions and fairness: A catalogue of choices, assumptions, and definitions. *arXiv preprint arXiv:1811.07867* (2018).
[21] Hussein Mouzannar, Mesrob I Ohannessian, and Nathan Srebro. 2019. From Fair Decision Making To Social Equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM.
[22] Judea Pearl. 2009. *Causality*. Cambridge university press.
[23] Judea Pearl and Dana Mackenzie. 2018. *The book of why: the new science of cause and effect*. Basic Books, New York.
[24] Julie Posselt, Ozan Jaquette, Rob Bielby, and Michael N Bastedo. 2012. Access without equity. *American Educational Research Journal* 49, 6 (2012), 1074–1111.
[25] John E Roemer. 1998. *Equality of opportunity*. Harvard University Press.
[26] Julia Stoyanovich, Ke Yang, and HV Jagadish. 2018. Online set selection with fairness and diversity constraints. In *Proceedings of the EDBT Conference*.
[27] Linda F Wightman. 1998. LSAC National Longitudinal Bar Passage Study. LSAC Research Report Series. (1998).
[28] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 1569–1578.