Composite Sorting*

Job Boerma
University of Wisconsin-Madison

Aleh Tsyvinski
Yale University

Ruodu Wang
University of Waterloo

Zhenyuan Zhang
Stanford University

August 2023

Abstract

We propose a new sorting framework: composite sorting. Composite sorting comprises of (1) distinct worker types assigned to the same occupation, and (2) a given worker type simultaneously being part of both positive and negative sorting. Composite sorting arises when fixed investments mitigate variable costs of mismatch. We completely characterize optimal sorting and additionally show it is more positive when mismatch costs are less concave. We then characterize equilibrium wages. Wages have a regional hierarchical structure − relative wages depend solely on sorting within skill groups. Quantitatively, composite sorting can generate a sizable portion of within-occupations wage dispersion in the US.

JEL-Codes: J01, D31, C78
Keywords: Composite Sorting, Assignment

*We thank Manuel Amador, Carter Braxton, Hector Chade, Philipp Kircher, Moritz Kuhn, Rasmus Lentz, Ilse Lindenlaub, Jeremy Lise, Paolo Martellini, Guido Menzio, Guiseppe Moscarini, Chris Moser, Andrea Ottolini, Luigi Pistaferri, Tommaso Porzio, Fabien Postel-Vinay, Stefan Steinerberger, Ethan Yang, and Alexander Zimin for comments.

1 Introduction

Sorting models have a prominent position in economics (Koopmans and Beckmann, 1957; Becker, 1973). An important insight of this literature is that when the output function is supermodular or submodular, sorting is positive or negative, implying that very similar or identical worker types work in the same occupation. Sattinger (1993), Chiappori and Salanié (2016), Chade, Eeckhout, and Smith (2017), and Eeckhout (2018) provide comprehensive reviews of this literature.[1]

Our paper provides the first complete characterization of optimal sorting and wages in a sorting model with an output function that is neither supermodular nor submodular. We establish that equilibrium sorting is significantly richer than in the canonical settings, yet we characterize it fully. We call the sorting pattern that emerges composite sorting. Composite sorting has two main features. First, distinct workers are sorted to the same occupation. This enables us to study wage dispersion within occupations. Second, a given worker type can simultaneously be in both positive and negative sorting. We theoretically characterize optimal sorting, its comparative statics, and a dual solution. We then present an example that captures defining features of the theoretical characterization and quantitatively demonstrate our model using American Community Survey data. The example generates 29 percent of overall wage dispersion and about 50 percent of wage dispersion at the top and the bottom of the wage distribution. In contrast, with either supermodular or submodular costs, the solution is positive or negative one-to-one sorting with no dispersion in wages.

We study a sorting model with heterogeneous workers, heterogeneous jobs, and mismatch. Mismatch is the difference between the skill of the worker and the difficulty of the job. When the job's difficulty exceeds the worker's skill level, that is, when a worker is underqualified, mismatch decreases output. The second type of mismatch (as in Lise and Postel-Vinay (2020)) arises when the worker's ability surpasses the job's demands, that is, when the worker is overqualified, resulting in a utility loss for a worker and a corresponding loss in the joint surplus. Following Stigler (1939) and Laffont and Tirole (1986, 1991), firms can incur fixed costs to reduce the variable costs of mismatch. Firms mitigate adverse effects of mismatch by investing in technologies or providing amenities. An example of such mitigation is investing in a cobot (collaborative robot) that enhances the output of a low-skill welder. The result of the technology decision is that the cost of mismatch is concave in the extent of underqualification. Similarly, employers may invest in amenities to decrease the disutility of workers (as in Rosen (1986)) when they are overqualified. An example of such mitigation is investing in a premium truck with advanced comfort and drivers support

[1] Antràs and Rossi-Hansberg (2009) and Costinot and Vogel (2015) provide an overview of the significance of assignment models in international trade.

Figure 1: Pairings Between Workers and Jobs Do Not Intersect

Figure 1 illustrates that for an optimal sorting, pairings between workers and jobs do not intersect. Positive sorting, in the left panel, pairs a low-skill worker and a high-skill job (x_L, z_H), and a medium-skill worker with an exceptional skill job (x_M, z_E), resulting in two medium-size mismatches. For concave costs, instead, it is preferable to have one large mismatch (x_L, z_E) and one small mismatch (x_M, z_H). It is optimal to have pairs that do not intersect as in the right panel.

system for a high-skill trucker. The result of investment in amenities is that the cost of mismatch is concave in the extent of overqualification. In sum, the cost of mismatch is concave in mismatch. The key characteristic of a generic concave function in mismatch is that it is neither supermodular nor submodular. We are able to provide a complete characterization of this sorting problem with neither supermodular nor submodular output by analyzing it using the tools of optimal transport theory.

We now describe the main features of an optimal assignment. First, an optimal assignment maximizes the number of perfect pairs, which are pairs without mismatch. Since mismatch costs are concave, it is preferable to have a combination of one pair with small mismatch and another with significant mismatch, as opposed to having two pairs with moderate mismatch. A combination of a perfect pair and a pair with significant mismatch exemplifies this.

The second feature of an optimal assignment is that pairings between workers and jobs do not intersect. Consider workers with low and medium skills $\{x_L, x_M\}$ and jobs that require high and exceptional skills $\{z_H, z_E\}$. Positive sorting assigns the low-skill worker to the high-skill job (x_L, z_H), and the medium-skill worker to the exceptional job (x_M, z_E). This results in two medium-size mismatch costs. Visualizing the pairing as arcs connecting a worker with a job reveals that the pairs intersect as in the left panel of Figure 1. For concave costs, in contrast, it is preferable to have one large mismatch (between a low-skill worker and a very high-skill job (x_L, z_E)) and one small mismatch (between a medium-skill worker and a high-skill job (x_M, z_H)). Thus, it is optimal to have pairs that do not intersect as in the right panel.

The feature of no intersecting pairs gives rise to two additional properties of optimal sorting. First, the assignment problem can be decomposed into independent problems by layers. Layers are constructed by first designing a measure of underqualification that evaluates the cumulative deficiency of worker skills compared to the cumulative demands of jobs at a given skill level. A layer contains all the workers and jobs in a particular slice of this measure of underqualification. An assignment in each layer optimally pairs workers and jobs within a given slice (given the same mismatch cost function), independent of all

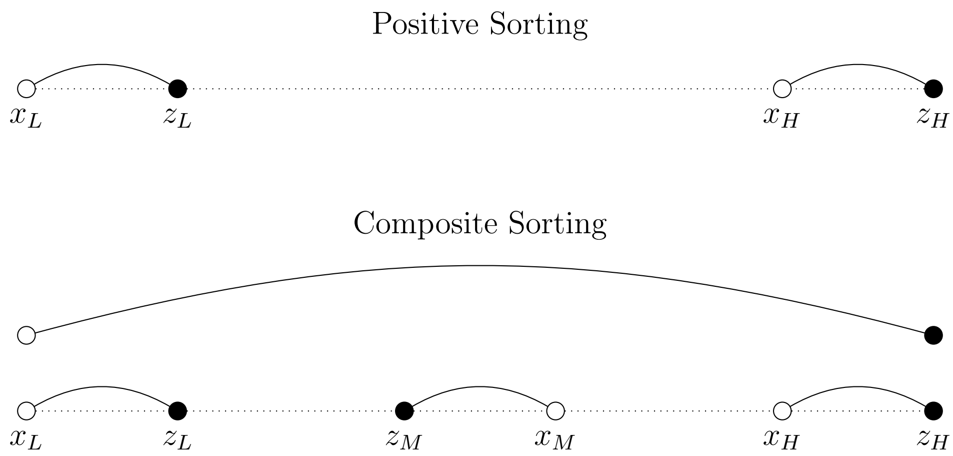


Figure 2: An Example of Composite Sorting

Figure 2 explains composite sorting with four workers (circles) and four jobs (dots). In the top panel, workers are optimally assigned to jobs in their skill group which results in two positively sorted pairs: (x_L, z_L) and (x_H, z_H). Adding an identical low-skill worker and an identical high-skill job does not allow to reduce the initial mismatch and results in a pair (x_L, z_H). The medium worker-job group $\{x_M, z_M\}$ can be used to rematch the low-skill worker and the high-skill job with the medium-skill job and worker forming pairs (x_L, z_M) and (x_M, z_H), but such reshuffling is not optimal due to the concavity of the costs. In the bottom panel, we hence also have two negatively sorted pairs (x_L, z_H) and (x_M, z_M) which delivers composite sorting (observe that worker type x_L is assigned to distant occupations; not to z_M).

workers and jobs in other layers. The full optimal assignment sums the optimal assignments of each layer. Second, we characterize an optimal assignment within a given layer through a Bellman equation.

The main result of our paper is composite sorting − the optimal assignment sorts identical workers into different jobs, some positively and some negatively. The intuition for composite sorting can be described through an example. Suppose a low worker-job group $\{x_L, z_L\}$ is far from a high worker-job group $\{x_H, z_H\}$ as in the top panel of Figure 2. When there is a significant gap between the groups, workers are optimally assigned to jobs within their skill groups resulting in two pairs (x_L, z_L) and (x_H, z_H). This results in two small mismatches and positive sorting. Suppose one identical low-skill worker x_L and one identical high-skill job z_H as well a medium-skill worker-job group $\{x_M, z_M\}$ are added as in the bottom panel of Figure 2. Adding an identical low-skill worker and an identical high-skill job does not allow to reduce the initial mismatched pairs (x_L, z_L) and (x_H, z_H), and so the added low-skill worker is assigned to the added high-skill job forming a pair (x_L, z_H). The medium worker-job group $\{x_M, z_M\}$ can, in principle, be used to rematch the added low-skill worker and the added high-skill job with the medium-skill job and worker forming pairs (x_L, z_M) and (x_M, z_H). However, such reshuffling is not optimal, because due to concavity of the costs, two medium mismatches (x_L, z_M) and (x_M, z_H) are worse than one small mismatch within the medium-skill group (x_M, z_M) and one large mismatch for (x_L, z_H). As a result, the medium-skill group is paired together and this results in the negative sorting within the added group.

We emphasize that composite sorting assigns the same worker type to very different occupations, with some positive and some negative sorting. In the example, the low-skill worker x_L is paired positively with the low-skill occupation z_L (as a part of the sorting (x_L, z_L) and (x_H, z_H)) but the same worker type is also paired negatively to the distant high-skill occupation z_H (as a part of the sorting (x_L, z_H) and (x_M, z_M)). Thus, the optimal full assignment is not one-to-one between types because the same worker type is assigned to different occupations as a part of simultaneously positive and negative sorting.

For general distributions, composite sorting may be very rich with very distinct worker types being assigned to the same occupation and may exhibit various local and global intervals of positive and negative sorting. We show how to completely characterize optimal sorting for general distributions.

We next analyze how these rich sorting patterns vary with changes in technology. In comparative statics analysis, we derive that sorting becomes more positive, by which we mean larger in concordance order, as the cost of mismatch becomes less concave. With a fixed pair of worker and job distributions, the assignment problem is split into a fixed set of layers irrespective of the cost function, and we show that for each layer sorting becomes more positive as the cost of mismatch becomes less concave. Moreover, we show that there exists a threshold of concavity for the cost function above which the optimal sorting within each layer is positive.

We determine equilibrium wages and firm values by characterizing the solution to the dual planning problem. We show that wages and job values exhibit a regional hierarchical structure. For a given skill group, relative wages are determined regionally; that is, they depend only on information within the regional skill group and do not depend on any other groups. The hierarchical structure aggregates the regional relative wages to wages for larger groups preserving the relative wages implied by the smaller regions.

An important implication of our results is frictionless wage dispersion within occupations. We develop an example that captures essential features of the complete characterization of composite sorting, and we investigate to what extent it can generate qualitative and quantitative patterns of wage dispersion in the American Community Survey. We find that the model delivers sizable wage dispersion within occupations with high and low mean wages. Composite sorting can generate 54 percent of wage dispersion within jobs at the bottom of the distribution, and 59 percent at the top. The model can generate 10 percent of wage dispersion in the middle of the distribution. The main reason is that the logarithmic wage profile is relatively flat in this range and, even though there may be substantial skill dispersion within a particular occupation, it does not give rise to substantial wage dispersion. Overall, composite sorting accounts for 29 percent of the wage dispersion in the sample. In contrast, one-to-one sorting models (such as models with positive or negative sorting arising from submodular or supermodular costs) would deliver no wage dispersion within occupations.

Literature. We briefly discuss additional relevant literature. Kremer and Maskin (1996) and Porzio (2017) study sorting of heterogeneous workers with a form of technology choice where there is selection in managerial and worker roles. Anderson and Smith (2023) depart from focusing on the conditions for assortative sorting by developing comparative statics for assortative sorting without having to solve for the optimal assignment. Chiappori, Fiorio, Galichon, and Verzillo (2022) and Fagereng, Guiso, and Pistaferri (2022) analyze non-assortative sorting patterns in income and wealth. Another recent development in the assignment literature is to study models with multiple agents combined together (Kremer, 1993; Chiappori, McCann, and Pass, 2017; Chade and Eeckhout, 2018; Eeckhout and Kircher, 2018; Boerma, Tsyvinski, and Zimin, 2021).

A general alternative that results in imperfect assortative sorting is the search and matching literature (for example, Shimer and Smith (2000), Postel-Vinay and Robin (2002), Cahuc, Postel-Vinay, and Robin (2006), Eeckhout and Kircher (2010), Lise and Robin (2017), Bagger and Lentz (2019)). Specifically, this approach generates wage dispersion within occupations due to the search frictions. Our work instead generates wage dispersion in a frictionless environment.

Our paper uses results from the optimal transport literature (see Galichon (2018) for a comprehensive overview of applications of optimal transport theory to economic problems). In the optimal transport literature the idea of perfect pairs is referred to as "mass stays in place if it can" (Gangbo and McCann, 1996; Villani, 2003). Villani (2009) states that the non-intersecting rule first appears in Monge (1781). This non-intersecting property is central to the literature on optimal transport with concave distance costs as well as to algorithmic sorting problems with distance costs (Werman, Peleg, Melter, and Kong, 1986; Aggarwal, Barnoy, Khuller, Kravets, and Schieber, 1995). Aggarwal, Barnoy, Khuller, Kravets, and Schieber (1995) proposed the first combinatorial algorithm that can be used to solve for an optimal assignment within a layer when the cost of mismatch is linear. The Bellman equation in our papers adopts a recursive algorithm developed by Nechaev, Sobolevski, and Valba (2013), designed to model statistical properties of polymer chains. The idea of layering was introduced in Delon, Salomon, and Sobolevski (2012a). Their work also develops a more computationally efficient version of the Bellman equation building on Aggarwal, Barnoy, Khuller, Kravets, and Schieber (1995). We provide a new, concise proof of this more efficient Bellman equation that also extends to case of asymmetric mismatch cost functions. Our comparative statics results for the primal problem and a complete characterization of the dual solution and its regional hierarchical structure are also a contribution to the optimal transport theory and, more specifically, to the literature with concave distance costs started by Gangbo and McCann (1996) and McCann (1999).

2 Model

We study an environment in which workers with heterogeneous skills sort into heterogeneous jobs, and where mismatch between their skills and the job difficulty leads to output losses. Firms mitigate the extent to which mismatch penalizes surplus by reducing variable mismatch costs with fixed cost investments.

2.1 Environment

The economy is populated by risk-neutral workers and jobs. The workers differ in their skills which are indexed by x. The set of worker skills X contains a finite number n of types $x_1 < x_2 < \cdots < x_n$. Workers are distributed according to the cumulative distribution function $F(x)$.

Jobs differ in their difficulty which is indexed by z. The set of occupations Z contains a finite number m of occupation types $z_1 < z_2 < \cdots < z_m$. Jobs are distributed according to the cumulative distribution function $G(z)$.

Surplus. Firms produce a single good. Production requires one worker for each job. Since both workers and firms are risk-neutral, what matters is the surplus generated by a worker-job pair (x, z). The surplus generated by a worker with skill x in an occupation with difficulty level z is:

$$s(x, z) = g(x) + h(z) - \gamma_p \max(z - x, 0) - \gamma_u \max(x - z, 0), \tag{1}$$

with $\gamma_p, \gamma_u \geqslant 0$. There are four terms in this technology specification. The first term $g(x)$ with $g'(x) > 0$ reflects that a more skilled worker contributes more to production, independent of the job. The second term $h(z)$ where $h'(z) > 0$ reflects that a more difficult job z produces more output and is thus more valuable, independent of the worker that fulfills the job. The third term $\gamma_p \max(z - x, 0)$ captures the idea that a worker with a skill x that is lower than the job demand z causes a loss of output. It is costly to have workers perform tasks for which they have limited talent. The fourth term $\gamma_u \max(x - z, 0)$ captures the idea that a worker with skills x that exceed the job demands z is overqualified and needs to be compensated for their utility cost (as in Rosen (1986) and Lise and Postel-Vinay (2020)). Mismatch is the distance between worker skill x and job difficulty z.

Technology Choice. We assumed so far that mismatch costs of a worker-job pair (x, z) are exogenous. We next endogenize the mismatch costs. A firm can reduce the costs of mismatch by making a fixed investment. In case a worker is underqualified, the firm can make technology investments to reduce the production costs of mismatch. When workers are overqualified, the firm can provide amenities to reduce the utility costs of mismatch. The main insight of Stigler (1939) and Laffont and Tirole (1986, 1991) is that this technology choice results in an effective output function with a concave cost of mismatch.

We model technology choice as a firm making fixed investments (for example, by purchasing better equipment) to reduce variable costs associated with mistakes and delays caused by a worker that is underqualified to fulfill the job, $x < z$. Specifically, firms choose the variable cost of production mismatch γ_p, which comes at an associated fixed cost $\Psi_p(\gamma_p) = \frac{1}{\eta_p}\gamma_p^{-\eta_p}$, where η_p is strictly positive.[2] By decreasing variable costs γ_p, the firm increases its fixed costs, or $\Psi_p' < 0$, where $\Psi_p'' > 0$. The effective output of worker x in occupation $z > x$ is then:

$$y(x, z) = \max_{\gamma_p \geqslant 0} \left(g(x) + h(z) - \gamma_p(z - x) - \frac{1}{\eta_p}\gamma_p^{-\eta_p} \right). \tag{2}$$

Investment increases in the distance between the worker skill and the job difficulty $-$ the technology choice is $\gamma_p = (z - x)^{-\frac{1}{1+\eta_p}}$. Firms choose a low variable cost of production mismatch if the worker is less qualified, that is, when the distance between skills and demands $(z - x)$ is large.

The effective output of worker x in job z is, using the technology investment decision, given by:

$$y(x, z) = g(x) + h(z) - \frac{1}{\zeta_p}(z - x)^{\zeta_p} \tag{3}$$

for underqualified workers $x < z$, with $\zeta_p = \frac{\eta_p}{1+\eta_p} \in (0, 1)$. The distinctive feature of this function is that the mismatch cost is concave in the distance between worker skill and job difficulty. In sum, the technology choice transforms a production function with linear mismatch costs into an effective output function with strictly concave mismatch costs.

Firms can similarly reduce the extent to which overqualificaton penalizes worker utility by providing amenities. We model amenity choices analogous to technology choices.[3] When workers are overqualified, $x > z$, their employers want to invest in amenities to lower the disutility from working. Firms choose the

[2] General convex cost functions are considered in the Technical Appendix.

[3] Alternative frameworks that incorporate the decision to provide amenities are presented in Hwang, Mortensen, and Reed (1998), Lang and Majumdar (2004), and Morchio and Moser (2021).

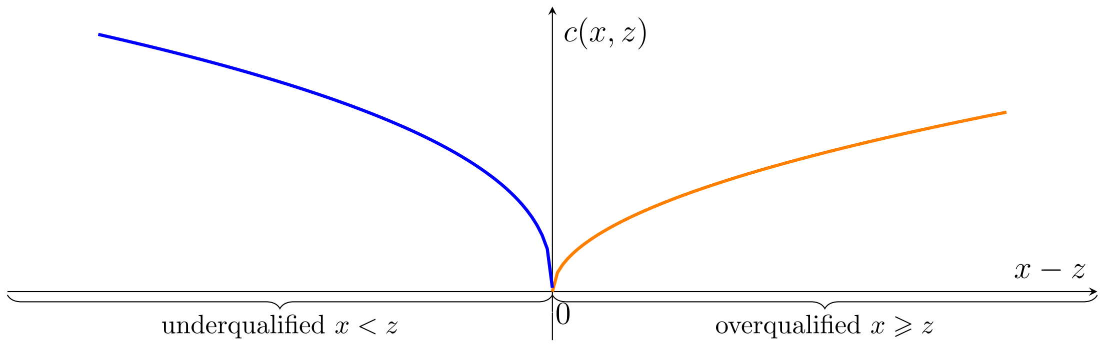


Figure 3: The Cost of Mismatch

Figure 3 illustrates the cost of mismatch $c(x, z)$, which is concave in the discrepancy between the worker skill x and the job difficulty z. The result of the firm technology decision is that the mismatch cost function is concave in the extent of underqualification $(z - x)$. The result of investment in amenities is that the mismatch cost function is concave in the extent of overqualification $(x - z)$.

variable utility cost due to mismatch γ_u at an associate fixed cost $\Psi_u(\gamma_u) = \frac{1}{\eta_u}\gamma_u^{-\eta_u}$, where η_u is strictly positive. The provision of amenities is thus characterized by $\gamma_u = (x - z)^{-\frac{1}{1+\eta_u}}$. Amenities increase with worker skills. As a result, the effective output of overqualified workers is $y(x, z) = g(x)+h(z)-\frac{1}{\zeta_u}(x-z)^{\zeta_u}$, with $\zeta_u = \frac{\eta_u}{1+\eta_u} \in (0, 1)$.

Effective Output. We write effective output as:

$$y(x, z) = g(x) + h(z) - \begin{cases} \frac{1}{\zeta_p}(z - x)^{\zeta_p} & \text{if } z \geqslant x \\ \frac{1}{\zeta_u}(x - z)^{\zeta_u} & \text{if } z < x, \end{cases} \tag{4}$$

where $\zeta_p, \zeta_u \in (0, 1)$. We use effective output (4) to define the cost of mismatch between worker x and job z as:

$$c(x, z) = g(x) + h(z) - y(x, z) = \begin{cases} \frac{1}{\zeta_p}(z - x)^{\zeta_p} & \text{if } z \geqslant x \\ \frac{1}{\zeta_u}(x - z)^{\zeta_u} & \text{if } z < x, \end{cases} \tag{5}$$

that is, maximal output of worker x and job z minus effective output $y(x, z)$. Thus, the mismatch cost function is concave in the discrepancy between worker x and job z. We plot an example of the mismatch cost function in Figure 3.

The output function is neither supermodular nor submodular. First, the cross-derivatives of the output function being negative for both $z > x$ and $z < x$ directly rules out supermodularity. Second, the output function is not submodular. Consider two workers and two jobs, each with skills a and b where $b \neq a$. Submodularity requires the combined output of pairs (a, b) and (b, a) to be greater than the combined output of pairs (a, a) and (b, b). However, pairs (a, b) and (b, a) have mismatch and consequently lower output than the positive sorting (a,a) and (b,b) which gives zero mismatch. Thus, the output function is not submodular either.

Assignment. An assignment pairs workers and jobs. Given a worker distribution F and a job distribution G, the set of feasible assignment functions is $\Pi := \Pi(F,G)$ which is the set of probability measures on the product space $X \times Z$ such that the marginal distributions of π onto X and Z are F and G. For an assignment π, we denote the support of this assignment by $\Gamma_\pi = \{(x,z) : \pi(\{(x,z)\}) > 0\} \subseteq \mathbb{R}^2$. Feasibility of an assignment is equivalent to labor market clearing; that is, all workers and jobs are sorted.

2.2 Planning Problem

We solve two planning problems to characterize an equilibrium.[4] We first solve a primal planning problem to characterize an equilibrium assignment. The primal planning problem is to choose an assignment to maximize aggregate output:

$$\max_{\pi \in \Pi} \int y(x,z)\, \mathrm{d}\pi \tag{6}$$

and is equivalent, in terms of choosing an optimal assignment, to a planning problem that minimizes the costs of mismatch:

$$\min_{\pi \in \Pi} \int c(x,z)\, \mathrm{d}\pi, \tag{7}$$

where $c(x,z)$ is the cost of mismatch (5). This is an optimal transport problem in the form of Kantorovich (1942) where the cost function is neither supermodular nor submodular.

Dual Problem. To obtain equilibrium wages w and firm value function v, we solve a dual problem. The dual problem is to choose functions w and v that solve:

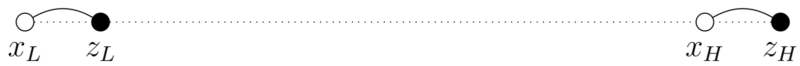
$$\min \int w(x)\, \mathrm{d}F + \int v(z)\, \mathrm{d}G, \tag{8}$$

subject to the constraint $w(x) + v(z) \geqslant y(x,z)$ for any $(x,z) \in X \times Z$. The Monge-Kantorovich duality states that the values from (6) and (8) are the same, or $\max \int y(x,z)\, \mathrm{d}\pi = \min \int w(x)\, \mathrm{d}F + \int v(z)\, \mathrm{d}G$. In Appendix A.2, we prove the following relation between the primal and dual solutions.

Lemma 1. Suppose that assignment $\pi \in \Pi$ and functions w and v are such that $w(x) + v(z) \geqslant y(x,z)$ for any $(x,z) \in X \times Z$ and that $w(x) + v(z) = y(x,z)$ for any $(x,z) \in \Gamma_\pi$. Then assignment π is an optimal assignment and (w,v) is an optimal dual pair.

[4] The equilibrium definition is standard and is presented in Appendix A.1 for completeness.

Positive Sorting



Negative Sorting

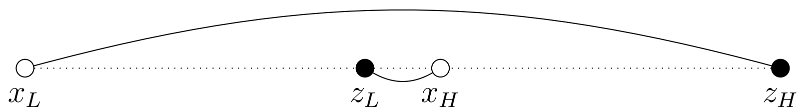


Figure 4: Assortative Sorting

Figure 4 shows assortative sorting for an example with two workers (indicated by the circles) and two jobs (indicated by the dots). The top panel illustrates the case of positive sorting, and the bottom panel illustrates the case of negative sorting.

3 Optimal Sorting

In this section, we characterize optimal sorting.

3.1 Intuition

Before providing the general characterization of an optimal assignment, we deliver the intuition behind composite sorting.

Assortative Sorting. Our sorting problem, which features neither a supermodular nor a submodular output function, can deliver both positive and negative sorting. Importantly, the sorting pattern, rather than being dictated by the shape of the production function alone as in the classic assignment problems, depends on the distributions of workers and jobs.

To show that an optimal assignment may feature both positive and negative sorting, we consider a problem with two workers and two jobs. Worker skills are given by x_L and x_H and job difficulties are given by z_L and z_H satisfying $x_L < z_L < x_H < z_H$. Let the distance between x_i and z_j be $d_{ij} := |x_i - z_j|$.

To explain the presence of positive sorting, consider the following configuration of distances: $d_{LL}^{\zeta} + d_{HH}^{\zeta} \leqslant d_{LH}^{\zeta} + d_{HL}^{\zeta}$ with the specific cost function $c(x, z) = |x - z|^{\zeta}$ with $\zeta \in (0, 1)$. This case is presented in the top panel of Figure 4, where workers are represented by circles and jobs are represented by dots. The low-skill worker x_L and the low-skill job z_L as well as the high-skill worker x_H and the high-skill job z_H are very close to each other, while the skill gap between the low-skill job z_L and high-skill worker x_H is large. It is natural to pair the low-skill worker to the low-skill job and to pair the high-skill worker to the high-skill job to minimize the cost of mismatch and, hence, the optimal assignment is positive.

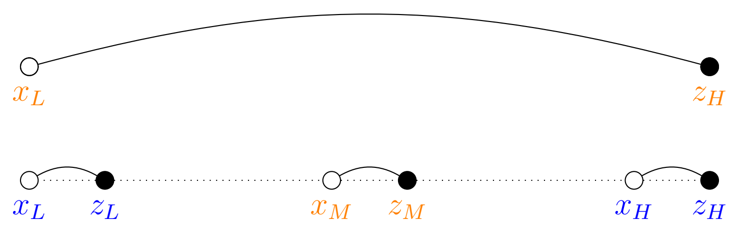


Figure 5: Composite Sorting

Figure 5 illustrates an example with four workers and four jobs that features composite sorting. First, distinct worker types are paired to identical jobs. A low-skill worker x_L and a high-skill worker x_H both work in the identical high-skill occupation z_H, while the medium-skill worker x_M does not work in this occupation. Second, a worker type is simultaneously part of both positive and negative sorting. A low-skill worker x_L is paired positively with a low-skill job z_L (part of positive sorting (x_L, z_L) and (x_H, z_H) in blue) but the same worker type is also paired negatively to a distant high-skill job z_H (part of negative sorting (x_L, z_H) and (x_M, z_M) in orange).

When worker-job skill groups are far apart, it is optimal to sort within groups. The optimal assignment is visualized using an arc between the low-skill worker and low-skill job, and an arc between the high-skill worker and high-skill job.

To explain the presence of negative sorting, consider the opposite configuration of distances where $d^{\zeta}_{LL} + d^{\zeta}_{HH} > d^{\zeta}_{LH} + d^{\zeta}_{HL}$. In this case, the high-skill worker and the low-skill job are close to each other, while the distance between the low-skill worker and the low-skill job as well as the distance between the high-skill worker and the high-skill job is large. Since the mismatch cost is concave in the distance between the worker skill and the job, it is intuitively optimal to pair the high-skill worker with the low-skill job since having one small mismatch and one large mismatch is better than having two medium-sized mismatches. While the cost function is identical, the optimal sorting pattern differs due to the differences in the underlying distributions of workers and jobs. This case is shown in the bottom panel of Figure 4.[5]

These two cases illustrate why general characterization of optimal sorting may be complex. The main reason is that both technology as well as the worker and job distributions jointly determine optimal sorting rather than technology alone (which is the case for assortative sorting results in the literature).

Composite Sorting. In order to show that an optimal assignment can feature composite sorting, we first consider an assignment problem between three workers and three jobs in the bottom half of Figure 5. Since the skill groups are very far apart, it is optimal to positively sort within groups as in the top panel of Figure 4.

[5]When $d^{\zeta}_{LL} + d^{\zeta}_{HH} = d^{\zeta}_{LH} + d^{\zeta}_{HL}$ positive and negative sorting both induce the same cost of mismatch and the optimal assignment is not unique. For any distributions of workers F and jobs G, we show that the set of values for ζ_p and ζ_u such that the optimal assignment is not unique has Lebesgue measure zero in the Technical Appendix.

The top half of Figure 5 introduces an additional low-skill worker x_L and an additional high-skill job z_H. One could break the medium-skill worker-job pair (x_M, z_M) such that the added low-skill worker is assigned to the medium-skill job forming a pair (x_L, z_M) and the added high-skill job is assigned to the medium-skill worker forming a pair (x_M, z_H).[6] This assignment gives two medium-sized mismatches. Instead, pairing the added low-skill worker to the added high-skill job forming (x_L, z_H), and preserving the medium-skill pair (x_M, z_M), results in one small mismatch and one large mismatch which is preferred by the concavity of the mismatch cost function. The added low-skill worker x_L and high-skill job x_H are thus optimally paired as indicated by the arc in Figure 5.

We emphasize that the optimal assignment features composite sorting. First, distinct worker types are paired to identical jobs. In Figure 5, a low-skill worker x_L and a high-skill type worker x_H both work in the identical high-skill occupation z_H, while the medium-skill worker x_M does not work in this occupation. Second, a worker type is simultaneously part of both positive and negative sorting. In Figure 5, a low-skill worker x_L is paired positively with the low-skill job z_L (part of the positive sorting (x_L, z_L) and (x_H, z_H) in blue) but the same worker type is also paired negatively to the distant high-skill job z_H (part of the negative sorting (x_L, z_H) and (x_M, z_M) in orange).[7]

3.2 Optimality Conditions

We next describe three important features that an optimal sorting satisfies: maximal number of perfect pairs, no intersecting pairs, and layering. A necessary condition for an optimal assignment is that aggregate output does not increase by a bilateral exchange of workers between jobs. For any two pairs in an optimal assignment (x, z) and $(\hat{x}, \hat{z})$:

$$y(x, z) + y(\hat{x}, \hat{z}) \geqslant y(x, \hat{z}) + y(x, \hat{z}). \tag{9}$$

We use this optimality condition to demonstrate the main features of an optimal assignment.

Maximal Number of Perfect Pairs. An optimal assignment maximizes the number of pairs that are perfectly sorted, i.e., the number of pairs with zero cost of mismatch between workers and jobs, or $x = z$.

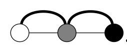
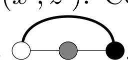
When mismatch costs are concave, it is preferable to have a pair with small mismatch and a pair with significant mismatch than to have two pairs with medium mismatch. A perfect pair is a demonstration of this concept since it has zero mismatch. Specifically, let an optimal assignment contain pairs (x, z)

[6] Adding a low-skill worker x_L and a high-difficulty job x_H does not allow to reduce the mismatch cost of pairs (x_L, z_L) and (x_H, z_H).

[7] In equilibrium, workers of a type x that is assigned to multiple occupations are indifferent between these occupations since the worker receives identical compensation $w(x)$.

Figure 6: Intersecting and Non-Intersecting Pairs

Figure 6 illustrates intersecting and non-intersecting pairs. The left panel shows an example of intersecting pairs as the arcs of pairs (x, z) and (x', z') intersect. The right panel shows an example of non-intersecting pairs as the arcs corresponding to pairs (x, z) and (x', z') do not intersect. Under an optimal assignment arcs never intersect.

and (x', z') but $x' = z$ where x and z' do not equal z. Suppose $x < x' = z < z'$, so the value $z = x'$ is in between x and z. This configuration can be visualized as , where the gray circle indicates the presence of both a worker x' and a job z, that is, the location of the potential perfect pair. The original cost of mismatch is two medium-size mismatches in pairs (x, z) and (x', z'). Consider reshuffling to form a perfect pair (x', z) and the pair (x, z'), which can be visualized as . The cost of mismatch of the reshuffled pairs is that of a large mismatch (x, z') and a zero mismatch for the perfect pair (x', z). By concavity of the cost, this is lower than two medium-size mismatches. The output loss due to mismatch is thus strictly reduced by assigning worker x to job z' and by perfectly assigning worker x' to job z, which contradicts the optimality condition (9).[8]

Maximal perfect pairing shows that workers and jobs that are part of the common component of the worker and job distributions are positively sorted. In analyzing the assignment problem between the remaining workers and jobs we can thus consider assignments between worker and job distributions for which the common components are removed. For brevity, we label the remaining worker distribution F and the remaining job distribution G. In the optimal transport literature the idea of the perfect pairs is referred to as "mass stays in place if it can" (Gangbo and McCann, 1996; Villani, 2003). That is, a planner who wants to transport mass, and is faced with transportation costs which are concave in distance, does not move mass that is already at a required destination.

No Intersecting Pairs. The second feature of an optimal assignment is that pairings between workers and jobs do not intersect. The result is a direct consequence of concavity of the cost function.

We first describe intersecting and non-intersecting pairs. Consider pairs (x, z) and (x', z') and illustrate their pairings by arcs. Figure 6 displays intersecting and non-intersecting pairs. The left panel shows an example of intersecting pairs as the arcs of pairs (x, z) and (x', z') intersect. The right panel shows an example of non-intersecting pairs as the arcs corresponding to pairs (x, z) and (x', z') do not

[8] We present a formal statement of this feature and a proof in Appendix A.3. Note that a strictly convex cost of mismatch instead implies that positive sorting is optimal which generically conflicts with maximal perfect pairing.

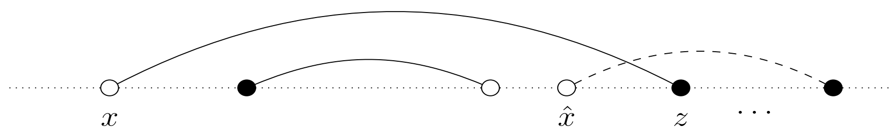
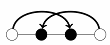
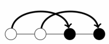
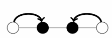
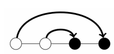


Figure 7: Layering

Figure 7 show that if worker x and job z are paired, then there is the same number of workers and jobs in between skill levels x and z. Suppose that the number of workers and jobs in between skill levels x and z instead is not the same so that there is some worker $\hat{x}$ that cannot be paired with a job inside the skill interval (x, z). Worker $\hat{x}$ then has to be paired with a job $\hat{z}$ outside the interval (x, z) which contradicts no intersecting pairs.

intersect.[9]

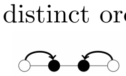
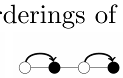
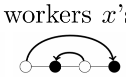
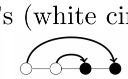
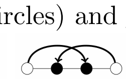
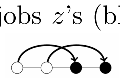
We argue that for any two pairs (x, z) and (x', z') in an optimal sorting, their arcs do not intersect as in the right panel of Figure 6.[10] Specifically, consider two unique configurations with intersecting pairs:  and .[11] The first configuration is $x < z' < z < x'$. Since the mismatch cost increases in the distance between the worker and the job, an improvement is to instead pair the closer points (x, z') and (x', z), or , as it reduces the mismatch cost for each worker and hence the total mismatch cost. The second configuration is $x < x' < z < z'$. In this case, we reshuffle so that we have one large mismatch (x, z') and one small mismatch (x', z), or . By concavity, the cost is smaller than two medium-size mismatches. The total mismatch loss strictly reduces by assigning worker x to job z' and worker x' to job z, which contradicts the optimality condition (9).

The observation on the impossibility of crossing first appeared in Monge (1781) as discussed by Villani (2009). The non-crossing arcs are also a feature of the literature on optimal transportation with concave distance costs (Gangbo and McCann, 1996; McCann, 1999) and of algorithmic sorting problems with distance costs (Aggarwal, Barnoy, Khuller, Kravets, and Schieber, 1995; Werman, Peleg, Melter, and Kong, 1986).

Layering. The result that an optimal sorting does not contain intersections leads to an observation that the assignment problem can be decomposed into a series of independent problems, or layers (Aggarwal,

[9] More formally, arcs (x, z) and (x', z') do not intersect if and only if the intervals (x, z) and (x', z') are either disjoint or one interval is a subset of the other interval. When referring to an interval (x, z), we do not require that the worker skills and job difficulties are ordered: we mean the set of real numbers lying between z and x on the real line.

[10] A formal statement of this feature and a proof are in Appendix A.4.

[11] There are six distinct orderings of workers x's (white circles) and jobs z's (black dots) to consider, which can be represented as:      . The first four configurations do not contain intersections. The final two configurations are discussed in the main text.

Barnoy, Khuller, Kravets, and Schieber, 1995; Delon, Salomon, and Sobolevski, 2012a).[12]

To establish layering, first observe that if worker x is paired with job z, then there is the same number of workers and jobs in between skill levels x and z. Suppose the number of workers and the number of jobs in between skill levels x and z are not the same, and suppose all points have equal weight. Then there exists a worker $\hat{x}$ that cannot be paired with a job inside the skill interval (x, z). We illustrate this argument in Figure 7, where worker $\hat{x}$ cannot be paired with a job inside the skill interval (x, z). As a result, worker $\hat{x}$ has to be paired with a job outside the interval (x, z), which would lead to intersection of pairs (x, z) and $(\hat{x}, \hat{z})$, contradicting no intersecting pairs. We conclude there is the same number of workers and jobs between worker x and job z, $F(z)-F(x) = G(z)-G(x)$. Alternatively, this requirement can be written as $F(z) - G(z) = F(x) - G(x)$.

We use the observation that the number of workers and jobs between an optimally paired worker x and job z is identical to decompose the overall sorting problem into sorting problems for different layers of the measure of underqualification. The measure of underqualification $H = F - G$ defines the extent to which workers up to skill level s outnumber the jobs requiring skill levels up to s. Since the number of workers and jobs in between optimally paired workers and jobs is identical, only workers and jobs within the same layer of the measure of underqualification can be paired. An optimal assignment between workers and jobs is thus equal to the sum of optimal assignments for each horizontal slice or layer of the measure of underqualification H. This observation decomposes the original problem into independent problems for each layer.[13] The solution to the full problem is given by aggregating the solutions to the assignment problems for each layer with the same cost function.[14]

3.3 Sorting Within a Layer

We now construct a recursive characterization for an optimal assignment within a given layer. This recursive formulation reflects on the salient features of optimal sorting stemming from concavity of the cost. We use the approach of Aggarwal, Barnoy, Khuller, Kravets, and Schieber (1995) that centers on the

[12] A formal statement of the layering feature and a proof are presented in Appendix A.5.

[13] In each layer, there is an alternating configuration of workers and jobs − every worker skill level is followed by a job difficulty level, possibly except for the last one. We define an alternating assignment problem as an assignment problem between alternating workers and jobs.

[14] The properties of maximal number of perfect pairs, no intersecting pairs, and layering on their own may be useful to construct simple algorithms to approximate optimal sorting mechanisms. Caracciolo, D'Achille, Erba, and Sportiello (2020) and Ottolini and Steinerberger (2023), for example, only use no intersecting pairs and layering to, respectively, construct a Dyck algorithm and greedy matching algorithm to study approximate optimal sorting for a random assignment problem. They show that the aggregate mismatch costs under the simple assignment scale similarly to the aggregate costs of mismatch for the optimal assignment, that is, achieves optimum on average up to a scaling constant, in the asymptotic limit with infinitely many points.

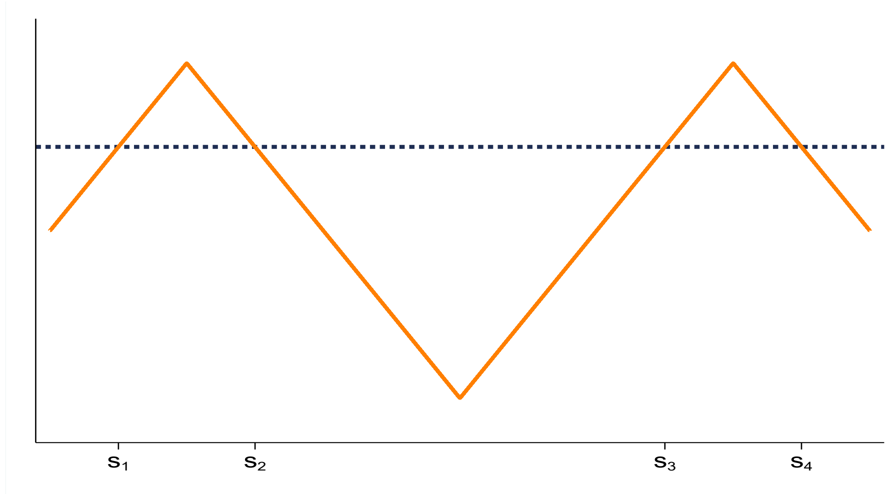


Figure 8: Measure of Underqualification

Figure 8 illustrates a layer given a measure of underqualification H (solid orange line). Layering breaks the original assignment problem into independent problems for each layer, such as the layer indicated by the blue dashed line. Since the number of workers and jobs in between an optimally paired worker x and job z is identical, only workers and jobs within the same layer can be paired. That is, only workers (s_1, s_3) and jobs (s_3, s_4) can optimally be paired.

property of no intersecting pairs. Specifically, we adopt the recursive algorithm developed by Nechaev, Sobolevski, and Valba (2013), designed to model statistical properties of polymer chains.[15]

The optimal assignment problem for a given layer is an alternating assignment problem. An optimal assignment within a layer matches one worker with precisely one job. For notational convenience, we order workers and jobs within each layer by their skill levels. Let there be n_ℓ workers and n_ℓ jobs in a given layer, and we denote the skill levels by $s_1 < s_2 < \cdots < s_{2n_\ell - 1} < s_{2n_\ell}$.

We write a Bellman equation to calculate the minimum aggregate cost of mismatch. The recursive component of the Bellman equation is that we consider assignment problems with an increasing number of skill levels. We start by solving all assignment problems between two consecutive elements: the assignment problem between one worker and one job. That is, we consider assignments between skill levels s_i and s_{i+1}, for each i. Using the solutions from the previous step, we proceed to solve all assignment problems between four consecutive elements (two workers and two jobs) and so on.

We denote by $V_{i,j}$ the minimum cost of mismatch when sorting all workers and jobs with skill levels between s_i and s_j (inclusive), where $j > i$. The difference $j - i$ is odd so that there are equal numbers of workers and jobs between s_i and s_j. Considering an assignment of workers and jobs with skill levels in

[15] All proofs are in Appendix A.6. From a mathematical perspective, we provide a novel proof of a computationally efficient version of the Bellman equation in Delon, Salomon, and Sobolevski (2012b). Our proof also extends these results to the case of asymmetric mismatch cost functions.

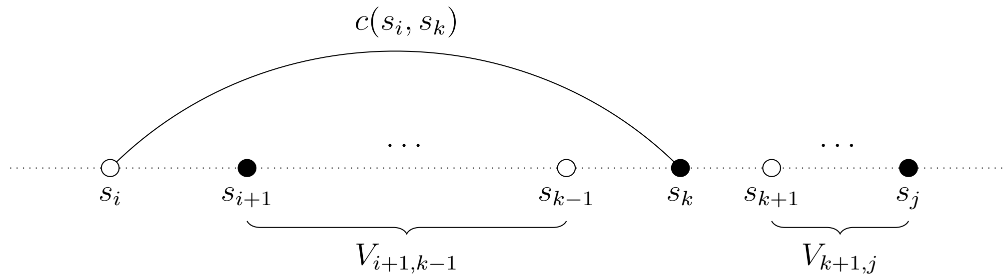


Figure 9: Bellman Equation

Figure 9 illustrates the Bellman equation for an optimal sorting. Consider an assignment problem between workers and jobs with skill levels in $[s_i, s_j]$. We can pair the leftmost skill s_i with any skill s_k such that $k - i$ is odd, as illustrated by the pair (s_i, s_k). Upon pairing s_i with s_k, the planner remains to optimally pair the other workers and jobs in $[s_{i+1}, s_{k-1}]$, and all workers and jobs with skill levels in $[s_{k+1}, s_j]$. There are no pairings between these two segments because this violates the property of no intersecting pairs.

$[s_i, s_j]$, the planner can pair the leftmost s_i with any s_k such that $k - i$ is odd. Upon pairing s_i with s_k, the planner remains to optimally pair the workers and jobs in $[s_{i+1}, s_{k-1}]$, and all workers and jobs with skill levels in $[s_{k+1}, s_j]$. The main observation that facilitates this characterization is that there are no pairings between these two segments because this violates the property of no intersecting pairs. Using the results from previous steps to obtain costs $V_{i+1,k-1}$ and $V_{k+1,j}$ delivers the Bellman equation:

$$V_{i,j} = \min_{k \in \{i+1, i+3, \ldots, j\}} \big(c(s_i, s_k) + V_{i+1,k-1} + V_{k+1,j} \big) \tag{10}$$

with boundary conditions $V_{i+1,i} = 0$ for all i.[16]

Finally, we construct an optimal assignment. Starting from $V_{1,2n_\ell}$, the optimal pairing of skill s_1 is given by skill s_k that solves equation (10). Then two corresponding continuation values, $V_{2,k-1}$ and $V_{k+1,2n_\ell}$, are evaluated to determine optimal pairings for skill s_2 and for skill s_{k+1}, respectively. This process of finding an optimal assignment continues until a full assignment is constructed.

3.4 Characterizing Optimal Sorting

The three-step approach to characterizing the optimal sorting is as follows. First, we maximize perfect pairs. This leads to positive sorting between identical workers and jobs and we can withdraw them from further analysis. Second, we construct a measure of underqualification to decompose the assignment problem into a sequence of independent layers with n_ℓ skills in each layer ℓ. Third, we characterize an optimal assignment for each layer using the Bellman algorithm. The optimal assignment is then given by

[16] The boundary conditions are invoked at either end of the choice interval. When $k = i + 1$, the minimum cost of mismatch is $c(s_i, s_{i+1}) + V_{i+2,j}$, the cost of pairing the first worker to the first job, together with optimally sorting all skill levels from s_{i+2} to s_j. When $k = j$, the minimum cost is $c(s_i, s_j) + V_{i+1,j-1}$, the cost of pairing the first worker to the last job, together with optimally sorting all intermediate skill levels between s_{i+1} and s_{j-1}.

the following theorem.

Theorem 1. An optimal assignment between workers F and jobs G sums optimal assignments for each layer, where an optimal assignment in each layer attains $V_{1,2n_\ell}$ in the Bellman equation (10).

3.5 Comparative Statics

We next show that optimal sorting becomes more positive, by which we formally mean larger in a natural ordering on assignments called concordance order, as the mismatch cost function becomes less concave. Moreover, we show that there exists a threshold in concavity of costs ζ_p and ζ_u beyond which the optimal assignment within each layer is positive.

For any two assignments π and $\hat{\pi}$ between a fixed pair of distributions of workers and jobs, we say that assignment π is smaller in concordance order than $\hat{\pi}$, which we denote by $\pi \preceq \hat{\pi}$, if for any coordinate (x_c, z_c), less mass is concentrated in both the top-right and bottom-left quadrants under assignment π than under $\hat{\pi}$. Intuitively, a more positive sorting corresponds to an assignment larger in concordance order, and this equivalence was made precise by Tchen (1980). When assignment $\hat{\pi}$ is larger in concordance order, it is implied that other measures of statistical association, such as the correlation coefficient, the rank correlation, and Kendall's τ coefficient are also larger for $\hat{\pi}$ than for assignment π (Joe, 1997).

Theorem 2 shows that optimal sorting is more positive when the costs of mismatch becomes less concave, which we prove in Appendix A.7.

Theorem 2. Suppose the mismatch cost $c(x, z)$ is of the form (5) and that for some increasing convex function κ, $\hat{c} = \kappa(c)$ is also of the form (5). If π is an optimal assignment with costs c, then there exists an optimal assignment $\hat{\pi}$ with the less concave mismatch cost $\hat{c}$ such that $\pi \preceq \hat{\pi}$.

We outline the main steps of the proof. Since the distributions of workers and jobs are identical, the sorting problem is split into identical layers irrespective of the cost function. Hence, we show that optimal sorting within each layer becomes more positive as the mismatch cost becomes less concave. If an optimal assignment is no longer optimal when the costs are less concave with the mismatch costs $\hat{c}$, there exists a pair (x_0, z_0) with positively sorted subpairs $\{(x_i, z_i)\}_{i=1}^{p}$, as in the top panel of Figure 10 (by Lemma 7), such that the local assignment problem with workers and jobs $\{(x_i, z_i)\}_{i=0}^{p}$ can be improved with more positive sorting shown in the bottom panel of Figure 10.[17] To prove this, suppose that in an optimal assignment with less concave costs, the worker $x_0 = \tilde{x}_1$ is instead optimally paired to job $z_k = \tilde{z}_1$ for

[17]Formally, we call a pair (x, z) a subpair of the pair (x_0, z_0) if the pair (x, z) is not a nested pair in the interval $[x_0, z_0]$ that is not equal to (x_0, z_0).

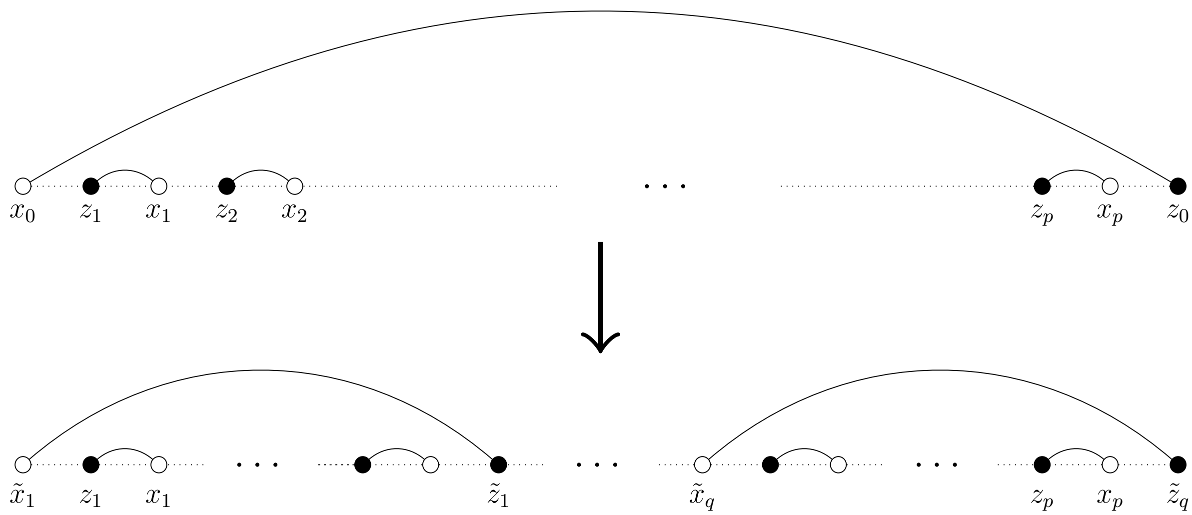


Figure 10: Improving the Assignment When Costs of Mismatch Become Less Concave

Figure 10 visualizes an optimal assignment that is no longer optimal when the costs of mismatch become less concave. In this case, there exists a pair (x_0, z_0) with positively sorted subpairs $\{(x_i, z_i)\}_{i=1}^p$, which we display in the top panel, such that the assignment with workers and jobs $\{(x_i, z_i)\}_{i=0}^p$ can be improved in a more positive fashion, as shown in the bottom panel.

some k. Since positive sorting is optimal on the interval (x_0, z_k) with more concave cost of mismatch, we show in Lemma 8 that positive sorting is also optimal on (x_0, z_k) with less concave cost of mismatch. We continue this procedure to the right, that is, we start with worker $x_{k+1} = \tilde{x}_2$ and repeat the argument, and indeed obtain the structure in the bottom panel of Figure 10.

To see that the optimal assignment becomes more positive, or larger in concordance order, we make two observations. First, observe that all successively positively sorted pairs in the bottom panel of Figure 10, such as (x_1, z_1), (x_2, z_2) and (x_p, z_p), are also formed in the top panel of Figure 10. Hence, they do not affect the concordance order. After removing these pairs, we secondly note that the bottom panel sorts the remaining workers and jobs positively, which has the largest concordance order among all assignments. Since the top panel does not sort the remaining workers and jobs positively, it indeed follows that the assignment for the bottom panel is larger in concordance order. Hence, all improvements increase the assignment in concordance order.

Another approach to analyze comparative statics might have drawn from recent results of Anderson and Smith (2023). Specifically, Anderson and Smith (2023) study comparative statics of optimal sorting with respect to the output function for general sorting models. They use the concordance order which they refer to as positive quadrant dependence order (Lehmann, 1966) and provide sufficient conditions under which sorting is larger in concordance order as the output function changes.[18] However, in Appendix A.7 we show an example in our economy for which their conditions are not satisfied.

Threshold for Layered Positive Assignment. The previous result shows that a more concave cost function yields more negative sorting, and a less concave cost function yields more positive sorting. We next derive a threshold for concavity beyond which optimal sorting is the most positive assignment for our economy. The most positive assignment for our economy is given by positive sorting in each layer, which we call a layered positive assignment. We note that positive sorting in each layer does not imply positive sorting overall.

We consider the assignment problem when the power indices ζ_p and ζ_u are close to one, that is, when the cost of mismatch is almost linear in the distance between the worker skill and the job. First, we maximize the number of perfect pairings. Second, we decompose the assignment problem into layers $0 \leqslant \ell \leqslant L$. Third, when ζ_p and ζ_u exceed the threshold $\bar{\zeta}$, the optimal assignment within each layer is simple. Specifically, we show in Theorem 3 below that the optimal sorting within each layer is positive sorting which we denote by π_ℓ^+. The solution to the full assignment problem is given by the sum of the positive assignments within each layer. We refer to this assignment as the layered positive assignment denoted by $\pi^+ = \sum \pi_\ell^+$.

Theorem 3. Given a worker distribution F and a job distribution G, there exists $\bar{\zeta} < 1$ such that for any $\zeta_p, \zeta_u \in [\bar{\zeta}, 1]$, the layered positive assignment π^+ is optimal.

The proof is in Appendix A.8. The implication is that for mismatch power indices above the threshold $\bar{\zeta}$, the solution can be directly constructed by evaluating the measure of underqualification, and by assigning positively within each layer. It is useful to contrast our result with Juillet (2020) who shows that the layered positive assignment is the limit of some optimal assignments as $\zeta \to 1^-$. Our results proves the existence of a threshold beyond which the layered positive assignment is optimal for our environment and is thus applicable away from the limit.[19]

[18] Positive quadrant dependence order is equivalent to concordance order when two assignments share the same pair of marginal distributions, which is satisfied for the sorting problem. Generally, the two orders can be different.

[19] In Appendix A.8, we also discuss the assignment problem with linear costs of mismatch ($\zeta_p = \zeta_u = 1$), which is the original formulation of the optimal transport problem due to Monge (1781) that has been well studied (Rachev and Rüschendorf, 1998; Villani, 2003).

4 Wages and Firm Values

We determine equilibrium wages and firm values by characterizing the solution to the dual problem (8). The first part of this section characterizes the dual solution for mismatched workers and jobs. We show that this shadow cost of mismatch has a regional hierarchical structure. The second part of the section constructs the dual for both mismatched and perfectly paired workers and jobs.

From the economic point of view, the dual solution allows to connect the model with directly observable variables such as wages. Our construction of the solution and characterization of its regional hierarchical structure are also new to the optimal transport literature. In the optimal transport literature, a local-global structure for the primal problem is one of the main results in McCann (1999), where the points of nondifferentiability of the shadow price in the dual problem are used to determine local regions for the primal problem.[20] Instead, we show that the dual solution has a regional-global structure. The technical difficulty in our construction is to ensure consistency at every scale from regional to global via aggregation of the regional relative wages. Ours is the first result in the literature to provide a complete constructive characterization of the dual solution for an optimal transport model with concave costs.

Mismatched Workers and Jobs. Let $S = I \cup J$ be the set of all skill levels, where I and J are disjoint sets of worker skills and job levels after the removal of perfect pairs. Suppose that an optimal assignment π consists of n worker-job pairs $(x_i, z_i) \in \Gamma_\pi$. Our goal is to first construct a shadow mismatch cost function $\phi : S \to \mathbb{R}$ such that for each worker x and every job z, $\phi(x) - \phi(z) \leqslant c(x, z)$, which holds with equality if the assignment π pairs worker x to job z.[21]

Figure 11 illustrates the construction of the regional hierarchical structure for the dual solution given the optimal sorting captured by the arcs. A lower skill group has skills in the interval between x_1 and z_1. A higher skill group has skills in the interval between x_2 and z_2. The relative shadow cost of mismatch for all skills within either the first or the second group is determined solely within each group. That is, wage determination is regional, meaning that the wage is determined within a group. Wage determination is also hierarchical within groups: at each stage, wages depend only on information from the skill group nested within the progressively larger group. In the low skill group, the wage is first determined for the innermost pair (x_3, z_3) which contains no nested skill groups, and then for the outer pair (x_1, z_1).

[20] McCann (1999) shows that for concave distance costs, optimal sorting has an intricate local-global hierarchical structure. This structure is then used to reduce the problem to a convex minimization problem and to derive a combinatorial algorithm for the solution as an optimization of a finite sequence of convex, separable network flows.

[21] Setting $\psi(z) = -\phi(z)$ for all mismatched jobs $z \in Z$, we equivalently construct the worker wage $\phi : I \to \mathbb{R}$ and the firm value $\psi : J \to \mathbb{R}$ so that for each worker x and job z, $\phi(x) + \psi(z) \leqslant c(x, z)$, where the equality holds if the optimal assignment π pairs worker type x with occupation z, that is $(x, z) \in \Gamma_\pi$.

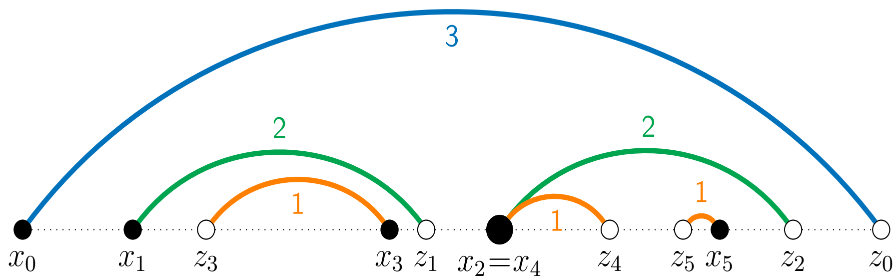


Figure 11: Regional Hierarchical Structure

Figure 11 illustrates the construction of the regional hierarchical structure for the dual solution given the optimal sorting captured by the arcs. A lower skill group has skills in the interval between x_1 and z_1 while a higher skill group has skills in the interval between x_2 and z_2. The hierarchical structure implies that the relative shadow cost of mismatch for all skills within either the first or the second group is determined exclusively within the group (region). Wages are determined hierarchically within groups. In the lower skill group, wages are first determined for the innermost pair (x_3, z_3), and then for the outer pair (x_1, z_1). The numbers on the arcs indicate the sequence by which we move from low-level pairs to high-level pairs.

In other words, hierarchical structure implies that the construction of wages moves sequentially from low-level pairs to high-level pairs, as indicated by the numbers on the arcs in Figure 11. For the high skill group, the relative wage for the pair (x_2, z_2) is thus constructed from the relative wages for both pairs (x_4, z_4) and (x_5, z_5). Finally, wages for the outermost pair (x_0, z_0) are constructed using the relative wages for the first and second skill group.

The formal definition of the regional hierarchical mechanism is analogous to this example but is notationally more involved. The main implication of the regional hierarchical structure is that relative wages are determined regionally − for any two skills s and s' in a given skill group, their relative wages $\phi(s) - \phi(s')$ can be computed only based on information in this skill group.[22] The mechanism then aggregates this regional structure by constructing suitable level shifts that preserve the regional structure of relative wages within groups while ensuring consistency across all groups.[23] We present a condensed statement of the theorem here and include the complete definition of the mechanism, explicit examples, the full statement of the theorem, and the proof in Appendix A.11.

Theorem 4. Given an optimal assignment, the regional hierarchical mechanism constructs an optimal

[22]Since the optimality of a dual potential is invariant to constant shifts, the relative difference $\phi(s) - \phi(s')$ contains all information regarding the solution to the dual problem.

[23]The regional groups generally do not appear for convex costs. For instance, when all worker skills are lower than each job difficulty, the optimal assignment for a convex cost is the positive sorting, which allows for no local regions (except for the whole set) since any two pairs intersect. In contrast, for a concave cost, every pair forms a local region regardless of how workers and jobs are located.

dual pair (ϕ, ψ) where $\psi = -\phi$. Within each skill group, relative wages are determined regionally: for any two points s and s' in a skill group, $\phi(s) - \phi(s')$ depends only on the pairs within the group.

We next use the dual functions to define worker wages and firm values for the non-overlapping segments of the worker distribution and the job distribution. Let wages $w(x) = g(x) - \phi(x)$ and firm values $v(z) = h(z) - \psi(z)$, where we recall from the technology (1) that g reflects the worker contribution to production independent of the occupation, and h reflects the value of the job independent of the worker that fulfills the job. The first observation is that the assignment π which solves the primal mismatch cost minimization problem (7) also solves the primal output maximization problem (6). Moreover, $w(x) + v(z) \geqslant y(x, z)$ holds for all (x, z) with equality if $(x, z) \in \Gamma_\pi$, where $y(x, z) = g(x) + h(z) - c(x, z)$. By Lemma 1, it follows that (w, v) is a dual optimizer for the output maximization problem. In sum, given the shadow mismatch cost (ϕ, ψ) for the minimization problem without overlapping parts, the dual pair (w, v) for the maximization problem without overlapping parts is obtained.

Adding Perfectly Paired Workers and Jobs. Up to this point, we addressed the issue of determining worker wages and firm values in the output maximization problem when there is no overlap between the distributions of workers and jobs. We now discuss how these wage and value functions can be used to sequentially construct worker wages and firm values for the problem with overlapping parts in the distributions.

We start with an equilibrium of only mismatched workers and jobs and denote the wages constructed above by $\tilde{w}$. After removing the common parts of the distributions, denote the sets of mismatched workers and jobs by I and J. We add perfectly matched firms and determine what income each firm could generate given mismatched workers and wages $\tilde{w}$. The first auxiliary firm problem is to choose an employee among only mismatched workers $x \in I$. Formally, a firm with job $z \in I \cup J$ solves: $\tilde{v}(z) := \max_{x \in I} \big(y(x, z) - \tilde{w}(x)\big)$. We refer to $\tilde{v}$ as firm mismatch compensation, that is, the profits firms can attain given a mismatched worker with wage $\tilde{w}$.

We next introduce perfectly paired workers and present both mismatched and perfectly paired workers with firm mismatch compensation $\tilde{v}$. We determine what wage income both the imperfectly and perfectly paired workers would generate given compensation required by firms. The auxiliary decision problem of a worker $x \in I \cup J$ is to choose any job, including the perfectly paired jobs, to solve:

$$\hat{w}(x) := \max_{z \in I \cup J} \big(y(x, z) - \tilde{v}(z)\big). \tag{11}$$

As a result, we obtain wages $\hat{w}$ for both mismatched and perfectly paired workers.

Finally, we determine what income $\hat{v}$ each firm would generate given all workers and their required compensation $\hat{w}$. We set up a second auxiliary firm problem, which is the problem of a mismatched job $z \in J$ choosing an employee among all workers (perfectly paired and mismatched) subject to the wage schedule $\hat{w}$:

$$\hat{v}(z) := \max_{x \in I \cup J} \big(y(x,z) - \hat{w}(x)\big). \tag{12}$$

We refer to $\hat{v}$ as mismatched firm compensation, since it represents the profits of firm type $z \in J$ that is mismatched in equilibrium.

Equilibrium wages are formulated using auxiliary wages for mismatched workers, $w(x) = \hat{w}(x)$ for all $x \in I$, as well as mismatched firm compensation, $v(z) = \hat{v}(z)$ for all $z \in J$, to set $w(x) = g(x)+h(x)-v(x)$ for all $x \in J$. Equilibrium firm values v are then given by $v(z) = g(z)+h(z)-w(z)$ for every job $z \in I \cup J$. Theorem 5 shows that the wage function w and the firm value function v indeed solve the dual problem for the full assignment problem.[24]

Theorem 5. The constructed functions (w, v) are a dual solution for the sorting problem between worker distribution F and job distribution G; that is, $w(x) + v(z) \geqslant y(x,z)$, where equality holds everywhere with respect to Γ_π.

The proof, as well as a formal analysis of the above mechanism, is in Appendix A.12.

5 Composite Sorting and Wage Dispersion

The distinguishing feature of composite sorting is wage dispersion within occupations. The analysis so far provides a general characterization of optimal sorting and wages. In this section, we develop an example that emphasizes essential elements of the characterization in Section 3 and evaluate its qualitative and quantitative implications for wage dispersion within occupations.

Data. Our data source is the American Community Survey. We consider all individuals between 25 and 60 years of age from 2010 to 2017. The final sample includes about 6.7 million individuals. We select this period to work with a single Standard Occupational Classification (SOC) code, which yields 497 distinct and consistent occupations. Our measure of income is annual wage and salary income before taxes.

Figure 12 shows wage dispersion within occupations. On the horizontal axis, we rank occupations by average earnings in each occupation. For every occupation, we calculate the dispersion in log wages

[24] We can extend the domain of wages w and values v to $K = X \setminus (I \cup J)$ by setting $w(x) = \max_{z \in I \cup J} \big(y(x,z) - v(z)\big)$ for $x \in K$ and $v(z) = g(z) + h(z) - w(z)$ for $z \in K$.

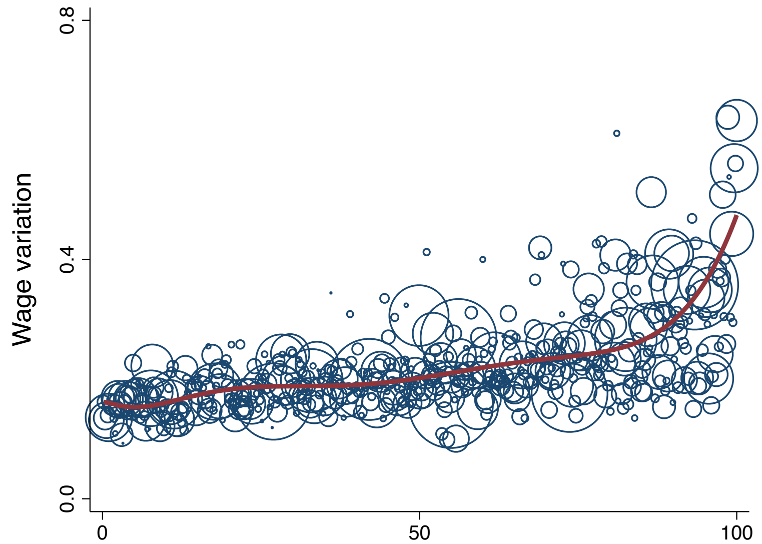


Figure 12: Wage Dispersion by Occupation

Figure 12 displays wage dispersion within occupation. On the horizontal axis, we rank occupations by the average wages earned in each occupation. For every occupation, we calculate the dispersion in logarithmic wages within that occupation, where the size of the circle indicates the share of employment within the occupation. The data pattern is summarized by the red solid line.

within that occupation, where circle size indicates the share of total employment within the occupation. The salient data pattern is captured by the red solid line which is a fractional polynomial fit. Wage dispersion within occupations is sizable, with average wage dispersion equal to 0.24 log points. Second, wage dispersion within occupations is relatively constant at the bottom 80 percent of occupations, but rapidly increases for the top fifth of occupations.

Example. Two elements are essential to require the general characterization of optimal sorting in Section 3.[25] First, there need to be mismatched workers and jobs. Specifically, there are at least two workers and two jobs in some layers of the measure of underqualification to necessitate the use of the Bellman equation (10). Second, there are perfectly paired workers and jobs.

We start by describing the mismatched workers and jobs. In Figure 13, we plot a numerical example of the measure of underqualification in Figure 8. An increasing measure of underqualification indicates mismatched workers; a decreasing measure of underqualification indicates mismatched jobs. Figure 13 thus indicates mismatched workers with skills between 0 and 500 as well as mismatched workers between 1500 and 2500. Similarly, there are mismatched jobs with skill difficulties between 500 and 1500 as well as mismatched jobs with skill difficulties between 2500 and 3000.

In order to characterize optimal sorting, we split the measure of underqualification into layers as

[25]For simplicity, we consider a symmetric cost of mismatch with $\zeta_p = \zeta_u \in (0, 1)$.

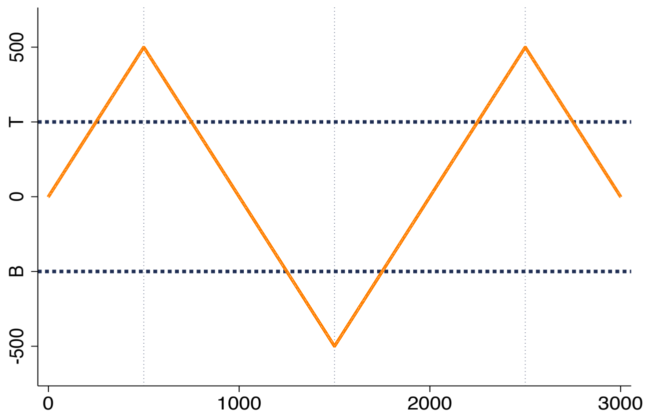


Figure 13: Measure of Underqualification

Figure 13 shows the measure of underqualification for our quantitative analysis. An increasing measure of underqualification indicates mismatched workers; a decreasing measure of underqualification indicates mismatched jobs.

illustrated by the blue dashed horizontal lines in Figure 13. The optimal sorting in layers corresponding to the bottom half, for example layer B in Figure 8, is simple as it contains a single worker and a single job, which necessarily are paired. The top layer, for example layer T, instead contains two workers and two jobs. In each top layer in Figure 8, the low-skill worker and the low-skill job as well as the high-skill worker and the high-skill job are close to each other, while the gap between the low-skill job and the high-skill worker is large. This resembles the configuration in the top panel of Figure 4. It is optimal to pair the low-skill worker to the low-skill job and to pair the high-skill worker to the high-skill job, and hence optimal sorting is positive in each layer.[26]

The optimal pairing of mismatched workers and jobs features significant variation in mismatch. The mismatch varies from almost zero for the worker at skill (slightly below) 500 and the job at skill (slightly above) 500 to mismatch equal to 1000 for the worker at skill 0 and the job at skill 1000.[27]

We then add common parts to the distributions of workers and jobs so that the resulting distribution of workers is uniform between skills 0 and 3000. Specifically, we add workers and jobs with skill levels between 500 and 1500 as well as workers and jobs with skill levels between 2500 and 3000. In Section 3, we showed than an optimal assignment perfectly pairs these workers and jobs to maximize the number

[26]Optimal sorting is positive in all layers in the top panel of Figure 13 since for all $\zeta = \zeta_p = \zeta_u \in (0, 1)$, it holds that $(z_L - x_L)^\zeta + (z_H - x_H)^\zeta \leqslant (z_H - x_L)^\zeta + (x_H - z_L)^\zeta$. Our numerical findings are thus invariant to ζ. Similarly, we do not specify the function $h(z)$ as all our findings are invariant to it. In principle, the function h could be directly inferred using a measure of the value of a job. Although optimal sorting in this example is intentionally simple and can be verified directly, in general when a layer has multiple workers and jobs, one needs to apply the Bellman equation (10).

[27]In contrast, if sorting were positive there would be no variation in mismatch – mismatch for each worker-job pair would equal 500.

of perfect pairs.

In order to describe the intuition for mismatch between workers and jobs, consider, for example, the region of skills between 0 and 1000. Workers with skills between 0 and 500 are sorted negatively to jobs between 1000 to 500. For example, worker 0 is paired with job 1000, and worker 400 is paired with job 600. Workers with skills between 500 and 1000 are perfectly positively paired to the jobs between 500 and 1000. This assignment is an example of composite sorting. First, there are two distinct workers in each job. For example, both a worker with skill 0 and a worker with skill 1000 are assigned to occupation 1000. Second, the same occupation is part of both positive and negative sorting. For example, occupation 1000 is a part of positive sorting (with worker 1000 as a part of positive sorting with perfect pairs) and negative sorting (with worker 0 as a part of negative sorting of jobs from 1000 to 500 and workers from 0 to 500).[28]

Negative sorting in this region implies that more valuable jobs feature larger fixed investments. The worker with the highest skills is paired with the job with the lowest skill demands meaning that mismatch $(z-x)$ between the worker and the job is small. Since mismatch is small and technology choice increases in mismatch, investments are small. However, the worker with the lowest skills is paired with the most valuable job. To ensure that the value of this job is not significantly diminished, a large investment is made. An example would be the case of welding (SOC 51-4120, mean wage of 45 thousand dollars) and the adoption of collaborative welding robots (cobots). A low-skill welder (with high mismatch between the skill and the job) is assisted by a cobot which ensures the quality of the output. A high-skill welder (with no mismatch between the skill and the job) instead completes the job without robotic assistance. In the right region of Figure 13, the intuition is similar for the case of surgeons (Physicians and Surgeons SOC 29-1060 with a mean wage of 213 thousand dollars is the top-ranked occupation). A low-skill surgeon is assisted by a robotic surgical system while a high-skill surgeon operates without robotic assistance. A welding company or a surgical practice makes investments in technology in order to produce valuable output with minimal mismatch costs. In the middle region of Figure 13, mismatched workers are more qualified than the job requires and firms invest in amenities for more skilled employees to reduce their utility cost of mismatch. Consider the case of semi-truck drivers (SOC 53-3030, mean wage 47 thousand dollars). A high-skill trucker drives a new truck with a quiet cabin, enhanced sleep space, and driver assistance technologies such as adaptive cruise control and digital mirrors. A low-skill trucker drives a basic model without such features. A high-skill computer user support specialist (SOC 15-1150, mean wage 64 thousand dollars) has an option to work from home using investments in remote access and

[28]The patterns of sorting are analogous in the regions 1000 to 2000 and 2000 to 3000.

control software. A low-skill computer support specialist instead works on-site in person.[29]

Equilibrium features wage dispersion within each occupation. Specifically, two distinct workers work in each occupation: one positively sorted worker in a perfect pair, and one negatively sorted worker. To illustrate, consider an occupation with skill demand 750. The perfectly positively sorted worker has skill 750. There is no mismatch for this worker. The negatively sorted worker is a worker with skill 250, which is assigned to the job within the top dashed layer of Figure 13. There is mismatch for this worker. No supplementary technology investment is necessary for a perfectly paired worker, while a worker with lower skills benefits from technology. Take the example of welders. The welder with skill 250 is assisted by a cobot to ensure that the output is not diminished due to skill mismatch. The perfectly paired welder with skill 750 does not need a cobot and instead works with a standard manual welding machine.[30]

The skill is a normalization that represents workers' rank. We now explain how sorting by skills can be linked to sorting by wages. In order to convert skills into wages we use a one-to-one mapping between workers' rank in the skill distribution and their corresponding rank in the earnings distribution. As an illustration, a worker located at the lowest point of the skill distribution (skill level 0) would earn a salary of 16 thousand dollars, which is the lowest earnings level. Similarly, a worker in the middle of the skill distribution has earnings of 47 thousand dollars (median earnings).

The key attribute of our example and our model is its ability to generate wage dispersion within the same occupation. The analysis above concluded that there is dispersion of skills in the same occupation and thus there is corresponding dispersion of wages within occupations. In contrast, a model that delivers either positive or negative sorting cannot result in variation in skill levels within a particular occupation, and as a consequence, does not create wage dispersion within that occupation. Furthermore, any model that pairs only one worker to each job does not generate any wage dispersion within occupations.

Quantitatively, our stylized example can generate 26 percent of the squared deviation and 29 percent of the absolute deviation of wages across the sample. For the middle of the distribution (20th to 80th percentiles), the model accounts for 3 percent of the squared deviation and 10 percent of the absolute

[29]For more empirical and theoretical discussion on the provision and valuation of amenities, see, amongst others, Gronberg and Reed (1994), Mas and Pallais (2017), Sorkin (2018), Wiswall and Zafar (2018), Morchio and Moser (2021), and Sockin (2022).

[30]In this example, the same job is performed by a high-skill welder and a low-skill welder assisted by a cobot. We emphasize that this situation differs from a scenario in which a high-skill welder is assigned to a more complex job (such as underwater welding requiring specialized skills) and a low-skill welder is assigned to a simpler job (such as assembling an industrial boiler). This would result in wage dispersion not within the same occupation, but rather across two different occupations: the complex and the simple occupation. Bayer and Kuhn (2023) argue empirically that differences in job execution in terms of responsibility and autonomy within the same occupation can account for a sizable portion of observed wage differences.

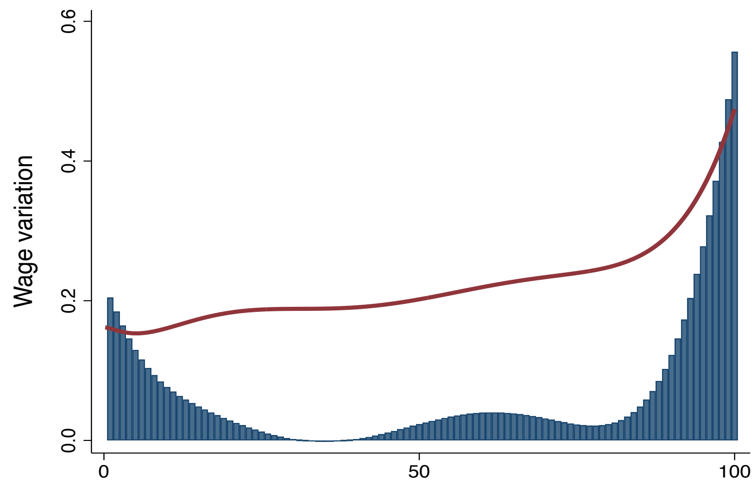


Figure 14: Wage Variation by Occupation in Model and Data

Figure 14 displays wage dispersion within occupation for the model and for the data. On the horizontal axis, we rank occupations by the average wages earned in each occupation. The red solid line is a fractional polynomial fit that captures the salient data patterns (Figure 12). The blue bars represent the variation in logarithmic wages across model occupations. Our model explains 26 percent of the squared deviation and 29 percent of the absolute deviation of wages across the sample.

deviation. In contrast, any one-to-one sorting model (for example, a model with positive or negative sorting arising from submodular or supermodular costs) would yield zero wage dispersion within occupations. For the bottom of the distribution (0th to 20th percentiles), the model accounts for 24 percent of the squared deviation and 54 percent of the absolute deviation. For the top of the distribution (80th to 100th percentiles), the model accounts for 58 percent of the squared deviation and 59 percent of the absolute deviation.

We now expand on the intuition for these quantitative results. Figure 14 displays wage dispersion within occupation for the model and for the data. We find that our stylized example delivers wage dispersion particularly in occupations with high and low mean wages. Consider a high-skill occupation such as surgeon. There is a large wage variation at the top of the income distribution. A high-skill surgeon earns a multiple of the wages of a low-skill surgeon. According to our model, a low-skill surgeon assisted by a robotic surgical system performs the same job as a high-skill surgeon. As a result, the salary variation in this profession is significant, since it encompasses low-skill surgeons as well as high-skill surgeons. A similar intuition holds at the bottom of the wage distribution. The wage distribution exhibits a lower level of dispersion at the bottom compared to the top, and the model accounts for a sizable part of this reduced level of dispersion. In contrast, the model only accounts for a moderate portion of wage dispersion in the middle of the income distribution. The reason is that the logarithmic wage profile is relatively flat in this range. As a result, even though there may be substantial skill dispersion in a particular occupation, it does not give rise to substantial wage dispersion.

6 Conclusion

We provide a complete solution to an assignment problem with heterogeneous workers and heterogeneous jobs when the underlying technology is neither supermodular nor submodular. Our analysis introduces composite sorting that yields both multiple workers sorted to the same occupation and workers being part of both positive and negative sorting. We derive that the optimal assignment is more positive when the costs of mismatch are less concave. We also show that wages have a regional hierarchical structure with relative wages determined within skill groups and aggregated to determine wages at different scales. We illustrate the composite sorting framework quantitatively to argue that it may generate a sizable portion of wage dispersion within jobs.

References

Aggarwal, A., A. Barnoy, S. Khuller, D. Kravets, and B. Schieber (1995): "Efficient Minimum Cost Matching and Transportation Using the Quadrangle Inequality," *Journal of Algorithms*, 19(1), 116–143.

Anderson, A., and L. Smith (2023): "The Comparative Statics of Sorting," University of Wisconsin-Madison Working Paper.

Antràs, P., and E. Rossi-Hansberg (2009): "Organizations and Trade," *Annual Review of Economics*, 1(1), 43–64.

Bagger, J., and R. Lentz (2019): "An Empirical Model of Wage Dispersion with Sorting," *Review of Economic Studies*, 86(1), 153–190.

Bass, R. F., and D. Khoshnevisan (1995): "Laws of the Iterated Logarithm for Local Times of the Empirical Process," *Annals of Probability*, 23(1), 388–399.

Bayer, C., and M. Kuhn (2023): "Job Levels and Wages," Bonn University Working Paper.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Birkhoff, G. (1946): "Tres Observaciones Sobre el Algebra Lineal," *Universidad Nacional de Tucumán Revista Series A*, 5, 147–154.

Boerma, J., A. Tsyvinski, and A. P. Zimin (2021): "Sorting with Team Formation," Discussion paper, NBER Working Paper No. 29290.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006): "Wage Bargaining with On-The-Job Search: Theory and Evidence," *Econometrica*, 74(2), 323–364.

Caracciolo, S., M. P. D'Achille, V. Erba, and A. Sportiello (2020): "The Dyck Bound in the Concave 1-Dimensional Random Assignment Model," *Journal of Physics A: Mathematical and Theoretical*, 53(6), 064001.

Chade, H., and J. Eeckhout (2018): "Matching Information," *Theoretical Economics*, 13(1), 377–414.

Chade, H., J. Eeckhout, and L. Smith (2017): "Sorting through Search and Matching Models in Economics," *Journal of Economic Literature*, 55(2), 493–544.

Chiappori, P.-A., C. Fiorio, A. Galichon, and S. Verzillo (2022): "Assortative Matching on Income," Columbia University Working Paper.

Chiappori, P.-A., R. J. McCann, and B. Pass (2017): "Multi-to One-Dimensional Optimal Transport," *Communications on Pure and Applied Mathematics*, 70(12), 2405–2444.

Chiappori, P.-A., and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54(3), 832–61.

Costinot, A., and J. Vogel (2015): "Beyond Ricardo: Assignment Models in International Trade," *Annual Review of Economics*, 7(1), 31–62.

Csörgő, M., Z. Shi, and M. Yor (1999): "Some Asymptotic Properties of the Local Time of the Uniform Empirical Process," *Bernoulli*, 5(6), 1035–1058.

Delon, J., J. Salomon, and A. Sobolevski (2012a): "Local Matching Indicators for Transport Problems with Concave Costs," *SIAM Journal on Discrete Mathematics*, 26(2), 801–827.

——— (2012b): "Minimum-Weight Perfect Matching for Non-Intrinsic Distances on the Line," *Journal of Mathematical Sciences*, 181(6), 782–791.

Eeckhout, J. (2018): "Sorting in the Labor Market," *Annual Review of Economics*, 10, 1–29.

Eeckhout, J., and P. Kircher (2010): "Sorting and Decentralized Price Competition," *Econometrica*, 78(2), 539–574.

——— (2018): "Assortative Matching with Large Firms," *Econometrica*, 86(1), 85–132.

Fagereng, A., L. Guiso, and L. Pistaferri (2022): "Assortative Mating and Wealth Inequality," Discussion paper, NBER Working Paper No. 29903.

Galichon, A. (2018): *Optimal Transport Methods in Economics*. Princeton University Press.

Gangbo, W., and R. J. McCann (1996): "The Geometry of Optimal Transportation," *Acta Mathematica*, 177(2), 113–161.

Gronberg, T. J., and W. R. Reed (1994): "Estimating Workers' Marginal Willingness to Pay for Job Attributes Using Duration Data," *Journal of Human Resources*, pp. 911–931.

Hwang, H.-s., D. T. Mortensen, and W. R. Reed (1998): "Hedonic Wages and Labor Market Search," *Journal of Labor Economics*, 16(4), 815–847.

Joe, H. (1997): *Multivariate Models and Multivariate Dependence Concepts*. Chapman & Hall.

Juillet, N. (2020): "On a Solution to the Monge Transport Problem on the Real Line Arising from the Strictly Concave Case," *SIAM Journal on Mathematical Analysis*, 52(5), 4783–4805.

Kantorovich, L. V. (1942): "On the Translocation of Masses," in *Dokl. Akad. Nauk. USSR*, vol. 37, pp. 227–229.

Khoshnevisan, D. (1992): "Level Crossings of the Empirical Process," *Stochastic Processes and their Applications*, 43(2), 331–343.

Koopmans, T. C., and M. Beckmann (1957): "Assignment Problems and the Location of Economic Activities," *Econometrica*, 25(1), 53–76.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108(3), 551–575.

Kremer, M., and E. Maskin (1996): "Wage Inequality and Segregation by Skill," Discussion paper, NBER Working Paper No. 5718.

Laffont, J.-J., and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94(3), 614–641.

——— (1991): "The Politics of Government Decision-Making: A Theory of Regulatory Capture," *Quarterly Journal of Economics*, 106(4), 1089–1127.

Lang, K., and S. Majumdar (2004): "The Pricing of Job Characteristics when Markets do not Clear: Theory and Policy Implications," *International Economic Review*, 45(4), 1111–1128.

Lehmann, E. L. (1966): "Some Concepts of Dependence," *The Annals of Mathematical Statistics*, 37(5), 1137–1153.

Lise, J., and F. Postel-Vinay (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110(8), 2328–76.

Lise, J., and J.-M. Robin (2017): "The Macrodynamics of Sorting between Workers and Firms," *American Economic Review*, 107(4), 1104–1135.

Mas, A., and A. Pallais (2017): "Valuing Alternative Work Arrangements," *American Economic Review*, 107(12), 3722–3759.

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.

Monge, G. (1781): "Mémoire sur la Théorie des Déblais et des Remblais," *Histoire de l'Académie Royale des Sciences de Paris*, pp. 666–704.

Morchio, I., and C. Moser (2021): "The Gender Pay Gap: Micro Sources and Macro Consequences," Columbia University Working Paper.

Nechaev, S. K., A. Sobolevski, and O. V. Valba (2013): "Planar Diagrams from Optimization for Concave Potentials," *Physical Review E*, 87(1), 1–9.

Ottolini, A., and S. Steinerberger (2023): "Greedy Matching in Optimal Transport with Concave Cost," University of Washington Working Paper.

Pegon, P., F. Santambrogio, and D. Piazzoli (2015): "Full Characterization of Optimal Transport Plans for Concave Costs," *Discrete & Continuous Dynamical Systems*, 35(12), 6113.

Porzio, T. (2017): "Cross-Country Differences in the Optimal Allocation of Talent and Technology," Columbia University Working Paper.

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Rachev, S. T., and L. Rüschendorf (1998): *Mass Transportation Problems: Volume I: Theory*. Springer Science.

Rosen, S. (1986): "The Theory of Equalizing Differences," *Handbook of Labor Economics*, 1, 641–692.

Santambrogio, F. (2015): *Optimal Transport for Applied Mathematicians*. Birkhäuser.

SATTINGER, M. (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature*, 31(2), 831–880.

SHIMER, R., AND L. SMITH (2000): "Assortative Matching and Search," *Econometrica*, 68(2), 343–369.

SOCKIN, J. (2022): "Show Me the Amenity: Are Higher-Paying Firms Better All Around?," University of Pennsylvania Working Paper.

SORKIN, I. (2018): "Ranking Firms Using Revealed Preference," *Quarterly Journal of Economics*, 133(3), 1331–1393.

STIGLER, G. (1939): "Production and Distribution in the Short Run," *Journal of Political Economy*, 47(3), 305–327.

TCHEN, A. H. (1980): "Inequalities for Distributions with Given Marginals," *Annals of Probability*, 8(4), 814–827.

TOSSAVAINEN, T. (2006): "On the Zeros of Finite Sums of Exponential Functions," *Australian Mathematical Society Gazette*, 33(1), 47.

VILLANI, C. (2003): *Topics in Optimal Transportation*, vol. 58. American Mathematical Society.

——— (2009): *Optimal Transport: Old and New*. Springer.

WERMAN, M., S. PELEG, R. MELTER, AND T. Y. KONG (1986): "Bipartite Graph Matching for Points on a Line or a Circle," *Journal of Algorithms*, 7(2), 277–284.

WISWALL, M., AND B. ZAFAR (2018): "Preference for the Workplace, Investment in Human Capital, and Gender," *Quarterly Journal of Economics*, 133(1), 457–507.

Composite Sorting

Online Appendix

Job Boerma, Aleh Tsyvinski, Ruodu Wang, and Zhenyuan Zhang

August 2023

A Proofs

In this appendix, we formally prove the results in the main text.

A.1 Equilibrium

We formally define an equilibrium for this economy.

To define an equilibrium, we specify the firm problem and the worker problem. A firm with job z employs worker x to maximize profits taking the wage schedule w as given. The firm problem is:

$$v(z) = \max_{x \in X} \big(y(x, z) - w(x)\big). \tag{A.1}$$

Taking firm compensation v as given, worker x chooses to fulfill job z to maximize their wage income:

$$w(x) = \max_{z \in Z} \big(y(x, z) - v(z)\big). \tag{A.2}$$

An equilibrium is a wage schedule w, firm value function v, and feasible assignment function π, such that firms solve their profit maximization problem (A.1), workers solve the worker problem (A.2), and a feasibility constraint is satisfied, $\int y(x, z)\,\mathrm{d}\pi = \int w(x)\,\mathrm{d}F + \int v(z)\,\mathrm{d}G$, which states that the total quantity of output produced, $\int y(x, z)\,\mathrm{d}\pi$, equals the total quantity of output distributed to workers and jobs.

A.2 Proof of Lemma 1

Suppose the assignment π and the functions (w, v) satisfy our assumptions above. Then it holds that $\int y(x, z)\,\mathrm{d}\pi = \int w(x)\,\mathrm{d}F + \int v(z)\,\mathrm{d}G$. By Monge-Kantorovich duality, we must have that the maximum for the primal problem is attained by π and the minimum for the dual problem is attained by (w, v), as required.

A.3 Maximal Perfect Pairs

In this appendix, we formally state and prove the result on maximal perfect pairs.

Lemma 2. *Maximal Perfect Pairs.* Let $F \wedge G$ denote the common component of the worker distribution F and the job distribution G. Any optimal assignment π between workers and jobs consists of perfect pairings on the support of $F \wedge G$ and an optimal assignment between workers $F - F \wedge G$ and jobs $G - F \wedge G$.

Proof. We show that a perfect pairing is made when feasible. By contradiction, suppose an optimal assignment contains pairings (x, z) and (x', z') when $x' = z$.[31]

By symmetry, it suffices to consider two cases. Consider first the case $x \leqslant z' < z = x'$. Since the cost of mismatch $\bar{c}$ is strictly increasing, $c(x, z') + c(x', z) = c(x, z') < c(x, z) \leqslant c(x, z) + c(x', z')$ where the equality follows since $c(x', z) = 0$. Thus, the cost of mismatch when making the perfect pairing is strictly lower than under the optimal configuration, which is a contradiction.

Second, we consider the case where $x < x' = z < z'$. In this case, the cost of mismatch $\bar{c}$ is given by $c(x, z') + c(x', z) = \bar{c}(z' - x)$ since $c(x', z) = 0$. To arrive at a contradiction, choose some weight $\lambda \in (0, 1)$ to scale the maximum distance such that $z - x = (1 - \lambda)(z' - x)$. Since the total distance is given by $(z - x) + (z' - x') = (z' - x)$, we also have $z' - x' = \lambda(z' - x)$. Since the cost of mismatch is strictly concave, we use strict concavity and add the two previous equations to obtain $\bar{c}(z - x) + \bar{c}(z' - x') > \bar{c}(z' - x)$.[32] The output loss can be strictly reduced by assigning worker x to job z' and by perfectly assigning worker x' to job z, which is a contradiction.[33] □

A.4 No Intersecting Pairs

In this appendix, we formally state and prove the result on no intersecting pairs.

Lemma 3. *No Intersecting Pairs.* Let π be an optimal assignment. For any two pairs (x, z) and (x', z') in the support Γ_π, their arcs do not intersect.

[31] If either $x = x'$ or $z = z'$, a perfect pairing is naturally made since $x' = z$. We thus restrict our attention to the cases where $x \neq x'$ and $z \neq z'$.

[32] We introduce the notation $\bar{c}$ to denote the cost of mismatch (5) when the worker is underqualified, or $z > x$. Similarly, we use the notation $\underline{c}$ to denote the cost of mismatch when the worker is overqualified, or $z < x$.

[33] Pegon, Santambrogio, and Piazzoli (2015) show that Lemma 2 extends with continuous worker and job distributions in higher dimensions. In these environments, the solution is partially characterized by two parts. Overlapping sets of the distributions are perfectly paired, and the remaining mass is assigned using Monge maps. In our economy featuring distributions with atoms the remaining mass is not assigned using one-to-one assignments as identical worker types fulfill different jobs. We highlight this feature of optimal sorting in the composite sorting example in Section 3.1.

Proof. To establish the result, we show that if two pairings (x, z) and (x', z') under an optimal assignment intersect, then the support of the assignment is not optimal.

By symmetry, it suffices to consider two cases. First, consider the case $x < z' < z < x'$. Since the cost function is increasing, $c(x, z') + c(x', z) = \bar{c}(z' - x) + \underline{c}(x' - z) < \bar{c}(z - x) + \underline{c}(x' - z') = c(x, z) + c(x', z')$. The output loss due to mismatch is strictly reduced by assigning worker x to job z' and worker x' to job z, which is a contradiction.

Second, consider the case $x < x' < z < z'$. In this case, the cost of mismatch is $c(x, z') + c(x', z) = \bar{c}(z' - x) + \bar{c}(z - x')$. To arrive at a contradiction, choose some weight $\lambda \in (0, 1)$ to average the minimum and maximum distance such that:

$$z - x = (1 - \lambda)(z' - x) + \lambda(z - x').$$

Since $(z - x) + (z' - x') = (z' - x) + (z - x')$, we moreover write:

$$z' - x' = \lambda(z' - x) + (1 - \lambda)(z - x').$$

Since the cost of mismatch is strictly concave, we can use strict concavity and add the two previous equations to obtain $\bar{c}(z - x) + \bar{c}(z' - x') > \bar{c}(z' - x) + \bar{c}(z - x')$. The output loss due to mismatch can be strictly reduced by assigning worker x to job z' and worker x' to job z, which is a contradiction. □

A.5 Layering

Figure A.1 illustrates the measure of underqualification. The top panel gives an example of the worker distribution F and the job distribution G. Distribution F has one worker at skill level s_1 and four workers at skill level s_3. Distribution G has two jobs at skill level s_2, a single job at s_4, and two jobs at s_5. The underqualification measure between s_1 and s_2 is $H = \frac{1}{5}$ as there is one worker at s_1 and no jobs less than s_2. Between s_2 and s_3, the underqualification measure is $H = -\frac{1}{5}$ as there are two jobs and only one worker below skill level s for each $s \in (s_2, s_3]$. The similar intuition holds for the other points yielding the measure of underqualification $H \in \{a_v\}_{v=0}^{4} = \{-\frac{1}{5}, 0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}\}$ at skill levels $\{s_i\}_{i=1}^{5}$. Generally, an underqualification measure has jumps at a finite number of skill levels $\{s_i\}$ and takes a finite number of values in $\{a_0, a_1, \ldots, a_L\}$ if the distributions of workers and jobs are discrete.

To illustrate the construction of layers we plot the blue dashed lines which indicate the different values of the measure of underqualification in the bottom panel of Figure 6. Layers are confined between two successive blue lines. There are thus four layers in this figure. The top layer (between $a_3 = \frac{2}{5}$ and $a_4 = \frac{3}{5}$) contains a worker at skill level s_3 and a job at skill level s_4. The second layer between $a_2 = \frac{1}{5}$ and $a_3 = \frac{2}{5}$

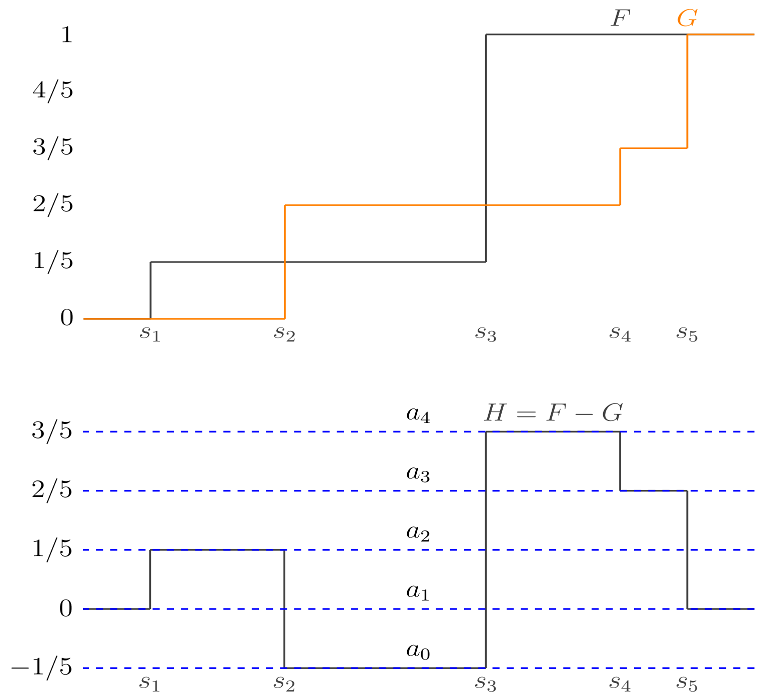


Figure A.1: Measure of Underqualification

Figure A.1 illustrates the construction of the measure of underqualification H. The top panel is an example of the worker distribution F (black line) and the job distribution G (orange line) with five workers and five jobs. Distribution F has one worker at skill level s_1 and four workers at skill level s_3. Distribution G has two jobs at skill level s_2, a single job at s_4, and two jobs at s_5. The bottom panel presents the measure of underqualification H defining the extent to which workers up to skill level s outweigh the jobs requiring skill levels up to s as $H = F - G$.

consists of one worker at skill level s_3, and one job at skill level s_5. The third layer between $a_1 = 0$ and $a_2 = \frac{1}{5}$ consists of one worker each at skill levels s_1 and s_3, and one job each at skill levels s_2 and s_5. The bottom layer (between $a_0 = -\frac{1}{5}$ and $a_1 = 0$) consists of one worker at skill level s_3 and one job at skill level s_2. We note that a worker or a job at the same skill level may generally belong to different layers. For example, there are two jobs at skill level s_5 which are split between the second layer and the third layer.

We observe that each layer consists of an alternating configuration of workers and jobs, that is, either $x_1 < z_1 < x_2 < z_2 < \cdots < x_n < z_n$ (for layers above 0) or $z_1 < x_1 < z_2 < x_2 < \cdots < z_n < x_n$ (for layers below 0). We define an alternating assignment problem as an assignment problem between n workers and n jobs, where workers and jobs are arranged in increasing order, and alternating such that every worker skill level is followed by a job difficulty level, except for the last one. Let F_ℓ and G_ℓ be the measures of the workers and the jobs in each layer.

After providing a decomposition into layers with alternating configurations, Lemma 4 shows how to solve the full assignment problem using the solutions to the assignment problem within each layer.

Lemma 4. *Layering.* Let π_ℓ be an optimal assignment between the worker distribution F_ℓ and the job distribution G_ℓ for the layer $\ell \in \{1, \dots, L\}$. Then, an optimal assignment between workers F and jobs G is their sum, $\pi := \sum_\ell \pi_\ell$.

To prove Lemma 4, we make use of the following result due to Villani (2009). We repeat the result here for completeness.

Lemma 5. *Stability of Optimal Assignment.* Let $c(x, z)$ be a continuous non-negative cost function, and $\{F_n\}_{n\in\mathbb{N}}$, $\{G_n\}_{n\in\mathbb{N}}$ be sequences of distributions of workers and jobs, respectively. Suppose $F_n \to F$, $G_n \to G$ weakly for some F, G,[34] and let π_n be an optimal assignment between F_n and G_n. If $\pi_n \to \pi$ in distribution, then π is an optimal assignment between F and G.

To establish Lemma 4, we first formally define F_ℓ and G_ℓ, the measures of the workers and the jobs in each layer. To obtain different layers for a general underqualification measure, we identify the skill levels where underqualification increases, and the skill levels where underqualification decreases. Recall that an underqualification measure H takes a finite number of values in $a_0 < a_1 < \cdots < a_L$. Underqualification increases from a_ℓ to $a_{\ell+1}$ at a skill level s if $H(s_-) \leqslant a_\ell < a_{\ell+1} \leqslant H(s)$, where s_- represents the limit from the left. Analogously, the measure of underqualification decreases from $a_{\ell+1}$ to a_ℓ at a skill level s if $H(s_-) \geqslant a_{\ell+1} > a_\ell \geqslant H(s)$. The set of skill levels where underqualification increases is denoted by $X_\ell^\uparrow := \{s\colon H \text{ increases from } a_{\ell-1} \text{ to } a_\ell \text{ at skill level } s\}$ for all $1 \leqslant \ell \leqslant L$. Similarly, the set of skill levels where underqualification decreases is denoted by $X_\ell^\downarrow := \{s\colon H \text{ decreases from } a_\ell \text{ to } a_{\ell-1} \text{ at skill level } s\}$. Define the discrete measures for all layers $1 \leqslant \ell \leqslant L$ by

$$F_\ell := (a_\ell - a_{\ell-1}) \sum_{x\in X_\ell^\uparrow} \delta_x \qquad \text{and} \qquad G_\ell := (a_\ell - a_{\ell-1}) \sum_{x\in X_\ell^\downarrow} \delta_x$$

where δ_x is the Dirac measure at x. It follows that $F = \sum F_\ell$ and $G = \sum G_\ell$. Note that F_ℓ and G_ℓ are generally not probability measures as they are not required to have total measure one.

Having defined the measures of workers and jobs in each layer, we next observe that the worker and job distributions are supported on disjoint sets and on a finite set of skills $\{s_j\}_{1\leqslant j\leqslant S}$. We smooth both the discrete distribution of workers and the discrete distribution of jobs by replacing each atom in the worker and job distribution at level s_j by a uniform distribution on $[s_j, s_j + \varepsilon]$ with the same mass for every $1 \leqslant j \leqslant S$, where ε is small enough such that the intervals $[s_j, s_j + \varepsilon]$ for all $1 \leqslant j \leqslant S$ do not

[34] This means that $F_n \to F$ on continuity points of F and respectively for G.

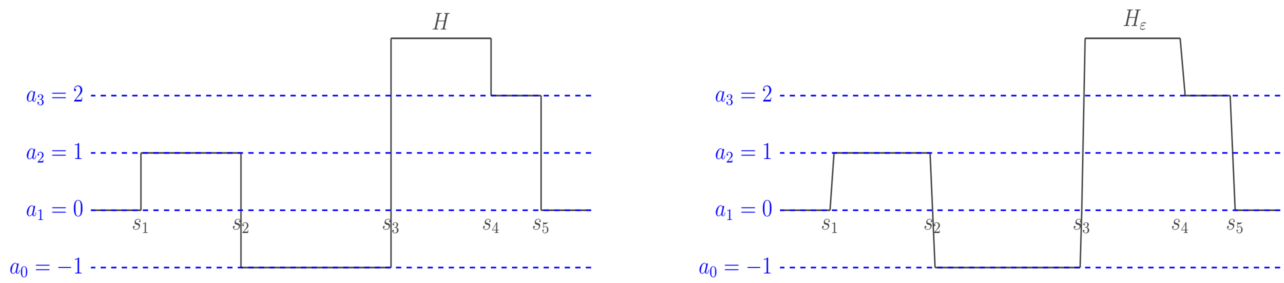


Figure A.2: Smoothed Measure of Underqualification H_ε.

Figure A.2 illustrates the smoothing of the measure of underqualification H displayed in the left panel. The corresponding smoothed measure of underqualification H_ε is presented in the right panel.

intersect. In Figure A.2, we provide an illustration of this procedure using the underqualification measure H of Figure A.1. We denote the smoothed measure of workers by F_ε and the smoothed measure of jobs by G_ε, and the corresponding underqualification measure by H_ε. An optimal assignment given worker measure F_ε and job measure G_ε is given by π_ε. Since the mismatch cost c is continuous, by stability of the optimal transport, $\pi_\varepsilon \to \pi$ weakly where π is the optimal assignment between workers F and jobs G.

Consider the support $A_\varepsilon^\ell := H_\varepsilon^{-1}((a_{\ell-1}, a_\ell))$ for all layers $1 \leqslant \ell \leqslant L$ and define the smoothed worker distribution $F_\varepsilon^\ell := F_\varepsilon\big|_{A_\varepsilon^\ell}$ and the smoothed job distribution $G_\varepsilon^\ell := G_\varepsilon\big|_{A_\varepsilon^\ell}$ for every layer such that $F_\varepsilon = \sum F_\varepsilon^\ell$ and $G_\varepsilon = \sum G_\varepsilon^\ell$.[35] Moreover, let an optimal assignment between workers F_ε^ℓ and jobs G_ε^ℓ in layer ℓ be denoted by π_ε^ℓ. Clearly, this assignment π_ε^ℓ is supported on the set $(A_\varepsilon^\ell)^2$.

Next, we establish that the sum of optimal assignments across layers $\sum \pi_\varepsilon^\ell$ is an optimal assignment between the smoothed worker distribution F_ε and the smoothed job distribution G_ε. Let π_ε be some optimal assignment between F_ε and G_ε. By cyclical monotonicity, the assignment π_ε is concentrated on a support Γ_ε that satisfies the property of no intersecting pairs. Since the smoothed distributions F_ε and G_ε are both atomless, this implies that any pairing $(x, z) \in \Gamma_\varepsilon$ where $x < z$ satisfies $F_\varepsilon([x, z]) = G_\varepsilon([x, z])$. In turn, by the definition of the measure of underqualification H, this implies $H_\varepsilon(x) = H_\varepsilon(z)$ meaning that x and z are both part of the same layer A_ε^ℓ. As a result, it follows that the support of the assignment π_ε is contained in the union of the support of all layers, or $\Gamma_\varepsilon \subseteq \bigcup (A_\varepsilon^\ell)^2$. Since all the supports $\{A_\varepsilon^\ell\}$ are disjoint, the assignment $\pi_\varepsilon\big|_{(A_\varepsilon^\ell)^2}$ transports between F_ε^ℓ and G_ε^ℓ. Since π_ε^ℓ is an optimal assignment between workers F_ε^ℓ and jobs G_ε^ℓ, it follows that the cost of mismatch for layer ℓ is greater under the assignment $\pi_\varepsilon\big|_{(A_\varepsilon^\ell)^2}$, that is, $\int c \, \mathrm{d}\pi_\varepsilon\big|_{(A_\varepsilon^\ell)^2} \geqslant \int c \, \mathrm{d}\pi_\varepsilon^\ell$. By summing over all layers $1 \leqslant \ell \leqslant L$, we can write

[35] The choice of an open or closed interval $(a_{\ell-1}, a_\ell)$ does not matter because the inverse of the boundary points is negligible with respect to the measure $F_\varepsilon + G_\varepsilon$.

that

$$\int c\,\mathrm{d}\pi_\varepsilon = \sum_{1\leqslant\ell\leqslant L} \int c\,\mathrm{d}\pi_\varepsilon\big|_{(A_\varepsilon^\ell)^2} \geqslant \sum_{1\leqslant\ell\leqslant L} \int c\,\mathrm{d}\pi_\varepsilon^\ell = \int c\,\mathrm{d}\Big(\sum_{1\leqslant\ell\leqslant L} \pi_\varepsilon^\ell\Big).$$

Since $\sum \pi_\varepsilon^\ell$ is a feasible assignment between the smoothed distributions F_ε and G_ε, and the mismatch cost is below the minimum mismatch cost, it follows that $\sum \pi_\varepsilon^\ell$ must be an optimal assignment.

To conclude the proof it follows from our construction and the stability of the optimal assignment that $\pi_\varepsilon^\ell \to \pi^\ell$ and $\sum \pi_\varepsilon^\ell \to \pi$ weakly. Thus,

$$\pi = \lim_{\varepsilon\to 0} \sum_{1\leqslant\ell\leqslant L} \pi_\varepsilon^\ell = \sum_{1\leqslant\ell\leqslant L} \lim_{\varepsilon\to 0} \pi_\varepsilon^\ell = \sum_{1\leqslant\ell\leqslant L} \pi^\ell.$$

A.6 Representation of the Bellman Algorithm

We can follow Delon, Salomon, and Sobolevski (2012b) to further simplify solving the Bellman equation. Rather than considering all potential pairings k as in equation (10), it suffices to compare only two alternatives. Specifically, the minimal mismatch cost can be described by the simple recursive equation:

$$V_{i,j} = \min(c(s_i, s_j) + V_{i+1,j-1}, V_{i,j-2} + V_{i+2,j} - V_{i+2,j-2}), \tag{A.3}$$

where $j-i$ is odd so that the assignment problem contains an equal mass of workers and jobs. We now illustrate Bellman equation (A.3). Consider, say, $V_{3,8}$ which is the cost of an optimal pairing for the skills in $[s_3, s_8]$. It is given as the minimum between two alternatives. The first alternative is $c(s_3, s_8) + V_{4,7}$. This is the cost of pairing s_3 and s_8 which is given by $c(s_3, s_8)$ and the minimal cost of pairing in $[s_4, s_7]$ given by the value $V_{4,7}$. The second alternative is $V_{3,6} + V_{5,8} - V_{5,6}$.

We use the Bellman equation (A.3), with initial conditions, to obtain the minimal cost of mismatch $V_{i,j}$ for every pair (i, j) such that $1 \leqslant i < j \leqslant 2n$.[36]

Proposition 6. An optimal assignment between workers F and jobs G sums an optimal assignment in each layer, where the optimal assignment in each layer attains $V_{1,2n_\ell}$ in the Bellman equation (A.3).

Proof of Proposition 6. The proposition is proved in two steps. First, we show the minimum aggregate cost of mismatch is described by the Bellman equation. Given the Bellman equation, we then characterize the assignment that attains the minimum cost of mismatch as discussed in the main text.

[36]For this Bellman formulation, the initial conditions are $V_{i,i-1} = 0$ and $V_{i+2,i-1} = -c(s_i, s_{i+1})$. To observe the role of initial conditions consider the assignment problems of size two and four. When $j - i = 1$, we obtain $V_{i,i+1} = \min(c(s_i, s_{i+1}) + V_{i+1,i}, V_{i,i-1} + V_{i+2,i+1} - V_{i+2,i-1}) = c(s_i, s_{i+1})$. When $j - i = 3$, we obtain $V_{i,i+3} = \min(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3} - V_{i+2,i+1}) = \min(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3})$. We note that both are indeed identical to the simple specification in equation (10).

We prove by induction that each mismatch cost $V_{i,j}$ in (A.3) represents the minimal cost of mismatching between the distribution of workers $F|_{[s_i,s_j]}$ and the distribution of jobs $G|_{[s_i,s_j]}$. Let us first check the base cases $|j-i| \leqslant 3$. Recall the initial conditions $V_{i,i-1} = 0$ and $V_{i+2,i-1} = -c(s_i, s_{i+1})$. First,

$$\begin{aligned} V_{i,i+1} &= \min\big(c(s_i, s_{i+1}) + V_{i+1,i}, V_{i,i-1} + V_{i+2,i+1} - V_{i+2,i-1}\big) \\ &= \min(c(s_i, s_{i+1}), c(s_i, s_{i+1})) = c(s_i, s_{i+1}). \end{aligned}$$

This is the correct cost of matching as there exists only a unique assignment with one worker and one job. Second,

$$\begin{aligned} V_{i,i+3} &= \min\big(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3} - V_{i+2,i+1}\big) \\ &= \min(c(s_i, s_{i+3}) + c(s_{i+1}, s_{i+2}), c(s_i, s_{i+1}) + c(s_{i+2}, s_{i+3})). \end{aligned}$$

This is the correct cost of matching since there are only two possibilities for an assignment in this case, match either the first and the fourth or the first and the second.

Suppose we know $V_{i,j}$ is the minimum cost of mismatch for every $|j-i| < 2n-1$. For the induction step, we show this also holds at $|j-i| = 2n-1$, or the complete alternating assignment problem, where:

$$V_{1,2n} = \min\big(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2}\big).$$

Denote by $F_t = F|_{[3,2n-2]}$ and $G_t = G|_{[3,2n-2]}$ the worker and job distributions on a thin support, and we let $F_a = F + F_t$ and $G_a = G + G_t$ denote the addition of the original worker and job distribution and their respective distributions on the thin support. In order to prove that the Bellman equation (A.3) holds, we show that both directions hold, that is, $V_{1,2n} \leqslant \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2})$ and $V_{1,2n} \geqslant \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2})$.

We establish the $\leqslant$ direction by showing $V_{1,2n} \leqslant c(s_1, s_{2n}) + V_{2,2n-1}$ as well as $V_{1,2n} \leqslant V_{1,2n-2} + V_{3,2n} - V_{3,2n-2}$. The first inequality follows immediately. Take the optimal assignment in the interval $[s_2, s_{2n-1}]$, and pair the remaining elements, s_1 and s_{2n}, which induces a cost of mismatch $c(s_1, s_{2n}) + V_{2,2n-1}$. Since this is a feasible assignment for all workers and jobs on the interval $[s_1, s_{2n}]$, the minimum mismatch cost on this interval is below the mismatch cost of the constructed assignment, or $V_{1,2n} \leqslant c(x_1, x_{2n}) + V_{2,2n-1}$.

To prove $V_{1,2n} + V_{3,2n-2} \leqslant V_{1,2n-2} + V_{3,2n}$, consider the assignment problem between workers F_a and jobs G_a. By Lemma 4, the minimal cost of mismatch is $V_{1,2n} + V_{3,2n-2}$. On the other hand, combining the two optimal assignments between $F|_{[1,2n-2]}$ and $G|_{[1,2n-2]}$ as well as between $F|_{[3,2n]}$ and $G|_{[3,2n]}$ gives a feasible assignment between workers F_a and jobs G_a. This shows that the optimal assignment cost between workers F_a and jobs G_a is less than $V_{1,2n-2} + V_{3,2n}$, which concludes the proof of the $\leqslant$ direction.

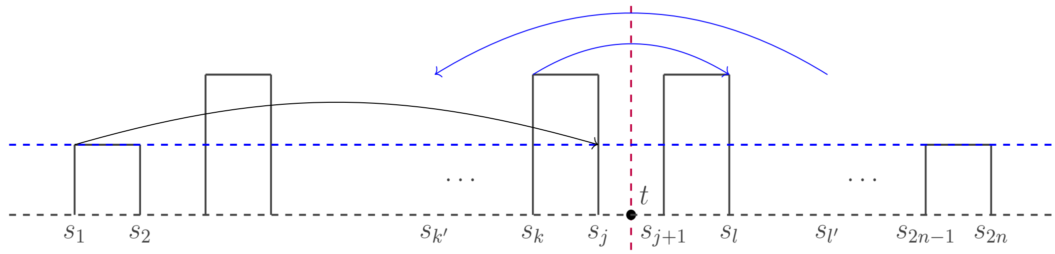


Figure A.3: Layer Decomposition with F_a and G_a

The blue dashed line indicates the layer decomposition given a worker distribution F_a and a job distribution G_a. The blue arcs indicate a nested pair in the optimal assignment π_t between a worker distribution F_t and a job distribution G_t. These blue arcs will intersect the arc corresponding to the pair (s_1, s_j) in the optimal assignment between distributions F and G, which gives a contradiction.

Next, we establish the $\geqslant$ direction, or $V_{1,2n} \geqslant \min\big(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2}\big)$. It suffices to establish that if $V_{1,2n} < c(s_1, s_{2n}) + V_{2,2n-1}$ then $V_{1,2n} + V_{3,2n-2} \geqslant V_{1,2n-2} + V_{3,2n}$.[37] Suppose $V_{1,2n} < c(s_1, s_{2n}) + V_{2,2n-1}$. This implies that an optimal assignment between workers F and jobs G does not contain the pairing (s_1, s_{2n}). Suppose that s_1 is paired with s_j where $j < 2n$. The absence of intersecting pairs implies that the optimal assignment is concentrated on square subsets to the left and right of skill s_j, that is, $\operatorname{supp} \pi \subseteq [s_1, s_j]^2 \cup [s_{j+1}, s_{2n}]^2$. We denote the skill threshold in between s_j and s_{j+1} by $t = (s_j + s_{j+1})/2$.

We next claim that an optimal assignment π_t between the thin distribution of workers F_t and the thin distribution of jobs G_t has no pairing with a corresponding interval that contains the point t. By contradiction, suppose there does exist a pair (s_k, s_l) crossing the threshold t in π_t. Since the masses to the left and the right of t are equal, there also exists a pairing $(s_{k'}, s_{l'})$ that contains (s_k, s_l). This is illustrated by the blue arrows in Figure A.3. Consider an optimal assignment π_a between the added distribution workers F_a and jobs G_a. By Lemma 4, $\pi_a = \pi + \pi_t$, and hence assignment π_a contains the pairings (s_1, s_j), (s_k, s_l), and $(s_{k'}, s_{l'})$. This contradicts the property of no intersecting pairs. Since the optimal assignment π_t does not contain a pair whose interval contains t, it follows that π_t is concentrated on the support $[s_3, s_j]^2 \cup [s_{j+1}, s_{2n-2}]^2$.

We combine the optimal assignment between workers F and jobs G with the optimal assignment between thinned distributions for workers F_t and jobs G_t to obtain the result. We construct a feasible assignment $\pi_{a,1}$ for the problem on $[s_1, s_{2n-2}]$ by combining the optimal assignment π on $[s_1, s_j]$ with the optimal assignment π_t on the interval $[s_{j+1}, s_{2n-2}]$, so that $\pi_{a,1} = \pi|_{[s_1,s_j]^2} + \pi_t|_{[s_{j+1},s_{2n-2}]^2}$. Analogously, we construct a feasible assignment $\pi_{a,2}$ for the problem on $[s_3, s_{2n}]$ by combining the optimal assignment

[37] Logically, we prove $V \geqslant \min(A, B)$ by showing that $V \geqslant B$ or $V \geqslant A$. We show this by proving the statement that if $V < A$, then $V \geqslant B$.

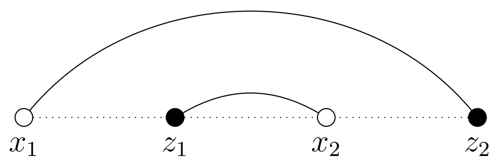


Figure A.4: Concealed Pair

Figure A.4 provides an example of a concealed pair. The pair (x_2, z_1) is concealed since the interval corresponding to this pair $[z_1, x_2]$, is contained within the interval corresponding to the pairing (x_1, z_2).

π on $[s_{j+1}, s_{2n}]$ with the optimal assignment π_t on the interval $[s_3, s_j]$ which generates $\pi_{a,2} = \pi_t|_{[s_3,s_j]^2} + \pi|_{[s_{j+1},s_{2n}]^2}$. Since the constructed assignments are feasible for the assignment problems on $[s_1, s_{2n-2}]$ and $[s_3, s_{2n}]$, respectively, we have $\mathcal{C}(\pi_{a,1}) \geqslant V_{1,2n-2}$ and $\mathcal{C}(\pi_{a,2}) \geqslant V_{3,2n}$. Finally, we observe that the addition of $\pi_{a,1}$ and $\pi_{a,2}$ solves the assignment problem with worker distribution F_a and job distribution G_a, such that $\pi_{a,1} + \pi_{a,2} = \pi_a$. These together yield $V_{1,2n-2} + V_{3,2n} \leqslant \mathcal{C}(\pi_{a,1}) + \mathcal{C}(\pi_{a,2}) = \mathcal{C}(\pi_a) = V_{1,2n} + V_{3,2n-2}$, completing the $\geqslant$ direction for (A.3). This concludes the proof.

A.7 Proof of Theorem 2

We provide a proof of Theorem 2. To do so, we first provide a formal definition of the concordance order, and then proceed to establish three intermediary results.

The distribution function π is smaller in concordance order than $\hat{\pi}$ if for any cutoff coordinate (x_c, z_c) we have $\pi((-\infty, x_c] \times (-\infty, z_c]) \leqslant \hat{\pi}((-\infty, x_c] \times (-\infty, z_c])$ and $\pi([x_c, \infty) \times [z_c, \infty)) \leqslant \hat{\pi}([x_c, \infty) \times [z_c, \infty))$. By definition, two measures are comparable in concordance order only when they have the same pair of marginal distributions, making the concordance order a natural tool to compare different assignments. For example, when π is the negative sorting or $\hat{\pi}$ is the positive sorting, $\pi \preceq \hat{\pi}$. We observe that $\preceq$ is a partial order on the probability measures with fixed marginals.

We first show another feature of an optimal sorting, which is the preservation of concealed pairs. Similar to Delon, Salomon, and Sobolevski (2012b), a pair (x, z) within an assignment is labeled concealed when the interval (x, z) is strictly contained within an interval (x', z') corresponding to some other pair (x', z') within the same assignment. Figure A.4 gives an example. The pair (x_2, z_1) is concealed since the interval (z_1, x_2) is contained within the interval corresponding to the pairing (x_1, z_2).

The next principle establishes that within each layer, every concealed pair is preserved, a term which we define precisely in the formulation of Lemma 6 following Delon, Salomon, and Sobolevski (2012b).

Lemma 6. *Preservation of Concealed Pairs.* Consider any interval $\mathcal{I}$ that has a balanced number of

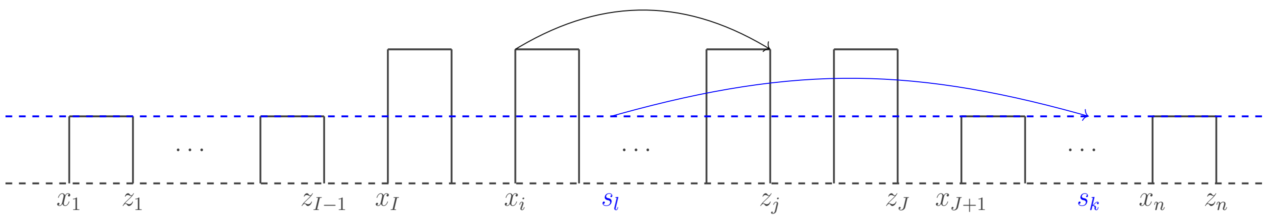


Figure A.5: Proof for Preservation of Concealed Pairs

The black arrow indicates the exposed pair (x_i, z_j) in the assignment π_1; the blue arrow indicates some pair (s_l, s_k) in the assignment π where $s_l \in (x_i, x_j)$. Since the arcs corresponding to these pairings cross, we obtain a contradiction.

workers and jobs in a layer ℓ. If, in an optimal assignment between F_ℓ and G_ℓ restricted to the interval $\mathcal{I}$, a pair (x_i, z_j) is concealed then it is optimal in the full assignment between F_ℓ and G_ℓ for that layer.

Proof. Consider the interval $\mathcal{I} = [x_I, z_J]$, the measures of workers $F_1 = F_\ell|_\mathcal{I}$ and jobs $G_1 = G_\ell|_\mathcal{I}$, and, additionally the measures of workers $F_2 = F_\ell + F_1$ and jobs $G_2 = G_\ell + G_1$, and let an optimal assignment between workers F_i and jobs G_i be given by π_i, for $i \in \{1, 2\}$. By Lemma 4, the optimal assignment between the measure of workers F_2 and the measure of jobs G_2 is the sum of the optimal assignments for each layers, or $\pi_2 = \pi_1 + \pi_\ell$.

We prove the result by contradiction. Suppose there is a concealed pair in the assignment π_1 that is not preserved in the assignment π_ℓ. This means there is at least one pair (x_i, z_j) in π_1 such that some skill level $s_l \in (x_i, z_j)$ is connected to some skill level s_k outside this interval in the assignment π_ℓ. This is represented by the blue arrow in Figure A.5. Otherwise, by replacing the assignment π_ℓ on $[x_i, z_j]$ by the assignment π_1 on $[x_i, z_j]$ decreases the total cost. This means that the corresponding intervals $[x_i, z_j]$ and $[s_l, s_k]$ intersect, which violates the property of no intersecting pairs, or equivalently, contradicts Lemma 3 applied to π_2. □

Figure A.6 illustrates the preservation of concealed pairs. The top panel shows the optimal assignment between workers, represented by hollow circles, and jobs, represented by solid dots, with skill levels in between x_2 and z_5. The assignment in the top panel features the pairs $(x_4, z_2), (x_3, z_3)$ and (x_5, z_4) that are concealed by the pair (x_2, z_5). The bottom panel shows part of the optimal assignment between all workers and jobs $\{x_i, z_i\}_{1 \leqslant i \leqslant 6}$. The preservation of concealed pairs implies that these concealed pairings in the top panel are preserved under the optimal assignment in the bottom panel.

Lemma 7. *Local Cyclical Monotonicity.* Suppose π_ℓ is an assignment between distributions F_ℓ and G_ℓ on a given layer satisfying the non-crossing property. Then π_ℓ is an optimal assignment if and only if the

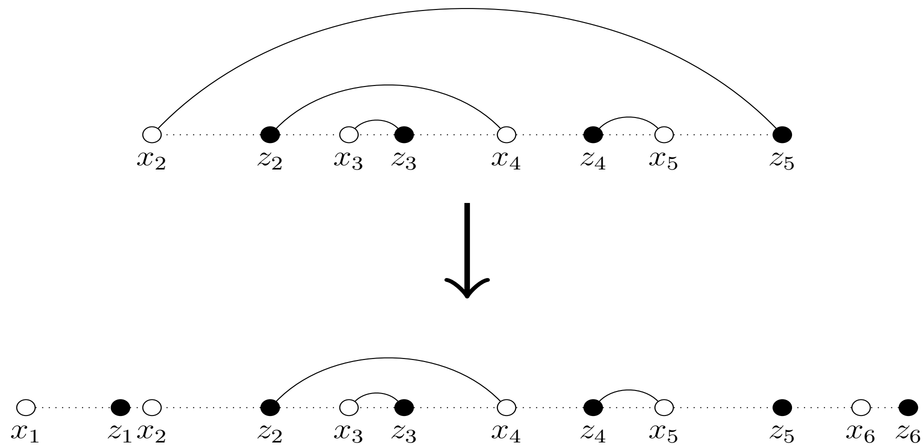


Figure A.6: Preservation of Concealed Pairs

Figure A.6 illustrates the preservation of concealed pairs of Lemma 6. The top panel shows the optimal assignment between the alternating workers (hollow circles) and jobs (solid dots) with skill levels $\{x_i, z_i\}_{2\leqslant i\leqslant 5}$, while the bottom panel shows part of the optimal assignment between all alternating workers and jobs $\{x_i, z_i\}_{1\leqslant i\leqslant 6}$. The assignment in the top panel features the pairs $(x_4, z_2), (x_3, z_3)$ and (x_5, z_4) that are concealed by pair (x_2, z_5). The preservation of concealed pairs implies that these concealed pairings are preserved under the optimal assignment in the bottom panel.

followings hold:

1. For any arc (x_0, z_0) in π_ℓ and subpairs $\{(x_i, z_i)\}_{i=1}^p$ of (x_0, z_0), π_ℓ is optimal on the assignment problem with workers $\{x_i\}_{i=0}^p$ and jobs $\{z_i\}_{i=0}^p$;
2. For exposed arcs $\{(\tilde{x}_i, \tilde{z}_i)\}_{i=1}^{\tilde{p}}$ in π_ℓ, π_ℓ is optimal on the assignment problem with workers $\{\tilde{x}_i\}_{i=0}^p$ and jobs $\{\tilde{z}_i\}_{i=0}^p$.

Proof. The "only if" direction follows directly. The remainder of the proof shows the "if" direction, where we iteratively eliminate concealed pairs to construct an optimal assignment between F_ℓ and G_ℓ, and show that such an assignment is precisely π_ℓ.

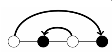
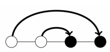
At each step of the procedure, consider pairs of the assignment π_ℓ whose subpairs contain no further subpairs. These pairs are mutually disjoint since the assignment satisfies the non-crossing property. Take such a pair (x_0, z_0) with subpairs $\{(x_i, z_i)\}_{i=1}^p$. The skill interval $\mathcal{I}_0$ formed by pairing (x_0, z_0) consists precisely of the workers x_i and jobs z_i, where $0 \leqslant i \leqslant p$. By assumption, $\pi_\ell|_{\mathcal{I}_0}$ is an optimal assignment between these workers and jobs. Observe that all pairs $\{(x_i, z_i)\}_{i=1}^p$ are therefore concealed under (x_0, z_0) in this assignment. By Lemma 6, there exists an optimal sorting between F_ℓ and G_ℓ that contains the pairs $\{(x_i, z_i)\}_{i=1}^p$. Since π_ℓ will contain the pairs $\{(x_i, z_i)\}_{i=1}^p$, we remove those pairs from our consideration, that is, we replace F_ℓ by $F_\ell \setminus \{x_i\}_{i=1}^p$ and G_ℓ by $G_\ell \setminus \{z_i\}_{i=1}^p$.

Since π_ℓ has finitely many arcs, we can continue this procedure until we have nothing left but exposed arcs $\{(\tilde{x}_i, \tilde{z}_i)\}_{i=1}^{\tilde{p}}$ in the assignment π_ℓ. By our assumption, the exposed arcs are locally optimal in π_ℓ, and hence we conclude π_ℓ is an optimal assignment between the distribution of workers F_ℓ and the distribution of jobs G_ℓ. □

Lemma 8. Suppose cost function $c(x,z)$ is of the concave form (5) and for some increasing and convex κ, $\hat{c}(x,z) = \kappa(c(x,z))$ is also of the concave form (5). On a layer ℓ, if positive sorting is optimal with cost c, then it is optimal with cost $\hat{c}$.

Proof. By contradiction, suppose positive sorting is not optimal for the cost function $\hat{c}$. Then there exists an optimal assignment for cost function $\hat{c}$ that contains negatively sorted pairs (x,z) and (x',z') where $x < x'$ and $z' < z$ and is such that

$$\hat{c}(x,z') + \hat{c}(x',z) > \hat{c}(x,z) + \hat{c}(x',z'). \tag{A.4}$$

Without loss of generality, we consider $x < z$, so that by non-crossing property and negative sorting we either have $x < z' < x' < z$ and we are in a  configuration, or we have $x < x' < z' < z$ so that we are in a  configuration. When $x < x' < z' < z$, the non-crossing principle directly contradicts that positive sorting is optimal under costs c.

Suppose $x < z' < x' < z$, and under the original cost function c, we have by assumption that

$$c(x,z') + c(x',z) \leqslant c(x,z) + c(x',z'). \tag{A.5}$$

Observe $\max(c(x,z'), c(x',z)) \leqslant c(x,z)$, and that by the optimality of negative sorting in (A.4), $\hat{c}(x',z') \leqslant \min(\hat{c}(x,z'), \hat{c}(x',z))$, and hence, $c(x',z') \leqslant \min(c(x,z'), c(x',z))$ since the function κ is increasing. By combining the two previous inequalities $c(x',z') \leqslant \min(c(x,z'), c(x',z)) \leqslant \max(c(x,z'), c(x',z)) \leqslant c(x,z)$.

To arrive at a contradiction, choose some weight $\lambda, \beta \in [0,1]$ to average the minimum and maximum cost such that:

$$c(x',z) = \lambda c(x',z') + (1-\lambda)c(x,z);$$
$$c(x,z') = \beta c(x',z') + (1-\beta)c(x,z).$$

By adding these two equalities and comparing to (A.5), it has to be true that $\lambda + \beta \geqslant 1$. Next, we use the convexity of the function κ to establish

$$\kappa(c(x',z)) = \kappa\big(\lambda c(x',z') + (1-\lambda)c(x,z)\big) \leqslant \lambda\kappa(c(x',z')) + (1-\lambda)\kappa(c(x,z));$$
$$\kappa(c(x,z')) = \kappa\big(\beta c(x',z') + (1-\beta)c(x,z)\big) \leqslant \beta\kappa(c(x',z')) + (1-\beta)\kappa(c(x,z)).$$

By adding these two inequalities, we write:

$$\hat{c}(x', z) + \hat{c}(x, z') \leqslant (\lambda + \beta)\hat{c}(x', z') + (2 - (\lambda + \beta))\hat{c}(x, z) \leqslant \hat{c}(x', z') + \hat{c}(x, z)$$

where the final equality follows since $\lambda + \beta \geqslant 1$ as well as $\hat{c}(x', z') \leqslant \hat{c}(x, z)$ because $c(x', z') \leqslant c(x, z)$. This contradicts (A.4), concluding the proof. □

Using the three Lemmas we now prove Theorem 2.

Proof of Theorem 2. Since the marginal distributions of workers F and jobs G are fixed, the optimal assignments π and $\hat{\pi}$ have the same layering structure. Hence, we fix a layer ℓ and consider the assignment problem between workers F_ℓ and jobs G_ℓ in that layer. We treat π_ℓ and $\hat{\pi}_\ell$ as assignments for layer ℓ.

If π_ℓ is an optimal assignment with cost $\hat{c}$, the result directly follows when we pick $\hat{\pi}_\ell = \pi_\ell$. Suppose instead π_ℓ is not optimal with cost $\hat{c}$. Then, by Lemma 7, there are two possibilities:

1. There exists a pair (x_0, z_0) of π_ℓ with subpairs $\{(x_i, z_i)\}_{i=1}^{p}$ such that the locally optimal assignment with workers x_i and jobs z_i, where $0 \leqslant i \leqslant p$, can be improved for cost $\hat{c}$.
2. The locally optimal assignment on the set of exposed arcs $\{(\tilde{x}_i, \tilde{z}_i)\}_{i=1}^{\tilde{p}}$ can be improved for cost $\hat{c}$.

In both cases, the assignment that improves upon π also satisfies the non-intersecting property, and has a strictly smaller total cost of mismatch. The improved assignment may still not be optimal. However, by iterating the procedure outlined in this paragraph, and since F_ℓ and G_ℓ are finite, we eventually reach an optimal assignment for the mismatch costs $\hat{c}$. In other words, step-by-step improvements on the local assignment problems gives an optimal assignment.

Before proceeding, we observe that the second case can be directly ruled out by Lemma 8. Because the set of exposed arcs $\{(\tilde{x}_i, \tilde{z}_i)\}_{i=1}^{\tilde{p}}$ is positively sorted, there is no way to improve this further since whenever positive sorting is optimal for the concave costs, positive sorting is also optimal for the less concave costs by Lemma 8. Hence, we focus on the first case in the remainder of this proof.

By the transitivity of the concordance order, it suffices to show that for each improvement step (on a pair and its subpairs), the concordance order is increased.

To show that for each improvement step, the concordance order is increased, fix a pair (x_0, z_0) with subpairs $\{(x_i, z_i)\}_{i=1}^{p}$ as in the top panel of Figure 10, and assume they form a locally optimal assignment π_c for the mismatch cost function c. Without loss of generality, take $x_0 < z_0$. Since we are working on a single layer, we must have $x_0 < z_1 < x_1 < \cdots < z_p < x_p < z_0$.

We claim that there exists some locally optimal assignment $\pi_{\hat{c}}$ for the more linear mismatch cost $\hat{c}$ with the following structure. For some $q \geqslant 1$, $\pi_{\hat{c}}$ consists of pairs $\{(\tilde{x}_i, \tilde{z}_i)\}_{i=1}^{q}$ that are not contained in any pair, and for each i, optimal sorting on the interval $(\tilde{x}_i, \tilde{z}_i)$ is positive, that is, $\pi_{\hat{c}}|_{(\tilde{x}_i,\tilde{z}_i)}$ is the positive sorting. This configuration is shown in the bottom panel of Figure 10. When this claim is true, it is straightforward to verify that $\pi_c \preceq \pi_{\hat{c}}$. Indeed, the positive sorting patterns within each interval $(\tilde{x}_i, \tilde{z}_i)$, for all $1 \leqslant i \leqslant q$ coincide under the assignments π_c and $\pi_{\hat{c}}$, and the remaining part is positive sorting for $\pi_{\hat{c}}$, where we recall that the positive sorting has the largest concordance order among all assignments.

To prove the above claim, suppose that in assignment $\pi_{\hat{c}}$, the worker x_0 is optimally paired to job z_k for some k. Consider the sorting problem between workers and jobs on the skill interval (x_0, z_k). Since π_c is optimal with the cost of mismatch c, positive sorting is locally optimal on (x_0, z_k) given the cost of mismatch c. By Lemma 8, positive sorting is also locally optimal on (x_0, z_k) for the cost $\hat{c}$. Therefore, $\pi_{\hat{c}}$ consists of positive sorting on the region (x_0, z_k).

We continue this procedure to the right, that is, we start with worker x_{k+1}, and repeat the above argument. Continuing this procedure thus yields the desired structure for $\pi_{\hat{c}}$, meaning that within each exposed arc there is positive sorting. This completes the proof. □

Comparative Statics of Sorting. In their work on the comparative statics of sorting, Anderson and Smith (2023) provides sufficient conditions for sorting to be more positive as the output function changes. However, their sufficient conditions do not apply in our setting. To show this, we introduce some of their definitions, formulate their sufficient conditions, and provide a counterexample.

A central object in their work is the difference between output under positive sorting and output under negative sorting, which they call synergy. By definition, a rectangle r is a combination of two workers (x_1, x_2) where $x_1 < x_2$ and two jobs (z_1, z_2) where $z_1 < z_2$, so that their representation in $\mathbb{R}^2$ is a rectangle. Rectangular synergy $S(r;\zeta)$ is the synergy inside the rectangle given by $y(x_1, z_1; \zeta) + y(x_2, z_2; \zeta) - y(x_1, z_2; \zeta) - y(x_2, z_1; \zeta)$, where ζ emphasizes that dependence of output on the concavity of the mismatch costs function. Summed rectangular synergy sums synergies on any finite set of disjoint rectangles. One of the assumptions that Anderson and Smith (2023) requires is that the summed rectangular synergy is up-crossing in ζ, where a function Υ is up-crossing if $\Upsilon(\zeta) \geqslant 0$ implies that $\Upsilon(\zeta') \geqslant 0$ for all $\zeta' \geqslant \zeta$.

To show that their assumption does not hold in our environment, we give an example that summed rectangular synergy is neither up-crossing nor down-crossing ($-\Upsilon$ is up-crossing) in ζ. Specifically, consider the case where the cost function (5) is symmetric in terms of the concavity of the mismatch function $\zeta = \zeta_p = \zeta_u$. Consider the example in Figure A.7 where we have a subset of three workers (x_1, x_2, x_3) and three jobs (z_1, z_2, z_3) of the alternating assignment problem within the layer, which are ordered such

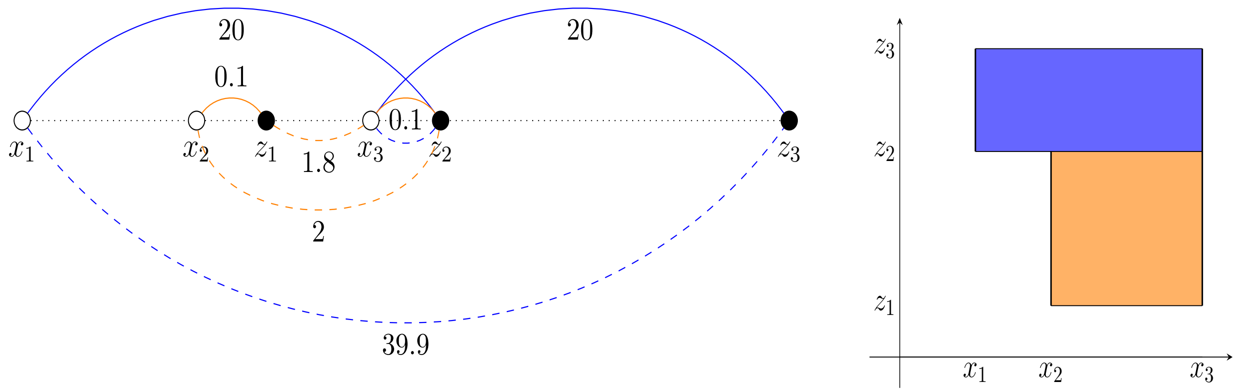


Figure A.7: Summed Rectangular Synergy is not One-Crossing with Concave Mismatch Costs

Figure A.7 shows that the condition of summed rectangular synergy is not satisfied in our setting with concave costs of mismatch. We construct two rectangles to show that summed rectangular synergy is not one-crossing. Each of the rectangles contains two workers and two jobs. The blue rectangle consists of workers (x_1, x_3) and jobs (z_1, z_3) while the orange rectangle consists of workers (x_2, x_3) and jobs (z_2, z_3). In the left panel, positive sorting for these rectangle is represented by solid arcs above the dotted line, while negative sorting is captured by blue dashed arcs below the dotted line. Distances between the paired workers and jobs are shown by the numbers on the arc (not in scale). Summed rectangular synergy equals approximately $\{0.09, -0.19, 0.30\}$ for $\zeta = \{0.2, 0.5, 0.8\}$, showing summed rectangular synergy is not one-crossing as a function of the concavity of the output function ζ.

that $x_1 < x_2 < z_1 < x_3 < z_2 < z_3$. The distances $|x - z|$ between the workers and the jobs are indicated by the numbers attached to the arcs.

We consider two distinct rectangles. The blue rectangle consists of workers (x_1, x_3) and jobs (z_1, z_3). Positive sorting for this rectangle is represented by the blue solid arcs above the dotted line, and negative sorting for this rectangle is captured by the blue dashed arcs below the dotted line. Similarly, the orange rectangle consists of workers (x_2, x_3) and jobs (z_2, z_3). Positive sorting for this rectangle is represented by the orange solid arcs above the dotted line, and negative sorting for this rectangle is captured by the orange dashed arcs below the dotted line. The rectangles are represented in $\mathbb{R}^2$ in the right panel of Figure A.7. Rectangular synergy measures the difference in costs of mismatch under positive sorting and the costs of mismatch under negative sorting in the rectangle. Summed rectangular synergies over the blue and the orange rectangles adds the synergies on the disjoint rectangles.

A numerical example shows directly that summed rectangular synergies are neither up-crossing nor down-crossing in our setting. Specifically, synergy for the blue rectangle is given by $39.9^\zeta + 0.1^\zeta - 20^\zeta - 20^\zeta$ while synergy for the orange rectangle is $1.8^\zeta + 2^\zeta - 0.1^\zeta - 0.1^\zeta$. The resulting summed synergies equal approximately $\{0.09, -0.19, 0.30\}$ for $\zeta = \{0.2, 0.5, 0.8\}$ respectively, meaning that the summed rectangular synergy is neither up-crossing nor down-crossing as a function of the concavity of the output function ζ.

A.8 Proof of Theorem 3

After maximizing perfect pairs, by Lemma 2, we can restrict attention to assignments between worker and job distributions that are supported on disjoint sets. This means that the distributions F and G are supported on a finite set S, and we denote by δ the smallest pairwise distance between elements in S, and by D the largest pairwise distance between elements.[38]

We show there exists $0 < \bar{\zeta} < 1$ such that for any $\zeta_p, \zeta_u \in [\bar{\zeta}, 1]$, the layered positive assignment π is optimal with respect to the mismatch cost $c(x, z)$. To prove the result, consider $\bar{\zeta}$ such that for any pair $(\delta', D') \in \{(\delta', D') \mid \delta \leqslant \delta' \leqslant D' \leqslant D,\ D' - \delta' \geqslant 2\delta\}$:

$$2^{1-\bar{\zeta}}(D' - \delta')^{\bar{\zeta}} \leqslant D'^{\bar{\zeta}}. \tag{A.6}$$

Such a $\bar{\zeta}$ exists because both sides of (A.6) are uniformly continuous in $\bar{\zeta}$ on $\{(\delta', D') \mid \delta \leqslant \delta' \leqslant D' \leqslant D,\ D' - \delta' \geqslant 2\delta\}$ and "$<$" holds uniformly when $\bar{\zeta} = 1$. Consider $\zeta_p, \zeta_u \in [\bar{\zeta}, 1]$. It suffices to prove that the optimal assignment within a layer does not contain any nested arc for the mismatch cost

$$c(x, z) = \begin{cases} B_p(z - x)^{\zeta_p} & \text{if } z \geqslant x; \\ B_u(x - z)^{\zeta_u} & \text{if } z < x. \end{cases} \tag{A.7}$$

By the principle of layering in Lemma 4, we decompose both measures F_n and G_n into layers. Here we assume without loss of generality that the lowest skill worker comes before the lowest skill job: $x_1 < z_1$. On each layer there are $2k$ equal masses on the skill levels $x_1 < z_1 < \cdots < x_k < z_k$ in S that are at least δ apart. The maximum distance within this layer is exceeded by $D \geqslant z_k - x_1$. Let $\{x_j\}_{1 \leqslant j \leqslant k}$ be the locations of mass on the layer for workers F_n, and let $\{z_j\}_{1 \leqslant j \leqslant k}$ be the locations of mass on the layer for jobs G_n. By contradiction, suppose that the optimal assignment within this layer instead does contain a nested arc, so it holds for some $x_1 \leqslant u < v < s < t \leqslant z_k$ – where we assume t and v are jobs and u and s are workers – that:

$$B_p(t - u)^{\zeta_p} + B_u(s - v)^{\zeta_u} \leqslant B_p(t - s)^{\zeta_p} + B_p(v - u)^{\zeta_p}.$$

By concavity of the function $x \mapsto B_p x^{\zeta_p}$ for $x \geqslant 0$,

$$B_p(t - s)^{\zeta_p} + B_p(v - u)^{\zeta_p} \leqslant 2B_p\left(\frac{1}{2}(t - s) + \frac{1}{2}(v - u)\right)^{\zeta_p} = 2^{1-\zeta_p}B_p\big((t - u) - (s - v)\big)^{\zeta_p}.$$

[38]This proof can be extended to the case of continuous distributions when F and G are compactly supported with the measure of underqualification $H := F - G$ satisfying that both H and $-H$ have finitely many local maxima and those maxima are strictly above zero.

Putting $\zeta = \zeta_p$, $t - u = D'$, and $s - v = \delta'$ in equation (A.6) leads to a contradiction. Hence the optimal assignment within this layer does not contain any nested arc for $\zeta_p \in [\bar{\zeta}, 1]$.

For the alternative case − where t and v are workers and u and s are jobs − the conclusion follows from the exact same steps, with the subscripts on B and ζ interchanged in the previous paragraph.

The implication of the proposition is that for mismatch power values close to one, the solution can be directly constructed by constructing the measure of underqualification, and constructing the positive alternating assignment by layer.[39] While this assignment generates positive sorting within each layer, we emphasize this does not imply positive sorting overall as we show in the example of two binomial distributions in Appendix A.9. In particular, there is negative sorting since x_3 is sorted with z_4 but x_4 is sorted with z_3.

A.9 Other Examples

In this appendix, we provide additional illustrations of our theory.

Reflecting Binomial Distributions. To illustrate layering, we consider the following example. Let the skill levels be given by $\{0, 1, 2, 3, 4\}$, the numbers of workers with these skill levels are given by $\{16, 32, 24, 8, 1\}$, and the numbers of jobs with these skill levels are given by $\{1, 8, 24, 32, 16\}$. This corresponds to an economy where both worker skills and job difficulties are distributed following binomial distributions, where we denote the worker skill distribution by $\mathrm{B}(n, p)$ and the job distribution by $\mathrm{B}(n, 1 - p)$ with $n = 4$ and $p = \frac{1}{3}$. In this case, the two binomial distributions are called reflecting as they have parameters p and $1 - p$.

We determine the solution to this sorting problem using the theory in Section 3. The first step is to maximize the number of perfect pairs. At skill level 0, there are 1 job and 16 workers. We perfectly pair this job with one of these 16 workers. Analogously, we perfectly pair $\{8, 24, 8, 1\}$ workers at each skill level in $\{1, 2, 3, 4\}$. After maximizing perfect pairings, we are left with 15 workers of skill level 0 and 24 workers of skill level 1 together with 24 remaining jobs of skill level 3, and 15 jobs of skill level 4.

Next, we apply the layer analysis to the remaining workers and jobs. The difference between the worker and job distribution, the measure of underqualification, is presented in the top panel of Figure A.8. In this case, we split the assignment problem into two layers. The top layer consists of 24 workers $(39 - 15)$ of skill level 1 and 24 jobs of skill level 3. The bottom layer consists of 15 workers of skill level 0

[39] Juillet (2020) calls the layered positive assignment an excursion coupling, and shows that the layered positive assignment is the limit of some optimal couplings as $\zeta \to 1^-$. We complement their result by proving the existence of a threshold $\bar{\zeta}$ beyond which the layered positive assignment is optimal for our environment.

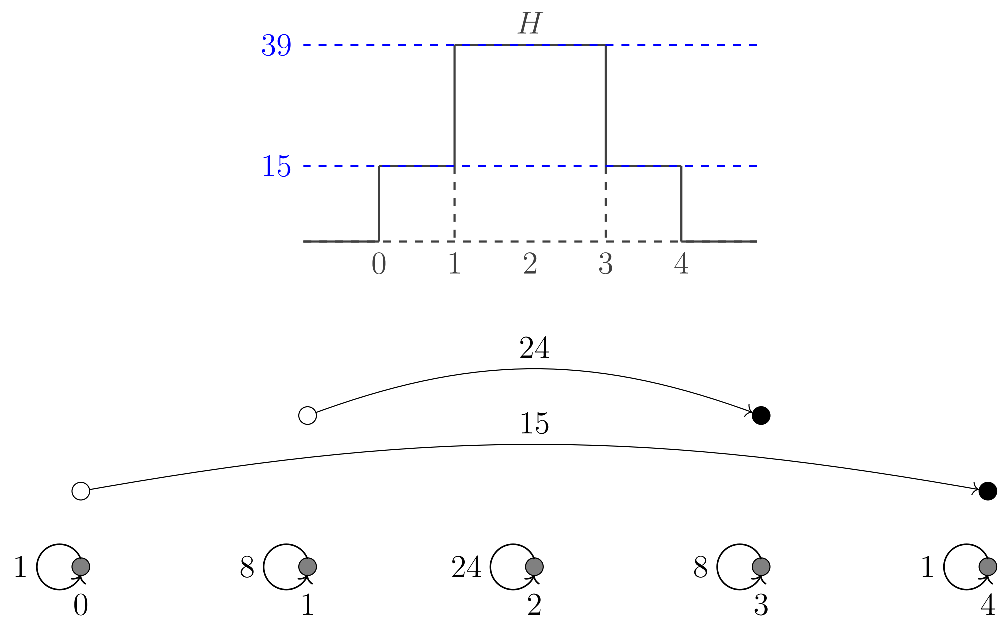


Figure A.8: Sorting with Reflecting Binomial Distributions

Figure A.8 explains composite sorting for an example with reflecting binomial distributions for workers (indicated by the hollow circles) and jobs (indicated by the dots). The top panel illustrates the measure of underqualification. The bottom panel visualizes the optimal assignment, where an arrow indicates that a given worker type is paired with a given occupation. The numbers next to the arrows indicate the number of workers of this type that are assigned to the corresponding occupation.

and 15 jobs of skill level 4. Within each layer there is only one feasible sorting. Across the layers, there is negative sorting, meaning that the best job is paired with the worst worker (bottom layer) and the worst job is paired with the top remaining worker (top layer). This negative sorting pattern within each layer is represented by the top two arcs in the bottom panel of Figure A.8. The circular arcs in the bottom of the second panel of Figure A.8 represent the optimal perfect pairing between workers and jobs to the extent that their distributions overlap.

Summarizing, the assignment problem with these reflecting binomial distributions simultaneously features both positive and negative sorting. Furthermore, we note that, as in the first example, the same job type is assigned to distinct worker types. For example, among the 32 jobs with skill level 3, 8 of them are matched perfectly with workers of skill level 3, while 24 of them are matched with workers of relatively low-skill level 1. In sum, the structure of the assignment is twofold: positive sorting of identical pairs and, for the remaining workers and jobs, pairing two points in each layer leading to negative sorting.[40]

Reflecting Binomial Mixture Distributions. We now develop a more general example that provides

[40]In fact, this result applies more broadly to any pair of worker and job distributions that lead to a unimodal measure of underqualification H.

a baseline intuition for the quantitative analysis in Section 5.

We take the previous case with the reflecting worker skill distribution and the job distribution parameterized by $p = \frac{1}{3}$ and $n = 4$, and mix these distributions with a reflecting binomial worker and job distribution with $\hat{p} = 1$ and $n = 4$ in the ratio of $3 : 1$. Again, we consider skill levels $\{0, 1, 2, 3, 4\}$ where now the number of jobs with these skill levels is $\{28, 8, 24, 32, 16\}$, and the number of workers with these skill levels is $\{16, 32, 24, 8, 28\}$.

We first maximize perfect pairs. For example, we sort 16 jobs with skill level 0 to all workers with skill level 0. Similarly, we perfectly pair $\{8, 24, 8, 16\}$ jobs with skill levels $\{1, 2, 3, 4\}$. After maximizing perfect pairs, we are left with 24 workers of skill level 1 and 12 workers of skill level 4 together with 12 remaining jobs of skill level 0, and 24 jobs of skill level 3.

Second, we decompose the problem into independent layers. The measure of underqualification H after normalization[41] at the respective skill levels is thus given by $\{-12, 12, 12, -12, 0\}$, and is depicted in Figure A.9. The top layer consists of 12 workers of skill level 1 and 12 jobs of skill level 3. The bottom layer contains four distinct elements: 12 jobs each at skill levels 0 and 3 and 12 workers each at skill levels 1 and 4.

Third, we characterize the optimal assignment for each layer independently. As in the case of the two reflecting binomial distributions, there is only one feasible sorting in the top layer. The problem for the bottom layer is identical to the example of positive and negative sorting where worker skills are represented by x_L and x_H, while jobs are represented by z_L and z_H satisfying $z_L < x_L < z_H < x_H$. The assumed distributions of skill levels imply that $d_{LL} = d_{HH} = 1, d_{LH} = 2$ and $d_{HL} = 4$. In this case, the sorting within the top layer is positive since $2 \leqslant 2^\zeta + 4^\zeta$ for all $\zeta \in (0, 1)$.

The full optimal assignment is shown in Figure A.9. Consider the jobs with skill difficulty 3. These jobs are paired with workers of skill 1, 3, and 4. Similarly, the workers with skill type 1 are paired with jobs of type 0, 1, and 3. We emphasize that, even without perfect pairing (the bottom row of the bottom panel, circular arcs), the job type with skill difficulty 3 is a part of both positive and negative pairing.

Linear Cost of Mismatch. The comparative statics with respect to the concavity of the mismatch cost function in Section 3.5 naturally beg the question what is optimal when the cost of mismatch is linear in the distance between worker skills and job difficulties. The linear case is exactly the specification of the mismatch cost prior to the technology choice in Section 2.1. This formulation is in fact the original formulation of the optimal transport problem due to Monge (1781) and has been well studied (Rachev

[41] That is, multiplied by the total number of workers or jobs, equal to 108.

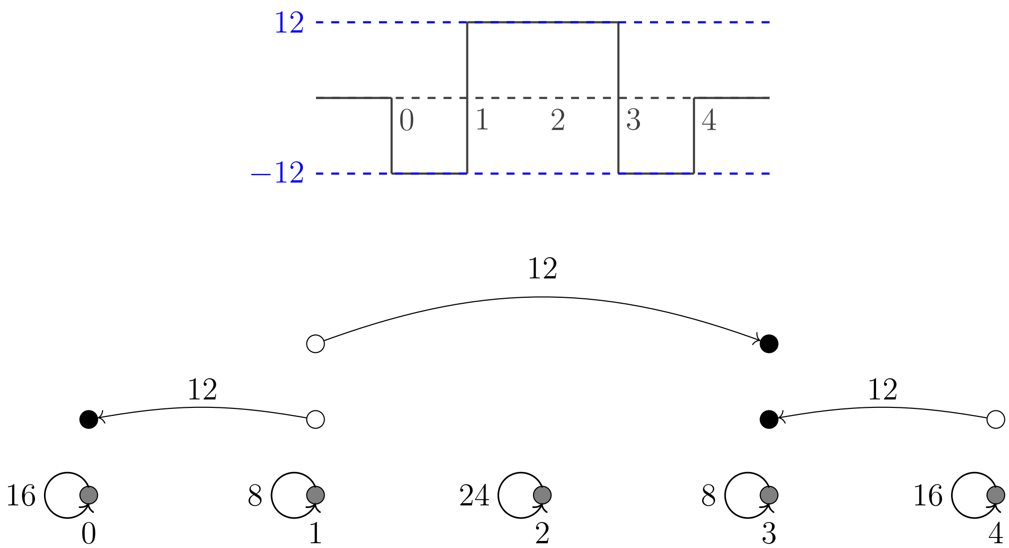


Figure A.9: Sorting with Reflecting Binomial Mixture Distributions

Figure A.9 explains composite sorting for an example with reflecting binomial mixture distributions for workers (indicated by the hollow circles) and jobs (indicated by the dots). The top panel illustrates the measure of underqualification. The bottom panel visualizes the optimal assignment, where an arrow indicates that a given worker type is paired with a given occupation. The numbers next to the arrows indicate the number of workers of this type that are assigned to the corresponding occupation.

and Rüschendorf, 1998; Villani, 2003). When the cost of mismatch is linear, an optimal assignment may not be unique. To illustrate this, we argue that both the Bellman equation and positive sorting are optimal.

To see that the Bellman equation also delivers the optimal assignment in the case of linear costs of mismatch, we note that the derivation of the Bellman equation did not require that the cost of mismatch was strictly concave, only that the cost of mismatch was concave. As a result, the same approach as before characterizes an optimal assignment.

Another solution to the assignment problem with linear costs of mismatch is positive sorting. In this case, as with convex costs of the distance, the production function is supermodular, and it follows that a positively sorted assignment is optimal.

A.10 Local Hierarchical Algorithm

We propose a new algorithm specifically tailored to the model of composite sorting, which has two distinct merits. First, it is typically more efficient than existing generic algorithms, as shown in Appendix B.4. Second, and more importantly, this new algorithm reveals a hierarchical structure of the dual potential functions, highlighting an implication of the absence of intersecting pairs to dual optimizers. This hierarchical structure means that a local dual optimizer within each region, which is the interval between two

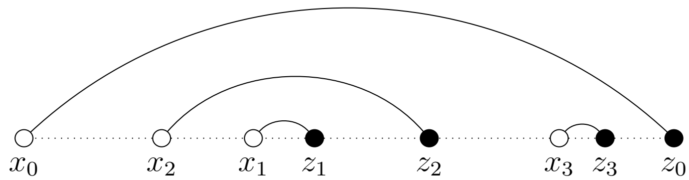


Figure A.10: Subpairs and Processing the Pairs

Figure A.10 illustrates the definition of the subpairs and the recursive computation structure in a setting of 4 workers and 4 jobs. For instance, The pair (x_1, z_1) is a subpair of (x_2, z_2) but not of (x_0, z_0). We process the pairs in the order $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 0$ to compute local dual optimizers $\phi_{[x_1,z_1]}$, $\phi_{[x_2,z_2]}$, $\phi_{[x_3,z_3]}$, and $\phi_{[x_0,z_0]}$. Note that the order of processing pairs 2 and 3 is irrelevant.

points in a pair, is computed from the local dual optimizers on its subregions. This hierarchical structure leads to the following separation property of the global dual optimizer ϕ: for any two points s and s' in one region, $\phi(s) - \phi(s')$ can be computed only based on points within this region, and thus it does not depend on points outside. We will describe this structure in detail and our algorithm below.

Recall that in our setting, we consider a problem with n workers with skill levels in X and n jobs with difficulty levels in Z, where X and Z are disjoint sets. We denote by $S = X \cup Z$ the set of all skill levels. Moreover, we recall that we construct the dual solution given an optimal sorting, or primal solution, π.

Our algorithm relies on recursive computations of ϕ constrained on smaller subsets of S. To explain such a recursive procedure, we need some preparation by introducing the notion of subpairs. A pair (x, z) is called a subpair of the pair (x_0, z_0) if (x, z) is a non-nested pair inside the interval $[x_0, z_0]$ that is not equal to (x_0, z_0). In Figure A.10, the pairs (x_2, z_2) and (x_3, z_3) are subpairs of the pair (x_0, z_0). The pair (x_1, z_1) is a subpair of (x_2, z_2) but not of (x_0, z_0) as it is nested in (x_2, z_2). For $(x, z) \in \Gamma_\pi$, let $X_{[x,z]}$ denote the set of all points in X between x and z inclusive of the boundary, and similarly for $Z_{[x,z]}$.

We process each pair $(x_0, z_0) \in \Gamma_\pi$ sequentially in a certain order described below to get a local dual optimizer on $[x_0, z_0]$, that is, a function $\phi_{[x_0,z_0]}$ such that for any $(x, z) \in X_{[x_0,z_0]} \times Z_{[x_0,z_0]}$, $\phi_{[x_0,z_0]}(x) - \phi_{[x_0,z_0]}(z) \leqslant c(x, z)$ which holds with equality when $(x, z) \in \Gamma_\pi$. We observe that this property is preserved if $\phi_{[x_0,z_0]}$ is shifted by any constant $a \in \mathbb{R}$.

A simple illustrative example is given in Figure A.10. We start with the pair (x_1, z_1) and found a local dual optimizer $\phi_{[x_1,z_1]}$. Here, the local dual optimizer is any function $\phi_{[x_1,z_1]}$ satisfying $\phi_{[x_1,z_1]}(x_1) - \phi_{[x_1,z_1]}(z_1) = c(x_1, z_1)$ as the only pair contained in $[x_1, z_1]$ is (x_1, z_1). We then process the pair (x_2, z_2). A local dual optimizer on $[x_2, z_2]$ is a function $\phi_{[x_2,z_2]}$ satisfying $\phi_{[x_2,z_2]}(x) - \phi_{[x_2,z_2]}(z) \leqslant c(x, z)$ for $(x, z) \in \{(x_1, z_1), (x_2, z_2), (x_1, z_2), (x_2, z_1)\}$ with equality holding for $(x, z) = (x_1, z_1)$ and $(x, z) = (x_2, z_2)$. We proceed with the subpair (x_3, z_3) and finally the remaining pair (x_0, z_0).

Below is a full description of the algorithm, where ϕ, ϕ_i, x_i and z_i are local variables which vary across each iteration, and $\phi_{[x,z]}$ for $(x,z) \in \Gamma_\pi$ are global variables which are the output of the algorithm.

1. Pick any pair $(x_0, z_0) \in \Gamma_\pi$ that has not been processed such that all subpairs of (x_0, z_0) have been processed. Let $(x_1, z_1), \ldots, (x_p, z_p)$ be the subpairs ordered in dictionary order[42] with potential functions $\phi_i := \phi_{[x_i, z_i]}$ for $i = 1, \ldots, p$.
2. If $p = 0$, then let $\phi(z_0) = 0$ and $\phi(x_0) = c(x_0, z_0)$.
3. If $p \geqslant 1$, then continue with the following sub-steps.
   (a) If $p > 1$, let $(\beta_2, \ldots, \beta_p) \in \mathbb{R}^{p-1}$ be a solution to the inequality system

$$\max(c_{00} - c_{0n} - c_{m0}, -c_{mn}) + c_{nn} \leqslant \sum_{k=n+1}^{m} \beta_k \leqslant \min(c_{0m} + c_{n0} - c_{00}, c_{nm}) - c_{mm} \tag{A.8}$$

   for all $1 \leqslant n < m \leqslant p$, where $c_{ij} = c(x_i, z_j)$ for $i, j \in \{1, \ldots, p\}$. We show the existence of such $(\beta_2, \ldots, \beta_p)$ in Lemma 9 below.

   (b) For $i = 1, \ldots, p$, let

$$\phi(s) = \phi_i(s) + \sum_{k=i+1}^{p} \beta_k + \phi_p(x_p) - \phi_i(x_i) \tag{A.9}$$

   for $s \in X_{[x_i, z_i]}$ or $s \in Z_{[x_i, z_i]}$. The above sum $\sum_{k=i+1}^{p} \beta_k$ is 0 if $i = p$.

   (c) Define $\phi(x_0)$ and $\phi(z_0)$ by the following equations

$$\phi(z_0) = \begin{cases} \max\limits_{i \in \{1, \ldots, p\}} (\phi(x_i) - c(x_i, z_0)) & \text{if } x_1 \neq x_0, \\ \min\limits_{i \in \{1, \ldots, p\}} (\phi(z_i) + c(x_0, z_i)) - c(x_0, z_0) & \text{elsewhere;} \end{cases} \tag{A.10}$$

$$\phi(x_0) = \phi(z_0) + c(x_0, z_0). \tag{A.11}$$

4. Let $\phi_{[x_0, z_0]}$ be equal to ϕ.
5. Return to step 1 with the next pair to process, or terminate if all pairs have been processed.

We note from step 3(b) that for $s, s' \in X_{[x_i, z_i]} \cup Z_{[x_i, z_i]}$, we have $\phi(s) - \phi(s') = \phi_i(s) - \phi_i(s')$. This means that after many iterations, the value of $\phi(s) - \phi(s')$ does not change, and therefore it depends only on points in the region $[x_i, z_i]$.

[42] A vector $(a_1, \ldots, a_d)$ is smaller than or equal to a vector $(b_1, \ldots, b_d)$ in dictionary order if either $(a_1, \ldots, a_d) = (b_1, \ldots, b_d)$ or there exists $p = 1, \ldots, d$ such that $a_i = b_i$ for $i = 1, \ldots, p-1$ and $a_{p+1} < b_{p+1}$.

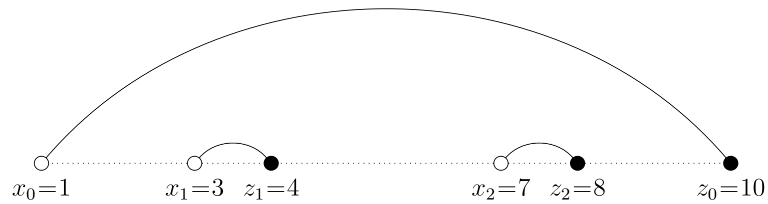


Figure A.11: Illustration of the Algorithm

Figure A.11 illustrates the construction of the dual potential ϕ in the example where $x_0 = 1,\ x_1 = 3,\ x_2 = 7,\ z_0 = 10,\ z_1 = 4,\ z_2 = 8$, with the cost function given by $c(x, z) = \sqrt{|x - z|}$. The pairs are processed in the order of $(x_1, z_1),\ (x_2, z_2),\ (x_3, z_3)$. The output of the algorithm is $\phi(x_1) = 5 - 2\sqrt{3},\ \phi(z_1) = 4 - 2\sqrt{3},\ \phi(x_2) = 1,\ \phi(z_2) = 0,\ \phi(x_0) = 4 - \sqrt{3}$, and $\phi(z_0) = 1 - \sqrt{3}$.

There is some flexibility in the above algorithm in choosing the order of processing the pairs in step 1 and in solving for $(\beta_2, \ldots, \beta_p)$ in (A.8) in step 3(a). This flexibility is natural as the dual optimizers are not unique in general. The order of processing the pairs does not affect the output of the algorithm because each $\phi_{[x_0,z_0]}$ only depends on the local dual optimizers of its subpairs, which are all processed before this pair. A default order is to always choose the unprocessed pair (x_0, z_0) with the smallest x_0 satisfying the condition in step 1. On the other hand, the choice of $(\beta_2, \ldots, \beta_p)$ does affect the output of the algorithm. As a default, $(\beta_2, \ldots, \beta_p)$ can be chosen as the solution of (A.8) which is the smallest in dictionary order.[43] In this way, we obtain a unique output of the algorithm. Nevertheless, in the next result, we will show that a dual potential ϕ is obtained from the algorithm with arbitrary choices of $(\beta_2, \ldots, \beta_p)$ satisfying (A.8) in each iteration.

We illustrate the construction of the dual potential ϕ in a simple example where $X = \{1, 3, 7\},\ Z = \{4, 8, 10\}$, and $x_0 = 1,\ x_1 = 3,\ x_2 = 7,\ z_0 = 10,\ z_1 = 4,\ z_2 = 8$, with the cost function given by $c(x, z) = \sqrt{|x - z|}$. By the principles established in Section 3, the optimal assignment π pairs x_i with z_i for all $i = \{0, 1, 2\}$ as displayed in Figure A.11. Our goal is to construct a function $\phi : S \to \mathbb{R}$ such that $\phi(x_i) - \phi(z_j) \leqslant c(x_i, z_j)$ for all i and j in $\{0, 1, 2\}$ with equality holding when $i = j$.

Following step 1 of the algorithm, we first process the pairs (x_1, z_1) and (x_2, z_2), as they do not have any subpair. Define $\phi_1(x_1) = \phi_2(x_2) = 1$ and $\phi_1(z_1) = \phi_2(z_2) = 0$, which are the local dual optimizers on $[x_1, z_1]$ and $[x_2, z_2]$, respectively. We now process the pair (x_0, z_0). According to (A.8), the second step is to find $\beta_2 \in \mathbb{R}$ such that

$$\max(c_{00} - c_{01} - c_{20}, -c_{21}) + c_{11} \leqslant \beta_2 \leqslant \min(c_{02} + c_{10} - c_{00}, c_{12}) - c_{22}. \tag{A.12}$$

[43]Note that such a smallest solution always exists since $(\beta_2, \ldots, \beta_p)$ satisfying (A.8) lies in a compact region.

Plugging in the values of x_i, z_i, we get

$$\max(c_{00} - c_{01} - c_{20}, -c_{21}) + c_{11} = 4 - 2\sqrt{3} \approx 0.536; \quad \min(c_{02} + c_{10} - c_{00}, c_{12}) - c_{22} = \sqrt{3} - 1 \approx 0.732.$$

Following the default choice of choosing the smallest β_2, we set $\beta_2 = 4 - 2\sqrt{3}$.[44] Following step 3(b) of the algorithm, we have

$$\begin{aligned}
&\phi(x_2) = \phi_2(x_2) + 0 + \phi_2(x_2) - \phi_2(x_2) = c_{22} = 1;\\
&\phi(z_2) = \phi_2(z_2) + 0 + \phi_2(x_2) - \phi_2(x_2) = 0;\\
&\phi(x_1) = \phi_1(x_1) + \beta_2 + \phi_2(x_2) - \phi_1(x_1) = c_{22} + \beta_2 = 5 - 2\sqrt{3};\\
&\phi(z_1) = \phi_1(z_1) + \beta_2 + \phi_2(x_2) - \phi_1(x_1) = c_{22} - c_{11} + \beta_2 = 4 - 2\sqrt{3}.
\end{aligned}$$

Further, step 3(c) yields

$$\begin{aligned}
&\phi(z_0) = \max(\phi(x_1) - c_{10}, \phi(x_2) - c_{20}) = 1 - \sqrt{3};\\
&\phi(x_0) = \phi(z_0) + c_{00} = 4 - \sqrt{3}.
\end{aligned}$$

One can verify numerically that the function ϕ defined above satisfies the conditions of a dual optimizer.

A.11 Proof of Theorem 4

We prove Theorem 4 in two parts. First, we prove there exists a solution $(\beta_2, \ldots, \beta_p)$ to (A.8), in Lemma 9. Second, we prove that the function ϕ defined in (A.9)-(A.11) is indeed a local dual optimizer on $S_{[x_0,z_0]}$.

Lemma 9. Suppose $(x_1, z_1), \ldots, (x_p, z_p)$ are ordered subpairs of pair (x_0, z_0) in the optimal assignment π. Then the system of inequalities, where for all $1 \leqslant n < m \leqslant p$:

$$\max(c_{00} + c_{nn} - c_{0n} - c_{m0}, c_{nn} - c_{mn}) \leqslant \sum_{k=n+1}^{m} \beta_k \leqslant \min(c_{0m} + c_{n0} - c_{00} - c_{mm}, c_{nm} - c_{mm}) \tag{A.13}$$

admits a solution $(\beta_2, \ldots, \beta_p)$.

Proof of Lemma 9. We use Farkas' Lemma to prove our existence result, specifically, to have a necessary and sufficient condition for a system of linear inequalities to have a solution. We state Farkas' Lemma for completeness.

[44] Any value in between $4 - 2\sqrt{3}$ and $\sqrt{3} - 1$ will produce a dual potential. That the left-hand side of (A.12) is no more than the right-hand side of (A.12) is not a coincidence. We will prove that the system of inequalities (A.8) always admits a solution in Lemma 9 in Appendix A.10. The set of solutions is always non-empty, although, intuitively, the larger $|S|$ is, the less freedom for the dual optimizer one has.

Lemma 10. *Farkas.* Let A be a $d_1 \times d_2$ matrix, $b \in \mathbb{R}^{d_1}$, and let $x = (x_1, \ldots, x_{d_2})^\top$ be a set of real-valued variables. Then the system $Ax \geqslant b$ allows a set of solutions if and only if for any $y \in [0, \infty)^{d_1}$ such that $y^\top A = 0$, it holds $y^\top b \leqslant 0$.

We aim to show that equation (A.13) admits a solution $(\beta_2, \ldots, \beta_p) \in \mathbb{R}^{p-1}$. We observe that we can think of (A.13) equivalently as the following set of inequalities:

$$\begin{aligned}
\sum_{k=n+1}^{m} \beta_k &\geqslant c_{00} + c_{nn} - c_{0n} - c_{m0} \\
\sum_{k=n+1}^{m} \beta_k &\geqslant c_{nn} - c_{mn} \\
-\sum_{k=n+1}^{m} \beta_k &\geqslant c_{mm} - c_{nm} \\
-\sum_{k=n+1}^{m} \beta_k &\geqslant c_{00} + c_{mm} - c_{0m} - c_{n0}
\end{aligned}$$

for all $1 \leqslant n < m \leqslant p$. All inequalities implied by (A.13) are thus linear in the variables $(\beta_2, \ldots, \beta_p)$. Matrix A is given by columns with values $(-1, 0, +1)$, while vector b is governed by the costs c.

By Lemma 10 it suffices to prove the following.[45] For any set of non-negative weights $(\lambda^+_{mn}, \lambda^-_{mn}, \omega^+_{mn}, \omega^-_{mn})$, $1 \leqslant n < m \leqslant p$ on each of the inequalities above such that

$$\sum_{1 \leqslant n < m \leqslant p} (\lambda^+_{mn} + \omega^+_{mn}) \sum_{k=n+1}^{m} \beta_k = \sum_{1 \leqslant n < m \leqslant p} (\lambda^-_{m,n} + \omega^-_{m,n}) \sum_{k=n+1}^{m} \beta_k, \text{ for all } (\beta_2, \ldots, \beta_p), \tag{A.14}$$

it holds that

$$\begin{aligned}
&\sum_{1 \leqslant n < m \leqslant p} \Big(\lambda^-_{mn}(c_{00} + c_{nn} - c_{0n} - c_{m0}) + \omega^-_{mn}(c_{nn} - c_{mn}) \Big) \\
&\qquad \leqslant \sum_{1 \leqslant n < m \leqslant p} \Big(\lambda^+_{mn}(c_{0m} + c_{n0} - c_{00} - c_{mm}) + \omega^+_{mn}(c_{nm} - c_{mm}) \Big).
\end{aligned} \tag{A.15}$$

We start by simplifying equations (A.14) and (A.15). We first simplify equation (A.14). Since (A.14) has to hold for all $(\beta_2, \ldots, \beta_p)$, we note that the coefficient on each β_k has to equal zero. For each $2 \leqslant k \leqslant p$, equating the coefficients for β_k requires

$$\sum_{m,n} \big(\lambda^+_{mn} + \omega^+_{mn}\big) = \sum_{m,n} \big(\lambda^-_{mn} + \omega^-_{mn}\big), \tag{A.16}$$

[45] Equation (A.14) is the analog of $y^\top A = 0$ in the statement of Farkas' Lemma. Specifically, we use $y^\top A = 0$ if and only if $y^\top A z = 0$ for all $z \in \mathbb{R}^{d_1}$. Applied to our setting, where β takes the position of x in Farkas' Lemma, this states that the weighted sum of all left-hand sides in the system of inequalities equals zero. Equation (A.15) below is similarly the analog of $y^\top b \leqslant 0$ in the statement of Farkas' Lemma.

where we sum over all (m, n) satisfying $1 \leqslant n < k \leqslant m \leqslant p$, that is, we sum over all equations where β_k appears. Furthermore, subtracting equation (A.16) evaluated at k from equation (A.16) evaluated at $k+1$ yields:

$$\sum_{k<m\leqslant p} (\lambda^+_{mk} + \omega^+_{mk}) - \sum_{k<m\leqslant p} (\lambda^-_{mk} + \omega^-_{mk}) = \sum_{1\leqslant n<k} (\lambda^+_{kn} + \omega^+_{kn}) - \sum_{1\leqslant n<k} (\lambda^-_{kn} + \omega^-_{kn}), \tag{A.17}$$

for all $2 \leqslant k < p$.

We next simplify (A.15). Rearranging (A.15) by collecting terms by coefficients in front of c_{ij} leads to the equivalent form:

$$\begin{aligned}
&\sum_{1\leqslant n<m\leqslant p} \big(\lambda^-_{mn} + \lambda^+_{mn}\big) c_{00} + \sum_{1<m\leqslant p} \big(\lambda^-_{m1} + \omega^-_{m1}\big) c_{11} + \sum_{1\leqslant n<p} \big(\lambda^+_{pn} + \omega^+_{pn}\big) c_{pp} \\
&\qquad + \sum_{k=2}^{p-1} \Big(\sum_{k<m\leqslant p} \big(\lambda^-_{mk} + \omega^-_{mk}\big) + \sum_{1\leqslant n<k} \big(\lambda^+_{kn} + \omega^+_{kn}\big) \Big) c_{kk} \\
&\leqslant \sum_{1<m\leqslant p} \lambda^-_{m1} c_{01} + \sum_{1\leqslant n<p} \lambda^+_{pn} c_{0p} + \sum_{1\leqslant n<p} \lambda^-_{pn} c_{p0} + \sum_{1<m\leqslant p} \lambda^+_{m1} c_{10} \\
&\qquad + \sum_{k=2}^{p-1} \Big(\sum_{1\leqslant n<k} \lambda^+_{kn} + \sum_{k<m\leqslant p} \lambda^-_{mk} \Big) c_{0k} + \Big(\sum_{1\leqslant n<k} \lambda^-_{kn} + \sum_{k<m\leqslant p} \lambda^+_{mk} \Big) c_{k0} \\
&\qquad + \sum_{1\leqslant n<m\leqslant p} \omega^-_{mn} c_{mn} + \sum_{1\leqslant n<m\leqslant p} \omega^+_{mn} c_{nm},
\end{aligned} \tag{A.18}$$

where the left-hand side of the inequality collects all "diagonal" elements, and the right-hand side collects all other elements.

Our next step in proving that equation (A.18) indeed holds, is to show that both sides of equation (A.18) represent transport costs of an assignment between a measure of workers $\tilde{F}$ and a measure of jobs $\tilde{G}$. Specifically, consider the assignment problem between a measure $\tilde{F}$ and a measure $\tilde{G}$, satisfying:

$$\begin{aligned}
\tilde{F} = &\sum_{1\leqslant n<m\leqslant p} \big(\lambda^-_{mn} + \lambda^+_{mn}\big) \delta_{x_0} + \sum_{1<m\leqslant p} \big(\lambda^-_{m1} + \omega^-_{m1}\big) \delta_{x_1} + \sum_{1\leqslant n<p} \big(\lambda^+_{pn} + \omega^+_{pn}\big) \delta_{x_p} \\
&\qquad + \sum_{k=2}^{p-1} \Big(\sum_{k<m\leqslant p} \big(\lambda^-_{mk} + \omega^-_{mk}\big) + \sum_{1\leqslant n<k} \big(\lambda^+_{kn} + \omega^+_{kn}\big) \Big) \delta_{x_k},
\end{aligned} \tag{A.19}$$

and, similarly,

$$\begin{aligned}
\tilde{G} = &\sum_{1\leqslant n<m\leqslant p} \big(\lambda^-_{mn} + \lambda^+_{mn}\big) \delta_{z_0} + \sum_{1<m\leqslant p} \big(\lambda^-_{m1} + \omega^-_{m1}\big) \delta_{z_1} + \sum_{1\leqslant n<p} \big(\lambda^+_{pn} + \omega^+_{pn}\big) \delta_{z_p} \\
&\qquad + \sum_{k=2}^{p-1} \Big(\sum_{k<m\leqslant p} \big(\lambda^-_{mk} + \omega^-_{mk}\big) + \sum_{1\leqslant n<k} \big(\lambda^+_{kn} + \omega^+_{kn}\big) \Big) \delta_{z_k}.
\end{aligned} \tag{A.20}$$

Both measures may not be probability measures, but they do have the same total mass.

The fact that the left-hand side of (A.18) represents a transport cost between workers $\tilde{F}$ and jobs $\tilde{G}$ is evident. Under this assignment each worker type is assigned to an identically indexed job, which have identical masses by construction of the worker distribution $\tilde{F}$ in (A.19) and the job distribution $\tilde{G}$ in (A.20). To establish the same on the right-hand side requires work. Consider first the worker x marginal on the right-hand side of (A.18).

1. The mass on x_0 is
$$\sum_{1<m\leqslant p}\lambda^-_{m1}+\sum_{1\leqslant n<p}\lambda^+_{pn}+\sum_{k=2}^{p-1}\Big(\sum_{1\leqslant n<k}\lambda^+_{kn}+\sum_{k<m\leqslant p}\lambda^-_{mk}\Big)=\sum_{1\leqslant n<m\leqslant p}\Big(\lambda^-_{mn}+\lambda^+_{mn}\Big).$$
2. Using equation (A.16) with $k=2$, the mass on x_1 is
$$\sum_{1<m\leqslant p}\lambda^+_{m1}+\sum_{1<m\leqslant p}\omega^+_{m1}=\sum_{1<m\leqslant p}\big(\lambda^-_{m1}+\omega^-_{m1}\big).$$
3. For $2\leqslant k<p$, using (A.17) and grouping terms, the mass on x_k is
$$\sum_{1\leqslant n<k}\lambda^-_{kn}+\sum_{k<m\leqslant p}\lambda^+_{mk}+\sum_{1\leqslant n<k}\omega^-_{k,n}+\sum_{k<m\leqslant p}\omega^+_{mk}=\sum_{k<m\leqslant p}(\lambda^-_{mk}+\omega^-_{mk})+\sum_{1\leqslant n<k}(\lambda^+_{kn}+\omega^+_{kn}).$$
4. Using equation (A.16) with $k=p$, the mass on x_p is
$$\sum_{1\leqslant n<p}\big(\lambda^-_{pn}+\omega^-_{pn}\big)=\sum_{1\leqslant n<p}\big(\lambda^+_{pn}+\omega^+_{pn}\big).$$

Combining these four terms we see that the x-marginal of the right-hand side of (A.18) corresponds with that of (A.19). We proceed to show that the same is true for the distribution of jobs.

1. The mass on z_0 is
$$\sum_{1\leqslant n<p}\lambda^-_{pn}+\sum_{1<m\leqslant p}\lambda^+_{m1}+\sum_{k=2}^{p-1}\Big(\sum_{1\leqslant n<k}\lambda^-_{kn}+\sum_{k<m\leqslant p}\lambda^+_{mk}\Big)=\sum_{1\leqslant n<m\leqslant p}(\lambda^-_{mn}+\lambda^+_{mn}),$$
where the equality follows by simple accounting.
2. The mass on z_1 is
$$\sum_{1<m\leqslant p}\lambda^-_{m1}+\sum_{1<m\leqslant p}\omega^-_{m1}=\sum_{1<m\leqslant p}\big(\lambda^-_{m1}+\omega^-_{m1}\big).$$
3. For $2\leqslant k<p$, the mass on z_k is
$$\sum_{1\leqslant n<k}\lambda^+_{kn}+\sum_{k<m\leqslant p}\lambda^-_{mk}+\sum_{k<m\leqslant p}\omega^-_{mk}+\sum_{1\leqslant n<k}\omega^+_{kn}=\sum_{k<m\leqslant p}\big(\lambda^-_{mk}+\omega^-_{mk}\big)+\sum_{1\leqslant n<k}\big(\lambda^+_{kn}+\omega^+_{kn}\big).$$

4. Finally, the mass on z_p is given by

$$\sum_{1\leqslant n<p} \lambda_{pn}^{+} + \sum_{1\leqslant n<p} \omega_{pn}^{+} = \sum_{1\leqslant n<p} (\lambda_{pn}^{+} + \omega_{pn}^{+}).$$

We have thus proved the marginal distributions on both sides of the costs (A.18) are the worker distribution $\tilde{F}$ and the job distribution $\tilde{G}$.

Why is the left-hand side of equation (A.18) the optimal transportation cost between the worker distribution $\tilde{F}$ and job distribution $\tilde{G}$? To characterize an optimal assignment between the constructed measures $\tilde{F}$ and $\tilde{G}$, we decompose the corresponding measure of underqualification $\tilde{H} := \tilde{F} - \tilde{G}$ into layers. By the definition of the worker measure $\tilde{F}$ in equation (A.19) and the job measure $\tilde{G}$ in equation (A.20), we know that for each k we have $\tilde{F}(x_k) = \tilde{G}(z_k)$. This means each layer ℓ will consist of a subset $S \subseteq \{0, \dots, p\}$ and the distributions within the layer F_ℓ and G_ℓ will be uniform on $\{x_k\}_{k\in S}$ and $\{z_k\}_{k\in S}$ respectively. From the assumption of the theorem we recall that the optimal assignment π pairs x_k with z_k for every k in the optimal assignment problem with uniform distributions on $\{x_k\}_{0\leqslant k\leqslant p}$ and $\{z_k\}_{0\leqslant k\leqslant p}$. Since a restriction of an optimal assignment is also optimal on the restricted marginals, we know that an optimal assignment between F_ℓ and G_ℓ matches x_k to z_k for each $k \in S$. After adding the layers, the same holds for an optimal assignment between $\tilde{F}$ and $\tilde{G}$ by the principle of layering. Therefore, the pairs $\{(x_k, z_k)\}$ are paired under an optimal assignment between $\tilde{F}$ and $\tilde{G}$. This establishes the inequality (A.18), hence we finally conclude (A.13) has a solution. □

Next, we continue to prove the second part of the result, that the function ϕ defined in (A.9)-(A.11) is indeed a local dual optimizer on $S_{[x_0,z_0]}$.

First, we record the following simple observation on our construction of ϕ in the case of overlapping masses.

We will extensively make use of Lemma 11.

Lemma 11. Suppose $h : [0,\infty) \to \mathbb{R}$ is concave. Then for $0 \leqslant x \leqslant y$ and $a > 0$ we have

$$h(x+a) - h(x) \geqslant h(y+a) - h(y).$$

Proof. From concavity, it follows that

$$h(y) + h(x+a) \geqslant \frac{(y-x)h(y+a) + ah(x)}{y+a-x} + \frac{(y-x)h(x) + ah(y+a)}{y+a-x} = h(x) + h(y+a),$$

completing the proof. □

Proof of Theorem 4. Suppose $(x_1, z_1), \ldots, (x_p, z_p)$ are ordered subpairs of pair (x_0, z_0) in the optimal assignment π, and that ϕ_i are dual potentials on $S_{[x_i,z_i]} := I_{[x_i,z_i]} \cup J_{[x_i,z_i]}$ for all $1 \leqslant i \leqslant p$. We first prove that, with the possibilities of multiple workers on the same skill level and multiple jobs on the same difficulty level, our ϕ in (A.9)-(A.11) is well-defined.

1. Suppose that $x_n = x_{n+1}$ or $z_n = z_{n+1}$ for some $1 \leqslant n < p$. Then any solution $(\beta_2, \ldots, \beta_p)$ to the system of inequalities (A.13) satisfies

$$\max(c_{00} - c_{0n} - c_{n+1,0}, -c_{n+1,n}) + c_{nn} \leqslant \beta_{n+1} \leqslant \min(c_{0,n+1} + c_{n0} - c_{00}, c_{n,n+1}) - c_{n+1,n+1}.$$

As a consequence,

$$c_{nn} - c_{n+1,n} \leqslant \beta_{n+1} \leqslant c_{n,n+1} - c_{n+1,n+1},$$

and hence we must have $\beta_{n+1} = c_{nn} - c_{n+1,n+1}$. In particular, the ϕ defined in (A.9) satisfies $\phi(x_n) = \phi(x_{n+1})$ in the case $x_n = x_{n+1}$, and $\phi(z_n) = \phi(z_{n+1})$ in the case $z_n = z_{n+1}$.

2. Suppose that $z_0 = z_1$ and $x_0 \neq x_1$. We first prove that

$$\max_{i \in \{1,\ldots,p\}} (\phi(x_i) - c(x_i, z_0)) = \phi(x_1) - c(x_1, z_0).$$

Indeed, for any $i \in \{1, \ldots, p\}$, it holds that

$$\phi(x_1) - c(x_1, z_0) = \phi(x_1) - c(x_1, z_1) = \phi(z_1) \geqslant \phi(x_i) - c(x_i, z_1) = \phi(x_i) - c(x_i, z_0).$$

Therefore, by (A.10),

$$\phi(z_0) = \phi(x_1) - c(x_1, z_0) = \phi(x_1) - c(x_1, z_1) = \phi(z_1).$$

3. Suppose that $x_0 = x_1$. Similarly as in step 2 above, we have

$$\phi(x_0) = \phi(z_1) + c(x_0, z_1) = \phi(z_1) + c(x_1, z_1) = \phi(x_1).$$

4. Suppose that $x_0 = x_1$ and $z_0 = z_1$. We need to show that

$$\max_{i \in \{1,\ldots,p\}} (\phi(x_i) - c(x_i, z_0)) = \min_{i \in \{1,\ldots,p\}} (\phi(z_i) + c(x_0, z_i)) - c(x_0, z_0). \tag{A.21}$$

This is because the left-hand side of (A.21) is equal to $\phi(z_1)$ by step 3 and the right-hand side of (A.21) is equal to $\phi(x_1) - c_{00}$ by step 2. Since $c_{00} = c_{11} - \phi(x_1) - \phi(z_1)$, this proves (A.21).

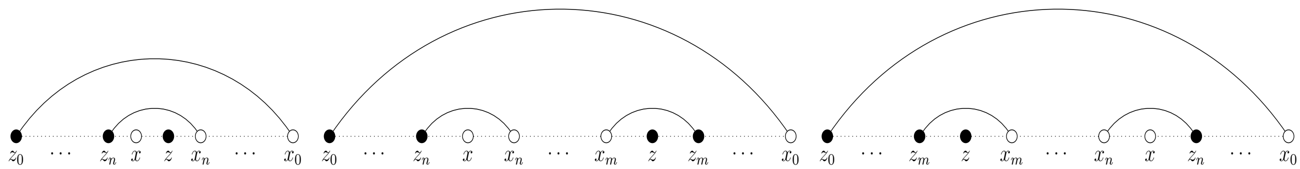


Figure A.12: Three Cases for the proof of Theorem 4

Figure A.12 illustrates the three different cases that we consider in the proof of Theorem 4. The first case is when $n = m$, the second case is when $m > n$, and the third case is when $n > m$. In each case, worker $x \in (x_n, z_n)$ and job $z \in (x_m, z_m)$.

We prove that ϕ is a local dual optimizer on the domain $S_{[x_0,z_0]}\backslash\{x_0, z_0\}$. Suppose that both $x \in I_{[x_n,z_n]}$ and $z \in J_{[x_m,z_m]}$. The equality $\phi(x) - \phi(z) = c(x, z)$ when (x, z) is a worker-job pair is obvious because the same condition is satisfied by ϕ_i for all $1 \leqslant i \leqslant p$. Our goal is to prove $\phi(x) - \phi(z) \leqslant c(x, z)$ when worker x and job z are not paired.

We consider three cases.

Case I: $n = m$. This follows immediately since ϕ_n is a dual potential on $S_{[x_n,z_n]}$.

Case II: $n < m$. Observe

$$\begin{aligned}
\phi(x) - \phi(z) &= \big(\phi(x) - \phi(z_n)\big) - \sum_{k=n}^{m}(\phi(x_k) - \phi(z_k)) + \sum_{k=n+1}^{m}(\phi(x_{k-1}) - \phi(z_k)) + (\phi(x_m) - \phi(z)) \\
&\leqslant c(x, z_n) - \sum_{k=n}^{m} c_{kk} + \sum_{k=n+1}^{m}(\beta_k + c_{kk}) + c(x_m, z) \\
&= c(x, z_n) - c_{nn} + c(x_m, z) + \sum_{k=n+1}^{m}\beta_k \\
&\leqslant c(x, z_n) - c_{nn} + c_{nm} - c_{mm} + c(x_m, z).
\end{aligned}$$

The first inequality follows since both $\phi(x) - \phi(z_n) \leqslant c(x, z_n)$ and $\phi(x_m) - \phi(z) \leqslant c(x_m, z)$ follow from the dual potential within the same part, and $\phi(x_k) - \phi(z_k) = c_{kk}$ follows by the dual potential within the same part for paired workers and jobs, and finally $\beta_k + c_{kk} = \phi(x_{k-1}) - \phi(z_k)$ by equation (A.9). The final inequality is implied by the upper bound on $\sum_{k=n+1}^{m}\beta_k$ from (A.13). To show $\phi(x) - \phi(z) \leqslant c(x, z)$, it suffices to prove

$$c(x, z_n) + c(x_m, z) - c(x, z) \leqslant c_{nn} + c_{mm} - c_{nm}. \tag{A.22}$$

Note that $c(x, z_n) \leqslant c(x_n, z_n)$ and $c(x_m, z) \leqslant c(x_m, z_m)$. There are four cases:

1. $z_n \leqslant x_n \leqslant z_m \leqslant x_m$. Then $c(x, z) \geqslant c(x_n, z_m)$ and the claim follows.

2. $x_n \leqslant z_n \leqslant z_m \leqslant x_m$. In this case $z_n - x \leqslant z_n - x_n$. Applying Lemma 11 with $a = z - z_n$ yields $c(x, z_n) - c(x, z) \leqslant c(x_n, z_n) - c(x_n, z) \leqslant c(x_n, z_n) - c(x_n, z_m)$. Using that $c(x_m, z) \leqslant c(x_m, z_m)$ equation (A.22) follows.
3. $z_n \leqslant x_n \leqslant x_m \leqslant z_m$. In this case $z - x_m \leqslant z_m - x_m$. Applying Lemma 11 with $a = x_m - x_n$ yields $c(x_m, z) - c(x, z) \leqslant c(x_m, z) - c(x_n, z) \leqslant c(x_m, z_m) - c(x_n, z_m)$, where the first inequality follows from $c(x_n, z) \leqslant c(x, z)$. Further using $c(x, z_n) \leqslant c(x_n, z_n)$, equation (A.22) follows.
4. $x_n \leqslant z_n \leqslant x_m \leqslant z_m$. In this case, the configuration between x_n and z_n is identical to Case 2, and the configuration between x_m and z_m is identical to Case 3. We apply Lemma 11 exactly as in Case 2 and in Case 3. First, $c(x, z_n) - c(x, z) \leqslant c(x_n, z_n) - c(x_n, z)$. Second, $c(x_m, z) - c(x_n, z) \leqslant c(x_m, z_m) - c(x_n, z_m)$. Summing the two inequalities delivers inequality (A.22).

This completes the proof of (A.22) for Case II.

Case III: $n > m$. Observe that

$$\begin{aligned}\phi(x) - \phi(z) &= (\phi(x) - \phi(z_n)) + \sum_{k=m+1}^{n-1} (\phi(x_k) - \phi(z_k)) - \sum_{k=m+1}^{n} (\phi(x_{k-1}) - \phi(z_k)) + (\phi(x_m) - \phi(z)) \\ &\leqslant c(x, z_n) + \sum_{k=m+1}^{n-1} c_{kk} - \sum_{k=m+1}^{n} (\beta_k + c_{kk}) + c(x_m, z) \\ &= c(x, z_n) - c_{nn} + c(x_m, z) - \sum_{k=m+1}^{n} \beta_k \\ &\leqslant c(x, z_n) - c_{nn} + c_{nm} - c_{mm} + c(x_m, z).\end{aligned}$$

The first inequality follows since both $\phi(x) - \phi(z_n) \leqslant c(x, z_n)$ and $\phi(x_m) - \phi(z) \leqslant c(x_m, z)$ follow from the dual potential within the same part, and $\phi(x_k) - \phi(z_k) = c_{kk}$ follows by the dual potential within the same part for paired workers and jobs, and finally $\beta_k + c_{kk} = \phi(x_{k-1}) - \phi(z_k)$ by equation (A.9). The final inequality is implied by the upper bound on $\sum_{k=n+1}^{m} \beta_k$ from (A.13). The rest follows similarly as in Case II.

It then remains to check that ϕ is a dual potential on $S_{[x_0,z_0]}$, that is, to show that

1. $\phi(x_0) - \phi(z_0) = c(x_0, z_0)$;
2. For $x \in I_{[x_0,z_0]} \setminus \{x_0\}$, $\phi(x) - \phi(z_0) \leqslant c(x, z_0)$;
3. For $z \in J_{[x_0,z_0]} \setminus \{z_0\}$, $\phi(x_0) - \phi(z) \leqslant c(x_0, z)$.

In view of (A.10) and (A.11), it remains to prove

$$\min_{1 \leqslant i \leqslant p} (\phi(z_i) + c(x_0, z_i)) - \max_{1 \leqslant i \leqslant p} (\phi(x_i) - c(x_i, z_0)) \geqslant c(x_0, z_0). \tag{A.23}$$

Equivalently, it suffices to show that

$$\phi(x) - \phi(z) \leqslant c(x_0, z) + c(x, z_0) - c_{0,0} \tag{A.24}$$

for all $x \in I_{[x_0,z_0]} \setminus \{x_0\}$ and $z \in J_{[x_0,z_0]} \setminus \{z_0\}$. Recall that $\phi(x_{i-1}) - \phi(z_i) = \beta_i + c_{ii}$ for $2 \leqslant i \leqslant p$ by equation (A.9). Again, we have the three cases of Figure A.12 to consider to show that the sufficient condition (A.24) is satisfied.

Case I: $n = m$. By symmetry, we may assume $x_0 < z_0$. Consider first the case $z_n \leqslant x_n$. Since ϕ_n is a dual potential on $S_{[x_n,z_n]}$, we have $\phi(x) - \phi(z) \leqslant c(x, z)$. Since π is optimal,

$$c(x, z) + c(x_0, z_0) \leqslant c(x_n, z_n) + c(x_0, z_0) \leqslant c(x_n, z_0) + c(x_0, z_n) \leqslant c(x, z_0) + c(x_0, z).$$

where the first inequality follows since $(x, z) \in (x_n, z_n)$, the second follows by optimality, and the third one follows since $z_n \leqslant x_n$ and $x_0 < z_0$. Using the above inequality, we write:

$$\phi(x) - \phi(z) \leqslant c(x, z) \leqslant c(x_0, z) + c(x, z_0) - c_{00},$$

verifying (A.24).

Next, we consider the case where $x_n < z_n$. By the property of no intersecting pairs, it follows that

$$c(x_n, z_n) + c(x_0, z_0) \leqslant c(x_0, z_n) + c(x_n, z_0). \tag{A.25}$$

Furthermore, two applications of Lemma 11 yield both

$$c(x_0, z_n) - c(x_0, z) \leqslant c(x_n, z_n) - c(x_n, z) \tag{A.26}$$

when $a = z_n - z$ and

$$c(x_n, z_0) - c(x, z_0) \leqslant c(x_n, z_n) - c(x, z_n). \tag{A.27}$$

when $a = z_0 - z_n$. Summing up equations (A.25), (A.26), and (A.27) yields

$$c(x_0, z_0) + c(x, z_n) + c(x_n, z) \leqslant c(x_0, z) + c(x, z_0) + c(x_n, z_n).$$

Since π is optimal, we have

$$\phi(x) - \phi(z) \leqslant c(x, z) \leqslant c(x, z_n) + c(x_n, z) - c(x_n, z_n) \leqslant c(x_0, z) + c(x, z_0) - c_{00},$$

as desired by (A.24).

Case II: $n < m$. Similar to the other Case II discussed above, we obtain

$$\phi(x) - \phi(z) \leqslant c(x, z_n) - c_{nn} + c(x_m, z) + \sum_{k=n+1}^{m} \beta_k$$
$$\leqslant c(x, z_n) - c_{nn} + c_{0m} + c_{n0} - c_{00} - c_{mm} + c(x_m, z), \tag{A.28}$$

where the second inequality follows from the upper bound in (A.13). To prove equation (A.24) it suffices to show that the upper bound in the previous equation is below the upper bound in equation (A.24), or

$$c(x, z_n) + c(x_m, z) - c(x_0, z) - c(x, z_0) \leqslant c(x_n, z_n) + c(x_m, z_m) - c(x_n, z_0) - c(x_0, z_m). \tag{A.29}$$

We observe that the arcs (x, z_0) and (x_n, z_n) intersect, but that the arcs (x, z_n) and (x_n, z_0) do not cross. By Lemma 3, describing that intersecting pairs are never optimal, we thus have

$$c(x, z_n) + c(x_n, z_0) \leqslant c(x, z_0) + c(x_n, z_n). \tag{A.30}$$

Similarly, the arcs (x_0, z) and (x_m, z_m) intersect, but the arcs (x_0, z_m) and (x_m, z) do not intersect. Thus, by the property of no intersecting pairs,

$$c(x_0, z_m) + c(x_m, z) \leqslant c(x_0, z) + c(x_m, z_m). \tag{A.31}$$

Summing up (A.30) and (A.31) yields (A.29) and hence completes the proof of Case II.

Case III: $n > m$. We have

$$\phi(x) - \phi(z) \leqslant c(x, z_n) - c_{nn} + c(x_m, z) - \sum_{k=m+1}^{n} \beta_k$$
$$\leqslant c(x, z_n) - c_{nn} - c_{00} - c_{mm} + c_{0m} + c_{n0} + c(x_m, z), \tag{A.32}$$

where the last step follows from (A.13). Inequality (A.32) coincides with inequality (A.28) in Case II. Then (A.24) follows in exactly the same way from (A.29) as in Case II. □

Finally, we emphasize that our construction relies on the concavity of the mismatch cost function $c(x, z)$ in two respects. First, the property of no intersecting pairs is essential for our induction structure. Second, (A.22) requires concavity as well.

A.12 Proof of Theorem 5

In this appendix, we prove Theorem 5. We make use of Lemma 12 and Lemma 13, which we prove first.

Lemma 12. *Triangle Inequality.* For all $x, y, z \in \mathbb{R}$, it holds that $c(x, y) + c(y, z) \geqslant c(x, z)$.

Proof. Our cost of mismatch takes the form $c(x,z) = h(z-x)$ where h is strictly concave and increasing on $[0,\infty)$, strictly concave and decreasing on $(-\infty,0]$, satisfying $h(0)=0$.

The case where $x=z$ is trivial, so we focus our attention to the case where $x \neq z$. By symmetry, we assume $x<z$ without loss of generality. If $z-y \geqslant z-x > 0$, then necessarily $c(y,z) \geqslant c(x,z)$ and hence $c(x,y)+c(y,z) \geqslant c(x,z)$. The same argument applies when $y-x \geqslant z-x > 0$. In the remaining scenario where both $(z-y)$ and $(y-x)$ are in $[0,z-x)$, we have by concavity of h that[46]

$$c(x,y)+c(y,z) = h(y-x)+h(z-y) \geqslant \frac{y-x}{z-x}\,h(z-x) + \frac{z-y}{z-x}\,h(z-x) = h(z-x) = c(x,z),$$

where the inequality follows since $y-x = \frac{y-x}{z-x} \times (z-x) + \frac{z-y}{z-x} \times 0$. This completes the proof. ☐

In the main text, we established the connection between the dual optimizers for the cost minimization problem and the dual optimizers for the output maximization problem. In this appendix, we exploit this connection to simplify the exposition to the proof of Theorem 5. Specifically, we use that we can equivalently characterize the dual functions (ϕ,ψ) for the overlapping segments of the worker and the job distribution, with the understanding that we can obtain wages and job values using $w(x) = g(x) - \phi(x)$ and $v(z) = h(z) - \psi(z)$.

To formulate Theorem 5 in terms of the dual potentials for the mismatch cost minimization problem, we need to describe our sequential construction of the functions. For the interpretation of these objects we refer the reader to Section 4. We define sequentially the dual maps, analogous to our previous definitions (11) and (12). Starting from $\tilde{\phi} = g - \tilde{w}$, where $\tilde{w}$ are the dual values for mismatched workers $x \in I$, let

$$\tilde{\psi}(z) := \min_{x\in I}\big(c(x,z) - \tilde{\phi}(x)\big) \qquad \text{and} \qquad \hat{\phi}(x) := \min_{z\in I\cup J}\big(c(x,z) - \tilde{\psi}(z)\big), \tag{A.33}$$

where we recall that J is the set of mismatched jobs. Moreover, let

$$\hat{\psi}(z) := \min_{x\in I\cup J}\big(c(x,z) - \hat{\phi}(x)\big), \tag{A.34}$$

$\phi(x) = \hat{\phi}(x)$ for $x \in I$, $\psi(z) = \hat{\psi}(z)$ for $z \in J$, and set $\phi(x) = -\psi(x)$ for $x \in J$ and $\psi(z) = -\phi(z)$ for $z \in I$. Finally, we define for $x \in K$

$$\phi(x) = \min_{z\in I\cup J}\big(c(x,z) - \psi(z)\big) \tag{A.35}$$

and $\psi(z) = -\phi(z)$ for $z \in K$. It is easy to check that with these definitions, $\phi = g - w$ and $\psi = h - v$, with w, v given in Theorem 5.

[46]The interval is bounded below by zero because when $z-y \geqslant z-x$ is not true, then $y-x>0$, and similarly when $y-x \geqslant z-x$ is not true then $z-y>0$.

To prove the result, we first define c-conjugate functions and analyze some of their properties.

For $\phi : I \to \mathbb{R}$, we define the c-conjugate function for all jobs $z \in J$ as

$$\phi^c(z) := \min_{x\in I} \big(c(x,z) - \phi(x)\big). \tag{A.36}$$

Denote by $\bar{c}(z,x) = c(x,z)$ and we further define for $x \in I$

$$\phi^{c\bar{c}}(x) = (\phi^c)^{\bar{c}}(x) = \min_{z\in J} \big(\bar{c}(z,x) - \phi^c(z)\big) = \min_{z\in J} \big(c(x,z) - \phi^c(z)\big). \tag{A.37}$$

Given these definitions, the following statements follow:[47]

1. $\phi^{c\bar{c}} \geqslant \phi$.

   This follows since for each $x \in I$ and $z \in J$, $\phi(x) + \phi^c(z) \leqslant c(x,z)$ or $\phi(x) \leqslant c(x,z) - \phi^c(z)$ by the definition of the c-conjugate function. By taking the infimum $z \in J$ this leads to $\phi^{c\bar{c}}(x) \geqslant \phi(x)$ by the definition (A.37).

2. If $\phi = \psi^{\bar{c}}$ for some ψ, then $\phi = \phi^{c\bar{c}}$.

   First, we observe that $\phi = \psi^{\bar{c}}$ naturally implies $\phi^c = \psi^{\bar{c}c}$. To see this, note that $\phi^c = \psi^{\bar{c}c} \geqslant \psi$ which follows from the previous statement. This inequality, by uniformly decreasing from ϕ^c to ψ, together with the definition (A.37), implies we uniformly increase the conjugate, or $\phi^{c\bar{c}} = (\phi^c)^{\bar{c}} \leqslant \psi^{\bar{c}} = \phi$. We establish $\phi = \phi^{c\bar{c}}$ by combining $\phi^{c\bar{c}} \leqslant \phi$ with Statement 1.

3. If (ϕ, ψ) is an optimal dual pair, then so is (ϕ, ϕ^c).

   Suppose (ϕ, ψ) is a dual pair, then $\phi(x) + \psi(z) \leqslant c(x,z)$. It holds by the definition in (A.36) that $\phi(x) + \phi^c(z) \leqslant c(x,z)$ as well as $\phi^c(z) \geqslant \psi(z)$. Since $\phi^c(z) \geqslant \psi(z)$ and (ϕ, ϕ^c) is a dual solution, it follows that if (ϕ, ψ) is a solution to the dual maximization problem, then so is (ϕ, ϕ^c).

Lemma 13. $\hat{\phi}(x) + \tilde{\psi}(z) \leqslant c(x,z)$ for all workers x and jobs z such that $x, z \in I \cup J$, and equality holds for $(x,z) \in \Gamma_\pi$.

Proof. That $\hat{\phi}(x) + \tilde{\psi}(z) \leqslant c(x,z)$ follows from definition of the dual optimizer for workers $\hat{\phi}(x)$ in (11). Next, we prove $\hat{\phi}(x) + \tilde{\psi}(z) = c(x,z)$ for all workers and jobs $(x,z) \in \Gamma_\pi$.

To prove that $\hat{\phi}(x)+\tilde{\psi}(z) = c(x,z)$ for workers and jobs $(x,z) \in \Gamma_\pi$, we fix some pair $(x,z) \in \Gamma_\pi$. Since $(\tilde{\phi}, \tilde{\phi}^c)$ is a dual solution to the assignment problem between remaining workers and jobs, $\tilde{\phi}(x) + \tilde{\phi}^c(z) = c(x,z)$ for all $(x,z) \in \Gamma_\pi$. Given the definition of the dual value for jobs $z \in I \cup J$ in (A.33) we obtain that $\tilde{\psi}(z) = \min_{x\in I} \big(c(x,z) - \tilde{\phi}(x)\big) = \tilde{\phi}^c(z)$ and hence that $\tilde{\phi}(x) + \tilde{\psi}(z) = c(x,z)$ for all $(x,z) \in \Gamma_\pi$. To conclude the proof it remains to show that $\hat{\phi}(x) = \tilde{\phi}(x)$ for every worker $x \in I$.

[47] See Chapter 1 of Santambrogio (2015) for further details.

We next show that $\hat{\phi}(x) = \tilde{\phi}(x)$ for every worker $x \in I$. Since we replaced, without loss of generality, the dual potential $\tilde{\phi}$ by the c-transform $\tilde{\phi}^{c\bar{c}}$,[48] it follows from the definition of the c-transform that for all $x \in I$:

$$\tilde{\phi}(x) = \tilde{\phi}^{c\bar{c}}(x) = \min_{z\in J}\big(c(x,z) - \tilde{\phi}^c(z)\big).$$

Further, since $z \in J$, by definition of the dual potential for jobs $\tilde{\psi}(z) = \min_{x\in I}\big(c(x,z) - \tilde{\phi}(x)\big) = \tilde{\phi}^c(z)$, where the second equality follows from the definition of the c-transform. We substitute this relationship into the previous expression for $\tilde{\phi}(x)$ to write

$$\tilde{\phi}(x) = \min_{z\in J}\big(c(x,z) - \tilde{\phi}^c(z)\big) = \min_{z\in J}\big(c(x,z) - \tilde{\psi}(z)\big).$$

We can use the definition of the dual optimizers (11) to write that for all workers $x \in I$:

$$\hat{\phi}(x) = \min\Big(\min_{z\in J}\big(c(x,z) - \tilde{\psi}(z)\big), \min_{z\in I}\big(c(x,z) - \tilde{\psi}(z)\big)\Big) = \min\Big(\tilde{\phi}(x), \min_{z\in I}\big(c(x,z) - \tilde{\psi}(z)\big)\Big)$$

where the first equality follows by splitting the sets in (A.36) and the second equality follows from the equation above.

Hence, we want to show for $(x,z) \in I$ the infimum is attained by $\tilde{\phi}(x)$. We show $c(x,z) \geqslant \tilde{\psi}(z) + \tilde{\phi}(x)$. This follows since the dual optimizer for all jobs is defined $\tilde{\psi}(z) := \min_{x\in I}\big(c(x,z) - \tilde{\phi}(x)\big)$ for all $z \in I \cup J$. □

Having established the two claims, we next prove Theorem 5.

Proof of Theorem 5. The proof is divided in three parts. We first show the inequality holds on $I \cup J$. To do so, we consider four cases:

1. If $x \in I$ and $z \in J$, we have

$$\phi(x) + \psi(z) = \phi(x) - \phi(z) = \hat{\phi}(x) - \max_{x\in I\cup J}(\hat{\phi}(x) - c(x,z)) \leqslant c(x,z)$$

   because in the final step we subtract the maximum, but a feasible deduction is $\hat{\phi}(x) - c(x,z)$.

2. If $x, z \in I$, then by Lemma 12 we have

$$\phi(x) + \psi(z) = \phi(x) - \phi(z) = \hat{\phi}(x) - \hat{\phi}(z) = \min_{y\in I\cup J}(c(x,y) - \tilde{\psi}(y)) - \min_{y\in I\cup J}(c(z,y) - \tilde{\psi}(y)) \leqslant c(x,z)$$

   where the final equality follows by (A.33). The concluding inequality is obtained as follows. Suppose y_0 attains the infimum for the second term, the same y_0 may not attain the infimum for the first

[48] We can always improve on the original $\tilde{\phi}$ by doing a double c-conjugate transform by Statement 1 that is weakly better in terms of the dual maximization problem, see Remark 1.13 in Santambrogio (2015).

term but is feasible, so the left hand side will be lower than when both terms are evaluated at y_0, or $\min_{y\in I\cup J}(c(x,y)-\tilde{\psi}(y))-\min_{y\in I\cup J}(c(z,y)-\tilde{\psi}(y)) \leqslant c(x,y_0)-c(z,y_0)$. We combine the right-hand side with the triangle inequality of Lemma 12 to write $c(x,z)+c(z,y_0) \geqslant c(x,y_0)$ or $c(x,z) \geqslant c(x,y_0)-c(z,y_0)$ to obtain the inequality.

3. If $x \in J$ and $z \in I$.

$$\phi(x) + \psi(z) = \max_{y\in I\cup J} \left(\hat{\phi}(y) - c(y,x)\right) - \hat{\phi}(z).$$

We next want to show that this expression is less than $c(x,z)$. This is equivalent to showing that $\hat{\phi}(y) - c(y,x) \leqslant c(x,z) + \hat{\phi}(z)$ for all $y \in I \cup J$. To establish this, fix y, and evaluate:

$$\hat{\phi}(y) - \hat{\phi}(z) = \min_{w\in I\cup J} (c(y,w) - \tilde{\psi}(w)) - \min_{w\in I\cup J} (c(z,w) - \tilde{\psi}(w))$$

where the equality follows from the definition of $\hat{\phi}$ in equation (A.33). Let w_0 be the value that attains the infimum in the second term on the right, which is also feasible for the first term so that $\hat{\phi}(y) - \hat{\phi}(z) \leqslant c(y,w_0) - c(z,w_0)$. To bound this further, we use the triangle inequality of Lemma 12 twice to write $c(y,w_0) - c(z,w_0) \leqslant c(y,z)$ as well as $c(y,z) \leqslant c(y,x) + c(x,z)$. Using the triangle inequalities, we thus write $\hat{\phi}(y) - \hat{\phi}(z) \leqslant c(y,x) + c(x,z)$ which is what we wanted to show since y is arbitrary.

4. If worker and job $x, z \in J$, use (A.34) to write

$$\phi(x) + \psi(z) = \max_{y\in I\cup J}(\hat{\phi}(y) - c(y,x)) - \max_{y\in I\cup J}(\hat{\phi}(y) - c(y,z)).$$

To bound the right-hand side, let y_0 denote the value that attains the supremum in the first term, that is also feasible for the second term. Hence, the right-hand side is bounded above by $-c(y_0,x) + c(y_0,z)$. By the triangle inequality of Lemma 12 it follows that $-c(y_0,x) + c(y_0,z) \leqslant c(x,z)$ and hence we have $\phi(x) + \psi(z) \leqslant c(x,z)$.

The second part of the proof shows that the equality holds everywhere on $I \cup J$ with respect to the optimal assignment π. We distinguish two cases:

1. The worker is perfectly matched to their job, or $(x,z) \in \{(x,x) : x \in \mathbb{R}\}$. Since the dual functions are defined as $\psi(x) = -\phi(x)$ we have $\phi(x) + \psi(x) = 0$. As a result, $\phi(x) + \psi(x) = 0 = c(x,x)$, as the cost of mismatch is zero.

2. The worker is mismatched in their job, or $(x,z) \in \Gamma_\pi$, implying worker $x \in I$ and job $z \in J$. Using definition (A.34), $\psi(z) = -\phi(z)$,

$$\phi(x) + \psi(z) = \hat{\phi}(x) - \max_{y\in I\cup J}(\tilde{\phi}(y) - c(y,z)).$$

By Lemma 13, $\hat{\phi}(x) + \tilde{\psi}(z) \leqslant c(x,z)$ for all (x,z). In particular, for a given job z, $\hat{\phi}(x) + \tilde{\psi}(z) \leqslant c(x,z)$ for all x, and $\min_{x\in I\cup J}(c(x,z) - \hat{\phi}(x)) \geqslant \tilde{\psi}(z)$, or, equivalently, $-\max_{x\in I\cup J}\big(\hat{\phi}(x) - c(x,z)\big) \geqslant \tilde{\psi}(z)$. Combining this inequality with the previous expression, we obtain the inequality

$$\phi(x) + \psi(z) \geqslant \hat{\phi}(x) + \tilde{\psi}(z) = c(x,z)$$

where the final equality follows by Lemma 13. Since we have shown the opposite inequality above in the first case of the first part to this proof, we obtain that $\phi(x) + \psi(z) = c(x,z)$.

In the third part of the proof we further establish that the dual inequality $\phi(x) + \psi(z) \leqslant c(x,z)$ holds when $x \in K$ or $z \in K$. There are three cases.

1. $x \notin K,\ z \in K$. For any $x' \notin K$, we have $\phi(x) + \psi(x') = \phi(x) - \phi(x') \leqslant c(x,x')$ when $x \notin K$ by the first part of this proof. Following the triangle inequality of Lemma 12, $\phi(x) - \phi(x') \leqslant c(x,x') \leqslant c(x,z) + c(z,x')$, giving $\phi(x) - c(x,z) \leqslant c(z,x') + \phi(x') = c(z,x') - \psi(x')$. Taking infimum over $x' \in I \cup J$ gives $\phi(x) - c(x,z) \leqslant \phi(z) = -\psi(z)$ using the definition of ϕ.
2. $x \in K,\ z \notin K$. For any $x' \in I \cup J$, by the definition of the wage function (A.35), we have that $\phi(x) = \min_{x'\in I\cup J}\big(c(x,x') - \psi(x')\big)$, such that $\phi(x) - \phi(x') \leqslant c(x,x') \leqslant c(x,z) + c(z,x')$, where the final step follows by the triangle inequality of Lemma 12. Alternatively, we write $\phi(x) - c(x,z) \leqslant c(z,x') - \psi(x')$. Taking infimum in $x' \in I \cup J$ gives $\phi(x) + \psi(z) \leqslant c(x,z)$ using the definition of ϕ.
3. $x, z \in K$. We want to establish $\phi(x) + \psi(z) \leqslant c(x,z)$. Using the definitions of the dual potentials in (A.35),

   $$\phi(x) + \psi(z) = \min_{x'\in I\cup J}\big(c(x,x') - \psi(x')\big) + \max_{x'\in I\cup J}\big(-c(z,x') + \psi(x')\big).$$

   Suppose the maximum in the second term is attained by the worker value x_0, and also evaluate the first term at x_0 where it may not attain the minimum, implying $\phi(x) + \psi(z) \leqslant c(x,x_0) - c(z,x_0)$. By the triangle inequality $c(x,x_0) - c(z,x_0) \leqslant c(x,z)$ and hence it indeed follows that $\phi(x) + \psi(z) \leqslant c(x,z)$.

By observing that $\phi = -\psi$ on the set K, the equality $\phi(x) + \psi(z) = 0 = c(x,z)$ holds when $x, z \in K$ and $(x,z) \in \Gamma_\pi$. This completes the proof in view of Lemma 1. □

Composite Sorting

Technical Appendix

Job Boerma, Aleh Tsyvinski, Ruodu Wang, and Zhenyuan Zhang

August 2023

B Additional Results

In this appendix we present additional technical results.

B.1 General Production Function

In this section, we present two generalizations of the model.

B.1.1 Concave Distance Function

Using Legendre transformations, we show that the indirect output function is generally a strictly concave function in mismatch given strictly convex cost functions. To be specific, consider the cost function Ψ as in Section 2.1, and use $d := |x - z|$ to denote the distance to obtain

$$\mathcal{C}(d) = \min_{\gamma \geqslant 0} \big(\gamma d + \Psi(\gamma)\big). \tag{A.38}$$

This problem has a unique solution characterized by $d = -\Psi'(\gamma(d))$. From the envelope condition, we obtain $\mathcal{C}'(d) = \gamma > 0$, showing that the cost function is strictly increasing in the distance.

To characterize the second derivative, we write the cost minimization problem as a maximization problem of the form:

$$\hat{\mathcal{C}}(d) = \max_{\gamma \geqslant 0} \big(- \gamma d - \Psi(\gamma)\big),$$

where $\hat{\mathcal{C}} = -\mathcal{C}$ which shows $\hat{\mathcal{C}}$ is the Legendre transformation of the strictly convex function Ψ. Since the Legendre transformation of a strictly convex function is also strictly convex, the indirect cost function $\mathcal{C}$ is a strictly concave function of the distance. As a result, choosing an assignment to maximize:

$$y(x, z) = z + x - \mathcal{C}(|x - z|)$$

where $\mathcal{C}$ is now our concave cost distance function.

B.1.2 Asymmetric Distance Function

Next, we incorporate differential distance functions for both $x - z > 0$ and $x - z < 0$. This is a trivial extension, let $\bar{\Psi}$ denote the cost function for $x - z > 0$ and $\underline{\Psi}$ denote the cost function for $x - z < 0$. In this case, the cost minimization problem is:

$$\bar{\mathcal{C}}(d) = \min_{\gamma \geqslant 0} \left(\gamma d + \bar{\Psi}(\gamma)\right)$$

when $d > 0$. By the same arguments on the Legendre transformation, this gives rise to a strictly concave function of the distance $\bar{\mathcal{C}}(d)$. Analogously, when $d < 0$, we generically obtain a distinct strictly concave function of the distance $\underline{\mathcal{C}}(d)$. As a result, we choose an assignment to maximize:

$$y(x, z) = z + x - \bar{\mathcal{C}}(\{x - z\}_+) - \underline{\mathcal{C}}(\{z - x\}_+).$$

B.2 Uniqueness of Optimal Sorting

In this appendix we discuss the uniqueness of optimal sorting for the mismatch cost function (5).

Proposition 7. For any fixed distributions of workers F and of jobs G, the set of $(\zeta_p, \zeta_u) \in (0,1)^2$ where the optimal assignment is not unique has Lebesgue measure zero.

Proof. First recall that every optimal assignment has the non-crossing property and the layering structure. In other words, the non-uniqueness problem arises only when we solve the assignment problems in each layer. Since layering does not depend on the cost function and there are finitely many layers, we consider without loss a fixed layer ℓ with $2n_\ell$ points, with an alternating pattern in the marginals F_ℓ and G_ℓ.

By Birkhoff's theorem (Birkhoff, 1946), every assignment between F_ℓ and G_ℓ is a mixture of bijective assignments. Therefore, it suffices to restrict to the set of bijective matchings between F_ℓ and G_ℓ. Since F_ℓ and G_ℓ have finite support, there exist finitely many assignments, and hence it suffices to prove that for any two assignments π and π', their costs equal on a set (ζ_p, ζ_u) of measure zero. In the following we fix π and $\hat{\pi}$. Their respective costs are of the form $\mathcal{C}(\pi) = \sum_{j=1}^{n_\ell} \left(a_j d_j^{\zeta_p} + b_j e_j^{\zeta_u}\right)$ and $\mathcal{C}(\hat{\pi}) = \sum_{j=1}^{n_\ell} \left(\hat{a}_j \hat{d}_j^{\zeta_p} + \hat{b}_j \hat{e}_j^{\zeta_u}\right)$ for some $d_j, e_j, \hat{d}_j, \hat{e}_j \geqslant 0$. Equating $\mathcal{C}(\pi) = \mathcal{C}(\hat{\pi})$ leads to an equation of the form

$$\sum_{j=1}^{2n_\ell} a_j d_j^{\zeta_p} = \sum_{j=1}^{2n_\ell} b_j e_j^{\zeta_u}, \tag{A.39}$$

where $d_j, e_j \geqslant 0$. Note that both sides are constant zero only if $F_\ell = G_\ell$, which is not feasible.

We next establish that the set of solutions $(\zeta_p, \zeta_u) \in (0,1)^2$ to (A.39) has zero Lebesgue measure. For a fixed value ζ_p, the equation (A.39) has finitely many solutions $\zeta_u \in (0,1)$ (see Tossavainen (2006)).

Similarly, for a fixed value ζ_u it has finitely many solutions $\zeta_p \in (0,1)$. Clearly, the zero set to equation (A.39) is a measurable set. Therefore, Fubini's theorem applied to the indicator function yields that such a set must have zero measure. □

B.3 Convex Cost of Mismatch

For completeness, we show that the optimal assignment features positive sorting when the mismatch cost function is strictly convex in the distance between worker skills and job difficulties. We establish this by showing that the cost function is submodular when the mismatch cost function is strictly convex. In turn, the production function is strictly supermodular, and hence the optimal sorting is positive following Becker (1973).

Lemma 14. *Convex Cost of Mismatch.* Suppose $c(x,z) = h(z-x)$ with $h : \mathbb{R} \to \mathbb{R}$ strictly convex. Then the mismatch cost function c is strictly submodular.

Proof. To establish that the cost of mismatch is submodular, we need to establish that

$$c(x_1, z_1) + c(x_2, z_2) < c(x_1, z_2) + c(x_2, z_1),$$

for any $x_1 < x_2$ and $z_1 < z_2$.

It follows from the strict convexity of the cost function that $h(z_1 - x_1) < c_1 h(z_1 - x_2) + c_2 h(z_2 - x_1)$ where the weighting coefficients are $c_1 = \frac{(z_2-x_1)-(z_1-x_1)}{(z_2-x_1)-(z_1-x_2)}$ and $c_2 = 1 - c_1 = \frac{(z_1-x_1)-(z_1-x_2)}{(z_2-x_1)-(z_1-x_2)}$.[49] Analogously, it follows that $h(z_2 - x_2) < c_2 h(z_1 - x_2) + c_1 h(z_2 - x_1)$.[50] We use these two inequalities to show that

$$c(x_1, z_2) + c(x_2, z_1) = h(z_2 - x_1) + h(z_1 - x_2) > h(z_1 - x_1) + h(z_2 - x_2) = c(x_1, z_1) + c(x_2, z_2),$$

where the inequality follows by addition of the two previous inequalities. □

B.4 Efficiency Properties of the Algorithm

Recall our algorithm for computing dual potentials from Appendix A.10. The total runtime is dominated by the runtime for solving the system of inequalities (A.8), which we recall as follows. Suppose that $(x_1, z_1), \ldots, (x_p, z_p)$ are ordered subpairs of pair (x_0, z_0) in the optimal assignment π. Define $c_{ij} :=$

[49] We observe that $c_1(z_1-x_2)+(1-c_1)(z_2-x_1) = z_2-x_1+c_1((z_1-x_2)-(z_2-x_1)) = z_2-x_1-((z_2-x_1)-(z_1-x_1)) = (z_1 - x_1)$ and further note that the coefficient c_1 is between 0 and 1 since the denominator is positive and the numerator is positive but smaller than the denominator.

[50] For completeness, we observe that $c_2(z_1 - x_2) + (1 - c_2)(z_2 - x_1) = z_2 - x_1 + c_2((z_1 - x_2) - (z_2 - x_1)) = z_2 - x_1 - ((z_1 - x_1) - (z_1 - x_2)) = (z_2 - x_2)$.

$c(x_i, z_j)$. Then the system of inequalities, where for all $1 \leqslant n < m \leqslant p$:

$$\max(c_{00} + c_{nn} - c_{0n} - c_{m0}, c_{nn} - c_{mn}) \leqslant \sum_{k=n+1}^{m} \beta_k \leqslant \min(c_{0m} + c_{n0} - c_{00} - c_{mm}, c_{nm} - c_{mm}) \tag{A.40}$$

(a) admits a solution $(\beta_2, \dots, \beta_p)$.

Since the system of inequalities (A.40) can be solved via linear programming in p steps where $p \leqslant n$, the worst-case runtime for our algorithm is $O(n^4)$ (when $p = n$). However, it becomes much more efficient when more layers of arcs are introduced. This is because compared with the trivial linear programming, our algorithm solves the problem in the order from bottom arcs to top, while at each step the values of ϕ in the hidden arcs need not be computed again, but only adjusted with constant factors. Typically, the number N will not be as large as n. The following proposition provides a general upper bound of the number N, which is a consequence of the absence of intersecting pairs. Define the number of crossings of $F - G$ at $x \in \mathbb{R}$ as

$$C_{F-G}(x) := \sum_{1 \leqslant k \leqslant n} \mathbb{1}_{\{(F-G)(x_k)=x\}} + \sum_{1 \leqslant k \leqslant n} \mathbb{1}_{\{(F-G)(z_k)=x\}}.$$

In other words, every time a flat part of $F - G$ takes value x, we count that as an x-crossing. Note that this is nonzero only if $nx \in \mathbb{Z}$ and $|x| \leqslant 1$.

Proposition 8. Suppose that $(x_1, z_1), \dots, (x_p, z_p)$ are the ordered subpairs of the pair (x_0, z_0) in the optimal matching π. Then there exists $x \in \mathbb{Z}/n$ such that the measure of underqualification $F - G$ crosses the level x for N times, i.e., $C_{F-G}(x) \geqslant N$.

Proof. Consider numbers $t_i \in (\max(x_i, z_i), \min(x_{i+1}, z_{i+1}))$, $1 \leqslant i < p$. By the property of no intersecting pairs, $F(t_i) - G(t_i)$ is a constant in i. On the other hand, $F - G$ cannot be constant on the interval $[t_i, t_{i+1}]$. Thus the claim follows. □

In fact, the number of crosses of $F - G$ at a certain level is typically much smaller if we consider empirical measures. By empirical we mean that $X_1, \dots, X_n$ are random samples drawn independently from a distribution F on $\mathbb{R}$, and the workers are uniformly distributed on $\{X_1, \dots, X_n\}$, and similarly for the jobs. In this case, we may further refine the bound for the runtime of our algorithm, as is remarked below.

Proposition 9. Suppose that F, G are independent empirical measures of the uniform distribution on $[0, 1]$. Then the runtime of the algorithm is $O(n^{2.5}(\log \log n)^{3/2})$ almost surely.

In order to prove Proposition 9, consider the (random) empirical cumulative densities F_n, G_n, drawn from two independent sequences $\{X_i\}_{1\leqslant i\leqslant n}$ and $\{Z_i\}_{1\leqslant i\leqslant n}$ uniformly in $[0,1]$, i.e.,

$$F_n(t) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k\leqslant t\}} \qquad \text{and} \qquad G_n(t) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{Z_k\leqslant t\}}.$$

It is well known that the scaled measure of underqualification $\sqrt{n}(F_n - G_n)$ can be well approximated by a Brownian bridge, where we recall that a (standard) Brownian bridge $B = \{B(t)\}_{t\in[0,1]}$ is a centered Gaussian process with covariance $\mathbb{E}[B(s)B(t)] = \min(s,t) - st$. We denote the local time of a standard Brownian bridge B on $[0,1]$ at $x \in \mathbb{R}$ by $L_B(x)$. By definition, the local time process $\{L_B(x)\}_{x\in\mathbb{R}}$ is such that for any bounded Borel function f,

$$\int_0^1 f(B(t))\,\mathrm{d}t = \int_{\mathbb{R}} f(x)L_B(x)\,\mathrm{d}x.$$

The following Lemma is a special case of Theorem 5 of Khoshnevisan (1992).

Lemma 15 (Theorem 5 of Khoshnevisan (1992))**.** There exists a suitable probability space carrying F_n, G_n, and a sequence of Brownian bridges $\{B_n\}$, such that

$$\lim_{n\to\infty} \max_{k\in\mathbb{Z}} \left| n^{-1/2} C_{F_n - G_n}\left(\frac{k}{n}\right) - \sqrt{2} L_{B_n}\left(\frac{k\sqrt{2}}{\sqrt{n}}\right)\right| = O(n^{-0.24}) \ \text{ a.s.}$$

We also have the following Lemma on fluctuations of the local time for Brownian bridges. This is taken from Lemma 3.2 of Bass and Khoshnevisan (1995) applied with $n_k = k$ and $\varepsilon_n = \sqrt{2/n}$.

Lemma 16. Let $\{B_n\}$ be any sequence of Brownian bridges. It holds that

$$\sup_{|x-y|<\sqrt{\frac{2}{n}}} |L_{B_n}(x) - L_{B_n}(y)| = O(n^{-0.24}) \ \text{ a.s.}$$

With a Borel-Cantelli argument in Csörgő, Shi, and Yor (1999) applied to the sequence of Brownian bridges $\{B_n\}$ (with the tail estimates supplied by Theorem 5.1 therein), the following lemma can be similarly established as Theorem 1.4 of Csörgő, Shi, and Yor (1999).

Lemma 17. Let $\{B_n\}$ be any sequence of Brownian bridges. There is a constant $C > 0$ such that

$$\mathbb{P}\left[\int_{\mathbb{R}} L_{B_n}(x)^4 \mathrm{d}x > y\right] \leqslant \exp\left(-\frac{y^{2/3}}{C}\right).$$

Moreover,

$$\int_{\mathbb{R}} L_{B_n}(x)^4 \mathrm{d}x = O\left((\log\log n)^{3/2}\right) \ \text{ a.s.}$$

Proof. The first claim is Theorem 5.1 of Csörgő, Shi, and Yor (1999) applied with $p = 4$. The second claim can be proved in a similar way as (3.7a) of Bass and Khoshnevisan (1995). $\square$

Proof of Proposition 9. Recall that solving (A.40) requires N^4 steps. In view of Proposition 8, the run-time of our algorithm has the upper bound

$$\sum_{|k|\leqslant n} C_{F-G}\left(\frac{k}{n}\right)^4.$$

Using Lemma 15, we get that almost surely,

$$C_{F-G}\left(\frac{k}{n}\right) = \sqrt{2n} L_{B_n}\left(\frac{\sqrt{2}k}{\sqrt{n}}\right) + O(n^{0.26}).$$

Therefore, using the elementary inequality $(A+B)^4 \leqslant 16(A^4+B^4)$ we have almost surely,

$$\begin{aligned}\sum_{|k|\leqslant n} C_{F-G}\left(\frac{k}{n}\right)^4 &\leqslant 64 \sum_{|k|\leqslant n}\left(n^2 L_{B_n}\left(\frac{\sqrt{2}k}{\sqrt{n}}\right)^4 + O(n^{1.04})\right) \\ &\leqslant O(n^{2.04}) + 1024 n^{2.5}\left(\int_{-\sqrt{n}}^{\sqrt{n}+1/\sqrt{n}} L_{B_n}(x)^4 \mathrm{d}x + \sum_{|k|\leqslant n}\frac{1}{\sqrt{n}} O(n^{-0.96})\right) \\ &\leqslant O(n^{2.04}) + 1024 n^{2.5}\int_{\mathbb{R}} L_{B_n}(x)^4 \mathrm{d}x,\end{aligned}$$

where we applied Lemma 16 in the second inequality. Applying now Lemma 17 concludes the proof. $\square$