Stable Matching as Transportation[1]

Federico Echenique[2] Joseph Root[3] Fedor Sandomirskiy[4]

[1]The paper has benefited from suggestions and comments by Sophie Bade, Peter Biro, Mark Braverman, Ben Brooks, Modibo Camara, Ryota Iijima, David Kempe, Axel Niemeyer, Afshin Nikzad, Piyush Panigrahi, Marek Pycia, Doron Ravid, Kirill Rudov, Utku Unver, Rakesh Vohra, Matt Weinberg, and audiences at ASU, U. of Arizona, Berkeley, Caltech, Emory, Harvard, NYU, MIT, MSRI, Princeton, Simons Institute, UPenn, Washington University, and Yale.

[2]Department of Economics, University of California, Berkeley

[3]Department of Economics, University of Chicago

[4]Department of Economics, Princeton University

Abstract

This paper links matching markets with aligned preferences and the theory of optimal transport. We show that common design objectives—stability, efficiency, and fairness—arise as solutions to a parametric family of optimal transportation problems, where the parameter captures society's preference for inequality. This connection yields new insights into the structural properties of matchings and the trade-offs between different objectives, revealing how pursuing stability can generate significant disparities in welfare distribution, even among similar agents. Our framework provides a tractable stylized model capturing supply-demand imbalances in a range of settings, including spatial matching markets, school choice, organ donor exchange, partnership formation, and markets post-match bargaining over transfers. We also demonstrate that large markets with idiosyncratic preferences can often be approximated by models with aligned preferences, broadening the applicability of our findings.

1 Introduction

Our paper establishes a connection between stability in matching markets and the class of optimization problems studied in the theory of *optimal transport*. Stability is the guiding principle in many practical instances of market design, especially in centralized labor markets and in school choice programs. The link to optimal transport unveils new phenomena in matching theory, providing insights into the consequences of stability for equity and welfare, the structure of stable matchings, and computational aspects of stability.

We introduce an optimal transportation problem that includes a transformation of the agents' utilities, modulated by a real-valued parameter α. This parameter α acts as a "tuning knob" that controls the trade-off between fairness and stability. When α is dialed up, the transportation problem delivers an approximately stable matching. The larger the value of α, the closer the matching is to being fully stable. Dialing α down steers the optimal transportation solution towards a more equitable matching. These findings highlight a fundamental tension between fairness and stability, a trade-off that can explain stylized facts observed in real-world matching markets, such as ride-sharing platforms and school choice programs.

The connection to optimal transport furnishes stability with a new economic interpretation. Stability is traditionally viewed as a game-theoretic equilibrium notion rooted in cooperative game theory. In the context of school choice, stability has also been interpreted as a normative fairness property. By demonstrating that stability may *implicitly* be optimizing a particular objective function, we can ascribe new, perhaps unintentional, goals to systems that implement stable matchings.

The objective function in question turns out to be a version of Atkinson's inequality index, where the parameter α reflects a preference for, or against, inequality. When α is dialed up, the objective becomes a convex transformation of agents' utilities, and therefore *favors* inequality. The larger α is, the more convex the objective function, and thus the stronger the preference towards inequality. As a result, stable matchings may exhibit highly unequal distributions of welfare, with discontinuities where very similar groups of agents obtain significantly dissimilar outcomes. While existing literature on stable matching has focused on unfairness *across* the market, highlighting the existence of extremal matchings that favor one side of the market over the other, our results emphasize the unfairness that can arise *within* the same side of the market due to supply and demand imbalances.

We study stability in a model where agents have aligned preferences—a framework that, as we argue, can capture the fundamental forces driving a variety of markets, at least approximately. In an aligned market, a single utility function $u(x, y)$ represents preferences on both sides of the market. When agents x and y are matched, they each receive a utility of $u(x, y)$. Importantly, alignment does not mean identical preferences over potential partners, but rather that the utility from a specific match is the same for both agents involved.

Spatial matching markets are an important special case of our model and a rich source of applications. Here, agents are represented as points in an Euclidean space, their "addresses." Aligned preferences arise when the agents prefer partners who are close to them, corresponding to u being the negative distance. For agents on the real line, stable matchings exhibit a simple and intuitive structure, characterized by a no-crossing property. This one-dimensional case, discussed in Section 2, motivates the general connection to optimal transport and illustrates the tension between stability, fairness, and efficiency, including a possible loss in overall welfare under stable matching.

The assumption of aligned preferences is, of course, restrictive. However, we contend that aligned markets constitute a useful and interesting class. They can arguably capture the main tensions in important applications, allowing us to focus on the basic driving forces in the market. A case in point is school choice. It is well documented that the distance between a student's home and the school is a key component of students' preferences, as well as of schools' priorities over students (Walters, 2018; Laverde, 2022). By focusing on the resulting aligned market, we can better understand the consequences of using stability as a solution in school choice. We also argue that our conclusions hold even if schools and students care about other aspects in addition to distance, such as a "vertical" quality component (see Section 4).

Aligned preferences are more general and cover more ground than one might initially think. In particular, we can allow for significant deviations from alignment and still conclude that the aligned model is a good approximation as long as the market is large. In Section 3.3, we consider agents with utilities that are the sum of an aligned and a non-aligned idiosyncratic component. The presence of an aligned component seems realistic, as many matching markets feature some objective notion of "match quality." We show that, even if the non-aligned component dominates, a large market is well approximated by a model that only uses the aligned utilities. The result in Section 3.3 is, we think, of independent interest and features methodological contributions that seem to be new in matching theory.

Our model is otherwise very general, allowing for both finite and infinite markets, including those with and without atoms. We consider this generality a strength, as it introduces no significant complications while offering a unified framework. In contrast, a more traditional approach taken by the stable matching literature is to extend the finite model to large economies and then deal with the complexities of large but finite markets. Our approach sidesteps these issues. Furthermore, our infinite market model delivers particularly stark conclusions, free from the combinatorial intricacies of finite markets, thus highlighting the core issues. For example, we show that, in a stable matching, two arbitrarily similar agents can obtain very dissimilar welfare, and so one can envy the other. While such discontinuities can only be possible with a continuum of agents, they crystallize forces also present, albeit less sharply, in finite markets.

1.1 Related literature

The economic literature on matching markets with aligned preferences is sparse. A systematic study of markets with aligned preferences was initiated by Ferdowsian, Niederle, and Yariv (2023) and Niederle and Yariv (2009), who examined the convergence of decentralized matching dynamics in such markets. The observation that a stable matching for a finite population with aligned preferences can be obtained via a greedy algorithm—iteratively matching the unmatched pair with the highest utility—is implicitly contained in Eeckhout (2000) and Clark (2006), who provide conditions for uniqueness of a stable matching; see also Gutin, Neary, and Yeo (2023); Reny (2021). Galichon, Ghelfi, and Henry (2023) extend the greedy algorithm to many-to-one stable matching and observe its unfairness in simulations, a phenomenon we explore further.

Our paper contributes to understanding of the stability-efficiency trade-off in markets with aligned preferences, complementing insights by Cantillon, Chen, and Pereyra (2022) and Lee and Yariv (2018). Cantillon, Chen, and Pereyra (2022) introduce a school choice environment where preference alignment becomes a descriptive property. Under a condition generalizing alignment, they demonstrate that stable matchings are Pareto optimal. By contrast, our work adopts a quantitative approach to efficiency, demonstrating that Pareto optimality may entail a loss of utilitarian welfare. Lee and Yariv (2018) study large markets with random preferences, showing that when a common utility for each pair is drawn independently from a continuous distribution, the stable matching achieves utilitarian welfare close to the optimum. Our results reveal that this phenomenon is specific to settings with no correlation across agents' preferences: in the presence of correlation, stability may reduce welfare by up to half of the optimal level.

A broader perspective on preference alignment is provided by Pycia (2012), who examines a general coalition-formation model that includes many-to-one matching as a special case. His work focuses on allowing for complementarities and peer effects, proving the existence of stable outcomes under a richness assumption on feasible coalitions and preferences. Echenique and Yenmez (2007) make a related point for many-to-one matching, building on the earlier ideas of Banerjee, Konishi, and Sönmez (2001) in coalition formation. Pycia's work also precedes our discussion in Section 4.3 on the connection between preference alignment and second-stage bargaining.

Our study touches on several other strains of literature. In computer science, aligned preferences have appeared under the name of "globally ranked pairs." Abraham, Levavi, Manlove, and O'Malley (2008) analyze aligned preferences in roommates problem and Lebedev, Mathieu, Viennot, Gai, Reynier, and De Montgolfier (2007) apply them to peer-to-peer networks. For an empirical use of markets with aligned preferences, see, for example, Agarwal (2015) and Sørensen (2007). Our findings are potentially applicable to models of international trade, where a growing literature emphasizes the spatial aspects of buyer-seller networks; see, for instance, Chaney (2014); Antras, Fort, and Tintelnot (2017); Panigrahi (2021).

The connection to optimal transport established in our paper reinforces the common wisdom that matching models with a continuum of agents can be more intuitive than their finite-population counterparts; see, for example, Ashlagi and Shi (2016); Azevedo and Leshno (2016); Leshno and Lo (2021); Arnosti (2022). With optimal transport methods, the cardinality of the space of agents becomes largely irrelevant for the problems we address. Connections to optimal transport have recently been discovered and used in various areas of economic theory, including mechanism design (Daskalakis, Deckelbaum, and Tzamos, 2015; Perez-Richet and Skreta, 2023; Kolesnikov, Sandomirskiy, Tsyvinski, and Zimin, 2022; McCann and Zhang, 2023), information design (Malamud, Cieslak, and Schrimpf, 2021; Arieli, Babichenko, and Sandomirskiy, 2023), and many others; see surveys by Ekeland (2010); Carlier (2012); Galichon (2018). We find a link between stability and a particular area of optimal transport known as concave transport, pioneered by McCann (1999); Gangbo and McCann (1996). To the best of our knowledge, concave transport has not appeared in economic applications, with the exception of Boerma, Tsyvinski, Wang, and Zhang (2023), who use it to study labor market sorting—a matching market with transferable utility. Their work is the closest to ours in the technical dimension. While the literature on markets with and without transfers is almost disjoint as such markets have few similarities, our shared reliance on the concave transport perspective highlights its generality and importance for understanding both settings.

2 Motivating examples: matching on the line

We begin by discussing three examples. In all of them, matching has a spatial interpretation. Agents are identified with a real number that describes their position on the real line. The utility for any agent $x \in \mathbb{R}$ who is matched with an agent $y \in \mathbb{R}$ is $-|x-y|$; so that everyone prefers to be matched with their closest possible partner. This simple setting is sufficiently rich to suggest many of our findings, and will motivate our use of optimal transport.

Matching on the line can be thought of as an approximation to those matching markets where distance is the key driver of preferences. For example, in school choice markets, schools generally prioritize students who live nearby. Likewise, students generally prefer close schools. In Section 4, we develop and further discuss the application to school choice.

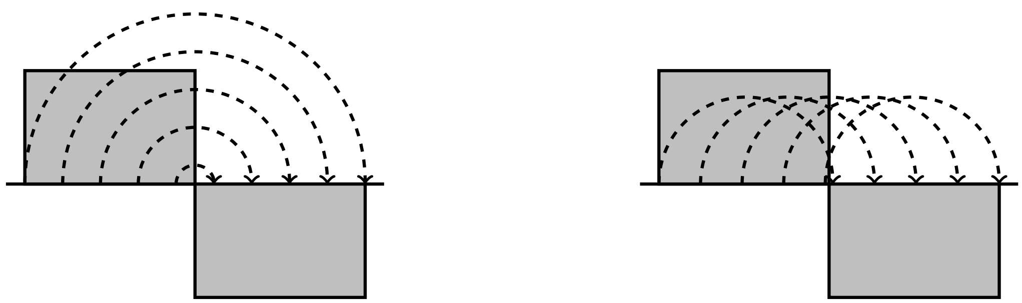
Example 1. The simplest useful example is a market with two agents on either side. Let $X = \{0, 1\}$ and $Y = \{2, 3\}$ be the two sides of the market: There are only two matchings. The first matching, π^{f}, matches 0 with 2 and 1 with 3. Then all agents are matched with a partner 2 units away, and utilitarian welfare is -8. Recall that a matching is stable if no two agents prefer to leave their partners and match together. Notice that π^{f} is not stable; 1 and 2 could leave their partners and match together.

The second matching, π^{s}, has 0 matched with 3 and 1 matched with 2. It is easy to check that this matching is stable. But note that, in π^{s}, two agents get utility -1 while two get utility -3. Utilitarian welfare is again -8, but more unevenly distributed than in π^{f}.

In fact, f may as well stand for "fair" instead of "first," and s for "stable," highlighting one of the central themes of our paper: the inherent tension between stability and fairness in matching markets.

Example 2. Now suppose that X is a population of agents uniformly distributed on the interval $[-1, 0]$, while Y is the population uniformly distributed on $[0, 1]$. We can visualize this market by plotting the distribution of X above the horizontal axis, and the population of Y below the horizontal axis. See Figure 1 for an illustration. It turns out that the matching π^{s} illustrated in Figure 1a by half-circles is the unique stable matching. In π^{s}, agents of type $-z \in [-1, 0]$ are matched with agents of type $z \in [0, 1]$. Notice that the stable matching induces substantial inequality in the value of the matches. Agents close to zero are paired with partners who have similar types to them, while agents far from zero are matched with those who have very different types.

Figure 1b, in contrast, illustrates the assortative matching π^{f}. All agents are matched with a partner who is one unit away. In π^{f} all agents obtain the same utility. It is, in particular, egalitarian, in the sense that it maximizes the utility of the worst-off agent. While the two matchings, stable and egalitarian, differ, utilitarian welfare is the same $(= -2)$ for both.



(a) Stable

(b) Egalitarian

Figure 1: Two solutions to the matching problem.

In Examples 1 and 2, there is a sharp tension between stability and fairness. However, both the stable matching and the egalitarian matching result in the same welfare. Example 2 also suggests a structural property of stable matchings on the line: in a stable matching, the half-circles do not cross. To what extent do these observations generalize? Our next example provides a step towards an answer.

Example 3. Suppose X has density 1 on the interval $[-2, -1]$ and density 2 on the interval $[0, 1]$. Assume that Y has density 1 on the intervals $[-1, 0]$ and $[1, 3]$. Figure 2 depicts the densities.

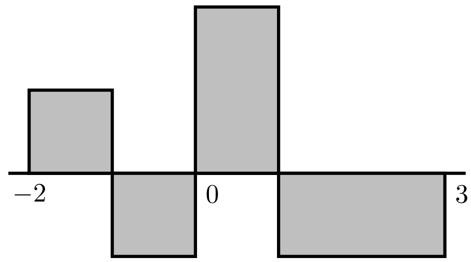
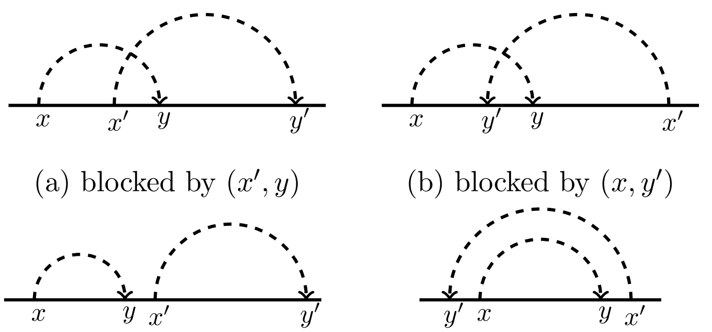


Figure 2: The density is above the axis for X, and below for Y.

Again, the assortative matching is egalitarian. However, the stable matching is harder to intuit. We make progress by establishing a structural property satisfied by stable matchings. In Examples 1 and 2, the stable matching satisfies a no-crossing property: the half-circles connecting two matched agents do not intersect. Formally, for any two real numbers z_1 and z_2, let $O(z_1, z_2)$ be the smallest circle containing the points $(z_1, 0)$ and $(z_2, 0)$ in $\mathbb{R}^2$.

A matching is defined as a probability measure on $X \times Y$ (see Section 3). A matching π satisfies **no-crossing** if, for any two pairs (x, y) and (x', y') in the support of π, the circles $O(x, y)$ and $O(x', y')$ do not intersect unless $x = x'$ or $y = y'$.



(a) blocked by (x', y) (b) blocked by (x, y')

(c) blocked by (x', y) (d) blocked by (x, y'), (x', y)

Figure 3: Forbidden patterns in stable matchings.

Lemma 1. *Any stable matching satisfies no-crossing.*

Proof. Let π be a stable matching. Towards a contradiction, suppose that (x, y) and (x', y') are in the support of π, where $x \neq x'$, $y \neq y'$, and the pair of circles $O(x, y)$ and $O(x', y')$ intersect. There are, up to symmetry, two cases to consider, depicted in Figures 3a and 3b. First, $x < x' < y < y'$ in which case $|x' - y| < |x - y|$

and $|x' - y| < |x' - y'|$, and so π is blocked by (x', y). Second, we could have $x < y' < y < x'$. Now $|x - y'| < |x - y|$ and $|x' - y| < |x' - y'|$. If π were stable we would have $|x - y| \leq |x' - y|$ and $|x' - y'| \leq |x - y'|$ however by transitivity $|x - y| < |x' - y'|$ and $|x' - y'| < |x - y|$. Thus π is not stable. $\square$

Note that, while no-crossing is necessary for stability, it is not sufficient: see Figures 3c and 3d.

Lemma 1 provides a first connection between stability and optimal transport. This no-crossing property has played a major role in the theory of one-dimensional optimal transport with concave costs developed by McCann (1999).[1] In particular, the set of matchings satisfying no-crossing is well-understood.

Returning to Example 3, no crossing narrows our search for the stable matching to a one-parametric family. This parameter, $\theta \in [0, 1]$, controls the portion of agents in the region $[-2, -1]$ who are matched with agents in the region $[1, 3]$. Three values of θ are shown in Figure 4.

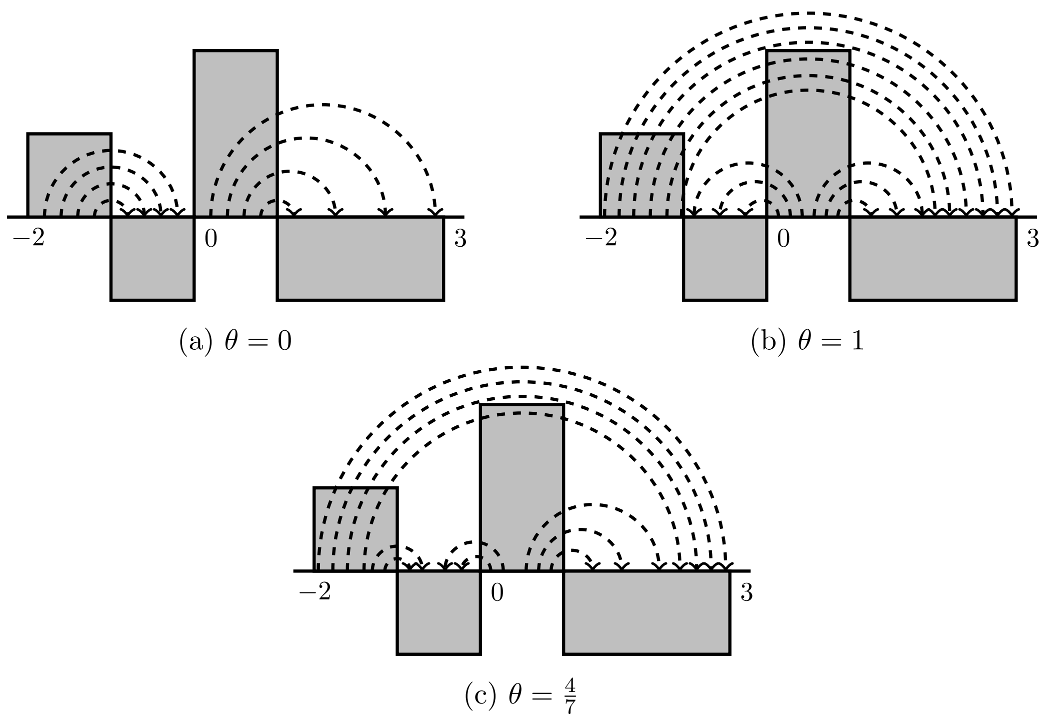


Figure 4: The matchings which satisfy no-crossing for three values of θ.

[1]According to Villani (2009), the no-crossing property dates back to papers by Monge, who initiated the study of optimal transport; see also the discussion by Boerma, Tsyvinski, Wang, and Zhang (2023).

One can verify that the only value of θ which results in a stable matching is $4/7$. In this stable matching, each type x is matched with $y = h(x)$ where

$$h(x) = \begin{cases} 1 - x, & x \in [-2,\ -1 - \frac{3}{7}] \\ -2 - x, & x \in (-1 - \frac{3}{7},\ -1] \\ -2x, & x \in (0,\ \frac{2}{7}] \\ 3 - 2x, & x \in (\frac{2}{7},\ 1] \end{cases}$$

as depicted by the half-circles in Figure 4c. Relative to the assortative matching, the stable matching induces inequality of outcomes. In the assortative matching, the worst-off couple is matched with a partner a distance 2 away. In the stable matching, any agent at -2 is matched with a partner at 3, giving utility -5. Unlike in our previous two examples, the stable match and the assortative match generate different levels of overall utilitarian welfare. The assortative matching results in total welfare of -5, while the stable matching generates welfare of roughly -7.2.

In the stable matching, arbitrarily similar agents can be matched with very different partners and experience vastly different levels of welfare. Indeed, one agent can envy another agent who is almost identical to them. For example, consider two agents x and x' with types $-\frac{10}{7} - \varepsilon$ and $-\frac{10}{7} + \varepsilon$ respectively, with $\varepsilon > 0$. Agent x is matched with a partner over 3 units away, while x' is matched within 1 unit, leading x to envy x'. Though considered "unjustified" in stable matching theory (as it does not cause instability), this envy highlights inherent inequality. It stems from supply and demand imbalances rather than the usual considerations in the field. Indeed, x is matched non-locally because of a scarcity of local partners $y \in [-1, 0]$, who have already formed matches with more preferred agents like x'. Thus, in stable matchings, high-utility couples impose negative externalities on low-utility ones reinforcing inequality.

Our main results (Section 3.1) establish a link between stability and optimal transport, demonstrating that approximate and exact stable matchings arise as solutions to specific optimal transport problems. This connection allows us to generalize the findings in the examples just discussed. After developing the general theory and examining some applications, we return to spatial matching in $\mathbb{R}^d$, including $d = 1$, in Section 6.

3 Stability, fairness, and optimal transport

We consider a general two-sided matching market, featuring two sets of agents that can be matched by forming pairs (an extension to multi-sided matching is in Section 7). Each side of the market is modeled as a population of types of agents.

Let X and Y be two sets of agents' types. The type of an agent encodes all the information that is relevant to calculate match utilities; for example their location in the real line in the examples of Section 2. We assume that X and Y are complete separable metric spaces endowed with their Borel sigma-algebras. For a space Z, we denote by $\mathcal{M}_+(Z)$ the set of positive measures on Z with a finite total mass. The distributions of agents' types over X and Y is represented by $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$. Accordingly, the two populations to be matched are given by the measure spaces (X, μ) and (Y, ν). For simplicity, we assume that the market is balanced, meaning that $\mu(X) = \nu(Y)$. It is, however, easy to accommodate unbalanced markets by adding "dummy" agents (see Example 5).

If agents with types $x \in X$ and $y \in Y$ are matched, they enjoy utilities $u(x, y)$ and $v(x, y)$, respectively, where u and v are measurable functions $X \times Y \to \mathbb{R}$. The marriage-market model of Gale and Shapley (1962) corresponds to finite X and Y with counting measures μ and ν. More generally, any market with atomic μ and ν having a finite number of atoms, all of the same mass, can be reduced to this classical model.

A distribution $\pi \in \mathcal{M}_+(X \times Y)$ is a **matching** if it has μ and ν as marginal distributions on, respectively, X and Y. We denote by $\Pi(\mu, \nu)$ the set of all matchings.

Stability. We first define stability without assuming that preferences are aligned, and then specialize the definition to the aligned case. A matching π is **ε-stable** with a parameter $\varepsilon \geq 0$ if for $\pi \times \pi$-almost all pairs (x_1, y_1) and (x_2, y_2) at least one of the following two inequalities holds

$$\begin{aligned} u(x_1, y_1) - u(x_1, y_2) &\geq -\varepsilon \\ v(x_2, y_2) - v(x_1, y_2) &\geq -\varepsilon \end{aligned} \tag{1}$$

If both inequalities are violated, (x_1, y_2) is called an **ε-blocking pair**. The definition effectively says that, in a (2×2) random submarket consisting of two couples (x_1, y_1) and (x_2, y_2) sampled from π independently, there is no ε-blocking pair almost surely.

For $\varepsilon = 0$, we will refer to ε-stable matchings as **stable**.[2]

If u and v are continuous, our notion of stability becomes equivalent to its more intuitive pointwise version resembling the classical definition of Gale and Shapley (1962). Recall that the support of π, denoted by $\mathrm{supp}(\pi)$, is the minimal closed set of full measure. The following lemma is proved in Supplementary Appendix B.

Lemma 2. *For continuous utilities u and v, a matching π is ε-stable if and only if for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$, at least one of the two inequalities (1) holds.*

The rest of the paper focuses on the special case of markets with aligned preferences. We say that the preferences of agents in X and Y are **aligned** if $u = v$, i.e., the utility of a match is the same for both sides of the market.

Beyond the motivation provided in the introduction, we further justify our focus on aligned preferences in Section 3.3 by demonstrating that, under certain conditions, even markets with significant preference misalignment can be approximated by assuming aligned preferences. Ordinal implications of alignment are explored in Section 5.

Under aligned preferences, stability boils down to a single inequality: π is stable if, for $\pi \times \pi$-almost all, pairs (x_1, y_1) and (x_2, y_2), the following inequality holds

$$u(x_1, y_2) \leq \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon. \tag{2}$$

Remark. When agents' types are identified with points on the real line, $X = Y = \mathbb{R}$ and $u(x, y) = -|x - y|$, the condition of stability (2) with $\varepsilon = 0$ reduces to the notion discussed in Section 2. The special case of the real line also illustrates that aligned preferences do not mean common preferences: two distinct types of agents, x and x' in X, will generally disagree on how they rank agents in Y.

Fairness. We now turn to a notion of fairness. To each matching π we assign a number $U_{\min}(\pi)$ equal to the minimal utility of a couple matched under π and denote by $U^*_{\min}(\mu, \nu)$ the best $U_{\min}(\pi)$ over all matchings[3]

$$U_{\min}(\pi) = \min_{(x,y)\in \mathrm{supp}[\pi]} u(x, y) \qquad \text{and} \qquad U^*_{\min}(\mu, \nu) = \max_{\pi \in \Pi(\mu,\nu)} U_{\min}(\pi).$$

[2]This notion of stability was introduced by Echenique, Lee, Shum, and Yenmez (2013), Kesten and Ünver (2015), and (in full generality) Greinecker and Kah (2021). For finite X and Y with a counting measures, the Birkhoff-von-Neumann theorem implies that any matching stable according to this notion can be represented as a convex combination of deterministic stable matchings.

[3]Under the assumption that X and Y are compact and u is continuous, $U_{\min}$ and $U^*_{\min}$ are well-defined. Indeed, the support of π is a closed set, a closed subset of a compact space is compact, and hence, the minimum is attained; the fact that the maximum is attained is established below in

This quantity $U^*_{\min}(\mu,\nu)$ is the egalitarian lower bound in the spirit of Rawls (1971): the highest utility level feasible for every couple in the population simultaneously.

A matching $\pi \in \Pi(\mu,\nu)$ is ε-**egalitarian** if there is a subset $S \subset X \times Y$ with $\pi(S) \geq (1-\varepsilon)\cdot \pi(X \times Y)$ such that

$$u(x,y) \geq U^*_{\min}(\mu,\nu) - \varepsilon \qquad \text{for any} \qquad (x,y) \in S.$$

In other words, for a large fraction of the whole population, the utility resulting from an ε-egalitarian matching almost satisfies the egalitarian lower bound. For $\varepsilon = 0$, we will refer to ε-egalitarian matchings as **egalitarian**.

Optimal transport. The canonical optimal transportation problem assumes a measurable cost function $c\colon X \times Y \to \mathbb{R}$, and marginal distributions $\mu \in M_+(X)$ and $\nu \in M_+(Y)$ with $\mu(X) = \nu(Y)$. The problem is to find a matching $\pi \in \Pi(\mu,\nu)$ that minimizes the total cost:

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x,y)\ \mathrm{d}\pi(x,y). \tag{3}$$

See Galichon (2018) and Villani (2009) for the basic theory of optimal transport.

The solutions to optimal transportation problems exhibit a "monotonicity" property. Given $c\colon X \times Y \to \mathbb{R}$, a set $\Gamma \subset X \times Y$ is called c-**cyclic monotone** if

$$\sum_{i=1}^{n} c(x_i,y_i) \leq \sum_{i=1}^{n} c(x_i,y_{i+1}) \tag{4}$$

for all $n \geq 2$ and pairs of points $(x_1,y_1),\ldots,(x_n,y_n) \in \Gamma$ with the convention $y_{n+1} = y_1$. The solutions to an optimal transportation problem are known to be supported on a c-cyclic monotone set.[4]

We shall see that the stability condition (2) arises naturally as a consequence of the c-cyclic monotonicity property in an auxiliary optimal transportation problem.

3.1 The main result

We now describe the link between stability, fairness, and optimal transport. For a matching market with aligned preferences represented by a utility function $u\colon X \times Y \to$

Corollary 1 even for non-compact X and Y. If u is discontinuous or X, Y are not compact, we replace the minimum with the essential infimum and the maximum with the supremum.

[4]See Kausamo, De Pascale, and Wyczesany (2023) for a survey of related results.

$\mathbb{R}$, we consider a parametric class of cost functions given by

$$c_\alpha(x,y) = \frac{1 - \exp(\alpha \cdot u(x,y))}{\alpha}. \tag{5}$$

A matching π induces a distribution of utility over the population. Then $\int c_\alpha(x,y)\, \mathrm{d}\pi(x,y)$ is a version of Atkinson's index of inequality.[5] The parameter α expresses a preference for, or aversion to, inequality. When $\alpha > 0$, c_α favors inequality; when $\alpha < 0$, c_α favors egalitarian outcomes.

Theorem 1. *Let π^* be a solution to the optimal transportation problem*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c_\alpha(x,y)\ \mathrm{d}\pi(x,y).$$

1. *If $\alpha > 0$, then π^* is ε-stable, with $\varepsilon = \frac{\ln 2}{\alpha}$.*
2. *If $\alpha < 0$, then π^* is ε-egalitarian, with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.*

Taking the limit $\alpha \to 0$ in (5) suggests that the cost c_α for $\alpha = 0$ must be defined by $c_0(x,y) = -u(x,y)$. Hence, for $\alpha = 0$, the transportation problem in Theorem 1 corresponds to maximizing the utilitarian social welfare

$$W(\pi) = \int_{X \times Y} u(x,y)\ \mathrm{d}\pi(x,y) \tag{6}$$

Thus, by changing α gradually from $-\infty$ to $+\infty$, we interpolate between *fairness*, *welfare*, and *stability* objectives. Theorem 2 below connects stability and utilitarian welfare maximization.

Remark (Existence of optimal matchings). To guarantee the existence of a solution in Theorem 1, it is enough to assume that u is continuous and bounded. No compactness assumptions on spaces X and Y are needed: a solution to (3) exist as long as c is lower semicontinuous and bounded from below (Villani, 2009, Theorem 4.1).

The proof of Theorem 1 is in Appendix A. Proving part 1 requires a result by Beiglböck, Goldstern, Maresch, and Schachermayer (2009), and boils down to checking that the cyclic monotonicity condition (4) for the cost function c_α from (5) implies

[5] It is a linear transformation of the negative (since it will be minimized) CARA version of Atkinson (1970). One can obtain a result similar to Theorem 1 with Atkinsons' actual CRRA index, as long as we impose that $u > 0$; see Appendix A.

the ε-stability condition (2). To prove part 2, we show that the contribution of low-utility couples to the transportation objective becomes dramatic for negative α, and thus π^* cannot place substantial weight on such couples.

Theorem 1 implies the existence of stable and egalitarian matchings under a continuity assumption on u. The idea is to construct these matchings as the weak limits of solutions to the optimal transportation problem from Theorem 1 for $\alpha \to \pm\infty$.

Denote by $\Pi^u_{+\infty}(\mu, \nu)$ the set of matchings π that can be obtained as the weak limit $\pi = \lim_{n\to+\infty} \pi_{\alpha_n}$ of sequences of solutions π_{α_n} to the transportation problem (3), with the cost c_{α_n} from (5), for some sequence of $\alpha_n \to +\infty$. Similarly define $\Pi^u_{-\infty}(\mu, \nu)$ to be the weak limits for some sequence $\alpha_n \to -\infty$.

Corollary 1. *For continuous and bounded utility u, the sets $\Pi^u_{+\infty}(\mu, \nu)$ and $\Pi^u_{-\infty}(\mu, \nu)$ are non-empty, convex, and weakly closed. All matchings in $\Pi^u_{+\infty}(\mu, \nu)$ are stable and all matchings in $\Pi^u_{-\infty}(\mu, \nu)$ are egalitarian.*

Corollary 1 provides the existence of stable and egalitarian matchings, and a computationally tractable way of finding them, as a limit of solutions to a sequence of linear programs.[6] The existence of stable matchings for markets with aligned preferences gives a special case of the general existence result proven in Greinecker and Kah (2021). In Appendix A, we check that Corollary 1 follows from Theorem 1.

Remark. For finite markets, it is enough to take α large enough to obtain exact stability. Indeed, if X and Y are finite, denote by δ the minimal change in utility that an agent can experience if they change partners:

$$\delta = \min\left\{ \min_{x,\ y\neq y'} |u(x,y) - u(x,y')|, \quad \min_{y,\ x\neq x'} |u(x,y) - u(x',y)| \right\}.$$

Then any ε-stable matching with $\varepsilon < \delta$ is automatically stable. Combining this observation with Theorem 1, we conclude that the solution of the optimal transportation problem with $\alpha > \delta/\ln 2$ is stable.

The conclusions of Theorem 1 are not specific to the exponential objective in c_α. As we demonstrate in Appendix A, it is enough to take

$$c(x,y) = -h\left(u(x,y)\right), \tag{7}$$

[6] The set of stable matchings does not change after a monotone parametrization of the utility function. Therefore, Corollary 1 implies the existence of stable matchings without boundedness assumption on utility u. Indeed, it is enough to replace u with a reparametrized utility $\tilde{u}(x,y) = \arctan(u(x,y))$.

where h is a monotone-increasing function with a large logarithmic derivative. Indeed, we show that if $h > 0$ is monotone increasing and $\frac{h'(t)}{h(t)} \geq \alpha$ for some positive α and t in the range of u, then any solution to the transportation problem with cost (7) is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. Similarly, if $h < 0$ is monotone increasing and satisfies $\frac{h'(t)}{h(t)} \geq \alpha$ with $\alpha < 0$, then the solution to the transportation problem is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.

3.2 Welfare and fairness of stable matchings

Theorem 1 shows that fairness and stability correspond to the two extremes of the α-spectrum, $+\infty$ and $-\infty$, respectively. This result suggests that stability may fail to provide any individual fairness guarantee. As we show here, even though the loss in fairness and welfare may be dramatic, there is an upper bound on it.

To intuitively understand why losses in welfare and fairness are bounded, we reinterpret the definition of stability. Consider a continuous utility u and a matching $\pi \in \mathcal{M}_+(X \times Y)$ with marginals μ and ν. By Lemma 2, matching π is ε-stable with $\varepsilon \geq 0$ if for all (x_1, y_1) and (x_2, y_2) in the support of π, inequality (2) holds, i.e., $u(x_1, y_2) \leq \max\{u(x_1, y_1),\, u(x_2, y_2)\} + \varepsilon$. The couple (x_1, y_2) can serve as a generic element of the product space $X \times Y$. This observation allows us to interpret formula (2) as follows:

> π is ε-stable if, for a generic couple (x, y)—not necessarily matched under π—the utility of at least one of the partners x or y in their respective match under π is at least the utility of the hypothetical match (x, y) minus ε.

Since the utility of a hypothetical couple (x, y) provides a lower bound to the utility of at least one of the partners in a stable match, the latter utilities cannot be too small simultaneously. This observation bounds the extent to which welfare and fairness are to be sacrificed in order to get stability.

Recall that the utilitarian welfare $W(\pi)$ of a matching π is given by (6) and let $W^*(\mu, \nu)$ be the maximal welfare over all $\pi \in \Pi(\mu, \nu)$. We prove the following result in Appendix A.

Theorem 2. *For a market with unit populations ($\mu(X) = \nu(Y) = 1$) and bounded* *continuous utility $u \geq 0$, any ε-stable matching π satisfies*[7]

$$W(\pi) \geq \frac{1}{2}\left(W^*(\mu,\nu) - \varepsilon\right).$$

Moreover, π is ε'-egalitarian with $\varepsilon' = \max\left\{\frac{1}{2}, \varepsilon\right\}$.

In particular, we obtain that any stable matching guarantees 1/2 of the optimal welfare and is 1/2-egalitarian. For markets with a finite number of agents, we recover the welfare guarantee obtained by Anshelevich, Das, and Naamad (2013). The bounds in Theorem 2 are conservative as they target ε-stable matchings that have the lowest welfare or that are least egalitarian.

3.3 Markets with non-aligned preferences

Our study of stability and fairness assumes aligned preferences. Here, we argue that, in some circumstances, preference alignment is a good approximation to the study of non-aligned markets.

First, we make a very simple point. When agents' preferences are not aligned but close to being aligned, then our results on ε-stability continue to hold, as long as the approximation contemplates the deviation from alignment. Specifically, if each x and y's utility from matching is within $\varepsilon > 0$ of an aligned utility $u(x, y)$, then any matching that is ε-stable for the aligned market with utility u is automatically 3ε stable in the non-aligned market.[8] So an an approximately stable matching remains approximately stable for nearby non-aligned markets.

Second, and more importantly, we show that in large finite markets, even dramatically misaligned preferences can be well approximated by aligned preferences, so long as the misalignment is idiosyncratic. Our result is non-asymptotic and provides explicit bounds on the quality of the approximation as a function of market size and other parameters.

Consider two finite populations of n agents $X_n = \{x_1, \ldots, x_n\} \subseteq X$ and $Y_n \in \{y_1, \ldots, y_n\} \subseteq Y$, with $X, Y \subset \mathbb{R}$ being two compact intervals. We assume that X_n

[7]If a bounded continuous utility u does not satisfy the non-negativity requirement, one can apply Theorem 2 to $\tilde{u}(x, y) = u(x, y) - \inf_{x',y'} u(x', y')$. For example, this argument applies to the distance-based utility of Section 2 if μ and ν have bounded support.

[8]Let the utility of $x \in X$ be $v_x : Y \to \mathbb{R}$ and $|v_x(y) - u(x, y)| \leq \varepsilon$. Then if x is matched with y and considers a potential blocking pair with y', we have $v_x(y') - v_x(y) \leq u(x, y) - u(x, y') + 2\varepsilon \leq 3\varepsilon$ when the matching is ε-stable in the aligned market.

and Y_n are i.i.d. samples from two full-support, non-atomic, distributions $\mu \in \Delta(X)$ and $\nu \in \Delta(Y)$. If a pair $(x_i, y_j) \in X_n \times Y_n$ is formed, agents i and j enjoy utilities

$$u_{i,j} = w(x_i, y_j) + \xi_{i,j} \qquad \text{and} \qquad v_{i,j} = w(x_i, y_j) + \eta_{i,j}.$$

Here, $w\colon X \times Y \to \mathbb{R}$ is a continuous function capturing the aligned component of agents' preferences. The idiosyncratic components $\xi_{i,j}$ and $\eta_{i,j}$ are given by i.i.d. shocks with continuous distribution functions F and G, independent across the two sides of the market. Note that the idiosyncratic components can be very large relative to the aligned component w. It is possible that the idiosyncratic components are even the main drivers of agents' preferences.

Theorem 3. *Let $\pi \in \Pi(\mu, \nu)$. With probability at least $1 - \alpha_n$, there exists a deterministic matching π_n of X_n and Y_n such that*

$$\left| \frac{\left|\left\{(x_i, y_j) \in [a,b] \times [c,d] \,:\, x_i \text{ and } y_j \text{ are matched in } \pi_n\right\}\right|}{n} - \pi\big([a,b] \times [c,d]\big) \right| \le \beta_n$$

for all intervals $[a,b] \subseteq X$, $[c,d] \subseteq Y$. Moreover, for all x_i and y_j matched under π_n,

$$F\big((-\infty, \xi_{i,j}]\big) \;\ge\; 1 - \gamma_n \quad \text{and} \quad G((-\infty, \eta_{i,j}])\;\ge\; 1 - \gamma_n.$$

Here α_n, β_n, and γ_n are constants that depend on n only and converge to 0, as $n \to \infty$.

Theorem 3 establishes that, under the given conditions, the finite non-aligned market can be approximated by a large aligned one. Specifically, for any matching π in the large market, there exists a finite-market matching π_n that approximates π while ensuring that individual agents receive close to their "maximum possible" idiosyncratic utilities, i.e., they are matched with a partner representing a large quantile in their idiosyncratic distribution.

The theorem does not assume that π is stable or egalitarian; indeed, it can be used to approximate a matching that optimizes any welfarist objective. When π is stable, π_n inherits approximate stability. Indeed, all agents in π_n are matched in a way that guarantees them a very high idiosyncratic utility. Consequently, any blocking pair in π_n would need to rely on exploiting the aligned component of preferences, implying the existence of a corresponding blocking pair in the original stable matching π.

In Appendix A, we present a stronger result, Theorem 8, which implies Theorem 3. Theorem 8 provides explicit formulas for α_n, β_n, and γ_n and does not rely on i.i.d.

sampling of the population. So it applies to any—possibly deterministic—set of points whose empirical distributions are close to μ and ν. Our inspiration for Theorem 3 came from Lee (2016), but we use different techniques and obtain a result that is independent of the limiting matching being stable. To prove Theorem 8, we use the probability method to obtain deterministic approximation results, leverage measure-concentration inequalities (Dvoretzky-Kiefer-Wolfowitz), and adapt the Erdos-Renyi classical result on the existence of perfect matching in random bipartite graphs.

4 Applications

4.1 School choice

In the standard model of school choice (Abdulkadiroğlu and Sönmez, 2003), students have preferences over schools, and schools have preferences over students. The latter are called priorities. Preferences and priorities may not be aligned, but we shall argue here that the aligned model is a reasonable and useful approximation to the problem.

The aligned model is a reasonable approximation because distance is a key component of student preferences over schools and of school priorities over students. First, Walters (2018) estimates student preferences over charter schools in Boston, and finds that distance traveled is a key driver of student preferences. Other papers find similar results (Laverde, 2022; Dinerstein and Smith, 2021; Agostinelli, Luflade, Martellini, et al., 2021). Second, school districts typically give priority to local, neighborhood, students. For example, in Boston, priority is given to students with siblings who attend the same school, and to students who live within a mile of the school (Angrist, Gray-Lobe, Idoux, and Pathak, 2022). As a first rough approximation, one can assume that student i and school s each get utility $u(i,s) = -d(i,s)$ from matching together where $d(i,s)$ is the physical distance between i and s.

The aligned model remains a good approximation when preferences and priorities have a "vertical" (quality) component. Suppose that students care about school quality, and that schools prioritize high-achieving students. Let q_s measure school s's quality and let q_i be a measure of student i's academic achievement. Letting $u(i,s) = -d(i,s) + f(q_s) + g(q_i)$ can capture such preferences. Note that

$$\begin{aligned} u(i,s) - u(i,t) &= d(i,t) - d(i,s) + f(q_s) - f(q_t) \text{ and} \\ u(i,s) - u(j,s) &= d(j,s) - d(i,s) + g(q_i) - g(q_j), \end{aligned}$$

so u accurately reflects the relative preferences of both sides. Using similar ideas, one can incorporate, for instance, priorities for siblings and diversity-related preferences. Of course, not all configurations of preferences and priorities are aligned. In section 5, we discuss ordinal conditions that guarantee aligned preferences.

The aligned model is useful. It provides some new insights for school choice. By recasting stable matching as the solution to a planner's optimization problem, we can compare the objective *implicit* in implementing stability with other, more traditional, measures of welfare. Theorem 1 shows that stability comes from maximizing an objective that places very high weight on matching students to nearby schools, but which places vanishing value on the externalities that these matches can cause on students who will have to travel farther as a result of stabilty. The stability objective is at the opposite extreme of some commonly used welfare measures in other areas of economics. For example, in public finance, it is common to model the planner as maximizing the average of a concave function—equivalently, minimizing a convex function—of utilities (Saez, 2001). The concave welfare measure reflects a planner's preference for fairness. When α is large, however, (9) corresponds to a planner who loves inequality.

Our discussion has two implications for school choice. First, stable matching can generate long average travel times; though Theorem 2 suggests a limit to the severity of this phenomenon. Second, stable matching can entail very long travel times for some students, and short travel times for others, even when these students have very similar characteristics. A common motivation for using stable mechanisms is that they produce fair outcomes in the sense of "no justified envy." No student will envy a student's placement at a school for which they have high priority. Theorem 1 suggests that minimizing unfairness, in the sense of eliminating instances of justified envy, can result in a different type of unfairness: a dispersion of outcomes across students.

Consistent with our findings, Angrist, Gray-Lobe, Idoux, and Pathak (2022) document that, since adopting the deferred acceptance algorithm, both Boston and New York have seen substantial increases in municipal expenditures on transportation, and a significant fraction of students travel long distances to school. For example, in the first decade after implementing the deferred acceptance algorithm, Boston saw more than a 50% increase in per-pupil expenditures on travel. They estimate that switching to neighborhood assignment would decrease average travel times by as much as 17 minutes in New York and 13 minutes in Boston. By contrast, Angrist, Gray-Lobe, Idoux, and Pathak (2022) estimate modest effects of centralized assignment on test scores and college attendance.

4.2 Ride-sharing

Ride-sharing platforms like Uber and Lyft operate large two-sided matching markets where thousands of riders are matched to drivers every day. While both riders and drivers, in principal, care about specific details of their match partner, the most important and salient feature is their physical distance $d(x, y)$ in space. Drivers and riders both prefer to be matched with partners close to them to avoid wasteful travel time. Initially, Uber's matching algorithm was greedy: they simply matched each rider with the closest available driver. The resulting matching is stable when $u(x, y) = -d(x, y)$. Quickly, however, Uber noticed that this greedy algorithm had some unappealing features. The following is a quote from Uber's website describing the evolution of their matching algorithm.

> In the early days, a rider was immediately matched with the closest available driver. It worked well for most riders but sometimes led to long wait times for others. Across a whole city, those longer wait times really added up [...] In the seconds after a rider requests a ride, we evaluate nearby drivers and riders in one batch. We then pair riders and drivers in the distribution, aiming to reduce the average wait time for everyone, not just the closest pair. This helps keep things moving and rides reliable across the network (Uber.com, 2024).

There are a few points worth emphasizing. First, Uber claims that the greedy match led to long wait times for some riders, while it worked well for most. This is consistent with the message that stability and fairness are in conflict (see the examples in Section 2). Second, Uber noticed that, not only did the greedy matching result in inequality, but the longer wait times "added up." This suggests that the greedy algorithm produced overall longer total waiting times, which is consistent with the message in Theorem 1. The objective implicitly being optimized in a stable matching does not minimize total wait times. Theorem 2 implies a bound on how poorly the stable matching can perform in terms of total wait time.

Uber now matches riders in batches. This is consistent with the growing literature on dynamic matching markets, which emphasizes the importance of market thickness.[9]

[9] See for example, Akbarpour, Li, and Gharan (2020); Baccara, Lee, and Yariv (2020); Doval (2022).

Within batches, the algorithm aims to reduce average weight times. That is, they do not use a stable matching, even within batches.

4.3 Markets with transfers and no commitment

We argue that aligned preferences are a natural assumption in models of matching with transfers and lack of commitment ability. Namely, we show that alignment arises when agents cannot commit to utility transfers at the stage where they bargain over matches, and the post-match surplus is shared according to Nash bargaining.

Consider the matching market with transferable utility introduced by Shapley and Shubik (1971) and used, for example, in marriage-market models of Becker (1973) and Galichon and Salanié (2022). If a couple (x, y) forms a match, then they generate a surplus $s(x, y)$ that they may share: x getting $\hat{u}(x, y)$ and y getting $\hat{v}(x, y)$, with $\hat{u}(x, y) + \hat{v}(x, y) = s(x, y)$. The model with transferable utilities assumes that the shares $\hat{u}(x, y)$ and $\hat{v}(x, y)$ are determined at the same time as the match. Effectively, the model assumes that transfers between couples are negotiated, agreed upon, and committed to, as part of the bargaining over who matches with whom.

We consider instead the possibility that agents cannot commit to a specific share of the surplus when they agree to form a pair with another agent. Specifically, we assume that, once a match has been formed, it cannot be broken without significant cost, and that the two members of a couple bargain over how to share the surplus according to the Nash bargaining model with fixed weights, β and $1 - \beta$, for the X and Y sides, respectively. This means that if a couple (x, y) forms, then $\hat{u}(x, y) = \beta \cdot s(x, y)$ while $\hat{v}(x, y) = (1 - \beta) \cdot s(x, y)$.

Thus a matching π is stable in the model with Nash bargaining surplus-sharing if and only if it is stable in a market with aligned preferences represented by $s(x, y)$. The connection extends to approximate stability: π is ε-stable in the model with Nash bargaining if it is ε'-stable in a market with aligned preferences and $\varepsilon' = \frac{\varepsilon}{\max\{\beta,\ 1-\beta\}}$.

5 Ordinal conditions for aligned preferences

We provide ordinal foundations for aligned preferences. Together with technical topological conditions, a certain acyclicity condition reminiscent of Monderer and Shapley (1996) suffices to guarantee the existence of aligned utility functions. In the absence of a monetary scale pinning down a particular utility representation of preferences, it becomes especially important to identify requirements on agents' preferences that enable the existence of a utility representation common to both sides of the market.

A binary relation is called a preference if it is complete and transitive. We denote preferences by $\succeq$. The two sides of the market are given by sets X and Y. For each $x \in X$ we have a preference $\succeq_x$ over Y. Likewise, for each $y \in Y$ we have a preference $\succeq_y$ over X. We collect these data into a tuple $(X, Y, \succeq_X, \succeq_Y)$ where $\succeq_X = (\succeq_x)_{x \in X}$ and $\succeq_Y = (\succeq_y)_{y \in Y}$.

We call a function $u : X \times Y \to \mathbb{R}$ a **potential** for $(X, Y, \succeq_X, \succeq_Y)$ if

$$\begin{aligned} u(x,y) \geq u(x,y') &\iff y \succeq_x y' \\ u(x,y) \geq u(x',y) &\iff x \succeq_y x' \end{aligned} \qquad \text{for all} \quad x, x', y, y'.$$

In other words, the potential is a single utility function representing the preferences of each side of the market. The key condition for the existence of a potential turns out to be a form of acyclicity. A sequence of couples, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, with $n > 2$ forms a cycle if $(x_n, y_n) = (x_1, y_1)$ and each couple (x_{i+1}, y_{i+1}) has exactly one agent in common with the preceding couple (x_i, y_i).

The tuple $(X, Y, \succeq_X, \succeq_Y)$ is **acyclic** if, for any cycle $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where each common agent prefers their partner in (x_{i+1}, y_{i+1}) to their partner in (x_i, y_i), all common agents are, in fact, indifferent between the two partners.

For finite markets, acyclicity is necessary and sufficient for a potential to exist. This result is due to Niederle and Yariv (2009) who established the connection with potential games where Monderer and Shapley (1996) and Voorneveld and Norde (1997) used a similar condition. We extend this result to general topological spaces. This generalization requires additional technical requirements, similar to those necessary for the existence of a utility function on a general topological space.

For a preference $\succeq$ on a set Z, we denote by $\succ$ the strict part of $\succeq$, i.e., $z \succ z'$ if and only if $z \succeq z'$ but $z' \not\succeq z$. For a topological space Z, a preference $\succeq$ is said to be **continuous** if the upper contour sets $U_{\succeq}(z) = \{z' \in Z : z' \succ z\}$ and lower contour sets $L_{\succeq}(z) = \{z' \in Z : z \succ z'\}$ are open.

In addition to the continuity of each agent's preferences, we consider the following requirements of continuity with respect to changing the agent:

1. If $b \succ_a b'$, then there is a neighborhood N_a of a for which $b \succ_{\tilde{a}} b'$ for all $\tilde{a} \in N_a$.

2. If $b \succ_a b'$, then there are neighborhoods N_a of a, N_b of b, and $N_{b'}$ of b' such that $\tilde{b} \succ_{\tilde{a}} \tilde{b}'$ for all $(\tilde{a}, \tilde{b}, \tilde{b}') \in N_a \times N_b \times N_{b'}$.

3. If $b' \succeq_a b$ and $b \succeq_{a'} b''$, where $b \neq b', b''$ and $a \neq a'$, then, in any neighborhood of b, there exists $\tilde{b}$ such that $b' \succ_a \tilde{b}$ and $\tilde{b} \succ_{a'} b''$.

Property 1 is a continuity assumption relating how one agents' preferences must be similar to the preferences of agents who are close by on the same side of the market. Property 2 is a stronger joint continuity property, and, under some additional assumptions on the spaces X and Y, amounts to saying that $a \mapsto \succeq_a$ is a continuous map in the topology of closed convergence; see Kannai (1970). Property 3 is in the spirit of local strictness by Border and Segal (1994).

Theorem 4. *Let $(X, Y, \succeq_X, \succeq_Y)$ be such that X and Y are complete, separable, and connected topological spaces. Suppose that preferences are continuous, satisfy acyclicity, and properties (1) and (3). Then there is a potential u. If, moreover, Property 2 is satisfied, then u can be taken to be upper semicontinuous.*

The proof of Theorem 4 is in Appendix E. The proof uses a construction borrowed from Debreu (1954, 1964), which works by extending a primitive ordering defined on a countable dense subset to the whole space. In the case of Theorem 4, the ordering admits a potential on a countable dense subset, but the extension is more challenging than in the canonical utility representation problem of Debreu. One has to show the existence of a strict comparison for the utility of each agent, even for those who are not included in the countable dense subset. This necessitates the continuity and local strictness properties listed as the hypotheses of the theorem.

6 Matching in $\mathbb{R}^d$

We consider markets with aligned preferences, where the two sides of the market X and Y are subsets of $\mathbb{R}^d$ with $d \geq 1$, and each agent prefers to be matched to someone as close as possible. This is captured by the distance-based utility function

$$u(x, y) = -\|x - y\| = -\sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}.$$

This setting is a benchmark model in school choice or ride-sharing, where X and Y correspond to physical locations. More generally, it captures matching driven by homophily, e.g., a preference for a partner with similar political views.

The connection between stability, fairness, welfare, and optimal transport established in Theorem 1 is not sensitive to the dimension d. However, the case $d = 1$ (Section 2) admits further structure.

We start with $d = 1$. The cost function (5) from the theorem specializes to

$$c_\alpha(x, y) = \frac{1 - \exp(-\alpha \cdot |x - y|)}{\alpha}. \tag{8}$$

which is a strictly convex function of $|x - y|$ for $\alpha < 0$ and strictly concave for $\alpha > 0$. Transportation problems on the line with costs given by concave and convex functions of distance are well understood.

For a cost function $c(x, y) = h(|x - y|)$ with strictly convex h, the solution to (3) is the assortative matching, which is uniquely optimal and does not depend on a particular form of h. Recall that the assortative matching is supported on the curve $F_\mu(x) = F_\nu(y)$, where F_μ and F_ν are the cumulative distribution functions of type distributions μ on X and ν on Y, respectively.

Thus, for any $\alpha < 0$, the assortative matching is the unique solution to the optimal transportation problem. Taking weak limits, and noting that as $\alpha \to 0$, (3) approaches the utilitarian objective, Theorem 1 implies that the assortative matching is fair and welfare-maximizing.

Corollary 2. *For $X = Y = \mathbb{R}$ and non-atomic μ, ν with bounded support, the assortative matching is egalitarian and is welfare-maximizing.*

We now consider $\alpha > 0$. In Section 2, stable matchings—corresponding to $\alpha \to +\infty$—satisfy a no-crossing property. McCann (1999) (Theorem 3.11) showed that for any cost $c(x, y) = h(|x - y|)$ with strictly concave h and non-atomic μ and ν on $\mathbb{R}$, optimal matchings satisfy no-crossing: first, the mass common to μ and ν is eliminated (these agents are matched with their ideal partners $x = y$) and then the remaining disjoint populations are matched in no-crossing fashion. Moreover, the set of all no-crossing matchings consists of a number of parametric families. For example, no-crossing matchings from Figure 4 form a single family parameterized by $\theta \in [0, 1]$.

The parametric structure of no-crossing matchings reduces the infinite-dimensional transportation problem with cost $c_\alpha(x, y)$ for $\alpha > 0$ to determining a finite number of

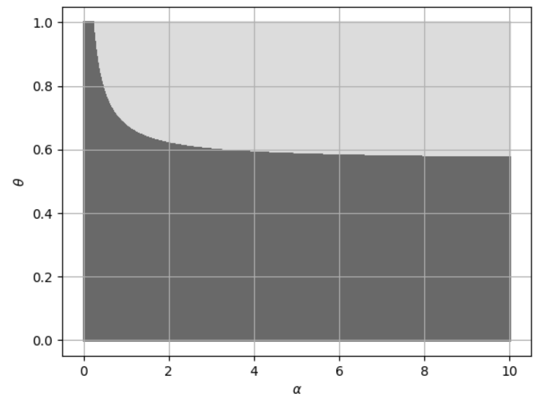


Figure 5: Dependence of θ on α in the optimal transportation problem with cost c_α and distribution from Figure 4.

parameters. Unlike the case of $\alpha < 0$, where the optimum is independent of α, for $\alpha > 0$ the optimum generally depends on α since the set of no-crossing matchings is not a singleton. For example, Figure 5 illustrates the dependence of the optimal θ on α for the market from Figure 4; the stable matching corresponds to $\theta = 4/7$, and welfare is maximized by $\theta = 1$.

The change in the optimal no-crossing matching as α ranges between 0 and $+\infty$ reflects the tension between welfare and stability. Indeed, for the market from Figure 4, the stable matching has a welfare of approximately -6.02, while the optimal welfare is -2.

The stability-welfare tension disappears if there are only a few regions of supply-demand imbalance. For example, if μ is uniform on $[-1, 0]$ and ν is uniform on $[0, 1]$, the stable matching (Figure 1a) and the egalitarian one (Figure 1b) have the same welfare. Since the egalitarian matching is welfare-maximizing (Corollary 2), so is the stable one. Thus, there is inequality but no welfare loss.

More generally, the stability-welfare tension is absent if $\mu - \nu$ changes sign at most twice (in the example, it changes sign once at 0). Indeed, McCann (1999) showed that for such μ and ν, a no-crossing matching is unique as in Figure 6). Thus, the same no-crossing matching solves the transportation problem with the cost c_α from (8) for any $\alpha \in (0, +\infty)$. Letting $\alpha \to +\infty$, we conclude that this matching is stable. Letting $\alpha = 0$, we get that it is also welfare-maximizing.

Problem complexity increases with the number of times $\mu - \nu$ changes sign. For at

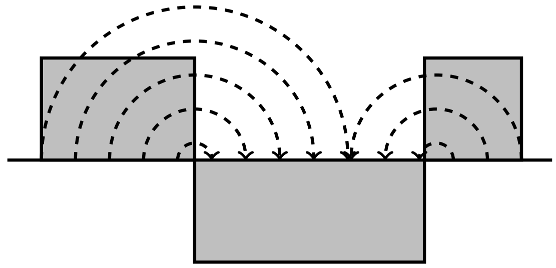


Figure 6: For $\mu - \nu$ changing sign two times, no-crossing matching is unique and thus is simultaneously stable and welfare-maximizing.

most two changes, a no-crossing matching is unique. For three changes, as in Figure 5, we get a single one-parameter family. More generally, we get multiple parametric families whose number grows exponentially with the number of sign changes.[10]

A designer pursuing stability can avoid the exploding search space by combining insights about no-crossing matchings with properties specific to stability. In Supplementary Appendix C, we describe an efficient algorithm that constructs a stable matching by sequentially identifying sub-markets similar to Figure 6, which are to be matched internally regardless of the rest of the population. If μ and ν have piecewise-constant density with at most m intervals of constancy, the algorithm runs in time $O(m^2)$.

Corollary 3. *For $X = Y = \mathbb{R}$ and non-atomic μ, ν, there is a unique stable matching, which can be computed efficiently. For each $x \in X$, there are at most two distinct types $y, y' \in Y$ such that (x, y) and (x, y') are in the support. The welfare-stability tension—while present in general—disappears if $\mu - \nu$ has at most two sign changes.*

Corollaries 2 and 3 imply that for $\mu - \nu$ with at most two sign changes, there are egalitarian and stable matchings with optimal welfare. However, these two matchings are never the same. More generally, for distinct non-atomic μ, ν egalitarian and stable matchings cannot coincide since the former is increasing, while the latter contains decreasing regions (as any no-crossing matching).

Corollary 4. *For $X = Y = \mathbb{R}$ and distinct non-atomic μ, ν, there is always a tension between stability and fairness.*

[10] No-crossing matchings are parameterized by partitioning the real line into intervals where $\mu - \nu \geq 0$ or $\mu - \nu \leq 0$. For each interval of positivity, we specify the fraction of agents matched non-locally to the right (θ in Figure 4) and the negativity intervals these matches come from. Thus, for each interval of positivity, we get a parameter and combinatorial data. The combinatorial data grows exponentially with the number of intervals.

We now consider the multidimensional setting. For $\mathbb{R}^d$ with $d > 1$, there may be a tension between individual and collective welfare, absent in the one-dimensional case.

Example 4 (Fairness-welfare tension). Let $X = Y = \mathbb{R}^2$. Let $B_\delta(x, y)$ denote the ball of radius δ centered at the point (x, y). Let μ have density 1 on $B_\delta(0, 0) \cup B_\delta(0, 1)$ and ν have density 1 on $B_\delta(0, 0) \cup B_\delta(1, 0)$. In the welfare maximizing matching, the agents X and Y in $B_\delta(0, 0)$ are matched together and the agents in X from $B_\delta(0, 1)$ are matched with those from $B_\delta(1, 0)$ in Y. This gives an average match distance of $\frac{1}{2}\sqrt{2}$ and the longest distance between matched agents is no less than $\sqrt{2}$. By contrast, the matching which pairs the X agents in $B_\delta(0, 1)$ with the Y agents in $B_\delta(0, 0)$ and the X agents in $B_\delta(0, 0)$ with the Y agents in $B_\delta(0, 1)$ gives an average match distance of 1 but has the longest distance between matched pairs no more than $1 + 2\delta$.

Concave and convex optimal transport are not prone to explicit solutions beyond $d = 1$, but some structural results can be obtained for $d > 1$. Now the cost function is

$$c_\alpha(x, y) = \frac{1 - \exp(-\alpha \cdot \|x - y\|)}{\alpha}. \tag{9}$$

Say that a matching π is deterministic if it is supported on the graph of an invertible map $s \colon X \to Y$. A matching π is diagonal if $x = y$ for π-almost all couples (x, y). Off-the-shelf results from optimal transport (Theorems 1.2 and 1.4 of Gangbo and McCann (1996)) address multidimensional optimal transport with concave or convex costs. Their results imply the following corollary.

Corollary 5. *Consider a market with $X = Y = \mathbb{R}^d$, compactly supported μ and ν absolutely continuous with respect to the Lebesgue measure, and utility $u(x, y) = -\|x - y\|$. The optimal transportation problem with cost (9) admits an optimal matching π^* and the following assertions hold:*

- *For $\alpha < 0$, π^* is unique and deterministic;*
- *For $\alpha > 0$, π^* is unique and is a convex combination of a deterministic matching and a diagonal one.*
- *There is a deterministic matching π^* which maximizes welfare.*

We note that the condition of absolute continuity can be weakened to requiring that μ and ν place zero mass on $(d - 1)$-dimensional surfaces. By imposing this on μ only, we obtain a version of Corollary 5 where the conclusion that π^* is deterministic is replaced with the Monge property (π^* is supported on a graph of $s\colon X \to Y$, but s may not be invertible).

The egalitarian matching obtained as the limit $\alpha \to -\infty$ corresponds to the L^∞-transportation problem Champion, De Pascale, and Juutinen (2008); Brizzi, De Pascale, and Kausamo (2023). The transportation literature has mostly focused on questions of existence, uniqueness, and cyclic monotonicity. For a distance-based cost, the L^∞-problem is interesting only for $d > 1$ since, on the real line, it corresponds to the assortative matching by Corollary 2. To the best of our knowledge, the limit problem for $\alpha \to +\infty$ has not been studied prior to our paper.

7 Multi-partner matching

Markets with more than two sides, referred to as multi-partner matching markets, are known to be challenging. Stable matchings may not exist, even in simple finite markets. It is therefore notable that our main results extend to this setting, opening up the theory to additional applications. Formally, we consider the formation of stable k-tuples. Given are k sets of types of agents, $X_1, \ldots, X_k$, and functions $u_i\colon X_1 \times \ldots \times X_k \to \mathbb{R}$, $i = 1, \ldots, k$. The interpretation is that each tuple $(x_1, \ldots, x_k)$ generates a utility $u_i(x_1, \ldots, x_k)$ to agent x_i.

First, we revisit the standard theory. Suppose that all X_i are finite and have the same cardinality. A matching is a collection μ of k-tuples $(x_1, \ldots, x_k)$ such that each x_i appears in exactly one tuple. We write $\mu(x_i)$ for the tuple that x_i belongs to. A k-tuple $(x_1, \ldots, x_k)$ blocks a matching μ if $u_i(x_1, \ldots, x_k) > u_i(\mu(x_i))$ for all $i = 1, \ldots, k$. A matching is stable if there is no blocking k-tuple.

The multi-partner matching problem is well known to be intractable. Even with three partners ($k = 3$) and additively separable preferences $u_i(x) = \sum_{j \neq i} v_i(x_i, x_j)$, there may not exist a stable matching (Alkan, 1988). Moreover, deciding whether a stable matching exists is NP-hard (Ng and Hirschberg, 1991). The case of aligned preferences—utilities $u_i = u$ for all agents i—turns out to be an exception: a stable matching is guaranteed to exist, and the theory we have developed for two-sided markets readily extends to multi-partner matching.

We turn to a generalization of the model introduced in Section 3. The sets X_i are assumed to be complete separable metric spaces, each X_i is endowed with a measure $\mu_i \in \mathcal{M}_+(X_i)$ representing the distribution of agents' types in X_i, and the total mass $\mu_i(X_i)$

is the same for all i. A **matching** π is a positive measure on $X = X_1 \times \ldots \times X_k$ with marginals μ_i on X_i; $\Pi(\mu_1, \ldots, \mu_k)$ denotes the set of all matchings. A measurable function $u \colon X \to \mathbb{R}$ is the agents' common utility. Each agent in a k-tuple $x = (x_1, \ldots, x_k)$ enjoys utility of $u(x)$.

For a matching π, let $\pi^{\times k}$ denote the product measure on X^k. So $\pi^{\times k}$ corresponds to k independent draws of a tuple from X. A matching π is ε-**stable** if, for $\pi^{\times k}$-almost all collections of tuples $x^1, \ldots, x^k$,

$$u(x_1^1, \ldots, x_k^k) \leq \max_{i=1,\ldots,k} u(x^i) + \varepsilon.$$

The notions of an ε-egalitarian matching and that of the optimal welfare W^* also extend straightforwardly to the multi-partner case.

Given a cost function $c \colon X \to \mathbb{R}$, the k-marginal optimal transportation problem is to find a matching π minimizing the cost

$$\min_{\pi \in \Pi(\mu_1,\ldots,\mu_n)} \int_X c(x) \, \mathrm{d}\pi(x). \tag{10}$$

As in the two-sided case, we consider the following cost function

$$c_\alpha(x) = \frac{1 - \exp(\alpha \cdot u(x))}{\alpha}. \tag{11}$$

Theorems 1 and 2 generalize to the k-partner case as follows.

Theorem 5. *Consider a k-partner matching market with aligned preferences and bounded continuous utility $u \colon X_1 \times \ldots \times X_k \to \mathbb{R}_+$. Let π^* be a solution to the optimal transportation problem (10) with the cost c_α from (11). The following assertions hold:*

1. *If $\alpha > 0$, then π^* is ε-stable with $\varepsilon = \frac{\ln k}{\alpha}$.*
2. *If $\alpha < 0$, then π^* is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.*
3. *For unit populations ($\mu_i(X_i) = 1$ for all $i = 1, \ldots, k$), any ε-stable matching π satisfies $W(\pi) \geq \frac{1}{k}\left(W^* - \varepsilon\right)$. Moreover, π is ε'-egalitarian with $\varepsilon' = \max\left\{\frac{1}{k}, \varepsilon\right\}$.*

The theorem is proved in Appendix D. As we see, the factor 2 in Theorems 1 and 2 is equal to the number of sides of the market. So 2 gets replaced with k in Theorem 5.

Theorem 5 can be applied to team, club, or coalition-formation models. The following example outlines a less immediate application to organ exchanges under compatibility constraints (Roth, Sönmez, and Ünver, 2005, 2007). The example is highly stylized and only meant as an illustration of what can be done with k-sided markets.

Example 5 (Organ exchanges). An agent is a patient-donor pair, where the donor is willing to donate an organ to the patient, but the donation is not feasible because the two are biologically incompatible. For practical reasons—see (Roth, Sönmez, and Ünver, 2005, 2007)—we rule out complicated organ trades and focus attention on pairwise exchanges. At first glance, the organ exchange problem appears to be a one-sided "roommate" problem, where any pair can be matched with any other pair. However, we can capture the compatibility constraints by recasting the model as a multi-sided matching problem.

Suppose that there are k types of agents. These could, for example, correspond to the blood types of the patient and the donor: (O,A), (A,B), and so on. Depending on the organ, other biological markers may be used in defining the type of an agent. The total number of types is k. In addition, assume that agents differ in the size $(s_1, s_2) \in \mathbb{R}^2$ of the organs of the patient and donor, respectively. Let μ_i represent the distribution of agents of type i in $\mathbb{R}^2$. Let $G = (\{1, \ldots, k\}, E)$ be a compatibility graph, so that $(i, j) \in E$ whenever the type of the donor from i is compatible with the type of the patient from j, and vice versa. Now for any compatible pair of types (i, j) and for any two vectors of sizes $s^i = (s^i_1, s^i_2)$ and $s^j = (s^j_1, s^j_2)$, assume that the match quality between the pairs can be captured by a function $q_{i,j}(s^i, s^j)$.

To make this into a k-sided matching market, we augment each type by adding a "dummy" agent d_i, and add a sufficient quantity of copies of the dummy agent so that $\mu_i(\mathbb{R}^2)$ is the same for all types i.

Now define a utility function $u : \bigtimes_{i=1}^{k}(\mathbb{R}^2 \cup \{d_i\}) \to \mathbb{R}$ by

$$u(x_1, \ldots, x_k) = \begin{cases} q_{i,j}(x_i, x_j), & \text{if } (i,j) \in E \text{, and } x_i, x_j \text{ are the only non-dummy types} \\ 0, & \text{if there is at most one non-dummy type} \\ -M, & \text{otherwise} \end{cases},$$

where $M > 0$ is a large number. This now forms a multi-sided matching market as considered above, and the theory developed can be used to find stable and egalitarian matchings.

References

Abdulkadiroğlu, A., and T. Sönmez (2003): "School choice: A mechanism design approach," *American economic review*, 93(3), 729–747.

Abraham, D. J., A. Levavi, D. F. Manlove, and G. O'Malley (2008): "The stable roommates problem with globally ranked pairs," *Internet Mathematics*, 5(4), 493–515.

AGARWAL, N. (2015): "An empirical model of the medical match," *American Economic Review*, 105(7), 1939–1978.

AGOSTINELLI, F., M. LUFLADE, P. MARTELLINI, ET AL. (2021): "On the spatial determinants of educational access," Discussion paper, University of Pennsylvania.

AKBARPOUR, M., S. LI, AND S. O. GHARAN (2020): "Thickness and information in dynamic matching markets," *Journal of Political Economy*, 128(3), 783–815.

ALKAN, A. (1988): "Nonexistence of stable threesome matchings," *Mathematical social sciences*, 16(2), 207–209.

ANGRIST, J., G. GRAY-LOBE, C. M. IDOUX, AND P. A. PATHAK (2022): "Still Worth the Trip? School Busing Effects in Boston and New York," Discussion paper, National Bureau of Economic Research.

ANSHELEVICH, E., S. DAS, AND Y. NAAMAD (2013): "Anarchy, stability, and utopia: creating better matchings," *Autonomous Agents and Multi-Agent Systems*, 26(1), 120–140.

ANTRAS, P., T. C. FORT, AND F. TINTELNOT (2017): "The margins of global sourcing: Theory and evidence from US firms," *American Economic Review*, 107(9), 2514–2564.

ARIELI, I., Y. BABICHENKO, AND F. SANDOMIRSKIY (2023): "Persuasion as transportation," *arXiv:2307.07672*.

ARNOSTI, N. (2022): "A continuum model of stable matching with finite capacities," *arXiv:2205.12881*.

ASHLAGI, I., AND P. SHI (2016): "Optimal allocation without money: An engineering approach," *Management Science*, 62(4), 1078–1097.

ATKINSON, A. B. (1970): "On the measurement of inequality," *Journal of Economic Theory*, 2(3), 244–263.

AZEVEDO, E. M., AND J. D. LESHNO (2016): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 124(5), 1235–1268.

BACCARA, M., S. LEE, AND L. YARIV (2020): "Optimal dynamic matching," *Theoretical Economics*, 15(3), 1221–1278.

BANERJEE, S., H. KONISHI, AND T. SÖNMEZ (2001): "Core in a simple coalition formation game," *Social Choice and Welfare*, 18, 135–153.

BECKER, G. S. (1973): "A theory of marriage: Part I," *Journal of Political economy*, 81(4), 813–846.

BEIGLBÖCK, M., M. GOLDSTERN, G. MARESCH, AND W. SCHACHERMAYER (2009): "Optimal and better transport plans," *Journal of Functional Analysis*, 256(6), 1907–1927.

BOERMA, J., A. TSYVINSKI, R. WANG, AND Z. ZHANG (2023): "Composite Sorting," *arXiv preprint arXiv:2303.06701*.

BORDER, K. C., AND U. SEGAL (1994): "Dynamic consistency implies approximately expected utility preferences," *Journal of Economic Theory*, 63(2), 170–188.

BRIZZI, C., L. DE PASCALE, AND A. KAUSAMO (2023): "L^∞-optimal transport for a class of strictly quasiconvex cost functions," *Journal of Mathematical Analysis and Applications*, 527(1), 127331.

CANTILLON, E., L. CHEN, AND J. S. PEREYRA (2022): "Respecting priorities versus respecting preferences in school choice: When is there a trade-off?," *arXiv preprint arXiv:2212.02881*.

CARLIER, G. (2012): "Optimal transportation and economic applications," MIT lecture notes.

CHAMPION, T., L. DE PASCALE, AND P. JUUTINEN (2008): "The ∞-Wasserstein Distance: Local Solutions and Existence of Optimal Transport Maps," *SIAM Journal on Mathematical Analysis*, 40(1), 1–20.

CHANEY, T. (2014): "The network structure of international trade," *American Economic Review*, 104(11), 3600–3634.

CLARK, S. (2006): "The uniqueness of stable matchings," *Contributions in Theoretical Economics*, 6(1).

DASKALAKIS, C., A. DECKELBAUM, AND C. TZAMOS (2015): "Strong duality for a multiple-good monopolist," in *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pp. 449–450.

DEBREU, G. (1954): "Representation of a preference ordering by a numerical function," in *Decision processes*, pp. 159–165. Wiley New York.

——— (1964): "Continuity properties of Paretian utility," *International Economic Review*, 5(3), 285–293.

DINERSTEIN, M., AND T. D. SMITH (2021): "Quantifying the supply response of private schools to public policies," *American Economic Review*, 111(10), 3376–3417.

DOVAL, L. (2022): "Dynamically stable matching," *Theoretical Economics*, 17(2), 687–724.

DVORETZKY, A., J. KIEFER, AND J. WOLFOWITZ (1956): "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *The Annals of Mathematical Statistics*, pp. 642–669.

ECHENIQUE, F., S. LEE, M. SHUM, AND M. B. YENMEZ (2013): "The revealed preference theory of stable and extremal stable matchings," *Econometrica*, 81(1), 153–171.

ECHENIQUE, F., AND M. B. YENMEZ (2007): "A solution to matching with preferences over colleagues," *Games and Economic Behavior*, 59(1), 46–71.

EECKHOUT, J. (2000): "On the uniqueness of stable marriage matchings," *Economics Letters*, 69(1), 1–8.

EKELAND, I. (2010): "Notes on optimal transportation," *Economic Theory*, pp. 437–459.

ERDŐS, P., AND A. RÉNYI (1964): "On random matrices," *Publ. Math. Inst. Hungar. Acad. Sci.*, 8, 455–461.

FERDOWSIAN, A., M. NIEDERLE, AND L. YARIV (2023): "Strategic Decentralized Matching: The Effects of Information Frictions," *Working Paper*.

GALE, D., AND L. S. SHAPLEY (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69(1), 9–15.

GALICHON, A. (2018): *Optimal transport methods in economics*. Princeton University Press.

GALICHON, A., O. GHELFI, AND M. HENRY (2023): "Stable and extremely unequal," *Economics Letters*, 226, 111101.

GALICHON, A., AND B. SALANIÉ (2022): "Cupid's invisible hand: Social surplus and identification in matching models," *The Review of Economic Studies*, 89(5), 2600–2629.

GANGBO, W., AND R. J. MCCANN (1996): "The geometry of optimal transportation," *Acta Mathematica*, 177(2), 113–161.

GREINECKER, M., AND C. KAH (2021): "Pairwise stable matching in large economies," *Econometrica*, 89(6), 2929–2974.

GUTIN, G. Z., P. R. NEARY, AND A. YEO (2023): "Unique stable matchings," *Games and Economic Behavior*, 141, 529–547.

KANNAI, Y. (1970): "Continuity properties of the core of a market," *Econometrica*, 88, 791–815.

KAUSAMO, A., L. DE PASCALE, AND K. WYCZESANY (2023): "60 years of cyclic monotonicity: a survey," *arXiv preprint arXiv:2308.07682*.

KESTEN, O., AND M. U. ÜNVER (2015): "A theory of school-choice lotteries," *Theoretical Economics*, 10(2), 543–595.

KIM, Y.-H., AND B. PASS (2014): "A general condition for Monge solutions in the multi-marginal optimal transport problem," *SIAM Journal on Mathematical Analysis*, 46(2), 1538–1550.

Kolesnikov, A. V., F. Sandomirskiy, A. Tsyvinski, and A. P. Zimin (2022): “Beckmann’s approach to multi-item multi-bidder auctions,” *arXiv preprint arXiv:2203.06837*.

Laverde, M. (2022): “Distance to schools and equal access in school choice systems,” *Working Paper*.

Lebedev, D., F. Mathieu, L. Viennot, A.-T. Gai, J. Reynier, and F. De Montgolfier (2007): “On using matching theory to understand P2P network design,” in *INOC 2007, International Network Optimization Conference*.

Lee, S. (2016): “Incentive compatibility of large centralized matching markets,” *The Review of Economic Studies*, 84(1), 444–463.

Lee, S., and L. Yariv (2018): “On the efficiency of stable matchings in large markets,” *Working Paper*.

Leshno, J. D., and I. Lo (2021): “The cutoff structure of top trading cycles in school choice,” *The Review of Economic Studies*, 88(4), 1582–1623.

Malamud, S., A. Cieslak, and A. Schrimpf (2021): “Optimal Transport of Information,” *arXiv preprint arXiv:2102.10909*.

Massart, P. (1990): “The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality,” *The annals of Probability*, pp. 1269–1283.

McCann, R. J. (1999): “Exact solutions to the transportation problem on the line,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1984), 1341–1380.

McCann, R. J., and K. S. Zhang (2023): “A duality and free boundary approach to adverse selection,” *arXiv preprint arXiv:2301.07660*.

Monderer, D., and L. S. Shapley (1996): “Potential games,” *Games and economic behavior*, 14(1), 124–143.

Naaman, M. (2021): “On the tight constant in the multivariate Dvoretzky–Kiefer–Wolfowitz inequality,” *Statistics & Probability Letters*, 173, 109088.

Ng, C., and D. S. Hirschberg (1991): “Three-dimensional stable matching problems,” *SIAM Journal on Discrete Mathematics*, 4(2), 245–252.

Niederle, M., and L. Yariv (2009): “Decentralized matching with aligned preferences,” *National Bureau of Economic Research*.

Panigrahi, P. (2021): “Endogenous spatial production networks: Quantitative implications for trade and productivity,” CESifo Working Paper.

Perez-Richet, E., and V. Skreta (2023): “Fraud-proof non-market allocation mechanisms,” *Working Paper*.

Petrov, F. (2016): “Probability bound for perfect matching,” MathOverflow, URL:https://mathoverflow.net/q/230047 (version: 2016-02-03).

Pycia, M. (2012): “Stability and preference alignment in matching and coalition formation,” *Econometrica*, 80(1), 323–362.

Rawls, J. (1971): “A theory of justice,” *Cambridge (Mass.)*.

Reny, P. J. (2021): “A simple sufficient condition for a unique and student-efficient stable matching in the college admissions problem,” *Economic Theory Bulletin*, 9(1), 7–9.

Roth, A. E., T. Sönmez, and M. U. Ünver (2005): “Pairwise kidney exchange,” *Journal of Economic theory*, 125(2), 151–188.

——— (2007): “Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences,” *American Economic Review*, 97(3), 828–851.

Saez, E. (2001): “Using elasticities to derive optimal income tax rates,” *The Review of Economic Studies*, 68(1), 205–229.

Shapley, L. S., and M. Shubik (1971): "The assignment game I: The core," *International Journal of game theory*, 1, 111–130.

Sørensen, M. (2007): "How smart is smart money? A two-sided matching model of venture capital," *The Journal of Finance*, 62(6), 2725–2762.

Uber.com (2024): "How does Uber match riders with drivers?," https://www.uber.com/us/en/marketplace/matching/, Accessed: 2024-07-19.

Villani, C. (2009): *Optimal transport: old and new*, vol. 338. Springer.

Voorneveld, M., and H. Norde (1997): "A characterization of ordinal potential games," *Games and Economic Behavior*, 19(2), 235–242.

Walters, C. R. (2018): "The demand for effective charter schools," *Journal of Political Economy*, 126(6), 2179–2223.

A Proofs

We prove two general results applicable to

$$c(x, y) = -h\left(u(x, y)\right), \tag{12}$$

and deduce Theorem 1 as a corollary. The first result addresses the case of positive α.

Theorem 6. *For a market with aligned preferences, assume that the utility function u takes values in an open interval $I \subset \mathbb{R}$, possibly infinite. Let $h\colon I \to (0, +\infty)$ be a differentiable function and assume that there exists $\alpha > 0$ such that*

$$\frac{h'(t)}{h(t)} \geq \alpha \quad \textit{for any} \quad t \in I. \tag{13}$$

Then any solution to the transportation problem with cost (12) is ε-stable with

$$\varepsilon = \frac{\ln 2}{\alpha}.$$

An example of h satisfying the requirements of Theorem 6 is $h(t) = \exp(\alpha \cdot t)$.

We use the following result by Beiglböck, Goldstern, Maresch, and Schachermayer (2009). Consider an optimal transportation problem

$$\min_{\pi} \int_{X \times Y} c(x, y)\ \mathrm{d}\pi(x, y).$$

with Polish spaces X and Y, measurable cost c. Assume that the value of this transportation problem is finite and is attained at some π^*. Beiglböck, Goldstern, Maresch, and Schachermayer (2009) establish the existence of a c-cyclic monotone Γ

such that $\pi^*(\Gamma) = 1$, i.e., the optimal transportation plan is supported on a cyclically-monotone set. Recall that c-cyclic monotonicity of Γ means that

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1})$$

for all $n \geq 2$ and pairs of points $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ with the convention $y_{n+1} = y_1$.

Proof of Theorem 6. Consider now the optimal transportation problem (3) with c from (12) let π^* be its solution. By the result of Beiglböck, Goldstern, Maresch, and Schachermayer (2009), π^* is supported on some c-monotone set Γ. The requirement of cyclic monotonicity with $n = 2$ implies that for any $(x_1, y_1), (x_2, y_2) \in \Gamma$

$$h\big(u(x_1, y_2)\big) + h\big(u(x_2, y_1)\big) \leq h\big(u(x_1, y_1)\big) + h\big(u(x_2, y_2)\big).$$

Dropping the second term on the left-hand side and replacing both terms on the right-hand side with their maximum, we obtain

$$h\big(u(x_1, y_2)\big) \leq 2 \cdot h\Big(\max\{u(x_1, y_1), u(x_2, y_2)\}\Big). \tag{14}$$

Consider a pair of points t, t' from I. Integrating the bound on the derivative (7) from t to t', we get

$$\ln\left(\frac{h(t')}{h(t)}\right) \geq \alpha \cdot (t' - t) \quad \text{for} \quad t \leq t'. \tag{15}$$

Pick $t = \max\{u(x_1, y_1), u(x_2, y_2)\}$ and $t' = u(x_1, y_2)$. Consider two cases depending on whether $t \leq t'$ or $t > t'$. If $t \leq t'$, plugging t and t' into into (15) and taking into account the bound (14), we obtain

$$\ln 2 \geq \alpha \cdot (u(x_1, y_2) - \max\{u(x_1, y_1), u(x_2, y_2)\}).$$

If $t > t'$, this inequality holds trivially as the right-hand side is negative. We conclude that

$$u(x_1, y_2) \leq \max\{u(x_1, y_1), u(x_2, y_2)\} + \frac{\ln 2}{\alpha}$$

for a set of (x_1, y_1) and (x_2, y_2) of full π^*-measure. Thus π^* is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. $\square$

The following theorem deals with negative α.

Theorem 7. *For a market with aligned preferences, assume that the utility function u takes values in an open interval $I \subset \mathbb{R}$, possibly infinite. Let $h\colon I \to (-\infty, 0)$ be a differentiable increasing function and assume that there exists $\alpha < 0$ such that*

$$\frac{h'(t)}{h(t)} \geq \alpha \quad \textit{for any} \quad t \in I.$$

Then any solution to the transportation problem with cost (12) is ε-egalitarian with

$$\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}.$$

The function $h(t) = -\exp(-|\alpha| \cdot t)$ satisfies the requirements of Theorem 7.

Proof. Without loss of generality, we assume that μ and ν are probability measures. Recall that $U_{\min}(\pi)$ is the essential infimum of $u(x, y)$ with respect to measure π, i.e.,

$$U_{\min}(\pi) = \inf\big\{\lambda \in \mathbb{R} : \ \pi(\{u(x, y) < \lambda\}) > 0\big\}.$$

The egalitarian lower bound is given by

$$U^*_{\min}(\mu, \nu) = \sup_{\pi \in \Pi(\mu,\nu)} U_{\min}(\pi). \tag{16}$$

Consider the set C of all hypothetical couples whose utility is below the egalitarian lower bound by more than ε

$$C = \{(x, y) \in X \times Y : u(x, y) < U^*_{\min}(\mu, \nu) - \varepsilon\}\,.$$

Let π^* be a solution to an optimal transportation problem with cost (12). Our goal is to show that $\pi^*(C)$ cannot be too big. Fix small $\delta > 0$ and find a matching $\pi' \in \Pi(\mu, \nu)$ such that

$$U_{\min}(\pi') \geq U^*_{\min}(\mu, \nu) - \delta.$$

Note that we may not be able to find such a matching for $\delta = 0$ since the supremum in (16) may not be attained. We get

$$\int_{X\times Y} h(u(x, u))\, \mathrm{d}\pi'(x, y) \leq \int_{X\times Y} h(u(x, u))\, \mathrm{d}\pi^*(x, y)$$

since π^* is the optimal matching for the transportation problem with cost $-h(u(x, y))$. Since h is increasing, $h(u(x, y)) \geq h(U^*_{\min}(\mu, \nu) - \delta)$ for π'-almost all pairs (x, y). Thus, the left-hand side admits the following bound

$$h(U^*_{\min}(\mu, \nu) - \delta) \leq \int_{X\times Y} h(u(x, u))\, \mathrm{d}\pi'(x, y)$$

and thus

$$h(U^*_{\min}(\mu,\nu)-\delta) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y).$$

Since h and u are continuous and $\delta > 0$ was arbitrary, we get

$$h(U^*_{\min}(\mu,\nu)) \leq \int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y). \tag{17}$$

Using monotonicity and negativity of h and the definition of C, we obtain

$$\begin{aligned}\int_{X\times Y} h(u(x,u))\,\mathrm{d}\pi^*(x,y) &= \int_C h(u(x,u))\,\mathrm{d}\pi^*(x,y) + \int_{X\times Y\setminus C} h(u(x,u))\,\mathrm{d}\pi^*(x,y)\\ &\leq \int_C h(u(x,u))\,\mathrm{d}\pi^*(x,y)\\ &\leq h(U^*_{\min}(\mu,\nu)-\varepsilon)\cdot\pi^*(C).\end{aligned}$$

Combining this bound with (17) gives

$$h(U^*_{\min}(\mu,\nu)) \leq h(U^*_{\min}(\mu,\nu)-\varepsilon)\cdot\pi^*(C)$$

and thus

$$\pi^*(C) \leq \frac{h(U^*_{\min}(\mu,\nu))}{h(U^*_{\min}(\mu,\nu)-\varepsilon)}.$$

Note that the inequality changes direction because of the negativity of h. Similarly to (15), integrating the bound on the logarithmic derivative of h, we obtain

$$\ln\left(\frac{|h(t')|}{|h(t)|}\right) \leq |\alpha|\cdot(t'-t) \quad \text{for} \quad t \leq t'$$

and conclude that

$$\pi^*(C) \leq \exp(-|\alpha|\cdot\varepsilon).$$

Plugging in $\varepsilon = \frac{\max\{1,\ln|\alpha|\}}{|\alpha|}$, we get

$$\pi^*(C) \leq \min\left\{\exp(-1),\, \frac{1}{|\alpha|}\right\}$$

and thus

$$\pi^*(C) \leq \varepsilon.$$

We conclude that π^* is ε-egalitarian with $\varepsilon = \frac{\max\{1,\ln|\alpha|\}}{|\alpha|}$. □

Proof of Theorem 1. Theorem 1 immediately follows from Theorems 6 and 7. Consider $h(t) = \exp(\alpha \cdot t)$. For $\alpha > 0$, Theorem 6 implies that the solution to the optimal transportation problem with the cost $c(x, y) = -h\big(u(x, y)\big)$ is ε-stable with $\varepsilon = \frac{\ln 2}{\alpha}$. For $\alpha < 0$, Theorem 6 gives ε-egalitarianism with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$. Multiplying a cost function by a positive factor and adding a constant does not affect the optimum. Thus a solution to the transportation problem with cost

$$c_\alpha(x, y) = \frac{1 - \exp(\alpha \cdot u(x, y))}{\alpha}$$

has the same properties.

Proof of Corollary 1. We first show that the sets $\Pi^u_{+\infty}(\mu, \nu)$ and $\Pi^u_{-\infty}(\mu, \nu)$, corresponding to $\alpha \to \pm\infty$, are non-empty. The argument for the two cases is identical, and so we focus on $\Pi^u_{+\infty}(\mu, \nu)$. Consider a sequence $\alpha_n \to +\infty$ and let π_n be a solution to the optimal transportation problem from Theorem 1 with $\alpha = \alpha_n$. Such a solution is guaranteed to exist under our assumptions on u; see Remark 3.1. The set of all transportation plans $\Pi(\mu, \nu)$ for $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with Polish X and Y is sequentially compact in the topology of weak convergence; this is a corollary of Prokhorov's theorem, see Lemma 4.4 in Villani (2009). Thus, possibly passing to a subsequence, we conclude that the sequence π_n converges weakly to some $\pi_{+\infty} \in \Pi(\mu, \nu)$. Thus $\Pi^u_{+\infty}(\mu, \nu)$ and $\Pi^u_{-\infty}(\mu, \nu)$ are non-empty.

We now show that $\Pi^u_{+\infty}(\mu, \nu)$ consists of stable matchings. In other words, we demonstrate that $\pi_{+\infty}$ is stable. Consider a continuous function $f\colon (X \times Y)^2 \to \mathbb{R}$ given by

$$f(x_1, y_1, x_2, y_2) = \max\Big\{0, \quad u(x_1, y_2) - \max\left\{u(x_1, y_1),\, u(x_2, y_2)\right\}\Big\}.$$

A matching π is stable if and only if $\int f \, \mathrm{d}\pi \times \mathrm{d}\pi = 0$. Note that for an ε-stable matching, this integral does not exceed $\varepsilon \cdot \mu(X) \cdot \nu(Y)$. We obtain

$$\int f \, \mathrm{d}\pi_{+\infty} \times \mathrm{d}\pi_{+\infty} = \lim_{n \to \infty} \int f \, \mathrm{d}\pi_n \times \mathrm{d}\pi_n \leq \lim_{n \to \infty} \frac{\ln 2}{\alpha_n} \cdot \mu(X) \cdot \nu(Y) = 0.$$

Since the left-hand side is non-negative, we get that it equals zero. Thus $\pi_{+\infty}$ is stable. We conclude that $\Pi^u_{+\infty}(\mu, \nu)$ consists of stable matchings, and so a stable matching exists.

To show that $\Pi^u_{-\infty}(\mu, \nu)$ consists of egalitarian matchings, consider a weak limit $\pi_{-\infty}$ of a sequence of matchings π_{α_n}, where $\alpha_n \to -\infty$. Fix $\varepsilon > 0$ and let C be the set of hypothetical couples whose utility is below the egalitarian lower bound by more than ε

$$C_\varepsilon = \{(x,y) \in X \times Y \colon u(x,y) < U^*_{\min}(\mu,\nu) - \varepsilon\}\,.$$

By the continuity of u, the set C is open and thus

$$\pi_{-\infty}(C_\varepsilon) \le \liminf_{n\to\infty} \pi_{\alpha_n}(C_\varepsilon)$$

By Theorem 1, the right-hand side goes to zero, and thus $\pi_{-\infty}(C_\varepsilon) = 0$ for any $\varepsilon > 0$. Sets C_ε are decreasing in ε. Hence,

$$\pi_{-\infty}(C_0) = \pi_{-\infty}\left(\cup_{\varepsilon>0} C_\varepsilon\right) = \lim_{\varepsilon\to 0} \pi_{-\infty}(C_\varepsilon) = 0.$$

Since $\pi_{-\infty}(C_0) = 0$, for $\pi_{-\infty}$-almost all couples (x,y) the utility $u(x,y)$ is at least $U^*_{\min}(\mu,\nu)$, i.e., $\pi_{-\infty}$ is an egalitarian matching. Therefore, all elements of $\Pi^u_{-\infty}(\mu,\nu)$ are egalitarian matchings.

Finally, the convexity of $\Pi^u_{+\infty}(\mu,\nu)$ and $\Pi^u_{-\infty}(\mu,\nu)$ follows from the convexity of the set of solutions to the optimal transportation problem. The weak closedness of $\Pi^u_{+\infty}(\mu,\nu)$ and $\Pi^u_{-\infty}(\mu,\nu)$ follows from the weak closeness of the set of solutions via the standard diagonal procedure.

Proof of Theorem 2. Let π be an ε-stable matching with marginals μ and ν. Since u is continuous, Lemma 2 implies that, for any $(x_1,y_1),(x_2,y_2) \in \mathrm{supp}(\pi)$,

$$u(x_1,y_2) \le \max\{u(x_1,y_1),\, u(x_2,y_2)\} + \varepsilon.$$

By non-negativity of u, we get

$$u(x_1,y_2) \le u(x_1,y_1) + u(x_2,y_2) + \varepsilon.$$

Let π' be any other matching with marginals μ and ν. Consider a distribution $\lambda \in \mathcal{M}_+\big((X\times Y)\times(X\times Y)\big)$ such that the marginals of λ on (x_1,y_1) and on (x_2,y_2) are equal to π and the marginal on (x_1,y_2) is π'. We get

$$\begin{aligned} W(\pi') &= \int_{X\times Y} u(x_1,y_2)\ \mathrm{d}\pi'(x_1,y_2) = \int_{(X\times Y)\times(X\times Y)} u(x_1,y_2)\ \mathrm{d}\lambda(x_1,y_1,x_2,y_2) \\ &\le \int_{(X\times Y)\times(X\times Y)} \left(u(x_1,y_1) + u(x_2,y_2) + \varepsilon\right)\ \mathrm{d}\lambda(x_1,y_1,x_2,y_2) = \\ &= \int_{X\times Y} u(x_1,y_1)\ \mathrm{d}\pi(x_1,y_1) + \int_{X\times Y} u(x_2,y_2)\ \mathrm{d}\pi(x_2,y_2) + \varepsilon = \\ &= 2W(\pi) + \varepsilon. \end{aligned}$$

We thus obtain

$$W(\pi) \geq \frac{1}{2}\left(W(\pi') - \varepsilon\right)$$

for any matching π'. In particular, this inequality holds for π' maximizing welfare. Thus $W(\pi) \geq \frac{1}{2}\left(W^*(\mu,\nu) - \varepsilon\right)$.

Now we show that a substantial fraction of agents in an ε-stable matching π have utilities above the egalitarian lower bound $U^*_{\min}(\mu,\nu)$. Consider the set of hypothetical couples whose utility is more than ε below $U^*_{\min}(\mu,\nu)$

$$C = \left\{(x,y) \in X \times Y \colon u(x,y) < U^*_{\min}(\mu,\nu) - \varepsilon\right\}.$$

Our goal is to show that $\pi(C)$ cannot be too big. Let π' be the egalitarian matching, which exists by Corollary 1. Take λ as in the construction above for the pair π and π'. In other words, λ is a distribution on $(X \times Y) \times (X \times Y)$ with marginals π on (x_1, y_1) and (x_2, y_2) and π' on (x_1, y_2). Thus $u(x_1, y_2) \geq U^*_{\min}(\mu,\nu)$ on a set of full λ-measure. By ε-stability, $u(x_1, y_2) \leq \max\left\{u(x_1,y_1),\, u(x_2,y_2)\right\} + \varepsilon$ and thus

$$\max\{u(x_1,y_1),\, u(x_2,y_2)\} + \varepsilon \geq U^*_{\min}(\mu,\nu).$$

Towards a contradiction, assume that $\pi(C) > \frac{1}{2}\pi(X \times Y)$. Then $\lambda(C \times C) > 0$ by Dirichlet's pigeonhole principle. On the other hand,

$$\max\{u(x_1,y_1),\, u(x_2,y_2)\} + \varepsilon < U^*_{\min}(\mu,\nu)$$

on $C \times C$. This contradiction implies that $\pi(C) \leq \pi(X \times Y)/2$ and thus any ε-stable matching π is ε'-egalitarian with $\varepsilon' = \max\{1/2,\, \varepsilon\}$.

Proof of Theorem 3. We shall prove a more general result that applies to finite populations X_n and Y_n that are close to the continuous distributions μ and ν, but may not be i.i.d. samples from these distributions. For example, X_n and Y_n can be given by the collections $\frac{k}{n}$-quantiles, $k = 1, \ldots, n$ of μ and ν. First, we formalize what is meant by "close."

Definition 1. Probability measures τ and τ' in $\Delta(\mathbb{R}^d)$ are ε-close if, for any $z \in \mathbb{R}^d$, the difference between their distribution functions $F(z) = \tau(\{w \in \mathbb{R}^d \colon w \leq z\})$ and $F'(z) = \tau'(\{w \in \mathbb{R}^d \colon w \leq z\})$ does not exceed ε.

We identify a finite collection of points $z_1, \ldots, z_n$ in $\mathbb{R}^d$ with its "empirical" distribution $\tau_n = \frac{1}{n}\sum_{i=1}^n \delta_{z_i}$. This allows us to say that $z_1, \ldots, z_n$ is ε-close to a distribution

$\tau \in \Delta(\mathbb{R}^d)$, or to another collection of points $z'_1, \ldots, z'_n$. Finally, when X_n and Y_n are finite and of the same cardinality n, we say that a matching π is deterministic if it is a matching in the usual sense. In an abuse of notation, $\Pi(X_n, Y_n)$ denotes the set of such matchings.

Theorem 8. *Suppose that X and Y are two compact (non-trivial) intervals in $\mathbb{R}$. Let $\mu \in \Delta(X)$ and $\nu \in \Delta(Y)$ be non-atomic. Let $\pi \in \Pi(\mu, \nu)$, and let X_n and Y_n be two finite populations of size n that are ε-close to μ and ν, respectively. Let γ be a parameter such that $1 \leq \gamma \leq \frac{\sqrt{n}}{2\sqrt{2}\ln n}$. Then, with probability at least*

$$1 - 5n^{1-2\gamma},$$

there exists $\pi_n \in \Pi(X_n, Y_n)$ such that, for all matched x_i and y_j in π_n,

$$\min\{F(\xi_{i,j}),\ G(\eta_{i,j})\} \geq 1 - 2\sqrt{2} \cdot \gamma \cdot \frac{\ln n}{\sqrt{n}}$$

and π_n is ε'-close to π with

$$\varepsilon' = 9 \max\left\{\varepsilon,\ \sqrt{\frac{\ln(4(n+1))}{2n}}\right\} + \frac{1}{\gamma \ln n}.$$

Recall the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky, Kiefer, and Wolfowitz, 1956), which states that the empirical distribution of a sample from a distribution is close to the distribution with high probability. An independent sample $z_1, \ldots, z_n$ from a distribution $\tau \in \Delta(\mathbb{R}^d)$ is ε-close to τ with probability at least $1 - 2 \cdot \exp(-2n\varepsilon^2)$ for $d = 1$ (Massart, 1990) and $1 - 2d(n+1) \cdot \exp(-2n\varepsilon^2)$ for $d \geq 2$ (Naaman, 2021).

By the Dvoretzky–Kiefer–Wolfowitz, if X_n and Y_n are i.i.d. samples from μ and ν, then, for large n these samples become arbitrarily close to μ and ν with high probability. Thus Theorem 3 follows from Theorem 8 applied to such i.i.d. samples.

We prove Theorem 8 in two steps. We first show that any continuous matching $\pi \in \Pi(\mu, \nu)$ can be approximated by a matching of finite populations X_n and Y_n, without providing any guarantee on agents' utilities with respect to their idiosyncratic components. Second, we demonstrate that any matching of X_n and Y_n, with high probability, can be modified so that the new matching is close to the original one and each agent is close to getting their best utilities with respect to the idiosyncratic component.

Proposition 1. *Let $\pi \in \Pi(\mu, \nu)$ with non-atomic $\mu, \nu \in \Delta(\mathbb{R})$. Let X_n and Y_n be ε-close to μ and ν, respectively. Then there exists a deterministic matching π_n of X_n and Y_n that is δ-close to π with*

$$\delta = 9 \max \left\{ \varepsilon,\ \sqrt{\frac{\ln(4(n+1))}{2n}} \right\}.$$

The first step in proving Proposition 1 is to show that any matching π can be approximated with an auxiliary deterministic matching of *some* finite populations X'_n and Y'_n, i.e., in contrast to Proposition 1 the finite populations are not given.

Lemma 3. *Let $\pi \in \Pi(\mu, \nu)$ with*[11] *$\mu, \nu \in \Delta(\mathbb{R})$ having compact support. Then there exist finite populations X'_n and Y'_n of size n and $\pi'_n \in \Pi(X'_n, Y'_n)$ such that π'_n is ε'-close to π, X'_n is ε'-close to μ, and Y'_n is ε'-close to ν with*

$$\varepsilon' = \sqrt{\frac{\ln(4(n+1))}{2n}}.$$

Proof. The existence of π'_n is established via the probabilistic method. Sample n points $z'_1, \dots, z'_n$ independently from π. By the Dvoretzky–Kiefer–Wolfowitz inequality, the sample $z'_1, \dots, z'_n$ is ε'-close to π with probability at least $1 - 4(n+1) \cdot \exp\left(-2n\varepsilon'^2\right)$. This probability is positive for any $\varepsilon' > \sqrt{\frac{\ln(4(n+1))}{2n}}$. Thus, for any such ε', there exists a collection of n points that is ε'-close to π. Letting ε' go to $\sqrt{\frac{\ln(4(n+1))}{2n}}$ from above, and choosing a convergent subsequence, there exists such $z'_1, \dots, z'_n$ for $\varepsilon' = \sqrt{\frac{\ln(4(n+1))}{2n}}$ as well. Each point z'_i is a pair (x'_i, y'_i). Since $z'_1, \dots, z'_n$ is ε'-close to π, the marginal $X'_n = \{x'_1, \dots, x'_n\}$ is ε'-close to μ and $Y'_n = \{y'_1, \dots, y'_n\}$ is ε'-close to ν. We demonstrate this for X'_n; the argument for Y'_n is analogous. Define the empirical distribution of X'_n as $\mu'_n = \sum_{i=1}^n \delta_{x'_i}$. Since π'_n is ε'-close to π, we have that $|\pi(\{(x,y) \le (t_x, t_y)\}) - \pi'_n((\{(x,y) \le (t_x, t_y))\}))| \le \varepsilon'$ for any (t_x, t_y)). This implies

$$|\mu(\{x \le t_x\}) - \mu'_n((\{x \le t_x\}))| \le \varepsilon'$$

and thus X'_n is ε'-close to μ. □

To prove Proposition 1, without loss of generality, we can assume that μ and ν are uniform on $[0, 1]$, which is equivalent to working in the space of quantiles for general μ and ν. The next lemma shows that if a finite population $w_1, \dots, w_n$ is close to the uniform distribution $[0, 1]$, then the points are close to the equidistant points i/n.

[11] Note that this lemma does not require μ and ν to be non-atomic. However, for atomic μ and ν, we are not guaranteed that X'_n and Y'_n contain n *distinct* points.

Lemma 4. *Let $w_1, \ldots, w_n$ be numbers in $[0,1]$ ordered so that $w_i \le w_{i+1}$. If they are ε-close to the uniform distribution on $[0,1]$, then for any $i = 1, \ldots, n$ we have*

$$\left| w_i - \frac{i}{n} \right| \le \varepsilon$$

Proof. Let $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{w_i}$ and μ be uniform on $[0,1]$. Since μ_n is ε-close to μ, we have that $|\mu([0,t]) - \mu_n([0,t])| \le \varepsilon$. Plugging in $t = w_i$ we get, $\left|w_i - \frac{i}{n}\right| \le \varepsilon$. □

The following result allows us to show that a deterministic matching of X_n and Y_n and a deterministic matching of X_n' and Y_n' are close to each other if matched couples are close.

Lemma 5. *Let X_n, Y_n, X_n', and Y_n' be finite collections n points in $[0,1]$ in non-decreasing order. Suppose X_n, X_n', Y_n, and Y_n' are α-close to the uniform distribution on $[0,1]$. If $\pi_n \in \Pi(X_n, Y_n)$ and $\pi_n' \in \Pi(X_n', Y_n')$ such that x_i is matched with y_j if and only if x_i' is matched with y_j', then π_n and π_n' are 8α-close to each other.*

Proof. By Lemma 4, we have that x_i, x_i' are within α from i/n and y_j, y_j' are within α from j/n. Fix $z = (z_1, z_2) \in [0,1]^2$ and let $z^+ = (z_1 + 2\alpha, z_2 + 2\alpha)$. We have

$$\pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right) \le \pi_n'\left(\{w \in [0,1]^2 \colon w \le z^+\}\right)$$

and, similarly,

$$\pi_n'\left(\{w \in [0,1]^2 \colon w \le z\}\right) \le \pi_n\left(\{w \in [0,1]^2 \colon w \le z^+\}\right).$$

Note that

$$\begin{aligned}\pi_n\left(\{w \in [0,1]^2 \colon z < w \le z^+\}\right) \le\ & \pi_n\left(\{w \in [0,1]^2 \colon z_1 < w_1 \le z_1^+\}\right) \\ & + \pi_n\left(\{w \in [0,1]^2 \colon z_2 < w_2 \le z_2^+\}\right).\end{aligned}$$

Consider the first summand on the right-hand side. We have

$$\pi_n\left(\{w \in [0,1]^2 \colon z_1 < w_1 \le z_1^+\}\right) = \frac{1}{n}\left|i = 1, \ldots n \colon z_1 < x_i \le z_1 + 2\alpha\right|,$$

as $\pi_n \in \Pi(X_n, Y_n)$. Since each x_i is within α of i/n, we get

$$\pi_n\left(\{w \in [0,1]^2 \colon z_1 < w_1 \le z_1^+\}\right) \le \frac{1}{n}\left|i = 1, \ldots n \colon z_1 - \alpha < \frac{i}{n} \le z_1 + 3\alpha\right| \le 4\alpha.$$

Using an analogous bound for the second summand, we get

$$\pi_n\left(\{w \in [0,1]^2 \colon z < w \le z^+\}\right) \le 8\alpha.$$

Analogous calculations for π'_n yield the same bound. Thus,

$$\left|\pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right) - \pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right)\right| \le 8\alpha$$

for any z and, therefore, π_n and π'_n are 8α-close to each other. □

We are now ready to prove Proposition 1.

Proof of Proposition 1. As discussed above, we can assume that μ and ν are uniform on $[0,1]$ by working in the space of quantiles. Let X'_n, Y'_n, and $\pi'_n \in \Pi(X'_n, Y'_n)$ be obtained from Lemma 3. Suppose points in X_n, X'_n, Y_n, and Y'_n are all ordered in non-decreasing order. Define a matching π_n of X_n and Y_n as follows: x_i and y_j are matched if and only if x'_i and y'_j are matched under π'_n. Taking $\alpha = \max\{\varepsilon, \varepsilon'\}$, we have that X_n and X'_n are α-close to the uniform distribution on $[0,1]$ and Y_n and Y'_n are α-close to the uniform distribution on $[0,1]$. By Lemma 5, we have that π_n and π'_n are 8α-close to each other and thus π_n is 9α-close to π, completing the proof. □

The next step in proving Theorem 8 is to show that any matching of X_n and Y_n can be modified so that the new matching is close to the original one and each agent is close to getting their best utilities with respect to the idiosyncratic component.

Proposition 2. *Let π_n be a deterministic matching of two finite populations X_n and Y_n of size $n \ge 8$, and let $\xi_{i,j}$ and $\eta_{i,j}$ be by i.i.d. random variables with continuous distribution functions F and G, respectively. For any γ such that $1 \le \gamma \le \frac{\sqrt{n}}{2\sqrt{2}\ln n}$, with probability of at least*

$$1 - 5n^{1-2\gamma}$$

there exists $\pi'_n \in \Pi(X_n, Y_n)$ such that for all matched x_i and y_j we have

$$\min\left\{F(\xi_{i,j}), \quad G(\eta_{i,j})\right\} \ge 1 - \delta, \qquad \textit{where} \qquad \delta = 2\sqrt{2} \cdot \gamma \cdot \frac{\ln n}{\sqrt{n}}$$

and π'_n is ε-close to π_n with

$$\varepsilon = \frac{1}{\gamma \ln n}.$$

We will need the following lemma resembling Lemma 5.

Lemma 6. *Let X_n and Y_n be two collections of n distinct numbers in non-decreasing order. Let $\pi_n, \pi'_n \in \Pi(X_n, Y_n)$, and t be a positive integer such that, if x_i is matched with y_j under π_n and to y_k under π'_n, then $|j-k| \le t-1$. Then π_n and π'_n are $\frac{t-1}{n}$-close to each other.*

Proof. Without loss of generality $x_i = y_i = i/n$. Denote $\delta = (t-1)/n$ and fix $z = (z_1, z_2) \in [0,1]^2$ and let $z^+ = (z_1, z_2 + \delta)$. We have

$$\pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right) \le \pi'_n\left(\{w \in [0,1]^2 \colon w \le z^+\}\right)$$

and similarly

$$\pi'_n\left(\{w \in [0,1]^2 \colon w \le z\}\right) \le \pi_n\left(\{w \in [0,1]^2 \colon w \le z^+\}\right).$$

Note that

$$\pi_n\left(\{w \in [0,1]^2 \colon z < w \le z^+\}\right) \le \pi_n\left(\{w \in [0,1]^2 \colon z_2 < w_2 \le z_2 + \delta\}\right).$$

The right-hand side is equal to the fraction of $i = 1, \ldots n$ such that $i/n \in (z_2, z_2 + \delta]$. Thus

$$\pi_n\left(\{w \in [0,1]^2 \colon z < w \le z^+\}\right) \le \delta.$$

Analogous calculations for π'_n yield the same bound, and thus

$$\left|\pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right) - \pi_n\left(\{w \in [0,1]^2 \colon w \le z\}\right)\right| \le \delta.$$

Since z was arbitrary, we have that π_n and π'_n are δ-close to each other. □

We are now ready to prove Proposition 2. The proof adapts the approach of Erdős and Rényi (1964), who studied the existence of a perfect matching in a random bipartite graph where each edge is traced independently with probability p, to graphs where some edges are never traced. Leveraging insights of Petrov (2016), we obtain the explicit bound that applies to any fixed n rather than the asymptotic result for $n \to \infty$ as in the classical paper.

Proof of Proposition 2. Suppose X_i and Y_i are ordered in non-decreasing order. Consider a bipartite graph $G = (X_n, Y_n, E)$ with vertices X_n and Y_n obtained by the edges in π_n and additional edges. Specifically, E contains an edge between all x_i and y_k such that x_i is matched with y_j in π_n and $|k - j| \le t - 1$. Here t is

$$t = \left\lfloor \frac{\sqrt{(2\ln 4) \cdot n}}{\delta} \right\rfloor,$$

where $\delta = 2\sqrt{2} \cdot \gamma \cdot \frac{\ln n}{\sqrt{n}}$, and $\lfloor z \rfloor$ denotes the integer part of z.

Consider a random subgraph $G_\delta = (X_n, Y_n, E_q)$ of G where only those edges $e = (i,k) \in E$ are left for which $F(\xi_{i,k}) \geq 1-\delta$ and $G(\eta_{i,k}) \geq 1-\delta$. In other words, G_δ is a subgraph of G where each edge is eliminated with probability $q = 1-\delta^2$. We aim to show that, with high probability, G_δ contains a perfect matching. Conditional on this high-probability event, we define a matching π'_n of X_n and Y_n as follows: x_i is matched with y_l if $e = (x_i, y_l)$ enters the perfect matching of G_δ. By Lemma 6, π_n and π'_n are ε-close to each other with

$$\varepsilon = \frac{t}{n} = \frac{1}{\gamma \ln n}.$$

By the definition of G_δ, $F(\xi_{i,l}) \geq 1-\delta$ and $G(\eta_{i,l}) \geq 1-\delta$ for any matched pair (x_i, y_l).

To complete the proof, we must bound the probability that G_δ admits a perfect matching. By Hall's lemma, G_δ contains a perfect matching if and only if for any subset $S \subset X_n$, the number of neighbors of S in Y_n is at least $|S|$. Equivalently, for any $S \subset X_n$ and $T \subset Y_n$ such that $|S| + |T| = n+1$, there must be an edge between S and T in G_δ. Consider S and T such that $|S| \leq |T|$ and let $m = |S|$. The number of edges $|E_G(S,T)|$ between such S and T in G is at least

$$|E_G(S,T)| \geq \begin{cases} m \cdot \frac{2t-m+1}{2}, & \text{if } m \leq t \\ t \cdot \frac{t+1}{2}, & \text{if } m > t \end{cases}. \tag{18}$$

This lower bound corresponds to the "diagonal" matching π_n—i.e., x_i is matched with y_i—and sets $S = \{x_1, \ldots, x_m\}$ and $T = \{y_m, y_{m+1}, \ldots, y_n\}$. By definition of G, there is an edge between x_i and each y_k with index k within $t-1$ from x_i's match. Thus for $m \leq t$, there are edges between each $x_i \in S$ and $y_m, \ldots, y_{t+i-1}$. For $m > t$, there are edges between $x_i \in S$ with $i \geq m-t+1$ and $y_m, \ldots, y_{t+i-1}$. Counting the numbers of such edges, we obtain (18).

Each edge of G is eliminated in G_δ with probability q. Thus there are no edges between S and T in G_δ with probability $q^{|E_G(S,T)|}$, and the union bound applied to the probability that G_δ does not contain a perfect matching gives the following

$$\mathbb{P}\left(G_\delta \text{ has no perfect matching}\right) \leq \sum_{m=1}^{n} \sum_{\substack{S,T: \\ |S|=m \\ |T|=n-m+1}} q^{|E_G(S,T)|}$$

Denote the right-hand side of (18) by $E^*(m,t)$. Thus the number of edges between S and T in G is at least $E^*(m,t)$ if $m = |S| \le |T| = n-m+1$. Since the roles of S and T are symmetric, the number of edges is at least $E^*(n+1-m,t)$ for $m \ge n-m+1$. Taking into account that the number of subsets S and T with $|S| = m$ is given by $\binom{n}{m}\binom{n}{n+1-m}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(G_\delta \text{ has no perfect matching}\right) &\le \sum_{m \le \frac{n+1}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{E^*(m,t)} \\
&\quad + \sum_{m \ge \frac{n+1}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{E^*(n+1-m,t)} \\
&= 2\sum_{m \le \frac{n+1}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{E^*(m,t)}
\end{aligned}$$

Plugging in the explicit value of $E^*(m,t)$, we get

$$\begin{aligned}
\mathbb{P}\left(G_\delta \text{ has no perfect matching}\right) &\le 2\sum_{m=1}^{t} \binom{n}{m}\binom{n}{n+1-m} q^{m\cdot\frac{2t-m+1}{2}} \\
&\quad + 2\sum_{t+1 \le m \le \frac{n+1}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{t\cdot\frac{t+1}{2}}.
\end{aligned} \tag{19}$$

Consider the first sum in (19). We have $m \cdot \frac{2t-m+1}{2} \ge m \cdot \frac{t}{2}$. Using a rough bound $\binom{n}{m} \le n^m$ and $\binom{n}{n+1-m} = \binom{n}{m-1} \le n^{m-1}$, and assuming $n^2 \cdot q^{\frac{t}{2}} < 1$, we get

$$\begin{aligned}
\sum_{m=1}^{t} \binom{n}{m}\binom{n}{n+1-m} q^{m\cdot\frac{2t-m+1}{2}} &\le \sum_{m=1}^{t} n^{2m-1} q^{m\cdot\frac{t}{2}} \\
&\le n \cdot q^{\frac{t}{2}} \sum_{l=0}^{\infty} \left(n^2 \cdot q^{\frac{t}{2}}\right)^l = \frac{nq^{\frac{t}{2}}}{1-n^2\cdot q^{\frac{t}{2}}}
\end{aligned} \tag{20}$$

We now ensure that $n^2 \cdot q^{\frac{t}{2}} < 1$, and thus the infinite geometric series converge. Indeed,

$$\begin{aligned}
n^2 q^{\frac{t}{2}} &= n^2(1-\delta^2)^{\frac{t}{2}} \le n^2\left(\exp(-\delta^2)\right)^{\frac{t}{2}} \\
&= \exp\left(2\ln n - \delta^2 \cdot \frac{t}{2}\right),
\end{aligned}$$

where we used the fact that $1 + z \le \exp(z)$ for any $z \in \mathbb{R}$. Since $t \ge \frac{\sqrt{(2\ln 4)\cdot n}}{\delta}$ and $\delta = 2\sqrt{2}\cdot\gamma\cdot\frac{\ln n}{\sqrt{n}}$, we get $\delta^2 \cdot \frac{t}{2} \ge 2\sqrt{\ln 4}\cdot\gamma\cdot\ln n$ and thus

$$n^2 q^{\frac{t}{2}} \le \exp\left(2\ln n - \delta^2\cdot\frac{t}{2}\right) \le n^{2(1-\gamma\cdot\sqrt{\ln 4})} \tag{21}$$

For $n = 8$, we get $n^{2(1-\gamma\cdot\sqrt{\ln 4})} \approx 0.478 < 1/2$. Therefore, for any $n \geq 8$, we get $n^2 q^{\frac{t}{2}} \leq 1/2$, and thus the infinite geometric series converges. Moreover, since $nq^{\frac{t}{2}} = \frac{1}{n} \cdot n^2 q^{\frac{t}{2}}$, we get

$$nq^{\frac{t}{2}} \leq n^{1-2\gamma} \tag{22}$$

from (21), where we replaced $\sqrt{\ln 4}$ with 1. . Thus the first sum in (19) can be bounded as follows:

$$\sum_{m=1}^{t} \binom{n}{m}\binom{n}{n+1-m} q^{m\cdot\frac{2t-m+1}{2}} \leq 2n^{1-2\gamma},$$

where we combined the bound (20) with (22) and $n^2 q^{\frac{t}{2}} \leq 1/2$.

Consider now the second sum in (19). The product of binomial coefficients is maximized at $m = \lfloor n/2 \rfloor$:

$$\binom{n}{m}\binom{n}{n+1-m} \leq \binom{n}{\lfloor n/2 \rfloor} \cdot \binom{n}{\lceil n/2 \rceil} = \binom{n}{\lfloor n/2 \rfloor}^2.$$

Applying the Stirling formula, we get

$$\binom{n}{\lfloor n/2 \rfloor} \leq \frac{\sqrt{2}\cdot 2^n}{\sqrt{\pi n}}.$$

Thus the second sum admits the following bound

$$\sum_{t+1 \leq m \leq \frac{n}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{t\cdot\frac{t+1}{2}} \leq \sum_{t+1 \leq m \leq \frac{n}{2}} \binom{n}{\lfloor n/2 \rfloor}^2 q^{t\cdot\frac{t+1}{2}} \leq \frac{n}{2} \cdot \frac{2\cdot 2^{2n}}{\pi n} q^{t\cdot\frac{t+1}{2}}.$$

Since $t \geq \frac{\sqrt{(2\ln 4)\cdot n}}{\delta}$, we have that

$$2^{2n} \cdot q^{\frac{t^2}{2}} = 2^{2n} \cdot (1-\delta^2)^{\frac{t^2}{2}} \leq 2^{2n} \cdot \exp\left(-\delta^2 \cdot \frac{t^2}{2}\right) \leq 2^{2n} \cdot e^{-n\cdot\ln 4} = 1.$$

Thus

$$\frac{n}{2} \cdot \frac{2\cdot 2^{2n}}{\pi n} q^{t\cdot\frac{t+1}{2}} \leq \frac{1}{\pi} \cdot q^{\frac{t}{2}}.$$

By (22), $q^{\frac{t}{2}} \leq n^{-2\gamma}$, and thus

$$\sum_{t+1 \leq m \leq \frac{n}{2}} \binom{n}{m}\binom{n}{n+1-m} q^{t\cdot\frac{t+1}{2}} \leq \frac{1}{\pi} \cdot n^{-2\gamma}.$$

Putting the bounds for the first and the second sum together, we obtain that

$$\mathbb{P}\left(G_\delta \text{ has no perfect matching}\right) \leq \left(4 + \frac{2}{\pi \cdot n}\right) n^{1-2\gamma} \leq 5 \cdot n^{1-2\gamma}$$

completing the proof. $\square$

Theorem 8 is an immediate corollary of already proved results.

Proof of Theorem 8. The theorem is a combination of Propositions 1 and 2. ∎

Supplemental Appendix

B Proof of Lemma 2

Proof. One direction is straightforward: the pointwise property implies the almost-everywhere property. We prove the opposite direction. Let π be an ε-stable matching of populations $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with continuous utilities u and v. Our goal is to show that, for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$, at least one of the two inequalities (1) holds. Towards a contradiction, suppose that there are $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \mathrm{supp}(\pi)$ such that both inequalities are violated. By continuity of u and v, we can find open neighborhoods $U_1 \subseteq X \times Y$ of (x_1^*, y_1^*) and $U_2 \subseteq X \times Y$ of (x_2^*, y_2^*) such that the inequalities are violated for all $(x_1, y_1), (x_2, y_2) \in U_1 \times U_2$. Since $(x_1^*, y_1^*), (x_2^*, y_2^*) \in \mathrm{supp}(\pi)$, we have $\pi(U_1) > 0$ and $\pi(U_2) > 0$. Thus $U_1 \times U_2$ has a positive $\pi \times \pi$-measure, which contradicts the ε-stability of π. We conclude that ε-stability of π implies that at least one of the two inequalities (1) holds for any $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$. □

C Exact algorithm for stable matching on $\mathbb{R}$

When X and Y are finite, a simple greedy algorithm will result in a stable matching: let (x, y) be the pair with the highest utility from matching. Match (x, y), remove the matched agents, and repeat. In this section, we construct an analogous procedure for the case where X and Y are possibly continuous.

Let $X = Y = \mathbb{R}$ and consider two populations represented by $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ with equal total measure $\mu(X) = \nu(Y)$ and utility $u(x, y) = -|x - y|$. We will refer to the pair (μ, ν) as a market and assume that μ and ν are non-atomic measures, and the difference $\rho = \mu - \nu$ changes sign a finite number of times. We say that ρ changes sign K times if there is a collection of K intervals $I_0, \ldots, I_K$ such that the restriction of ρ to I_k is either a positive or a negative measure, and K is the minimal number with this property. We will refer to the intervals I_j as the imbalance regions.

Proposition 3. *There is a unique stable matching π, which can be constructed via an algorithm. For each $x \in X$, there are at most two distinct types $y, y' \in Y$ such that (x, y) and (x, y') are in the support of π. If μ and ν have piecewise-constant density with at most m intervals of constancy, the algorithm runs in time of the order of m^2.*

A stable matching for (μ, ν) is constructed by sequential simplification the market. The idea is to identify submarkets (μ', ν') that are to be matched together in any stable matching the same way they would be matched if the rest of the population $(\mu-\mu', \nu-\nu')$ did not exist. As we will see, there is always a way to find such (μ', ν') that are to be matched in a simple monotone way. Repeating this procedure, we construct the stable matching for (μ, ν). This requires only a finite number of steps since the number of imbalance regions is finite and monotonically decreases. For piecewise-constant density with m intervals of constancy, the number of sign changes is bounded by m and constructing each simple submarket requires a number of operations proportional to m, hence the quadratic total runtime.

We start with two lemmas describing benchmark markets for which a stable matching can be obtained in one step. For general markets (μ, ν), these benchmark cases will correspond to the submarkets of interest.

If $\mu = \nu$, we call (μ, ν) a *diagonal market*. In such markets, everyone can be matched with their ideal partner $x = y$. The opposite extreme is when nobody can be matched with their hypothetical ideal partner $x = y$. Populations (μ, ν) form a *disjoint market* if μ and ν are mutually singular, i.e., there is a measurable subset $A \subset \mathbb{R}$ such that $\mu(\mathbb{R}\setminus A) = 0$ and $\nu(A) = 0$. An important special case corresponds to $A = (-\infty, t]$ or $A = [t, \infty)$ for some $t \in \mathbb{R}$. In this case, the supports of μ and ν are ordered on the line, i.e., either $\mathrm{supp}(\mu) \leq \mathrm{supp}(\nu)$ or $\mathrm{supp}(\nu) \leq \mathrm{supp}(\mu)$. We call such (μ, ν) an *anti-diagonal market*. Equivalently, (μ, ν) is an anti-diagonal market if μ and ν are mutually singular and ρ changes sign only once.

Lemma 7. *If (μ, ν) are diagonal or anti-diagonal markets, then the stable matching π is unique and is assortative or anti-assortative, respectively.*

In other words, for diagonal markets, π is supported on the solution $F_\mu(x) = F_\nu(y)$, where F_μ and F_ν are the cumulative distribution functions of μ and ν. Since $\mu = \nu$, π is supported on the 45-degree line. In the anti-diagonal case, π is supported on the solution to $F_\mu(x) = 1 - F_\nu(y)$.

Proof. For diagonal and anti-diagonal markets, $\rho = \mu - \nu$ changes sign only once. As shown by McCann (1999), a matching satisfying the no-crossing condition is unique whenever the sign changes at most two times. Assortative and anti-assortative matchings satisfy the no-crossing condition and, thus, are unique matchings with this property. By Lemma 1, they are unique stable ones. $\square$

A market (μ', ν') is a *submarket* of (μ, ν) if $\mu' \leq \mu$ and $\nu' \leq \nu$. Note that $\mu'(\mathbb{R}) =$

$\nu'(\mathbb{R})$ by the assumption that (μ', ν') is a market. For a submarket (μ', ν'), the *residual submarket* is defined by $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$.

Lemma 8. *For any market (μ, ν), there exists a unique diagonal submarket (μ', ν') such that the residual submarket $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$ is disjoint.*

A matching π is a stable matching of (μ, ν) if and only if $\pi = \pi' + \pi''$, where π' is the (unique) stable matching of (μ', ν') and π'' is a stable matching of (μ'', ν'').

Proof. We first construct a diagonal submarket (μ', ν') such that the residual submarket (μ'', ν'') is disjoint. Let $\rho = \mu - \nu$. We get that ρ can also be expressed as $\mu'' - \nu''$. By Hahn's theorem, there is a unique way to represent a signed measure τ as $\tau_+ - \tau_-$, where $\tau_\pm$ are positive measures that are mutually singular. We get that (μ'', ν'') must provide the Hahn decomposition of ρ, i.e., $\mu'' = \rho_+$ and $\nu'' = \rho_-$. Thus (μ', ν') is given by $\mu' = \mu - \rho_+$ and $\nu' = \nu - \rho_-$. It is a diagonal submarket since

$$\mu' - \nu' = (\mu - \nu) - (\rho_+ - \rho_-) = \rho - \rho = 0.$$

Its uniqueness follows from the uniqueness of the Hahn decomposition.

Let π' be the stable matching of the diagonal submarket (μ', ν'). By Lemma 7, matching π' is unique and is given by the assortative matching. If π'' is a stable matching of (μ'', ν''), then $\pi' + \pi''$ is a stable matching of (μ, ν). Indeed, by combining the two markets, we do not create any cross-market blocking pairs as any agent in π' is matched with their best partner.

It remains to show that any stable matching π of (μ, ν) can be represented as $\pi' + \pi''$ with stable π' and π'', where π' is the stable matching of the diagonal submarket (μ', ν'). We adapt an argument from the optimal transport literature with metric costs (see Gangbo and McCann, 1996, Proposition 2.9). Let π_{diag} be the restriction of π to the diagonal $\{(x, y) \colon x = y\}$, i.e., $\pi_{\mathrm{diag}}(A) = \pi'(A \cap \{(x, y) \colon x = y\})$ for any measurable A. Our goal is to show that $\pi' = \pi_{\mathrm{diag}}$. Towards contradiction, assume this equality does not hold. Hence, the marginals μ_{diag} and ν_{diag} of π_{diag} satisfy $\mu_{\mathrm{diag}} \neq \mu'$ and $\nu_{\mathrm{diag}} \neq \nu'$. By the uniqueness of the Hahn decomposition, $\mu'' = \mu - \mu_{\mathrm{diag}}$ and $\nu'' = \nu - \nu_{\mathrm{diag}}$ do not form a disjoint market, i.e., there is a set $B \subset \mathbb{R}$ with $\mu''(B) > 0$ and $\nu''(B) > 0$. Consider a set $S = \mathrm{supp}(\pi) \setminus \{(x, y) \colon x = y\}$. Let X_S and Y_S be the projections of the set S. These are sets of full measure with respect to μ'' and ν'', and thus, the intersection $X_S \cap Y_S$ is non-empty. Pick $t \in X_S \cap Y_S$. There are two couples $(x_1, y_1), (x_2, y_2) \in \mathrm{supp}(\pi)$ with $x_1 = t$ and $y_2 = t$ and $x_1 \neq y_1$ and $x_2 \neq y_2$

which violate the no-crossing condition. By Lemma 1, π cannot be stable. This contradiction implies that $\pi_{\text{diag}} = \pi'$. We conclude that any stable π can be represented as $\pi' + \pi''$, where π' is a stable matching of the diagonal submarket and π'' is a stable matching of a disjoint submarket. □

Lemma 8 reduces the problem of constructing and showing the uniqueness of a stable matching for general (μ, ν) to these questions for disjoint markets.

A submarket (μ', ν') is *independent* if members of (μ', ν') top-rank each other within (μ, ν), i.e., $|x - y'| > |x - y|$ for μ'-almost all x, ν'-almost all y, and $(\nu - \nu')$-almost all y' and $|x' - y| > |x - y|$ for μ'-almost all x, ν'-almost all y, and $(\mu - \mu')$-almost all x'. Independent submarkets are important because they can be matched myopically, i.e., without worrying about the residual populations.

Lemma 9. *Let (μ', ν') be an independent submarket of (μ, ν) and let $(\mu'', \nu'') = (\mu - \mu', \nu - \nu')$ be the residual submarket. A matching π is a stable matching of (μ, ν) if and only if $\pi = \pi' + \pi''$, where π' is a stable matching of (μ', ν') and π'' is a stable matching of (μ'', ν'').*

Proof. Let π' be a stable matching of an independent submarket (μ', ν') and π'' be a stable matching of the residual submarket (μ'', ν''). Then $\pi = \pi' + \pi''$ is a stable matching of (μ, ν) since combining the two markets cannot create cross-market blocking pairs by the independence of (μ', ν').

Now, let π be a stable matching of (μ, ν). We show that π can be represented as $\pi' + \pi''$, where π' is a stable matching of (μ', ν') and π'' be a stable matching of (μ'', ν''). Note that the stability of a submarket follows from the stability of a market, and thus we only need to check that $\pi = \pi' + \pi''$, where π' and π'' are matchings.

Strict inequalities in the definition of independence imply that μ' and μ'' cannot have any mass in common. In other words, μ', μ'' and ν', ν'' are mutually singular. By Hahn's theorem, there are disjoint sets X' and X'' such that μ' and μ'' are given by restricting μ on X' and X'', respectively. The disjoint sets Y' and Y'' are constructed analogously.

We now show that π cannot place positive mass on $X' \times Y''$ and $X'' \times Y'$. First, observe that $\pi(X' \times Y'') = \pi(X'' \times Y')$. Indeed,

$$\pi(X' \times Y'') - \pi(X'' \times Y') = \pi(X' \times (Y' \cup Y'')) - \pi((X' \cup X'') \times Y') = \mu(X') - \nu(Y') = 0.$$

Second, we show that the common value is zero. Towards contradiction, assume that $\pi(X' \times Y'') = \pi(X'' \times Y') > 0$. Thus $\text{supp}(\pi)$ contains $(x_1, y_2) \in X' \times Y''$

and $(x_2, y_2) \in (X'', Y')$. By independence of (μ', ν'), the couple (x_1, y_2) is a blocking pair which contradicts stability of π. We conclude that π has no mass outside of $X' \times Y'$ and $X'' \times Y''$. Consequently, $\pi = \pi' + \pi''$, where π' is the restriction of π to $X' \times Y'$ and π'' is the restriction to $X'' \times Y''$. By the construction, π' is a matching of (μ', ν') and π'' is a matching of (μ'', ν''), which are both stable by the stability of π. □

We say that a matching π is Monge if π is supported on a graph of some function $f\colon \mathbb{R} \to \mathbb{R}$. The following lemma is the key step in the proof of Proposition 3. Starting from disjoint (μ, ν) such that $\rho = \mu - \nu$ changes sign $K \geq 2$ times, this lemma allows us to reduce K sequentially until we reach an anti-diagonal market ($K = 1$).

Lemma 10. *Let (μ, ν) be a disjoint market such that ρ changes sign $K \geq 2$ times. Then μ and ν can be represented as $\mu = \mu_1' + \mu_2' + \mu''$ and $\nu = \nu_1' + \nu_2' + \mu''$ so that*

1. *(μ_1', ν_1') and (μ_2', ν_2') are independent anti-diagonal submarkets of (μ, ν);*
2. *a matching π is a stable matching of (μ, ν) if and only if $\pi = \pi_1' + \pi_2' + \pi''$, where π_i' is the (unique) stable matching of (μ_i', ν_i'), $i = 1, 2$, and π'' is a stable matching of (μ'', ν'')*
3. *such a stable matching π is Monge if and only if π'' is Monge;*
4. *the residual submarket (μ'', ν'') is a disjoint market with $\rho' = \mu' - \nu'$ changing sign at most $K - 1$ times.*
5. *if μ and ν have piecewise constant density with m intervals of constancy, then (μ_1', ν_1') and (μ_2', ν_2') can be constructed in time of the order of m.*

Proof. Since (μ, ν) is a non-atomic disjoint market with $K \geq 2$, we can find points $a_0 < a_1 < a_2 < \ldots < a_{K-1} < a_{K+1}$ with $a_0 = -\infty$ and $a_K = +\infty$ such that either μ is supported on intervals $I_k = (a_k, a_{k+1})$ with even k and ν is supported on I_k with odd k, or the other way around. By the minimality of K, each interval I_k carries strictly positive μ-mass or strictly positive ν-mass.

Consider $\rho = \mu - \nu$ and let δ be a non-negative number. In each closed interval $\bar{I}_k = [a_k, a_{k+1}]$ we aim to pick a pair of points $a_k^+ \leq a_{k+1}^-$ so that the following conditions are satisfied:

1. equal-weight condition: $\rho([a_k^-, a_k^+]) = 0$ for $k = 1, \ldots, K$;

2. equal-distance condition: $|a_k^+ - a_k^-| = \delta$ for $k = 1, \dots, K$.

The conditions on the collection of points $a_k^\pm$ and δ are closed. Consider the maximal $\delta \geq 0$ such that points satisfying the conditions exist and let $a_k^\pm$ be the corresponding collection of points.

As δ cannot be increased further, there is an interval $I_{k^*} = (a_{k^*}, a_{k^*+1})$ with $k^* = 1, \dots, K-1$ such that $a_{k^*}^+ = a_{k^*+1}^-$. In other words, the two points $a_{k^*}^+$ and $a_{k^*+1}^-$ hit each other, which does not allow us to increase δ.

Consider the two submarkets (μ_1', ν_1') and (μ_2', ν_2') cut from (μ, ν) by the intervals $J_1^* = [a_{k^*}^-, a_{k^*}^+]$ and $J_2^* = [a_{k^*+1}^-, a_{k^*+1}^+]$, i.e.,

$$\mu_i'(A) = \mu(A \cap J_i^*), \qquad \nu_i'(A) = \mu(A \cap J_i^*) \qquad i = 1, 2$$

for any measurable $A \subset \mathbb{R}$. Denote by (μ'', ν'') the residual submarket $(\mu'', \nu'') = (\mu - \mu_1' - \mu_2', \nu - \nu_1' - \nu_2')$.

The submarket (μ_1', ν_1') is an independent anti-diagonal submarket of (μ, ν). To show this, assume without loss of generality that the interval I_{k^*} carries a positive μ-weight. Hence, μ_1' is supported on $[a_{k^*}, a_{k^*}^+]$ and ν_1' on $[a_{k^*}^-, a_{k^*}]$. By the equal-weight condition, $\mu'(\mathbb{R}) = \nu'(\mathbb{R})$ and so (μ_1', ν_1') is an anti-diagonal submarket of (μ, ν). The equal-distance condition implies that any $x \in [a_{k^*}, a_{k^*}^+]$ and $y \in [a_{k^*}^-, a_{k^*}]$ are within δ from each other, while the distances between x and $y' \in \operatorname{supp}(\nu'')$ and between x' and $y \in \operatorname{supp}(\mu'')$ are at least δ. Hence, (μ_1', ν_1') is independent. Similarly, (μ_2', ν_2') is an independent anti-diagonal submarket of (μ, ν). We get assertion 1.

Applying Lemma 9, we conclude that π is a stable matching of (μ, ν) if and only if $\pi = \pi_1' + \pi_2' + \pi''$, where π_i' is the unique stable matching of (μ_i', ν_i'), $i = 1, 2$, and π'' is a stable matching of (μ'', ν''). We obtain assertion 2.

The stable matching for anti-diagonal markets is Monge (Lemma 7). Since μ_1', μ_2', and μ'' are mutually singular (and similarly for ν), we conclude that a stable matching π is Monge if and only if π'' is Monge. Assertion 3 is proved.

The residual submarket (μ'', ν'') is a disjoint market, and $\rho' = \mu' - \nu'$ changes sign at most $K - 1$ times. Indeed, the interval I_{k^*} is removed from both μ and ν, and so the number of imbalance regions is reduced by at least 1. Thus assertion 4 holds.

Finally, μ and ν having piecewise constant densities with at most m intervals of constancy. Without loss of generality, these intervals are common and consecutive, i.e., there is a collection of points $b_0 < b_1 < \ldots < b_m$ such that the densities of μ and ν on an interval $J_i = (b_i, b_{i+1})$ are constants f_i and g_i, respectively. We can also assume that $(b_i)_{i=0,\ldots,m}$ is a subsequence of the sequence $(a_k)_{k=0,\ldots,k+1}$, which defines intervals I_k.

To construct the submarket (μ_1', ν_1') and (μ_2', ν_2'), we need to identify an interval I_{k^*} defined above. This can be done as follows. For each three consecutive intervals I_{k-1}, I_k, I_{k+1}, we look for a triplet of points $\alpha_k^- \in \bar{I}_{k-1}$, $\alpha_k \in \bar{I}_k$ and $\alpha_k^+ \in \bar{I}_{k+1}$ such that

$$\rho([\alpha_k^-, \alpha_k]) = \rho([\alpha_k, \alpha_k^+]) = 0 \qquad \text{and} \qquad |\alpha_k - \alpha_k^-| = |\alpha_k - \alpha_k^+|. \tag{23}$$

If $I_{k-1},\ I_k,\ I_{k+1}$ contain m_k intervals of constancy J_i, then finding a solution $(\alpha_k^-, \alpha_k, \alpha_k^+)$ or checking that no such solution exists can be done in $O(m_k)$ operations. Indeed, to solve the system (23), it is enough to test each of the subintervals of constancy for whether it contains a solution. Indeed, the cumulative distribution function $R_k(x) = \rho([a_{k-1}, x])$ for $x \in I_{k-1} \cup I_k \cup I_{k+1}$ is piecewise linear with m_k intervals of linearity. Testing corresponds to solving a linear system with three unknowns. We conclude that solving (23) boils down to solving of the order of m_k linear systems of given size and thus requires $O(m_k)$ operations. After finding a solution for each triplet of consecutive intervals, we pick $k = k^*$ that minimizes $\delta_k = |\alpha_k - \alpha_k^-|$. The interval I_{k^*} and points $(a_{k^*}^-, a_{k^*}^+, a_{k^*+1}^-, a_{k^*+1}^+) = (\alpha_{k^*}^-, \alpha_{k^*}, \alpha_{k^*}, \alpha_{k^*}^+)$ determine (μ_1', ν_1') and (μ_2', ν_2').

We need $O(m_k)$ operations per each interval I_k, and thus, the total number of operations is of the order $\sum_k m_k$, i.e., of the order of the number of intervals of constancy m. Assertion 5 is proved. $\square$

We are now ready to prove Proposition 3.

Proof of Proposition 3. Consider (μ, ν) and assume that $\rho = \mu - \nu$ changes sign K times. By Lemma 8, any stable matching $\pi = \pi_{\text{diag}} + \pi_1$, where π_{diag} is a unique stable matching of a diagonal market $(\mu - \rho_+, \nu - \rho_-)$ and π_1 is a stable matching of the residual submarket (μ_1, ν_1) with $\mu_1 = \rho_+$ and $\nu_1 = \rho_-$.

The residual submarket (μ_1, ν_1) is disjoint and $\rho_1 = \mu_1 - \nu_1$ changes sign K times since $\rho_1 = \rho$. We construct a sequences of submarkets (μ_k, ν_k) inductively starting from (μ_1, ν_1). Suppose (μ_k, ν_k) is already constructed and let (μ_k', ν_k') to be its independent submarket from Lemma 9. We then define (μ_{k+1}, ν_{k+1}) by

$$\mu_{k+1} = \mu_k - \mu_k' \quad \text{and} \quad \nu_{k+1} = \nu_k - \nu_k'.$$

Let π_k' be a stable matching of (μ_k', ν_k') which is unique since $\rho_k' = \mu_k' - \nu_k'$ changes sign at most two times. Thus any stable matching of (μ_ν) can be represented as

$$\pi = \pi_{\text{diag}} + \pi_1' + \pi_2' + \ldots + \pi_{k-1}' + \pi_k,$$

where π_k is a stable matching of (μ_k, ν_k). By Lemma 9, ρ_k changes sign at most $K-k+1$ times, there is $L \le K$ such that ρ_L changes sign at most one time. Abusing the notation, denote the stable matching of (μ_L, ν_L) by π_L' is unique by Lemma 7.

We conclude that a stable matching π of (μ, ν) is unique and has the following form

$$\pi = \pi_{\text{diag}} + \underbrace{\pi_1' + \pi_2' + \ldots \pi_{L-1}' + \pi_L'}_{L \le K \text{ terms}}.$$

By Lemma 10, $pi_1' + \pi_2' + \ldots \pi_{L-1}' + \pi_L'$ is a Monge matching. Thus π is a convex combination of a diagonal matching and a Monge matching. Consequently, for each x there are at most two distinct y such that (x, y) is in the support of π. Moreover, if there are two such y, one of them necessarily equals x.

We now consider the computational complexity of constructing π for μ and ν having piecewise constant densities with at most m intervals of constancy. Without loss of generality, these intervals are common and consecutive, i.e., there is a collection of points $b_0 < b_1 < \ldots < b_m$ such that the densities of μ and ν on an interval $J_i = (b_i, b_{i+1})$ are constants f_i and g_i, respectively. The collection of all these numbers is the input of the algorithm.

Constructing the diagonal π_{diag} requires a linear number of operations in m. Indeed, $\mu - \rho_+ = \nu - \rho_-$ have density $\min\{f_i, g_i\}$ on J_i and π_{diag} has the corresponding density on the diagonal of $J_i \times J_i$.

The complexity bottleneck corresponds to finding independent submarkets from Lemma 9. By Assertion 5, this requires of the order of m operations for each (μ_k, ν_k). The number of steps k is bounded by the number of times ρ changes its sign, and this number cannot exceed $m - 1$. We conclude that stable matching can be constructed in $O(m^2)$ operations.

D Proof of Theorem 5

We extend the proof techniques used in the two-sided case.

Proof of Assertion 1. Kim and Pass (2014) generalize two-marginal c-cyclic monotonicity condition that we used in the proof of Theorem 1. They consider a k-marginal optimal transportation problem with a bounded continuous cost $c\colon X_1 \times \ldots \times X_k \to \mathbb{R}$ and show that π is optimal if and only if $\mathrm{supp}(\pi)$ is c-cyclically monotone. Here, a set $T \subset X_1 \times \ldots \times X_k$ is c-cyclically monotone if for any collection of k-tuples $(x_1^i, \ldots, x_k^i) \in T$, $i = 1, \ldots, n$, and a collection of k permutations $\sigma_1, \ldots, \sigma_k$ of the set $\{1, \ldots, n\}$, the following inequality holds:

$$\sum_{i=1}^{n} c(x_1^i, \ldots, x_k^i) \le \sum_{i=1}^{n} c(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)}).$$

We now apply this result to $c = c_\alpha$ with

$$c_\alpha(x) = \frac{1 - \exp(\alpha \cdot u(x))}{\alpha}.$$

Let π^* be a solution to the optimal transportation problem with this cost. Take the number of points n equal to the number of marginals k and put $\sigma_i(1) = i$. We obtain that for any any collection $x^i = (x_1^i, \ldots, x_k^i) \in \mathrm{supp}(\pi^*)$, $i = 1, \ldots, k$

$$\begin{aligned}\sum_{i=1}^{k} \exp\left(\alpha \cdot u(x^i)\right) &\ge \sum_{i=1}^{k} \exp\left(\alpha \cdot u(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)})\right)\\ &= \exp\left(\alpha \cdot u(x_1^1, \ldots, x_k^k)\right) + \sum_{i=2}^{k} \exp\left(\alpha \cdot u(x_1^{\sigma_1(i)}, \ldots, x_k^{\sigma_k(i)})\right).\end{aligned}$$

Thus

$$k \cdot \max_i \exp\left(\alpha \cdot u(x^i)\right) \ge \exp\left(\alpha \cdot u(x_1^1, \ldots, x_k^k)\right).$$

Taking logarithm, we get

$$\max_i u(x_i) \ge u(x_1^1, \ldots, x_k^k) + \frac{\ln k}{\alpha}$$

and conclude that π^* is ε-stable with $\varepsilon = \frac{\ln k}{\alpha}$.

Proof of Assertion 2. Without loss of generality, we assume that there is a unit amount of agents on each side of the market, i.e., $\mu_i \in \Delta(X_i)$ for all $i = 1, \ldots, k$.

Let $U^*_{\min}(\mu_1, \ldots, \mu_k)$ be the egalitarian lower bound and π' be a matching such that $u(x) \ge U^*_{\min} - \delta$ where $\delta > 0$ is small. Consider a set

$$C = \left\{x \in X \colon u(x) < U^*_{\min} - \varepsilon\right\}.$$

A solution π^* to the optimal transportation problem with cost c_α maximizes $\int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi$ over all matchings and thus

$$\int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \geq \int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi'(x) \geq -\exp\left(\alpha \cdot (U^*_{\min} - \delta)\right).$$

Since δ is arbitrary, we obtain

$$\begin{aligned}
-\exp\left(\alpha \cdot U^*_{\min}\right) &\leq \int -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\
&= \int_C -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) + \int_{X\setminus C} -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\
&\leq \int_C -\exp(\alpha \cdot u(x))\,\mathrm{d}\pi(x) \\
&\leq -\exp\left(\alpha \cdot (U^*_{\min} - \varepsilon)\right) \cdot \pi(C).
\end{aligned}$$

Thus

$$\alpha \cdot U^*_{\min} \geq \alpha \cdot (U^*_{\min} - \varepsilon) + \ln(\pi(C)),$$

and so $\alpha \cdot \varepsilon \geq \ln(\pi(C))$. Plugging in $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$, we get

$$\pi(C) \leq \min\left\{\exp(-1),\ \frac{1}{|\alpha|}\right\}$$

and thus $\pi(C) \leq \varepsilon$. We conclude that π is ε-egalitarian with $\varepsilon = \frac{\max\{1, \ln|\alpha|\}}{|\alpha|}$.

Proof of Assertion 3. Let π be an ε-stable matching. Since u is continuous, ε-stability implies that, for all collections of points $x^i = (x^i_1, \ldots, x^i_k) \in \mathrm{supp}(\pi)$, $i = 1, \ldots, k$,

$$u(x^1_1, \ldots, x^k_k) \leq \max_{i=1,\ldots,k} u(x^i) + \varepsilon \leq u(x^1) + \ldots + u(x^k) + \varepsilon.$$

Let π' be any other matching. Denote $X = X_1 \times \ldots \times X_k$ and consider a distribution

$$\lambda \in \mathcal{M}_+\big(\underbrace{X \times \ldots \times X}_{k \text{ times}}\big)$$

such that the marginals of λ on each X are equal to π, and the marginal on the set $\{(x^1_1, \ldots, x^k_k) : (x^1, \ldots, x^k) \in X \times \ldots \times X\}$ coincides with π'. We get

$$\begin{aligned}
W(\pi') &= \int_X u(x^1_1, \ldots, x^k_k)\ \mathrm{d}\pi'(x^1_1, \ldots, x^k_k) = \int_{X\times\ldots\times X} u(x^1_1, \ldots, x^k_k)\ \mathrm{d}\lambda(x^1, \ldots, x^k) \\
&\leq \int_{X\times\ldots\times X} \left(u(x^1) + \ldots + u(x^k) + \varepsilon\right)\ \mathrm{d}\lambda(x^1, \ldots, x^k) = \\
&= k \cdot W(\pi) + \varepsilon.
\end{aligned}$$

We thus obtain $W(\pi) \geq \frac{1}{k}\left(W(\pi') - \varepsilon\right)$ for any matching π'. In particular, this inequality holds for π' maximizing welfare. Thus $W(\pi) \geq \frac{1}{k}\left(W^* - \varepsilon\right)$.

We now show that a substantial fraction of agents in an ε-stable matching π have utilities above the egalitarian utility level $U^*_{\min}$. Consider

$$C_\delta = \left\{x \in X \colon u(x) < U^*_{\min} - \varepsilon - \delta\right\},$$

where δ is a small positive number. Our goal is to bound $\pi(C_0)$, but we will bound $\pi(C_\delta)$ with $\delta > 0$ first. Fix $\delta > 0$. As above, let λ be a distribution on $X \times \ldots \times X$ with marginals π on each copy of X and π' on $(x_1^1, \ldots, x_k^k)$. Let π' in this construction be a matching such that $u(x) \geq U^*_{\min} - \delta$ for π'-almost all x. Thus $u(x_1^1, \ldots, x_k^k) \geq U^*_{\min} - \delta$ on a set of full λ-measure. By ε-stability, $u(x_1^1, \ldots, x_k^k) \leq \max_i u(x^i) + \varepsilon$ and thus

$$\max_i u(x^i) + \varepsilon \geq U^*_{\min} - \delta.$$

Towards a contradiction, assume that $\pi(C_\delta) > \frac{1}{k}\pi(X)$. Then

$$\begin{aligned}\lambda(C_\delta \times \ldots \times C_\delta) &\geq \lambda(X \times \ldots \times X) - \sum_i \lambda(\{(x^1, \ldots, x^i) \colon x_i \notin C_\delta\}) \\ &= \pi(X)^{k-1}\left(\pi(X) - k \cdot \pi(C_\delta)\right) > 0.\end{aligned}$$

On the other hand,

$$\max_i u(x_i) + \varepsilon < U^*_{\min} - \delta$$

on $C_\delta \times \ldots \times C_\delta$. This contradiction implies that $\pi(C_\delta) \leq \frac{1}{k}\pi(X)$. The sequence of sets C_δ is decreasing in δ and $C_0 = \cup_{\delta>0} C_\delta$. Hence,

$$\pi(C_0) = \lim_{\delta \to 0} \pi(C_\delta) \leq \frac{1}{k}\pi(X)$$

and thusany ε-stable matching π is ε'-egalitarian with $\varepsilon' = \max\{\frac{1}{k}, \varepsilon\}$.

E Proof of Theorem 4

The existence of a potential without continuity guarantees. Let $\geq$ be the binary relation on $Z = X \times Y$ defined by $(x, y) \geq (x', y')$ if and only if $x = x'$ and $y \succeq_x y'$, or $y = y'$ and $x \succeq_y x'$. Our goal is to show the existence of a potential, i.e., a function $u \colon X \times Y \to \mathbb{R}$ such that for any $(x, y), (x', y')$ with $x = x'$ or $y = y'$,

$$u(x, y) \geq u(x', y') \Longleftrightarrow (x, y) \geq (x', y').$$

We write $>$ for the strict part of $\geq$.

Define the binary relation $\geq^*$ as the transitive closure of $\geq$, which is constructed as follows. Say that $z \geq^* z'$ if there exists a sequence $z^1, \ldots, z^k$ in Z with

$$z = z^1 \geq z^2 \geq \ldots \geq z^{k-1} \geq z^k = z'.$$

The binary relation $\geq^*$ is a preorder, i.e., it is transitive and reflexive but may not be complete. We also define $>^*$ by $z >^* z'$ if $z \geq^* z'$ and at least one of the $\geq$ comparisons in the above sequence is $>$.

A function u represents $\geq^*$ if

$$u(x,y) \geq u(x',y') \Longleftrightarrow (x,y) \geq^* (x',y')$$

for any (x,y) and (x',y'). It is easy to see that any such function is a potential, and thus, it is enough to establish a representation of $\geq^*$.

We construct the representation of $\geq^*$ as follows. Fix two countable dense subsets $\bar{X}$ of X and $\bar{Y}$ of Y and denote $\bar{X} \times \bar{Y}$ by $\bar{Z}$. Consider

$$L_{\geq^*}(z) = \{z' \in \bar{Z} : z \geq^* z'\}$$

and let $r \colon \bar{Z} \to \mathbb{N}$ be an enumeration of $\bar{Z}$. Define a function $u \colon X \times Y \to \mathbb{R}$ by

$$u(z) = \sum_{z' \in L_{\geq^*}(z)} \left(\frac{1}{2}\right)^{r(z')}. \tag{24}$$

To show that u represents $\geq^*$, we need the following technical lemma, which is the key step in the proof.

Lemma 11. *If $(x,y) >^* (x',y')$ then there exists $(\bar{x},\bar{y}) \in \bar{Z}$ with*

$$(x,y) >^* (\bar{x},\bar{y}) >^* (x',y').$$

Proof. By definition of $>^*$, if $(x,y) >^* (x',y')$ then there exists $(\hat{x},\hat{y})$ and $(\hat{x}',\hat{y}')$ with $(x,y) \geq^* (\hat{x},\hat{y}) > (\hat{x}',\hat{y}') \geq^* (x',y')$. We may without loss assume that $\hat{x} = \hat{x}'$ and that $\hat{y} \succ_{\hat{x}} \hat{y}'$.

In fact, by transitivity of the agents' preferences, we may assume that no consecutive terms in a sequence that defines the transitive closure correspond to the same agents' preference. Indeed, if $z^{k-1} \geq z^k \geq z^{k+1}$ and x is the common agent to all three couples, then we would also have $z^{k-1} \geq z^{k+1}$ and so the sequence can be shortened by dropping

z^k. Thus the comparisons in the sequence must then alternate between the common agent being from X and from Y.

So there exists x^1 and x^2 with

$$x^1 \succeq_{\hat{y}} \hat{x}, \quad \hat{y} \succ_{\hat{x}} \hat{y}', \quad \hat{x} \succeq_{\hat{y}'} x^2.$$

By the continuity of agent's preferences with respect to the agent (Property 1), there exists a neighborhood N of $\hat{x}$, so that $\hat{y} \succ_w \hat{y}'$ for all $w \in N$.

Now, by Property 3, $x^1 \succeq_{\hat{y}} \hat{x} \succeq_{\hat{y}'} x^2$ means that there is $\bar{x} \in N$ with $x^1 \succ_{\hat{y}} \bar{x} \succ_{\hat{y}'} x^2$. The condition $\hat{y} \neq \hat{y}'$ needed to apply Property 3 follows from $\hat{y} \succ_{\hat{x}} \hat{y}'$.

The continuity of preferences then implies that there is a neighborhood of $\bar{x}$ which consists of alternatives that are strictly worse for $\hat{y}$ than x^1, and one neighborhood with alternatives that are strictly better for $\hat{y}'$ than x^2. Thus there is a neighborhood N' of $\bar{x}$ with $x^1 \succ_{\hat{y}} w \succ_{\hat{y}'} x^2$ for all $w \in N'$.

Since $\bar{X}$ is dense and $N' \cap N$ is open and non-empty (as it contains $\bar{x}$), there exists $x^* \in \bar{X} \cap N' \cap N$. Thus,

$$x^1 \succ_{\hat{y}} x^*, \quad \hat{y} \succ_{x^*} \hat{y}', \quad x^* \succ_{\hat{y}'} x^2.$$

Denote by $L_{\succeq_{x^*}}(\hat{y}) = \{\tilde{y} \in Y : \hat{y} \succ_{x^*} \tilde{y}\}$ the set of agents in Y that are strictly worse than $\hat{y}$ for x^*, and by $U_{\succeq_{x^*}}(\hat{y}') = \{\tilde{y} \in Y : \tilde{y} \succ_{\hat{x}} \hat{y}'\}$ the set of agents that are strictly better than $\hat{y}'$. These sets are open sets by the continuity of $\succeq_{x^*}$ and non-empty as they contain, respectively, $\hat{y}'$ and $\hat{y}$.

For any $y \in Y$, by completeness of $\succeq_{x^*}$, either $y \succeq_{x^*} \hat{y} \succ_{x^*} \hat{y}'$ and thus $y \in U_{\succeq_{x^*}}(\hat{y}')$, or $\hat{y} \succ_{x^*} y$ and thus $y \in L_{\succeq_{x^*}}(\hat{y})$. Therefore, $Y = U_{\succeq_{x^*}}(\hat{y}') \cup L_{\succeq_{x^*}}(\hat{y})$. Then $U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y}) \neq \emptyset$, as Y is connected and therefore cannot be the union of disjoint, non-empty, open sets.

Since $\bar{Y}$ is dense in Y, and $U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y})$ is open, there exists $y^* \in \bar{Y} \cap U_{\succeq_{x^*}}(\hat{y}') \cap L_{\succeq_{x^*}}(\hat{y})$. So we obtain $(x^*, y^*) \in \bar{X} \times \bar{Y}$ for which

$$(x, y) \geq^* (x^1, \hat{y}) \geq (x^*, \hat{y}) > (x^*, y^*) > (x^*, \hat{y}') \geq (x^2, \hat{y}') \geq^* (x', y'),$$

completing the proof of the lemma. □

Lemma 11 shows that $\geq^*$ admits an order-dense countable subset $\bar{Z}$ (using standard terminology). Once this property is established, one can use Theorem 3.1 and Lemma 2.3 from Voorneveld and Norde (1997) to deduce the existence of a potential u (without continuity guarantees). We include a short proof not relying on the results by Voorneveld and Norde (1997) for the reader's convenience.

Lemma 12. *Function u given by (24) represents $\geq^*$.*

Proof. If $z \geq^* z'$ then $L_{\geq^*}(z') \subseteq L_{\geq^*}(z)$, and so $u(z) \geq u(z')$. And if $z >^* z'$ then, by Lemma 11, there exists $\bar{z} \in \bar{Z}$ with $z >^* \bar{z} >^* z'$. This means that $\bar{z} \in L_{\geq^*}(z)$ while $\bar{z} \notin L_{\geq^*}(z)$ because, by acyclicity, $\bar{z} >^* z'$ implies that $z' \geq^* \bar{z}$ cannot hold. Thus $L_{\geq^*}(z') \subsetneq L_{\geq^*}(z)$, and so $u(z) > u(z')$.

Finally, by definition of $\geq^*$, if $y \succeq_x y'$ then $(x,y) \geq^* (x,y')$ and thus $u(x,y) \geq u(x,y')$. Similarly, if $y \succ_x y'$ then $(x,y) >^* (x,y')$ and thus $u(x,y) > u(x,y')$. For $\succeq_y$, the argument is analogous and thus omitted. $\square$

The existence of an upper semicontinuous potential. Here, we prove the second statement of Theorem 4 establishing the existence of an upper semicontinuous potential.

As before, we consider a countable dense set $\bar{Z} \subseteq X \times Y$ and its enumeration $r \colon \bar{Z} \to \mathbb{N}$. We define u by

$$u(z) = (-1) \cdot \sum_{z' \in \bar{Z} \colon z' > z} \left(\frac{1}{2}\right)^{r(z')}, \tag{25}$$

and $u(z) = 0$ if there is no $z' > z$. Note that, in contrast to (24), this formula uses on $>$ but not $>^*$. We will need several technical lemmas to demonstrate that u represents $\geq^*$ and is upper semicontinuous.

Lemma 13. *If $(x_1, y) \geq (x, y) > (x, y') \geq (x_2, y')$, then there exists $z \in X$ such that $(x_1, y) > (z, y) > (z, y') > (x_2, y')$.*

Proof. First, $(x,y) > (x,y')$ and, hence, Property 1 implies that there is a neighborhood N_x of x in X such that $(z,y) > (z,y')$ for all $z \in N_x$. Second, $(x_1,y) \geq (x,y)$ and $(x,y') \geq (x_2,y')$ combined with Property 3 imply that there exists $z \in N_x$ with $(x_1,y) > (z,y)$ and $(z,y') > (x_2,y')$. Note that $y \neq y'$ follows from $(x,y) > (x,y')$, and so Property 3 applies. $\square$

Lemma 14. *If $(x,y) >^* (x',y')$ then there are neighborhoods N_x, $N_{x'}$, N_y and $N_{y'}$ of, respectively, x in X, x' in X, y in Y, and y' in Y such that $(\tilde{x},\tilde{y}) > (\tilde{x}',\tilde{y}')$ for all $(\tilde{x},\tilde{x}',\tilde{y},\tilde{y}') \in N_x \times N_{x'} \times N_y \times N_{y'}$.*

Proof. By definition of $>^*$, we may without loss assume that there is a sequence (x_i, y_i),

$i = 1, \ldots, k$, with

$$\begin{aligned}(x, y) = (x_1, y_1) \geq (x_2, y_1) \geq \ldots \\ \ldots \geq (x_{t-1}, y_{t-1}) \geq (x_t, y_{t-1}) > (x_t, y_t) \geq (x_{t+1}, y_t) \geq \ldots \\ \ldots \geq (x_{k-1}, y_{k-1}) \geq (x_k, y_k) = (x', y').\end{aligned} \tag{26}$$

By repeatedly applying Lemma 13, we may replace all $\geq$ in (26) with $>$. Then, for each t, we have

$$(x_{t-1}, y_{t-1}) > (x_t, y_{t-1}) > (x_t, y_t) > (x_{t+1}, y_t).$$

By Property 2, there are neighborhoods N_z^1, N_z^2, and N_z^3 of $z \in \{x_{t-1}, x_t, x_{t+1}, y_{t-1}, y_t\}$ with the property that $(\tilde{x}, \tilde{y}) > (\tilde{x}', \tilde{y})$ for all $(\tilde{x}, \tilde{x}', \tilde{y}) \in N^3_{x_{t-1}} \times N^1_{x_t} \times N^2_{y_{t-1}}$ and all $(\tilde{x}, \tilde{x}', \tilde{y}) \in N^3_{x_t} \times N^1_{x_{t+1}} \times N^2_{y_t}$, while $(\tilde{x}, \tilde{y}) > (\tilde{x}, \tilde{y}')$ for all $(\tilde{x}, \tilde{y}, \tilde{y}') \in N^2_{x_t} \times N^3_{y_{t-1}} \times N^1_{y_t}$.

Denote $N_z = N_z^1 \cap N_z^2 \cap N_y^3$ for each $z \in \cup_{t=1}^k \{x_t, y_t\}$. Then (26) holds for any selection of $\tilde{x}_t \in N_{x_t}$ and $\tilde{y}_t \in N_{y_t}$, $t = 1, \ldots, k$, completing the proof. $\square$

The second statement of Theorem 4 is the content of the following lemma.

Lemma 15. *A function u given by (25) represents $\geq^*$ and is upper semicontinuous.*

Proof. The argument that u represents $\geq^*$ repeats that in Lemma 12 and thus is omitted. We prove that u is upper semicontinuous.

Consider a sequence $z_k = (x_k, y_k)$ converging to $(x, y) = z$. Let $V_k = \{\tilde{z} \in X \times Y : \tilde{z} > z_k\}$ and $V = \{\tilde{z} \in X \times Y : \tilde{z} > z\}$. First note that $V \subseteq \cup_n \cap_{k \geq n} V_k$. Indeed, by Lemma 14, for $z' \in V$, there is a neighborhood N_z of z for which $z' > \tilde{z}$ for all $\tilde{z} \in N_z$. Hence, for k large enough, we have $z' > z_k$ since $z_k \to z$. Thus $z' \in \cap_{k \geq n} V_k$ for n large enough. We obtain

$$u(z) = (-1) \cdot \sum_{z \in \bar{Z} \cap V} \left(\frac{1}{2}\right)^{r(z)} \geq (-1) \cdot \sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)}.$$

On the other hand,

$$u(z_n) = (-1) \cdot \sum_{z \in \bar{Z} \cap V_n} \left(\frac{1}{2}\right)^{r(z)} \leq (-1) \cdot \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)}.$$

Observe that

$$\sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)} - \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} \left(\frac{1}{2}\right)^{r(z)} \leq 2 \cdot \left(\frac{1}{2}\right)^{r_n^*},$$

where

$$r_n^* = \min\left\{r(z) \ : \ z \in \bar{Z} \cap (\cup_m \cap_{k \geq m} V_k) \setminus \bar{Z} \cap (\cap_{k\geq n} V_k)\right\}$$

under the convention $\min \emptyset = +\infty$. For any decreasing sequence of sets $C_n \subseteq \mathbb{N}$ such that $\cap C_n = \emptyset$, the minimum $\min C_n$ converges to $+\infty$ as n gets large. Therefore, $r_n^* \to +\infty$ as $n \to \infty$ because r takes values in $\mathbb{N}$ and is injective. We obtain that

$$\begin{aligned}
u(z) &\geq \sum_{z \in \bar{Z} \cap (\cup_n \cap_{k \geq n} V_k)} (-1) \left(\frac{1}{2}\right)^{r(z)} \\
&= \lim_{n\to\infty} \sum_{z \in \bar{Z} \cap (\cap_{k \geq n} V_k)} (-1) \left(\frac{1}{2}\right)^{r(z)} \\
&\geq \limsup_{n\to\infty} u(z_n),
\end{aligned}$$

and thus u is upper semicontinuous. $\square$