Sorting with Team Formation[*]

Job Boerma

University of Wisconsin-Madison

Aleh Tsyvinski

Yale University and NES

Alexander P. Zimin

MIT and HSE

September 2021

Abstract

We fully solve an assignment problem with heterogeneous firms and multiple heterogeneous workers whose skills are imperfect substitutes, that is, when production is submodular. We show that sorting is neither positive nor negative and is characterized sufficiently by two regions. In the first region, mediocre firms sort with mediocre workers and coworkers such that output losses are equal across all these pairings (complete mixing). In the second region, high skill workers sort with a low skill coworker and a high productivity firm, while high productivity firms employ a low skill worker and a high skill coworker (pairwise countermonotonicity). The equilibrium assignment is also necessarily characterized by product countermonotonicity, meaning that sorting is negative for each dimension of heterogeneity with the product of heterogeneity in the other dimensions. The equilibrium assignment as well as wages and firm values are completely characterized in closed form. We illustrate our theory with an application to show that our model is consistent with the observed dispersion of earnings within and across U.S. firms. Our counterfactual analysis gives evidence that the change in the firm project distribution between 1981 and 2013 has a larger effect on the observed change in earnings dispersion than the change in the worker skill distribution.

JEL-Codes: J01, D31, C78.

Keywords: Sorting, Team Formation, Assignment.

1 Introduction

How do heterogeneous workers sort to work in teams at heterogeneous firms? What individuals work together, and what firm do they work for? How do worker earnings depend on their employer, their coworkers, as well as other firms and workers in the economy? How does the firm value vary with the labor they hire? To answer these questions, we study the sorting of multiple heterogeneous workers into teams at heterogeneous firms when workers' skills are imperfect substitutes.

The traditional approach to one worker-one firm assignment problems is to provide conditions under which sorting is positive or negative. Becker (1973) shows that optimal sorting is positive under supermodular technologies and negative under submodular technologies. For many-to-one assignment problems, such as sorting multiple workers into teams to work at firms, Kremer (1993) shows that the analysis of Becker (1973) straightforwardly extends when the technology is supermodular, in which case equilibrium sorting is positive.

An important open question is how to characterize sorting of heterogeneous firms with multiple heterogeneous workers when technology is submodular. While positive sorting, or comonotonicity, easily extends to matching with multiple workers, Ahlin (2017), Chade and Eeckhout (2018), and Eeckhout (2018) argue that negative sorting, or countermonotonicity, does not have such a simple multi-type counterpart. Galichon (2021) posits that considering multi-type sorting problems in economic applications is a considerable outstanding challenge. Bonhomme (2021), describing the literature on sorting and teams, concludes his review by stating that incorporating complementarities in matching problems with many types is an important venue for future research. We show that this problem can be analyzed as a multimarginal optimal transport problem and fully characterize the solution in closed form when production is submodular.

We are the first to fully solve a sorting problem with firm heterogeneity and multiple heterogeneous workers whose skills are imperfect substitutes, that is, when production is submodular. In doing so, this paper makes two contributions. First, we show that an equilibrium assignment is sufficiently characterized by two types of assignment regions. In the first region, mediocre firms sort with mediocre workers and coworkers such that output losses are equal across all these pairings (*complete mixing*). In the second region, high skill workers sort with a low skill coworker and a high productivity firm, while high productivity firms employ a low skill worker with a high skill coworker (*pairwise countermonotonicity*). The equilibrium assignment is also necessarily char-

acterized by product countermonotonicity, meaning that sorting is negative in every dimension of heterogeneity with the product of heterogeneity in the remaining dimensions. For example, a worker's skill is negatively sorted with the product of firm productivity and their coworkers' skills. In contrast to most of the assignment literature, sorting is neither positive nor negative. Second, we characterize the equilibrium assignment as well as workers' wages and firm values in closed form. We illustrate the use of our theory with a quantitative application to earnings dispersion within and across U.S. firms.

We develop our findings using a transferable utility assignment model with worker peer effects and firm heterogeneity. To illustrate our theory, and introduce the notions of product- and pairwise countermonotonicity, consider a simple production and team technology.[1] The output of a firm is the product of the value of their project, or productivity z, and the probability their worker team is able to solve production problems q. When workers independently do not know how to solve a problem with probability x_i, team quality with workers x_1 and x_2 is $q = 1 - x_1 x_2$, and output is:

$$y(x_1, x_2, z) = zq = z(1 - x_1 x_2). \tag{1}$$

Production is supermodular in team quality and productivity, and submodular in worker skills and productivity. Workers are imperfect substitutes, with peer effects in output scaling with the project value.

To characterize equilibrium in our assignment model with worker and firm heterogeneity under technology (1), and to relate our findings to prior work, we proceed in three steps. We first discuss one-to-one assignment models nested in our framework to contrast our setup and findings to prior work. We then analyze a discretized version of our problem to formulate a conjecture for the set of potential matches under an equilibrium assignment. Finally, we formally characterize the solution to the continuous version of our problem.

Absent any interaction between different workers at the same firm, equilibrium is characterized by positive sorting when the production function is supermodular, as in Becker (1973). If output is supermodular, or workers and firms are complements, it is optimal to pair high-skill workers with high-skill firms, and low-skill workers with low-skill firms. Instead, if a high-skill worker is paired

---

[1]In Section 2, we show the logic extends to any number of workers per firm, and general submodular production functions. Consistent with this setup, Grossman and Maggi (2000) argue that a submodular production function accurately characterizes processes that require problem solving. In their example, it only takes one good idea of two software engineers to solve a programming problem. In Sah and Stiglitz (1986), team quality is the probability of correctly approving projects when workers are organized in a polyarchy.

with a low productivity firm, and a low-skill worker is paired with a high-productivity firm, output can be increased by pairing the high skill worker and high productivity firm together and pairing the low skill worker and low productivity firm together. By the definition of supermodularity, this improves output. The same logic applies to negative sorting when production is submodular.

How do the sorting patterns look when multiple workers work and interact at the same firm? When production is supermodular in worker qualities and firm productivity, as is the case for his "O-Ring" technology, Kremer (1993) shows that the argument of Becker (1973) straightforwardly extends. Equilibrium is characterized by positive sorting, or self-matching, meaning that workers are paired with identical coworkers. If any two pairings of workers and firms are not positively sorted in all dimensions, again, the definition of supermodularity tells that the positively sorted assignment is output improving. Equilibrium thus features heterogeneity of workers and earnings across firms since better firms pair with better workers. There is no worker and earnings heterogeneity within firms, however, since the workers are identical within firms.

In this paper, we instead characterize sorting when multiple workers work and interact at firms with a submodular production technology. How does negative sorting generalize to dimensions exceeding two? In our model, importantly, the assignment is very different from the one-to-one sorting pattern. To build intuition, we work with a discrete sample of workers and firms. We study both the planning problem and the dual problem for the discrete samples to highlight the main theoretical insights: the support for an equilibrium assignment and the characterization of the dual formulation to understand equilibrium wages and firm values.

It is useful to first derive a simple necessary condition − product countermonotonicity − for an optimal assignment. It is obtained from the minimal requirement of stability. By product countermonotonicity we mean that when the firm's project value z is greater, the product of their workers' skills, $x_1 x_2$, is lower. For any two triplets of worker, coworker, and firm, it cannot be beneficial to bilaterally exchange either one or both workers between the firms. Given the production technology (1), this condition translates into product countermonotonicity. The requirement that bilateral exchange of both workers between the two firms cannot be beneficial, for example, implies that the product of workers' skills, $x_1 x_2$, is lower when firm productivity z is greater. Thus, more productive firms work with better teams. Importantly, this does not imply both workers are individually more skilled as is the case under positive sorting. Similarly, if a worker is more skilled, they work with a coworker and firm that all else equal have greater expected losses. The worker's

marginal contribution is high either because the project is valuable or because the coworker is less skilled.

Deriving a sufficient condition and thus fully characterizing an optimal assignment is significantly more involved. We show that there are two regions to an optimal assignment. The first region satisfies complete mixing. Complete mixing means the expected output loss, $x_1 x_2 z$, is identical for all triplets. We start with the observation that complete mixing is optimal whenever it is feasible. Equalizing expected output loss $x_1 x_2 z$ across triplets is optimal as average expected losses (a geometric mean) attain their lower bound if and only if output losses are equal across all triplets following the inequality of arithmetic and geometric means. Complete mixing implies that firms with identical projects may hire different worker pairings with the same implied team quality. Low value projects that are completely mixed are, necessarily due to product countermonotonicity, assigned low quality teams.

Complete mixing does not fully describe the equilibrium assignment because it is not feasible everywhere. Equilibrium instead features the maximal completely mixed set combined with the second region of the pairwise countermonotonic extrema. To develop intuition for the optimality of pairwise countermonotonic sets, we note that at the extreme of the completely mixed set, the lowest value project is matched with identical low-skill workers. Extending this logic to lower valued projects, projects which cannot be completely mixed, we find that these are sorted with identically low-skill workers. In the limit, two workers who never know how to solve a problem work on a project which has no value.

To establish equilibrium properties for worker wages and firm values we characterize the dual problem. We show that multipliers on the constraints for the planning problem are equilibrium wages and firm values. Then we show that the derivative of the wage schedule is equal to the marginal worker product, or $-x_2 z$. The increase in output obtained by replacing a worker with a slightly more skilled worker, keeping their coworker and their firm unchanged, has to equal the increase in wages necessary to hire the higher-skill worker. Since more skilled workers work with a coworker and firm that has greater expected output losses by product countermonotonicity, the marginal worker product is increasing, implying a convex wage schedule.

Armed with insights from the discretized version of our economy, we then prove that any assignment with its support on the complete mixed and pairwise countermonotonic set is indeed an equilibrium assignment with continuous marginal distributions. We first establish that there

exists a probability measure concentrated on this support. To prove existence, we connect our assignment model to work on risk aggregation by Wang and Wang (2016) that describes when there exists a dependence structure, a measure, such that the sum of idiosyncratic risks is constant. We then prove optimality of our assignment by using the integral version of Karamata's inequality with respect to output losses implied by alternative assignments. Having characterized the equilibrium assignment, we use a duality argument for multidimensional optimal transport problems, see Kellerer (1984), to start deriving that the marginal wage is the marginal worker product. To characterize the wages and firm values, we combine our description of optimal assignments with the observation that the marginal worker product is given by the product of the coworker's skills and project value. This allows us to write equilibrium wages only as a function of a worker's skill.

Having characterized the model equilibrium in closed form, we quantitatively illustrate the theory using administrative U.S. earnings records. Since a distinctive feature of our work among assignment models is that we obtain distinct distributions of earnings within heterogeneous firms, we apply our model to evaluate the dispersion in earnings within and across firms, which were empirically documented by Barth, Bryson, Davis, and Freeman (2016) and Song, Price, Guvenen, Bloom, and Von Wachter (2019) for the United States, by Card, Heining, and Kline (2013) for Germany, and by Alvarez, Benguria, Engbom, and Moser (2018) for Brazil. Song, Price, Guvenen, Bloom, and Von Wachter (2019) attribute about 40 percent of the variation in log earnings, a total of 0.8 log points, to between-firm dispersion and 60 percent to within-firm dispersion. From 1981 to 2013, they document an increase in the variance of log earnings of about 0.2 log points, two-thirds of which is driven by an increase in earnings dispersion between firms.

Our model endogenously matches the decomposition of earnings dispersion into between and within-firm components. We invert the closed-form wage equation to back out the workers' skill distribution using only the distribution of earnings, that is, without using data on their coworkers. We then compare the model decomposition of earnings variation to their empirical counterpart. By construction, our model replicates the total dispersion of earnings in the data. We then find that the stylized model correctly attributes about two-thirds of earnings variation in 1981 and 2013 to within-firm variation. Our counterfactual analysis gives evidence that the change in the firm project distribution between 1981 and 2013 has a larger effect on the observed change in earnings dispersion than the change in the worker skill distribution.

Related Literature. There is an extensive literature on one-to-one assignment models in the tradition of Becker (1973), which is surveyed in Sattinger (1993), Chiappori and Salanié (2016), Chade, Eeckhout, and Smith (2017), and Eeckhout (2018). This literature, in particular, analyzes conditions under which sorting is positive or negative. Our contribution is to characterize equilibrium sorting for a multimarginal assignment problem with heterogeneous firms and multiple heterogeneous workers whose skills are imperfect substitutes. We provide a complete equilibrium characterization for this problem, and show this equilibrium features neither positive nor negative sorting.

While the one-to-one assignment models described so far sort two exogenous distributions, such as workers and firms, related work considers only a single population of workers which also has to be split into different roles at different teams. Kremer and Maskin (1996) study sorting of heterogeneous workers into managerial and worker roles. In this case, production is supermodular yet asymmetric, which proves to be a challenge in terms of characterizing equilibrium in case of a single distribution. Porzio (2017) uses optimal transport, similar to us, to extend this framework with a continuous technological investment choice for each worker and manager which scales the strength of the complementary between them. Different from these papers, our team technology is symmetric between workers, so there are no distinct roles within a team, and our economy is endowed with an exogenous distribution of firm productivity that simultaneously has to be matched to teams. Our problem thus becomes a many-to-one assignment model rather than one-to-one assignment model, but we can nonetheless characterize the assignment as well as wages and firm values in closed form.

The closest to our work is a small literature that studies negative sorting with more than two agents. Ahlin (2017), Chade and Eeckhout (2018) and Eeckhout (2018) argue it is neither evident how to define negative sorting in this case, nor how to characterize an optimal assignment when production is submodular, and also derive insightful partial characterizations of the equilibrium. Ahlin (2017) also develops a careful quantitative application to group formation in villages.[2] In our paper, we fully characterize optimal sorting for dimensions greater than two for a submodular production technology.

Kremer (1993) departs from the one-to-one assignment framework by analyzing the assignment

---

[2]See also Ahlin (2015) for a study of a role of group sizes in group lending and Saint-Paul (2001) for an assignment model with different types of intrafirm spillovers.

of multiple heterogeneous workers to a single firm with a fixed size. Kremer (1993) shows that the analysis Becker (1973) straightforwardly extends when the production technology is supermodular. Recent work by Herkenhoff, Lise, Menzio, and Phillips (2018) and Jarosch, Oberfield, and Rossi-Hansberg (2021) quantitatively examines production and learning of workers and coworkers in teams in labor markets with and without frictions.

An alternative approach to multiworker firms in assignment models is due to Eeckhout and Kircher (2018).[3] They develop a model in which firms sort with a single type of worker, and choose how many workers of this type to employ. Eeckhout and Kircher (2018) develop conditions under which sorting is positive or negative. Importantly, they assume the firm technology is additively separable between workers of different types, so the marginal worker product is independent from other types of workers within the firm. This paper differs from Eeckhout and Kircher (2018) by focusing on firms of fixed size similar to Kremer (1993), and by incorporating peer effects. Workers are instead imperfect substitutes under a submodular production technology, and their marginal product does depend on their coworkers.

This paper relates to a growing literature which uses optimal transport theory to solve economic problem, see Galichon (2018) for a comprehensive overview. Specifically, there is an important line of work on multidimensional sorting and multi-market problems, working with optimal transport theory, such as Choo and Siow (2006), Dupuy and Galichon (2014), McCann, Shi, Siow, and Wolthoff (2015), Lindenlaub (2017), Chiappori, McCann, and Pass (2017), Chiappori, Salanié, and Weiss (2017), Ocampo Díaz (2018), Edmond and Mongey (2019), Lindenlaub and Postel-Vinay (2020) and Galichon and Salanié (2021).[4] Our assignment problem instead results in a sorting problem where each dimension of heterogeneity sorts with an endogenous multidimensional distribution.

Another literature, following Garicano and Rossi-Hansberg (2006), solves hierarchical assignment models with heterogeneous workers.[5] Similar to Garicano (2000), workers differ in their knowledge, which governs the probability that they know how to solve a production problem. Knowledge is assumed to be cumulative, so that more skilled workers know how to solve a prob-

---

[3] Kelso and Crawford (1982) provide the gross substitutes conditions for equilibrium existence in a related many worker-to-one firm assignment model with finite workers and firms and a more general production technology.

[4] Related to this is research on the identification of multi-attribute hedonic models using optimal transport, see Chernozhukov, Galichon, Henry, and Pass (2021).

[5] A simple exposition of the hierarchical assignment problem is presented in Garicano and Rossi-Hansberg (2004). Antras, Garicano, and Rossi-Hansberg (2006) apply the knowledge-based hierarchy structure to study international trade and the international structure to production. See Garicano and Rossi-Hansberg (2015) for a review.

lem whenever less skilled workers do. A key implication is that production is supermodular in worker skill, so that equilibrium sorting is positive. In line with these papers, our workers differ in the probability that they know how to solve production problems. Knowledge, however, is not cumulative, allowing for the possibility that less skilled workers know how to solve a problem when more skilled workers do not. As a result, our production technology is instead not supermodular, and hence optimal sorting is not positive.

To characterize our equilibrium, we build on the optimal transport literature in mathematics that develops tools to solve multimarginal problems, see, for example, a review in Pass (2011). He concludes that the extension of the classical Monge-Kantorovich problem to the case of more than two marginal distributions is not well understood. While this is an active area of recent research with a number of applications, understanding of the problem is far from complete and is fractured (for examples of different settings and costs, see Gangbo and Świech (1998), Carlier and Ekeland (2010), Cotar, Friesecke, and Klüppelberg (2013), Kim and Pass (2014), Colombo, De Pascale, and Di Marino (2015), and Gladkov and Zimin (2020)). Our paper studies a submodular cost function and introduces firm production, non-uniform distributions of workers and firms, and uses this to understand equilibrium sorting, worker wages, and firm values. In the literature on risk aggregation and insurance, Bernard, Jiang, and Wang (2014), Embrechts, Puccetti, Rüschendorf, Wang, and Beleraj (2014), Puccetti and Wang (2015) and Wang and Wang (2016) study properties of aggregate risk as the sum of individual risks. Similar in spirit, we study the minimization of output losses (risks) incurred by teams of workers and firms in the economy. Different from them, our interest lie in distribution of output losses and the sorting patterns of workers that generate maximum aggregate output which we characterize using optimal transport theory. Moreover, we use the dual transport problem to characterize the distribution of equilibrium wages and firm values.

2   Model

We study an economy in which agents with heterogeneous skills choose their firm and coworker to work with. This is a static assignment problem extended to incorporate the sorting of two workers to each firm. Our setup results in a sorting problem between two workers with each firm for a submodular production technology. In this section, we set this problem as a Monge-Kantorovich multimarginal optimal transport problem.

2.1 Environment

Agents. The economy is populated by risk-neutral workers and firms. There exist two groups of workers with mass equal to 1 each and a continuum of firms with mass equal to 1. Workers differ in skills such that high skill workers are more likely to know how to solve a problem.[6] The skill of every worker is indexed by a single number $x_i \in X := [0, 1]$, which captures the probability that the worker does not know how to solve a problem. To illustrate, $x_i = 0.10$ implies that the worker knows how to solve a problem with 90 percent probability. High skill workers have a low x_i. Each group of workers are distributed according to the cumulative distribution function $F_x(x_i)$, with a corresponding continuous density function $f_x(x_i)$.

Firms are indexed by a productivity type $z \in Z := [0, 1]$, capturing the value of their projects. High value firms have a high z. Firms are distributed according to cumulative distribution function $F_z(z)$, with a continuous density function $f_z(z)$. We assume that cumulative distribution functions are strictly increasing and continuous, so that the inverse distribution function, which we denote by I_z, is defined so that $I_z(p_z)$ for percentile $p_z \in [0, 1]$ is the unique number z such that $F_z(z) = p_z$. Thus, p_z is the percentile corresponding to project value z in the firm distribution. The inverse distribution functions for workers I_x are defined analogously.

Technology. Firms produce a single homogeneous good. Production requires a project and a team of workers. Teams solve problems that arise in order to produce. In particular, each team encounters a single problem in production. Firm output is the project value z when the problem is solved, and zero otherwise.

A firm of type z that employs a team of quality q, which represents the probability with which the team solves production problems, produces expected output according to:

$$g(q, z) = qz. \tag{2}$$

The production function g is supermodular in team quality q and firm type z. Team quality and the project value are complements in production.

A team can solve a problem if and only if at least one of the workers knows how to solve it. Since the probability that worker i does not know the solution is x_i, independent of their coworker

---

[6]Worker heterogeneity in terms of probability to know how to solve a problem, or complete a task, is found in Kremer (1993), Garicano (2000), and Garicano and Rossi-Hansberg (2006), amongst others.

knowing the solution, the probability of successful production with workers x_1 and x_2 is:

$$q = h(x_1, x_2) = 1 - x_1 x_2. \tag{3}$$

The team quality function h is a symmetric, submodular function in worker types, so workers are substitutes. Team production of this nature appears, amongst others, in Sah and Stiglitz (1986) and Grossman and Maggi (2000).

Combining the production technology (2) with the team quality function (3), the firm output function y is:

$$y(x_1, x_2, z) = (1 - x_1 x_2)z. \tag{1}$$

The firm output function is a submodular function in worker and firm type. The marginal product of worker x_1 is defined as:

$$m(x_1) := y_1(x_1, x_2, z) = -x_2 z \leq 0. \tag{4}$$

When a worker is less skilled, so less likely to know how to solve a problem, expected output decreases. The marginal product of every worker exclusively depends on the skills of their coworker and the firm they work with. That is, the model is set up so that the marginal product only depends on peer effects.

Our specification of technology nests standard one-to-one assignment models with positive and negative sorting. If the distribution for team quality q is exogenous, sorting between teams and firms is positive. As in Becker (1973), high quality teams work on the most valuable projects. On the other hand, absent firm heterogeneity, $z = \bar{z}$, sorting is negative. Low and high skill workers form teams, as discussed in Grossman and Maggi (2000) and Legros and Newman (2002).

Assignment. An assignment function prescribes for every worker both a coworker to work with and a firm to work for. Formally, an assignment is a probability measure π over workers, coworkers and firms. Given a distribution of workers F_x, a distribution of coworkers F_x, and a distribution of firms F_z, the set of feasible assignment functions is $\Pi := \Pi(F_x, F_x, F_z)$ which is the set of probability measures π on the product space $X \times X \times Z$ such that the marginal distributions of π onto X and Z are equal to F_x and F_z respectively. Feasibility of an assignment function is thus equivalent to labor market clearing, or all workers and firms are sorted.

10

2.2 Equilibrium

A firm with the project value z chooses worker x_1 and coworker x_2 to maximize profits taking the wage schedule for workers w as given. The firm problem is:

$$v(z) = \max_{x_1, x_2} \big(y(x_1, x_2, z) - w(x_1) - w(x_2)\big). \tag{5}$$

Workers choose to work for firm z with a coworker x_2 to maximize their wage income. The worker takes the wage schedule for the coworker and the firm value v as given:

$$w(x_1) = \max_{x_2, z} \big(y(x_1, x_2, z) - w(x_2) - v(z)\big). \tag{6}$$

Similarly, coworkers choose to work with a firm z and a worker x_1 to maximize their wage income, taking the wage schedule for the worker and the firm value v as given:

$$w(x_2) = \max_{x_1, z} \big(y(x_1, x_2, z) - w(x_1) - v(z)\big). \tag{7}$$

Equilibrium. An equilibrium is a wage schedule w, firm value function v, and feasible assignment function π, such that firms solve their profit maximization problem (5), workers solve the worker problem (6), coworkers solve the coworker problem (7), and a feasibility constraint:

$$\int y(x_1, x_2, z) d\pi = \int w(x_1) dF_x + \int w(x_2) dF_x + \int v(z) dF_z, \tag{8}$$

which states that the total quantity of output produced, $\int y(x_1, x_2, z) d\pi$, equals the total quantity of output distributed to workers and firms.

2.3 Planning Problem

We solve a planning problem to characterize the equilibrium assignment. The planning problem is to choose an assignment to maximize production:

$$\max_{\pi \in \Pi} \int y(x_1, x_2, z) d\pi(x_1, x_2, z) = \max_{\pi \in \Pi} \ \mu_z - \int x_1 x_2 z d\pi(x_1, x_2, z) \tag{9}$$

where the equality follows by the production technology (1) and $\mu_z := \int z dF_z$ is the mean project value. The planning problem to maximize expected output is equivalent, in terms of choosing an optimal assignment, to a planning problem that minimizes expected output losses. Expected output losses are the product of the project value z and the probability of failure by a team of

workers which is given by $x_1 x_2$, that is, a loss $\ell(x_1, x_2, z) := x_1 x_2 z$. When we characterize the optimal assignment, we can equivalently represent the planning problem as:

$$\min_{\pi \in \Pi} \int \ell(x_1, x_2, z) \mathrm{d}\pi(x_1, x_2, z), \tag{10}$$

which is an optimal transport problem in the tradition of Monge (1781) and Kantorovich (1942), with multiple marginal distributions as in Pass (2011).

The key issue to note is that our production technology (1) is submodular in both worker and firm types. This specification differentiates our economy from Kremer (1993), which features a supermodular technology, the O-Ring technology, which corresponds to a firm technology under which both workers need to know the solution to the problem in order to produce. Kremer (1993) shows that positive sorting straightforwardly extends from a one-to-one sorting environment to a multimarginal environment when production is supermodular. In contrast, we show that negative sorting does not extend to a multimarginal environment when the production function is submodular, and that the optimal sorting is significantly more involved.

Discussion. Before proceeding to the dual formulation, we pause to make two points. First, we emphasize that the production technology (1) can be generalized substantially. The production function can include any linearly additive function of coworker skills x_i and firm value z, leaving the equilibrium assignment unaffected. This addition does not affect expected output losses so the equilibrium assignment can still be obtained from the planning problem (10).

The production technology can further be generalized to incorporate interaction terms between workers and coworkers $x_1 x_2$, workers and firms $x_1 z$, and coworkers and firms $x_2 z$, as long as the technology is submodular. To see why, note that any production technology with some coefficients φ satisfying $(\varphi_1, \varphi_2, \varphi_3) \geq 0$ which ensure submodularity:

$$y(x_1, x_2, z) = z(1 - x_1 x_2) - \varphi_1 x_2 z - \varphi_2 x_1 z - \varphi_3 x_1 x_2 - \varphi_4 x_1 - \varphi_5 x_2 - \varphi_6 z - \varphi_7, \tag{11}$$

can, using a change of variables, be written as:

$$\tilde{y}(\tilde{x}_1, \tilde{x}_2, \tilde{z}) = \tilde{z}(1 - \tilde{x}_1 \tilde{x}_2) + \varphi_{x_1} \tilde{x}_1 + \varphi_{x_2} \tilde{x}_2 + \varphi_z \tilde{z} + \varphi_y, \tag{12}$$

where $\tilde{x}_1 = x_1 + \varphi_1, \tilde{x}_2 = x_2 + \varphi_2, \tilde{z} = z + \varphi_3 \geq 0.$[7] Because the optimal assignment is unaffected by linearly additive functions of coworker skills and firm values, the solution to this problem is

---

[7] The coefficients are determined as $\varphi_{x_1} = \varphi_2 \varphi_3 - \varphi_4$, $\varphi_{x_2} = \varphi_1 \varphi_3 - \varphi_5$, $\varphi_z = \varphi_1 \varphi_2 - \varphi_6$, and $\varphi_y = \varphi_1 \varphi_2 \varphi_3 - \varphi_1 \varphi_{x_1} - \varphi_2 \varphi_{x_2} - \varphi_3 \varphi_z - \varphi_3 - \varphi_7$.

identical to the minimizing argument for modified losses $\tilde{x}_1\tilde{x}_2\tilde{z}$. This problem, in turn, has the same qualitative features as our baseline problem. To a first order, the observations imply that our results extend to any technology with all potential interaction effects $(x_1x_2, x_2z, x_1z, x_1x_2z)$ for which the first-order Taylor expansion around these interaction effects is submodular.

Finally, while we choose to work with teams of two workers for expositional purposes, none of our arguments are limited to this case. Our results extend to any number of workers greater than two as we show in Appendix B.[8]

Dual Problem. The solution to the planning problem is an assignment, which we argue below is an equilibrium assignment. To obtain the corresponding equilibrium wage schedule and firm value function, we solve a dual problem. The dual formulation for the assignment planning problem is to choose functions w and v that solve:

$$\min \int w(x_1)\mathrm{d}F_x + \int w(x_2)\mathrm{d}F_x + \int v(z)\mathrm{d}F_z, \tag{13}$$

subject to the constraint that $w(x_1) + w(x_2) + v(z) \geq y(x_1, x_2, z)$ for any $(x_1, x_2, z) \in X \times X \times Z$. To simplify notation, we define the surplus function S as output minus payments to workers and firms:

$$S(x_1, x_2, z) := y(x_1, x_2, z) - w(x_1) - w(x_2) - v(z). \tag{14}$$

The constraint to the dual formulation is that the surplus function $S(x_1, x_2, z)$ is negative for any triplet $(x_1, x_2, z) \in X \times X \times Z$.

We use the following relation between the planning problem and the dual formulation.

Lemma 1. Let $\pi \in \Pi(F_x, F_x, F_z)$ be a joint probability measure, and w and v be functions such that $S(x_1, x_2, z) \leq 0$ for all (x_1, x_2, z). If there is a set $M \subset X \times X \times Z$ such that $S(x_1, x_2, z) = 0$ on M with the additional property that $\pi(M) = 1$, then assignment π is a planning solution and functions w and v are a dual solution.

---

[8]We observe that an economy with two identical worker distributions F_x with mass equal to one for each role is equivalent to an economy with a single distribution of workers F_x with mass equal to two. Owing to the symmetry of worker skills in the team quality function (3), these are equivalent. Intuitively, any equilibrium worker assignment for a planning problem with distinct worker distributions can be made symmetric. By assigning half of the mass of equilibrium worker pairings to one distribution and half to another, we obtain two identical worker distributions. The optimum value is unaffected as the same distribution of worker pairings can be made due to symmetry. When the distribution of worker skills and firms projects are identical, this argument also implies that our framework is a self-matching problem with fixed group size, which is the problem studied by Ahlin (2017).

The proof to Lemma 1 only makes use of a notion of weak duality and is presented in Appendix A.1. Any probability measure, together with wage and firm value functions satisfying the dual constraint for any triplet (x_1, x_2, z), solve the planning and dual problem provided there is a set M on which the assignment is concentrated and on which the surplus, or excess resources, are zero. Since the assignment in Lemma 1 is concentrated on set M, we refer to M as the set of potential matches.

While Lemma 1 connects the planning problem and the dual formulation, it does not connect either to our definition of an equilibrium. The link is Proposition 1.

Proposition 1. *Planning and Equilibrium Allocation.* Let π be a solution to the planning problem on set M of potential matches and let (w, v) solve the dual problem. Then, wage schedule w, firm value function v, and assignment function π are an equilibrium.

The proof is in Appendix A.2. Proposition 1 implies that assignment function π, wage schedule w, and firm value function v are an equilibrium. Having established this connection, we henceforth refer to w as the equilibrium wage schedule and to v as the equilibrium firm value. The proposition proposes a clear path for characterizing equilibrium. We separately solve the planning problem and the dual problem, which are linked through the primitive project value distribution, worker distribution, and production technology. Moreover, Proposition 1 shows that the decentralized equilibrium coincides with the solution to the planning problem so the competitive equilibrium is efficient. We now proceed to characterize the equilibrium assignment π, wage schedule w, and firm value function v.

3 Discretized Analysis

We first study equilibrium using a discretized approach to formulate our conjecture for the set of potential matches, the set on which an equilibrium assignment is concentrated, and to understand equilibrium wages.

3.1 Planning Problem

Discrete Planning Problem. We formulate a discrete approximation to the planning problem given worker samples $\{x_{is}\}_{s=1}^n \sim F_x$ for $i \in \{1, 2\}$ and a firm type sample $\{z_s\}_{s=1}^n \sim F_z$. Given these samples, there are $n!^2$ different combinations to pair workers $\{x_{is}\}$ to firms, or $n!$ different combinations for the worker and the coworker sample to sort with firms. The discrete planning

problem is to choose an assignment, or equivalently choose permutations σ_i, to maximize aggregate output:

$$\max_{\sigma_1, \sigma_2} \sum_s y\left(x_{1\sigma_1(s)}, x_{2\sigma_2(s)}, z_s\right). \tag{15}$$

To explicitly relate our environment to one-to-one assignment problems, and introduce concepts, we first examine the nested frameworks.

One-to-One Assignment Model. Our specification of the technology nests the standard one-to-one assignment models with positive and negative sorting. Suppose that the distribution for team quality is exogenously given. In this case, the discrete planning problem to maximize aggregate output simplifies to:

$$\max_{\sigma_q} \sum_s g\left(q_{\sigma_q(s)}, z_s\right), \tag{16}$$

where σ_q is a permutation for team quality.

An optimal assignment necessarily requires that the assignment is stable, by which we mean that aggregate output is not increased by changing the assignment of teams between firms. For any two optimal assignment pairs (q, z) and $(\hat{q}, \hat{z})$, it should hold that $g(q, z) + g(\hat{q}, \hat{z}) \geq g(\hat{q}, z) + g(q, \hat{z})$. Given the output function (2), this implies:

$$(q - \hat{q})(z - \hat{z}) \geq 0, \tag{17}$$

that is if $q \geq \hat{q}$ then $z \geq \hat{z}$. Since the production function g is supermodular, more productive teams work on more valuable projects. As discussed in Becker (1973), high quality teams optimally pair with high value projects. The optimal assignment is positive, or *comonotonic*.

Alternatively, suppose there is no firm heterogeneity, $z = \bar{z}$. In this case, the discrete planning problem to maximize aggregate output simplifies to choosing a permutation σ_1 to solve:

$$\max_{\sigma_1} \sum_s h\left(x_{1\sigma_1(s)}, x_{2s}\right). \tag{18}$$

The necessary stability condition for an optimal assignment implies that for any two optimal pairings (x_1, x_2) and $(\hat{x}_1, \hat{x}_2)$, it has to be true that $h(x_1, x_2) + h(\hat{x}_1, \hat{x}_2) \geq h(\hat{x}_1, x_2) + h(x_1, \hat{x}_2)$. Given the team quality function (3), this implies:

$$(x_1 - \hat{x}_1)(x_2 - \hat{x}_2) \leq 0. \tag{19}$$

The optimal sorting is negative, or *countermonotonic*, that is, if $x_1 \geq \hat{x}_1$, then $x_2 \leq \hat{x}_2$. Low and high skill workers form teams, as discussed in Grossman and Maggi (2000) and Legros and Newman (2002).

For one-to-one assignment models with supermodular or submodular technology, we observe that stability is both necessary and sufficient to characterize optimal sorting. We next use a notion of stability for multimarginal assignment problems, and show that while stability is still necessary and sufficient to obtain optimal comonotonicity with supermodular production, stability is no longer sufficient to characterize optimal sorting with submodular technology.

Multimarginal Assignment Model. The team-to-firm and worker-to-coworker problems are two-dimensional assignment problems. With three dimensions or more, for an assignment to be optimal also necessarily requires that the assignment is stable. In this case, by stability we mean a minimal requirement that expected output is not increased by a bilateral exchange of either one or both workers between two firms. For any two optimal assignment triplets (x_1, x_2, z) and $(\hat{x}_1, \hat{x}_2, \hat{z})$, the following inequalities should thus be satisfied:

$$y(x_1, x_2, z) + y(\hat{x}_1, \hat{x}_2, \hat{z}) \geq y(\hat{x}_1, x_2, z) + y(x_1, \hat{x}_2, \hat{z}),$$

$$y(x_1, x_2, z) + y(\hat{x}_1, \hat{x}_2, \hat{z}) \geq y(x_1, \hat{x}_2, z) + y(\hat{x}_1, x_2, \hat{z}), \tag{20}$$

$$y(x_1, x_2, z) + y(\hat{x}_1, \hat{x}_2, \hat{z}) \geq y(\hat{x}_1, \hat{x}_2, z) + y(x_1, x_2, \hat{z}).$$

When the production technology y is supermodular, the optimal assignment features positive sorting as in Kremer (1993). In this case, optimal comonotonicity straightforwardly extends to multi-worker assignment problems.[9] The most productive firms employ the highest skill workers. Under this specification, the optimal assignment thus features self-sorting meaning that workers are paired with identical coworkers. An equilibrium thus features heterogeneity of workers and earnings *across* firms since the higher quality firms match with the higher quality workers. However, there is no heterogeneity of workers and earnings *within* firms since the workers within each team are identical.

Our production technology is instead submodular. In this case, the necessary set of inequalities

---

[9]Formally, suppose (x_1, x_2, z) and $(\hat{x}_1, \hat{x}_2, \hat{z})$ are pairings of an optimal assignment. Optimal sorting is positive, so if $z > \hat{z}$, then both $x_1 \geq \hat{x}_1$ and $x_2 \geq \hat{x}_2$. By contradiction, suppose $x_1 \geq \hat{x}_1$ and $x_2 < \hat{x}_2$. By supermodularity of the output function $y(x_1, x_2, z) + y(\hat{x}_1, \hat{x}_2, \hat{z}) < y(x_1, \hat{x}_2, z) + y(\hat{x}_1, x_2, \hat{z})$, contradicting that the initial assignment is optimal by violating necessary conditions (20). Similar contradictions are obtained when $x_1 < \hat{x}_1$ and $x_2 \geq \hat{x}_2$ as well as when $x_1 < \hat{x}_1$ and $x_2 < \hat{x}_2$.

(20) is not sufficient to characterize the equilibrium assignment.[10] We can, however, use the set of inequalities (20) to make two observations. Given the functional form of the output function (1), these equations read:

$$(\hat{x}_2\hat{z} - x_2 z)(\hat{x}_1 - x_1) \leq 0,$$

$$(\hat{x}_1\hat{z} - x_1 z)(\hat{x}_2 - x_2) \leq 0, \tag{21}$$

$$(\hat{x}_1\hat{x}_2 - x_1 x_2)(\hat{z} - z) \leq 0.$$

The first observation is that better teams work on more valuable projects. When firm $\hat{z}$ holds a more valuable project than firm z, so $\hat{z} > z$, the final inequality in (21) implies that $\hat{x}_1\hat{x}_2 < x_1 x_2$. Optimal sorting is negative between the firm z and the product of the workers' skills $x_1 x_2$. Since there is complementarity between project values and team quality in the production function (2), sorting between firms and team quality is comonotonic. Similarly, if a worker is more skilled, $\hat{x}_1 < x_1$, the first inequality in (21) implies that they work with a coworker and firm that ceteris paribus has a greater expected output loss, $\hat{x}_2\hat{z} \geq x_2 z$. Equivalently, when a worker is more skilled, their marginal product m defined in (4) is more negative.[11] The marginal product increases with x, the probability that the worker does not know how to solve a problem.[12] We refer to condition (21) that the optimal sorting is product countermonotonic following Puccetti and Wang (2015).

Definition 1. Set $\{(x_{1s}, x_{2s}, z_s)\}_{s=1}^n$ is product countermonotonic if for any subset $\mathcal{D} \subset \{x_1, x_2, z\}$ the product of its components are countermonotonic with respect to the product of the remaining components.

While product countermonotonicity is a necessary condition for an optimal assignment, it is

---

[10]See, for example, footnote 13.

[11]A related observation is that the inequalities (21) present a rearrangement algorithm to numerically identify a stable assignment. Suppose we are given some set of triplets $\{(x_{1s}, x_{2s}, z_s)\}_{s=1}^n$, then the rearrangement algorithm works as follows. We update the set of triplets to $\{(x_{1s}, x_{2\sigma(s)}, z_{\sigma(s)})\}$, where σ is the perturbation that negatively assorts $\{x_{1s}\}$ with $\{x_{2s}z_s\}$ as suggested by inequalities (21). We next order the second element, $\{x_{2s}\}$, and similarly negatively sort these with the updated set of points $\{x_{1s}z_s\}$. We keep cycling through elements until the assignment converges.

[12]We observe that pairwise countermonotonicity (Dall'Aglio, 1972), an alternative potential generalization of countermonotonicity to dimensions greater than two, is not feasible. An assignment is defined as pairwise countermonotonic if $x_1 < \hat{x}_1$ implies both $x_2 \geq \hat{x}_2$ and $z \geq \hat{z}$, and symmetrically that $x_2 < \hat{x}_2$ implies $x_1 \geq \hat{x}_1$ as well as $z \geq \hat{z}$, and that $z < \hat{z}$ implies $x_1 \geq \hat{x}_1$ as well as $x_2 \geq \hat{x}_2$. To see that this configuration is not feasible, suppose that $z < \hat{z}$ without loss of generality. By countermonotonicity, and because all elements sampled from continuous distributions are distinct, $x_1 > \hat{x}_1$, which implies that $x_2 < \hat{x}_2$. We arrive at a contradiction, as this would imply $z > \hat{z}$.

generally not sufficient.[13] To identify sufficient conditions, we introduce a notion of complete mixing, as defined in Wang and Wang (2011).

Definition 2. Set $\{(x_{1s}, x_{2s}, z_s)\}_{s=1}^n$ is <u>completely mixed</u> if $x_{1s}x_{2s}z_s = \mathcal{C}$ for all s for constant $\mathcal{C}$.

An assignment is completely mixed when expected losses are constant across all triplets, $x_{1s}x_{2s}z_s = \mathcal{C}$ for all s for some constant $\mathcal{C}$. It follows immediately that an assignment that is completely mixed is product countermonotonic because $z_s = \frac{\mathcal{C}}{x_{1s}x_{2s}}$ shows higher project values z_s sort with more productive teams, that is, lower $x_{1s}x_{2s}$. The same is true for the two other combinations of teams and workers.

Importantly, a completely mixed assignment is optimal by the inequality of arithmetic and geometric means. The inequality states that the arithmetic mean for the output losses, which is proportional to the objective of the assignment problem in (10), is bounded below by the n-th root of the product of all elements. Formally, with a set of elements $\{(x_{1s}, x_{2s}, z_s)\}_{s=1}^n$, the inequality reads:

$$\frac{1}{n}\sum x_{1s}x_{2\sigma_2(s)}z_{\sigma_z(s)} \geq \sqrt[n]{\prod x_{1s}x_{2s}z_s} =: \mathcal{C}_n, \tag{22}$$

where the geometric mean, the constant on the right-hand side, depends on sampled distributions, but is invariant to any permutation.[14] Given a sample of size n, the arithmetic mean attains lower bound $\mathcal{C}_n$ if and only if the expected output loss, x_1x_2z, is equal across all triplets and equal to $\mathcal{C}_n$. When an assignment is completely mixed, the assignment is optimal as it minimizes aggregate output losses (10).

While complete mixing of triplets is optimal when it is feasible, it is not feasible everywhere. In fact, there is no optimal completely mixed assignment. For example, consider a triplet (x_{1s}, x_{2s}, z_s) part of some optimal assignment with $z_s = 0$. In this case, the constant $\mathcal{C}_n$ is zero. Under complete mixing it would therefore have to be that $x_{1s}x_{2s}z_s = 0$ for all s. More generally, one can show that there exists no optimal completely mixed assignment for atomless distributions with full support on the unit interval.

---

[13]For example, suppose the sampled distributions for worker skills x_1 and x_2 and for firm type z are identical and are equal to $\{0.1, 0.2, 0.4\}$. The assignment $\{(0.1, 0.4, 0.4), (0.2, 0.2, 0.2), (0.4, 0.1, 0.1)\}$ is product countermonotonic, but it is not optimal since the feasible assignment $\{(0.1, 0.4, 0.2), (0.2, 0.1, 0.4), (0.4, 0.2, 0.1)\}$ delivers strictly greater aggregate output (0.676 versus 0.672).

[14]Equivalently, the statement follows directly from Jensen's inequality. Taking logarithms on both sides of (22) we obtain $\log\left(\frac{1}{n}\sum x_{1s}x_{2\sigma_2(s)}z_{\sigma_z(s)}\right) \geq \frac{1}{n}\sum \log x_{1s}x_{2s}z_s$.

We now argue that the optimal assignment for our problem is instead the largest possible set of completely mixed triplets combined with pairwise countermonotonic sets. To see this, we first consider the largest possible percentile interval $[\underline{p}, \bar{p}]$, all the percentiles such that $I(\underline{p}) \leq I(p) \leq I(\bar{p})$, for which the distribution F is completely mixable. The percentile interval implies corresponding level intervals for mediocre workers, defined as workers in the skill interval $[\underline{x}_1, \bar{x}_1]$, for mediocre coworkers $[\underline{x}_2, \bar{x}_2]$ as well as for mediocre firms $[\underline{z}, \bar{z}]$. For example, the lowest project value is given by the inverse distribution function at the lower percentile cutoff, $\underline{z} = I(\underline{p})$. Let $\hat{\pi}$ denote an optimal assignment concentrated on the completely mixed set where output losses are constant with $x_1 x_2 z = \mathcal{C}$, that is, an assignment that pairs mediocre workers and coworkers with mediocre firms giving identical losses for all triplets. Since $\log x_1 + \log x_2 + \log z = \log \mathcal{C}$ almost everywhere with respect to assignment $\hat{\pi}$, integration yields the constant:[15]

$$\log \mathcal{C} = \int_{\underline{x}_1}^{\bar{x}_1} \log x_1 \mathrm{d}\hat{\pi} + \int_{\underline{x}_2}^{\bar{x}_2} \log x_2 \mathrm{d}\hat{\pi} + \int_{\underline{z}}^{\bar{z}} \log z \mathrm{d}\hat{\pi}. \tag{23}$$

The expression for the constant output loss on the completely mixed set simplifies by observing that the marginal distributions of feasible assignment $\hat{\pi}$ on X and Z are given by distribution F on the percentile interval $[\underline{p}, \bar{p}]$, scaled by $\bar{p} - \underline{p}$ to normalize to a unit mass. To illustrate, for the firm distribution this implies $\int_{\underline{z}}^{\bar{z}} \log z \mathrm{d}\hat{\pi} = \frac{1}{\bar{p}-\underline{p}} \int_{\underline{z}}^{\bar{z}} \log z \mathrm{d}F(z)$. Since the minimum and maximum project values within the completely mixed set are $\underline{z} = I(\underline{p})$ and $\bar{z} = I(\bar{p})$, $\frac{1}{\bar{p}-\underline{p}} \int_{\underline{z}}^{\bar{z}} \log z \mathrm{d}F(z) = \frac{1}{\bar{p}-\underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p$. To simplify exposition, we assume the distributions for worker and firm types are identical, so the constant is given by:

$$\log \mathcal{C} = \frac{3}{\bar{p} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p. \tag{24}$$

To describe triplets on the maximal completely mixed set, we consider a firm $z \in [I(\underline{p}), I(\bar{p})]$ and describe the workers that can work with this firm. Since the triplet is completely mixed, their losses are constant, $x_1 x_2 z = \mathcal{C}$, and the paired coworker is $x_2 = \frac{\mathcal{C}}{x_1 z}$. By Wang and Wang (2016), which we discuss more below and use in conjunction with part (a) of Assumption 1, the maximal

---

[15] Another intuitive way to obtain (23) is by considering the limit of the geometric mean in equation (22) as the sample size n tends to infinity. We observe that constant $\mathcal{C}_n = \exp\left(\frac{1}{n} \sum \log x_{1s}\right) \exp\left(\frac{1}{n} \sum \log x_{2s}\right) \exp\left(\frac{1}{n} \sum \log z_s\right)$. When the number of sample observations increases, $n \to \infty$, $\frac{1}{n} \sum \log z_s \to \int_{\underline{z}}^{\bar{z}} \log(z) \mathrm{d}F_z$, and similarly for the worker distribution. As the sample size tends to infinity, the constant for completely mixed workers and firms thus tends to $\mathcal{C} = \exp(\int_{\underline{x}}^{\bar{x}} \log(x) \mathrm{d}F_x + \int_{\underline{x}}^{\bar{x}} \log(x) \mathrm{d}F_x + \int_{\underline{z}}^{\bar{z}} \log(z) \mathrm{d}F_z)$. Since assignment $\hat{\pi}$ is feasible, so that its projections onto X and Z are respectively given by F_x and F_z on the interval $[\underline{p}, \bar{p}]$, $\mathcal{C} = \exp(\int_{\underline{x}}^{\bar{x}} \log(x) \mathrm{d}\hat{\pi} + \int_{\underline{x}}^{\bar{x}} \log(x) \mathrm{d}\hat{\pi} + \int_{\underline{z}}^{\bar{z}} \log(z) \mathrm{d}\hat{\pi})$. This is equivalent to (23).
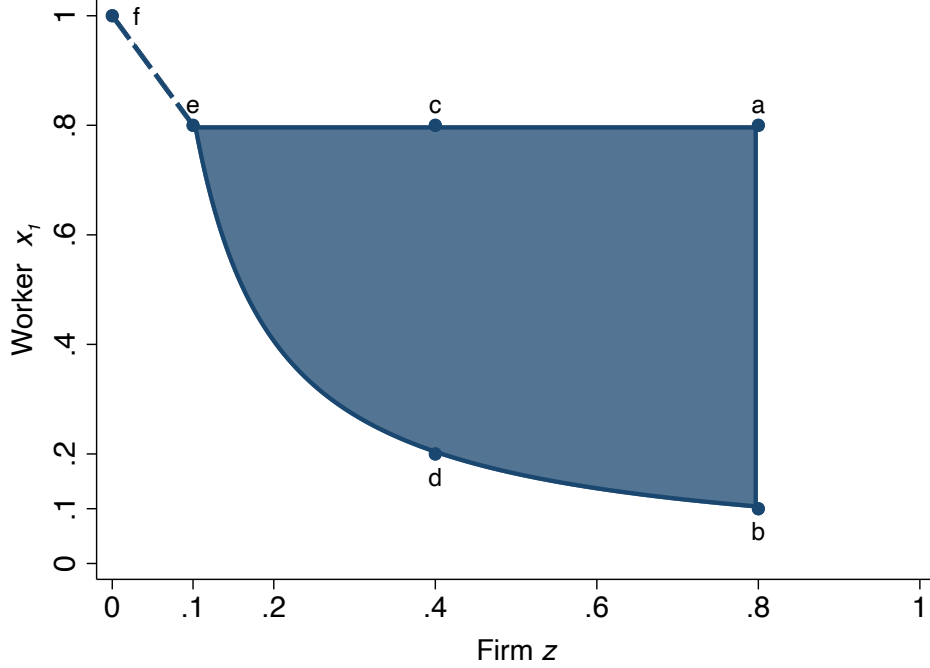
Figure 1: Set of Potential Matches

Figure 1 shows the range of worker types employable by firm $z \in [0, I(\bar{p})]$. The x-axis shows the percentile of the firm in the project value distribution while the y-axis shows the percentile range for the workers in the worker skill distribution it can employ. The description of employable workers is given by (25). For our illustration, the lower bound is $I(\underline{p}) = 0.10$ and the upper bound is $I(\bar{p}) = 0.80$.

completely mixed set satisfies $\mathcal{C} = I^2(\bar{p})I(\underline{p})$. Since both the worker and coworker lie between the lower and the upper skill boundary, $I(\underline{p}) \leq x_1, x_2 \leq I(\bar{p})$, it follows that a firm with project z can employ worker x_1 with skills that are bounded by:[16]

$$\frac{I(\underline{p})I(\bar{p})}{z} \leq x_1 \leq I(\bar{p}). \tag{25}$$

Any firm $z \in [I(\underline{p}), I(\bar{p})]$ can employ a low skill worker $I(\bar{p})$ and attain loss $\mathcal{C}$. When the project is valuable, $z = I(\bar{p})$, this requires pairing with a most skilled completely mixed coworker $I(\underline{p})$. If the project is not valuable, $z = I(\underline{p})$, this requires pairing with a least skilled completely mixed coworker $I(\bar{p})$. On the other hand, firm z cannot employ a high skill worker $x_1 < I(\underline{p})I(\bar{p})/z$ as it takes a coworker outside the completely mixed set, $x_2 > I(\bar{p})$, to attain loss $I^2(\bar{p})I(\underline{p}) = x_1 x_2 z$, if it is feasible at all.

Figure 1 provides an example to illustrate the range of worker types employable by firm z. A high type firm $z = I(\bar{p})$ can employ workers in the skill interval $x_1 \in [I(\underline{p}), I(\bar{p})]$. If the project

---

[16]Since $I(\underline{p}) \leq x_2 \leq I(\bar{p})$ and $x_2 = I^2(\bar{p})I(\underline{p})/(x_1 z)$, necessarily $I(\underline{p})I(\bar{p})/z \leq x_1 \leq I^2(\bar{p})/z$. The upper bound, however, exceeds $I(\bar{p})$ because $I(\underline{p}) \leq z \leq I(\bar{p})$, and hence the upper bound is $I(\bar{p})$.

is valuable, the firm that hires low skill worker $I(\bar{p})$ also hires coworker $I(\underline{p})$ to incur loss $\mathcal{C}$. In Figure 1, where $I(\underline{p}) = 0.1$ and $I(\bar{p}) = 0.8$ imply a constant loss $\mathcal{C} = 0.064$, when the project value is $z = 0.8$, the firm that hires low skill worker $x_1 = 0.8$ also hires coworker $x_2 = 0.1$. Similarly, the firm can hire an alternative worker and coworker pair such as (0.2,0.4), illustrated by the range between a and b. Firms with identical project values can thus employ different teams of workers. As the project value decreases, the range of potential workers shrinks from below as in (25). For example, when the project value is $z = 0.4$, the firm that hires low skill worker $x_1 = 0.8$ also hires coworker $x_2 = 0.2$, illustrated by the range from c to d. Firms with lower project values tend to employ workers with lower skills. In the limit where $z = I(\underline{p}) = 0.1$, the lowest value projects in the completely mixed set employ only a single type of worker, which have the lowest skills in the middle quality interval $x_1 = x_2 = 0.8$, which is illustrated by point e in Figure 1.

How does the optimal assignment extend beyond the maximum interval which permits complete mixing? In the limit of completely mixed sets, firms with the lowest project values employ only the lowest skill worker, as in point e of Figure 1. Extrapolating this logic, we pair firms with project values below $I(\underline{p})$ pairwise countermonotonically to pairs of identical low skill workers (workers with identical probabilities of not knowing how to solve a problem exceeding $I(\bar{p})$). This generates a pairwise countermonotic set for low value firms $z \in [0, I(\underline{p}))$ with self-sorted low skill workers $x_1 = x_2 \in (I(\bar{p}), 1]$, which we display by the dashed line from point e to point f in Figure 1. For example, a firm with project value $z = 0.05$ works with self-sorted low skill workers $x_1 = x_2 = 0.90$. By symmetry, high skill workers are paired with low skill coworkers and a high value project, that is, $x_1 \in [0, I(\underline{p}))$ pairs with $x_2 = z \in (I(\bar{p}), 1]$ and similarly coworker $x_2 \in [0, I(\underline{p}))$ pairs with $x_1 = z \in (I(\bar{p}), 1]$. We pair very high skill workers with low skill coworkers on valuable firm projects. Since both low value firms and high skill workers work with low skill coworkers, percentiles $[0, p)$ of firms z and high-skill workers x_1 absorb measure $2p$ of low-skill coworkers. Percentile $p \in [0, \underline{p})$ of the firm distribution is thus sorted with percentile $1 - 2p$ of the coworker distribution. A firm at percentile rank $\underline{p}$ is thus sorted with identical worker and coworker ranked $1 - 2\underline{p} = \bar{p}$, where the equality follows since all workers and coworkers outside the completely mixed set are to be assigned. This triplet incurs loss $I^2(\bar{p})I(\underline{p}) = \mathcal{C}$, identical to the loss for all triplets on the completely mixed set.

We finally verify that our construction is indeed product countermonotonic. To see this, recall that product countermonotonicity is equivalent to increasing marginal products following
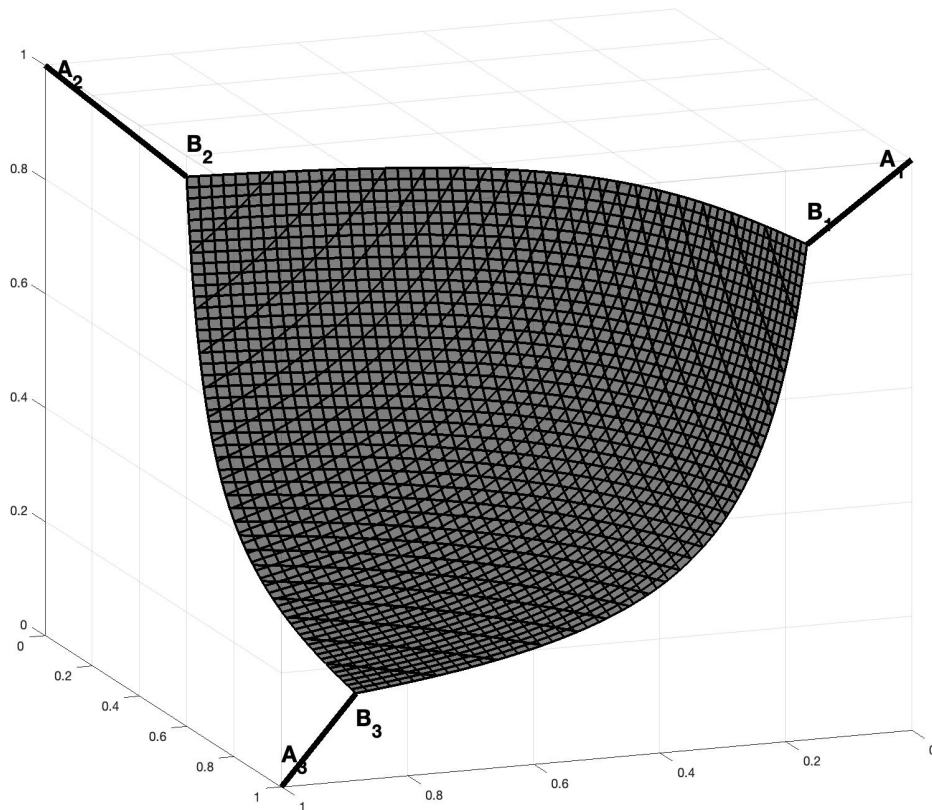
21

Figure 2: Support of the Assignment Function

Figure 2 plots the support for an optimal assignment, or set of potential matches M, in terms of percentile coordinates. The x-axis and y-axis show the percentile of the worker p_{x_1} and the percentile of their coworker p_{x_2} in the worker skill distribution, while the z-axis plots the percentile of the firm in the firm distribution p_z. The formal description of the set of potential matches is given by (26).

(4) and (21). Since top percentile $p \in [0, \underline{p})$ from the worker distribution is countermonotonically paired with identically ranked low-skill coworkers and high-value projects, the marginal worker product indeed increases for high-skill workers. For medium-skill workers, $p \in [\underline{p}, \bar{p}]$, this follows as $x_1 x_2 z = \mathcal{C}$ implies that $x_2 z$ decreases with x_1. To prove that the marginal worker product also increases for low-skill workers under our construction, we use part (b) of Assumption 1 to prove Proposition 5 in Section 4.

In sum, the analysis of the discretized planning problem provides a candidate solution which follows a simple description. Pair very high skill workers with low skill workers on valuable firm projects, until, at some point, mediocre workers can be paired together with firms such that expected output losses are identical for all these triplets. This intuition informs the conjecture for the solution to planning problem (9) that we show is optimal.

Conjecture. The analysis of the discretized planning problem informs a conjecture for the support of an optimal assignment. Figure 2 summarizes the candidate optimal assignment. The candidate

support consists of three symmetric pairwise countermonotonic sets (A_1B_1, A_2B_2, A_3B_3), and a completely mixed set $(B_1B_2B_3)$. Workers who always know how to solve a problem, $x_1 = 0$, have coworkers that never know how to solve a problem, $x_2 = 1$, and tackle the most valuable projects. The team always knows how to solve a problem, so that production succeeds. Similarly, the second most skilled workers have coworkers that rarely know how to address a problem. The team still almost always knows how to solve a problem and are hence allocated to only slightly less valuable projects. This logic leads to pairwise countermonotonic sets, such as the dashed line in Figure 1. Low values for x_1, or skilled workers, are paired with high values for x_2 and z, low skill coworkers and valuable projects. By symmetry, we have three pairwise countermonotonic sets, or segments in Figure 2. Since both low value firms and high skill workers work with low skill coworkers, percentiles $[0, \underline{p})$ for firms and high-skill workers jointly absorb measure $2\underline{p}$ of low-skill coworkers, implying $\bar{p} = 1 - 2\underline{p}$. For a worker ranked $\underline{p}$, the coworker and the project value are ranked $\bar{p}$, and the output loss is equal to $I^2(\bar{p})I(\underline{p})$.

Firms and workers outside the pairwise countermonotonic sets are mediocre workers, coworkers and firms ranked between percentiles $\underline{p}$ and $\bar{p}$ in their respective distributions. Mediocre workers and firms are sorted into completely mixed triplets on the set $B_1B_2B_3$. Triplets on this set have identical output losses equal to $\mathcal{C} = I^2(\bar{p})I(\underline{p})$, which thus connect with the output loss on the pairwise countermonotonic sets at $(\underline{p}, \bar{p}, \bar{p})$.

We can now specify the support of an optimal assignment, the set of possible matches M. As discussed, the set of possible matches combines three symmetric pairwise countermonotonic sets and a mixed middle:

$$M := M_{x_1} \cup M_{x_2} \cup M_z \cup M_{\mathcal{C}}, \tag{26}$$

where:

$$M_{x_1} := \left\{ \left(p_{x_1}, 1 - 2p_{x_1}, 1 - 2p_{x_1}\right) \mid p_{x_1} \in [0, \underline{p}) \right\}$$
$$M_{x_2} := \left\{ \left(1 - 2p_{x_2}, p_{x_2}, 1 - 2p_{x_2}\right) \mid p_{x_2} \in [0, \underline{p}) \right\}$$
$$M_z \ := \left\{ \left(1 - 2p_z, 1 - 2p_z, p_z\right) \mid p_z \in [0, \underline{p}) \right\}$$
$$M_{\mathcal{C}} := \left\{ \left(p_{x_1}, p_{x_2}, p_z\right) \mid \underline{p} \le p_{x_1}, p_{x_2}, p_z \le \bar{p} \ \text{ and } \ I(p_{x_1})I(p_{x_2})I(p_z) = \mathcal{C} \right\}$$

The pairwise countermonotonic set which corresponds to pairings for high skill workers x_1 is M_{x_1} (A_1B_1), the pairwise countermonotonic set which corresponds to pairings for high skill coworkers

x_2 is M_{x_2} $(A_2 B_2)$, while the pairwise countermonotonic set which corresponds to pairings for low value firms z is M_z $(A_3 B_3)$. Finally, $M_{\mathcal{C}}$ is the completely mixed set $(B_1 B_2 B_3)$, the set of matches with constant expected output losses.

One of the main results of this paper, Theorem 3 in Section 4, shows that any feasible assignment with its support being a subset of the set of potential matches M is optimal.

3.2 Dual

To develop intuition for the equilibrium wages and firm values, we study the discrete approximation to the dual problem. Given worker samples $\{x_{is}\}_{s=1}^{n}$ for $i \in \{1, 2\}$, and a firm type sample $\{z_s\}_{s=1}^{n}$, the discrete dual problem is to choose w_1, w_2 and v to solve:

$$\min \ \sum w_1(x_1) + \sum w_2(x_2) + \sum v(z), \tag{27}$$

subject to the constraint that the surplus $S(x_1, x_2, z)$ is negative for any triplet (x_1, x_2, z).

We establish two equilibrium properties. First, we show that the multipliers on the constraints for the planning problem are equilibrium wages and firm values. Second, we show that the derivative of the wage schedule is the marginal worker product (4). The output increase obtained by replacing a worker with a slightly more skilled worker, keeping the coworker and project value unchanged, has to equal the increase in wages necessary to hire the high-skill worker.

To establish that the multipliers on the constraints for the planning problem are equilibrium wages and values, we first note that the dual value is bounded below. By the constraint to the dual problem, the surplus is negative for any triplet (x_1, x_2, z), $S(x_1, x_2, z) \leq 0$. The negative surplus is equivalent to $w(x_1) + w(x_2) + v(z) \geq y(x_1, x_2, z)$ and therefore the total surplus is bounded below $\sum w_1(x_1) + \sum w_2(x_2) + \sum v(z) \geq \sum \pi_{ijk} y(x_1, x_2, z)$, for some π that satisfies constraints (29). In particular, this inequality holds for any assignment π^* that solves the planning problem. To establish that the multipliers are a solution to the dual problem, we show that the multipliers attain this lower bound.

The discrete analog to the planning problem is to choose an assignment π that solves:

$$\max_{\pi \geq 0} \ \sum \pi_{ijk} y_{ijk}, \tag{28}$$

subject to constraints:

$$\sum_{jk} \pi_{ijk} = 1, \qquad \sum_{ik} \pi_{ijk} = 1, \qquad \sum_{ij} \pi_{ijk} = 1, \tag{29}$$

24

where i indexes sampled workers, j indexes sampled coworkers, and k indexes sampled projects. The constraints require that every worker and coworker are employed and that every firm employs two workers. The constraint does allow workers to work at multiple firms with different coworkers, that is, $\pi_{ijk} > 0$ and $\pi_{i\hat{j}\hat{k}} > 0$ for $(j, k) \neq (\hat{j}, \hat{k})$. The multipliers on the non-negativity constraints are u_{ijk}, while w_{1i}, w_{2j} and v_k are multipliers for the constraints (29). By the Karush-Kuhn-Tucker conditions, $y_{ijk} - w_{1i} - w_{2j} - v_k + u_{ijk} = 0$ and either $\pi_{ijk}^* = 0$ or $u_{ijk} = 0$. When $\pi_{ijk} > 0$ the surplus is equal to zero, $S_{ijk} = 0$. When $\pi_{ijk}^* = 0$ the surplus is negative $S_{ijk} \leq 0$, showing that the planning multipliers are a feasible dual solution. Summing over triplets $\sum \pi_{ijk}^* y(x_1, x_2, z) = \sum w_1(x_1) + \sum w_2(x_2) + \sum v(z)$. This implies that the multipliers on constraints to the planning problem (29) indeed solve the dual problem.

Finally, the derivative of the wage schedule is the marginal worker product. Since the surplus is negative $S(x_1, x_2, z) \leq 0$ and equal to zero when $\pi_{ijk}^* > 0$, any equilibrium triplet attains the maximum with respect to the surplus function. When the surplus function is differentiable this gives marginal earnings as $w_1'(x_1) = y_1(x_1, x_2, z) = -x_2 z = m(x_1) \leq 0$, where the final equality follows from the definition of the marginal worker product (4). For the continuous distribution, we prove this property in Proposition 6.

4   Main Results

We characterize equilibrium with continuous marginal distributions in closed form building on the discretized analysis of the previous section. To simplify exposition, we assume the distributions for worker and firm types are identical, and make the following assumption for this distribution.

Assumption 1. The distribution of worker and firm types F satisfies:

(a) $x f(x)$ is nondecreasing on the middle interval $[\underline{p}, \bar{p}]$.

(b)   Logarithmic loss $L(p) := 2 \log I(1 - 2p) + \log I(p)$ is concave on the interval $p \in [0, \frac{1}{3}]$.

We use part (a) when we show that there exists an assignment π concentrated on a subset of the set of potential matches M, and part (b) when we show optimality of any assignment function π that is concentrated on a subset of M.

A more restrictive but simpler condition is Assumption 2.

Assumption 2. For all x, $x f(x)$ is nondecreasing.

When the probability density function is differentiable, Assumption 2 is equivalent to bounding the elasticity of the density function from below by -1. Distributions with nondecreasing density functions trivially satisfy this restriction. Assumption 2 is satisfied by Beta distributions as well as truncated exponential, log-normal, and Pareto distributions for a subset of parameters, amongst others. Naturally, Assumption 2 also holds for linear combinations of density functions that satisfy it. The assumption implies that the logarithm of the inverse distribution function is concave.[17] Assumption 2 is indeed more restrictive as it implies that logarithmic loss $L(p)$ is concave on the interval $p \in [0, \frac{1}{3}]$.[18]

4.1 Planning Problem

In the discretized analysis, we asserted existence of an assignment concentrated on a subset of the set of potential matches M. We start our analysis of the continuous planning problem showing there indeed exists such an assignment.

Proposition 2. *Existence.* There exists an assignment π concentrated on a subset of the potential matches M.

The proof is in Appendix A.3. The existence of a feasible assignment concentrated on the pairwise countermonotonic sets M_{x_1}, M_{x_2}, M_z follows immediately. To prove the existence of an assignment concentrated on the completely mixed set $M_{\mathcal{C}}$ we connect our assignment framework to Wang and Wang (2016). Wang and Wang (2016) study aggregate risk, the sum of idiosyncratic risks with known marginal distributions, but with an unknown dependence structure between idiosyncratic risks. For idiosyncratic risks with continuous and decreasing density function $\hat{g}$, Wang and Wang (2016) describe when there exists a dependence structure, a measure, such that the aggregate risk is constant, $x_1 + x_2 + z = \mathcal{C}_e$.[19]

To prove existence of an assignment, we let $\hat{x}_1 = \exp(-x_1)$, $\hat{x}_2 = \exp(-x_2)$ and $\hat{z} = \exp(-z)$, and use that our distribution for workers and firms is identical. Under these conditions, constant

---

[17]This observation follows by the inverse function theorem because $\frac{\mathrm{d}}{\mathrm{d}p} \log I(p) = \frac{\mathrm{d}}{\mathrm{d}p} I(p)/I(p) = 1/\big(I(p)f(I(p))\big)$. Therefore, $xf(x)$ is increasing if and only if $\log I(p)$ is concave.

[18]The logarithmic loss $L(p)$ is concave as it sums two concave functions. By Assumption 2, see also footnote 17, the second term, the logarithm of the inverse distribution function $I(p)$ is concave. The first term is also concave, as its derivative, $\frac{\mathrm{d}}{\mathrm{d}p} \log I^2(1 - 2p) = -\frac{4}{I(1-2p)f(I(1-2p))} < 0$, increases with $I(1 - 2p)$ and hence decreases with p.

[19]An early treatment on this problem, whether risks can attain a constant sum, is found in Gaffke and Rüschendorf (1981).

losses, $\hat{x}_1 \hat{x}_2 \hat{z} = \mathcal{C}$, is equivalent to $x_1 + x_2 + z = -\log\mathcal{C}$. Since the distribution functions for x and $\hat{x}$ are linked by $\hat{G}(x) = 1 - \hat{F}(\exp(-x))$, where $\hat{F}$ is the type distribution restricted to mediocre workers and firms $[I(\underline{p}), I(\bar{p})]$, the density functions are linked through $\hat{g}(x) = \exp(-x)\hat{f}(\exp(-x))$. Assumption 1 implies $x\hat{f}(x)$ is increasing, showing the density function $\hat{g}$ is indeed decreasing. In the remainder of the proof we show the existence condition in Wang and Wang (2016) holds when $I^2(\bar{p})I(\underline{p}) \geq \exp\left(\frac{3}{\bar{p}-\underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p)\mathrm{d}p\right)$, implying that there indeed exists an assignment concentrated on our set of potential matches M.

Optimality. To prove the optimality of a candidate assignment, let γ be the distribution of losses implied by a feasible assignment. To illustrate, the assignment $\{(0.4, 0.1, 0.1), (0.1, 0.4, 0.4)\}$ has an implied loss distribution $\{0.004, 0.016\}$ each with mass equal to $\frac{1}{2}$. The implied loss distribution has an associated inverse cumulative distribution function I_γ, with losses by a triplet at percentile p in the loss ranking given by $I_\gamma(p)$ when the triplets are sorted in ascending loss order. In the illustration, the implied loss at percentile $p = \frac{1}{4}$ in the loss ranking is 0.004. Since it is equivalent to integrate output losses with respect to the assignment function and with respect to ranks of the loss distribution, $\int \ell(x_1, x_2, z)\mathrm{d}\pi = \int_0^1 I_\gamma(p)\mathrm{d}p = \int_0^1 \exp(\log I_\gamma(p))\mathrm{d}p$. To show that an assignment π^* is optimal, we show aggregate losses are greater under any other feasible assignment π:

$$\int \ell(x_1, x_2, z)\mathrm{d}\pi^* \leq \int \ell(x_1, x_2, z)\mathrm{d}\pi \iff \int_0^1 \exp(\log I_{\gamma^*}(p))\mathrm{d}p \leq \int_0^1 \exp(\log I_\gamma(p))\mathrm{d}p. \quad (30)$$

The proof of optimality is constructed around the integral version of Karamata's inequality (see Karamata (1932) and Pečarić (1984)). To establish optimality, we note that the exponential function is convex and that logarithmic losses, $\log I_\gamma$, are increasing as the logarithm of a positive increasing function is itself increasing. As a result, we can use the integral version of Karamata's inequality with $h = \exp$, $s_1(p) = \log I_{\gamma^*}(p)$ and $s_2(p) = \log I_\gamma(p)$.

Karamata's Inequality. Let h be a continuous convex function, and s_1 and s_2 be non-decreasing functions on the unit interval. Assume that

$$1. \int_0^1 s_1(p)\mathrm{d}p = \int_0^1 s_2(p)\mathrm{d}p \qquad (31)$$

$$2. \int_0^t s_1(p)\mathrm{d}p \geq \int_0^t s_2(p)\mathrm{d}p, \qquad (32)$$

for all $t \in [0, 1]$. Then $\int_0^1 h(s_1(p))\mathrm{d}p \leq \int_0^1 h(s_2(p))\mathrm{d}p$.

27

Invoking Karamata's inequality, it suffices to verify that the two assumptions are satisfied. The first condition requires that aggregate logarithmic losses under the contrasted assignments align. As aggregate logarithmic losses are invariant across feasible assignments this condition is verified. That is, for any feasible assignment π:

$$\int_0^1 \log I_\gamma(p)\mathrm{d}p = \int \log \ell(x_1, x_2, z)\mathrm{d}\pi = \int \log x_1 \mathrm{d}F + \int \log x_2 \mathrm{d}F + \int \log z \mathrm{d}F. \tag{33}$$

The second condition is that cumulative logarithmic losses at any percentile in the triplet rank t are larger under optimal assignment π^*. We observe that cumulative logarithmic losses are convex in triplet rank t as they integrate the (negative) increasing function $\log I_\gamma$. For the completely mixed set, or triplets ranked $t \in [3\underline{p}, 1]$, cumulative logarithmic losses decrease linearly as losses are constant. By the fact that cumulative logarithmic output losses for the $t \in [0, 3\underline{p})$ lowest ranked triplets are maximized by three pairwise countermonotonic sets (by Lemma 2 and Lemma 3 in Appendix A.4), and since aggregate logarithmic losses are equal, logarithmic cumulative output losses are larger at any triplet rank $t \in [0, 1]$. We formalize the verification of the second condition, the final step of the proof to Theorem 3.

Theorem 3. *Support of the Assignment.* Any assignment function $\pi \in \Pi$ which is concentrated on a subset of the potential matches M solves the planning problem (9).

Theorem 3 does not imply uniqueness of an optimal assignment, only that any assignment function on a subset of the set of potential matches solves the planning problem. We next analyze the dual problem to characterize equilibrium wages and firm values, that are both uniquely determined.

4.2   Dual Problem

To characterize the solution to the dual problem, we first evaluate the worker's marginal product, or peer effects, under an optimal assignment.

Marginal Worker Product. The marginal worker product (4) equals the marginal loss induced by a worker, which is the negative of the value of the firm project and the coworker skills. Given that an optimal assignment is concentrated on a subset of potential matches M, which describes for every worker pairings with firm types and coworker's skills, the equilibrium marginal worker product can now be fully described.

High-skill workers are pairwise countermonotonically sorted. High-skill workers (low x_1) work with low-skill coworkers on valuable projects (high x_2, z). For a high-skill worker at a top percentile $p \in [0, \underline{p}]$, the pairwise countermonotonic triplet is $(p, n(p), n(p))$, where $n(p)$ denotes the negatively sorted percentile corresponding to worker p's coworker and firm in their respective distributions. By symmetry, high-skill coworkers negatively sort with low-skill workers on high value projects $(n(p), p, n(p))$, while low-value projects sort with low-skill workers and coworkers $(n(p), n(p), p)$. Since both high-skill coworkers and low-value firms employ low-skill workers, at top percentile p measure $2p$ of low-skill workers are employed. Since the mass of workers with skills below $n(p)$ is $1 - n(p)$, we obtain $n(p) = 1 - 2p$. The pairwise countermonotonic triplet for a high-skill worker at top percentile p is thus $(p, 1 - 2p, 1 - 2p)$. The product of the project value and the coworker skills is thus $I^2(1 - 2p)$.

Workers and project values in the middle of their distribution, at percentiles $p \in (\underline{p}, \bar{p})$, work in completely mixed triplets with identical expected output losses across all such mediocre triplets $\ell(x_1, x_2, z) = x_1 x_2 z = \mathcal{C}$. The marginal worker product is thus equal to $-\frac{\mathcal{C}}{x_1}$. Low-skill workers are, as discussed, paired with an equally low-skilled coworker to work a low value project. Given some low-skill worker percentile $p \in [\bar{p}, 1]$, their triplet is $\left(p, p, \frac{1-p}{2}\right)$, which is another representation of the low value project $p_z = \frac{1-p}{2} \in [0, \underline{p})$ employing two low skill workers $p = 1 - 2p_z$, or $(1 - 2p_z, 1 - 2p_z, p_z)$ as for M_z in the matching set (26). The marginal product at low-skill worker percentile p is thus equal to $-I(p)I\left(\frac{1-p}{2}\right)$. This description of the marginal product is summarized by Proposition 4.

Proposition 4. *Marginal Product.* The marginal product of worker x is:

$$
m(x) = \begin{cases} -I^2(1 - 2p) & \text{if } p \in [0, \underline{p}] \\ -\mathcal{C}/I(p) & \text{if } p \in (\underline{p}, \bar{p}) \\ -I(p)I\left(\frac{1-p}{2}\right) & \text{if } p \in [\bar{p}, 1] \end{cases} \tag{34}
$$

where constant $\mathcal{C} = \exp\left(\frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p\right)$, with lower percentile bound $\underline{p} \in \left(0, \frac{1}{3}\right]$ satisfying:

$$
\mathcal{C} = I^2(\bar{p}) I(\underline{p}), \tag{35}
$$

and where the upper percentile bound is given by $\bar{p} = 1 - 2\underline{p}$.

The proof is in Appendix A.5. To understand (35), we use our observation from Section 3, which we deduced from stability condition (21), that the marginal worker product increases with worker skill x. By evaluating the marginal product (34) at the lower percentile cutoff $\underline{p}$, and multiplying with worker skill $I(\underline{p})$, we obtain that the output losses at the lower percentile cutoff exceed the output losses on the completely mixed set, that is, $I^2(\bar{p})I(\underline{p}) \geq \mathcal{C}$. Similarly, we assess the marginal worker product (34) at the upper percentile cutoff $\bar{p}$, and multiply by worker skill $I(\bar{p})$ to derive that the output losses on the completely mixed set exceed the output losses at the upper percentile cutoff $I^2(\bar{p})I(\underline{p}) \leq \mathcal{C}$, which implies $I^2(\bar{p})I(\underline{p}) = \mathcal{C}$. As we construct the marginal worker product using the support of an equilibrium assignment (26), the constant is equal to (24) which we conjectured for the discretized analysis. We formalize the observation that the marginal worker product m is increasing, and establish continuity, in Proposition 5.

Proposition 5. *Properties Marginal Product.* The marginal product m is continuous and strictly increasing.

The proof is in Appendix A.6. High-skill workers are paired with a better combination of coworker and firm. There is countermonotonic sorting between worker errors and expected losses incurred by the firm and coworker absent the worker's input. The logic follows from the stability condition. Using the first condition in (21), together with the definition of the marginal worker product, we have $(m(x_1) - m(\hat{x}_1))(x_1 - \hat{x}_1) \geq 0$. When a worker is more skilled, say $\hat{x}_1 < x_1$, they work with a coworker and firm that has a greater expected output loss all else equal, and equivalently a more negative marginal product, $m(\hat{x}_1) \leq m(x_1)$.

Wages. Given the marginal product we formulate equilibrium wages. We first use a generalization of Kantorovich duality for multiple marginal distributions, due to Kellerer (1984). The duality result states that the dual solution exists in the class of integrable functions. We establish that a solution exists in the class of continuous functions in Appendix A.7. This implies that the surplus function (14) is continuous also.

The duality result further states that the optimal value to the planning problem is equal to the optimal value for the dual problem, $\int y(x_1, x_2, z)\mathrm{d}\pi = \int w(x_1)\mathrm{d}F_x + \int w(x_2)\mathrm{d}F_x + \int v(z)\mathrm{d}F_z$, where π solves the planning problem and w and v solve the dual. This implies the surplus is zero for every triplet (x_1, x_2, z) in the support of an optimal assignment. The surplus is non-positive

for every triplet by the dual constraint, hence given continuity $\int S(x_1, x_2, z) \mathrm{d}\pi = 0$ if and only if the surplus is zero for all triplets in the support of the assignment, or $S(x_1, x_2, z) = 0$ for all triplets $(x_1, x_2, z) \in \mathrm{supp}(\pi)$.

We use that there are no excess resources in equilibrium to characterize the derivative for the wage schedule and the firm value function.

Proposition 6. *Marginal Wages and Firm Values.* Wages and firm values are both continuously differentiable, with derivatives $w'(x) = m(x)$ and $v'(z) = m(z) + 1$.

The proof is in Appendix A.8. As discussed for the discretized dual problem, workers' marginal earnings are equal to their marginal product. The increase in expected output which is obtained by replacing a worker with a marginally higher skill worker leaving the coworker and firm value unchanged must equal the increase in the cost necessary to recruit the higher skill worker. While the technological marginal product, $-x_2 z$, depends only on peer effects, the equilibrium level of these peer effects can be directly related to the worker's own skills through Proposition 4.

While a worker's marginal earnings reflect their marginal product, earnings levels reflect the marginal product of all other workers in the economy that are more skilled. Equivalently, worker earnings depend on marginal losses incurred by the product of coworker skills and project values for workers that are more skilled. Specifically, workers incur an earnings penalty relative to the most skilled worker. Equilibrium wages are characterized in Proposition 7.[20]

Proposition 7. *Wages.* The wage schedule is, up to additive worker constant $\mathcal{C}_w$, given by:

$$w(x) = \mathcal{C}_w + \int_0^x m(s) \mathrm{d}s. \tag{36}$$

To obtain intuition, observe the surplus is zero almost everywhere with respect to π, $S(x_1, x_2, z) = z(1 - x_1 x_2) - w(x_1) - w(x_2) - v(z) = 0$. Further, since w is a differentiable function, fix a coworker x_2 and firm z. Since the surplus function is negative $S(x_1, x_2, z) \leq 0$, equilibrium triplet (x_1, x_2, z) is a local maximum with respect to the surplus function, and hence:

$$S_1(x_1, x_2, z) = 0 = -x_2 z - w'(x_1) \qquad \Longrightarrow \qquad w'(x_1) = -x_2 z = m(x_1), \tag{37}$$

---

[20]Alternatively, we could project our primal solution onto the space of worker pairings $X \times X$ to write a two worker-to-one firm assignment problem. In case the implied distribution of worker pairings has a density function, Chiappori, McCann, and Pass (2017) provides an alternative approach to establish properties of the dual solution. However, under our optimal assignment the distribution of worker pairings does not have a density function because of the pairwise countermonotonic sets.

where the final equality follows from the specification of the marginal worker product (4). Since there are no excess resources and the surplus is differentiable, every worker's marginal earnings is their marginal product. By integrating, equilibrium wages follow (36).

Equilibrium wages are decreasing and convex in the skill ranking. The marginal cost of workers is equal to their marginal product $w'(x) = m(x)$ for all $x \in (0, 1)$. By Proposition 5, this marginal worker product is negative, continuous, and strictly increasing, thereby establishing our claim.

Corollary. Wages are decreasing and convex in skills x.

Firm. Similar to wages on the worker side of the labor market, we characterize firm values closed-form. We obtain formal intuition for the equilibrium firm value similar to our intuition for wages. If there are no excess resources almost everywhere for assignment π, then, by fixing some pair of workers x_1 and x_2 as well as by differentiability of the dual function v:

$$S_3(x_1, x_2, z) = 0 = (1 - x_1 x_2) - v'(z) \qquad \Longrightarrow \qquad v'(z) = 1 - x_1 x_2 = m(z) + 1, \tag{38}$$

where the final equality again follows from the definition of the marginal product (4). A firm its marginal reward is its marginal product. When a firm tackles a marginally more valuable project, its output is its expected probability of success, or team quality, $1 - x_1 x_2$. In equilibrium, the firm's marginal product is similar to the worker product, up to the unit constant, due to the symmetry of expected loss function ℓ and the identical worker and project value distributions. By integrating, firm values follow Proposition 8.

Proposition 8. *Firm Value.* The firm value v is, up to additive firm constant $\mathcal{C}_v$, given by:

$$v(z) = \mathcal{C}_v + \int_0^z k(s) \mathrm{d}s, \tag{39}$$

where $k(s) := m(s) + 1$ is the marginal firm product.

By Proposition 8, the firm value shares properties with the wage function. By Proposition 5, the marginal firm product k is continuous and strictly increasing. The marginal worker product m is continuous and strictly increasing, and $m(z) \geq -1$ implying $v'(z) \geq 0$. The firm value function is thus increasing and convex.

Corollary. Firm values are increasing and convex in project value z.

Given the description of wages in Proposition 7 and the firm value function in Proposition 8, we observe that for any constants $\mathcal{C}_w$ and $\mathcal{C}_v$ that satisfy:

$$0 = \mathcal{C}_w + \mathcal{C}_v/2 + \int_0^1 m(s)\mathrm{d}s \tag{40}$$

the functions (w, v) are a dual solution. Equation (40) ensures the surplus is zero for the triplet $(p_{x_1}, p_{x_2}, p_z) = (1, 1, 0)$. Since the surplus is equal to zero everywhere in the support of an optimal assignment, this condition ensures that the value for the planning problem and the dual problem coincide. By Proposition 1, any assignment π with the support contained in the set of potential matches M, together with dual functions (w, v) satisfying (40) is thus an equilibrium.[21]

5 Quantitative Analysis

We now turn to the quantitative analysis of the model calibrated to administrative records for the US. We evaluate the ability of the model to generate the observed wage dispersion in the United States, as well as its decomposition into within and between-firm components. We compare the predictions of our model against data moments by Song, Price, Guvenen, Bloom, and Von Wachter (2019).

5.1 Data

Our main data source are administrative earnings records for every individual from the US Social Security Administration (SSA). We use data moments for 1981 and 2013 provided by Song, Price, Guvenen, Bloom, and Von Wachter (2019). Individual earnings are derived from W-2 forms, which capture wages and salaries, bonuses, tips and other forms of income defined as compensation for labor services by the Internal Revenue Service. The data sample considers all employed individuals between 20 and 60 years of age.[22] All amounts are in 2013 dollars.

Our main reason for using the SSA records is that the data matches the universe of individuals

---

[21]Equation (40) determines the equilibrium in terms of wages and firm values up to a constant. When functions w and v solve the dual problem, so do functions $\hat{w}$ and $\hat{v}$ which differ from w and v only in terms of the constants as long as they satisfy (40). In our quantitative analysis, we set the worker constant to the maximum earnings observed in the U.S. earnings distribution in order for the model to match the top of the earnings distribution (36).

[22]An individual is considered employed when their earnings exceed the minimum wage earned full-time for one quarter of the year, that is, for 520 hours (40 hours per week times 13 weeks). The data exclude firms and workers in the government and education sector. While the baseline sample of Song, Price, Guvenen, Bloom, and Von Wachter (2019) considers firms with over 20 employees, we confirm our results extend to firms of all sizes.
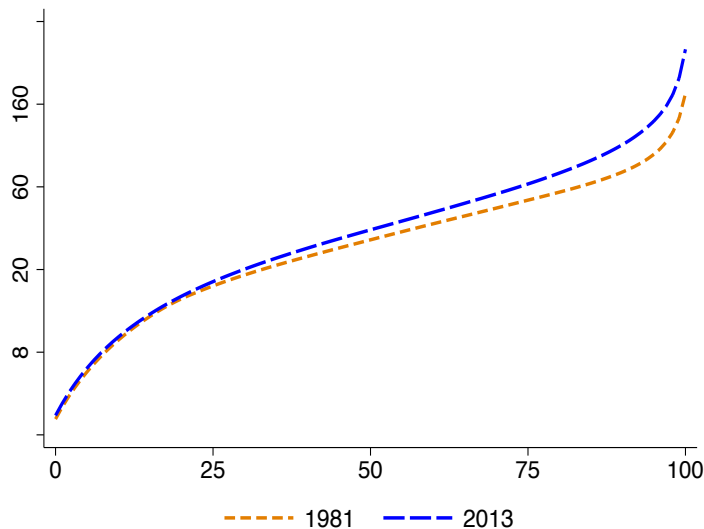
Figure 3: Distribution of Individual Earnings

Figure 3 displays the distribution of annual earnings in the SSA, replicating Figure I in Song, Price, Guvenen, Bloom, and Von Wachter (2019). The figure shows individual earnings levels, on the y-axis in thousands of dollars, for each percentile of the income distribution, on the x-axis. Between 1981 and 2013, earnings in the bottom third have seen little change, while earnings at the top of the distribution increased.

to their employer and thus speaks directly to the sorting of workers with coworkers and firms.[23] Workers with multiple jobs in a year are linked to the firm that provides their largest share of labor earnings. Most important for our purposes, the matched employer-employee dataset contains information on the earnings distribution within every firm.

Figure 3 shows individual earnings by percentile of the earnings distribution for 1981 and 2013. In 1981, earnings vary between 18 thousand dollars at the 25th percentile, to 31 thousand at the median, to 89 thousand at the 95th percentile. By 2013, earnings are 19 thousand dollars at the 25th percentile, 35 thousand at the median, and 132 thousand at the 95th percentile. Figure 3 shows that the earnings distribution features considerable dispersion in levels, which has increased over time. Between 1981 and 2013, earnings in the bottom third of the earnings distribution have seen little change, while earnings at the top strongly increased.

We assess the ability of our model to generate the observed dispersion in labor income, the decomposition of the variance between and within firms, as well as changes in the distribution of earnings over time. We then use the calibrated model to structurally decompose increased income

---

[23]A firm in the data corresponds to a corporate unit for tax purposes. A corporate tax unit is a more aggregated notion compared to an establishment. Using US Census data, Barth, Bryson, Davis, and Freeman (2016) split the variance of income between and across establishments. They similarly document an increase in the variance of income between 1992 and 2007, of which two-thirds is accounted for by the variation between establishments.

dispersion between changes on the worker side and the firm side of the labor market.

5.2 Inference

The model economy is parsimonious. Only the worker skill distribution F_x, the project value distribution F_z, and the constant $\mathcal{C}_w$ in the worker earnings equation (36) are exogenous.[24] Consistent with our presentation of the theory, we assume that the worker and firm type distributions are identical. With this assumption, we only require the cross-sectional distribution of worker earnings to inform the skill distribution F. Since backing out the type distribution does not require any information about coworkers, we can then use the information about coworkers to evaluate non-targeted moments.

We back out the underlying type distribution using the level of earnings in the cross-section by percentile. We denote the empirical earnings distribution by percentile p as $e(p)$, where empirical earnings are arranged in descending order. We directly inform the empirical earnings distribution using the data underlying Figure 3. For example, $e(0.05)$ in 1981 is 89 thousand dollars and $e(0.5)$ in 2013 is 35 thousand dollars.

The closed-form equilibrium characterization is leveraged to back out the type distribution as well as the constant in the earnings equation using the empirical earnings distribution by percentile. Specifically, we back out the type distribution point-by-point and obtain the constant in the earnings equation by exactly matching the empirical earnings distribution. Intuitively, the constant $\mathcal{C}_w$ in the earnings equation (36) is pinned down by earnings at the top of the distribution. The difficulty in backing out the skill distribution lies in exactly matching observed earnings at all lower percentiles because earnings by workers at lower percentiles depend on the marginal worker product of all individuals more skilled than themselves. In turn, the marginal worker product of every individual depends on their coworkers and their firm. Our solution to this inference problem is given by Proposition 9.

Proposition 9. *Inferring the Skill Distribution.* If there exists an inverse distribution function I that satisfies:

---

[24]Given a constant $\mathcal{C}_w$ in the earnings schedule (36), the characterization of the marginal product in Proposition 4, and the worker and firm value distributions, the equilibrium cross-equation restriction (40) determines the constant on the firm side $\mathcal{C}_v$ in the firm value equation (39).

$$
\log I(p) = \begin{cases} - \int_0^p \frac{e'(t)}{\omega(t)} \mathrm{d}t & \text{if } p \in [0, \underline{p}] \\[2ex] - \int_0^p \frac{e'(t)}{\omega(t)} \mathrm{d}t - \frac{e(p) - e(\underline{p})}{I(\underline{p})I(\bar{p})^2} & \text{if } p \in (\underline{p}, \bar{p}) \\[2ex] -2 \int_0^p \frac{e'(1-2t)}{\omega(t)} \mathrm{d}t & \text{if } p \in [\bar{p}, 1] \end{cases} \tag{41}
$$

where $\omega(p) = e(t)\big|_0^p - 2e(1-2t)\big|_0^p$ and $\underline{p} = \frac{1-\bar{p}}{2}$ solves:

$$
\log I(\underline{p}) I(\bar{p})^2 = \frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p
$$

Then the observed earnings distribution is the equilibrium outcome for an economy with constant $\mathcal{C}_w = e(0)$ and type distribution $F = I^{-1}$.

The proof to Proposition 9 consists of two parts, and is presented in Appendix A.9. We first show that if the empirical earnings function is a decreasing and convex function in percentile rank p, then Assumption 1 is satisfied. This implies that a decreasing and convex earnings schedule can be rationalized by our model, and provides an empirical criterion to determine if Assumption 1 is restrictive.

Second, Proposition 9 pins down the inverse cumulative distribution function. The distribution function is directly identified from the cross-sectional earnings distribution, as shown by equation (41). We take the empirical distribution function e, and feed this in for both the bottom and the top percentiles of the distribution, $[0, \underline{p}]$ and $[\bar{p}, 1]$, to obtain $I(\underline{p})$ and $I(\bar{p})$. We use these values to back out the skill distribution for the medium percentiles. We back out the distribution function for 1981 to exactly match the earnings distribution in 1981 and back out another distribution function to exactly match the observed earnings distribution for 2013. The resulting model fit for cross-sectional earnings is summarized by Figure 4 for the distribution excluding the top 1 percent and in Figure 5 for the top 1 percent. Taken together, Figure 4 and Figure 5 show that our framework generates the observed dispersion in earnings, as well as changes in the distribution of earnings over time.

5.3 Earnings Dispersion Within and Between Firms

We next decompose the overall cross-sectional variance of log earnings into within- and between-firm components. Let w_{ij} be log earnings of worker i employed by firm j. Earnings in any year
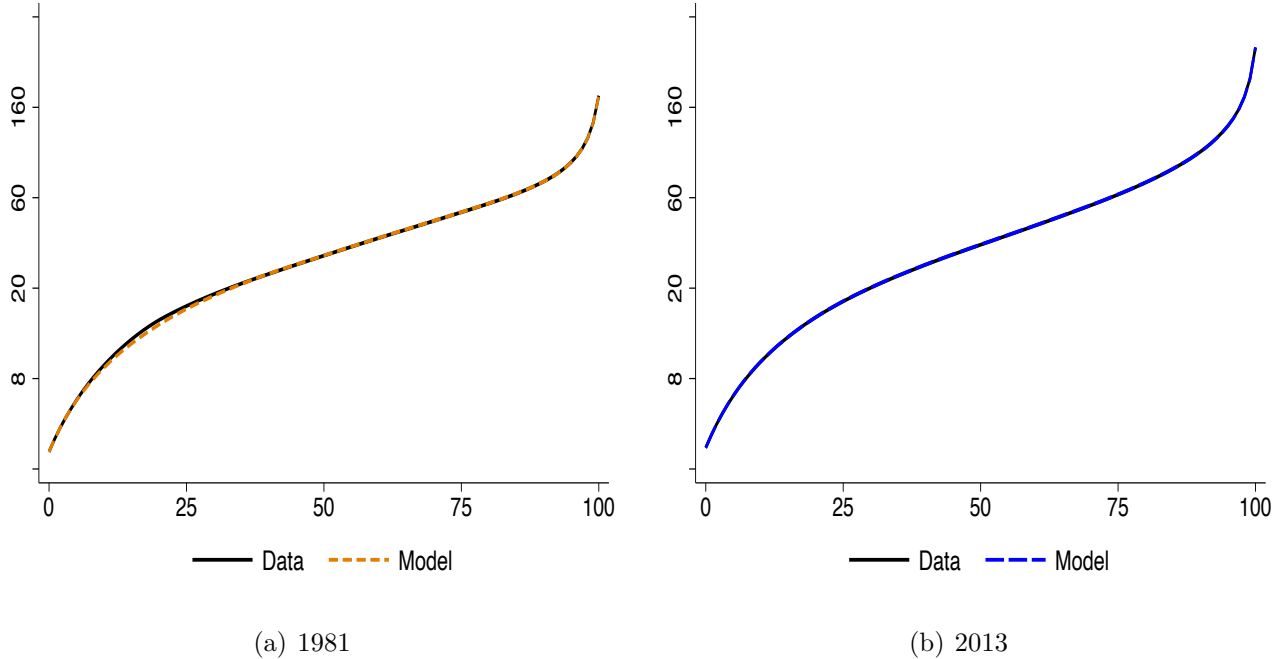
(a) 1981　　　　　　　　　　　　　　　　(b) 2013

Figure 4: Data and Model Distribution of Individual Earnings

Figure 4 compares the empirical earnings distribution to the model earnings distribution. The empirical distributions, which we display by solid black lines, follow Figure 3, while the model distributions are presented by dashed lines. The left panel shows the empirical and model distribution for 1981, the right panel for 2013.

can be decomposed as:

$$w_{ij} := \overline{w}_j + \left(w_{ij} - \overline{w}_j \right), \tag{42}$$

where $\bar{w}_j$ is average log earnings within firm j. The total cross-sectional variance is decomposed into within- and between firm components:

$$\mathrm{Var}(w_{ij}) = \mathrm{Var}(\overline{w}_j) + \sum_j \theta_j \mathrm{Var}(w_{ij}|i \in j), \tag{43}$$

where θ_j is the cross-sectional employment share of firm j.[25] The first term is the variance of mean earnings across firms, or the between-firm variance. The second term is the employment weighted average of within-firm dispersion of employee earnings.

Table 1 displays the decomposition of the variance of log earnings for 1981 and 2013. The left panel shows that the total empirical variance of log earnings in 1981, 0.65 log points, is for one-third attributed to differences in mean earnings between firms and for two-thirds to within-firm

---

[25]For a derivation of the variance decomposition formula (43) see, for example, Davis and Haltiwanger (1991) and Alvarez, Benguria, Engbom, and Moser (2018). The decomposition is employment-weighted and thus invariant to splitting a single firm into two replicas with half the firm's original size.
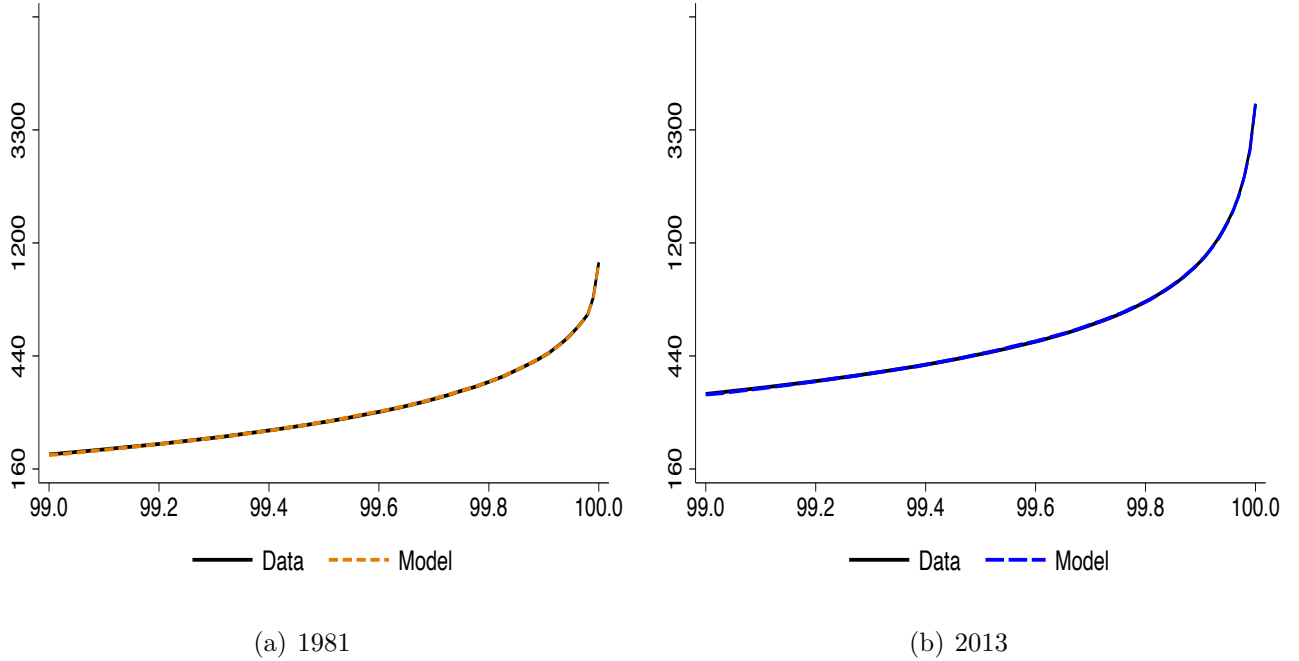
37

Figure 5: Data and Model Distribution of Individual Earnings

Figure 5 compares the top of the empirical earnings distribution to the top of the model earnings distribution. The empirical distributions are presented by solid black lines, while the model distributions are presented by dashed lines. The left panel shows the empirical and model distribution for 1981, the right panel for 2013.

Table 1: Model and Data Earnings Decomposition

	Data			Model		
Moment	1981	2013	change	1981	2013	change
Total	0.65	0.85	0.20	0.66	0.84	0.18
Between	0.22	0.36	0.14	0.21	0.26	0.05
Within	0.43	0.49	0.07	0.45	0.58	0.13

Table 1 compares the empirical and model decomposition of earnings dispersion. The left panel shows the empirical decomposition of earnings variation in dispersion between firms and dispersion within firms following (43), while the right panel shows the model analog. All numbers are log points.

earnings dispersion. From 1981 to 2013 the variance of log earnings increased by 0.20 log points to 0.85. Two-thirds of this increase is accounted for by increased differences in mean earnings across firms, while only a third is attributed to an increase in within-firm earnings dispersion.

The right panel of Table 1 displays the nontargeted model decomposition for the variance of

Table 2: Model and Data Coworker Earnings

	Data			Model		
Percentile	1981	2013	change	1981	2013	change
25	10.01	10.07	0.06	10.18	10.31	0.14
50	10.30	10.45	0.15	10.44	10.62	0.18
75	10.56	10.75	0.19	10.57	10.77	0.20
90	10.67	10.98	0.31	10.47	10.62	0.15

Table 2 compares the model and data in terms of mean coworker earnings. The left panel shows mean coworker earnings at selected percentiles of the individual earnings distribution. The right panel shows their model analog. All numbers are log points.

log earnings between 1981 and 2013. Our stylized model endogenously accounts for most of the earnings decomposition in 1981. The fifth column shows that the model correctly attributes a third of the earnings variation to dispersion between firms, and two-thirds to dispersion within firms. By construction, the model generates an increase in the variance of log earnings of 0.18 log points between 1981 and 2013, and endogenously attributes about two-thirds of earnings dispersion in 1981 and 2013 to within-firm variation.[26]

In addition to studying the variance decomposition, we examine average earnings of coworkers at different percentiles of the earnings distribution. We evaluate mean log earnings of coworkers, and changes herein, at different percentiles of the earnings distribution. The results are in Table 2. Model and data align qualitatively up to the 75th percentile with quantitative deviations of roughly 0.15 log points. At the top of the skill distribution, the data display a stronger correlation between worker earnings compared to our more negatively sorted model economy.

The rise in earnings dispersion between 1981 and 2013 could be driven by changes in the worker or the firm side of the labor market, or a combination thereof. We use our framework to evaluate the drivers of the increase in earnings dispersion. To do this, we analyze counterfactual changes in earnings dispersion by only changing the worker distribution and by only changing the project value distribution.

Table 3 shows a structural decomposition of changes in earnings dispersion from 1981 to 2013.

---

[26]The difference between model and data in terms of total variation is due to rounding.

Table 3: Model and Data Earnings Decomposition

Moment	Model			Firm Effect		Worker Effect	
	1981	2013	change	2013	change	2013	change
Total	0.66	0.84	0.18	0.80	0.14	0.73	0.07
Between	0.21	0.26	0.05	0.24	0.03	0.20	-0.01
Within	0.45	0.58	0.13	0.55	0.12	0.52	0.07

Table 3 compares the baseline and counterfactual model decomposition of earnings dispersion. The left panel shows the baseline model decomposition of earnings as in Table 1, while the middle and right panel show counterfactual decompositions. For the firm effect counterfactual, we evaluate the model with the worker distribution for 1981 and the firm distribution for 2013, while the worker effect counterfactual evaluates the model using the firm distribution for 1981 and the worker distribution for 2013. All numbers are log points.

The left panel repeats the baseline decomposition, while the middle panel and right panel present counterfactual results. The firm effect counterfactual evaluates the model for the distribution of workers in 1981 and the firm distribution in 2013. The middle panel demonstrates that by only changing the distribution of firm projects, we obtain most of the increase in earnings dispersion. The worker effect counterfactual similarly evaluates the model using the distribution of firms in 1981 and the worker distribution in 2013. The right panel shows that the change in the worker skill distribution has a smaller effect on the observed change in earnings than the change in the firm productivity distribution. That is, our counterfactual analysis gives evidence that the change in the firm project distribution between 1981 and 2013 has a larger effect on the observed change in earnings dispersion than the change in the worker skill distribution.

6 Conclusion

We provide a complete solution to an assignment problem with heterogeneous firms and multiple heterogeneous workers whose skills are imperfect substitutes, that is, for a submodular technology. While most assignment problems feature sorting which is either positive or negative, we show that equilibrium sorting in our multi-type model is neither and characterize it fully in closed form.

References

AHLIN, C. (2015): "The Role of Group Size in Group Lending," *Journal of Development Economics*, 115, 140–155.

———— (2017): "Matching Patterns when Group Size Exceeds Two," *American Economic Journal: Microeconomics*, 9(1), 352–84.

ALVAREZ, J., F. BENGURIA, N. ENGBOM, AND C. MOSER (2018): "Firms and the Decline in Earnings Inequality in Brazil," *American Economic Journal: Macroeconomics*, 10(1), 149–89.

ANTRAS, P., L. GARICANO, AND E. ROSSI-HANSBERG (2006): "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121(1), 31–77.

BARTH, E., A. BRYSON, J. C. DAVIS, AND R. FREEMAN (2016): "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States," *Journal of Labor Economics*, 34(2), 67–97.

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

BERNARD, C., X. JIANG, AND R. WANG (2014): "Risk Aggregation with Dependence Uncertainty," *Insurance: Mathematics and Economics*, 54, 93–108.

BONHOMME, S. (2021): "Teams: Heterogeneity, Sorting and Complementarity," University of Chicago Working Paper.

CARD, D., J. HEINING, AND P. KLINE (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128(3), 967–1015.

CARLIER, G., AND I. EKELAND (2010): "Matching for Teams," *Economic Theory*, 42(2), 397–418.

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.

CHERNOZHUKOV, V., A. GALICHON, M. HENRY, AND B. PASS (2021): "Single Market Non-Parametric Identification of Multi-Attribute Hedonic Equilibrium Models," MIT 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.

CHIAPPORI, P.-A., B. SALANIÉ, AND Y. WEISS (2017): "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, 107(8), 2109–67.

CHOO, E., AND A. SIOW (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114(1), 175–201.

COLOMBO, M., L. DE PASCALE, AND S. DI MARINO (2015): "Multimarginal Optimal Transport Maps for One-Dimensional Repulsive Costs," *Canadian Journal of Mathematics*, 67(2), 350–368.

COTAR, C., G. FRIESECKE, AND C. KLÜPPELBERG (2013): "Density Functional Theory and Optimal Transportation with Coulomb Cost," *Communications on Pure and Applied Mathematics*, 66(4), 548–599.

DALL'AGLIO, G. (1972): "Fréchet Classes and Compatibility of Distribution Functions," in *Symposia Mathematica*, vol. 9, pp. 131–150.

DAVIS, S. J., AND J. HALTIWANGER (1991): "Wage Dispersion Between and Within US Manufacturing Plants, 1963-86," *Brookings Papers on Economic Activity*, 1991, 115–200.

DUPUY, A., AND A. GALICHON (2014): "Personality Traits and the Marriage Market," *Journal of Political Economy*, 122(6), 1271–1319.

EDMOND, C., AND S. MONGEY (2019): "Unbundling Labor," University of Chicago Working Paper.

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

EECKHOUT, J., AND P. KIRCHER (2018): "Assortative Matching with Large Firms," *Econometrica*, 86(1), 85–132.

EMBRECHTS, P., G. PUCCETTI, L. RÜSCHENDORF, R. WANG, AND A. BELERAJ (2014): "An Academic Response to Basel 3.5," *Risks*, 2(1), 25–48.

GAFFKE, N., AND L. RÜSCHENDORF (1981): "On a Class of Extremal Problems in Statistics," *Mathematische Operationsforschung und Statistik. Series Optimization*, 12(1), 123–135.

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

——— (2021): "The Unreasonable Effectiveness of Optimal Transport in Economics," New York University Working Paper.

GALICHON, A., AND B. SALANIÉ (2021): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," New York University Working Paper.

GANGBO, W., AND A. ŚWIECH (1998): "Optimal Maps for the Multidimensional Monge-Kantorovich Problem," *Communications on Pure and Applied Mathematics*, 51(1), 23–45.

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108(5), 874–904.

GARICANO, L., AND E. ROSSI-HANSBERG (2004): "Inequality and the Organization of Knowledge," *American Economic Review Papers & Proceedings*, 94(2), 197–202.

——— (2006): "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121(4), 1383–1435.

——— (2015): "Knowledge-Based Hierarchies: Using Organizations to Understand the Economy," *Annual Review of Economics*, 7(1), 1–30.

GLADKOV, N. A., AND A. P. ZIMIN (2020): "An Explicit Solution for a Multimarginal Mass Transportation Problem," *SIAM Journal on Mathematical Analysis*, 52(4), 3666–3696.

GROSSMAN, G. M., AND G. MAGGI (2000): "Diversity and Trade," *American Economic Review*, 90(5), 1255–1275.

HERKENHOFF, K., J. LISE, G. MENZIO, AND G. M. PHILLIPS (2018): "Production and Learning in Teams," NBER Working Paper No. 25179.

JAROSCH, G., E. OBERFIELD, AND E. ROSSI-HANSBERG (2021): "Learning From Coworkers," *Econometrica*, 89(2), 647–676.

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

KARAMATA, J. (1932): "Sur une Inégalité Relative aux Fonctions Convexes," *Publications Mathématiques de l'Université de Belgrade*, 1(1), 145–147.

Kellerer, H. G. (1984): "Duality Theorems for Marginal Problems," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4), 399–432.

Kelso, A. S., and V. P. Crawford (1982): "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50(6), 1483–1504.

Kim, Y.-H., and B. Pass (2014): "A General Condition for Monge Solutions in the Multi-Marginal Optimal Transport Problem," *SIAM Journal on Mathematical Analysis*, 46(2), 1538–1550.

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," NBER Working Paper No. 5718.

Legros, P., and A. F. Newman (2002): "Monotone Matching in Perfect and Imperfect Worlds," *Review of Economic Studies*, 69(4), 925–942.

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *Review of Economic Studies*, 84(2), 718–789.

Lindenlaub, I., and F. Postel-Vinay (2020): "Multidimensional Sorting under Random Search," Yale University Working Paper.

McCann, R. J., X. Shi, A. Siow, and R. Wolthoff (2015): "Becker meets Ricardo: Multisector Matching with Communication and Cognitive skills," *Journal of Law, Economics, and Organization*, 31(4), 690–720.

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

Ocampo Díaz, S. (2018): "A Task-Based Theory of Occupations with Multidimensional Heterogeneity," Western University Working Paper.

Pass, B. (2011): "Uniqueness and Monge Solutions in the Multimarginal Optimal Transportation Problem," *SIAM Journal on Mathematical Analysis*, 43(6), 2758–2775.

Pečarić, J. E. (1984): "On Some Inequalities for Functions with Nondecreasing Increments," *Journal of Mathematical Analysis and Applications*, 98(1), 188–197.

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

PUCCETTI, G., AND R. WANG (2015): "Extremal Dependence Concepts," *Statistical Science*, 30(4), 485–517.

RACHEV, S. T., AND L. RÜSCHENDORF (1998): *Mass Transportation Problems: Volume I: Theory.* Springer Science and Business.

SAH, R. K., AND J. E. STIGLITZ (1986): "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review*, 76(4), 716–727.

SAINT-PAUL, G. (2001): "On the Distribution of Income and Worker Assignment under Intrafirm Spillovers, with an Application to Ideas and Networks," *Journal of Political Economy*, 109(1), 1–37.

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

SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2019): "Firming Up Inequality," *Quarterly Journal of Economics*, 134(1), 1–50.

WANG, B., AND R. WANG (2011): "The Complete Mixability and Convex Minimization Problems with Monotone Marginal Densities," *Journal of Multivariate Analysis*, 102(10), 1344–1360.

——— (2016): "Joint Mixability," *Mathematics of Operations Research*, 41(3), 808–826.

Sorting with Team Formation

Appendix

Job Boerma, Aleh Tsyvinski and Alexander Zimin

September 2021

A   Proofs

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

A.1   Lemma 1

To prove Lemma 1, we require weak duality.

Weak Duality. Let $\pi \in \Pi(F_{x_1}, F_{x_2}, F_z)$ be a joint probability measure, and (f, g, h) be functions such that $y(x_1, x_2, z) \leq f(x_1) + g(x_2) + h(z)$ for all (x_1, x_2, z). Then

$$\min_{f,g,h} \int f(x)\mathrm{d}F_{x_1} + \int g(x)\mathrm{d}F_{x_2} + \int h(z)\mathrm{d}F_z \geq \max_{\pi \in \Pi} \int y(x_1, x_2, z)\mathrm{d}\pi. \tag{A.1}$$

Proof. For any functions (f, g, h) so that $y(x_1, x_2, z) \leq f(x_1) + g(x_2) + h(z)$ we have:

$$\max_{\pi \in \Pi} \int y(x_1, x_2, z)\mathrm{d}\pi \leq \int \big(f(x_1) + g(x_2) + h(z)\big)\mathrm{d}\pi = \int f(x)\mathrm{d}F_{x_1} + \int g(x)\mathrm{d}F_{x_2} + \int h(z)\mathrm{d}F_z,$$

where the equality follows as $\pi \in \Pi(F_{x_1}, F_{x_2}, F_z)$. Since the above inequality holds for any (f, g, h) it holds for (f, g, h) that minimize the right-hand side.

We use weak duality to establish Lemma 1 by contradiction.

Proof of Lemma 1. Suppose by contradiction that $\hat{\pi}$ does not solve the planning problem. Then there exists another probability measure π such that

$$\max_{\pi \in \Pi} \int y(x_1, x_2, z)\mathrm{d}\pi > \int y(x_1, x_2, z)\mathrm{d}\hat{\pi} = \int \hat{f}(x)\mathrm{d}F_x + \int \hat{g}(x)\mathrm{d}F_x + \int \hat{h}(z)\mathrm{d}F_z$$

$$\geq \min_{f,g,h} \int f(x)\mathrm{d}F_x + \int g(x)\mathrm{d}F_x + \int h(z)\mathrm{d}F_z, \tag{A.2}$$

where the equality follows by assumption. This contradicts weak duality (A.1).

Suppose by contradiction that the functions $\hat{f}, \hat{g}$, and $\hat{h}$ do not solve the dual problem. Then there exists functions f, g, and h such that

$$\min_{f,g,h} \int f(x)\mathrm{d}F_x + \int g(x)\mathrm{d}F_x + \int h(z)\mathrm{d}F_z < \int \hat{f}(x)\mathrm{d}F_x + \int \hat{g}(x)\mathrm{d}F_x + \int \hat{h}(z)\mathrm{d}F_z$$
$$= \int y(x_1, x_2, z)\mathrm{d}\hat{\pi} \leq \max_{\pi \in \Pi} \int y(x_1, x_2, z)\mathrm{d}\pi, \quad \text{(A.3)}$$

where the equality follows by the assumption. This inequality contradicts weak duality (A.1).

A.2 Proposition 1

This appendix proves Proposition 1. The proof follows from the constraints on the dual planning problem, which imply:

$$w(x_1) \geq y(x_1, x_2, z) - w(x_2) - v(z) \implies w(x_1) \geq \max_{x_2, z} \big(y(x_1, x_2, z) - w(x_2) - v(z) \big). \quad \text{(A.4)}$$

The equality follows since $y(x_1, x_2, z) = w(x_1) + w(x_2) + v(z)$ on the set of potential matches M on which the assignment function is concentrated. The argument for the firm value function v is identical. The equality also implies that the goods market clears. A similar argument for one-to-one sorting is in Proposition 2.3 of Galichon (2018) and discussed in Chade, Eeckhout, and Smith (2017).

A.3 Proposition 2

We show there exists a transport plan π concentrated on the set of potential matches M. We prove this using Theorem 3.2 of Wang and Wang (2016), which we map to our environment.

Suppose we are given some cumulative distribution functions G_i for $i = 1, 2, 3$, with corresponding continuous density functions g_i that are all decreasing on their support $[a_i, b_i]$. Wang and Wang (2016) show there exists a measure $\pi \in \Pi(G_1, G_2, G_3)$ so that $x_1 + x_2 + x_3 = \mathcal{C}_e$ almost everywhere with respect to the measure π if and only if:

$$\sum a_i + \max_i \ l_i \ \leq \sum \mu_i \leq \sum b_i - \max_i \ l_i, \quad \text{(A.5)}$$

with length of support $l_i := b_i - a_i$, and mean of the distribution $\mu_i := \int x_i \mathrm{d}G_i$.

To map our economy to this result, we let $x_1 = \exp(-\hat{x}_1)$, $x_2 = \exp(-\hat{x}_2)$ and $z = \exp(-\hat{x}_3)$, and use that our distributions for worker and firm types are identical. With this notation, $x_1 x_2 z =$

$\mathcal{C}$ is equivalent to $\hat{x}_1 + \hat{x}_2 + \hat{x}_3 = -\log \mathcal{C}$. The distribution function for $\hat{x}$ is related to the distribution for x through $G(t) = 1 - \hat{F}(\exp(-t))$, where $\hat{F}$ is the distribution function F restricted to the interval $[x(\underline{p}), x(\bar{p})]$. The distribution function G then has a corresponding density function:

$$g(t) = \frac{\mathrm{d}}{\mathrm{d}t} G(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \hat{F}(\exp(-t)) = \hat{f}(\exp(-t)) \exp(-t). \tag{A.6}$$

Since $x f(x)$ is increasing on the interval $[\underline{p}, \bar{p}]$ by assumption, and therefore $x \hat{f}(x)$ is increasing, the density function g is decreasing.

To apply the existence result of Wang and Wang (2016), we verify condition (A.5) is satisfied. To show this condition holds we observe $\hat{x} = -\log x = -\log I(p)$. The lower bound of the support is thus $a = -\log I(\bar{p})$ while the upper bound is $b = -\log I(\underline{p})$. The left-hand side of (A.5) is thus $-2\log I(\bar{p}) - \log I(\underline{p})$.

To evaluate the middle of (A.5), we examine the mean of the transformed distribution:

$$\mu = \int \hat{x} \mathrm{d}G = \int \left(-\log x\right) \mathrm{d}\hat{F} = \int_0^1 \left(-\log \hat{I}(\hat{p})\right) \mathrm{d}\hat{p} = \frac{1}{\bar{p} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \left(-\log I(p)\right) \mathrm{d}p, \tag{A.7}$$

where $p = \underline{p} + (\bar{p} - \underline{p})\hat{p}$ and $\hat{I}$ is the inverse cumulative distribution function for $\hat{F}$. Since the distributions are identical, the middle of condition (A.5) is $\mu = -\frac{3}{\bar{p} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p$, and we can write the first inequality in (A.5) as:

$$2\log I(\bar{p}) + \log I(\underline{p}) \geq \frac{3}{\bar{p} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p = \frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p, \tag{A.8}$$

where the equality follows since $\bar{p} = 1 - 2\underline{p}$. This inequality is satisfied by the specification of the constant in (35) of Proposition 4.

We confirm that the second inequality of condition (A.5) is satisfied using the first inequality. Since the density function g decreases on support $[a, b]$, the mean is below the median of the interval, that is, $\mu \leq \frac{a+b}{2}$, implying $b - \mu \geq \mu - a$. Intuitively, the average equals the midpoint of the support for a uniform distribution, and decreasing density functions put more weight on low values. It thus follows that $l \leq \sum(\mu - a) \leq \sum(b - \mu)$ implying the second inequality is also satisfied as $\sum b - l \geq \sum \mu$. Hence, there exists a measure $\pi \in \Pi(\hat{F}_x, \hat{F}_x, \hat{F}_z)$ such that $x_1 x_2 z = \mathcal{C}$ almost everywhere with respect to the measure π.

A.4   Proposition 3

In this appendix, we verify Assumption 2 to Karamata's inequality.

For any assignment π, let $S_\gamma(t)$ denote the cumulative logarithmic losses for the t teams with the smallest losses:

$$S_\gamma(t) = \int_0^t \log I_\gamma(p)\mathrm{d}p. \tag{A.9}$$

To verify the assumption we thus compare S_{γ^*} to S_γ.

We first construct the cumulative logarithmic losses attained by the pairwise countermonotonic sets under the candidate optimal assignment S_{γ^*}.

Lemma 2. For any $t \in [0, 3\underline{p}]$,

$$S_{\gamma^*}(t) = 3 \int_0^{\frac{t}{3}} L(p)\mathrm{d}p. \tag{A.10}$$

where $L(p)$ is the logarithm of output losses on the negatively assorted sets:

$$L(p) = 2\log I(1 - 2p) + \log I(p) \tag{A.11}$$

Proof. The statement is equivalent to the statement that the t teams with the smallest losses under assignment π^* are in the three pairwise countermonotonic sets:

$$\big(I(p), I(1 - 2p), I(1 - 2p)\big), \quad \big(I(1 - 2p), I(p), I(1 - 2p)\big), \quad \big(I(1 - 2p), I(1 - 2p), I(p)\big), \tag{A.12}$$

for percentiles $p \in \left[0, \frac{t}{3}\right)$. We need to show that losses $I(p)I^2(1 - 2p)$ increase on the interval $[0, \underline{p})$ which is equivalent to verifying that the logarithmic losses $L(p)$ increase on this interval.

We establish that logarithmic losses $L(p)$ are strictly increasing in the interval $[0, \underline{p})$. Observe that by taking logs of equation (35), which pins down the lower percentile bound $\underline{p}$:

$$L\left(\underline{p}\right) = \frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p)\mathrm{d}p = \frac{1}{\frac{1}{3} - \underline{p}} \left[\int_{\underline{p}}^{\frac{1}{3}} \log I(p)\mathrm{d}p - \int_{\bar{p}}^{\frac{1}{3}} \log I(p)\mathrm{d}p\right]. \tag{A.13}$$

We simplify the expression by changing the bounds of integration on the final integral so that:[27]

$$L\left(\underline{p}\right) = \frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\frac{1}{3}} \big(2\log I(1 - 2p) + \log I(p)\big)\mathrm{d}t = \frac{1}{\frac{1}{3} - \underline{p}} \int_{\underline{p}}^{\frac{1}{3}} L(p)\mathrm{d}p, \tag{A.14}$$

where the second equality follows by (A.11). The output loss at the lower percentile cutoff $\underline{p}$ is equal to the average output loss for all remaining negative sorting possibilities. Since the log output loss L is concave by footnote 18, L is strictly increasing for percentiles $p \in [0, \underline{p}]$. $\qquad\square$

---

[27]Let $\tilde{p} := \frac{1-p}{2}$ implying $\mathrm{d}p = -2\mathrm{d}\tilde{p}$, so that we can write $\int_{\bar{p}}^{\frac{1}{3}} \log I(p)\mathrm{d}p = \int_{1-2\underline{p}}^{\frac{1}{3}} \log I(p)\mathrm{d}p = -2 \int_{\underline{p}}^{\frac{1}{3}} \log I(1-2\tilde{p})\mathrm{d}\tilde{p}$ by change of variables.

We next bound the cumulative logarithmic losses for the t teams with the smallest losses, $S_\gamma(t)$, from above for any feasible assignment π.

Lemma 3. For any feasible assignment π and for all $t \in [0, 1]$,

$$S_\gamma(t) \le 3 \int_0^{\frac{t}{3}} L(p)\mathrm{d}p \tag{A.15}$$

Proof. The function S_γ is the sum of logarithmic losses for the t teams with smallest output losses. For any other set U with mass $\pi(U) = t$, by definition:

$$S_\gamma(t) \le \int \log \ell(x_1, x_2, z)\mathrm{d}\pi_U. \tag{A.16}$$

where π_U represent the restriction of assignment π to the set V.

Define the set with the $\frac{t}{3}$ highest skill workers as $V_1 = \left[0, I\left(\frac{t}{3}\right)\right] \times [0, 1] \times [0, 1]$. The measure of this set with respect to assignment π is $\frac{t}{3}$. Analogously, define the set with highest skill coworkers as $V_2 = [0, 1] \times \left[0, I\left(\frac{t}{3}\right)\right] \times [0, 1]$ and the set of low value firms as $V_3 = [0, 1] \times [0, 1] \times \left[0, I\left(\frac{t}{3}\right)\right]$, which similarly have measure $\frac{t}{3}$ with respect to assignment π. The union of sets $V = V_1 \cup V_2 \cup V_3$ thus has measure of at most t with respect to assignment π. We construct a set $U \supset V$ such that $\pi(U) = t$ by adding some set of triplets to the set U.

Given that π_U represents the restriction of the assignment function π to the set U, we write:

$$S_\gamma(t) \le \int \log \ell(x_1, x_2, z)\mathrm{d}\pi_U = \int \log x_1 \mathrm{d}\mathrm{Pr}_{x_1}(\pi_U) + \int \log x_2 \mathrm{d}\mathrm{Pr}_{x_2}(\pi_U) + \int \log z \mathrm{d}\mathrm{Pr}_z(\pi_U),$$

where Pr_z is the projection of measure π_U onto the firm productivity dimension. To characterize the upper bound more precisely, we bound $\int \log x_1 \mathrm{d}\mathrm{Pr}_{x_1}(\pi_U)$. Since $V_1 \subset U$, the density function corresponding to measure $\mathrm{Pr}_{x_1}(\pi_U)$ is equal to the skill density f for all $0 \le x \le I\left(\frac{t}{3}\right)$ and is less than skill density f otherwise. As a result,

$$\int \log x_1 \mathrm{d}\mathrm{Pr}_{x_1}(\pi_U) \le \int_0^{\frac{t}{3}} \log I(p)\mathrm{d}p + \int_{1-\frac{2}{3}t}^1 \log I(p)\mathrm{d}p = \int_0^{\frac{t}{3}} \log I(p)\mathrm{d}p + 2 \int_0^{\frac{t}{3}} \log I(1 - 2p)\mathrm{d}p.$$

where the inequality follows as the second term on the right-hand side captures the contribution to logarithmic losses by the $\frac{t}{3}$ lowest skill workers. The equality follows by change of variables. As a result, we write the upper bound $\int \log x_1 \mathrm{d}\mathrm{Pr}_{x_1}(\pi_U) \le \int_0^{\frac{t}{3}} L(p)\mathrm{d}p$, and therefore,

$$S_\gamma(t) \le 3 \int_0^{\frac{t}{3}} L(p)\mathrm{d}p, \tag{A.15}$$

which was want we wanted. $\square$

We use Lemma 2 and Lemma 3 to verify Assumption 2 of Karamata's inequality.

Proposition 10. For any assignment π, $S_\gamma(t) \le S_{\gamma^*}(t)$ for all $t \in [0,1]$.

Proof. For teams $t \in [0, 3\underline{p}]$, this inequality directly follows by combining Lemma 2 and Lemma 3.

For teams with the highest losses $t \in [3\underline{p}, 1]$, we first observe that cumulative logarithmic losses S_γ is a convex function as it is the integral over an increasing function. Since S_γ is convex, for all $t \in [3\underline{p}, 1]$:

$$S_\gamma(t) \le \frac{1-t}{1-3\underline{p}} S_\gamma(3\underline{p}) + \frac{t-3\underline{p}}{1-3\underline{p}} S_\gamma(1) \le \frac{1-t}{1-3\underline{p}} S_{\gamma^*}(3\underline{p}) + \frac{t-3\underline{p}}{1-3\underline{p}} S_{\gamma^*}(1). \tag{A.17}$$

where the final inequality follows because $S_\gamma(3\underline{p}) \le S_{\gamma^*}(3\underline{p})$ by combining Lemma 2 with Lemma 3, and $S_\gamma(1) = S_{\gamma^*}(1)$ by the verification of Assumption 1 in (33).

The integral over logarithmic losses $S_{\gamma^*}(t)$ is linear for teams with the highest losses $t \in [3\underline{p}, 1]$ as losses are constant outside the pairwise countermonotonic sets with $\mathcal{C} = I(\underline{p})I^2(1-2\underline{p})$ implying $\log I_{\gamma^*}(t) = \log \mathcal{C}$. Hence, for teams $t \in [3\underline{p}, 1]$:

$$S_{\gamma^*}(t) = \frac{1-t}{1-3\underline{p}} S_{\gamma^*}(3\underline{p}) + \frac{t-3\underline{p}}{1-3\underline{p}} S_{\gamma^*}(1) \tag{A.18}$$

Combining (A.17) and (A.18) the result follows. □

A.5 Proposition 4

If a triplet (x_1, x_2, z) is contained in the set of potential matches M, then function m reflects the marginal worker product (4), or $m(x_1) = -x_2 z$, $m(x_2) = -x_1 z$, and $m(z) = -x_1 x_2$. Since the matching set (26) consists of four subsets, we prove the result by establishing the claim for each subset.

If triplet (x_1, x_2, z) is in the negative sorted set corresponding to high skill worker x_1, $(x_1, x_2, z) \in M_{x_1}$ then $(x_1, x_2, z) = (I(p_x), I(1-2p_x), I(1-2p_x))$ for percentile $p_x \in [0, \underline{p})$. Hence, the marginal worker products are $m(x_1) = -I^2(1 - 2p_x)$ and $m(z) = -I(p_x)I(1 - 2p_x) = -I\left(\frac{1-p_z}{2}\right)I(p_z)$, with the argument for lower percentiles of coworkers $m(x_2)$ following the logic for $m(z)$. By symmetry of the negatively sorted branches, the same logic applies for triplets (x_1, x_2, z) contained in either subset M_{x_2} or M_z.

If triplet (x_1, x_2, z) is in the mixed middle of potential matches $M_{\mathcal{C}}$ then the expected output loss equals constant $\mathcal{C}$, or $\ell(x_1, x_2, z) = x_1 x_2 z = \mathcal{C}$. The marginal products in the mixed middle are then given by $m(x_1) = -\mathcal{C}/x_1 = -x_2 z$, $m(x_2) = -\mathcal{C}/x_2 = -x_1 z$, and $m(z) = -\mathcal{C}/z = -x_1 x_2$.

A.6 Proposition 5

To establish continuity, it suffices to verify that the marginal worker product m is continuous at the cutoffs $\underline{p}$ and $\bar{p}$. At both the lower and the upper percentile cutoff, continuity follows directly by substituting the condition that pins down the cutoff (35).

To verify that the marginal worker product m is increasing, we show its derivative is positive everywhere except at both cutoffs. At both the cutoffs, the marginal worker product features a kink and the derivative does not exist. We show the derivative is positive by equivalently establishing that the derivative of $\log(-m(x))$ is negative. To see this, start by using the marginal product in (34) to write:

$$\log\big(-m(x)\big) = \begin{cases} 2 \log I\,(1-2p) & \text{if } p \in [0, \underline{p}] \\[2mm] \log \mathcal{C} - \log I(p) & \text{if } p \in (\underline{p}, \bar{p}) \\[2mm] \log I(p) + \log I\left(\frac{1-p}{2}\right) & \text{if } p \in [\bar{p}, 1] \end{cases} \tag{A.19}$$

For high skill workers in interval $[0, \underline{p}]$, the derivative is negative since the inverse function for any cumulative distribution function is nondecreasing, that is, higher values are associated with higher percentiles. For workers in the mixed middle in the interval $(\underline{p}, \bar{p})$, the derivative is for the same reason also negative, $-\frac{I'(p)}{I(p)} < 0$.

To show that the marginal worker product is increasing for low skill workers write $\log(-m(x))$, by change of variables with $t = \frac{1-p}{2}$, as:

$$\log I(1-2t) + \log I\,(t) = L(t) - \log I(1-2t), \tag{A.20}$$

where the loss L is defined by (A.11). By Lemma 2, L increases for all $t \in [0, \underline{p}]$ and so decreases for $p \in [\bar{p}, 1]$. The second term also increases in $t \in [0, \underline{p})$ and therefore decreases in $p \in [\bar{p}, 1]$. As a result, the derivative is negative for all $p \in [0, 1]$.

A.7 Dual Problem

To characterize the solution to the dual problem, we use a generalization of Kantorovich duality for multiple marginal distributions (Kellerer (1984)).[28] Kantorovich duality states that the optimal value for the planning problem coincides with the optimal value for the dual problem.

Duality. Let X and Z be compact subsets of the set of real numbers, and F_x and F_z be probability measures on X and Z respectively. Let loss function $\ell(x_1, x_2, z)$ be continuous on $X \times X \times Z$ and $\Pi(F_x, F_x, F_z)$ denote the set of measures π on $X \times X \times Z$ with fixed projections $\mathrm{pr}_x(\pi) = F_x$ and $\mathrm{pr}_z(\pi) = F_z$. Then,

$$\min_{\pi \in \Pi} \int \ell(x_1, x_2, z) \mathrm{d}\pi = \sup \int \hat{w}_1(x_1) \mathrm{d}F_x + \int \hat{w}_2(x_2) \mathrm{d}F_x + \int \hat{v}(z) \mathrm{d}F_z, \quad (A.21)$$

where the supremum is taken with respect to the collection of integrable functions $\hat{w}_i \in L^1(X, F_x)$ and $v \in L^1(z, F_z)$ subject to the constraint $\hat{w}_1(x_1) + \hat{w}_2(x_2) + \hat{v}(z) \leq \ell(x_1, x_2, z)$ for any triplet $(x_1, x_2, z) \in X \times X \times Z$.

Applying the duality theorem to our problem, we establish that the dual solution attains the same value as the planning problem and that the dual solution exists in the class of integrable functions. We next establish that the solution exists in the class of continuous functions.

Lemma 4. For any dual solution (w_1, w_2, v), there exists a continuous functions $(\hat{w}_1, \hat{w}_2, \hat{v})$ such that $\hat{w} = w$ and $\hat{v} = v$ almost everywhere that solves the dual problem.

Proof. For any dual solution (w_1, w_2, v) there exists:

$$\hat{w}_1(x_1) := \sup_{x_2, z} \ y(x_1, x_2, z) - w_2(x_2) - v(z), \quad (A.22)$$

where $\hat{w}_1(x_1) \leq w_1(x_1)$ since $w_1(x_1) \geq y(x_1, x_2, z) - w_2(x_2) - v(z)$. Similarly, we define $\hat{w}_2$ and $\hat{v}$ as $\hat{w}_2(x_2) := \sup_{x_1, z} \ y(x_1, x_2, z) - \hat{w}_1(x_1) - v(z)$ and $\hat{v}(z) := \sup_{x_1, x_2} \ y(x_1, x_2, z) - \hat{w}_1(x_1) - \hat{w}_2(x_2)$. As a consequence, there exists:

$$\tilde{w}_1(x_1) := \sup_{x_2, z} \ y(x_1, x_2, z) - \hat{w}_2(x_2) - \hat{v}(z). \quad (A.23)$$

---

[28] An instructive exposition of Kellerer (1984) is presented in Rachev and Rüschendorf (1998).

Since $\hat{w}_2(x_2) \leq w_2(x_2)$ and $\hat{v}(z) \leq v(z)$, it follows that $\tilde{w}_1(x_1) := \sup_{x_2, z} y(x_1, x_2, z) - \hat{w}_2(x_2) - \hat{v}(z) \geq \hat{w}_1(x_1)$. Simultaneously, by the definition of $\hat{v}$, it follows that $\hat{w}_1(x_1) \geq y(x_1, x_2, z) - \hat{w}_2(x_2) - \hat{v}(z)$, which implies $\hat{w}_1(x_1) \geq \tilde{w}_1(x_1)$. As a result $\hat{w}_1(x_1) = \tilde{w}_1(x_1)$, which implies:

$$\hat{w}_1(x_1) := \sup_{x_2, z} y(x_1, x_2, z) - \hat{w}_2(x_2) - \hat{v}(z), \tag{A.24}$$

and similarly for $\hat{w}_2(x_2)$.

Next we establish $(\hat{w}_1, \hat{w}_2, \hat{v})$ is a dual solution. By (A.24), $(\hat{w}_1, \hat{w}_2, \hat{v})$ is a permissible solution, while through $\hat{w}_1(x_1) \leq w_1(x_1)$, $\hat{w}_2(x_2) \leq w_2(x_2)$ and $\hat{v}(z) \leq v(z)$, it follows that $\int \hat{w}_1(x_1) \mathrm{d}F_x + \int \hat{w}_2(x_2) \mathrm{d}F_x + \int \hat{v}(z) \mathrm{d}F_z \leq \int w_1(x_1) \mathrm{d}F_x + \int w_2(x_2) \mathrm{d}F_x + \int v(z) \mathrm{d}F_z$. Thus, $(\hat{w}_1, \hat{w}_2, \hat{v})$ is a dual solution. Further, $\hat{w}_1 = w_1$, $\hat{w}_2 = w_2$ and $\hat{v} = v$ almost everywhere. Suppose instead $\hat{w}_1 < w_1$, for example, then $\int \hat{w}_1(x_1) \mathrm{d}F_x + \int \hat{w}_2(x_2) \mathrm{d}F_x + \int \hat{v}(z) \mathrm{d}F_z < \int w_1(x_1) \mathrm{d}F_x + \int w_2(x_2) \mathrm{d}F_x + \int v(z) \mathrm{d}F_z$, which contradicts (w_1, w_2, v) is a dual solution.

We next establish that $\hat{w}_1$ is a continuous function. To see this, let $\hat{w}_1(x_1; x_2, z) := y(x_1, x_2, z) - \hat{w}_2(x_2) - \hat{v}(z)$, that is, equation (A.24) holding constant both coworker x_2 and the firm z. Given the specification of the production technology (1), and since both $x_2 \in [0, 1]$ and $z \in [0, 1]$, $\hat{w}_1(x_1; x_2, z)$ is a linear function with a slope less than one in absolute value. This implies that $\hat{w}_1(x_1; x_2, z)$ is Lipschitz continuous. Since $\hat{w}_1(x_1)$ can be represented as $\hat{w}_1(x_1) = \sup_{x_2, z} \hat{w}_1(x_1; x_2, z)$, $\hat{w}_1(x_1)$ is Lipschitz continuous, and hence $\hat{w}_1(x_1)$ is continuous. The same argument shows both $\hat{w}_2$ and $\hat{v}$ are continuous. □

A.8 Proposition 6

For any worker $x_{1,0}$ there exists a coworker $x_{2,0}$ and firm z_0 so that triplet $(x_{1,0}, x_{2,0}, z_0)$ is contained in the support of an optimal assignment, which implies that:

$$w(x_{1,0}) + w(x_{2,0}) + v(z_0) = y(x_{1,0}, x_{2,0}, z_0) = z_0 + x_{1,0}m(x_{1,0}), \tag{A.25}$$

with the second equality following from the marginal worker product (4). For any other worker x_1, it must similarly be that $w(x_1) + w(x_{2,0}) + v(z_0) \geq z_0 + x_1 m(x_{1,0})$. Combining these expressions, it holds that for any x_0 and $x_1 \in [0, 1]$:

$$w(x_1) - w(x_{1,0}) \geq m(x_{1,0})(x_1 - x_{1,0}). \tag{A.26}$$

Similarly, by interchanging $x_{1,0}$ and x_1 in the previous paragraph, we obtain:

$$w(x_{1,0}) - w(x_1) \geq m(x_1)(x_{1,0} - x_1). \tag{A.27}$$

9

Combining (A.26) and (A.27), and dividing by $x_1 - x_{1,0}$ yields

$$m(x_1) \geq \frac{w(x_1) - w(x_{1,0})}{x_1 - x_{1,0}} \geq m(x_{1,0}). \tag{A.28}$$

Taking limits as $x_1 \to x_{1,0}$, this equation shows that the function $w(x)$ has a derivative at $x_1 = x_{1,0}$ and that this derivative is equal to $m(x_{1,0})$. Since the marginal worker product m is a continuous function by Proposition 5, the derivative of w is continuous. The result is similarly established for $v'(z)$.

A.9 Proposition 9

We establish the model can rationalize the empirical wage distribution using Lemma 5.

Lemma 5. If (a) the empirical earnings function is convex in percentile rank, and (b) if the model rationalizes the empirical earnings distribution, then Assumption 1 is satisfied.

Proof of Lemma 5. To show that both parts of Assumption 1 are satisfied, we develop the model earnings equation:

$$w(I(p)) = \mathcal{C}_w + \int_0^{I(p)} m(s)\mathrm{d}s, \tag{36}$$

which is equation (36) with $x = I(p)$. The derivative of the model earnings equation with respect to percentile rank p, which is the model analog of the slope of the empirical earnings function, is:

$$w'(I(p))I'(p) = m(I(p))I'(p).$$

For the model to generate the observed earnings distribution as an equilibrium outcome, it has to hold that $m(I(p))I'(p) = e'(p)$. Using the expression for the equilibrium marginal worker product (34) from Proposition 4, we thus require:

$$e'(p) = \begin{cases} -I^2 \left(1 - 2p\right) I'(p) & \text{if } p \in [0, \underline{p}] \\ -\mathcal{C}/I(p)I'(p) & \text{if } p \in (\underline{p}, \bar{p}) \\ -I(p)I\left(\frac{1-p}{2}\right) I'(p) & \text{if } p \in [\bar{p}, 1] \end{cases} \tag{A.29}$$

We prove that if the empirical earnings function is decreasing and convex, and if the model rationalizes this empirical earnings function, then Assumption 1 is satisfied. We start with the

10

condition that $\log I(p)$ is concave for medium skill workers $(\underline{p}, \bar{p})$. In this case, we can write (A.29) as:

$$-\frac{e'(p)}{\mathcal{C}} = \frac{\mathrm{d}}{\mathrm{d}p} \log I(p). \tag{A.30}$$

When the earnings function is convex the left-hand side decreases with p, implying $\frac{\mathrm{d}}{\mathrm{d}p} \log I(p)$ is decreasing or $\log I(p)$ is concave. By footnote 17, this is equivalent to part (a) of Assumption 1.

We establish part (b) of Assumption 1 is satisfied starting with the low skill workers $p \in [\bar{p}, 1]$. We use equation (A.29) for $p \in [\bar{p}, 1]$, and apply a change of variables $q = \frac{1-p}{2}$, to write:

$$e'(1 - 2q) = -I(q) I(1 - 2q)I'(1 - 2q), \tag{A.31}$$

for $q \in [0, \underline{p}]$. To show the concavity of the logarithmic output loss $L(p)$ for the high skill workers $p \in [0, \underline{p}]$ one can equivalently show concavity of $I(p)I^2(1 - 2p)$ since a logarithmic transformation of a positive and concave function is concave. Using the first part of equation (A.29) and (A.31), the derivative of $I(p)I^2(1 - 2p)$ can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}p}\big(I(p)I^2(1 - 2p)\big) = -\frac{\mathrm{d}}{\mathrm{d}p}\big(e(p) + 2e(1 - 2p)\big), \tag{A.32}$$

and hence the second derivative is $-e''(p) - 8e''(1 - 2p) \leq 0$. Since the empirical earnings function is convex, the function $L(p)$ is concave for high skill workers on interval $p \in [0, \underline{p}]$. Concavity of the logarithmic output loss $L(p)$ on the medium skill interval $(\underline{p}, \bar{p})$ is established using that $\log I(p)$ is concave for medium workers $(\underline{p}, \bar{p})$, implying that $\log I(1 - 2p)$ is concave on the same interval $(\underline{p}, \bar{p})$. Since the sum of concave functions is concave, $L(p)$ is concave for medium skill workers.

We established that $L(p)$ is concave for high skill workers $p \in [0, \underline{p}]$ as well as for medium skill workers $p \in (\underline{p}, \bar{p})$. To establish that the function is concave on the full interval $[0, \frac{1}{3}]$, we further verify continuity of the derivative at cutoff percentile $\underline{p}$. From (A.32) we have $I(p)I^2(1 - 2p) = \mathcal{C}_d - e(p) - 2e(1-2p)$, implying that the derivative at percentile $\underline{p}$ is given by $L'(p) = \frac{-e'(p)+4e'(1-2p)}{I(p)I^2(1-2p)}$. For medium skill workers $(\underline{p}, \bar{p})$, the derivative of $L(p)$ at the cutoff $\underline{p}$ can, using (A.30), be shown to be identical. Therefore, $L(p)$ is concave on the interval $[0, \frac{1}{3}]$.

Proof of Proposition 9. To prove the result, we define $g'(p) := -e'(p)$ and $u(p) := I^2(1 - 2p)$,

implying $u'(p) = -4I(1 - 2p)I'(1 - 2p)$. Using these definitions, we can write (A.29) as:

$$
g'(p) = \begin{cases}
u(p)I'(p) & \text{if } p \in [0, \underline{p}] \\
\mathcal{C}/I(p)I'(p) & \text{if } p \in (\underline{p}, \bar{p}) \\
I(p)I\left(\frac{1-p}{2}\right)I'(p) & \text{if } p \in [\bar{p}, 1]
\end{cases}
\tag{A.29}
$$

We simplify the third part of equation (A.29) by change of variables $p = 1 - 2q$, so that $g'(1 - 2q) = I(1 - 2q)I(q)I'(1 - 2q) = -\frac{1}{4}u'(q)I(q)$ for $q \in [0, \underline{p}]$. As a result, we have the following system of differential equations on the interval $[0, \underline{p}]$:

$$
I'(p)u(p) = g'(p) \tag{A.33}
$$

$$
I(p)u'(p) = -4g'(1 - 2p) \tag{A.34}
$$

We solve the system of differential equations by collapsing it to a single differential equation. We add equations (A.33) and (A.33) to write:

$$
\frac{\mathrm{d}}{\mathrm{d}p}\big(I(p)u(p)\big) = \frac{\mathrm{d}}{\mathrm{d}p}\big(g(p) + 2g(1 - 2p)\big). \tag{A.35}
$$

By integrating between 0 and $p \in [0, \underline{p}]$, setting $I(0)$ equal to small $\varepsilon > 0$ and noting that $u(0) = 1$, we have:

$$
I(p)u(p) = \varepsilon + g(p) + 2g(1 - 2p) - g(0) - 2g(1) \tag{A.36}
$$

We divide (A.34) by (A.36) to obtain a single differential equation for u:

$$
\frac{u'(p)}{u(p)} = \frac{\mathrm{d}}{\mathrm{d}p}\log u(p) = -\frac{4g'(1 - 2p)}{\varepsilon + g(p) + 2g(1 - 2p) - g(0) - 2g(1)}, \tag{A.37}
$$

which is solved using initial condition $u(0) = 1$, or $\log u(0) = 0$. Given a solution for u, and given the initial condition for $I(0)$, we use (A.33) to obtain $I(p)$. As a result, we obtain $I(p)$ on the set $[0, \underline{p}] \cup [\bar{p}, 1]$. Finally, to obtain the type distribution on the interval $[\underline{p}, \bar{p}]$, we directly use (A.29) for medium skill workers, where the boundary conditions are pinned down by $I(\underline{p})$ and $I(\bar{p})$.

B   Generalization

In this appendix we lay out how our main findings extend to n workers per team by constructing the set of potential matches M and the marginal worker product. Since the logic is analogous to the discretized analysis in Section 3, we only provide minimal discussion here.

In the generalization with n workers per team, the firm output function y is:

$$y(x_1, x_2, \ldots, x_n, z) = \left(1 - \prod_{i=1}^{n} x_i\right) z. \tag{A.38}$$

To explore sufficient conditions, we first generalize the notion of complete mixing.

Definition 3. Set $\{(x_{1s}, x_{2s}, \ldots, x_{ns}, z_s)\}_{s=1}^{m}$ is <u>completely mixed</u> if $x_{1s}x_{2s}\ldots x_{ns}z_s = \mathcal{C}$ for all s for constant $\mathcal{C}$.

The assignment is completely mixed when losses are constant across all pairings, $x_{1s}x_{2s}\ldots x_{ns}z_s = \mathcal{C}$ for all s. A completely mixed assignment is product countermonotonic because $z_s = \frac{\mathcal{C}}{x_{1s}x_{2s}\ldots x_{ns}}$ shows higher project values z_s sort with more productive teams, that is, lower $x_{1s}x_{2s}\ldots x_{ns}$. The same holds true for the n other combinations of teams and workers.

By the inequality of arithmetic and geometric means, a completely mixed assignment is optimal. With a set of elements $\{(x_{1s}, x_{2s}, \ldots, x_{ns}, z_s)\}_{s=1}^{m}$, the inequality reads:

$$\frac{1}{m} \sum x_{1s} x_{2\sigma_2(s)} \ldots x_{n\sigma_n(s)} z_{\sigma_z(s)} \geq \sqrt[m]{\prod x_{1s}x_{2s}\ldots x_{ns}z_s} =: \mathcal{C}_m, \tag{A.39}$$

When an assignment is completely mixed, it is optimal since it minimizes aggregate output losses. As in our baseline, however, there exists no optimal completely mixed assignment with atomless distributions that have full support on the unit interval.

The optimal assignment is the largest possible set of completely mixed triplets combined with pairwise countermonotonic sets. Let $\hat{\pi}$ be an optimal assignment concentrated on the completely mixed set, so $x_1 x_2 \ldots x_n z = \mathcal{C}$. Since $\sum_{i=1}^{n} \log x_i + \log z = \log \mathcal{C}$ almost everywhere with respect to assignment $\hat{\pi}$, integration yields:

$$\log \mathcal{C} = \sum_{i=1}^{n} \int_{\underline{x}_i}^{\bar{x}_i} \log x_i \mathrm{d}\hat{\pi} + \int_{\underline{z}}^{\bar{z}} \log z \mathrm{d}\hat{\pi} = \frac{n+1}{\bar{p} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p)\mathrm{d}p, \tag{A.40}$$

where the final expression follows since the marginal distributions of feasible assignment $\hat{\pi}$ on X and Z are distribution F on the percentile interval $[\underline{p}, \bar{p}]$, scaled by $\bar{p} - \underline{p}$ to normalize to unit mass. By Wang and Wang (2016), the maximal completely mixed set in this case satisfies $\mathcal{C} = I^n(\bar{p})I(\underline{p})$.

Conjecture. We specify the support of an optimal assignment, the set of possible matches M. As discussed, the set of possible matches combines $n + 1$ symmetric pairwise countermonotonic sets and a mixed middle:

$$M := M_{x_1} \cup M_{x_2} \cup \cdots \cup M_{x_n} \cup M_z \cup M_{\mathcal{C}}, \tag{A.41}$$

where:

$$M_{x_i} := \left\{ \left(\underbrace{1 - np_{x_i}, \ldots, 1 - np_{x_i}}_{i-1 \text{ workers}}, p_{x_i}, \underbrace{1 - np_{x_i}, \ldots, 1 - np_{x_i}}_{n-i \text{ workers}} \right) \mid p_{x_i} \in [0, \underline{p}) \right\}$$

$$M_z := \left\{ \left(\underbrace{1 - np_z, \ldots, 1 - np_z}_{n \text{ workers}}, p_z \right) \mid p_z \in [0, \underline{p}) \right\}$$

$$M_{\mathcal{C}} := \left\{ (p_{x_1}, p_{x_2}, \ldots, p_{x_n}, p_z) \mid \underline{p} \leq p_{x_1}, p_{x_2}, \ldots, p_{x_n}, p_z \leq \bar{p} \text{ and } I(p_{x_1}) \ldots I(p_{x_n}) I(p_z) = \mathcal{C} \right\}$$

The pairwise countermonotonic set which corresponds to pairings for high skill workers x_i is M_{x_i}. Finally, $M_{\mathcal{C}}$ is the completely mixed set, the set of matches with constant expected output losses.

Using this expression for the support of an optimal assignment π, we can formulate the marginal worker product.

Proposition 11. *Marginal Product.* The marginal product of worker x is:

$$m(x) = \begin{cases} -I^n(1 - np) & \text{if } p \in [0, \underline{p}] \\ -\mathcal{C}/I(p) & \text{if } p \in (\underline{p}, \bar{p}) \\ -I(p)I\left(\frac{1-p}{n}\right) & \text{if } p \in [\bar{p}, 1] \end{cases} \tag{A.42}$$

where $\mathcal{C} = \exp\left(\frac{1}{\frac{1}{n+1} - \underline{p}} \int_{\underline{p}}^{\bar{p}} \log I(p) \mathrm{d}p \right)$, with lower percentile bound $\underline{p} \in \left(0, \frac{1}{n+1}\right]$ satisfying:

$$\mathcal{C} = I^n(\bar{p}) I(\underline{p}), \tag{A.43}$$

and the upper percentile bound is given by $\bar{p} = 1 - n\underline{p}$.

Using an analogous proof, any assignment concentrated on a subset of the set of potential matches M solves the planning problem. To characterize the dual solution with n workers per team, we use the marginal product under an optimal assignment (A.42). Invoking the generalized duality argument of Kellerer (1984), we can use the marginal product to formulate equilibrium wages and firm values as in Section 4. As in our baseline analysis, wages are decreasing and convex in skills, while firm values are increasing and convex in firm productivity.