Statistical Discrimination in Stable Matchings

Rémi Castera, Patrick Loiseau, Bary Pradelski

Statistical Discrimination in Stable Matchings

Rémi Castera [*,1], Patrick Loiseau[1], and Bary S.R. Pradelski[1]

[1]Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, Grenoble, France

April 2022

Abstract

Statistical discrimination results when a decision-maker observes an imperfect estimate of the quality of each candidate dependent on which demographic group they belong to. Prior literature is limited to simple selection problems with a single decision-maker. In this paper, we initiate the study of statistical discrimination in matching, where multiple decision-makers are simultaneously facing selection problems from the same pool of candidates (e.g., colleges admitting students). We propose a model where two colleges observe noisy estimates of each candidate's quality. The estimation noise controls a new key feature of the problem, namely the correlation between the estimates of the two colleges: if the noise is high, the correlation is low and vice-versa. We consider stable matchings in an infinite population of students. We show that a lower correlation (i.e., higher estimation noise) for one of the groups worsens the outcome for all groups. Further, the probability that a candidate is assigned to their first choice is independent of their group. In contrast, the probability that a candidate is assigned to a college at all depends on their group, revealing the presence of discrimination coming from the correlation effect alone. Somewhat counter-intuitively the group that is subjected to more noise is better off.

1 Introduction

Discrimination in matching problems such as college admission has been the subject of frequent and continued controversy over the past decades. On the one hand, there are myriads of news articles and research papers reporting empirical observations of discrimination in the outcomes of various college admission procedures [Gersen 2019, Cortes 2019, Bonneau et al. 2021]—that is, the fact that certain demographic groups defined by sensitive attributes such as race or gender receive less favorable outcomes. On the other hand, preventive measures such as affirmative actions are the subject of heated debates in the context of college admission in many countries. In the US for

[*]Corresponding author: remi.castera@protonmail.com

instance, they are banned in several states and they are at the center of an upcoming case in the supreme court Liptak and Hartocollis (2022); and other countries have similar debates Vairet (2021). A common thread that emerges from those debates is that the precise causes of discrimination in matching problems are not well understood and clearly identified.

To explain discrimination, economic theory distinguishes between *taste-based discrimination* and *statistical discrimination*. While the former refers to actual preferences being based on demographic groups, the latter posits that discrimination may occur from the imperfect information a decision-maker has about individuals' qualities. Statistical discrimination theory was originally proposed by Phelps (1972) and Arrow (1973) in the context of employment (where the decision-maker is an employer attempting to assess the productivity of workers) to explain racial disparities (in particular in wages); and it was subsequently extended and refined in the economic literature, see e.g., Aigner and Cain (1977); Lundberg and Startz (1983); Coate and Loury (1993) or a survey in Fang and Moro (2011). Recently, the computer-science community proposed a similar model in the context of selection problems, under the term *differential variance* Emelianov et al. (2020, 2022); Garg et al. (2021). In this model, a decision-maker has a noisy estimate of the quality of candidates, with a noise variance that depends on the demographic group of the candidate. Emelianov et al. (2022) and Garg et al. (2021) show that it leads to some groups being underrepresented and study the effect of imposing fairness constraints and/or standardized tests on the selection outcome.

The aforementioned works on statistical discrimination and differential variance are limited to the selection setting with a single decision-maker, and hence a single quality estimate for each candidate. In a matching problem, there are multiple decision-makers (say, the colleges) who each have their own imperfect estimate of the candidates' qualities. The noise therefore has a new effect, beyond the possible group-dependent variance. For a given candidate, the quality estimates of different colleges are the same if the noise is zero, i.e., the colleges perfectly observe the candidates quality. Generally, quality estimates may be more or less *correlated* depending on the amount of noise: a small noise would lead to a high correlation of estimates between the different colleges, whereas a high noise would lead to a low correlation (assuming the noise is independent for each college). How statistical discrimination, modeled via the described noise and correlation process, affects matching outcomes for different groups of students is, however, a non-trivial and totally open question, which we address in this work.

1.1 Our contribution

In this paper, we initiate the study of *statistical discrimination in matching*. We consider a setting with a population of students divided in two groups G_1 and G_2, applying to two colleges $\mathbf{A}$ and $\mathbf{B}$. We first propose a model where the colleges get noisy estimates of each candidate's quality $W_{\mathbf{A}}$ and $W_{\mathbf{B}}$ such that the vector $(W_{\mathbf{A}}, W_{\mathbf{B}})$ is a bivariate normal random variable with correlation coefficient ρ. This model flexibly describes the information on which the matching is based. In particular, it describes the case where the quality estimates at each college are equal to a latent quality of the candidate plus a centered measurement noise drawn independently for each college (all being normal). If the measurement noise is large, ρ will be close to zero as the quality estimates at each college are close to independent. On the other hand, if the measurement noise is small, then the estimate is almost equal to the latent quality, hence both colleges have almost identical estimates and ρ is close to one. The range of values of $\rho \in (0,1)$ parameterizes all intermediate situations. To isolate the effect of correlation (which is specific to matching), we focus on the case where the marginals of $W_{\mathbf{A}}$ and $W_{\mathbf{B}}$ have the same variance, but the correlation coefficient depends on the demographic group of the candidate (ρ_{G_1} for group G_1 and ρ_{G_2} for group G_2)—we term this *differential correlation.*

We focus on characterizing the outcome of (centralized) stable matchings. Such procedures are used in increasingly many countries such as France, Brazil, or Hungary Parcoursup (2021); Machado and Szerman (2021); Biró (2008) in college admission or other education markets (e.g., school assignment).[1] For simplicity, we consider an infinite population of students whose preferences are simply captured by the proportion of students of each group that prefer college **A** to **B**, and use techniques from Azevedo and Leshno (2016) to characterize the stable matching (which is found to be unique in this model with a continuum of students). Then, we are able to compute welfare metrics such as the probability that a student gets their first or second choice or remains unmatched; and we have the following main results:

1. We show that the probability that a student is assigned to their first choice is independent of the student's group, but that it decreases when the correlation of either group decreases. This means that higher measurement noise (inducing lower correlation) on one group hurts not only the students of that group, but the students of all groups.

2. We show that the probability that a student is assigned to their second choice and the probability that they remain unassigned both depend on the student's group, which reveals the presence of discrimination coming from the correlation effect alone. Specifically, we find that the probability that a student remains unmatched is decreasing when the correlation of their group decreases (higher measurement noise) and when the correlation of the other group increases. In other words, the higher the measurement noise of their own group, the better off students are with regard to getting assigned a college at all. This may sound counter-intuitive at first, but is explained by the observation that with high noise (i.e., low correlation) the fact that a student is rejected from one college gives only little information about the outcome at the other college. That is, a student has an independent second chance for admission.

These two comparative static results give insights on the effect of correlation on the stable matching outcome for different demographic groups and show that indeed, statistical discrimination is an important theory to understand discrimination in matching problems. We also analyze a

[1] See `https://www.matching-in-practice.eu/` for more examples in the European Union.

number of special cases of our model, in particular the case of a single group, to show that even in this case the correlation affects the stable matching outcome.

Our work is the first to investigate statistical discrimination in the context of matching. Overall it shows that group-dependent measurement noises of the candidates quality—and the resulting group-dependent correlation between the colleges' estimates—plays an important role in leading to unequal outcomes for different demographic groups, and in particular underrepresentation of one of the groups. Of course, we do not argue that statistical discrimination is the only possible cause of discrimination. In particular, if there is bias in the quality estimates for one group, then it will naturally also hurt the representation of that group. We do not model bias since our primary purpose is to isolate in the cleanest possible way the effect of statistical discrimination. Throughout the paper, we make a number of other simplifying assumptions (e.g., focusing on two colleges) whose purpose is also to simplify our results and isolate the effect of correlation. Our analysis, however, can be extended to remove these assumptions—we discuss these extensions in Section 5.

Throughout the paper, we discuss the matching problem of college admission, and use the terminology of *grade* to denote the estimate that a college has about the quality of a candidate. Our model and results, however, are generic and apply to any matching problem where there is a notion of quality of candidates on one side that can be represented by a numeric score. This includes for instance public school choice problems, assignments in hospital residency programs, or labor markets (see e.g., examples in Abdulkadiroğlu (2005)).[2]

1.2 Related work

Matching The matching literature is today a cornerstone of the increasing connection between economic theory and computer science. The college admission problem, i.e., how a centralized authority can fairly assign prospective students to colleges given each agent's preferences and capacity constraints of colleges, was first modeled and studied by Gale and Shapley (1962). It was then followed by many other works including for instance Roth (1986), and Abdulkadiroğlu et al. (2009), see also Roth and Sotomayor (1992) for the standard text book.[3] The idea of considering a continuum of students and a finite number of colleges has previously been exploited in Chade et al. (2014) and Azevedo and Leshno (2016). In particular, we shall follow Azevedo and Leshno (2016) in analyzing the continuum model as a supply and demand problem.

Matching with incomplete information In our model, we assume that the colleges have imperfect (noisy) information about their preferences. Similar assumptions are made in Chade et al. (2014) and Azevedo and Leshno (2016). There are also works on matching under incomplete

[2]We assume that quality can be mapped to a mono-dimensional grade and that all colleges have the same ordinal preferences.

[3]For the variant of the model commonly called *school choice problem*, see [Balinski and Sonmez 1999, Abdulkadiroğlu and Sönmez 2003, Abdulkadiroğlu 2005, Ergin and Sönmez 2006, Yenmez 2013].

information, studying various models. Rastegari et al. (2013) study matching under partially ordered preferences, Aziz et al. (2020) assume a known distribution of preference profiles, and in Immorlica et al. (2020); Liu et al. (2020) additional information can be acquired. Some models, e.g., in Liu et al. (2014), Kloosterman and Troyan (2020), and Liu (2020), suppose that information is partial only for one side (colleges or students). Finally, Chakraborty et al. (2010) and Bikhchandani (2017) show that a matching based on incomplete information can be stable if agents do not know to the full matching. We do not include incomplete information considerations in our study. In Ashlagi et al. (2019), it is assumed that colleges' rankings contain ties, that are broken using a lottery. The authors then compare the welfare of students in two settings: either one common lottery is used by all colleges, or all colleges draw an independent lottery. The notion of correlation between colleges' signals and its influence on students' welfare is quite similar to the results we provide in this paper, and the results of both papers point in the same direction, even though the models are quite different. A recent paper by Arnosti (2022) study a very similar model of tie-breaking, that allows students to choose the size of their preference list and derive the influence of the tie-breaking rule on students' welfare depending on the size of their preference list.

Fairness Statistical discrimination in selection problems was first studied by Kleinberg and Raghavan (2018), followed by Emelianov et al. (2020, 2022) and Garg et al. (2021). They suppose that candidates have a latent quality, and that the college or company they apply to only has access to a biased and/or noisy estimator of this quality. We depart from their models by considering several colleges instead of one—that is, we consider the matching problem instead of the selection problem. Works on fairness in matching have considered various affirmative action policies, including upper and lower quotas, to reduce discrimination [Abdulkadiroğlu 2005, Kamada and Kojima 2015, Delacrétaz et al. 2020, Kamada and Kojima 2022, Krishnaa et al. 2019]. These works, however, focus on finding stable matchings under distributional constraints representing fairness notions; in contrast, our work intends to explain discrimination that naturally occurs in stable matchings without constraints. Chade et al. (2014) study a model with application costs for students but without group-dependent variance which is at the center of our work. Recently, reducing discrimination in ranking rather than in the final selection has been an emerging way of pursuing fairness [Celis et al. 2020, Yang et al. 2021, Zehlike et al. 2021]. However, Karni et al. (2021) show that fairness of the ranking does not imply fairness of the matching. Finally, Monachou and Ashlagi (2019) and Che et al. (2020) both study discrimination in online markets using ratings based on reviews.[4]

[4]For the general question of bias and fairness in algorithms and machine learning, see [Dwork et al. 2012, Hardt et al. 2016, Zafar et al. 2017, Blum and Stangl 2020, Barocas et al. 2019, Chouldechova and Roth 2020, Finocchiaro et al. 2021, Kleinberg and Raghavan 2021, Mehrabi et al. 2021].

1.3 Outline

The remainder of this paper is organized as follows. Section 2 presents the model, the matching mechanism (deferred acceptance for a continuum of students), and the supply and demand framework for matching. In Section 3 we first introduce welfare metrics to evaluate different outcomes and then present our main results. Section 4 treats the special cases of capacity excess and a single group, thus allowing to build further intuition. Finally, Section 5 concludes with a discussion on the generality of our findings and future avenues of research. To improve the flow and readability of the paper, we only include essential elements of the proofs of our results in the body of the paper and defer longer and more technical details to an appendix.

2 Model and preliminaries

Most models of statistical discrimination consider that students have a latent quality, and that the college only sees a noisy and biased estimate of this quality.[5] They also suppose that the population is partitioned into groups with different bias and variance of estimation. As a consequence, the distribution of estimated qualities is different for each group. In a matching context, i.e., with several colleges, it would also lead to different levels of correlation between the estimate of the same student by each college: without noise the correlation would be 1, but with very high noise it would tend to 0. We aim to isolate the effect of this difference of correlation from the effects of variance and bias. This section defines a model that achieves this goal by equalizing the marginal distributions of estimations for each group at each college, leaving only a difference in correlation.

2.1 Students and colleges

Let there be two colleges, $\mathbf{A}$ and $\mathbf{B}$, which are to be matched with a continuum unit mass of students $\mathbf{S}$. Colleges cannot admit more than a certain mass of students, their respective capacities are $\alpha_{\mathbf{A}}$ and $\alpha_{\mathbf{B}}$, both in $(0, 1]$. The students are divided into two groups, denoted G_1 and G_2, a proportion $\gamma \in [0, 1]$ of students belonging to G_1 and $1 - \gamma$ belonging to G_2. The group to which a student $s \in \mathbf{S}$ belongs is denoted by $G(s)$.

Students have a preference over colleges: among group G_1, a proportion $\beta_{G_1} \in [0, 1]$ prefers college $\mathbf{A}$ to college $\mathbf{B}$, the remaining $1 - \beta_{G_1}$ preferring $\mathbf{B}$, similarly in G_2 a proportion $\beta_{G_2} \in [0, 1]$ prefers $\mathbf{A}$ and $1 - \beta_{G_2}$ prefer $\mathbf{B}$. When student s prefers college $\mathbf{A}$ to college $\mathbf{B}$, we write $\mathbf{A} \succ_s \mathbf{B}$, and vice versa. All students prefer attending any college rather than remaining unmatched. Preferences are assumed to be strict. Note that these quantities are proportions and not masses; for instance, the subset of students that are in group G_1 and prefer college $\mathbf{A}$ is therefore of measure $\gamma\beta_{G_1}$.

[5] see Section 1.2

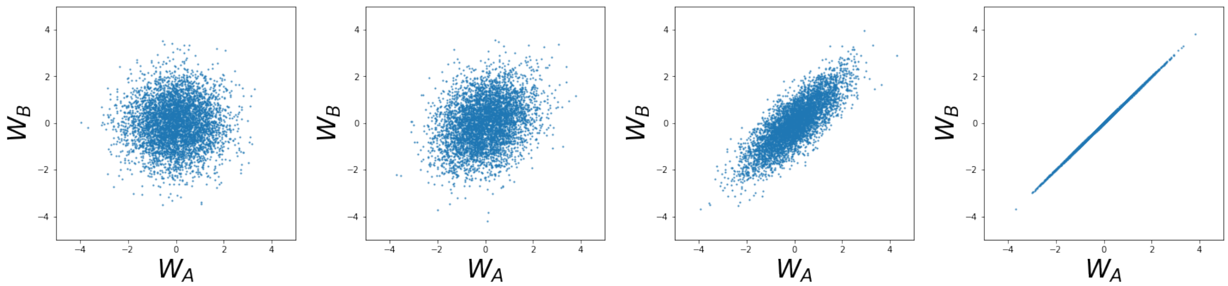


Figure 1: Grade distributions for $\rho = 0, 0.3, 0.8$ and 1.

Colleges preferences over students are based on grades and each college assigns grades according to a standard normal distribution. If both colleges are able to measure a student's quality accurately, they will assign the same grade to them. On the other hand, if one (or both) of the colleges is not able to assess a student's quality at all the two respective grades will be uncorrelated. More generally, given a student, the correlation of the grades they receive from each college are an indication of the (joint) capacity to evaluate a student's quality.

Colleges have variable accuracies in evaluating students, depending on whether they belong to group G_1 or group G_2. Grades of students from G_1 have a correlation of ρ_{G_1}, grades of students from G_2 have a correlation of ρ_{G_2}. We call this feature of the model *differential correlation*, linked to the notion of differential variance studied in Emelianov et al. (2022) and Garg et al. (2021). Formally, for each student $s \in \mathbf{S}$, we assume that their grades at colleges $\mathbf{A}$ and $\mathbf{B}$ form a vector $(W^s_{\mathbf{A}}, W^s_{\mathbf{B}})$ drawn randomly following the normal bivariate distribution with mean $(0,0)$ and covariance matrix

$$\begin{pmatrix} 1 & \rho_{G(s)} \\ \rho_{G(s)} & 1 \end{pmatrix}.$$

The obtained grades thus have a correlation of $\rho_{G(s)}$, and the marginals, i.e., the grade distribution of each college, remain standard normal distributions. Figure 1 displays the distributions obtained for various values of the correlation. The preference of college $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$ over students is then: $\mathbf{C}$ prefers $s \in \mathbf{S}$ to $s' \in \mathbf{S}$ if and only if $W^s_{\mathbf{C}} > W^{s'}_{\mathbf{C}}$.

From now on, $\gamma, \beta_{G_1}, \beta_{G_2}, \alpha_{\mathbf{A}}, \alpha_{\mathbf{B}}, \rho_{G_1}$ and ρ_{G_2} will be referred to as the *parameters* of a given college admission problem.

Remark. *The choice of the normal bivariate distribution stems from the following observation. In order to isolate the effect of noise in a matching setting it is critical that the considered noise process does not lead to bias, that is, does not change the expectation or any marginal of the resulting grade distribution. For concreteness, consider the model studied by Emelianov et al. (2022) or Garg et al. (2021) where the score is a sum of a latent quality and a normally distributed noise term. In this* *model, a group whose noise variance is higher would also have an overall higher grade variance, and thus the students of this group would be overrepresented among the top grades. Our model could encompass this one, by allowing the covariance matrix's diagonal terms to differ from one group to another (since the sum of two normal random variables is normal too). However, in this paper we choose to only allow the correlation to differ in order to isolate its effect.*

We now formally define the notion of mass for a subset of students. The following definitions may be omitted to understand the model and the results, but are required in their proofs. We identify $\mathbf{S}$ to $\Theta = \mathbb{R}^2 \times \{G_1, G_2\} \times \{\mathbf{A}, \mathbf{B}\}$. We partition Θ into 4 subsets: $\Theta_{G,\mathbf{C}} = \{s \in \Theta : s = ((x, y), G, \mathbf{C}), x, y \in \mathbb{R}\}$ is the subset of students belonging to group G and preferring college $\mathbf{C}$. A subset $I \subseteq \Theta$ is measurable if and only if $\{(W^s_{\mathbf{A}}, W^s_{\mathbf{B}}) : s \in I\}$ is for the Borel σ-algebra of $\mathbb{R}^2$. We can partition I into 4 subsets $I_{G,\mathbf{C}} = I \cap \Theta_{G,\mathbf{C}}$. On each $\Theta_{G,\mathbf{C}}$ we define a measure $\eta_{G,\mathbf{C}}$ as follows: for $I \subseteq \Theta$ measurable,

$$\begin{aligned}
\eta_{G_1,\mathbf{A}}(I_{G_1,\mathbf{A}}) &= \gamma\beta_{G_1}\mathbb{P}_{\rho_{G_1}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W^s_{\mathbf{A}}, W^s_{\mathbf{B}}) : s \in I_{G_1,\mathbf{A}}\}),\\
\eta_{G_1,\mathbf{B}}(I_{G_1,\mathbf{B}}) &= \gamma(1-\beta_{G_1})\mathbb{P}_{\rho_{G_1}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W^s_{\mathbf{A}}, W^s_{\mathbf{B}}) : s \in I_{G_1,\mathbf{B}}\}),\\
\eta_{G_2,\mathbf{A}}(I_{G_2,\mathbf{A}}) &= (1-\gamma)\beta_{G_2}\mathbb{P}_{\rho_{G_2}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W^s_{\mathbf{A}}, W^s_{\mathbf{B}}) : s \in I_{G_2,\mathbf{A}}\}),\\
\eta_{G_2,\mathbf{B}}(I_{G_2,\mathbf{B}}) &= (1-\gamma)(1-\beta_{G_2})\mathbb{P}_{\rho_{G_2}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W^s_{\mathbf{A}}, W^s_{\mathbf{B}}) : s \in I_{G_2,\mathbf{B}}\}),
\end{aligned} \tag{1}$$

where $\mathbb{P}_\rho$ is the probability measure associated to the bivariate normal distribution with correlation ρ. We define over Θ the probability measure $\eta : \mathbf{S} \to [0, 1]$ such that for any measurable subset I of $\mathbf{S}$,

$$\eta(I) = \eta_{G_1,\mathbf{A}}(I_{G_1,\mathbf{A}}) + \eta_{G_1,\mathbf{B}}(I_{G_1,\mathbf{B}}) + \eta_{G_2,\mathbf{A}}(I_{G_2,\mathbf{A}}) + \eta_{G_2,\mathbf{B}}(I_{G_2,\mathbf{B}}). \tag{2}$$

This definition is consistent with the definition of the parameters, as it verifies $\eta(G_1) = \gamma$, $\eta(\{s \in G_1 : \mathbf{A} \succ_s \mathbf{B}\})/\gamma = \beta_{G_1}$ and $\eta(\{s \in G_2 : \mathbf{A} \succ_s \mathbf{B}\})/(1-\gamma) = \beta_{G_2}$.

2.2 Matching mechanism

To define matching in a continuum context, we follow Azevedo and Leshno (2016).

Definition 2.1. A matching is an assignment of students to colleges, described by a mapping $\mu : \mathbf{S} \cup \{\mathbf{A}, \mathbf{B}\} \to \mathscr{P}(\mathbf{S}) \cup \mathbf{C} \cup \mathbf{S}$, with the following properties:

1. for all $s \in \mathbf{S}$, $\mu(s) \in \{\mathbf{A}, \mathbf{B}\} \cup \{s\}$;
2. for $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$, $\mu(\mathbf{C}) \subseteq \mathbf{S}$ is measurable and $\eta(\mu(\mathbf{C})) \leq \alpha_{\mathbf{C}}$;
3. $\mathbf{C} = \mu(s)$ if and only if $s \in \mu(\mathbf{C})$;
4. for $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$, the set $\{s \in \mathbf{S} : \mu(s) \preceq_s \mathbf{C}\}$ is open.

The first three conditions are common to almost all definitions of matching in discrete or continuous models. Condition (1) ensures that a student is either matched to a college or to themselves, which means that they remain unmatched. Condition (2) ensures that colleges are assigned to a subset of students that respects the capacity constraint. Condition (3) ensures that the matching is consistent, i.e., if a student is matched to a college, then this college is also matched to the student. Condition (4) was introduced by Azevedo and Leshno (2016) and is necessary to ensure there does not exist several stable matchings that only differ by a set of measure 0.

We next define the notions of blocking and stability.

Definition 2.2 (Blocking)**.** The pair $(s, \mathbf{C})$ *blocks* a matching μ if s would prefers $\mathbf{C}$ to her current match, and either $\mathbf{C}$ has remaining capacity or it admitted a student with a lower score than s. Formally, if $\mu(s) \prec_s \mathbf{C}$ and either $\eta(\mu(\mathbf{C})) < \alpha_{\mathbf{C}}$ or $\exists s' \in \mu(\mathbf{C})$ such that $W_{\mathbf{C}}^{s'} < W_{\mathbf{C}}^{s}$.

Definition 2.3 (Stability)**.** A matching is *stable* if it is not blocked by any student-college pair.

To produce a stable matching, one can then extend the classic Deferred Acceptance algorithm from Gale and Shapley (1962) to the continuum model. This algorithm is described in Algorithm 1.

Algorithm 1 Deferred Acceptance Algorithm

First step: All students apply to their favorite college, they are temporarily accepted. If the mass of students applying to college $\mathbf{C}$ is greater than its capacity $\alpha_{\mathbf{C}}$, then $\mathbf{C}$ only keeps the $\alpha_{\mathbf{C}}$ best

while A positive mass of students are unmatched and have not yet been rejected from every college **do**

  Each student who has been rejected at the previous step proposes to her preferred college among those which have not rejected her yet

  Each college $\mathbf{C}$ keeps the best $\alpha_{\mathbf{C}}$ mass of students among those it had temporarily accepted and those who just applied, and rejects the others

end while

End: If the mass of students that are either matched or rejected from every college is 1, the algorithm stops. However it could happen that it takes an infinite number of steps to converge.

If the algorithm stops, the matching it outputs is stable; Abdulkadiroğlu et al. (2015) show that even when the number of steps is infinite, the algorithm converges to a stable matching.

2.3 Supply and Demand

In a model with infinitely many students, a matching problem can be alternatively viewed as a supply and demand setup, where a stable matching is a Walrasian equilibrium Azevedo and Leshno (2016).

Definition 2.4 (Cutoffs and demand)**.** If μ is a stable matching, define the *cutoffs* at $\mathbf{A}$ and $\mathbf{B}$ as the quantities $P_{\mathbf{A}} := \inf\{W^s_{\mathbf{A}} : \mu(s) = \mathbf{A}\}$ and $P_{\mathbf{B}} := \inf\{W^s_{\mathbf{B}} : \mu(s) = \mathbf{B}\}$. The cutoff of a college represents the grade above which a student who applies gets admitted.
Given $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$, we say that student s' *demand* $D_s(P_{\mathbf{A}}, P_{\mathbf{B}})$ is the college they prefer among those where they pass the cutoff, or themselves if they does not pass the cutoff at any college.
The *aggregate demand* at college $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$ is the mass of students demanding it: $D_{\mathbf{C}}(P_{\mathbf{A}}, P_{\mathbf{B}}) = \eta(\{s : D_s(P_{\mathbf{A}}, P_{\mathbf{B}}) = \mathbf{C}\}$.

In our model, the "supply" associated to this demand is simply the capacity of each college. Now consider the equilibria of this problem:

Definition 2.5 (Market clearing)**.** The cutoffs are *market clearing* if for $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$, $D_{\mathbf{C}}(P_{\mathbf{A}}, P_{\mathbf{B}}) \leq \alpha_{\mathbf{C}}$, with equality if $P_{\mathbf{C}} \neq -\infty$.

A cutoff vector is therefore market clearing if it induces a demand that is equal to colleges capacities when they reach their capacity constraint, and lower for colleges that are not full. When $P_{\mathbf{A}}, P_{\mathbf{B}} \neq -\infty$, the system of equations

$$\begin{cases} D_{\mathbf{A}}(P_{\mathbf{A}}, P_{\mathbf{B}}) & = \alpha_{\mathbf{A}} \\ D_{\mathbf{B}}(P_{\mathbf{A}}, P_{\mathbf{B}}) & = \alpha_{\mathbf{B}} \end{cases} \tag{3}$$

is called the *market clearing equations*, and the cutoffs $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ can be computed by solving the system.

The reason to introduce this notion is that there is a one to one correspondence between stable matchings and market clearing cutoffs.

Lemma 2.6 ((Azevedo and Leshno, 2016, Lemma 1))**.**

1. *If μ is a stable matching, the associated cutoffs $P_{\boldsymbol{A}}$ ad $P_{\boldsymbol{B}}$ are market clearing;*
2. *If $P_{\boldsymbol{A}}$ ad $P_{\boldsymbol{B}}$ are market clearing cutoffs, we define μ such that for all $s \in \boldsymbol{S}$, $\mu(s) = D_s(P_{\boldsymbol{A}}, P_{\boldsymbol{B}})$. Then μ is stable.*

We can therefore study stable matchings by studying the cutoffs of each college, which will be critical to prove our results. The theorem that follows will be of great use:

Theorem 2.7 (Special case of (Azevedo and Leshno, 2016, Theorem 1))**.** *For any college admission problem (defined by its parameters $\gamma, \beta_{G_1}, \beta_{G_2}, \alpha_{\boldsymbol{A}}, \alpha_{\boldsymbol{B}}, \rho_{G_2}$ and ρ_{G_2}), there exists a unique stable matching.*

The original theorem specifies conditions on the distribution of grades, such as being continuous and having full support, which our model verifies. Unlike the finite case where several stable matchings typically coexist, in the continuum model the stable matching is unique and therefore no considerations regarding selection among the set of stable matchings are necessary. From now on, we will therefore consider the cutoffs $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ as the ones uniquely determined by the parameters of the problem and the market clearing equation. We shall say "student s goes to college $\mathbf{C}$" to mean that they are matched in the unique stable matching.

Finally, Azevedo and Leshno (2016, Theorem 2) show that the stable matching varies continuously in the parameters of the problem. Furthermore, they prove that the set of stable matchings from a college admission problem with a finite number of students converges to the unique stable matching of the continuum problem with same parameters. This tends to indicate that our results are still relevant when the number of students is finite but large.

3 Analysis of the Model

This section contains our main results on the impact of noise—and thus correlation between grades—on the quality of the (unique) stable matching.

3.1 Welfare metrics

To measure students' satisfaction, we consider an individual's likelihood of getting their first choice.

Definition 3.1 (Welfare metrics)**.** For correlation levels ρ_{G_1} and ρ_{G_2} associated to groups G_1 and G_2, we define $V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2})$, $V_1^{G_1,\mathbf{B}}(\rho_{G_1},\rho_{G_2})$, $V_1^{G_2,\mathbf{A}}(\rho_{G_1},\rho_{G_2})$ and $V_1^{G_2,\mathbf{B}}(\rho_{G_1},\rho_{G_2})$ as the proportion of students from each group-preference profile who get their first choice. Equivalently, it is the probability of a student to get their first choice conditionally on their profile. Formally,

$$
\begin{aligned}
V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{\gamma\beta_{G_1}}\eta(\{s \in G_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\}),\\
V_1^{G_1,\mathbf{B}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{\gamma(1-\beta_{G_1})}\eta(\{s \in G_1 : \mathbf{B} \succ_s \mathbf{A}, \mu(s) = \mathbf{B}\}),\\
V_1^{G_2,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{(1-\gamma)\beta_{G_2}}\eta(\{s \in G_2 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\}),\\
V_1^{G_2,\mathbf{B}}(\rho_{G_1},\rho_{G_2}) &:= \frac{1}{(1-\gamma)(1-\beta_{G_2})}\eta(\{s \in G_2 : \mathbf{B} \succ_s \mathbf{A}, \mu(s) = \mathbf{B}\}).
\end{aligned}
$$

Consider $V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2})$; it is the probability of a student from group G_1 who prefers $\mathbf{A}$ to $\mathbf{B}$ to get their first choice, i.e., $\mathbf{A}$. We also define the following aggregated quantities:

$$
\begin{aligned}
V_1^{G_1}(\rho_{G_1},\rho_{G_2}) &:= \beta_{G_1}V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) + (1-\beta_{G_1})V_1^{G_1,\mathbf{B}},\\
V_1^{G_2}(\rho_{G_1},\rho_{G_2}) &:= \beta_{G_2}V_1^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) + (1-\beta_{G_2})V_1^{G_1,\mathbf{B}},
\end{aligned}
$$

which represent the probability of a student getting their first choice conditionally on belonging to group G_1 or G_2 respectively.

We define similarly the proportions of students getting their second choice or staying unmatched, using respectively V_2 and $V_\emptyset$ instead of V_1. For instance, $V_2^{G_1,\mathbf{A}}(\rho_{G_1}, \rho_{G_2})$ is the probability of a student from group G_1 who prefers college **A** to get their second choice, i.e., **B**. We now provide expressions for these metrics as functions of $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$.

Lemma 3.2. *Given the cutoffs $P_{\boldsymbol{A}}$, $P_{\boldsymbol{B}}$, we can compute the quantities V_1, V_2 and $V_\emptyset$ as follows: Let $\boldsymbol{C} \in \{\boldsymbol{A}, \boldsymbol{B}\}$ be a college and let $\overline{\boldsymbol{C}}$ be the other college, let $G \in \{G_1, G_2\}$. Then:*

$$V_1^{G,\boldsymbol{C}}(\rho_{G_1}, \rho_{G_2}) = \mathbb{P}_{\rho_G}(W_{\boldsymbol{C}} \geq P_{\boldsymbol{C}}) = \int_{P_{\boldsymbol{C}}}^{\infty} \phi(x)\,\mathrm{d}x, \tag{4}$$

$$V_2^{G,\boldsymbol{C}}(\rho_{G_1}, \rho_{G_2}) = \mathbb{P}_{\rho_G}(W_{\boldsymbol{C}} < P_{\boldsymbol{C}}, W_{\overline{\boldsymbol{C}}} \geq P_{\overline{\boldsymbol{C}}}) = \int_{-\infty}^{P_{\boldsymbol{C}}} \int_{P_{\overline{\boldsymbol{C}}}}^{\infty} \phi(x, y, \rho_G)\,\mathrm{d}x\mathrm{d}y, \tag{5}$$

$$V_\emptyset^{G,\boldsymbol{C}}(\rho_{G_1}, \rho_{G_2}) = \mathbb{P}_{\rho_G}(W_{\boldsymbol{C}} < P_{\boldsymbol{C}}, W_{\overline{\boldsymbol{C}}} < P_{\overline{\boldsymbol{C}}}) = \int_{-\infty}^{P_{\boldsymbol{C}}} \int_{-\infty}^{P_{\overline{\boldsymbol{C}}}} \phi(x, y, \rho_G)\,\mathrm{d}x\mathrm{d}y, \tag{6}$$

where $\phi(x)$ is the probability density function of the standard normal distribution and $\phi(x, y, \rho)$ is the one of the standard normal bivariate distribution with correlation ρ.

Sketch of Proof. The proof mainly consists in writing the metrics V_1, V_2 and $V_\emptyset$ in terms of the measure $\eta(\cdot)$ of an appropriate set of students, and identifying it with the distribution of the corresponding set of grades. The details are provided in Appendix A.1. ■

This lemma will allow us to compare chances of admission of different types of students, and derive comparative statics regarding the differential correlation.

3.2 Impact of differential correlation

Recall that groups are differentiated by the correlation of their grades between the two colleges. We shall next consider the chances of admission dependent on a student's group.

Theorem 3.3. *The probability that a student gets their first choice is independent from the group they belong to. Formally, for any γ, β_{G_1}, β_{G_2}, $\alpha_{\boldsymbol{A}}$, $\alpha_{\boldsymbol{B}}$, ρ_{G_1}, $\rho_{G_2} \in [0, 1]$, $V_1^{G_1,\boldsymbol{A}}(\rho_{G_1}, \rho_{G_2}) = V_1^{G_2,\boldsymbol{A}}(\rho_{G_1}, \rho_{G_2})$ and $V_1^{G_1,\boldsymbol{B}}(\rho_{G_1}, \rho_{G_2}) = V_1^{G_2,\boldsymbol{B}}(\rho_{G_1}, \rho_{G_2})$. If $\beta_{G_1} = \beta_{G_2}$, then $V_1^{G_1}(\rho_{G_1}, \rho_{G_2}) = V_1^{G_2}(\rho_{G_1}, \rho_{G_2})$.*

Proof. The result follows directly by applying (4) from Lemma 3.2 to both groups, and by observing that the integral in (4) does not depend on the correlation coefficients and hence does not depend on the group in our model. ■

This result, albeit quite simple, is an unexpected property of the model. A student, given the college they prefer, have the same chances of getting admitted there whatever group they belong to. Notice that $V_1^{G_1}(\rho_{G_1}, \rho_{G_2})$ might not be equal to $V_1^{G_2}(\rho_{G_1}, \rho_{G_2})$, but the difference would only be due to different proportions of students preferring each college, not to the differential correlation.

From now and until the end of Section 3, we assume $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. This condition is further discussed in Section 4.1.

In order to further analyze the model, we will make use of the following result by Sibuya (1960):

Lemma 3.4 (Appendix of Sibuya (1960))**.** *Let $\Phi(x, y, \rho) = \mathbb{P}_\rho(W_{\boldsymbol{A}} \leq x, W_{\boldsymbol{B}} \leq y)$ the cumulative distribution function of the normal bivariate distribution, and $\phi(x, y, \rho)$ its probability density. Then the partial derivative of $\Phi(x, y, \rho)$ with respect to ρ is $\phi(x, y, \rho)$:*

$$\frac{\partial \Phi(x, y, \rho)}{\partial \rho} = \phi(x, y, \rho). \tag{7}$$

In particular, $\Phi(x, y, \rho)$ is strictly increasing in ρ for any $x, y \in \mathbb{R}$. Notice that if $x = \pm\infty$ or $y = \pm\infty$, Φ becomes constant in ρ rather than increasing.

This lemma allows us to derive qualitative results about the role of differential correlation and the difference in the outcomes of students depending on their group. Next, consider the difference between groups regarding students getting their second choice or remaining unmatched.

Theorem 3.5. *The proportion of students getting their second choice and remaining unmatched is* not *the same across both groups: students from the group with higher correlation coefficient have a lower probability of getting their second choice and a higher probability of staying unmatched. Formally, if $\rho_{G_1} < \rho_{G_2}$, then $V_2^{G_1,\boldsymbol{A}} > V_2^{G_2,\boldsymbol{A}}$, $V_2^{G_1,\boldsymbol{B}} > V_2^{G_2,\boldsymbol{B}}$, $V_\emptyset^{G_1,\boldsymbol{A}} < V_\emptyset^{G_2,\boldsymbol{A}}$ and $V_\emptyset^{G_1,\boldsymbol{B}} < V_\emptyset^{G_2,\boldsymbol{B}}$.*

Proof. Suppose, without loss of generality, that $\rho_{G_1} < \rho_{G_2}$. Then, by using Lemma 3.2 to write the quantities V_2 we have

$$\begin{aligned} V_2^{G_1,\mathbf{A}} &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) \\ &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}) - \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}}) \\ &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}) - \Phi(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}) \\ &= \mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < P_{\mathbf{A}}) - \Phi(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}) \\ &\geq \mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < P_{\mathbf{A}}) - \Phi(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_2}) \\ &= V_2^{G_2,\mathbf{A}}, \end{aligned} \tag{8}$$

and the same holds for $V_2^{G_1,\mathbf{B}}$ and $V_2^{G_2,\mathbf{B}}$. The inequality at the second-to-last line is a direct application of Lemma 3.4 and it is strict if and only if $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are not infinite. $P_{\mathbf{A}} = +\infty$ would happen if and only if $\alpha_{\mathbf{A}} = 0$, which we excluded in the definition of the model (because it would not be a matching problem anymore); and $P_{\mathbf{A}} = -\infty$ would mean that college **A** accepts all students applying to it, which we also ruled out by assuming $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. The same arguments hold for $P_{\mathbf{B}}$. Hence the inequality is strict.

The inequalities between the $V_{\emptyset}$ quantities are direct consequences of Φ being increasing in ρ by using the formulation in (6) from Lemma 3.4. ■

This result shows that while the chance of getting their first choice is the same for students of both groups, it differs for the second choice and staying unmatched. Specifically, the students from the group with higher correlation have lower chances of getting their second choice and hence higher chances of not getting matched to any college at all. While this may seem surprising at first since a higher correlation is associated with a lower noise in the quality estimation, it is in fact quite intuitive: as the marginals are the same, there are as many good students in each group, but a student with high correlation that has been rejected from their first choice has a high chance of also being rejected from the second one. On the other hand, the fact that a student with low correlation has been rejected from her first choice does not give a lot of information on her chances at the other college. A possible interpretation of this result is that differential correlation levels will not hurt good students, but will hurt intermediate students who have high grade correlation and might have been admitted to their second choice had they been in the other group.

3.3 Comparative statics

With the understanding of the effect of differential correlation on students of each group, we now turn to analyze how noise—that is correlation—influences the overall quality of the matching. We first consider how the probability of getting one's first choice varies when changing the correlation for one group (recall that this probability is the same for both groups).

Theorem 3.6. *The probability that a student of either group gets their first choice is increasing in both ρ_{G_1} and ρ_{G_2}. Formally, let $\gamma, \beta_{G_1}, \beta_{G_2} \in [0,1]$, $\alpha_{\boldsymbol{A}}, \alpha_{\boldsymbol{B}} \in (0,1)$ such that $\alpha_{\boldsymbol{A}} + \alpha_{\boldsymbol{B}} < 1$, and $\rho_{G_1}, \rho_{G_2} \in [0,1)$. Then for $\boldsymbol{C} \in \{\boldsymbol{A}, \boldsymbol{B}\}$, $G \in \{G_1, G_2\}$,*

$$\frac{\partial V_1^{G,\boldsymbol{C}}}{\partial \rho_{G_1}} > 0 \quad \textit{and} \quad \frac{\partial V_1^{G,\boldsymbol{C}}}{\partial \rho_{G_2}} > 0$$

Sketch of Proof. The proof works in multiple steps. First, we rewrite the market clearing equation (3) using Lemma 3.2. We obtain a system of two equations, where the variables are the cutoffs $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$, parameterized by the correlation coefficients. We then apply the implicit function theorem to a mapping whose roots are the solution of this system of equations. We first check that its Jacobian has a positive determinant, and then compute the partial derivatives. To characterize the sign of the derivatives with respect to ρ, we use Lemma 3.4. Through analytical derivations, we then arrive at the conclusion of Theorem 3.6. The details are provided in Appendix A.2. ■

Note that the formal statement of the positivity of partial derivatives in Theorem 3.6 excludes the cases where $\rho_{G_1} = 1$ or $\rho_{G_2} = 1$. That is because for these values some demand functions are not differentiable. As the first choice functions are continuous, however, they are increasing nonetheless on the whole interval $[0, 1]$.

Theorem 3.6 implies that, if the estimation noise increases (i.e., the correlation decreases) for the grades of one of the groups, then both groups suffer the same decrease in first-choice admittance. By contrast, decreasing the noise (i.e., increasing the correlation) for one group leads to an increase in first-choice admittance for all groups. This means that both groups may benefit (at least as far as the probability of first choice assignment is concerned) from colleges reducing the estimation noise on either group. In particular, if there is a majority group with low estimation noise (i.e., high correlation) and a minority group with high estimation noise (i.e., low correlation), then also the majority group will benefit from colleges reducing the estimation noise on the minority group.

Consider the increase of the proportion of students getting their first choice when the correlation of either group increases. This maps to a decrease of the cutoffs of both colleges. How do both cutoffs decrease, while the colleges capacities remain the same? We offer a geometric intuition: consider the real plane divided into 4 areas by the lines $x = P_{\mathbf{A}}$ and $y = P_{\mathbf{B}}$. Students in the upper-right quadrant get their favorite college. Students in upper-left quadrant who prefer **B** and students in the lower-right quadrant preferring **A** also get their first choice. However, students from these quadrants who have the opposite preference get their second choice. Finally, students from the lower-left quadrant are rejected from both colleges. When the correlation increases, students grade vectors accumulate close to the diagonal, and therefore in the lower-left and upper-right quadrants, while the other two quadrants are increasingly empty. This phenomenon is illustrated in Figure 2. If the cutoffs did not change, then the amount of unmatched students would increase, which is not feasible as the capacities are constant. Therefore, at least one of the cutoffs needs to decrease to compensate this (in fact, Theorem 3.6 implies that both are decreasing). As a consequence, some students who would previously get their second choice now get their first one.

Theoreom 3.6 allows us to derive the following corollary:

Corollary 3.7. *The proportion of students from a given group remaining unmatched is increasing in its own correlation level and decreasing in the correlation level of the other group:*
for $\boldsymbol{C} \in \{\boldsymbol{A}, \boldsymbol{B}\}$, $G \in \{G_1, G_2\}$,

$$\frac{\partial V_{\emptyset}^{G,\boldsymbol{C}}}{\partial \rho_G} > 0 \quad \textit{and} \quad \frac{\partial V_{\emptyset}^{G,\boldsymbol{C}}}{\partial \rho_{\overline{G}}} < 0,$$

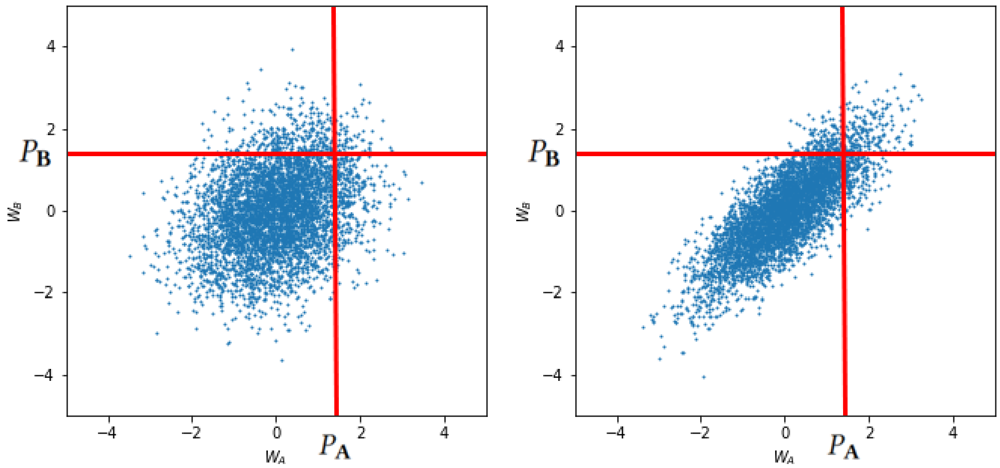


Figure 2: Illustration of the proportions of students in each quadrant for different correlations. The red lines correspond to a vertical line at $P_{\mathbf{A}}$ and a horizontal line at $P_{\mathbf{B}}$. On the left $\rho = 0.3$, on the right $\rho = 0.8$. We can see that the amount of points in the bottom-left quadrant increases with the correlation.

where $\overline{G}$ *is the other group.*

Sketch of Proof. The proof reuses some of the partial derivatives with respect to the correlation coefficient computed in the proof of Thereom 3.6 and relies and the mass conservation to derive the sign of the partial derivatives of the $V_{\emptyset}$ terms. The details are provided in Appendix A.3. ■

This corollary is enlightening to understand what happens when the correlation of a given group varies unilaterally: while it benefits good students from both groups equally, it hurts the intermediate students from this group to the benefit of those from the other group.

4 Special Cases

In our model, the exact solutions of the market clearing equation, and thus the metrics, are in general intractable because the probability density function of the normal distribution has no primitive. In this section, we focus on some notable special cases for which these calculations are nonetheless possible and which thus allow us to gain further intuition for the mechanisms at work.

4.1 Capacity excess

In the previous section we assumed that the total capacity of colleges was smaller than the number of students, i.e., $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. The following proposition shows that if the total capacity exceeds the number of students, the noise no longer has any influence on the matching.

Proposition 4.1. *If there is sufficient capacity to admit all students, i.e., $\alpha_{\boldsymbol{A}} + \alpha_{\boldsymbol{B}} \geq 1$, then correlation has no effect on the matching, and students from both groups have the exact same outcome distributions. In other words, for $k \in \{1,2\}$, $\boldsymbol{C} \in \{\boldsymbol{A}, \boldsymbol{B}\}$, we have $V_k^{G_1,\boldsymbol{C}} = V_k^{G_2,\boldsymbol{C}}$. Moreover, for $G \in \{G_1, G_2\}$, $\boldsymbol{C} \in \{\boldsymbol{A}, \boldsymbol{B}\}$, $V_\emptyset^{G,\boldsymbol{C}} = 0$. Finally, these quantities are constant in ρ_{G_1} and ρ_{G_2}.*

Sketch of Proof. The relations between the V_1 are from Theorem 3.3 (which did not assume $\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}} < 1$). To prove the relations between the V_2, we write them as in Lemma 3.2 and use the fact that $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$ implies that at least one of the cutoff $P_{\mathbf{A}}$ or $P_{\mathbf{B}}$ is $-\infty$. The details are provided in Appendix A.4. ■

In fact, when $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$, it is possible to compute explicitly the proportions of students getting their first or second choice by analyzing Algorithm 1. We consider three (partitioning) cases:

(1) *There is not enough room in $\boldsymbol{A}$ for all students preferring it to $\boldsymbol{B}$, i.e., $\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2} \geq \alpha_{\boldsymbol{A}}$.* In this case, there is necessarily enough room in $\mathbf{B}$ for all students preferring it, since $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$. Therefore, following the steps of Algorithm 1, we find:

- (*i*) At step one, $\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2}$ students preferring $\mathbf{A}$ apply there and $\alpha_{\mathbf{A}}$ are temporarily admitted, and $\gamma(1-\beta_{G_1}) + (1-\gamma)(1-\beta_{G_1})$ students preferring $\mathbf{B}$ apply there and are all temporarily admitted.

- (*ii*) At step two, the $\gamma\beta_{G_1}+(1-\gamma)\beta_{G_2}-\alpha_{\mathbf{A}}$ students rejected from $\mathbf{A}$ apply to $\mathbf{B}$, and are admitted since there is enough room for them (considering the students previously admitted).

This results in the following chances of a student to get their first or second choice:

$$V_1^{G_1,\mathbf{A}} = V_1^{G_2,\mathbf{A}} = \frac{\alpha_{\mathbf{A}}}{\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2}}, \qquad V_1^{G_1,\mathbf{B}} = V_1^{G_2,\mathbf{B}} = 1, \tag{9}$$

$$V_2^{G_1,\mathbf{A}} = V_2^{G_2,\mathbf{A}} = 1 - \frac{\alpha_{\mathbf{A}}}{\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2}}, \qquad V_2^{G_1,\mathbf{B}} = V_2^{G_2,\mathbf{B}} = 0. \tag{10}$$

Finally, as every student is admitted somewhere, $V_\emptyset^{G,\mathbf{C}} = 0$ for $G \in \{G_1, G_2\}$ and $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$.

(2) *There is not enough room in $\boldsymbol{B}$ for all students preferring it to $\boldsymbol{A}$, i.e., $\gamma(1-\beta_{G_1})+(1-\gamma)(1-\beta_{G_1}) \geq \alpha_{\boldsymbol{B}}$* This implies that $\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2} \leq \alpha_{\mathbf{A}}$ and, by symmetry, we obtain the same results as in (9) and (10) with $\mathbf{A}$ and $\mathbf{B}$ exchanging roles, β_{G_1} becoming $1-\beta_{G_1}$ and β_{G_2} becoming $1-\beta_{G_2}$.

(3) *There is enough room in each college to admit all students who prefer attending it, i.e., $\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2} \leq \alpha_{\boldsymbol{A}}$ and $\gamma(1-\beta_{G_1}) + (1-\gamma)(1-\beta_{G_1}) \leq \alpha_{\boldsymbol{B}}$.* It follows that everyone gets their first choice: for $G \in \{G_1, G_2\}$ and $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$,

$$V_1^{G,\mathbf{C}} = 1, \tag{11}$$

$$V_2^{G,\mathbf{C}} = V_\emptyset^{G,\mathbf{C}} = 0. \tag{12}$$

Note that Equations (9) to (12) are consistent with Proposition 4.1.

4.2 One group

Suppose now that there is only one group of students (i.e., $\gamma = 1$) and therefore all students have the same correlation coefficient ρ. In this section we consider the two boundary cases, where the colleges are either able to perfectly assess qualities—the correlation is 1—or are not able to assess qualities at all—the correlation is 0. These cases will allow us to understand the matching's dependencies on the capacities and preferences of students. We assume that there is no excess of capacity, i.e., $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. Since there is only one group, there is only one parameter β for the proportion of students preferring $\mathbf{A}$, and the metrics V_1, V_2 and $V_\emptyset$ do not depend on the group.

4.2.1 Full correlation

We first study the case $\rho = 1$, i.e., students have the same grade in both colleges.

Proposition 4.2. *When $\rho = 1$, we have:*

(i) If $\beta \leq \frac{\alpha_{\boldsymbol{A}}}{\alpha_{\boldsymbol{A}}+\alpha_{\boldsymbol{B}}}$,

$$V_1^{\boldsymbol{A}} = \alpha_{\boldsymbol{A}} + \alpha_{\boldsymbol{B}}, \qquad V_1^{\boldsymbol{B}} = \frac{\alpha_{\boldsymbol{B}}}{1-\beta},$$

$$V_2^{\boldsymbol{A}} = 0, \qquad V_2^{\boldsymbol{B}} = \alpha_{\boldsymbol{A}} - \frac{\beta}{1-\beta}\alpha_{\boldsymbol{B}},$$

$$V_\emptyset^{\boldsymbol{A}} = 1 - \alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}}, \qquad V_\emptyset^{\boldsymbol{B}} = 1 - \alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}};$$

(ii) If $\beta \geq \frac{\alpha_{\boldsymbol{A}}}{\alpha_{\boldsymbol{A}}+\alpha_{\boldsymbol{B}}}$

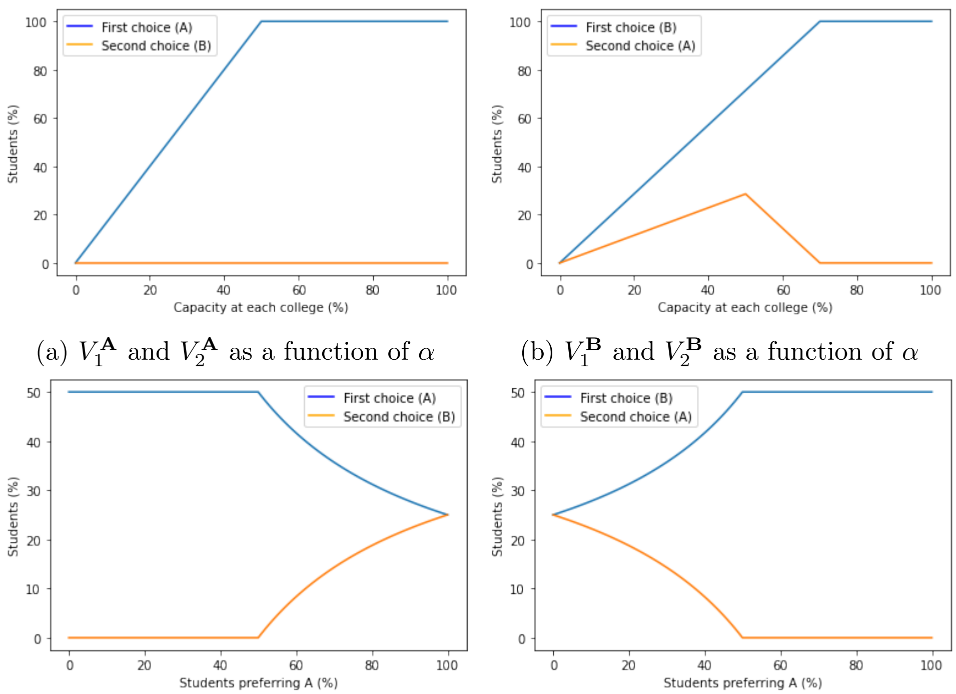
$$V_1^{\boldsymbol{A}} = \frac{\alpha_{\boldsymbol{A}}}{\beta}, \qquad V_1^{\boldsymbol{B}} = \alpha_{\boldsymbol{A}} + \alpha_{\boldsymbol{B}},$$

$$V_2^{\boldsymbol{A}} = \alpha_{\boldsymbol{B}} - \frac{1-\beta}{\beta}\alpha_{\boldsymbol{A}}, \qquad V_2^{\boldsymbol{B}} = 0$$

$$V_\emptyset^{\boldsymbol{A}} = 1 - \alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}}, \qquad V_\emptyset^{\boldsymbol{B}} = 1 - \alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}}.$$

Sketch of Proof. The proof amounts to solving the market clearing equation, the details are provided in Appendix A.5. ■

To illustrate this result, Figure 3 shows V_1 for particular values of the parameters. For the top row, $\beta = 0.3$ and $V_1^{\mathbf{A}}$ and $V_1^{\mathbf{B}}$ are computed for $\alpha_{\mathbf{A}} = \alpha_{\mathbf{B}}$ varying from 0 to 1. Unsurprisingly, the chances of getting one's first choice is increasing in the capacity. We can notice a change of expression of the function when the capacity becomes excessive, and on the right plot we also see a change happening when $\alpha_{\mathbf{B}}$ becomes greater than $1-\beta$, as predicted in Proposition 4.1. For the bottom row, $\alpha_{\mathbf{A}} = \alpha_{\mathbf{B}} := \alpha = 0.25$ and we computed $V_1^{\mathbf{A}}$ and $V_1^{\mathbf{B}}$ for β varying from 0 to 1. We observe that students who prefer the least popular college have higher chances of getting their first choice.



(a) $V_1^{\mathbf{A}}$ and $V_2^{\mathbf{A}}$ as a function of α

(b) $V_1^{\mathbf{B}}$ and $V_2^{\mathbf{B}}$ as a function of α

(c) $V_1^{\mathbf{A}}$ and $V_2^{\mathbf{A}}$ as a function of β

(d) $V_1^{\mathbf{B}}$ and $V_2^{\mathbf{B}}$ as a function of β

Figure 3: Proportion of students getting their first and second choice with $\rho = 1$, and $\alpha_{\mathbf{A}} = \alpha_{\mathbf{B}} := \alpha$.

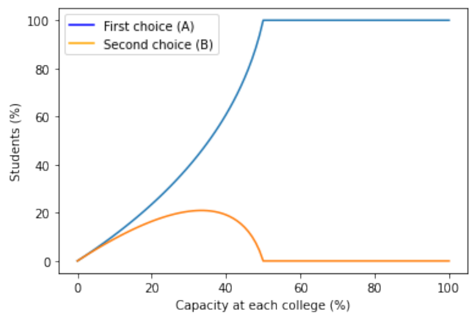
4.2.2 No correlation

Now, consider the case $\rho = 0$, i.e., the grades a student gets at $\mathbf{A}$ and $\mathbf{B}$ are independent random variables.

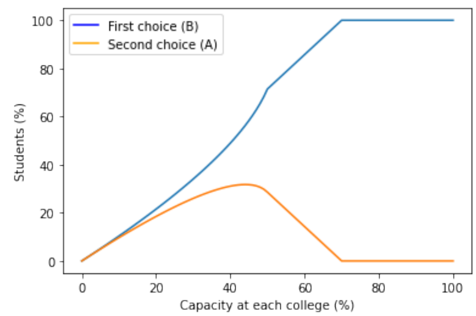
Proposition 4.3. *When $\rho = 0$, we have:*

$$
\begin{aligned}
V_1^{\boldsymbol{A}} &= 1 - \frac{1-\beta}{2\beta}(\Delta - \zeta), & V_1^{\boldsymbol{B}} &= 1 - \frac{1}{2}(\Delta + \zeta), \\
V_2^{\boldsymbol{A}} &= \frac{1-\beta}{2\beta}(\Delta - \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_2^{\boldsymbol{B}} &= \frac{1}{2}(\Delta + \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), \\
V_\emptyset^{\boldsymbol{A}} &= \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_\emptyset^{\boldsymbol{B}} &= \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2);
\end{aligned}
$$

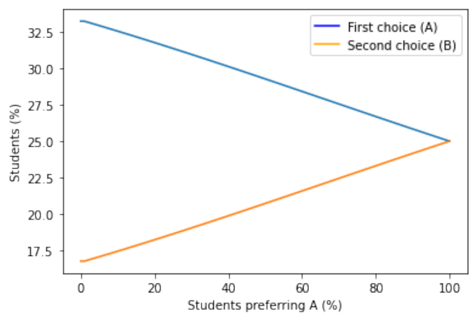
with $\zeta = \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}}$ *and* $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha_{\boldsymbol{A}} - \alpha_{\boldsymbol{B}})}$



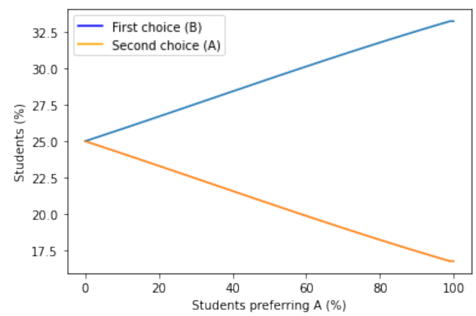
(a) $V_1^{\mathbf{A}}$ and $V_2^{\mathbf{A}}$ as a function of α



(b) $V_1^{\mathbf{B}}$ and $V_2^{\mathbf{B}}$ as a function of α



(c) $V_1^{\mathbf{A}}$ and $V_2^{\mathbf{A}}$ as a function of β



(d) $V_1^{\mathbf{B}}$ and $V_2^{\mathbf{B}}$ as a function of β

Figure 4: Proportion of students getting their first and second choice with $\rho = 0$, and $\alpha_{\mathbf{A}} = \alpha_{\mathbf{B}} := \alpha$.

Sketch of Proof. Once again, the proof amounts to solving the market clearing equation, the details are provided in Appendix A.6. ■

We illustrate this result in Figure 4 by plotting the same quantities as in Figure 3, but with $\rho = 0$ this time. A quick comparison between these plots and those displayed in Figure 3 shows that the first choice curves of Figure 4 are always below those of Figure 3. This was expected, as Theorem 3.6 states that V_1 is increasing in the correlation, for any values of the parameters. As a consequence, the proportion of students getting their first choice is always higher for $\rho = 1$ than for $\rho = 0$. For intermediate values of ρ, the curves would be contained between the ones for $\rho = 0$ and $\rho = 1$. Computing these two extreme cases therefore provides a good intuition of the dynamics even for intermediate values of the correlation.

5 Conclusion

In this work, we introduced a model for the college admission problem that accounts for statistical discrimination, i.e., when colleges' grading accuracy differs across different groups of the population. Statistical discrimination results in noisy estimates of students' qualities, and, in particular, may lead to varying assessments across colleges. Our model isolates the effect of this differential correlation by equalizing the marginal grade distributions of each group at each college. We can therefore study the difference in outcomes for students based on their group caused by differential correlation.

We first showed that students of all groups have the same probability of getting their first choice and this probability is decreasing as quality estimates become noisier. This implies, that 'good' students—i.e., students who should have gotten their first choice—from both groups are equally suffering from noisy quality estimates. Moreover, we show that students from the group with high correlation have a lower probability of getting their second choice and a higher probability of being rejected from both colleges compared to the other group. That is, students from the group whose true qualities are more accurately estimated by colleges in fact suffer more from differential correlation. Overall, differential correlation does not induce statistical discrimination for good students, but does for all the others. Thus, perfect accuracy in estimating students' qualities would be optimal for all students, and a drop in accuracy for one group not only hurts both groups, but hurts the other group even more.

Several assumptions made for clarity can be relaxed, while preserving results qualitatively. First, we chose to consider only two colleges, as it is sufficient to exhibit the effects of differential correlation while keeping the calculations reasonably simple, but the proof techniques allow for straightforward–albeit cumbersome–extensions. Second, we assumed that the marginals, i.e., the grade distribution of each college, are normal distributions. As the matching procedure is solely based on ordinal comparisons, this assumption is not restrictive and other distributions could be used without altering the results. Third, we also assumed the marginals to be the same at each college and for each group. As long as they are equal for all groups, the marginals being different for each college would not change anything since the ranking would stay the same. If we allow marginals to be different across groups, then the effects of differential correlation we described in this paper would be mixed with the effects of differential bias and variance studied by Emelianov et al. (2022) and Garg et al. (2021). The resulting model would be a very general framework to study statistical discrimination, encompassing models from the aforementioned authors as well as ours.

Our work is, to our knowledge, the first to consider differential correlation as a process of statistical discrimination in matching. Several avenues for future research present themselves, both theoretical and applied. We have assumed that students have no influence on the capability of a college to evaluate their quality. A natural extension would be to allow student's to invest in accurate assessment, e.g., via acquiring certifications or doing in-person interviews. From an empirical point of view one key testable prediction of our model is that it is in fact the group which is more accurately assessed that suffers more from noise than the one subjected to the noise. This finding appears counter-intuitive and merits study.

References

Abdulkadiroğlu, Atila (2005), "College admissions with affirmative action." *International Journal of Game Theory*, 33, 535–549.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2015), "Expanding "Choice" in School Choice." *American Economic Journal: Microeconomics*, 7, 1–42.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth (2009), "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review*, 99, 1954–1978.

Abdulkadiroğlu, Atila and Tayfun Sönmez (2003), "School Choice: A Mechanism Design Approach." *American Economic Review*, 93, 729–747.

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

Arnosti, Nick (2022), "LOTTERY DESIGN FOR SCHOOL CHOICE." 38.

Arrow, Kenneth J. (1973), "The theory of discrimination." In *Discrimination in Labor Markets* (Orley Ashenfelter and Albert Rees, eds.), 3–33, Princeton University Press.

Ashlagi, Itai, Afshin Nikzad, and Assaf Romm (2019), "Assigning more students to their top choices: A comparison of tie-breaking rules." *Games and Economic Behavior*, 115, 167–187, URL `https://www.sciencedirect.com/science/article/pii/S089982561930034X`.

Azevedo, Eduardo M. and Jacob D. Leshno (2016), "A Supply and Demand Framework for Two-Sided Matching Markets." *Journal of Political Economy*, 124, 1235–1268.

Aziz, Haris, Péter Biró, Serge Gaspers, Ronald de Haan, Nicholas Mattei, and Baharak Rastegari (2020), "Stable Matching with Uncertain Linear Preferences." *Algorithmica*, 82, 1410–1433.

Balinski, Michel and Tayfun Sonmez (1999), "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory*, 84, 73–94.

Barocas, Solon, Moritz Hardt, and Arvind Narayanan (2019), *Fairness and Machine Learning*. `http://www.fairmlbook.org`.

Bikhchandani, Sushil (2017), "Stability with one-sided incomplete information." *Journal of Economic Theory*, 168, 372–399.

Biró, Péter (2008), "Student admissions in hungary as gale and shapley envisaged." Technical Report TR-2008-291, Computing Science Department of Glasgow University.

Blum, Avrim and Kevin Stangl (2020), "Recovering from Biased Data: Can Fairness Constraints Improve Accuracy?" In *Proceedings of the Symposium on Foundations of Responsible Computing (FORC 2020)*.

Bonneau, Cécile, Pauline Charousset, Julien Grenet, and Georgia Thebault (2021), "Les grandes écoles: de fortes inégalités d'accès entre élèves franciliens et non franciliens." *Education et Formations*, 102, 367–393.

Celis, L. Elisa, Anay Mehrotra, and Nisheeth K. Vishnoi (2020), "Interventions for ranking in the presence of implicit bias." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*.

Chade, H., G. Lewis, and L. Smith (2014), "Student Portfolios and the College Admissions Problem." *The Review of Economic Studies*, 81, 971–1002.

Chakraborty, Archishman, Alessandro Citanna, and Michael Ostrovsky (2010), "Two-sided matching with interdependent values." *Journal of Economic Theory*, 145, 85–105.

Che, Yeon-Koo, Kyungmin Kim, and Weijie Zhong (2020), "Statistical discrimination in ratings-guided markets." Working paper.

Chouldechova, Alexandra and Aaron Roth (2020), "A snapshot of the frontiers of fairness in machine learning." *Commun. ACM*, 63, 82–89.

Coate, Stephen and Glenn Loury (1993), "Will affirmative-action policies eliminate negative stereotypes?" *American Economic Review*, 83, 1220–40.

Cortes, Anthony (2019), "Parcoursup : l'édition 2019 toujours soupçonnée de discrimination selon le lycée orientation." *Marianne*, URL `https://www.marianne.net/societe/parcoursup-une-nouvelle-edition-encore-marquee-par-des-risques-de-discrimination`. Jan. 22, 2019.

Delacrétaz, David, Scott Duke Kominers, and Alexander Teytelboym (2020), "Matching Mechanisms for Refugee Resettlement." Working paper.

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel (2012), "Fairness through awareness." In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS '12)*.

Emelianov, Vitalii, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau (2020), "On fair selection in the presence of implicit variance." In *Proceedings of the 21st ACM Conference on Economics and Computation (EC)*.

Emelianov, Vitalii, Nicolas Gast, Krishna P. Gummadi, and Patrick Loiseau (2022), "On fair selection in the presence of implicit and differential variance." *Artificial Intelligence*, 302, 103609.

Ergin, Haluk and Tayfun Sönmez (2006), "Games of school choice under the Boston mechanism." *Journal of Public Economics*, 90, 215–237.

Fang, Hanming and Andrea Moro (2011), "Chapter 5 - theories of statistical discrimination and affirmative action: A survey." volume 1 of *Handbook of Social Economics*, 133–200, North-Holland.

Finocchiaro, Jessie, Roland Maio, Faidra Monachou, Gourab K. Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis (2021), "Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*.

Gale, D. and L. S. Shapley (1962), "College Admissions and the Stability of Marriage." *The American Mathematical Monthly*, 69, 9–15.

Garg, Nikhil, Hannah Li, and Faidra Monachou (2021), "Standardized tests and affirmative action: The role of bias and variance." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*.

Gersen, Jeannie Suk (2019), "The many sins of college admissions." *The New Yorker*, URL `https://www.newyorker.com/news/our-columnists/the-many-sins-of-college-admissions`. Oct. 7, 2019.

Hardt, Moritz, Eric Price, and Nathan Srebro (2016), "Equality of opportunity in supervised learning." In *Advances in Neural Information Processing Systems (NIPS '16)*.

Immorlica, Nicole, Jacob Leshno, Irene Lo, and Brendan Lucier (2020), "Information Acquisition in Matching Markets: The Role of Price Discovery." Working paper.

Kamada, Yuichiro and Fuhito Kojima (2015), "Efficient Matching under Distributional Constraints: Theory and Applications." *American Economic Review*, 105, 67–99.

Kamada, Yuichiro and Fuhito Kojima (2022), "Fair Matching under Constraints: Theory and Applications." Review of Economic Studies, forthcoming.

Karni, Gili, Guy N. Rothblum, and Gal Yona (2021), "On Fairness and Stability in Two-Sided Matchings." Working paper.

Kleinberg, Jon and Manish Raghavan (2018), "Selection Problems in the Presence of Implicit Bias." In *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS '18)*.

Kleinberg, Jon and Manish Raghavan (2021), "Algorithmic monoculture and social welfare." *Proceedings of the National Academy of Sciences*, 118.

Kloosterman, Andrew and Peter Troyan (2020), "School choice with asymmetric information: Priority design and the curse of acceptance." *Theoretical Economics*, 15, 1095–1133.

Krishnaa, Prem, Girija Limaye, Meghana Nasre, and Prajakta Nimbhorkar (2019), "Envy-freeness and Relaxed Stability under lower quotas." Working paper.

Liptak, Adam and Anemona Hartocollis (2022), "Supreme court will hear challenge to affirmative action at harvard and u.n.c." *The New York Times*, URL `https://www.nytimes.com/2022/01/24/us/politics/supreme-court-affirmative-action-harvard-unc.html`. Jan. 24, 2022.

Liu, Lydia T., Horia Mania, and Michael I. Jordan (2020), "Competing Bandits in Matching Markets." In *International Conference on Artificial Intelligence and Statistics (AISTATS '20)*.

Liu, Qingmin (2020), "Stability and Bayesian Consistency in Two-Sided Markets." *American Economic Review*, 110, 2625–2666.

Liu, Qingmin, George J. Mailath, Andrew Postlewaite, and Larry Samuelson (2014), "Stable Matching With Incomplete Information." *Econometrica*, 82, 541–587.

Lundberg, Shelly J. and Richard Startz (1983), "Private discrimination and social intervention in competitive labor market." *The American Economic Review*, 73, 340–347.

Machado, Cecilia and Christiane Szerman (2021), "Centralized college admissions and student composition." *Economics of Education Review*, 85, 102184.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan (2021), "A survey on bias and fairness in machine learning." *ACM Comput. Surv.*, 54, 1–35.

Monachou, Faidra and Itai Ashlagi (2019), "Discrimination in online markets: effects of social bias on learning from reviews and policy design." In *Advances in Neural Information Processing Systems (NeurIPS '19)*, 192.

Parcoursup (2021), "Algorithmes de parcoursup." `https://framagit.org/parcoursup/algorithmes-de-parcoursup`.

Phelps, Edmund (1972), "The statistical theory of racism and sexism." *American Economic Review*, 62, 659–661.

Rastegari, Baharak, Anne Condon, Nicole Immorlica, and Kevin Leyton-Brown (2013), "Two-sided matching with partial information." In *Proceedings of the fourteenth ACM Conference on Electronic Commerce (EC '13)*.

Roth, Alvin E. (1986), "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." *Econometrica*, 54, 425–427.

Roth, Alvin E. and Marilda Sotomayor (1992), *Two-sided matching: A study in game-theoretic modeling and analysis*, 1 edition. Handbook of game theory with economic applications, Cambridge University Press.

Sibuya, Masaaki (1960), "Bivariate extreme statistics, I." *Annals of the Institute of Statistical Mathematics*, 11, 195–210.

Vairet, Florent (2021), "Discrimination positive au concours des grandes écoles : les premières mesures sont désormais connues." *Les Echos*, URL `https://start.lesechos.fr/apprendre/universites-ecoles/discrimination-positive-au-concours-des-grandes-ecoles-les-premieres-mesures-sont-desormais-connues-1311179`. Apr. 29, 2021.

Yang, Ke, Joshua R. Loftus, and Julia Stoyanovich (2021), "Causal intersectionality for fair ranking." In *Proceedings of the Symposium on Foundations of Responsible Computing (FORC 2021)*.

Yenmez, M. Bumin (2013), "Incentive-Compatible Matching Mechanisms: Consistency with Various Stability Notions." *American Economic Journal: Microeconomics*, 5, 120–41.

Zafar, Muhammad Bilal, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi (2017), "Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment." In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*.

Zehlike, Meike, Ke Yang, and Julia Stoyanovich (2021), "Fairness in Ranking: A Survey." Working paper.

A Omitted proofs

A.1 Proof of Lemma 3.2

Proof of Lemma 3.2. Consider student $s \in G_1$ who prefers $\mathbf{A}$ to $\mathbf{B}$. By Lemma 2.6, s is admitted to $\mathbf{A}$ if and only if $s \in D_{\mathbf{A}}(P_{\mathbf{A}}, P_{\mathbf{B}})$, i.e., if and only if their grade at $\mathbf{A}$ is greater than $P_{\mathbf{A}}$. Then by definition of η,

$$V_1^{G_1,\mathbf{A}}(\rho_{G_1}, \rho_{G_2}) = \frac{\eta(\{s \in G_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\})}{\gamma \beta_{G_1}} = \mathbb{P}_{\rho_{G_1}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in [P_{\mathbf{A}}, +\infty) \times \mathbb{R}).$$

The same reasoning applies to all quantities of type V_1, which proves (4).

The same student s is admitted to $\mathbf{B}$ if and only if $s \in D_{\mathbf{B}}(P_{\mathbf{A}}, P_{\mathbf{B}})$, i.e., if and only if $W_{\mathbf{B}}^s \geq P_{\mathbf{B}}$ and $W_{\mathbf{A}}^s < P_{\mathbf{A}}$. Then we have

$$V_2^{G_1,\mathbf{A}}(\rho_{G_1}, \rho_{G_2}) = \frac{\eta(\{s \in G_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{B}\})}{\gamma \beta_{G_1}} = \mathbb{P}_{\rho_{G_1}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in (-\infty, P_{\mathbf{A}}) \times [P_{\mathbf{B}}, +\infty)).$$

The same reasoning applies to all quantities of type V_2, which proves (5).

Students s remains unmatched if and only if $W^s_{\mathbf{A}} < P_{\mathbf{A}}$ and $W^s_{\mathbf{B}} < P_{\mathbf{B}}$. Then we have

$$V_{\emptyset}^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) = \frac{\eta(\{s \in G_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = s\})}{\gamma\beta_{G_1}} = \mathbb{P}_{\rho_{G_1}}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in (-\infty, P_{\mathbf{A}}) \times (-\infty, P_{\mathbf{B}})).$$

which proves (6). ■

A.2 Proof of Theorem 3.6

Proof of theorem 3.6. Let $\gamma, \beta_{G_1}, \beta_{G_2} \in [0,1]$, $\alpha_{\mathbf{A}}, \alpha_{\mathbf{B}} \in (0,1)$ such that $\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}} < 1$, and $\rho_{G_1}, \rho_{G_2} \in [0,1)$. Let $P_{\mathbf{A}}, P_{\mathbf{B}} \in \mathbb{R}$ be the cutoffs of colleges **A** and **B**.

By definition of the quantities V_1 and V_2, the market clearing equation (3) can be written as

$$\begin{cases} \gamma\beta_{G_1}V_1^{G_1,\mathbf{A}} + \gamma(1-\beta_{G_1})V_2^{G_1,\mathbf{B}} + (1-\gamma)\beta_{G_2}V_1^{G_2,\mathbf{A}} + (1-\gamma)(1-\beta_{G_2})V_2^{G_2,\mathbf{B}} & = \alpha_{\mathbf{A}}, \\ \gamma\beta_{G_1}V_2^{G_1,\mathbf{A}} + \gamma(1-\beta_{G_1})V_1^{G_1,\mathbf{B}} + (1-\gamma)\beta_{G_2}V_2^{G_2,\mathbf{A}} + (1-\gamma)(1-\beta_{G_2})V_1^{G_2,\mathbf{B}} & = \alpha_{\mathbf{B}}. \end{cases}$$

Then, using Lemma 3.2, we can rewrite is as

$$\begin{cases} (\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2})\mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}}) + \gamma(1-\beta_{G_1})\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}}) & \\ \qquad + (1-\gamma)(1-\beta_{G_2})\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}}) & = \alpha_{\mathbf{A}}, \\ \gamma\beta_{G_1}\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) + (1-\gamma)\beta_{G_2}\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) & \\ \qquad + (\gamma(1-\beta_{G_1}) + (1-\gamma)(1-\beta_{G_2}))\mathbb{P}(W_{\mathbf{B}} \geq P_{\mathbf{B}}) & = \alpha_{\mathbf{B}}, \end{cases}$$

which is finally equivalent to

$$\begin{cases} (\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2})\int_{P_{\mathbf{A}}}^{\infty}\phi(x)\,\mathrm{d}x + \gamma(1-\beta_{G_1})\int_{P_{\mathbf{A}}}^{\infty}\int_{-\infty}^{P_{\mathbf{B}}}\phi(x,y,\rho_{G_1})\,\mathrm{d}x\mathrm{d}y & \\ \qquad + (1-\gamma)(1-\beta_{G_2})\int_{P_{\mathbf{A}}}^{\infty}\int_{-\infty}^{P_{\mathbf{B}}}\phi(x,y,\rho_{G_2})\,\mathrm{d}x\mathrm{d}y & = \alpha_{\mathbf{A}}, \\ \gamma\beta_{G_1}\int_{-\infty}^{P_{\mathbf{A}}}\int_{P_{\mathbf{B}}}^{\infty}\phi(x,y,\rho_{G_1})\,\mathrm{d}x\mathrm{d}y + (1-\gamma)\beta_{G_2}\int_{-\infty}^{P_{\mathbf{A}}}\int_{P_{\mathbf{B}}}^{\infty}\phi(x,y,\rho_{G_2})\,\mathrm{d}x\mathrm{d}y & \\ \qquad + (\gamma(1-\beta_{G_1}) + (1-\gamma)(1-\beta_{G_2}))\int_{P_{\mathbf{B}}}^{\infty}\phi(x)\,\mathrm{d}x & = \alpha_{\mathbf{B}}. \end{cases} \tag{13}$$

We fix ρ_{G_2}, and we want to study how the solution $(P_{\mathbf{A}}, P_{\mathbf{B}})$ of the above equation vary as a function of ρ_{G_1}. Let us define $f : \mathbb{R}^2 \times [0,1) \to \mathbb{R}^2$, $(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}) \mapsto (D_{\mathbf{A}}(P_{\mathbf{A}}, P_{\mathbf{B}}) - \alpha_{\mathbf{A}}, D_{\mathbf{B}}(P_{\mathbf{A}}, P_{\mathbf{B}}) - \alpha_{\mathbf{B}})$. (We will denote by f_1 and f_2 its two components.) Then for each $\rho_{G_1} \in [0,1)$, $(P_{\mathbf{A}}, P_{\mathbf{B}})$ is the solution of the equation $f(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}) = 0_{\mathbb{R}^2}$. In order to show that $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are decreasing in ρ_{G_1}, we apply the implicit function theorem. Let $P_{\mathbf{A}}, P_{\mathbf{B}} \in \mathbb{R}$ and $\rho_{G_1} \in [0,1)$ such that $f(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}) = 0$. Function f is of class C^1. We first verify that the partial

Jacobian $J_{f,(P_{\mathbf{A}},P_{\mathbf{B}})}(P_{\mathbf{A}},P_{\mathbf{B}},\rho_{G_1})$ is invertible, where

$$J_{f,(P_{\mathbf{A}},P_{\mathbf{B}})}(P_{\mathbf{A}},P_{\mathbf{B}},\rho_{G_1}) = \begin{pmatrix} \frac{\partial f_1}{\partial P_{\mathbf{A}}} & \frac{\partial f_1}{\partial P_{\mathbf{B}}} \\ \\ \frac{\partial f_2}{\partial P_{\mathbf{A}}} & \frac{\partial f_2}{\partial P_{\mathbf{B}}} \end{pmatrix}. \tag{14}$$

To show that the determinant $\frac{\partial f_1}{\partial P_{\mathbf{A}}}\frac{\partial f_2}{\partial P_{\mathbf{B}}} - \frac{\partial f_1}{\partial P_{\mathbf{B}}}\frac{\partial f_2}{\partial P_{\mathbf{A}}} \neq 0$, we will show that it is in fact (strictly) positive. From (13), it is clear that f_1 is decreasing in $P_{\mathbf{A}}$ and increasing in $P_{\mathbf{B}}$, and that f_2 is increasing in $P_{\mathbf{A}}$ and decreasing in $P_{\mathbf{B}}$. Therefore, to prove that $\frac{\partial f_1}{\partial P_{\mathbf{A}}}\frac{\partial f_2}{\partial P_{\mathbf{B}}} - \frac{\partial f_1}{\partial P_{\mathbf{B}}}\frac{\partial f_2}{\partial P_{\mathbf{A}}} > 0$, we only need to prove that $\left|\frac{\partial f_1}{\partial P_{\mathbf{A}}}\right| > \frac{\partial f_2}{\partial P_{\mathbf{A}}}$ and $\left|\frac{\partial f_2}{\partial P_{\mathbf{B}}}\right| > \frac{\partial f_1}{\partial P_{\mathbf{B}}}$.

By symmetry, we will only prove the first one. We can compute each term separately:

$$\begin{aligned}
\frac{\partial f_1}{\partial P_{\mathbf{A}}} =& (\gamma\beta_{G_1} + (1-\gamma)\beta_{G_2})\frac{\partial\,\mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}} + \gamma(1-\beta_{G_1})\frac{\partial\,\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} \\
&+ (1-\gamma)(1-\beta_{G_2})\frac{\partial\,\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}, \\
\frac{\partial f_2}{\partial P_{\mathbf{A}}} =& \gamma\beta_{G_1}\frac{\partial\,\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} + (1-\gamma)\beta_{G_2}\frac{\partial\,\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.
\end{aligned}$$

All terms of f_1 are decreasing in $P_{\mathbf{A}}$ and all terms of f_2 are increasing in $P_{\mathbf{A}}$, therefore we can proceed term by term:

$$\begin{aligned}
\left|\gamma\beta_{G_1}\frac{\partial\,\mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}}\right| &= \gamma\beta_{G_1}\frac{\partial\,\mathbb{P}(W_{\mathbf{A}} < P_{\mathbf{A}})}{\partial P_{\mathbf{A}}}, \\
&= \gamma\beta_{G_1}\left(\frac{\partial\,\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} + \frac{\partial\,\mathbb{P}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}\right), \\
&> \gamma\beta_{G_1}\frac{\partial\,\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.
\end{aligned} \tag{15}$$

The same reasoning, when replacing ρ_{G_1} by ρ_{G_2} in the (15) shows that

$$\left|(1-\gamma)\beta_{G_2}\frac{\partial\,\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}}\right| > (1-\gamma)\beta_{G_2}\frac{\partial\,\mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.$$

We conclude that $\left|\frac{\partial f_1}{\partial P_{\mathbf{A}}}\right| > \frac{\partial f_2}{\partial P_{\mathbf{A}}}$, and similarly $\left|\frac{\partial f_2}{\partial P_{\mathbf{B}}}\right| > \frac{\partial f_1}{\partial P_{\mathbf{B}}}$. Therefore the Jacobian in (14) has positive determinant and is invertible.

By the implicit function theorem, there exists a neighborhood U of $(P_{\mathbf{A}}, P_{\mathbf{B}})$, a neighborhood V

of ρ_{G_1}, and a function $\psi : V \to U$ such that for all $(x, y) \in \mathbb{R}^2$, $\rho \in [0, 1)$,

$$((x, y, \rho) \in U \times V \text{ and } f(x, y, \rho) = 0) \Leftrightarrow (\rho \in V \text{ and } (x, y) = \psi(\rho)).$$

In particular, $(P_{\mathbf{A}}, P_{\mathbf{B}}) = \psi(\rho_{G_1})$, and we can compute the derivative of ψ:

$$\begin{aligned}
J_\psi(\rho_{G_1}) &= -J_{f,(P_{\mathbf{A}},P_{\mathbf{B}})}(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1})^{-1}\, J_{f,\rho_{G_1}}(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho_{G_1}), \\
&= \frac{-1}{\frac{\partial f_1}{\partial P_{\mathbf{A}}}\frac{\partial f_2}{\partial P_{\mathbf{B}}} - \frac{\partial f_1}{\partial P_{\mathbf{B}}}\frac{\partial f_2}{\partial P_{\mathbf{A}}}} \begin{pmatrix} \frac{\partial f_2}{\partial P_{\mathbf{B}}} & -\frac{\partial f_1}{\partial P_{\mathbf{B}}} \\ -\frac{\partial f_2}{\partial P_{\mathbf{A}}} & \frac{\partial f_1}{\partial P_{\mathbf{A}}} \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial \rho_{G_1}} \\ \frac{\partial f_2}{\partial \rho_{G_1}} \end{pmatrix}, \\
&= \frac{-1}{\frac{\partial f_1}{\partial P_{\mathbf{A}}}\frac{\partial f_2}{\partial P_{\mathbf{B}}} - \frac{\partial f_1}{\partial P_{\mathbf{B}}}\frac{\partial f_2}{\partial P_{\mathbf{A}}}} \begin{pmatrix} \frac{\partial f_2}{\partial P_{\mathbf{B}}}\frac{\partial f_1}{\partial \rho_{G_1}} - \frac{\partial f_1}{\partial P_{\mathbf{B}}}\frac{\partial f_2}{\partial \rho_{G_1}} \\ -\frac{\partial f_2}{\partial P_{\mathbf{A}}}\frac{\partial f_1}{\partial \rho_{G_1}} + \frac{\partial f_1}{\partial P_{\mathbf{A}}}\frac{\partial f_2}{\partial \rho_{G_1}} \end{pmatrix}.
\end{aligned} \tag{16}$$

We only need to know the sign of each term to conclude about the variations of ψ. We already know the sign of the derivatives in $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$, so we one need those in ρ_{G_1}. The only term in f_1 that depends on ρ_{G_1} is $\gamma(1 - \beta_{G_1})\mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})$. Therefore,

$$\begin{aligned}
\frac{\partial f_1}{\partial \rho_{G_1}} &= \gamma(1 - \beta_{G_1}) \frac{\partial \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial \rho_{G_1}} \\
&= \gamma(1 - \beta_{G_1}) \left(\frac{\partial \mathbb{P}(W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial \rho_{G_1}} - \frac{\partial \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial \rho_{G_1}} \right) \\
&= -\gamma(1 - \beta_{G_1}) \frac{\partial \Phi(P_{\mathbf{A}}, P_{\mathbf{B}}, \rho)}{\partial \rho_{G_1}}.
\end{aligned}$$

and $\frac{\partial f_1}{\partial \rho_{G_1}} < 0$ by Lemma 3.4. By the same argument, $\frac{\partial f_2}{\partial \rho_{G_1}}$ is also negative. (Note that here the implicit functions theorem requires that we compute the partial derivatives of f as if $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ were not functions of ρ_{G_1}.)

If we replace each term of the last line of (16) by its sign, we get

$$-\frac{1}{+} \begin{pmatrix} (- \times -) - (+ \times -) \\ -(+ \times -) + (- \times -) \end{pmatrix} = \begin{pmatrix} - \\ - \end{pmatrix}.$$

We conclude that ψ and therefore $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are decreasing in ρ_{G_1}, and finally:

$$\frac{\partial V_1^{G_1,\mathbf{A}}}{\partial \rho_{G_1}}, \frac{\partial V_1^{G_1,\mathbf{B}}}{\partial \rho_{G_1}}, \frac{\partial V_1^{G_2,\mathbf{A}}}{\partial \rho_{G_1}}, \frac{\partial V_1^{G_2,\mathbf{B}}}{\partial \rho_{G_1}} > 0.$$

The problem being symmetric between ρ_{G_1} and ρ_{G_2}, we also conclude that

$$\frac{\partial V_1^{G_1,\mathbf{A}}}{\partial \rho_{G_2}}, \frac{\partial V_1^{G_1,\mathbf{B}}}{\partial \rho_{G_2}}, \frac{\partial V_1^{G_2,\mathbf{A}}}{\partial \rho_{G_2}}, \frac{\partial V_1^{G_2,\mathbf{B}}}{\partial \rho_{G_2}} > 0,$$

which concludes the proof of the theorem. ■

A.3 Proof of Corollary 3.7

Proof of corollary 3.7. We showed in the proof of Theorem 3.6 that $P_\mathbf{A}$ and $P_\mathbf{B}$ are decreasing in both ρ_{G_1} and ρ_{G_2}. By using this, we have:

$$\frac{\partial V_\emptyset^{G_2,\mathbf{A}}}{\partial \rho_{G_1}} = \frac{\partial V_\emptyset^{G_2,\mathbf{B}}}{\partial \rho_{G_1}} = \frac{\partial \, \mathbb{P}_{\rho_{G_2}}(W_\mathbf{A} < P_\mathbf{A}, W_\mathbf{B} < P_\mathbf{B})}{\partial \rho_{G_1}} < 0.$$

Since the total capacity (of the two colleges) is constant, the mass of unmatched student must also be constant. Therefore, we have

$$\gamma \beta_{G_1} V_\emptyset^{G_1,\mathbf{A}} + \gamma(1-\beta_{G_1}) V_\emptyset^{G_1,\mathbf{B}} + (1-\gamma)\beta_{G_2} V_\emptyset^{G_2,\mathbf{A}} + (1-\gamma)(1-\beta_{G_1}) V_\emptyset^{G_2,\mathbf{B}} = 1 - \alpha_\mathbf{A} - \alpha_\mathbf{B}.$$

Rearranging the terms gives

$$\underbrace{\gamma \beta_{G_1} \frac{\partial V_\emptyset^{G_1,\mathbf{A}}}{\partial \rho_{G_1}} + \gamma(1-\beta_{G_1}) \frac{\partial V_\emptyset^{G_1,\mathbf{B}}}{\partial \rho_{G_1}}}_{>0} = -((1-\gamma)\beta_{G_2} \frac{\partial V_\emptyset^{G_2,\mathbf{A}}}{\partial \rho_{G_1}} + (1-\gamma)(1-\beta_{G_1}), \frac{\partial V_\emptyset^{G_2,\mathbf{B}}}{\partial \rho_{G_1}}).$$

Since $V_\emptyset^{G_1,\mathbf{A}} = V_\emptyset^{G_1,\mathbf{B}}$, we conclude that both $\frac{\partial V_\emptyset^{G_1,\mathbf{A}}}{\partial \rho_{G_1}}$ and $\frac{\partial V_\emptyset^{G_1,\mathbf{B}}}{\partial \rho_{G_1}}$ are positive.

The same argument shows that $\frac{\partial V_\emptyset^{G_1,\mathbf{A}}}{\partial \rho_{G_2}} = \frac{\partial V_\emptyset^{G_1,\mathbf{B}}}{\partial \rho_{G_2}} < 0$ and $\frac{\partial V_\emptyset^{G_1,\mathbf{A}}}{\partial \rho_{G_1}} = \frac{\partial V_\emptyset^{G_1,\mathbf{B}}}{\partial \rho_{G_1}} > 0$, which concludes the proof. ■

A.4 Proof of Proposition 4.1

Proof of Proposition 4.1. The relations between the V_1 are from Theorem 3.3. If $\alpha_\mathbf{A} + \alpha_\mathbf{B} \geq 1$, then all students are admitted to some college, therefore either $P_\mathbf{A} = -\infty$ or $P_\mathbf{B} = -\infty$. Let us suppose it is $P_{\mathbf{A}}$. Then we have

$$\begin{aligned} V_2^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}) &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} < -\infty, W_{\mathbf{B}} \geq P_{\mathbf{B}}), \\ &= 0, \\ &= \mathbb{P}_{\rho_{G_2}}(W_{\mathbf{A}} < -\infty, W_{\mathbf{B}} \geq P_{\mathbf{B}}), \\ &= V_2^{G_1,\mathbf{A}}(\rho_{G_1},\rho_{G_2}), \end{aligned}$$

and

$$\begin{aligned} V_2^{G_1,\mathbf{B}}(\rho_{G_1},\rho_{G_2}) &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{A}} \geq -\infty, W_{\mathbf{B}} < P_{\mathbf{B}}), \\ &= \mathbb{P}_{\rho_{G_1}}(W_{\mathbf{B}} < P_{\mathbf{B}}), \\ &= \mathbb{P}_{\rho_{G_2}}(W_{\mathbf{B}} < P_{\mathbf{B}}), \\ &= V_2^{G_2,\mathbf{B}}(\rho_{G_1},\rho_{G_2}). \end{aligned}$$

Since every student is matched, all the $V_\emptyset$ are 0. We assumed $P_{\mathbf{A}} = -\infty$, if it is $P_{\mathbf{B}}$ instead that is $-\infty$ we obtain the same result as the problem is symmetric in **A** and **B**. Finally, all these quantities are constant in ρ_{G_1} and ρ_{G_2}: the only ones that are non-zero are of the type $\mathbb{P}_\rho(W_{\mathbf{C}} < P_{\mathbf{C}})$ and therefore constant in the correlation. ■

A.5 Proof of Proposition 4.2

Proof of Proposition 4.2. For this proof, to simplify computations, assume without loss of generality that grades follow a uniform distribution on $[0,1]$. Since the Deferred Acceptance algorithm only depends on ordinal comparisons, this assumption is indeed not restrictive and switching to a uniform distribution will greatly solving the solution of the market clearing equation. The students' grade vectors are therefore uniformly distributed along the diagonal of the square $[0,1]^2$. The cutoffs $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ belong to $[0,1]$, and the metrics are given by:

$$\begin{aligned} V_1^{\mathbf{A}} &= 1 - P_{\mathbf{A}}, & V_1^{\mathbf{B}} &= 1 - P_{\mathbf{B}}, \\ V_2^{\mathbf{A}} &= \max(P_{\mathbf{A}} - P_{\mathbf{B}}, 0), & V_2^{\mathbf{B}} &= \max(P_{\mathbf{B}} - P_{\mathbf{A}}, 0), \\ V_\emptyset^{\mathbf{A}} &= \min(P_{\mathbf{A}}, P_{\mathbf{B}}), & V_\emptyset^{\mathbf{B}} &= \min(P_{\mathbf{A}}, P_{\mathbf{B}}). \end{aligned} \tag{17}$$

Therefore, the market clearing equation is

$$\begin{cases} \beta(1 - P_{\mathbf{A}}) + (1-\beta)\max(P_{\mathbf{B}} - P_{\mathbf{A}}, 0) & = \alpha_{\mathbf{A}}, \\ \beta\max(P_{\mathbf{A}} - P_{\mathbf{B}}, 0) + (1-\beta)(1 - P_{\mathbf{B}}) & = \alpha_{\mathbf{B}}. \end{cases}$$

Assume that $P_{\mathbf{B}} \geq P_{\mathbf{A}}$. Then we have

$$\begin{cases} \beta(1-P_{\mathbf{A}}) + (1-\beta)(P_{\mathbf{B}} - P_{\mathbf{A}}) & = \alpha_{\mathbf{A}}, \\ (1-\beta)(1-P_{\mathbf{B}}) & = \alpha_{\mathbf{B}}, \end{cases} \tag{18}$$

which is equivalent to

$$\begin{cases} P_{\mathbf{A}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}, \\ P_{\mathbf{B}} = 1 - \frac{\alpha_{\mathbf{B}}}{1-\beta}. \end{cases}$$

Moreover, the assumption $P_{\mathbf{B}} \geq P_{\mathbf{A}}$ implies that $\beta \leq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}}}$. $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are well-defined, that is, they are in $[0,1]$. For $P_{\mathbf{A}}$, this follows from the assumption $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$, and for $P_{\mathbf{B}}$ it is implied by the relation $\beta \leq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}}}$. If $P_{\mathbf{A}} \geq P_{\mathbf{B}}$ instead, we have:

$$\begin{cases} P_{\mathbf{A}} = 1 - \frac{\alpha_{\mathbf{A}}}{\beta}, \\ P_{\mathbf{B}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}. \end{cases} \tag{19}$$

Similarly, this implies that $\beta \geq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}}}$, and using this we can verify that $P_{\mathbf{A}}, P_{\mathbf{B}} \in [0,1]$.

We can then conclude that if $\beta \leq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}}}$, then

$$\begin{aligned} V_1^{\mathbf{A}} &= \alpha_{\mathbf{A}} + \alpha_{\mathbf{B}}, & V_1^{\mathbf{B}} &= \frac{\alpha_{\mathbf{B}}}{1-\beta}, \\ V_2^{\mathbf{A}} &= 0, & V_2^{\mathbf{B}} &= \alpha_{\mathbf{A}} - \frac{\beta}{1-\beta}\alpha_{\mathbf{B}}, \\ V_\emptyset^{\mathbf{A}} &= 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}, & V_\emptyset^{\mathbf{B}} &= 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}; \end{aligned}$$

and if $\beta \geq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}}+\alpha_{\mathbf{B}}}$, then

$$\begin{aligned} V_1^{\mathbf{A}} &= \frac{\alpha_{\mathbf{A}}}{\beta}, & V_1^{\mathbf{B}} &= \alpha_{\mathbf{A}} + \alpha_{\mathbf{B}}, \\ V_2^{\mathbf{A}} &= \alpha_{\mathbf{B}} - \frac{1-\beta}{\beta}\alpha_{\mathbf{A}}, & V_2^{\mathbf{B}} &= 0, \\ V_\emptyset^{\mathbf{A}} &= 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}, & V_\emptyset^{\mathbf{B}} &= 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}. \end{aligned}$$

This is obtained by replacing in (17) the values of $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ found in (18) and (19). ■

A.6 Proof of Proposition 4.3

Proof of Proposition 4.3. As in the proof of Proposition 4.2, we assume for this proof that the grades follow a uniform distribution on $[0,1]$; which is without loss of generality. Then the grades at colleges **A** and **B** are independent random variables with a uniform distribution over [0, 1]. Students grade vectors are thus uniformly distributed on the whole area of the square $[0,1]^2$. Therefore the metrics as functions of $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are:

$$\begin{array}{ll} V_1^{\mathbf{A}} = 1 - P_{\mathbf{A}}, & V_1^{\mathbf{B}} = 1 - P_{\mathbf{B}}, \\ V_2^{\mathbf{A}} = P_{\mathbf{A}}(1 - P_{\mathbf{B}}), & V_2^{\mathbf{B}} = P_{\mathbf{B}}(1 - P_{\mathbf{A}}), \\ V_\emptyset^{\mathbf{A}} = P_{\mathbf{A}} P_{\mathbf{B}}, & V_\emptyset^{\mathbf{B}} = P_{\mathbf{A}} P_{\mathbf{B}}. \end{array} \tag{20}$$

The market clearing equation is:

$$\begin{aligned} & \begin{cases} \beta(1 - P_{\mathbf{A}}) + (1-\beta) P_{\mathbf{B}}(1 - P_{\mathbf{A}}) = \alpha_{\mathbf{A}}, \\ \beta P_{\mathbf{A}}(1 - P_{\mathbf{B}}) + (1-\beta)(1 - P_{\mathbf{B}}) = \alpha_{\mathbf{B}}, \end{cases} \\ \Leftrightarrow & \begin{cases} P_{\mathbf{B}} = 1 - \alpha_{\mathbf{B}} - \frac{\beta}{1-\beta}(1 - P_{\mathbf{A}} - \alpha_{\mathbf{A}}), \\ P_{\mathbf{A}} P_{\mathbf{B}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}, \end{cases} \\ \Leftrightarrow & \begin{cases} P_{\mathbf{B}} = 1 - \alpha_{\mathbf{B}} - \frac{\beta}{1-\beta}(1 - P_{\mathbf{A}} - \alpha_{\mathbf{A}}), \\ \frac{\beta}{1-\beta} P_{\mathbf{A}}^2 + (\frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}) P_{\mathbf{A}} - (1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}) = 0. \end{cases} \end{aligned} \tag{21}$$

Let $\zeta = \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}$ and $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}})}$. From (21) and the fact that $P_{\mathbf{A}} \geq 0$ we deduce that

$$\begin{cases} P_{\mathbf{A}} = \frac{1-\beta}{2\beta}(\Delta - \zeta), \\ P_{\mathbf{B}} = \frac{1}{2}(\Delta + \zeta). \end{cases}$$

Injecting this in equation (20) finally gives

$$\begin{array}{ll} V_1^{\mathbf{A}} = 1 - \frac{1-\beta}{2\beta}(\Delta - \zeta), & V_1^{\mathbf{B}} = 1 - \frac{1}{2}(\Delta + \zeta), \\ V_2^{\mathbf{A}} = \frac{1-\beta}{2\beta}(\Delta - \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_2^{\mathbf{B}} = \frac{1}{2}(\Delta + \zeta) - \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), \\ V_\emptyset^{\mathbf{A}} = \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), & V_\emptyset^{\mathbf{B}} = \frac{1-\beta}{4\beta}(\Delta^2 - \zeta^2), \end{array}$$

which concludes the proof of the proposition. ■