



HAL
open science

Statistical Discrimination in Stable Matching

Rémi Castera, Patrick Loiseau, Bary Pradelski

► **To cite this version:**

Rémi Castera, Patrick Loiseau, Bary Pradelski. Statistical Discrimination in Stable Matching. 2023.
hal-03672270v3

HAL Id: hal-03672270

<https://hal.science/hal-03672270v3>

Preprint submitted on 15 Jun 2023 (v3), last revised 5 Mar 2024 (v5)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical Discrimination in Stable Matching*

Rémi Castera ^{†,1}, Patrick Loiseau², and Bary S.R. Pradelski¹

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

²Inria, FairPlay Team, Palaiseau, France

First version: April 2022; This version: April 2023

Abstract

We study statistical discrimination in matching, where multiple decision-makers are simultaneously facing selection problems from the same pool of candidates. We propose a model where decision-makers observe different, but correlated estimates of each candidate’s quality. The candidate population consists of several groups that represent gender, ethnicity, or other attributes. The correlation differs across groups and may, for example, result from noisy estimates of candidates’ latent qualities, a weighting of common and decision-maker specific evaluations, or different admission criteria of each decision maker. We show that lower correlation (e.g., resulting from higher estimation noise) for one of the groups worsens the outcome for all groups, thus leading to efficiency loss. 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 at all depends on their group, and — against common intuition — the group that is subjected to lower correlation is better off. The resulting inequality reveals a novel source of statistical discrimination.

1 Introduction

Outcome inequalities for different demographic or social groups are ubiquitous, for example, in college admission, job assignment, or investment allocation. Arcidiacono et al. (2022) find that Asian-American applicants have lower admission chances at Harvard than white applicants for

*We would like to thank Nick Arnosti, Julien Combe, Faidra Monachou, Vitalii Emelianov, Simon Finster, Simon Jantschi and Jakob Weissteiner, as well as the participants of the 23rd ACM Conference on Economics and Computation (EC’ 22), the 12th Conference on Economic Design, and the 33rd Stony Brook International Conference on Game Theory for their remarks and advice, all of which greatly helped us improving this work. All remaining errors are ours. This work was partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003) and by the French National Research Agency (ANR) through grants ANR-19-CE48-0018 and ANR-20-CE23-0007.

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

a similar academic record, Niessen-Ruenzi and Ruenzi (2019) find significantly lower inflows in female-managed mutual funds than in male-managed mutual funds, and Bertrand and Mullainathan (2004) find race-based discrimination in callback decisions by job advertisers. Consequently, the sources of observed outcome inequalities and whether and how to employ affirmative action policies remain the subject of frequent and continued controversy and political debate. The US Supreme Court is currently re-evaluating the lawfulness of affirmative action policies in college admissions, as ruled in *Grutter v. Bollinger* in 2003. *Students for Fair Admissions Inc.* is challenging admissions programs at Harvard University and the University of North Carolina, which consider race as part of their efforts to obtain a diverse student body (Howe, 2022). A common thread that emerges from these debates is that the causes of outcome inequalities in matching markets are not well understood, making them difficult to address (Longhofer, 1995).

This paper examines how imperfect information that differs across applicants dependent on their demographic or social groups affects outcomes in matching markets. We study multiple decision-makers who must select a subset of applicants from an applicant pool. We depart from existing work by allowing for multiple decision-makers, e.g., multiple schools, companies, or investment funds. We show that a new source of outcome inequality arises, namely the correlation between the evaluations of different decision-makers of any given applicant.

The sources of outcome inequalities and discrimination — beyond intrinsic differences — can be grouped into two categories. On the one hand, *taste-based discrimination* posits that the decision-maker’s preferences are based on demographic or social groups, which leads to intentional discrimination (cf. Becker 1957). On the other hand, *statistical discrimination* posits that the decision-maker may be faced with imperfect information (that may vary across groups) and thus discrimination can arise unintentionally (cf. Phelps 1972; Arrow 1973). Falling in the latter category, Garg et al. (2021); Emelianov et al. (2020, 2022) study selection problems, where imperfect information is modeled via *differential variance*, that is, a decision-maker has a noisy estimate of the quality of candidates and the noise variance depends on the group of the candidate. They show that differential variance leads to outcome inequalities and study various affirmative action policies. Our work extends the latter studies on statistical discrimination from a single decision-maker selection problem to matching markets and we identify a matching-specific source of discrimination.

We consider the college-admissions problem (Gale and Shapley, 1962) and use the continuous framework by Azevedo and Leshno (2016). Multiple decision-makers each select applicants from the same pool. We consider stability as the solution concept. Each decision-maker gives each applicant a score, and the scores received by an applicant may be correlated, e.g., due to the latent quality of a student in the case of college admissions or due to similar risk-reward preferences of investors choosing between different funds. Notably, the effect of a specific form of correlation has been studied in school choice problems that emerges from tie-breaking (Abdulkadiroğlu et al., 2015; Ashlagi et al., 2019; Arnosti, 2022b). When students have the same ranking at a given college a

tie-breaking rule describes who should receive priority. Two natural choices are that each college breaks ties independently or colleges use a common order to break ties. Intuitively, the former leads to 0 correlation and the latter to correlation 1 (among those students for whom tie-breaking is required). We go beyond the latter papers by allowing any correlation level between -1 and 1 and, to our knowledge, are the first to consider different demographic or social groups with different correlation levels.

1.1 Our contribution

We study the effect of ranking correlation on stable matchings with different groups, and find that it impacts students’ welfare. Moreover, we show that different correlation levels across groups lead to outcome inequalities even when the rankings by each college are fair with respect to these groups.

Suppose an infinite population of students divided into groups \mathbf{G}_1 and \mathbf{G}_2 apply to colleges \mathbf{A} and \mathbf{B} . Groups can, for example, represent protected attributes, such as gender or race. Each college assigns a priority score to each student. A given student gets priority scores $W_{\mathbf{A}}$ at college \mathbf{A} and $W_{\mathbf{B}}$ at college \mathbf{B} . We propose an original model for the distributions of these scores, designed to study the correlation between the rankings made by the different colleges.

Our model allows for any intermediate level of correlation between two extreme cases: a unique ranking shared by every college, or statistically independent rankings made by each college. To isolate the effect of correlation from effects of other features of the priority distributions, we focus on the case where the marginals of $(W_{\mathbf{A}}, W_{\mathbf{B}})$ (i.e., the univariate distributions of $W_{\mathbf{A}}$ and $W_{\mathbf{B}}$) are the same for all candidates, but the correlation level depends on a candidate’s group—we term this *differential correlation*. Note that the effect of correlation is only present in matching problems, while it is not present in selection problems with a single decision-maker.

We use techniques from Azevedo and Leshno (2016) to characterize the stable matching (which is generically unique in the model with a continuum of students). We focus on ordinal welfare, and use as metrics the probability of a student obtaining the first choice, second choice, or staying unmatched, conditionally on their group and preferences. Our metric for aggregate efficiency of the matching is the total amount of students obtaining their first choice, and we measure the inequality between groups by the difference in their probability of not matching anywhere. We provide a theoretical justification for the choice of these metrics. Our main results summarize as follows:

1. The probability that a student is assigned to their first choice is independent of the student’s group.
2. The efficiency of the matching is increasing in the correlation of either group: increasing the correlation level of one or both groups increases the amount of students getting their first choice without increasing the amount staying unmatched.

3. With different correlation levels, inequality arises: students from the high-correlation group have a higher probability of remaining unmatched.

Note that the second result gives a direct link between correlation and students' welfare. Building on this we further show that the probability of staying unmatched increases for the group whose correlation increases, and that a given global efficiency level can be reached by different pairs of correlations, yielding different levels of inequality. We also use our results to recover existing results in matching theory, such as tie-breaking results by Ashlagi et al. (2019), Ashlagi and Nikzad (2020) and Arnosti (2022b). Note that our results imply that it is — against common intuition — advantageous to belong to the low-correlation group. An explanation is that with low correlation, the fact that a student gets rejected from their first choice has a lower impact on their chances at the other college. In our model, colleges have no preference for either group as each college has the same score distribution for both groups. Nevertheless, we find inequalities in the matching outcomes, thus identifying a novel source of statistical discrimination. Note that the correlation difference can have many sources and our findings thus have a different flavor than prior identified statistical discrimination, which is usually arising from the varying accuracy of colleges when assessing candidates' qualities.

Our work is the first to investigate the effect of correlation on outcome inequalities in the context of matching. We show that group-dependent correlation between colleges' rankings plays an important role in leading to unequal outcomes for different demographic groups. This is the case even as colleges have fair rankings, that is, all demographic groups are represented in each college's ranking as they are in the total application population.

Throughout the paper, we use the example of college admission, and use the terminology of *priority scores* to denote the value assigned by a college to a student in order to rank them. Our model and results, however, are generic and apply to any matching problem without transfers, cf. examples in the introduction. Moreover, since we use a definition of correlation based on rankings, our results remain valid when colleges assign ordinal rather than cardinal scores.

1.2 Related literature

Matching. The college admission problem, i.e, how a centralized authority can assign prospective students to colleges given each student's preferences and colleges' priorities over students and capacities such that the outcome is stable, was introduced by Gale and Shapley (1962) and then studied by many authors, see Roth and Sotomayor (1992) for the standard text book. A variant of this model where colleges do not have priorities over students is commonly called the school choice problem, and has been investigated in (Balinski and Sonmez 1999, Abdulkadiroğlu and Sönmez 2003, Abdulkadiroğlu 2005, Ergin and Sönmez 2006, Yenmez 2013). The idea of considering a continuum of students and a finite number of colleges has previously been exploited. Chade et al. (2014) use it for a model of noisy priorities, Abdulkadiroğlu et al. (2015) show the inefficiency of Deferred Acceptance for cardinal utilities, and Azevedo and Leshno (2016) develop a supply and demand

framework allowing to easily analyze the quality of a matching and deriving comparative statics. We shall follow Azevedo and Leshno (2016) in analyzing our model as a supply and demand problem. Arnosti (2022a) recently proposed a similar framework, giving better approximations when applied to study finite markets.

Matching with correlated types. We study matching in the presence of correlation between the priority scores given by each college to a given student. A special case of this problem has been studied for centralized school choice problems, where a lot of students have the same priority and therefore ties are broken at random.¹ In Ashlagi et al. (2019); Ashlagi and Nikzad (2020); Arnosti (2022b), the authors compare the welfare of students in two settings: either one common lottery is used by all colleges, or all colleges draw independent lotteries. In our model, this corresponds to correlation 1 or 0, respectively, and our results nest these prior papers. Another line of work has considered correlation between other features, e.g., correlation between students' preferences and priorities. In this context, Brilliantova and Hosseini (2022) focus on one-to-one matching and try to find a matching that does not favor one side over the other, while Che and Tercieux (2019) and Leshno and Lo (2020) study the stability-efficiency trade-off by comparing Deferred Acceptance and Top Trading Cycles (see Shapley and Scarf 1974), and how the magnitude of this trade-off depends on the correlation between agents preferences and priorities.

Fairness. The computer science literature on fairness in selection problems was initiated by Kleinberg and Raghavan (2018) who study the effect of bias and the efficiency of affirmative action policies. Emelianov et al. (2020, 2022) and Garg et al. (2021) study statistical discrimination. Candidates have a latent quality, and a 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 decision-makers 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, 2022; Delacrétaz et al. 2020; Krishnaa et al. 2019; Dur et al. 2020). These works, however, focus on finding stable matchings under some constraints, accounting for different fairness notions. In contrast, we aim to explain outcome inequalities that naturally occur in stable matchings without constraints. Recently, reducing outcome inequalities 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. (2022) show that a fair ranking does not necessarily lead to a fair matching, a conclusion that can also be drawn from our results. Finally, Monachou and Ashlagi (2019) and Che et al. (2020) both study outcome inequality in online markets using ratings based on reviews.²

¹The implications of this feature on students' welfare have been studied by Erdil and Ergin (2008); Abdulkadiroğlu et al. (2009) and Abdulkadiroğlu et al. (2015).

²For the general question of bias and fairness in algorithms and machine learning, see, for example, Dwork et al. (2012); Hardt et al. (2016); Zafar et al. (2017); Blum and Stangl (2020); Orwat (2020); Barocas et al. (2019);

1.3 Outline

The remainder of this paper is organized as follows. Section 2 introduces the model and the concept of differential correlation, as well as several motivating examples. Section 3 presents the theoretical tools we use to study matching outcomes. In Section 4 we present our main results. Section 5 treats some special cases and applications: capacity excess, a single group, and tie breaking. Finally, Section 6 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 long proofs of our results in the body of the paper and defer longer and more technical details to Appendix A.4. Further theoretical concepts used only in proofs are presented in Appendices A.1, A.2 and A.3.

2 Setup

We study the effect of correlation between colleges' rankings of students on the matching. To isolate this effect from other sources of inequalities, we assume that each college's ranking is fair, that is, in any segment of the ranking, all demographic groups are represented in the same proportion than in the total applicant population. We formally define *correlation* in our context in Section 2.2.

A table of notation is provided in Appendix A.1.1 for readers' convenience.

2.1 Model

The college admission problem (Gale and Shapley 1962) describes a many-to-one matching market with non-transferable utility. We consider a variant of this model with a continuum of students (cf. Chade et al. 2014; Azevedo and Leshno 2016).

Students and colleges.

Let \mathbf{A} and \mathbf{B} be two colleges to which a continuum unit mass of students, \mathbf{S} , is to be matched. The mass of a subset of \mathbf{S} is measured with a function η .³ Colleges have maximum capacities of the mass of students they can admit, $(\alpha_{\mathbf{A}}, \alpha_{\mathbf{B}}) := \boldsymbol{\alpha} \in (0, 1]^2$. The students are divided into two groups \mathbf{G}_1 and \mathbf{G}_2 , with a fraction $\gamma \in [0, 1]$ of students belonging to \mathbf{G}_1 and a fraction $1 - \gamma$ belonging to \mathbf{G}_2 . We denote the group to which a student $s \in \mathbf{S}$ belongs by $\mathbf{G}(s)$.

Preferences.

Students have strict preferences over colleges, and the amount of students preferring college \mathbf{A} might differ between groups: among group \mathbf{G}_1 , a share $\beta_{\mathbf{G}_1} \in [0, 1]$ prefer college \mathbf{A} to college \mathbf{B} , the remaining $1 - \beta_{\mathbf{G}_1}$ prefer \mathbf{B} . When student s prefers college \mathbf{A} to college \mathbf{B} , we write $\mathbf{A} \succ_s \mathbf{B}$, and

Chouldechova and Roth (2020); Finocchiaro et al. (2021); Kleinberg and Raghavan (2021); Mehrabi et al. (2021).

³The formal definition of this measure is deferred to Appendix A.1.2 and uses elements defined later in Section 2.1.

vice versa. Note that β is a share that is conditional on the group, and not a mass: for instance, $\eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}\}) = \gamma\beta_{\mathbf{G}_1}$. Similarly in \mathbf{G}_2 a share $\beta_{\mathbf{G}_2} \in [0, 1]$ prefer \mathbf{A} and $1 - \beta_{\mathbf{G}_2}$ prefer \mathbf{B} . All students prefer attending some college to remaining unmatched.

Priorities.

Each college assigns a priority score to each student, the higher the better. Each student s thus is assigned a vector of priority scores $(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s)$. This means that college \mathbf{C} prefers $s \in \mathbf{S}$ to s' if and only if $W_{\mathbf{C}}^s > W_{\mathbf{C}}^{s'}$.

The (marginal) distribution of scores given by college \mathbf{C} is described by a probability density function (pdf) $f_{\mathbf{C}}$ defined over the support $\mathbf{I}_{\mathbf{C}} \subseteq \mathbb{R}$. Let $\mathbf{I} = \mathbf{I}_{\mathbf{A}} \times \mathbf{I}_{\mathbf{B}}$. We will sometimes require $\mathbf{I}_{\mathbf{A}}$ and $\mathbf{I}_{\mathbf{B}}$ to be intervals, resulting in \mathbf{I} being convex. Denote by $\mathbf{I}_{\mathbf{A}}^l$ and $\mathbf{I}_{\mathbf{A}}^u$ the lower and upper bounds of $\mathbf{I}_{\mathbf{A}}$, $\mathbf{I}_{\mathbf{B}}^l$ and by $\mathbf{I}_{\mathbf{B}}^u$ those of $\mathbf{I}_{\mathbf{B}}$. These bounds might be finite or not.

We make two assumption on these marginal distributions: First, that the preferences of a student are independent from their priority vector, i.e., the score distribution of students who prefer \mathbf{A} is the same as the one of students preferring \mathbf{B} . Second, we assume that the distribution $f_{\mathbf{A}}$ is the same for both groups and the distribution $f_{\mathbf{B}}$ is also the same for both groups. The joint distribution of the priority vector may however differ. This implies that each college has a fair ranking: for any $\omega \in [0, 1]$, the proportion of students from a given group in the top ω fraction of the ranking is equal to the proportion of this group in the total population.

Differential correlation.

Further, we assume that the joint distributions of grades belongs to a family of probability density functions $(f_{\theta} : \mathbf{I} \rightarrow \mathbb{R})_{\theta \in \Theta}$. Here, θ is a real parameter that we will use to represent the level of correlation of the distribution, and Θ is the set of possible values for θ ; some of our results will require Θ to be an interval. For consistency with previous definitions and assumptions, we impose that $\forall \theta \in \Theta$, the marginals of f_{θ} are $f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ (i.e., they do *not* depend on θ). We assume that all students from group \mathbf{G}_1 have their priority vectors drawn from f_{θ_1} and students from \mathbf{G}_2 have their priority vector drawn from f_{θ_2} , with $\theta_1, \theta_2 \in \Theta$. We call this feature of the model *differential correlation*.⁴ Finally, we assume that all the distributions in $(f_{\theta})_{\theta \in \Theta}$ have full support over \mathbf{I} .

We denote by F_{θ} the cumulative density function (cdf) associated to the pdf f_{θ} , and $F_{\mathbf{A}}, F_{\mathbf{B}}$ the marginals' cdfs. We shall refer to the tuple $(\gamma, \beta_{\mathbf{G}_1}, \beta_{\mathbf{G}_2}, \alpha, f_{\theta_1}, f_{\theta_2})$ as a *college admission problem*.

Remark. *The assumptions that distributions admit a density and have full support, that the distribution family is parameterized by a scalar, and that the marginals remain the same for any θ , are not very restrictive. To support this assertion, and to make the notion of distribution family*

⁴This is in spirit of the notion of differential variance studied in Emelianov et al. (2022) and Garg et al. (2021)

more concrete, we present a few examples of classical distributions fitting our model, that we will refer to throughout the article.

1. *Standard bivariate Gaussian: the distribution family is the set of bivariate normal distributions with covariance matrix $\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. Here, the parameter is r and controls the covariance.*
2. *Gaussian copula: given two marginal distributions, define a distribution family by associating these marginals to a Gaussian copula with the same covariance matrix as above. Once again, the parameter is r .*
3. *Archimedean copulas: Archimedean copulas are a broad family of copulas, each element of this family being itself a parametric family of copulas with parameter θ . Thus, any choice of an Archimedean copula and a pair of marginals yields a valid distribution family for our model, with parameter θ . In particular, after fixing a copula, we could choose any marginals as long as their cdfs are continuous. This includes a very wide range of distributions, with bounded or unbounded support, light or heavy tail, etc.*

2.2 Correlation and the coherence assumption

We use a condition, namely coherence, on the family of distributions and show that whenever $(f_\theta)_{\theta \in \Theta}$ is coherent, there exists a bijection between θ and classical measures of correlation.

Definition 2.1 (Coherence). We say that $(f_\theta)_{\theta \in \Theta}$ is *coherent* if for all $(x, y) \in \overset{\circ}{\mathbf{I}}$, $F_\theta(x, y)$ is increasing in θ on Θ , where $\overset{\circ}{\mathbf{I}}$ is the interior of \mathbf{I} .

The following lemma states that under coherence θ is naturally interpreted as a measure of correlation.

Lemma 2.2. *If $(f_\theta)_{\theta \in \Theta}$ is coherent, then $\forall (x, y) \in \overset{\circ}{\mathbf{I}}$, $\mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y)$ and $\mathbb{P}(W_{\mathbf{A}} > x, W_{\mathbf{B}} > y)$ are both increasing in θ , while $\mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} > y)$ and $\mathbb{P}(W_{\mathbf{A}} > x, W_{\mathbf{B}} < y)$ are decreasing.*

The proof of the Lemma relies on the fact that the marginals do not depend on θ . Intuitively, when scores are highly correlated, when, for example, $W_{\mathbf{A}}$ is small, the other grade $W_{\mathbf{B}}$ is likely also small. The proof is provided in Appendix A.4.1.

We now compare θ to classical measures of correlation, namely *Spearman's correlation*, denoted ρ , and *Kendall's correlation*, denoted τ . Both these measures are defined on ordinal rankings and not on cardinal scores and are therefore invariant under monotonic transformations. This is suited to our model as stability in matching is an ordinal solution concept. For further details about classical correlation measures and their properties, see Appendix A.2.

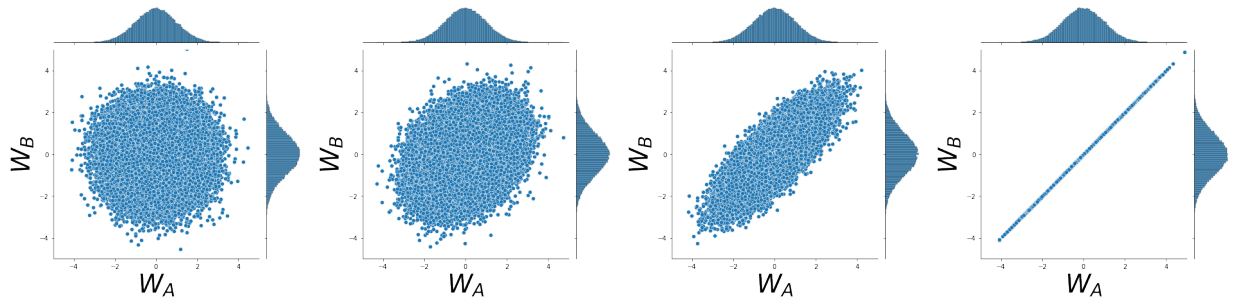


Figure 1: Gaussian priority distributions for $\theta = 0, 0.3, 0.8$, and 1.

Lemma 2.3 (Scarsini 1984, Theorems 4 and 5). *If $(f_\theta)_{\theta \in \Theta}$ is coherent, then Spearman's and Kendall's correlation coefficients ρ and τ are strictly increasing functions of θ .*

Lemma 2.3 shows that, under the coherence assumption, θ is in agreement with classical measures of correlation, justifying our choice to use it to describe the correlation of a distribution from $(f_\theta)_{\theta \in \Theta}$. Figure 1 illustrates the effect of correlation on the joint distribution, displaying bivariate Gaussian distributions with different values of θ , with θ being the anti-diagonal coefficient of the covariance matrix.

We finally introduce a technical assumption that will be required for some of our results, especially when considering comparative statics in θ .

Definition 2.4 (Differentiability). We say that $(f_\theta)_{\theta \in \Theta}$ is *differentiable* if for all $(x, y) \in \mathring{\mathbf{I}}$ and for all $\theta \in \mathring{\Theta}$, $f_\theta(x, y)$ is differentiable in θ .

The coherence and differentiability assumptions are not particularly restrictive: consider the examples of distribution families in Remark 2.1. The bivariate Gaussian, the Gaussian copula, as well as most classical Archimedean copulas such as Clayton's, Gumbel's or Frank's, are all coherent and differentiable.

2.3 Motivating Examples

We consider three motivating examples where differential correlation naturally occurs.

2.3.1 Noisy estimates with normalization.

Assume that all students have a latent quality W , and their score $\widehat{W}_A, \widehat{W}_B$ are the sum of the latent quality and a noise term ε drawn independently at each college, that is,

$$\forall s \in \mathbf{S}, \text{ for } \mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}, \widehat{W}_\mathbf{C}^s = W^s + \varepsilon_\mathbf{C}^s.$$

Further, assume that the latent qualities of all students are drawn from a (group-independent) normal distribution, and that the noises are also normally distributed and depend on the group:

$$\forall s \in \mathbf{S}, \text{ for } \mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}, W^s \sim \mathcal{N}(0, \chi^2), \varepsilon_{\mathbf{C}}^s \sim \mathcal{N}(0, \sigma_{G(s)}^2).$$

The fact that the noise’s variance is different for the two groups can be interpreted as colleges having different accuracies when evaluating students from different groups. Consider two examples: First, \mathbf{G}_1 could consist of students from well-known high schools, which colleges can evaluate well since they have a lot of applicants from there each year; and \mathbf{G}_2 could consist of students from unknown high schools, for which colleges do not have a lot of prior information. Second, \mathbf{G}_1 could consist of students from a demographic majority group and \mathbf{G}_2 from a minority groups, defined by sensitive attributes such as gender, ethnicity, or social class.

Suppose that colleges know each student’s group, and are aware of the difference that exists in noise variance across groups. Further, assume that colleges implement equal opportunity policies, that is, ceteris paribus the probability for a student from group \mathbf{G}_1 to be admitted to the college is the same as for a student from group \mathbf{G}_2 .⁵ To do so while maximizing the expected quality of admitted candidates, it is optimal for a college to not change the order of grades within groups, but to only fit each group’s grades $\widehat{W}_{\mathbf{C}}$ to the same, standardized distribution. Thus, suppose that colleges divide each student’s score by the standard deviation of their group’s scores:

$$\forall s \in \mathbf{S}, \text{ for } \mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}, \widetilde{W}_{\mathbf{C}}^s = \frac{\widehat{W}_{\mathbf{C}}^s}{\sqrt{\chi^2 + \sigma_{G(s)}^2}}.$$

With these new standardized scores, the marginal priority distribution of each group is $\mathcal{N}(0, 1)$ at each college. The priority vectors then follow a centered bivariate normal distribution with variance 1 and a correlation that is different between the two groups: formally, if $\forall s \in \mathbf{S}, \text{ for } \mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}, W^s \sim \mathcal{N}(0, \chi^2), \varepsilon_{\mathbf{C}}^s \sim \mathcal{N}(0, \sigma_{G(s)}^2), \widehat{W}_{\mathbf{C}}^s = W^s + \varepsilon_{\mathbf{C}}^s$ and $\widetilde{W}_{\mathbf{C}}^s = \frac{\widehat{W}_{\mathbf{C}}^s}{\sqrt{\chi^2 + \sigma_{G(s)}^2}}$, then

$$\left(\widetilde{W}_{\mathbf{A}}^s, \widetilde{W}_{\mathbf{B}}^s \right) \sim \mathcal{N} \left((0, 0), \begin{pmatrix} 1 & r_{G(s)} \\ r_{G(s)} & 1 \end{pmatrix} \right), \quad (1)$$

with

$$r_{G(s)} = \frac{\chi^2}{\chi^2 + \sigma_{G(s)}^2}. \quad (2)$$

Equation (1) shows that this example is covered by our model. Here, the parameter is the noise variance, and, from Equation (2), it follows that correlation is a decreasing function of noise. This is expected: higher noise variance leads to lower correlation, and vice-versa. This also implies that the

⁵Without the equal opportunity assumption, colleges would compute the expected true qualities based on the different variances for each group, see Emelianov et al. (2022); Garg et al. (2021).

group with higher noise is the one with lower correlation.

2.3.2 Common technical test.

Suppose a firm has two branches **A** and **B** inside a country, each branch looking to hire new workers with fixed salaries (e.g., for entry-level positions). Candidates are either local (\mathbf{G}_1) or foreign (\mathbf{G}_2), and apply to the firm, declaring which branch they prefer. To assess candidates' qualities, the firm asks them to take part in a technical test, the result W^{tech} of which is used by both branches. Moreover, candidates are interviewed by each branch separately, leading to a score vector $(W_{\mathbf{A}}^{inter}, W_{\mathbf{B}}^{inter})$ for their performance at these interviews.⁶ The final score of a candidate at branch **C** is a weighted average of their two scores W^{tech} and $W_{\mathbf{C}}^{inter}$.

It is natural to assume that foreign candidates are more difficult to evaluate at the interview stage, due to, for example, language barriers or cultural differences. A firm may thus deem the technical score more informative for foreign candidates and put a higher weight on it than for local candidates. Therefore, foreign applicants' score vectors will be more correlated than local applicants' ones. Supposing that the final matching of workers to branches is stable and that W^{tech} and $W_{\mathbf{C}}^{inter}$'s distributions have the same marginals for both groups, our model fits.

2.3.3 Different criteria.

Suppose that colleges **A** and **B** have different criteria to assess applicants. College **A** ranks students by their grade in mathematics, $W_{\mathbf{A}}$, while college **B** ranks them by their grade in physics, $W_{\mathbf{B}}$. Further, students come from two high schools. In high school \mathbf{G}_1 , students have similar grades in mathematics and physics because physics is taught in a theoretical manner. On the other hand, in high school \mathbf{G}_2 physics is taught in a more experimental manner, therefore the grades in mathematics and physics are less correlated. Finally, if the grade distributions are different between the two high schools, suppose that colleges, as in the first example (Section 2.3.1) normalize them to avoid bias against one high school. Observe that the marginals are the same for both high schools and only the difference in correlation remains. Thus our model applies to this setting.

Note that the two latter examples do not feature the "classical" notion of statistical discrimination, since the scores are not noisy.

3 The Supply and Demand Framework

We now introduce the key elements of matching theory used throughout the paper. A matching μ is a mapping associating a student to a college (or themselves if they are unmatched) and a college to a subset of students. We use the common definition of stability: we say that a matching is stable if,

⁶Another example consists of a centralized college admission exam with a common written test and college-specific oral interviews.

for any student s that would prefer college \mathbf{C} to their current match, s has a lower score at \mathbf{C} than all the currently admitted students in that college. For the formal definitions of those notions and of the Deferred Acceptance Algorithm, see Appendix A.3.

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 3.1 (Cutoffs and demand). If μ is a stable matching, define the *cutoff* at $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$ as $P_{\mathbf{C}} := \inf\{W_{\mathbf{C}}^s : \mu(s) = \mathbf{C}\}$. Given $\mathbf{P} = (P_{\mathbf{A}}, P_{\mathbf{B}})$, we call the *demand* of student s , denoted $D_s(\mathbf{P}) \in \{\mathbf{A}, \mathbf{B}\} \cup \{s\}$, the college they prefer among those where their score is above the cutoff, or themselves if their score does not exceed the cutoff at any college. The *aggregate demand* at college \mathbf{C} is the mass of students demanding it: $D_{\mathbf{C}}(\mathbf{P}) = \eta(\{s : D_s(\mathbf{P}) = \mathbf{C}\})$.

The cutoff of a college represents the score above which a student who applies gets admitted. Recall that $\mathbf{I} = \mathbf{I}_{\mathbf{A}} \times \mathbf{I}_{\mathbf{B}}$ is the support of f_{θ} , $\mathbf{I}_{\mathbf{A}}^l$ and $\mathbf{I}_{\mathbf{A}}^u$ the lower and upper bounds of $\mathbf{I}_{\mathbf{A}}$, $\mathbf{I}_{\mathbf{B}}^l$ and $\mathbf{I}_{\mathbf{B}}^u$ those of $\mathbf{I}_{\mathbf{B}}$. If $P_{\mathbf{C}} = \mathbf{I}_{\mathbf{C}}^l$, then college \mathbf{C} rejects no one. In our model, the “supply” associated to this demand is simply the capacity of each college. Now consider the equilibria of this problem:

Definition 3.2 (Market clearing). The cutoff vector \mathbf{P} is **market clearing** if for $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$, $D_{\mathbf{C}}(\mathbf{P}) \leq \alpha_{\mathbf{C}}$, with equality if $P_{\mathbf{C}} > \mathbf{I}_{\mathbf{C}}^l$.

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 the constraint is reached at both colleges, i.e., when $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$, the system

$$\mathbf{D}(\mathbf{P}) = \boldsymbol{\alpha} \tag{3}$$

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

The following Lemma from Azevedo and Leshno (2016) establishes the link between market-clearing cutoffs and stable matchings:

Lemma 3.3 (Azevedo and Leshno 2016, Lemma 1).

1. If μ is a stable matching, the associated cutoff vector \mathbf{P} is market-clearing;
2. If \mathbf{P} is market-clearing, we define μ such that for all $s \in \mathbf{S}$, $\mu(s) = D_s(\mathbf{P})$. Then μ is stable.

This allows to analyze stable matchings by studying the cutoffs of each college. Figure 2 illustrates the link between the cutoffs and the matching: students who prefer \mathbf{A} get admitted there if and only if their score $W_{\mathbf{A}}$ is higher than the cutoff $P_{\mathbf{A}}$. Otherwise, they get admitted to \mathbf{B} if their score $W_{\mathbf{B}}$ is higher than $P_{\mathbf{B}}$, and stay unmatched if it is not. The situation is symmetric for students preferring college \mathbf{B} .

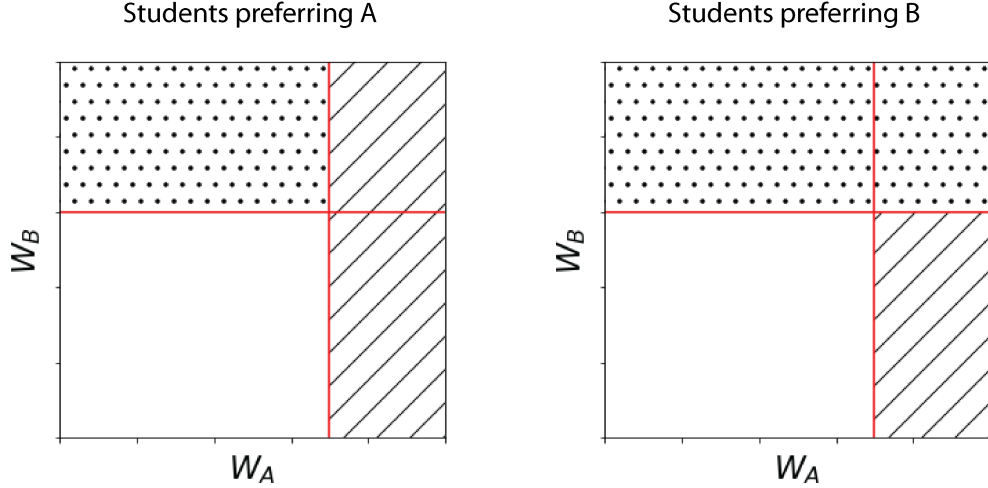


Figure 2: Illustration of the match of students depending on their preferences. Students in the hashed area are matched to college **A**, those in the dotted area to college **B**, and those in the white area remain unmatched.

In the continuous college admissions problem the same authors show that there is a unique stable matching.

Lemma 3.4 (Special case of Azevedo and Leshno 2016, Theorem 1). *For any college admission problem $(\gamma, \beta_{\mathbf{G}_1}, \beta_{\mathbf{G}_2}, \boldsymbol{\alpha}, f_{\theta_1}, f_{\theta_2})$, there exists a unique stable matching.*

Note that the original theorem specifies conditions on the distribution of students' types, such as being continuous and having full support, which holds in our definition of a college admission problem. Unlike the finite case where typically several stable matchings exist, 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 cutoff vector \mathbf{P} as the one 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 to college \mathbf{C} in the unique stable matching.

Azevedo and Leshno (2016) further show that the stable matching varies continuously in the parameters of the problem and 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 the same parameters. The latter result justifies the approximation of large finite instances through their limit.⁷

⁷For a better approximation for instances with a small number of students, Arnosti (2022a) proposes a related framework.

4 Main results

This section contains our main results on the impact of differential correlation on the properties of stable matchings.

4.1 Welfare metrics

In selection problems, inequalities between groups are measured by the proportion of admitted candidates in each group. In a matching setting, the situation is more complex: on the one hand, one group might have a higher proportion of unmatched students than the other, but on the other hand, the proportion of students getting their first choice might also differ. If all students in a group get their first choice and all students in the other get their second choice, the matching may be deemed unfair. In this section, we define metrics that allow us to quantify the satisfaction of students from each group.

Consider an individual's likelihood of getting their first choice in a stable matching defined by cutoffs \mathbf{P} . Denote the probability for a student to get their first choice, conditionally on belonging to group \mathbf{G}_1 and preferring \mathbf{A} to \mathbf{B} , by $V_1^{\mathbf{G}_1, \mathbf{A}}(\mathbf{P})$.

Definition 4.1 (Welfare metrics). Under a stable matching μ with cutoffs \mathbf{P} , define $V_1^{\mathbf{G}_1, \mathbf{A}}(\mathbf{P})$, $V_1^{\mathbf{G}_1, \mathbf{B}}(\mathbf{P})$, $V_1^{\mathbf{G}_2, \mathbf{A}}(\mathbf{P})$ and $V_1^{\mathbf{G}_2, \mathbf{B}}(\mathbf{P})$ as the proportion of students from each group-preference profile who get their first choice. Formally,

$$\begin{aligned} V_1^{\mathbf{G}_1, \mathbf{A}}(\mathbf{P}) &:= \frac{1}{\gamma\beta_{\mathbf{G}_1}} \eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\}), \\ V_1^{\mathbf{G}_1, \mathbf{B}}(\mathbf{P}) &:= \frac{1}{\gamma(1-\beta_{\mathbf{G}_1})} \eta(\{s \in \mathbf{G}_1 : \mathbf{B} \succ_s \mathbf{A}, \mu(s) = \mathbf{B}\}), \\ V_1^{\mathbf{G}_2, \mathbf{A}}(\mathbf{P}) &:= \frac{1}{(1-\gamma)\beta_{\mathbf{G}_2}} \eta(\{s \in \mathbf{G}_2 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\}), \\ V_1^{\mathbf{G}_2, \mathbf{B}}(\mathbf{P}) &:= \frac{1}{(1-\gamma)(1-\beta_{\mathbf{G}_2})} \eta(\{s \in \mathbf{G}_2 : \mathbf{B} \succ_s \mathbf{A}, \mu(s) = \mathbf{B}\}). \end{aligned}$$

Those metrics can be thought of in two ways: they are the relative masses of students getting their first choice among those with the same group and preferences, or equivalently, they are the probabilities of a randomly drawn student to get their first choice conditionally on their group and favorite college.⁸

We define similarly the proportions of students getting their second choice or staying unmatched, writing respectively V_2 and V_\emptyset instead of V_1 . For example, $V_2^{\mathbf{G}_1, \mathbf{A}}(\theta_1, \theta_2)$ is the probability of a

⁸The choice of conditioning over preferences stems from the following observation: if students from group \mathbf{G}_1 all prefer college \mathbf{A} , but only half of the students from group \mathbf{G}_2 prefer \mathbf{A} , \mathbf{A} has a low capacity and \mathbf{B} a large one, then very few students from \mathbf{G}_1 will get their first choice while half of the students from \mathbf{G}_2 are likely to get their first choice since it is a less demanded college. Students' satisfaction might differ across groups only due to their own preferences, and not because of differential correlation.

student from group \mathbf{G}_1 who prefers college \mathbf{A} to get their second choice, i.e., \mathbf{B} .

We now provide expressions for these metrics as functions of the cutoffs $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$.

Lemma 4.2. *Let $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$ be a college, $\bar{\mathbf{C}}$ be the other college, and $\mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\}$ be a group. For a stable matching defined by the cutoffs \mathbf{P} , we have:*

$$V_1^{\mathbf{G}, \mathbf{C}}(\mathbf{P}) = \mathbb{P}_{\theta_{\mathbf{G}}}(W_{\mathbf{C}} \geq P_{\mathbf{C}}) = \int_{P_{\mathbf{C}}}^{\infty} f_{\mathbf{C}}(x) dx, \quad (4)$$

$$V_2^{\mathbf{G}, \mathbf{C}}(\mathbf{P}) = \mathbb{P}_{\theta_{\mathbf{G}}}(W_{\mathbf{C}} < P_{\mathbf{C}}, W_{\bar{\mathbf{C}}} \geq P_{\bar{\mathbf{C}}}) = \int_{-\infty}^{P_{\mathbf{C}}} \int_{P_{\bar{\mathbf{C}}}}^{\infty} f_{\theta_{\mathbf{G}}}(x, y) dx dy, \quad (5)$$

$$V_{\emptyset}^{\mathbf{G}, \mathbf{C}}(\mathbf{P}) = \mathbb{P}_{\theta_{\mathbf{G}}}(W_{\mathbf{C}} < P_{\mathbf{C}}, W_{\bar{\mathbf{C}}} < P_{\bar{\mathbf{C}}}) = \int_{-\infty}^{P_{\mathbf{C}}} \int_{-\infty}^{P_{\bar{\mathbf{C}}}} f_{\theta_{\mathbf{G}}}(x, y) dx dy. \quad (6)$$

Lemma 4.2 allows to compare probabilities of admission of different types of students, and derive comparative statics with respect to differential correlation. The proof is provided in Appendix A.4.2.

Regarding the probability of staying unmatched, we can derive a simple yet important result.

Lemma 4.3. *The probability that a student remains unmatched is independent of their preference. Moreover, the total mass of unmatched students is constant in both groups' correlation levels. Formally, let \mathbf{P} be associated to a stable matching. Then $V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}(\mathbf{P}) = V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}(\mathbf{P})$ and $V_{\emptyset}^{\mathbf{G}_2, \mathbf{A}}(\mathbf{P}) = V_{\emptyset}^{\mathbf{G}_2, \mathbf{B}}(\mathbf{P})$; and $\eta(\{s \in \mathbf{S} : \mu(s) = \emptyset\}) = \max(0, 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}})$ (which does not depend on θ_1 and θ_2).*

Proof. It is sufficient to notice that Equation (6) is symmetric in \mathbf{C} and $\bar{\mathbf{C}}$ to obtain the first part of the lemma. The second one comes from the fact that either there is excess capacity and everyone is matched, or both colleges are full and the mass of matched students is the sum of the capacities. ■

Remark. *The first part of the lemma could also be derived from the strategy-proofness for students of student-proposing Deferred Acceptance (Roth 1985). Indeed, the fact that students cannot improve their outcome by modifying the order of their preferences implies that them being unmatched or not does not depend on which college they reported to be their first choice.*

From now on, we will use the notation $V_{\emptyset}^{\mathbf{G}_1}$ and $V_{\emptyset}^{\mathbf{G}_2}$ since these quantities do not depend on the preference of students. By Lemma 3.4, for a given college admission problem the stable matching is unique, and so is the vector of market clearing cutoffs. We can therefore, to focus on the influence of θ and consider all other parameters as exogenous, write $V_i^{\mathbf{G}, \mathbf{C}}(\theta_1, \theta_2)$ instead of $V_i^{\mathbf{G}, \mathbf{C}}(\mathbf{P})$ for $i \in \{\emptyset, 1, 2\}$, $\mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\}$, $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$. When there is no ambiguity, we also omit the dependence on (θ_1, θ_2) and write $V_i^{\mathbf{G}, \mathbf{C}}$ instead.

We now define two global metrics, i.e., metrics that are not conditioned on the groups and preferences of student, namely efficiency and inequality:

Definition 4.4 (Efficiency and Inequality). Define the *efficiency* $V_1(\theta_1, \theta_2)$ of a matching as the proportion of students getting their first choice, and the *inequality* $L(\theta_1, \theta_2)$ of a matching as the difference of the probability of staying unmatched between the two groups:

$$\begin{aligned} V_1(\mathbf{P}) &= \eta(\{s \in \mathbf{S} : \mu(s) = \mathbf{C} \text{ and } \mathbf{C} \succ_s \bar{\mathbf{C}}\}) \\ &= (\gamma\beta_{\mathbf{G}_1} + (1 - \gamma)\beta_{\mathbf{G}_2})V_1^{\mathbf{A}} + (\gamma(1 - \beta_{\mathbf{G}_1}) + (1 - \gamma)\beta_{\mathbf{G}_2})V_1^{\mathbf{B}} \end{aligned} \quad (7)$$

$$L(\theta_1, \theta_2) = |V_\theta^{\mathbf{G}_1} - V_\theta^{\mathbf{G}_2}|. \quad (8)$$

By Lemma 4.3 the mass of unmatched students is constant. The mass of matched student is therefore also constant, and matched students get either their first or second choice, so it is clearly desirable to maximize the mass of students getting their first choice V_1 , which is equivalent to minimizing the amount mass of students getting their second choice. Inequality can be measured by the difference in unassigned students as we shall see in Theorem 4.5 that the probability to receive one's first choice is the same for both groups.

4.2 Impact of differential correlation

We now study the influence of differential correlation on the previously defined metrics. First, we show that differential correlation has no effect on the probability of getting one's first choice.

Theorem 4.5. *For any college admission problem, in any stable matching \mathbf{P} , the probability that a student gets their first choice is independent of the group they belong to. Formally, $V_1^{\mathbf{G}_1, \mathbf{A}}(\mathbf{P}) = V_1^{\mathbf{G}_2, \mathbf{A}}(\mathbf{P})$ and $V_1^{\mathbf{G}_1, \mathbf{B}}(\mathbf{P}) = V_1^{\mathbf{G}_2, \mathbf{B}}(\mathbf{P})$.*

Proof. The result follows directly by applying (4) to both groups, and by observing that the integral in (4) does not depend on the parameter θ and hence does not depend on the group a student belongs to. ■

This result, albeit simple, is an important property of the model, and a direct consequence of our choice to keep the marginals identical for both groups. A student, given the college they prefer, has the same probability of getting admitted there independently of the group they belong to — so differential correlation has no effect on this metric. From now on we may refer to the first choice probabilities as $V_1^{\mathbf{A}}$ and $V_1^{\mathbf{B}}$, since they are the same for both groups.

Theorem 4.5 also justifies the choice of L as a measure of inequality: the proportion of students getting their first choice is the same in both groups, so students who do not get their first choice either get their second one or remain unmatched. Consequently, when the proportion of unmatched students is higher in one group than the other then the matching is unequal.⁹

⁹Another choice could have been to compare the proportions of students getting their second choice in each group, however Theorem 4.5 implies that this quantity is equal to L .

From now until the end of Section 4, we assume that the colleges joint capacity is constrained, i.e., $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. The unique stable matching is then given by solving the market clearing equation (3).

Consider a student's probability of staying unmatched, depending on their group.

Theorem 4.6. *For any college admission problem, suppose that $(f_{\theta})_{\theta \in \Theta}$ is coherent and capacity is constrained, $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. The stable matching exhibits inequality, $L(\theta_1, \theta_2) > 0$, if and only if $\theta_1 \neq \theta_2$. Students from the group with higher correlation have a lower probability of getting their second choice and a higher probability of staying unmatched. Formally, if $\theta_1 < \theta_2$, for $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$, then $V_2^{\mathbf{G}_1, \mathbf{C}} > V_2^{\mathbf{G}_2, \mathbf{C}}$ and $V_{\emptyset}^{\mathbf{G}_1} < V_{\emptyset}^{\mathbf{G}_2}$.*

Proof. Suppose, without loss of generality, that $\theta_1 < \theta_2$. Then, using Lemma 4.2 to express V_2 and applying Lemma 2.2 we have

$$\begin{aligned} V_2^{\mathbf{G}_1, \mathbf{A}} &= \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) \\ &\geq \mathbb{P}_{\theta_2}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) \\ &= V_2^{\mathbf{G}_2, \mathbf{A}} \end{aligned} \tag{9}$$

and the same holds for $V_2^{\mathbf{G}_1, \mathbf{B}}$ and $V_2^{\mathbf{G}_2, \mathbf{B}}$. The inequality in (9) is strict if and only if $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are in the interior of the support \mathbf{I} . $P_{\mathbf{A}} = \mathbf{I}_A^u$ would happen if and only if $\alpha_{\mathbf{A}} = 0$, which we excluded in the definition of the model (as it would no longer be a matching model); and $P_{\mathbf{A}} = \mathbf{I}_A^l$ would mean that college \mathbf{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 F_{θ} being increasing in θ by the definition of coherence. ■

The probability of staying unmatched is different for students from different groups. This is in contrast to Theorem 4.5. Different levels of correlation lead to an unequal matching, and this inequality benefits students from the low correlation group who are at risk of staying unmatched due to a low score at one or both colleges. This happens at the expense of similar students from the other group. This may seem surprising, for example, when considering the motivating example from Section 2.3.1, where a higher correlation is associated with a lower noise in the quality estimation. However, 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 probability of also being rejected from their second choice. On the other hand, a student with low correlation who has been rejected from their first choice has a "more independent" second chance at the other college. Theorem 4.6 also shows that a fair ranking (each college's ranking is fair since the marginals are the same for both groups) does not imply an equitable — or fair — matching.

4.3 Comparative statics

With the understanding of the effect of differential correlation on the inequality between students of each group, we next analyze how correlation influences the overall quality of the matching. We first consider how the efficiency of the matching, i.e., the probability of getting one's first choice, varies when changing the correlation for one group (recall from Theorem 4.5 that this probability is the same for both groups).

Theorem 4.7. *Suppose that $(f_\theta)_{\theta \in \Theta}$ is coherent and differentiable, and that $\alpha_A + \alpha_B < 1$. Then the efficiency $V_1(\theta_1, \theta_2)$ of the matching is increasing in both arguments. More precisely, suppose that Θ, \mathbf{I}_A and \mathbf{I}_B are intervals, and $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$. Then $V_1^A(\theta_1, \theta_2)$ and $V_1^B(\theta_1, \theta_2)$ are differentiable and for any $C \in \{A, B\}$, for any $G \in \{G_1, G_2\}$,*

$$\frac{\partial V_1^C}{\partial \theta_G} > 0.$$

Proof sketch. The proof follows several steps. First, we rewrite the market-clearing equation (3) using Lemma 4.2. We obtain a system of two equations, where the variables are the cutoffs P_A and P_B , parameterized by θ_1, θ_2 . We then apply the Implicit Function Theorem to a mapping whose roots are the solution of this system of equations. We next compute the partial derivatives. To characterize the sign of the derivatives with respect to θ , we use the coherence assumption. Through analytical derivations, we can conclude. The details are provided in Appendix A.4.3. ■

Theorem 4.7 implies that, if the correlation decreases for the priorities of one of the groups, then both groups suffer the same decrease in first-choice admittance. Conversely, increasing the correlation for one group leads to an increase in first-choice admittance for both groups. Thus, both groups may benefit (at least as far as the probability of first choice assignment is concerned) from colleges working on increasing the correlation for either group. Consider the example presented in Section 2.3.1. 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 of the minority group.

When the correlation increases, students' score 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 3 with a bivariate normal distribution. If the cutoffs did not change, then the amount of unmatched students would increase, which is not possible as the capacities are constant and the matching would become unstable. Therefore, at least one of the cutoffs decreases (in fact, Theorem 4.7 implies that both are decreasing). As a consequence, the mass of matched students remains the same but more of them get their first choice. Notice that, as predicted by Theorem 4.5, the amount of students on the right of the red vertical line in the left figure is equal to the amount of students on the right of the green line in the right figure. Same goes

for the horizontal lines.

Remark. *The formal statement of Theorem 4.7 excludes the extremities of Θ . This assumption is made only to avoid the case where rankings are fully correlated, which would mean that f does not have full support. However, since V_1^A and V_1^B are continuous in θ , they are increasing on the whole interval Θ .*

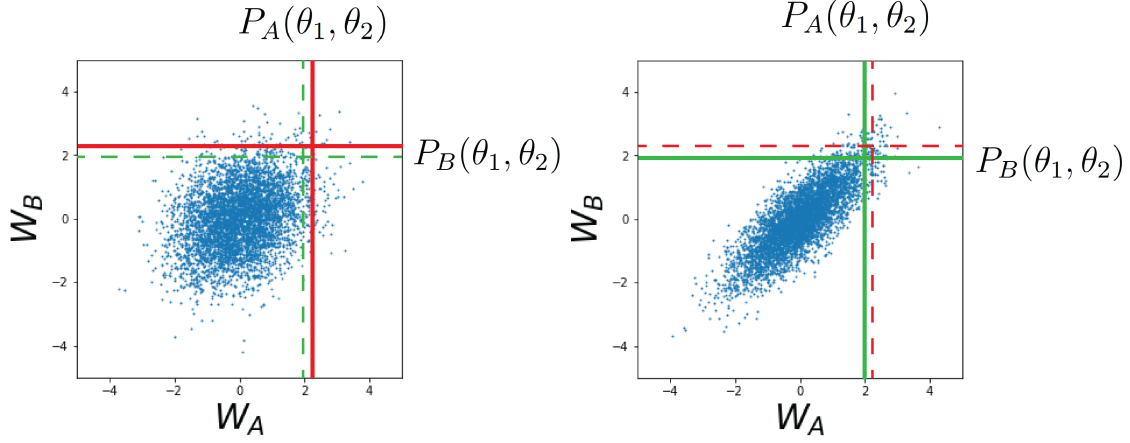


Figure 3: Illustration of the cutoffs decrease when correlation increases. P_A is represented as a vertical line and P_B as an horizontal one. Correlation of group \mathbf{G}_1 varies from $\theta_1 = 0.3$ (left, with cutoffs in red) to $\theta'_1 = 0.8$ (right, cutoffs in green). In each figure, the cutoffs corresponding to the other figure are plotted in dashes for comparison.

Theorem 4.7 allows us to derive the following corollary regarding a student's probability of remaining unmatched.

Corollary 4.8. *Suppose that $(f_\theta)_{\theta \in \Theta}$ is coherent and differentiable, assume Θ, \mathbf{I}_A and \mathbf{I}_B are intervals, $\theta_1, \theta_2 \in \Theta$ and $\alpha_A + \alpha_B < 1$. Then 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 $\mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\}$,*

$$\frac{\partial V_\emptyset^{\mathbf{G}}}{\partial \theta_{\mathbf{G}}} > 0 \quad \text{and} \quad \frac{\partial V_\emptyset^{\mathbf{G}}}{\partial \theta_{\overline{\mathbf{G}}}} < 0,$$

where $\overline{\mathbf{G}}$ is the other group. Moreover, the inequality of the matching is increasing in the high-correlation group's correlation level and decreasing in the low-correlation group's correlation level. Formally, assume $\theta_1 < \theta_2$. Then

$$\frac{\partial L(\theta_1, \theta_2)}{\partial \theta_1} < 0 \quad \text{and} \quad \frac{\partial L(\theta_1, \theta_2)}{\partial \theta_2} > 0.$$

Proof sketch. The proof leverages the partial derivatives with respect to the correlation coefficient computed in the proof of Theorem 4.7 and relies on the mass conservation to derive the sign of the partial derivatives of the V_\emptyset terms. The details are provided in Appendix A.4.4. ■

This result helps understand the influence of correlation on inequality. Recall that Theorem 4.6 states that the low-correlation group is advantaged compared to the other group; Corollary 4.8 completes it by showing that the extent of this advantage increases as the gap between correlation levels widens. Linking this to Theorem 4.5, we conclude that increasing the correlation level of a group hurts the intermediate students of this group while benefiting all good students (those who get either their first or second choice due to high grades at both colleges) as well as intermediate students of the other group.

In a situation with only one group, Theorem 4.7 states that the efficiency of the matching, i.e., the proportion of students getting their first choice, is increasing in the correlation of the rankings. This phenomenon is consistent with results from the tie-breaking literature (Abdulkadiroğlu et al., 2015; Ashlagi et al., 2019; Arnosti, 2022b), that we further discuss in Section 5.3. However, when there are several groups, our results reveal an additional insight. The following corollary shows that when holding the level of efficiency of the stable matching constant, different correlation pairs exist for which the stable matching achieves this efficiency but favors one of the groups. Thus, the effect we find goes beyond the efficiency loss of stable matchings when preferences are less correlated.

Corollary 4.9. *There exist infinitely many correlation pairs achieving a given efficiency. Formally, suppose that Θ, \mathbf{I}_A and \mathbf{I}_B are intervals, and $(f_\theta)_{\theta \in \Theta}$ is coherent and differentiable. Fix $\gamma, \beta_{G_1}, \beta_{G_2} \in [0, 1]$, $\alpha_A, \alpha_B \in (0, 1)$ such that $\alpha_A + \alpha_B < 1$, and let (θ_1, θ_2) vary. Then the possible values for the efficiency of the matching form an interval $[V_1^{\min}, V_1^{\max}]$. Moreover, for any efficiency level V inside the open interval (V_1^{\min}, V_1^{\max}) , there exists an interval $U \subset \Theta$ and a decreasing function ϕ such that for any $\theta_1 \in U$, $V_1(\theta_1, \phi(\theta_1)) = V$. There exists $\bar{\theta}_1 \in U$ such that $\phi(\bar{\theta}_1) = \bar{\theta}_1$, leading to $L(\bar{\theta}_1, \phi(\bar{\theta}_1)) = 0$, and $L(\theta_1, \phi(\theta_1))$ is maximal on the extremities of U .*

The proof is provided in Appendix A.4.5. This corollary shows that a given global efficiency level can hide different admission levels across groups, leading to inequalities. For any reachable efficiency V , there exists a continuous line of correlation couples achieving it, each extremity being optimal for one group and pessimal for the other. Moreover, it is always possible to reach this efficiency level while maintaining $L = 0$. The reach of Theorem 4.7 therefore goes beyond the intuition that correlation favors efficiency, it provides a precise insight to the relation between efficiency and inequality and shows that there is no inherent trade-off between them.

5 Special Cases

In our model, the exact solutions of the market-clearing equation, and thus the metrics, usually do not admit closed-form expressions. In this section, we focus on some notable special cases for which these calculations are possible and allow us to have a quantitative view of the effects of correlation. In particular, since Theorem 4.7 and 4.8 state that the metrics are monotonic, we are able to compute them for correlation levels of 0 and 1 which provides bounds for all correlation values in between. We also discuss the link between our model and the tie-breaking problem in school choice, and show how our model nests some existing results on this subject.

5.1 Excess capacity

In the previous section we mostly 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, correlation no longer has any influence on the matching.

Proposition 5.1. *If capacity is not constrained, i.e., $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$, then correlation has no effect on the stable matching, and students from both groups have the exact same outcome probabilities. Formally, for $C \in \{\mathbf{A}, \mathbf{B}\}$ $V_1^{\mathbf{G}_1, C} = V_1^{\mathbf{G}_2, C}$ and $V_2^{\mathbf{G}_1, C} = V_2^{\mathbf{G}_2, C}$. Moreover, $V_{\emptyset}^{\mathbf{G}_1} = V_{\emptyset}^{\mathbf{G}_2} = 0$. Finally, these quantities are constant in θ_1 and θ_2 .*

Proof sketch. The relations between the V_1 are from Theorem 4.5 (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 4.2 and use the fact that $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$ implies that at least one of the two colleges does not reject any student. The details are provided in Appendix A.4.6. ■

In fact, when $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$, we can explicitly compute the proportions of students getting their first or second choice by analyzing the steps of the Deferred Acceptance Algorithm (see Algorithm 1 in the Appendix). We consider three (partitioning) cases:

- (1) *There is not enough room in college \mathbf{A} for all students preferring it to college \mathbf{B} , i.e., $\gamma\beta_{\mathbf{G}_1} + (1 - \gamma)\beta_{\mathbf{G}_2} \geq \alpha_{\mathbf{A}}$. In this case, there is necessarily enough room in college \mathbf{B} for all students preferring it, since $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$. Therefore, following the steps of DA, we find:

 - (i) At step one, $\gamma\beta_{\mathbf{G}_1} + (1 - \gamma)\beta_{\mathbf{G}_2}$ students preferring \mathbf{A} apply there and the best $\alpha_{\mathbf{A}}$ are temporarily admitted, and $\gamma(1 - \beta_{\mathbf{G}_1}) + (1 - \gamma)(1 - \beta_{\mathbf{G}_2})$ students preferring \mathbf{B} apply there and are all temporarily admitted.
 - (ii) At step two, the $\gamma\beta_{\mathbf{G}_1} + (1 - \gamma)\beta_{\mathbf{G}_2} - \alpha_{\mathbf{A}}$ 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 probabilities of a student to get their first or second choice:

$$V_1^{\mathbf{G}_1, \mathbf{A}} = V_1^{\mathbf{G}_2, \mathbf{A}} = \frac{\alpha_{\mathbf{A}}}{\gamma\beta_{\mathbf{G}_1} + (1-\gamma)\beta_{\mathbf{G}_2}}, \quad V_1^{\mathbf{G}_1, \mathbf{B}} = V_1^{\mathbf{G}_2, \mathbf{B}} = 1, \quad (10)$$

$$V_2^{\mathbf{G}_1, \mathbf{A}} = V_2^{\mathbf{G}_2, \mathbf{A}} = 1 - \frac{\alpha_{\mathbf{A}}}{\gamma\beta_{\mathbf{G}_1} + (1-\gamma)\beta_{\mathbf{G}_2}}, \quad V_2^{\mathbf{G}_1, \mathbf{B}} = V_2^{\mathbf{G}_2, \mathbf{B}} = 0. \quad (11)$$

Finally, as every student is admitted somewhere, $V_\emptyset^{\mathbf{G}, \mathbf{C}} = 0$ for $\mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\}$ and $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$.

(2) *There is not enough room in college B for all students preferring it to A, i.e., $\gamma(1 - \beta_{\mathbf{G}_1}) + (1 - \gamma)(1 - \beta_{\mathbf{G}_2}) \geq \alpha_{\mathbf{B}}$.* This case is symmetric to the previous case and is omitted for brevity.

(3) *There is enough room in each college to admit all students who prefer attending it, i.e., $\gamma\beta_{\mathbf{G}_1} + (1 - \gamma)\beta_{\mathbf{G}_2} \leq \alpha_{\mathbf{A}}$ and $\gamma(1 - \beta_{\mathbf{G}_1}) + (1 - \gamma)(1 - \beta_{\mathbf{G}_2}) \leq \alpha_{\mathbf{B}}$.* It follows that everyone gets their first choice: for $\mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\}$ and $\mathbf{C} \in \{\mathbf{A}, \mathbf{B}\}$,

$$V_1^{\mathbf{G}, \mathbf{C}} = 1, \quad (12)$$

$$V_2^{\mathbf{G}, \mathbf{C}} = V_\emptyset^{\mathbf{G}, \mathbf{C}} = 0. \quad (13)$$

Note that Equations (10) to (13) are consistent with Proposition 5.1.

5.2 One group

Suppose that there is only one group of students (i.e., $\gamma = 1$) and therefore all students have the same correlation parameter θ . In this section we consider two boundary cases, where the colleges have either full correlation (they use the same ranking of students) or no correlation at all (their rankings are statistically independent). These special cases will allow us to understand the matching's dependencies on the capacities and preferences of students. We assume that there is no capacity excess, 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.

Full correlation.

We first study the case where students have the same rank in both colleges.

Proposition 5.2. *When both colleges use the same ranking, the metrics V_1, V_2, V_\emptyset can be computed exactly, and their expressions are:*

(i) *If $\beta \leq \frac{\alpha_{\mathbf{A}}}{\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}}}$ (College A is under-demanded):*

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

(ii) If $\beta \geq \frac{\alpha_A}{\alpha_A + \alpha_B}$ (College A is over-demanded):

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

Proof sketch. The proof amounts to solving the market-clearing equation, the details are provided in Appendix A.4.7. ■

This result is illustrated by the blue and orange lines in Figure 4. For the top row, $\beta = 0.3$, and V_1^A and V_1^B are computed for $\alpha_A = \alpha_B := \alpha$ varying from 0 to 1. The probability of getting one's first choice is increasing in the capacity. Students preferring college **A** get either their first choice or nothing, as shown in Proposition 5.2. Indeed, college **A** is easier to get in than college **B** since $\beta \leq \frac{\alpha_A}{\alpha_A + \alpha_B}$, so a student rejected from college **A** is necessarily rejected from college **B**. For $\alpha > 0.5$, Proposition 5.2 does not apply anymore because there is capacity excess, and we need to refer to Equations (10) and (11). For the bottom row, $\alpha_A = \alpha_B = 0.25$ and V_1^A and V_1^B are shown for β varying from 0 to 1. The two figures are mirrored images of each other, which is natural as the problem is symmetric in **A** and **B**. Observe that students who prefer the least popular college have a higher probability of getting their first choice (the blue line is higher on the left plot than on the right for $\beta < 0.5$, and lower for $\beta > 0.5$).

Remark. *The full correlation case does not fit our model's assumptions, since the distribution does not have full support. The consequence is that uniqueness of the stable matching is not guaranteed. However, by solving the market clearing equation, we showed that there is indeed only one solution, so the stable matching is still unique.*

No correlation.

Now, consider the case where the ranks of a student at **A** and **B** are not correlated at all.

Proposition 5.3. *When colleges' rankings are independent, the metrics V_1, V_2, V_\emptyset can be computed exactly, and their expressions are:*

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

with $\zeta = \frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha_A - \alpha_B$ and $\Delta = \sqrt{\zeta^2 + \frac{4\beta}{1-\beta}(1 - \alpha_A - \alpha_B)}$.

Proof sketch. Again, the proof amounts to solving the market-clearing equation, the details are provided in Appendix A.4.8. ■

These results are illustrated by the dashed green and red lines in Figure 4. For the top row, once again, Proposition 5.3 applies only for $\alpha < 0.5$. Note that V_2^A and V_2^B are both strictly positive, because even though college **A** is easier to get in than college **B**, students rejected from college **A** still have an independent second chance at **B**.

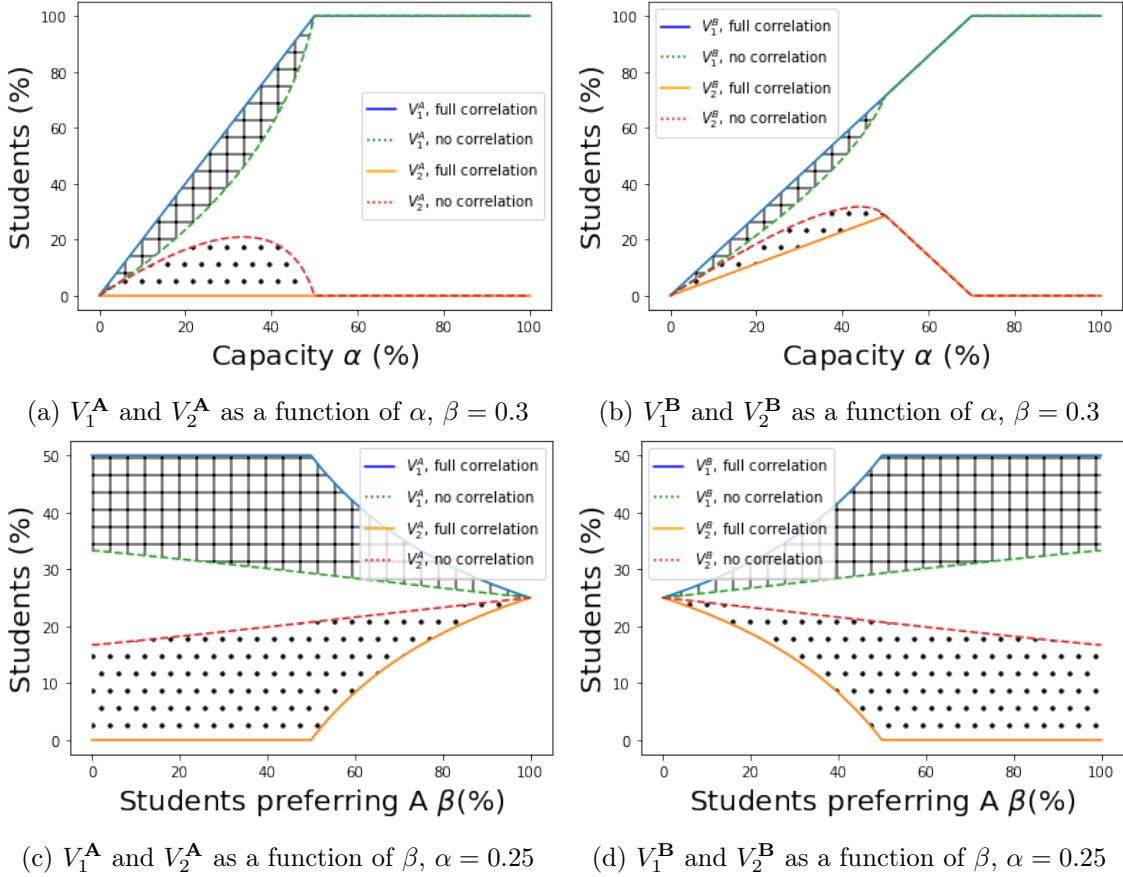


Figure 4: Proportion of students getting their first and second choice with one group, and $\alpha_A = \alpha_B = \alpha$. The solid blue and orange lines are for the full-correlation case, the dashed green and red ones for the no-correlation case. As the blue and green lines represent the same metric in two different settings, the gap between them is hashed to highlight the welfare increase when switching from one to the other. The same applies to the orange and red lines, where the gap is dotted.

Comparing the blue and green lines, note that the amount of students getting their first choice is always larger with full correlation. This follows from Theorem 4.7 as V_1 is increasing in the correlation, for any values of the parameters. Correspondingly, the amount of students getting their second choice is always lower with full correlation. For values of θ for which the correlation is between 0 and 1, it is generally not possible to obtain closed-form expressions for these metrics. However,

Theorem 4.7 implies that their graphs have to be contained between the two extreme cases, i.e., in the colored area highlighted in Figure 4. These areas show the extent of the influence of correlation on students' welfare. It appears that for intermediate values of the problem's fundamentals, the gap is substantial. For instance, looking at the bottom row, we see that for $\beta = 0.5$ and $\alpha_{\mathbf{A}} = \alpha_{\mathbf{B}} = 0.25$, increasing the correlation can make the number of students getting their first choice grow from 30% to 50%.

5.3 Tie-Breaking

Some recent papers have studied the impact of tie-breaking rules on school choice problems, which has a strong link with correlation. In this section, we apply our model to this setting and recover a result similar to several existing results from this literature.

Assume there is only one group, and each school \mathbf{C} has $n_{\mathbf{C}}$ priority classes, i.e., there exists a partition of $\mathbf{S} = K_{\mathbf{C}}^1 \sqcup \dots \sqcup K_{\mathbf{C}}^{n_{\mathbf{C}}}$ such that for $i, j \in \{1, \dots, n_{\mathbf{C}}\}$, $s \in K_{\mathbf{C}}^i$, $s' \in K_{\mathbf{C}}^j$, s has higher priority than s' at \mathbf{C} if $i < j$. Students belonging to the same priority class at a school are assumed to have the same priority at this school, but due to limited capacity the school might need to choose between them. To achieve this, schools use a random ranking of students to which they refer each time they need to choose between students from the same priority class; this random ranking is called a *tie-breaker*.

A natural question that has been actively studied in recent years is whether there is a difference in students' welfare if schools use the same tie-breaker (called single tie-breaker, or STB), instead of each producing an independent one (multiple tie-breakers, or MTB)? Ashlagi et al. (2019); Ashlagi and Nikzad (2020); Arnosti (2022b) show — with slightly different models and assumptions (and among other results) — that when the total capacity of schools is lower than the number of students, then students are better off under STB than MTB. To ease the comparison, we restate their results here in a simplified form:

Proposition 5.4 (Ashlagi et al. 2019; Ashlagi and Nikzad 2020; Arnosti 2022b). *Given n students and m schools:*

- *Suppose that there is capacity shortage, students' preferences are drawn uniformly at random and there is only one priority class (the whole ranking is random), then for any $k < m$ the fraction of students matched to one of their top k choices approaches 0 under MTB but approaches a positive constant under STB (Ashlagi et al. 2019, Main Theorem)*
- *Suppose there is one slot per school, only one priority class, and schools are divided into two tiers (top and bottom) with students' preferences inside a given tier drawn uniformly at random and a capacity shortage at top schools, then, with high probability, STB stochastically Pareto-dominates MTB (Ashlagi and Nikzad 2020, Theorem 3.2).*

- Suppose there is only one priority class and students only list $l < m$ schools in a uniform random order, then the number of students matched to their first choice is greater under STB than under MTB (Arnosti 2022b, Theorem 2).

A similar result can also be proven using our model.

Proposition 5.5. *Let there be a continuum mass of students. Let \mathbf{A}, \mathbf{B} be two schools with priority classes, capacities $\alpha_{\mathbf{A}}$ and $\alpha_{\mathbf{B}}$ such that $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$. Assume that students prefer any school rather than staying unmatched. Then:*

1. *If there is only one priority class, the mass of students getting their first choice is greater with STB than MTB.*
2. *If there are several priority classes, assume that all products of priority classes $K_{\mathbf{A}}^i \times K_{\mathbf{B}}^j$ contain a positive mass of students. Then the mass of students getting their first choice with STB is greater or equal compared to MTB. Further assume that students' preferences or schools' capacities are randomly drawn from a distribution that admits a density. Then the mass of students getting their first choice is almost surely strictly greater with STB than MTB.*

Proof sketch. We build a distribution family $(f_{\theta})_{\theta \in \Theta}$ that describes the priorities of students at each school taking into account priority classes as well as tie-breakers, such that MTB and STB correspond to values of $\theta = 0$ and $\theta = 1$. To achieve this, we consider the unit square. We divide the segment $[0, 1]$ on the x axis in $n_{\mathbf{A}}$ parts (the number of classes at \mathbf{A}), and we do the same for the y axis with $n_{\mathbf{B}}$. Each box inside the unit square defined by those divisions corresponds to a product of priority classes, and we impose that a student's score vector belongs to the box corresponding to their priority classes. To incorporate the tie-breaking, inside each box we draw the grades of the corresponding students from a Gaussian copula with correlation θ contained inside the box. The obtained distribution, while being complex, still satisfies most of the assumptions required by our model, and with some adjustments we are able to apply Theorem 4.7 and conclude. Details are provided in Appendix A.4.9. ■

Proposition 5.5 is in some regards more restrictive than the results presented in Proposition 5.4, because it only applies to two schools and assumes students list both schools. On the other hand, it is more general in that it applies to cases with several priority classes and does not require students preferences to be uniform (in our model, we can have any fraction β of students preferring school \mathbf{A}). Finally, Proposition 5.5 can also be extended to intermediary tie-breaking rules between MTB and STB—that is, any level of correlation between tie-breakers—and to two groups of students with different tie-breaking rules, with all results from Section 4 remaining valid.

6 Discussion

Selection problems are ubiquitous, and consequently, researchers and practitioners have focused on understanding and mitigating sources of discrimination in these settings. However, in prior work a single decision-maker is considered, which is in contrast to many situations where several decision-makers compete over the same pool of candidates.

Our work is the first to consider a matching model where differential correlation is the source of outcome inequalities between different groups. Our results appear counter-intuitive for the model where differential correlation comes from noisy estimates of students' qualities: the group with higher noise is overrepresented. This contrasts the intuition that the group with higher noise is the minority group and is often underrepresented in practice. To fully capture real-life situations one would need to consider other sources of differences, including varying noise variance, taste-based discrimination, and risk averse decision-makers. However, the aim of our work was to isolate the effect of differential correlation.

From a technical point of view, to isolate the effect of correlation from other distributional effects, we assumed that the marginals are the same for both groups. When the marginals are different, Theorem 4.7, and Corollaries 4.8 and 4.9 remain valid, while Theorems 4.5 and 4.6 no longer apply. This means that estimating which group is advantaged is more difficult when marginals are different, and it might depend on colleges' capacities, but the comparative statics still apply. Therefore, our results can be used to compare settings with different levels of correlation. We also focus on a setting with two colleges. When there are more colleges, Theorems 4.5 and 4.6 are still valid, allowing us to determine which group is advantaged. The proof of Theorem 4.7, however, does not extend straightforwardly. An open question is then whether the result is still valid or if stronger assumptions are required.

We encourage researchers to take the matching context as the basis and inquire themes that have already been considered in the single decision-maker setting. This includes, allowing applicants to invest in accurate assessment, e.g., via acquiring certifications or doing in-person interviews, or considering the effects of risk aversion. Other possible directions could include allowing applicants to not list all colleges in their preferences, making applications costly, or adding a correlation between an applicant's preferences and their rank at each decision-maker.

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.
- Arcidiacono, Peter, Josh Kinsler, and Tyler Ransom (2022), “Asian american discrimination in harvard admissions.” *European Economic Review*, 144, 104079.
- Arnosti, Nick (2022a), “A continuum model of stable matching with finite capacities.” In *Proceedings of the 23rd ACM Conference on Economics and Computation*, 960.
- Arnosti, Nick (2022b), “Lottery design for school choice.” *Management Science*, forthcoming.
- 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 and Afshin Nikzad (2020), “What matters in school choice tie-breaking? how competition guides design.” *Journal of Economic Theory*, 190, 105120.
- 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.
- 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.
- 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>.
- Becker, Gary S (1957), *The economics of discrimination*. University of Chicago press.
- Bertrand, Marianne and Sendhil Mullainathan (2004), “Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination.” *American Economic Review*, 94.
- 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)*.
- Brilliantova, Angelina and Hadi Hosseini (2022), “Fair stable matching meets correlated preferences.” *CoRR*, abs/2201.12484.

- 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.
- Che, Yeon-Koo, Kyungmin Kim, and Weijie Zhong (2020), “Statistical discrimination in ratings-guided markets.” Working paper.
- Che, Yeon-Koo and Olivier Tercieux (2019), “Efficiency and stability in large matching markets.” *Journal of Political Economy*, 127, 2301–2342.
- Chouldechova, Alexandra and Aaron Roth (2020), “A snapshot of the frontiers of fairness in machine learning.” *Commun. ACM*, 63, 82–89.
- Delacrétaz, David, Scott Duke Kominers, and Alexander Teytelboym (2020), “Matching Mechanisms for Refugee Resettlement.” Working paper.
- Dur, Umut, Parag A. Pathak, and Tayfun Sönmez (2020), “Explicit vs. statistical targeting in affirmative action: Theory and evidence from chicago’s exam schools.” *Journal of Economic Theory*, 187, 104996.
- 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.
- Erdil, Aytek and Haluk Ergin (2008), “What’s the matter with tie-breaking? improving efficiency in school choice.” *American Economic Review*, 98, 669–89.
- Ergin, Haluk and Tayfun Sönmez (2006), “Games of school choice under the Boston mechanism.” *Journal of Public Economics*, 90, 215–237.
- 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)*.
- Hardt, Moritz, Eric Price, and Nathan Srebro (2016), “Equality of opportunity in supervised learning.” In *Advances in Neural Information Processing Systems (NIPS ’16)*.
- Howe, Amy (2022), “Affirmative action cases up first in november argument calendar.” www.scotusblog.com.
- 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 (2022), “On Fairness and Stability in Two-Sided Matchings.” In *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, Leibniz International Proceedings in Informatics (LIPIcs), 92:1–92:17.
- 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.
- Krishnaa, Prem, Girija Limaye, Meghana Nasre, and Prajakta Nimbhorkar (2019), “Envy-freeness and Relaxed Stability under lower quotas.” Working paper.
- Leshno, Jacob D and Irene Lo (2020), “The Cutoff Structure of Top Trading Cycles in School Choice.” *The Review of Economic Studies*, 88, 1582–1623.
- Longhofer, Stanley D. (1995), “Rooting out discrimination in home mortgage lending.” *Economic Commentary*, 1.
- 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)*.

- Niessen-Ruenzi, Alexandra and Stefan Ruenzi (2019), “Sex matters: Gender bias in the mutual fund industry.” *Management Science*, 65, 3001–3025.
- Orwat, Carsten (2020), *Risks of Discrimination through the Use of Algorithms*.
- Phelps, Edmund (1972), “The statistical theory of racism and sexism.” *American Economic Review*, 62, 659–661.
- Roth, A. E. and H. Vande Vate (1990), “Random paths to stability in two-sided matching.” *Econometrica*, 58, 1475–1480.
- Roth, Alvin E (1985), “The college admissions problem is not equivalent to the marriage problem.” *Journal of Economic Theory*, 36, 277–288.
- 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.
- Scarsini, Marco (1984), “On measures of concordance.” *Stochastica*, 8, 201–218.
- Shapley, Lloyd and Herbert Scarf (1974), “On cores and indivisibility.” *Journal of Mathematical Economics*, 1, 23–37.
- 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 and definitions

A.1 Definitions

A.1.1 Table of notation.

Table 1 provides a summary of the notation used throughout the paper.

Table 1: Notation

<u>Agents:</u>	
\mathbf{A}, \mathbf{B}	Colleges (generic: \mathbf{C})
s	An arbitrary student
\mathbf{S}	Students set
$\mathbf{G}_1, \mathbf{G}_2$	Groups of students, partition of \mathbf{S} (generic: \mathbf{G})
η	Measure for student masses
<u>Agents' features:</u>	
$\alpha_{\mathbf{A}}, \alpha_{\mathbf{B}}$	Colleges' capacities ($\in (0, 1)$)
γ	Mass of students in group \mathbf{G}_1 ($\in [0, 1]$)
$\beta_{\mathbf{G}}$	Share of students in group \mathbf{G} preferring college \mathbf{A} ($\in [0, 1]$)
<u>Priority scores:</u>	
$W_{\mathbf{C}}^s$	Score at \mathbf{C} of student s (generic: W)
f	Score vectors' joint probability distribution function
F	Joint cumulative distribution function
$f_{\mathbf{C}}, F_{\mathbf{C}}$	Marginal pdf and cdf of college \mathbf{C}
θ	Parameter for a distribution family
Θ	Set of possible values for θ
$\mathbf{I}, \mathbf{I}_{\mathbf{A}}, \mathbf{I}_{\mathbf{B}}$	Support of $f, f_{\mathbf{A}}$ and $f_{\mathbf{B}}$ respectively. $\mathbf{I} = \mathbf{I}_{\mathbf{A}} \times \mathbf{I}_{\mathbf{B}}$
$\mathbf{I}_{\mathbf{C}}^l, \mathbf{I}_{\mathbf{C}}^u$	Lower and upper bounds of \mathbf{C} 's component of \mathbf{I}
<u>Correlation:</u>	
r	Pearson's correlation
ρ	Spearman's correlation
τ	Kendall's correlation
<u>Matching:</u>	
μ	Matching
$V_1^{\mathbf{G}, \mathbf{C}}$	Share of students of group \mathbf{G} and preferring \mathbf{C} who get their first choice
$V_2^{\mathbf{G}, \mathbf{C}}$	Share of students of group \mathbf{G} and preferring \mathbf{C} who get their second choice
$V_{\emptyset}^{\mathbf{G}, \mathbf{C}}$	Share of students of group \mathbf{G} and preferring \mathbf{C} who are unassigned

A.1.2 Definition of the mass η .

Here we formally define the notion of mass for a subset of students. This section is quite technical and is not necessary to understand the results, and the notations introduced here are not used anywhere else in the paper. We identify \mathbf{S} to $\Sigma := \mathbb{R}^2 \times \{\mathbf{G}_1, \mathbf{G}_2\} \times \{\mathbf{A}, \mathbf{B}\}$. We partition Σ into 4 subsets: $\Sigma_{\mathbf{G}, \mathbf{C}} := \{s \in \Sigma : s = ((x, y), \mathbf{G}, \mathbf{C}), x, y \in \mathbb{R}\}$ is the subset of students belonging to group \mathbf{G} and preferring college \mathbf{C} . Given θ_1, θ_2 and priorities $W_{\mathbf{A}}, W_{\mathbf{B}}$ distributed according to f_{θ_1} for \mathbf{G}_1 and f_{θ_2} for \mathbf{G}_2 , we say that a subset $K \subseteq \Sigma$ is measurable if and only if $\{(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s) : s \in K\}$ is

Borel-measurable in \mathbb{R}^2 . We can partition K into 4 subsets $K_{\mathbf{G},\mathbf{C}} := K \cap \Sigma_{\mathbf{G},\mathbf{C}}$. On each $\Sigma_{\mathbf{G},\mathbf{C}}$ we define a measure $\eta_{\mathbf{G},\mathbf{C}}$ as follows: for $K \subseteq \Sigma$ measurable,

$$\begin{aligned}\eta_{\mathbf{G}_1,\mathbf{A}}(K_{\mathbf{G}_1,\mathbf{A}}) &= \gamma\beta_{\mathbf{G}_1}\mathbb{P}_{\theta_1}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s) : s \in K_{\mathbf{G}_1,\mathbf{A}}\}), \\ \eta_{\mathbf{G}_1,\mathbf{B}}(K_{\mathbf{G}_1,\mathbf{B}}) &= \gamma(1 - \beta_{\mathbf{G}_1})\mathbb{P}_{\theta_1}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s) : s \in K_{\mathbf{G}_1,\mathbf{B}}\}), \\ \eta_{\mathbf{G}_2,\mathbf{A}}(K_{\mathbf{G}_2,\mathbf{A}}) &= (1 - \gamma)\beta_{\mathbf{G}_2}\mathbb{P}_{\theta_2}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s) : s \in K_{\mathbf{G}_2,\mathbf{A}}\}), \\ \eta_{\mathbf{G}_2,\mathbf{B}}(K_{\mathbf{G}_2,\mathbf{B}}) &= (1 - \gamma)(1 - \beta_{\mathbf{G}_2})\mathbb{P}_{\theta_2}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in \{(W_{\mathbf{A}}^s, W_{\mathbf{B}}^s) : s \in K_{\mathbf{G}_2,\mathbf{B}}\}),\end{aligned}\tag{14}$$

Let $\mathcal{B}(\mathbf{S})$ be the set of measurable subsets of \mathbf{S} . We define over $\mathcal{B}(\mathbf{S})$ the probability measure $\eta : \mathcal{B}(\mathbf{S}) \rightarrow [0, 1]$ such that for any measurable subset K of \mathbf{S} ,

$$\eta(K) = \eta_{\mathbf{G}_1,\mathbf{A}}(K_{\mathbf{G}_1,\mathbf{A}}) + \eta_{\mathbf{G}_1,\mathbf{B}}(K_{\mathbf{G}_1,\mathbf{B}}) + \eta_{\mathbf{G}_2,\mathbf{A}}(K_{\mathbf{G}_2,\mathbf{A}}) + \eta_{\mathbf{G}_2,\mathbf{B}}(K_{\mathbf{G}_2,\mathbf{B}}).\tag{15}$$

This definition is consistent with the definition of the parameters, as it verifies $\eta(\mathbf{G}_1) = \gamma$, $\eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}\}) = \gamma\beta_{\mathbf{G}_1}$ and $\eta(\{s \in \mathbf{G}_2 : \mathbf{A} \succ_s \mathbf{B}\}) = (1 - \gamma)\beta_{\mathbf{G}_2}$.

A.2 Elements of correlation theory

In this section, we present common measures of correlation used in the literature, and some of their properties.

Definition A.1 (Common measures of correlation). Let (X, Y) be two random variables with respective cdfs F_X, F_Y . Define:

1. Pearson's correlation: Assume X, Y have finite standard deviations σ_X and σ_Y . Then $r_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$.
2. Spearman's correlation: let $rk_X = F_X(X)$ and $rk_Y = F_Y(Y)$. We can think of rk_X as describing the ranking of X inside a sample. Then Spearman's correlation is $\rho_{X,Y} = r_{rk_X, rk_Y}$.
3. Kendall's correlation: let (X_1, Y_1) and (X_2, Y_2) be two independent pairs of random variables with the same joint distribution as (X, Y) . Then Kendall's correlation is

$$\begin{aligned}\tau_{X,Y} &= \mathbb{P}[(X_1 > X_2 \cap Y_1 > Y_2) \cup (X_1 < X_2 \cap Y_1 < Y_2)] - \\ &\quad \mathbb{P}[(X_1 > X_2 \cap Y_1 < Y_2) \cup (X_1 < X_2 \cap Y_1 > Y_2)].\end{aligned}$$

We use the same letter r for the covariance of the standard bivariate Gaussian and for Pearson's correlation as they are equal. Moreover, for this distribution simple expressions exist for the two other correlation coefficients:

$$\rho = \frac{6}{\pi} \arcsin(r/2), \tau = \frac{2}{\pi} \arcsin(r).$$

A correlation measure should be zero when variables are independent, and reach its maximum when the variables are totally dependent on each other. The following lemma provides these properties for the measures we just introduced.

Lemma A.2 (Scarsini 1984, Theorems 1, 4, and 5). *Let X, Y be two real random variables.*

1. $r_{X,Y}, \rho_{X,Y}, \tau_{X,Y} \in [-1, 1]$.
2. $\rho_{X,Y} = 1$ if and only if $Y = g(X)$ with $g : \mathbb{R} \rightarrow \mathbb{R}$ increasing. The same holds for $\tau_{X,Y}$.
 $r_{X,Y} = 1$ if and only if the relation is affine.
3. If X and Y are independent, then $r_{X,Y} = \rho_{X,Y} = \tau_{X,Y} = 0$.

In stable matchings, the crucial information is the rank of each student, not directly their score. Therefore, we are mostly interested in Spearman's and Kendall's coefficients as they are more informative of the correlation between rankings. Indeed, two variables yield the same ranking if there is an increasing relation between them, but it does not need to be affine.

A.3 Stable matching

We now introduce some elements of matching theory used throughout the paper.

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

Definition A.3. A *matching* is an assignment of students to colleges, described by a mapping $\mu : \mathbf{S} \cup \{\mathbf{A}, \mathbf{B}\} \rightarrow 2^{\mathbf{S}} \cup \mathbf{C} \cup \mathbf{S}$, 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 constraints. 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 that there do not exist several stable matchings that only differ by a set of students of measure 0.

We next define the notions of blocking and stability.

Definition A.4 (Stability). The pair (s, \mathbf{C}) *blocks* a matching μ if s would prefer \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 $W_{\mathbf{C}}^{s'} < W_{\mathbf{C}}^s$. A matching is *stable* if it is not blocked by any student-college pair.

To produce a stable matching, one can 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 them 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.

Remark. *Note that stable matchings do not only result from centralized algorithms but are often the result of a decentralized process (see, e.g., Roth and Vande Vate 1990).*

A.4 Proofs

A.4.1 Proof of Lemma 2.2.

Since $F(x, y) = \mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y)$ by definition, then the first part of the lemma is just a rewriting of the definition of coherence. For the second part, we have

$$\begin{aligned} \mathbb{P}(W_{\mathbf{A}} > x, W_{\mathbf{B}} > y) &= \mathbb{P}(W_{\mathbf{B}} > y) - \mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} > y) \\ &= \mathbb{P}(W_{\mathbf{B}} > y) - \mathbb{P}(W_{\mathbf{A}} < x) + \mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y) \end{aligned}$$

and $\mathbb{P}(W_{\mathbf{B}} > y), \mathbb{P}(W_{\mathbf{A}} < x)$ are constant in θ by assumption while $\mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y)$ is increasing, so $\mathbb{P}(W_{\mathbf{A}} > x, W_{\mathbf{B}} > y)$ is also increasing. Finally, we also get that

$$\mathbb{P}(W_{\mathbf{A}} > x, W_{\mathbf{B}} < y) = \mathbb{P}(W_{\mathbf{B}} < y) - \mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y)$$

and

$$\mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} > y) = \mathbb{P}(W_{\mathbf{A}} < x) - \mathbb{P}(W_{\mathbf{A}} < x, W_{\mathbf{B}} < y)$$

and both are therefore decreasing.

A.4.2 Proof of Lemma 4.2.

Consider student $s \in \mathbf{G}_1$ who prefers \mathbf{A} to \mathbf{B} . By Lemma 3.3, 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 score at \mathbf{A} is greater than $P_{\mathbf{A}}$. Then by definition of η ,

$$V_1^{\mathbf{G}_1, \mathbf{A}}(\theta_1, \theta_2) = \frac{\eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{A}\})}{\gamma\beta_{\mathbf{G}_1}} = \mathbb{P}_{\theta_1}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in [P_{\mathbf{A}}, +\infty) \times \mathbb{R}) = \mathbb{P}(W_{\mathbf{A}} > P_{\mathbf{A}}).$$

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

$$V_2^{\mathbf{G}_1, \mathbf{A}}(\theta_1, \theta_2) = \frac{\eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = \mathbf{B}\})}{\gamma\beta_{\mathbf{G}_1}} = \mathbb{P}_{\theta_1}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in (-\infty, P_{\mathbf{A}}) \times [P_{\mathbf{B}}, +\infty)).$$

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

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

$$V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}(\theta_1, \theta_2) = \frac{\eta(\{s \in \mathbf{G}_1 : \mathbf{A} \succ_s \mathbf{B}, \mu(s) = s\})}{\gamma\beta_{\mathbf{G}_1}} = \mathbb{P}_{\theta_1}((W_{\mathbf{A}}, W_{\mathbf{B}}) \in (-\infty, P_{\mathbf{A}}) \times (-\infty, P_{\mathbf{B}})).$$

which proves (6).

A.4.3 Proof of Theorem 4.7.

Let $\gamma, \beta_{\mathbf{G}_1}, \beta_{\mathbf{G}_2} \in [0, 1]$, $\alpha_{\mathbf{A}}, \alpha_{\mathbf{B}} \in (0, 1)$ such that $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} < 1$, and $\theta_1, \theta_2 \in \mathring{\Theta}$. Let $P_{\mathbf{A}}, P_{\mathbf{B}} \in \mathbb{R}$ be the cutoffs of colleges \mathbf{A} and \mathbf{B} .

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

$$\begin{cases} \gamma\beta_{\mathbf{G}_1}V_1^{\mathbf{G}_1, \mathbf{A}} + \gamma(1 - \beta_{\mathbf{G}_1})V_2^{\mathbf{G}_1, \mathbf{B}} + (1 - \gamma)\beta_{\mathbf{G}_2}V_1^{\mathbf{G}_2, \mathbf{A}} + (1 - \gamma)(1 - \beta_{\mathbf{G}_2})V_2^{\mathbf{G}_2, \mathbf{B}} & = \alpha_{\mathbf{A}}, \\ \gamma\beta_{\mathbf{G}_1}V_2^{\mathbf{G}_1, \mathbf{A}} + \gamma(1 - \beta_{\mathbf{G}_1})V_1^{\mathbf{G}_1, \mathbf{B}} + (1 - \gamma)\beta_{\mathbf{G}_2}V_2^{\mathbf{G}_2, \mathbf{A}} + (1 - \gamma)(1 - \beta_{\mathbf{G}_2})V_1^{\mathbf{G}_2, \mathbf{B}} & = \alpha_{\mathbf{B}}. \end{cases}$$

Then, using Lemma 4.2, we can rewrite is as

$$\left\{ \begin{array}{l} (\gamma\beta_{\mathbf{G}_1} + (1-\gamma)\beta_{\mathbf{G}_2})\mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}}) + \gamma(1-\beta_{\mathbf{G}_1})\mathbb{P}_{\theta_1}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}}) \\ \quad + (1-\gamma)(1-\beta_{\mathbf{G}_2})\mathbb{P}_{\theta_2}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}}) = \alpha_{\mathbf{A}}, \\ \gamma\beta_{\mathbf{G}_1}\mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) + (1-\gamma)\beta_{\mathbf{G}_2}\mathbb{P}_{\theta_2}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}}) \\ \quad + (\gamma(1-\beta_{\mathbf{G}_1}) + (1-\gamma)(1-\beta_{\mathbf{G}_2}))\mathbb{P}(W_{\mathbf{B}} \geq P_{\mathbf{B}}) = \alpha_{\mathbf{B}}, \end{array} \right.$$

which is finally equivalent to

$$\left\{ \begin{array}{l} (\gamma\beta_{\mathbf{G}_1} + (1-\gamma)\beta_{\mathbf{G}_2}) \int_{P_{\mathbf{A}}}^{\infty} f_{\mathbf{A}}(x) dx + \gamma(1-\beta_{\mathbf{G}_1}) \int_{P_{\mathbf{A}}}^{\infty} \int_{-\infty}^{P_{\mathbf{B}}} f_{\theta_1}(x, y) dx dy \\ \quad + (1-\gamma)(1-\beta_{\mathbf{G}_2}) \int_{P_{\mathbf{A}}}^{\infty} \int_{-\infty}^{P_{\mathbf{B}}} f_{\theta_2}(x, y) dx dy = \alpha_{\mathbf{A}}, \\ \gamma\beta_{\mathbf{G}_1} \int_{-\infty}^{P_{\mathbf{A}}} \int_{P_{\mathbf{B}}}^{\infty} f_{\theta_1}(x, y) dx dy + (1-\gamma)\beta_{\mathbf{G}_2} \int_{-\infty}^{P_{\mathbf{A}}} \int_{P_{\mathbf{B}}}^{\infty} f_{\theta_2}(x, y) dx dy \\ \quad + (\gamma(1-\beta_{\mathbf{G}_1}) + (1-\gamma)(1-\beta_{\mathbf{G}_2})) \int_{P_{\mathbf{B}}}^{\infty} f_{\mathbf{B}}(x) dx = \alpha_{\mathbf{B}}. \end{array} \right. \quad (16)$$

We fix θ_2 , and we want to study how the solution $(P_{\mathbf{A}}, P_{\mathbf{B}})$ of the above equation varies as a function of θ_1 . Let us define $h : \mathbb{R}^2 \times [0, 1) \rightarrow \mathbb{R}^2$, $(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1) \mapsto (D_{\mathbf{A}}(P_{\mathbf{A}}, P_{\mathbf{B}}) - \alpha_{\mathbf{A}}, D_{\mathbf{B}}(P_{\mathbf{A}}, P_{\mathbf{B}}) - \alpha_{\mathbf{B}})$. (We will denote by h_1 and h_2 its two components.) Then for each $\theta_1 \in \Theta$, $(P_{\mathbf{A}}, P_{\mathbf{B}})$ is the solution of the equation $h(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1) = (0, 0)$. In order to show that $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are decreasing in θ_1 , we wish to apply the implicit function theorem. Let $P_{\mathbf{A}}, P_{\mathbf{B}} \in \mathbb{R}$ and $\theta_1 \in \Theta$ such that $h(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1) = 0$. Function h is of class C^1 . We first verify that the partial Jacobian $J_{h, (P_{\mathbf{A}}, P_{\mathbf{B}})}(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1)$ is invertible, where

$$J_{h, (P_{\mathbf{A}}, P_{\mathbf{B}})}(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1) = \begin{pmatrix} \frac{\partial h_1}{\partial P_{\mathbf{A}}} & \frac{\partial h_1}{\partial P_{\mathbf{B}}} \\ \frac{\partial h_2}{\partial P_{\mathbf{A}}} & \frac{\partial h_2}{\partial P_{\mathbf{B}}} \end{pmatrix}. \quad (17)$$

To show that the determinant $\frac{\partial h_1}{\partial P_{\mathbf{A}}} \frac{\partial h_2}{\partial P_{\mathbf{B}}} - \frac{\partial h_1}{\partial P_{\mathbf{B}}} \frac{\partial h_2}{\partial P_{\mathbf{A}}} \neq 0$, we will show that it is in fact (strictly) positive. From (16), it is clear that h_1 is decreasing in $P_{\mathbf{A}}$ and increasing in $P_{\mathbf{B}}$, and that h_2 is increasing in $P_{\mathbf{A}}$ and decreasing in $P_{\mathbf{B}}$. Therefore, to prove that $\frac{\partial h_1}{\partial P_{\mathbf{A}}} \frac{\partial h_2}{\partial P_{\mathbf{B}}} - \frac{\partial h_1}{\partial P_{\mathbf{B}}} \frac{\partial h_2}{\partial P_{\mathbf{A}}} > 0$, we only need to prove that $\left| \frac{\partial h_1}{\partial P_{\mathbf{A}}} \right| > \frac{\partial h_2}{\partial P_{\mathbf{A}}}$ and $\left| \frac{\partial h_2}{\partial P_{\mathbf{B}}} \right| > \frac{\partial h_1}{\partial P_{\mathbf{B}}}$.

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

$$\begin{aligned}\frac{\partial h_1}{\partial P_{\mathbf{A}}} &= (\gamma\beta_{\mathbf{G}_1} + (1-\gamma)\beta_{\mathbf{G}_2}) \frac{\partial \mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}} + \gamma(1-\beta_{\mathbf{G}_1}) \frac{\partial \mathbb{P}_{\theta_1}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} \\ &\quad + (1-\gamma)(1-\beta_{\mathbf{G}_2}) \frac{\partial \mathbb{P}_{\theta_2}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}, \\ \frac{\partial h_2}{\partial P_{\mathbf{A}}} &= \gamma\beta_{\mathbf{G}_1} \frac{\partial \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} + (1-\gamma)\beta_{\mathbf{G}_2} \frac{\partial \mathbb{P}_{\theta_2}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.\end{aligned}$$

All terms of h_1 are decreasing in $P_{\mathbf{A}}$ and all terms of h_2 are increasing in $P_{\mathbf{A}}$, therefore we can proceed term by term:

$$\begin{aligned}\left| \gamma\beta_{\mathbf{G}_1} \frac{\partial \mathbb{P}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}} \right| &= \gamma\beta_{\mathbf{G}_1} \frac{\partial \mathbb{P}(W_{\mathbf{A}} < P_{\mathbf{A}})}{\partial P_{\mathbf{A}}}, \\ &= \gamma\beta_{\mathbf{G}_1} \left(\frac{\partial \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} + \frac{\partial \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}} \right), \\ &> \gamma\beta_{\mathbf{G}_1} \frac{\partial \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.\end{aligned}\tag{18}$$

The same reasoning, when replacing θ_1 by θ_2 in the (18) shows that

$$\left| (1-\gamma)\beta_{\mathbf{G}_2} \frac{\partial \mathbb{P}_{\theta_2}(W_{\mathbf{A}} \geq P_{\mathbf{A}})}{\partial P_{\mathbf{A}}} \right| > (1-\gamma)\beta_{\mathbf{G}_2} \frac{\partial \mathbb{P}_{\theta_2}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} \geq P_{\mathbf{B}})}{\partial P_{\mathbf{A}}}.$$

We conclude that $\left| \frac{\partial h_1}{\partial P_{\mathbf{A}}} \right| > \frac{\partial h_2}{\partial P_{\mathbf{A}}}$, and similarly $\left| \frac{\partial h_2}{\partial P_{\mathbf{B}}} \right| > \frac{\partial h_1}{\partial P_{\mathbf{B}}}$. Therefore the Jacobian in (17) 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 θ_1 , and a function $\psi : V \rightarrow U$ such that for all $(x, y) \in \mathbb{R}^2$, $\theta \in \Theta$,

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

In particular, $(P_{\mathbf{A}}, P_{\mathbf{B}}) = \psi(\theta_1)$, and we can compute the derivative of ψ :

$$\begin{aligned}
J_\psi(\theta_1) &= -J_{h,(P_{\mathbf{A}}, P_{\mathbf{B}})}(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1)^{-1} J_{h,\theta_1}(P_{\mathbf{A}}, P_{\mathbf{B}}, \theta_1), \\
&= \frac{-1}{\frac{\partial h_1}{\partial P_{\mathbf{A}}} \frac{\partial h_2}{\partial P_{\mathbf{B}}} - \frac{\partial h_1}{\partial P_{\mathbf{B}}} \frac{\partial h_2}{\partial P_{\mathbf{A}}}} \begin{pmatrix} \frac{\partial h_2}{\partial P_{\mathbf{B}}} & -\frac{\partial h_1}{\partial P_{\mathbf{B}}} \\ -\frac{\partial h_2}{\partial P_{\mathbf{A}}} & \frac{\partial h_1}{\partial P_{\mathbf{A}}} \end{pmatrix} \begin{pmatrix} \frac{\partial h_1}{\partial \theta_1} \\ \frac{\partial h_2}{\partial \theta_1} \end{pmatrix}, \\
&= \frac{-1}{\frac{\partial h_1}{\partial P_{\mathbf{A}}} \frac{\partial h_2}{\partial P_{\mathbf{B}}} - \frac{\partial h_1}{\partial P_{\mathbf{B}}} \frac{\partial h_2}{\partial P_{\mathbf{A}}}} \begin{pmatrix} \frac{\partial h_2}{\partial P_{\mathbf{B}}} \frac{\partial h_1}{\partial \theta_1} - \frac{\partial h_1}{\partial P_{\mathbf{B}}} \frac{\partial h_2}{\partial \theta_1} \\ -\frac{\partial h_2}{\partial P_{\mathbf{A}}} \frac{\partial h_1}{\partial \theta_1} + \frac{\partial h_1}{\partial P_{\mathbf{A}}} \frac{\partial h_2}{\partial \theta_1} \end{pmatrix}. \tag{19}
\end{aligned}$$

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 only need those in θ_1 . The only term in h_1 that depends on θ_1 is $\gamma(1 - \beta_{\mathbf{G}_1})\mathbb{P}_{\theta_1}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})$. By Lemma 2.2, $\mathbb{P}_{\theta_1}(W_{\mathbf{A}} \geq P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})$ is decreasing in θ_1 , and thus $\frac{\partial h_1}{\partial \theta_1} < 0$. By the same argument, $\frac{\partial h_2}{\partial \theta_1}$ is also negative. (Note that here the implicit functions theorem requires that we compute the partial derivatives of h as if $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ were not functions of θ_1 .)

If we replace each term of the last line of (19) by its signs, 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 θ_1 , and finally:

$$\frac{\partial V_1^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_1}, \frac{\partial V_1^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_1}, \frac{\partial V_1^{\mathbf{G}_2, \mathbf{A}}}{\partial \theta_1}, \frac{\partial V_1^{\mathbf{G}_2, \mathbf{B}}}{\partial \theta_1} > 0.$$

The problem being symmetric between θ_1 and θ_2 , we also conclude that

$$\frac{\partial V_1^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_2}, \frac{\partial V_1^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_2}, \frac{\partial V_1^{\mathbf{G}_2, \mathbf{A}}}{\partial \theta_2}, \frac{\partial V_1^{\mathbf{G}_2, \mathbf{B}}}{\partial \theta_2} > 0,$$

which concludes the proof of the theorem.

Notice that we require $\theta_1, \theta_2 \in \overset{\circ}{\Theta}$ because if, when one of the θ is maximal, the distribution is fully correlated (i.e., $W_{\mathbf{B}}$ is a deterministic function of $W_{\mathbf{A}}$), then the V_1 metrics are not differentiable at this point. However, they are continuous, therefore they are increasing on the whole interval Θ . Moreover, if the distribution is not fully correlated when θ is maximal, then we can replace $\overset{\circ}{\Theta}$ by Θ in the statement of the theorem.

A.4.4 Proof of Corollary 4.8.

We showed in the proof of Theorem 4.7 that $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ are decreasing in both θ_1 and θ_2 . By using this, we have:

$$\frac{\partial V_{\emptyset}^{\mathbf{G}_2, \mathbf{A}}}{\partial \theta_1} = \frac{\partial V_{\emptyset}^{\mathbf{G}_2, \mathbf{B}}}{\partial \theta_1} = \frac{\partial \mathbb{P}_{\theta_2}(W_{\mathbf{A}} < P_{\mathbf{A}}, W_{\mathbf{B}} < P_{\mathbf{B}})}{\partial \theta_1} < 0.$$

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

$$\gamma \beta_{\mathbf{G}_1} V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}} + \gamma(1 - \beta_{\mathbf{G}_1}) V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}} + (1 - \gamma) \beta_{\mathbf{G}_2} V_{\emptyset}^{\mathbf{G}_2, \mathbf{A}} + (1 - \gamma)(1 - \beta_{\mathbf{G}_1}) V_{\emptyset}^{\mathbf{G}_2, \mathbf{B}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}.$$

Rearranging the terms gives

$$\begin{aligned} \gamma \beta_{\mathbf{G}_1} \frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_1} + \gamma(1 - \beta_{\mathbf{G}_1}) \frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_1} &= -((1 - \gamma) \beta_{\mathbf{G}_2} \frac{\partial V_{\emptyset}^{\mathbf{G}_2, \mathbf{A}}}{\partial \theta_1} + (1 - \gamma)(1 - \beta_{\mathbf{G}_1}) \frac{\partial V_{\emptyset}^{\mathbf{G}_2, \mathbf{B}}}{\partial \theta_1}). \\ &> 0 \end{aligned}$$

Since $V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}} = V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}$, we conclude that both $\frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_1}$ and $\frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_1}$ are positive.

The same argument shows that $\frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_2} = \frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_2} < 0$ and $\frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{A}}}{\partial \theta_1} = \frac{\partial V_{\emptyset}^{\mathbf{G}_1, \mathbf{B}}}{\partial \theta_1} > 0$. From there, it is trivial that L , which is the distance between $V_{\emptyset}^{\mathbf{G}_1}$ and $V_{\emptyset}^{\mathbf{G}_2}$ is increasing in the correlation level of the high correlation group and decreasing in that of the low-correlation group.

A.4.5 Proof of Corollary 4.9.

V_1 is a convex combination of the first choice functions that are increasing in both θ_1 and θ_2 , so it is increasing in those two parameters too. Moreover it is continuous, and we assumed Θ to be an interval, so the set of possible values for V_1 is an interval. Let us fix $V \in (V_1^{\min}, V_1^{\max})$, and consider the solutions of the equation $V(\theta_1, \theta_2) = V$. By continuity, this equation has a solution, and the implicit function theorem immediately shows that there exists an interval $(\theta_{\min}^V, \theta_{\max}^V)$ and a function ϕ such that $\forall \theta_1 \in (\theta_{\min}^x, \theta_{\max}^x), V(\theta_1, \phi(\theta_1)) = V$. Since $V(\theta_1, \theta_2)$ is increasing in both arguments, ϕ is necessarily decreasing. Finally, since by Theorem 4.7 V_1^{\min} is achieved for $\theta_1 = \theta_2 = \min(\Theta)$ and V_1^{\max} for $\theta_1 = \theta_2 = \max(\Theta)$, then any $V \in [V_1^{\min}, V_1^{\max}]$ can be achieved with $\theta_1 = \theta_2$, which implies $L = 0$.

A.4.6 Proof of Proposition 5.1.

The relations between the V_1 come from Theorem 4.5. If $\alpha_{\mathbf{A}} + \alpha_{\mathbf{B}} \geq 1$, then all students are admitted to some college, therefore either $P_{\mathbf{A}} = \mathbf{I}_{\mathbf{A}}^l$ or $P_{\mathbf{B}} = \mathbf{I}_{\mathbf{B}}^l$. Let us suppose it is $P_{\mathbf{A}}$. Then we have

$$\begin{aligned} V_2^{\mathbf{G}^1, \mathbf{A}}(\theta_1, \theta_2) &= \mathbb{P}_{\theta_1}(W_{\mathbf{A}} < \mathbf{I}_{\mathbf{A}}^l, W_{\mathbf{B}} \geq P_{\mathbf{B}}), \\ &= 0, \\ &= \mathbb{P}_{\theta_2}(W_{\mathbf{A}} < \mathbf{I}_{\mathbf{A}}^l, W_{\mathbf{B}} \geq P_{\mathbf{B}}), \\ &= V_2^{\mathbf{G}^1, \mathbf{A}}(\theta_1, \theta_2), \end{aligned}$$

and

$$\begin{aligned} V_2^{\mathbf{G}^1, \mathbf{B}}(\theta_1, \theta_2) &= \mathbb{P}_{\theta_1}(W_{\mathbf{A}} \geq \mathbf{I}_{\mathbf{A}}^l, W_{\mathbf{B}} < P_{\mathbf{B}}) \\ &= \mathbb{P}(W_{\mathbf{B}} < P_{\mathbf{B}}) \\ &= \mathbb{P}_{\theta_2}(W_{\mathbf{A}} \geq \mathbf{I}_{\mathbf{A}}^l, W_{\mathbf{B}} < P_{\mathbf{B}}) \\ &= V_2^{\mathbf{G}^2, \mathbf{B}}(\theta_1, \theta_2). \end{aligned}$$

Since every student is matched, all the V_{\emptyset} are 0. We assumed $P_{\mathbf{A}} = \mathbf{I}_{\mathbf{A}}^l$, if it was $P_{\mathbf{B}}$ instead we would obtain the same result as the problem is symmetric between \mathbf{A} and \mathbf{B} . Finally, all these quantities are constant in θ_1 and θ_2 : the only ones that are non-zero are of the type $\mathbb{P}(W_{\mathbf{C}} < P_{\mathbf{C}})$ and therefore constant in the correlation.

A.4.7 Proof of Proposition 5.2

For this proof, to simplify computations, assume without loss of generality that marginals 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 help solving the market-clearing equation 16. The students' score 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:

$$\begin{aligned} V_1^{\mathbf{A}} &= 1 - P_{\mathbf{A}}, & V_1^{\mathbf{B}} &= 1 - P_{\mathbf{B}}, \\ V_2^{\mathbf{A}} &= \max(P_{\mathbf{A}} - P_{\mathbf{B}}, 0), & V_2^{\mathbf{B}} &= \max(P_{\mathbf{B}} - P_{\mathbf{A}}, 0), \\ V_{\emptyset}^{\mathbf{A}} &= \min(P_{\mathbf{A}}, P_{\mathbf{B}}), & V_{\emptyset}^{\mathbf{B}} &= \min(P_{\mathbf{A}}, P_{\mathbf{B}}). \end{aligned} \tag{20}$$

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

$$\begin{cases} \beta(1 - P_{\mathbf{A}}) + (1 - \beta)(P_{\mathbf{B}} - P_{\mathbf{A}}) &= \alpha_{\mathbf{A}}, \\ (1 - \beta)(1 - P_{\mathbf{B}}) &= \alpha_{\mathbf{B}}, \end{cases} \quad (21)$$

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:

$$\begin{cases} P_{\mathbf{A}} = 1 - \frac{\alpha_{\mathbf{A}}}{\beta}, \\ P_{\mathbf{B}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}. \end{cases} \quad (22)$$

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 (20) the values of $P_{\mathbf{A}}$ and $P_{\mathbf{B}}$ found in (21) and (22).

A.4.8 Proof of Proposition 5.3

As in the proof of Proposition 5.2, we assume without loss of generality for this proof that the marginals are uniform over $[0, 1]$. Then the grades at colleges \mathbf{A} and \mathbf{B} are independent random variables with a uniform distribution over $[0, 1]$. Students' score 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:

$$\begin{aligned} V_1^{\mathbf{A}} &= 1 - P_{\mathbf{A}}, & V_1^{\mathbf{B}} &= 1 - P_{\mathbf{B}}, \\ V_2^{\mathbf{A}} &= P_{\mathbf{A}}(1 - P_{\mathbf{B}}), & V_2^{\mathbf{B}} &= P_{\mathbf{B}}(1 - P_{\mathbf{A}}), \\ V_{\emptyset}^{\mathbf{A}} &= P_{\mathbf{A}}P_{\mathbf{B}}, & V_{\emptyset}^{\mathbf{B}} &= P_{\mathbf{A}}P_{\mathbf{B}}. \end{aligned} \quad (23)$$

The market-clearing equation is:

$$\begin{aligned} &\begin{cases} \beta(1 - P_{\mathbf{A}}) + (1 - \beta)P_{\mathbf{B}}(1 - P_{\mathbf{A}}) = \alpha_{\mathbf{A}}, \\ \beta P_{\mathbf{A}}(1 - P_{\mathbf{B}}) + (1 - \beta)(1 - P_{\mathbf{B}}) = \alpha_{\mathbf{B}}, \end{cases} \\ \Leftrightarrow &\begin{cases} P_{\mathbf{B}} = 1 - \alpha_{\mathbf{B}} - \frac{\beta}{1-\beta}(1 - P_{\mathbf{A}} - \alpha_{\mathbf{A}}), \\ P_{\mathbf{A}}P_{\mathbf{B}} = 1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}, \end{cases} \\ \Leftrightarrow &\begin{cases} P_{\mathbf{B}} = 1 - \alpha_{\mathbf{B}} - \frac{\beta}{1-\beta}(1 - P_{\mathbf{A}} - \alpha_{\mathbf{A}}), \\ \frac{\beta}{1-\beta}P_{\mathbf{A}}^2 + (\frac{1-2\beta}{1-\beta} + \frac{\beta}{1-\beta}\alpha_{\mathbf{A}} - \alpha_{\mathbf{B}})P_{\mathbf{A}} - (1 - \alpha_{\mathbf{A}} - \alpha_{\mathbf{B}}) = 0. \end{cases} \end{aligned} \quad (24)$$

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 (24) 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 (23) finally gives

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

which concludes the proof of the proposition.

A.4.9 Proof of Proposition 5.5.

We start by building a distribution family that can represent both STB and MTB for two values of the parameter. The priority classes are $K_{\mathbf{A}}^1, \dots, K_{\mathbf{A}}^{n_{\mathbf{A}}}$ and $K_{\mathbf{B}}^1, \dots, K_{\mathbf{B}}^{n_{\mathbf{B}}}$, and we denote by $\kappa_{\mathbf{C}}^j = \eta(K_{\mathbf{C}}^j)$ the mass of students inside class j of college \mathbf{C} . Let $a_0 = 0, a_1 = \kappa_{\mathbf{A}}^1, a_2 = \kappa_{\mathbf{A}}^1 + \kappa_{\mathbf{A}}^2, \dots, a_{n_{\mathbf{A}}} = 1$, such that they form a partition of $[0, 1]$ with the j -th segment having length $\kappa_{\mathbf{A}}^j$. Define $b_0, \dots, b_{n_{\mathbf{B}}}$ similarly. Finally, for any $i \leq n_{\mathbf{A}}, j \leq n_{\mathbf{B}}$, let $\kappa^{i,j} = \eta(K_{\mathbf{A}}^i \times K_{\mathbf{B}}^j)$ be the mass of students belonging to class i for \mathbf{A} and class j for \mathbf{B} .

Let ϕ_r be the pdf of the Gaussian copula with uniform marginals on $[0, 1]^2$ and covariance r . For any $r \in [0, 1]$, let $f_r : [0, 1]^2 \rightarrow \mathbb{R}$ be defined as:

$$f_r(x, y) = \kappa^{i,j} \phi_r\left(\frac{x - a_{i-1}}{\kappa_{\mathbf{A}}^i}, \frac{y - b_{j-1}}{\kappa_{\mathbf{B}}^j}\right) \text{ with } a_{i-1} \leq x \leq a_i, b_{j-1} \leq y \leq b_j$$

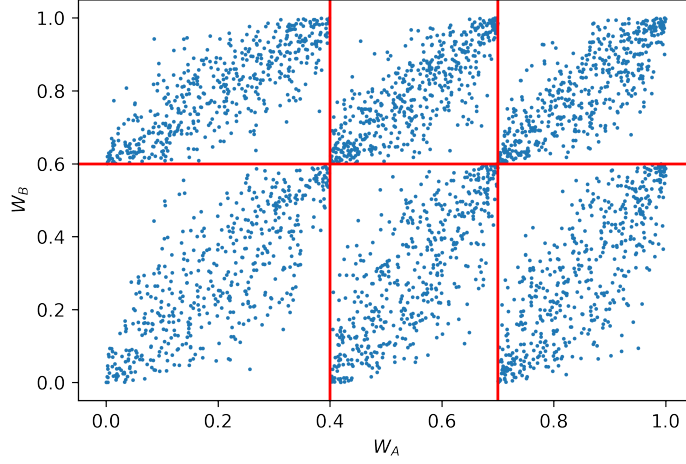


Figure 5: Example of the distribution f_r with three classes at **A**, two at **B**, and $r = 0.8$.

Defined this way, f_r is a pdf since it is non-negative and has integral 1. The marginals are uniform and do not depend on r . Moreover, the integral of f_r over each rectangle $K_{\mathbf{A}}^i \times K_{\mathbf{B}}^j$ is $\kappa^{i,j}$, and each rectangle contains a “copy” of the Gaussian copula adjusted to its dimensions. There is no “spill” between classes: if student s is in a higher priority class at college **C** than student s' , then s will have a higher score with probability 1. If for all i, j , $\kappa^{i,j} > 0$, and $r \notin \{-1, 1\}$, then f_r has full support. This distribution is depicted in Figure 5.

When $r = 0$, this distribution represents MTB: if two students are in the same priority class for a college, they each have an ex-ante probability of 0.5 of winning the tie-breaking, and if they also are in the same priority class for the other college (i.e., they are in the same rectangle $K_{\mathbf{A}}^i \times K_{\mathbf{B}}^j$), the result of this second tie-breaking is independent from the first one.

When $r = 1$, this distribution represents STB: if two students are in the same priority class for a college, they once again each have an ex-ante probability of 0.5 of winning the tie-breaking. But if they also are in the same priority class for the other college (i.e., they are in the same rectangle $K_{\mathbf{A}}^i \times K_{\mathbf{B}}^j$), the winner of the tie-breaking is the same as for the first college since scores inside the rectangle are perfectly correlated. In that case, the correlation does not have full support but this is not an issue as explained in Remark 4.3.

The family $(f_r)_{r \in [0,1]}$ is differentiable by differentiability of the Gaussian copula. It is also coherent (except for the (x, y) such that $x = a_i$ or $y = b_j$, i.e., on the sides of rectangles, in that case the cdf is constant and not increasing). Therefore we can apply Theorem 4.7 and deduce the first part of the proposition.

Suppose that $\exists \theta \in \Theta$ such that $P_{\mathbf{A}}(\theta) \neq a_i$ and $P_{\mathbf{B}}(\theta) \neq b_j$ for all i, j . We can then apply Theorem 4.7, and deduce that $V_1^{\mathbf{A}}$ and $V_1^{\mathbf{B}}$ are increasing they are greater for $r = 1$ than for $r = 0$. If there exists no such θ , it implies that $P_{\mathbf{A}}, P_{\mathbf{B}}$ are constant in θ and so are $V_1^{\mathbf{A}}$ and $V_1^{\mathbf{B}}$, and

thus STB and MTB are equivalent. However, if we assume that students' preferences or colleges' capacities are drawn from distributions that admit densities, the probability of this event is zero, which gives the second part of the proposition.