

Sorting with Teams*

Job Boerma

University of Wisconsin-Madison

Aleh Tsyvinski

Yale University

Alexander P. Zimin

MIT and HSE

November 2023

Abstract

We fully solve a sorting problem with heterogeneous firms and multiple heterogeneous workers whose skills are imperfect substitutes. We show that optimal sorting, which we call mixed and countermonotonic, is comprised of two regions. In the first region, mediocre firms sort with mediocre workers and coworkers such that the output losses are equal across all these teams (mixing). In the second region, a high skill worker sorts with low skill coworkers and a high productivity firm (countermonotonicity). We characterize the equilibrium wages and firm values. Quantitatively, our model can generate the dispersion of earnings within and across US firms.

JEL-Codes: J01, D31, C78.

Keywords: Sorting, Teams, Assignment.

*We thank Anmol Bhandari, Carter Braxton, Hector Chade, Jan Eeckhout, Fatih Guvenen, Kyle Herkenhoff, Rasmus Lentz, Ilse Lindenlaub, Jeremy Lise, Paolo Martellini, Simon Mongey, Chris Moser, Guillaume Sublet, and Ruodu Wang for useful discussion.

1 Introduction

How do heterogeneous workers sort to work in teams at heterogeneous firms? Which workers 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 workers 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 assortative. [Becker \(1973\)](#) shows that optimal sorting is positive under supermodular technologies and negative under submodular technologies.¹ Positive sorting, or comonotonicity, readily extends to settings with multiple workers and is generally optimal with supermodular technologies. On the other hand, negative sorting, or countermonotonicity, does not have a simple multi-type equivalent.

An important open question is how to characterize sorting of heterogeneous firms with multiple heterogeneous workers when technology is submodular. We show that this problem can be analyzed as a multimarginal optimal transport problem and fully characterize the solution.

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 the technology is submodular. In doing so, this paper makes two contributions. First, we show that an equilibrium assignment is characterized by two types of assignment regions. In the first region, mediocre firms sort with mediocre workers and coworkers so that output losses are equal across all these teams (mixed assignment). In the second region, high skill workers sort with low skill coworkers and a high productivity firm, while high productivity firms employ low skill workers and a single high skill coworker (countermonotonic assignment). We call this a mixed and countermonotonic assignment. Second, we fully characterize the equilibrium assignment as well as workers' wages and firm values. We illustrate 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. In order to illustrate our theory, consider a simple production and team technology. The output of a firm is the product of the value of their project, or productivity

¹[Sattinger \(1993\)](#), [Chiappori and Salanié \(2016\)](#), [Chade, Eeckhout, and Smith \(2017\)](#), and [Eeckhout \(2018\)](#) provide a comprehensive review of this literature.

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, x_2, \dots, x_m) is $q = 1 - \prod_{i=1}^m x_i$, and output is $y(x_1, x_2, \dots, x_m, z) = zq = \left(1 - \prod_{i=1}^m x_i\right)z$. 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.

We show there are two regions to an optimal assignment. The first region is the mixed set. Mixing means that the output loss, $z \prod x_i$, is identical for all teams. We start with the observation that mixing is optimal whenever it is feasible. Equalizing output loss $z \prod x_i$ across teams is optimal as average expected losses attain their lower bound if and only if output losses are equal across all teams following Jensen’s inequality. Mixing implies that firms with identical projects may hire different teams of workers which have the same overall quality.

Mixing, however, does not fully describe equilibrium assignment as it is not feasible everywhere. We show that optimal sorting features the maximal mixed set combined with countermonotonic regions. To develop intuition for the optimality of countermonotonic sets, we note that at the boundary of the mixed set, the lowest value project is sorted with identical low-skill workers (high x_i). Similarly, a project which has zero value employs workers who do not know how to solve a problem ($x_i = 1$). Applying this reasoning to other low value projects outside the mixed set, these projects are countermonotonically sorted with low-skill workers.

In order to establish equilibrium properties of worker wages and firm values we characterize the dual problem. We show that the derivative of the wage schedule is equal to the marginal worker product, or $-z \prod_{j \neq i} x_j$. The increase in output obtained by replacing a worker with a slightly more skilled worker, keeping their coworkers and their firm unchanged, has to equal the increase in wages necessary to hire the higher-skill worker. More skilled workers work with coworkers and a firm that has greater expected output losses, implying a convex wage schedule.

Our main result is derived in a setting with heterogeneous distributions of workers and firms and teams of $n \geq 2$ members. We prove that a mixed and countermonotonic assignment is optimal, derive its dual, and thus characterize an equilibrium. We establish the optimality of a mixed and countermonotonic assignment by using a majorization inequality (Hardy, Littlewood, and Pólya, 1929; Karamata, 1932; Pečarić, 1984).² We combine our characterization of optimal assignments

²We prove there exists a mixed set by connecting our assignment model to work on risk aggregation by Wang

with a duality argument for multimarginal optimal transport problems (Kellerer, 1984) to derive worker wages and firm values. Finally, we discuss how our results apply to more general production structures.

Having characterized equilibrium, we quantitatively illustrate the theory using administrative U.S. earnings records. Since a characteristic feature of our paper 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 U.S. firms.

Our model can generate the dispersion of earnings as well as its decomposition between and within firms as empirically documented by Song, Price, Guvenen, Bloom, and Von Wachter (2019). We use the cross-sectional distribution of earnings and the decomposition of earnings variation into within-firm and between-firm variation to estimate the underlying distributions of worker skills and firm values. We show that only changing the distribution of firm projects between 1981 and 2013, the share of within-firm earnings dispersion would have increased by 22 percentage points. Changing the distribution of workers, we show that the share of within-firm earnings dispersion would have decreased by 38 percentage points. That is, our counterfactual analysis shows that both the changes in the worker and project distributions are important in the model in generating the observed change in earnings dispersion.

Related Literature. There is an extensive literature on one-to-one assignment models in the tradition of Becker (1973). This literature, in particular, analyzes conditions under which sorting is positive or negative.³ Our contribution is to characterize sorting for an 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.

The closest to our work is a 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)

and Wang (2016) that describes when there exists a dependence structure such that the sum of idiosyncratic risks is constant.

³Boerma, Tsyvinski, Wang, and Zhang (2023) provides a closed-form solution to a one-to-one assignment that is neither positive sorting nor negative sorting, which they call composite sorting.

also develops a quantitative application to group formation in villages.⁴ In our paper, we fully characterize optimal sorting for dimensions greater than two with a submodular technology.

An alternative approach to multiworker firms in assignment models is due to [Eeckhout and Kircher \(2018\)](#).⁵ 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 and by incorporating peer effects. Workers are instead imperfect substitutes under a submodular technology, and their marginal product does depend on their coworkers.⁶

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 [Dupuy and Galichon \(2014\)](#), [McCann, Shi, Siow, and Wolthoff \(2015\)](#), [Lindenlaub \(2017\)](#), [Chiappori, McCann, and Pass \(2017\)](#), [Chiappori, Salanié, and Weiss \(2017\)](#), [Ocampo Díaz \(2022\)](#), [Lindenlaub and Postel-Vinay \(2020\)](#) and [Galichon and Salanié \(2021\)](#). 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.⁷ 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 problem 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

⁴See 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 intra-firm spillovers.

⁵[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.

⁶Recent quantitative work by [Jarosch, Oberfield, and Rossi-Hansberg \(2021\)](#) and [Herkenhoff, Lise, Menzio, and Phillips \(2023\)](#) studies production and learning of workers in teams in labor markets with and without frictions.

⁷An exposition of the hierarchical assignment problem is presented in [Garicano and Rossi-Hansberg \(2004\)](#). See [Garicano and Rossi-Hansberg \(2015\)](#) for a review.

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 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 transport problems.⁸ 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 \(2010\)](#), [Kim and Pass \(2014\)](#), 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 to them, 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 the distribution of output losses and characterizing sorting patterns of workers that generate maximum aggregate output. Moreover, we characterize the dual problem to characterize the distribution of wages and firm values.

2 Model

We study an economy in which agents with heterogeneous skills choose their firm and coworkers to work with. This is an assignment problem extended to incorporate multiple workers in each firm. Our setup results in a sorting problem between multiple workers with each firm for a submodular production technology.

2.1 Environment

Agents. There are $m \geq 2$ groups of risk-neutral workers and a single group of risk-neutral firms. Each group has mass one.

Workers differ in skills. The skill 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. For example, a worker

⁸See, 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.

with skill $x_i = 0.1$ knows how to solve a problem with 90 percent probability. High skill workers have low x_i . The distribution of each group of workers has cumulative distribution function $F_x(x_i)$ and corresponding continuous density function $f_x(x_i)$.

Firms differ in productivity $z \in Z = [0, 1]$, capturing the value of their projects. High value firms have a high z . The distribution of firms has cumulative distribution function $F_z(z)$ and continuous density function $f_z(z)$. The cumulative distribution functions are strictly increasing and continuous. The inverse distribution function for firms is so that $I_z(p_z)$ for percentile $p_z \in [0, 1]$ is the unique number z that satisfies $F_z(z) = p_z$. Thus, the percentile p_z corresponds to the value z in the firm distribution. The inverse distribution function for workers I_x is defined analogously.

Technology. A team comprises of m workers that encounter a problem in production. These workers solve a problem when at least one of them knows how to solve it. Since the probability that worker i does not know the solution is x_i , independent of their coworker knowing the solution, the probability of successful production with workers (x_1, x_2, \dots, x_m) is:

$$q = h(x_1, x_2, \dots, x_m) = 1 - \prod_{i=1}^m x_i. \quad (1)$$

The team quality function h is a symmetric, submodular function in worker types, so workers are substitutes.

A firm of type z that employs a team of quality q produces output according to:

$$g(q, z) = qz. \quad (2)$$

The production function g is supermodular in team quality q and firm type z . Team quality and the project are complements in production. Combining the technology (2) with the team quality function (1), firm output y is:

$$y(x_1, x_2, \dots, x_m, z) = \left(1 - \prod_{i=1}^m x_i\right)z. \quad (3)$$

Firm output is a submodular function in worker and firm type.⁹ The marginal product of worker x_i is:

$$m(x_i) = y_i(x_1, x_2, \dots, x_m, z) = -z \prod_{j \neq i} x_j \leq 0. \quad (4)$$

⁹This technology nests 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. High quality teams work on the most valuable projects. On the other hand, absent firm heterogeneity, $z = \bar{z}$, sorting is negative in the case of two workers, or $m = 2$. Low and high skill workers form teams.

When a worker is less skilled, that is, less likely to know how to solve a problem, output decreases. The marginal product of every worker depends only on the skills of their coworker and the firm they work with.

Assignment. An assignment prescribes for every worker coworkers to work with and a firm to work for. Given a distribution of workers F_x and a distribution of firms F_z , the set of feasible assignment functions is $\Pi := \Pi(F_x, \dots, F_x, F_z)$ which is the set of probability measures π on $X \times \dots \times X \times Z$ 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 equivalent to labor market clearing, that is, all workers and firms are sorted.

2.2 Assignment Problem

We solve two problems to characterize an equilibrium.¹⁰

Primal Problem. We first solve a primal problem to find an optimal sorting:

$$\max_{\pi \in \Pi} \int y(x_1, x_2, \dots, x_m, z) d\pi. \quad (5)$$

This problem is to choose an assignment π to maximize production. It is equivalent to, in terms of choosing an optimal assignment, minimizing output losses. Output losses are the product of the project value z and the probability of failure by a team of workers which is given by $x_1 \cdots x_m$, that is, a loss $x_1 \cdots x_m z$. We thus equivalently represent the planning problem as:

$$\min_{\pi \in \Pi} \int x_1 \cdots x_m z d\pi, \quad (6)$$

which is an optimal transport problem in the tradition of [Monge \(1781\)](#) and [Kantorovich \(1942\)](#), with multiple marginal distributions. The number of marginal distributions is given by $n = m + 1$, comprising m workers in a team and a project.

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

$$\min \sum_{i=1}^m \int w(x_i) dF_x + \int v(z) dF_z, \quad (7)$$

subject to the constraint that $\sum w(x_i) + v(z) \geq y(x_1, x_2, \dots, x_m, z)$ for any $(x_1, x_2, \dots, x_m, z)$.

¹⁰The equilibrium definition is standard and is presented in [Appendix A.1](#) for completeness.

3 Intuition for the Main Result

This section provides intuition for optimal sorting. We consider the case where all distributions are identical.

The core idea of the optimal sorting is based on three observations. First, Jensen's inequality bounds aggregate output losses from below:¹¹

$$\int x_1 \cdots x_m z d\pi \geq \exp \left(\int \log x_1 \cdots x_m z d\pi \right). \quad (8)$$

Second, the lower bound is independent of assignment π due to feasibility:

$$\exp \left(\int \log x_1 \cdots x_m z d\pi \right) = \exp \left(\int \log x_1 dF + \dots + \int \log x_m dF + \int \log z dF \right). \quad (9)$$

The right hand side is a constant \mathcal{C} as $\int \log z dF$ and $\int \log x_i dF$ depend only on the marginal distributions. Combining (8) and (9), the aggregate output loss is bounded below by \mathcal{C} :

$$\int x_1 \cdots x_m z d\pi \geq \mathcal{C}. \quad (10)$$

Third, by Jensen's inequality the minimum given by the right-hand side is attained when the output loss is equal across all teams and equal to \mathcal{C} . We refer to the set of teams with the same output loss \mathcal{C} as mixed.¹² When an assignment is mixed, it is optimal as it minimizes aggregate losses (6).

While mixing is optimal when it is feasible, it is not feasible everywhere. Consider a project $z = 0$ in a mixed assignment. Then the output loss $x_1 x_2 \dots x_m z$ has to be equal to zero for all teams. Thus, mixed is not feasible everywhere.¹³

We show that the optimal assignment is the largest possible set of mixed teams combined with countermonotonic sets.¹⁴ The intuition for this result is as follows.

Jensen's inequality suggests that it is optimal to have the largest set of mixed teams. Consider a mixed set with the workers and firms between percentiles \underline{p} and \bar{p} in their distribution, that is, the values $[I(\underline{p}), I(\bar{p})]$ from each distribution. We now show that $\mathcal{C} = I(\underline{p})I^m(\bar{p})$ for the largest

¹¹For any convex function ζ and any random variable X , Jensen's inequality states that $E[\zeta(X)] \geq \zeta(E[X])$. Here, ζ is the exponential function, and the random variable $X = \log x_1 \cdots x_m z$ is the logarithmic output loss.

¹²Wang and Wang (2011) define assignments with equal output loss as completely mixed. Such sortings are also studied by Gaffke and Rüschemdorf (1981), Rüschemdorf and Uckelmann (2002), and Knott and Smith (2006).

¹³Generally, there exists no mixed assignment for atomless distributions with full support on the unit interval.

¹⁴The teams $(x_1, x_2, \dots, x_m, z)$ and $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m, \hat{z})$ are countermonotonic in k if $x_k < \hat{x}_k$ implies that $x_i \geq \hat{x}_i$ for all $i \neq k$ and $z \geq \hat{z}$, or if $x_k > \hat{x}_k$ implies that $x_i \leq \hat{x}_i$ for all $i \neq k$ and $z \leq \hat{z}$, or if $\hat{x}_k = x_k$. A set of teams is countermonotonic in k if all pairs of teams are countermonotonic in k .

possible set of mixed teams. The maximal output loss on a low-value project $z = I(\underline{p})$ is attained by pairing this project with the lowest-skill workers $x_i = I(\bar{p})$ for all i . This team incurs an output loss $I(\underline{p})I^m(\bar{p})$ and thus $\mathcal{C} \leq I(\underline{p})I^m(\bar{p})$. For the maximal feasible mixed set, this condition has to hold with equality $\mathcal{C} = I(\underline{p})I^m(\bar{p})$.

We next describe the reasoning behind the countermonotonic sets. First, note that the lowest value project $z = I(\underline{p})$ in the mixed set is paired with m identical low skill workers $I(\bar{p})$ in this set to attain the loss $\mathcal{C} = I(\underline{p})I^m(\bar{p})$. Second, it is naturally optimal to pair the lowest value project $z = 0$ with m lowest skill workers $x_i = 1$, attaining zero loss. Interpolating this logic between the project values $z = 0$ and $z = I(\underline{p})$, projects below $I(\underline{p})$ are sorted countermonotonically to m identical low skill workers. This generates a countermonotonic set for low value projects $z \in [0, I(\underline{p}))$ with m identical low skill workers $x_i \in (I(\bar{p}), 1]$. By symmetry, high skill workers $x_j \in [0, I(\underline{p}))$ are paired with $m - 1$ low skill coworkers $x_i \in (I(\bar{p}), 1]$ and a high value project $z \in (I(\bar{p}), 1]$. Since both low value projects z and high skill workers x_j for all $j \neq i$ are assigned to teams with low skill workers x_i , percentiles $[0, p)$ of firms z and high-skill workers x_j absorb measure mp of low-skill coworker x_i . Percentile $p \in [0, \underline{p})$ of the project distribution is thus sorted with percentile $1 - mp$ of the worker distributions. The least valuable project $p = 0$ employs the least productive workers at percentile 1; a project at percentile \underline{p} is sorted with identical workers ranked $1 - m\underline{p} = \bar{p}$. This team incurs loss $I^m(\bar{p})I(\underline{p}) = \mathcal{C}$, identical to the loss for all teams on the mixed set.

In sum, we provided intuition for the following solution. Pair high skill workers with low skill workers on valuable projects, until, at some point, mediocre workers can be paired with firms such that output losses are identical for all mediocre teams. Figure 1 visualizes the support of this candidate solution with two workers. The countermonotonic sets corresponding to teams with high skill workers x_1 and x_2 is represented by A_1B_1 and A_2B_2 , while the countermonotonic set that corresponds to teams with low value projects z is A_3B_3 . The mixed set is represented by $B_1B_2B_3$.

4 Main Results

We next show that a mixed and countermonotonic assignment solves the planning problem with heterogeneous continuous marginal distributions and characterize the dual solution for wages and firm values.

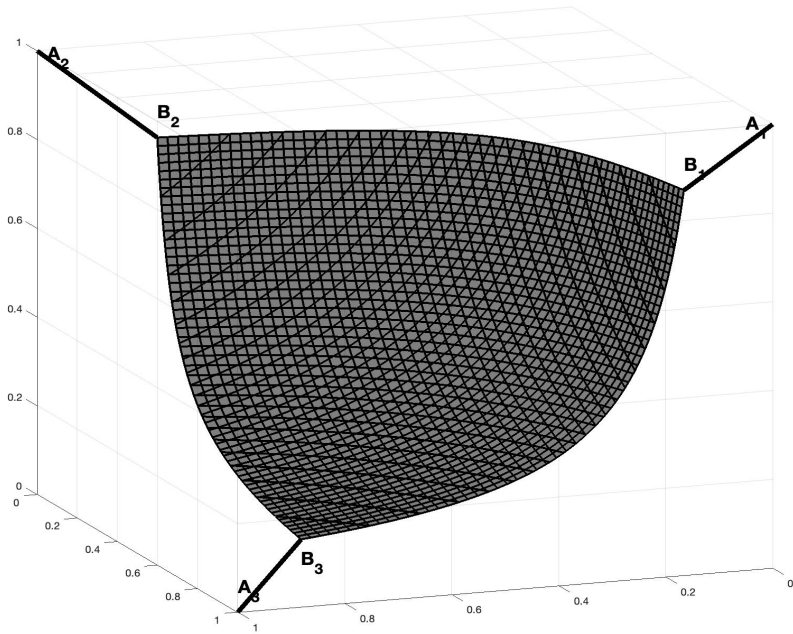


Figure 1: Support of the Assignment

Figure 1 plots an example for the support for an optimal assignment when each team employs two workers, $m = 2$. The x -axis and y -axis show the skill of the worker x_1 and their coworker x_2 , while the z -axis plots the value of the firm project z .

4.1 Mixed and Countermonotonic Assignment

We generalize the assignment in Section 3 to heterogeneous marginal distributions by developing three insights from the intuition for the case of homogenous marginal distributions.

First, in the case of homogenous distributions, each distribution is segmented into three parts: low values $[0, I(\underline{p})]$, medium values $[I(\underline{p}), I(\bar{p})]$, and high values $[I(\bar{p}), 1]$. The measure of medium values is $\bar{p} - \underline{p} = 1 - q$. With heterogeneous distributions, we segment all distributions into three parts, each indexed by $i \in \{1, \dots, n\}$: low values $[0, I_i(\underline{p}_i)]$, medium values $[I_i(\underline{p}_i), I_i(\bar{p}_i)]$, and high values $[I_i(\bar{p}_i), 1]$.¹⁵ The measure of medium values is $\bar{p}_i - \underline{p}_i = 1 - q$, identical across distributions.

Second, we consider the ranking of output losses across teams. In the homogenous case, the loss is largest for teams in the mixed set which has measure $1 - q$. The remaining teams are in countermonotonic sets with measure q . For each percentile $t \leq q$, there are n teams located at the t -th percentile of the loss distribution, one in each countermonotonic set. Each team consists of a low value $I(\frac{t}{n})$ from one of the distributions and m high values $I(1 - \frac{m}{n}t)$ from the remaining distributions. With heterogeneous distributions, the loss is as well largest for teams in the mixed

¹⁵When we use $i \in \{1, \dots, n\}$ to enumerate the different marginal distributions, we use $1 \leq i \leq m$ to refer to the respective worker distributions, and n to reflect the firm distribution.

set which has measure $1 - q$. There are also n teams located at every percentile $t \leq q$. Each team consists of a low value $I_i(\underline{t}_i(t))$ from one of the distributions and m high values $I_j(1 - \bar{t}_j(t))$ from the remaining distributions $j \neq i$, where the functions \underline{t}_i and \bar{t}_i are both increasing. The function $\underline{t}_i(t)$ gives the percentile of the low value drawn from distribution i at the t -th percentile of the team loss distribution. Similarly, the function $\bar{t}_j(t)$ yields the percentile of the high value drawn from distribution j at the t -th percentile of the team loss distribution. In the homogenous case, $\underline{t}_i(t) = \frac{t}{n}$ and $\bar{t}_j(t) = \frac{m}{n}t$.

Finally, in the case of homogeneous distributions, mixed teams contain agents in $[I(\underline{p}), I(\bar{p})]$, and their losses are all identical and equal to $I(\underline{p})I(\bar{p})^m$. With heterogeneous distributions, mixed teams contain values in $[I_i(\underline{p}_i), I(\bar{p}_i)]$ and attain identical losses $\mathcal{C} = \prod x_i$, where for all i :

$$\mathcal{C} = I_i(\underline{p}_i) \prod_{j \neq i} I_j(\bar{p}_j). \quad (11)$$

We use the generalization of features of the optimal sorting with homogeneous distributions to define a mixed and countermonotonic assignment for the case of heterogeneous distributions.

Definition 1. A mixed and countermonotonic assignment is a collection of continuous increasing functions $\{(\underline{t}_i, \bar{t}_i)\}_{i=1}^n$, thresholds $\{(\underline{p}_i, \bar{p}_i)\}_{i=1}^n$, and an assignment $\pi \in \Pi(\{F_i\}_{i=1}^n)$ satisfying:

1. The mass of agents in the mixed sets $[I_i(\underline{p}_i), I(\bar{p}_i)]$ is identical for all distributions $1 - q = \bar{p}_i - \underline{p}_i$ for all i .
2. The support of assignment π consists of (1) a mixed assignment; and (2) n countermonotonic sets so that for all $0 \leq t \leq q$ the assignment is countermonotonic:
 $b_i(t) = (b_{i1}(t), \dots, b_{in}(t))$ with $b_{ii}(t) = I_i(\underline{t}_i(t))$ and $b_{ij}(t) = I_j(1 - \bar{t}_j(t))$ if $i \neq j$.
3. For each i , $b_i(t)$ starts at $(1, \dots, 1, 0, 1, \dots, 1)$ and ends at $(I_1(\bar{p}_1), \dots, I_i(\underline{p}_i), \dots, I_n(\bar{p}_n))$. Equivalently, this means that $\underline{t}_i(0) = \bar{t}_i(0) = 0$, $\underline{t}_i(q) = \underline{p}_i$, and $1 - \bar{t}_i(q) = \bar{p}_i$ for each i .
4. For each percentile $t \leq q$, there are n teams located at the t -th percentile, when teams are ranked according to their output losses. Team i is located on the countermonotonic set at $b_i(t)$. All other teams are mixed and have identically large losses equal to \mathcal{C} given by (11).

4.2 Planning Problem

In this section, we prove our main result that a mixed and countermonotonic assignment, following Definition 1, is optimal.

Theorem 1. Any mixed and countermonotonic assignment $\pi \in \Pi$ solves the planning problem (5).

We provide a sketch of the proof. Let γ be the distribution of logarithmic output losses induced by assignment π . The associated inverse cumulative distribution function is I_γ . That is, logarithmic losses by a team at percentile p in the loss ranking are given by $I_\gamma(p)$ when teams are sorted in increasing loss order. We show that an assignment π^* is optimal by showing that output losses are greater under any other assignment π . That is, $\int \ell(x_1, x_2, \dots, x_n, z) d\pi^* \leq \int \ell(x_1, x_2, \dots, x_m, z) d\pi$ or, equivalently, $\int_0^1 \exp(I_{\gamma^*}(p)) dp \leq \int_0^1 \exp(I_\gamma(p)) dp$.¹⁶

Specifically, we establish optimality by proving that logarithmic output losses under the optimal assignment are smaller in convex order than the logarithmic output losses under any other feasible assignment.¹⁷ The proof is constructed using the integral version of the majorization inequality (Hardy, Littlewood, and Pólya, 1929; Karamata, 1932; Pečarić, 1984).

Majorization Inequality. Let h be a continuous convex function, and let s_1 and s_2 be non-decreasing functions on the unit interval. Suppose

$$1. \int_0^1 s_1(p) dp = \int_0^1 s_2(p) dp \tag{12}$$

$$2. \int_0^t s_1(p) dp \geq \int_0^t s_2(p) dp, \tag{13}$$

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

We use the majorization inequality with $h = \exp$, $s_1(p) = I_{\gamma^*}(p)$ and $s_2(p) = I_\gamma(p)$. The exponential function is convex, and logarithmic losses I_γ are increasing as the logarithm of a positive increasing function is itself increasing.

For application of the majorization inequality, it suffices to verify that (12) and (13) are satisfied. The first condition (12) requires that aggregate logarithmic losses under the assignments are equal. As aggregate logarithmic losses are invariant across feasible assignments this condition is verified. That is, for any feasible assignment π : $\int_0^1 \log I_\gamma(p) dp = \int \log \ell(x_1, x_2, \dots, x_m, z) d\pi = \sum_{i=1}^m \int \log x_i dF_i + \int \log z dF_z$.

¹⁶It is equivalent to integrate output losses with respect to the assignment function and with respect to ranks in the logarithmic loss distribution, $\int \ell(x_1, x_2, \dots, x_m, z) d\pi = \int_0^1 \exp(I_\gamma(p)) dp$.

¹⁷A random variable A is less in convex order than a random variable B if and only if for any convex function h : $E[h(A)] \leq E[h(B)]$. This is equivalent to $\int_0^1 h(I_A(p)) dp \leq \int_0^1 h(I_B(p)) dp$ given that I is the quantile function.

The second condition (13) is that cumulative logarithmic losses at any percentile in the team rank t are larger under optimal assignment π^* . We observe that cumulative logarithmic losses are convex in team rank t as they integrate the increasing negative function $\log I_\gamma$. For the mixed set, that is teams ranked $t \in [q, 1]$, cumulative logarithmic losses decrease linearly as losses are constant. Since cumulative logarithmic losses for the lowest ranked teams $t \in [0, q)$ are maximized by the countermonotonic assignments (Lemma 2 in Appendix A.2), and since aggregate logarithmic losses are equal, cumulative logarithmic losses are larger for any team rank $t \in [0, 1]$. We formalize the verification of condition (13) in Appendix A.2.¹⁸

General Production Functions. We next discuss how the mixed and countermonotonic sorting is also an optimal assignment with more general production technologies.

Consider a general technology $y(x_1, \dots, x_m, z)$. Its second-order Taylor expansion is a second-order polynomial in worker skills $\{x_1, \dots, x_m\}$ and project value z . Assume that the matrix of the coefficients in this expansion has rank one and is elementwise non-positive to ensure submodularity. In Appendix A.5, we establish that a mixed and countermonotonic assignment is optimal. This shows that our results generalize in the second order sense to general production functions.

In the case of teams comprising of two workers and a project, we derive a complementary generalization. Let a technology function be given by $y(x_1, x_2, z, x_1x_2, x_2z, x_1z, x_1x_2z)$ which additionally incorporates interactions effects between workers and coworkers x_1x_2 , and workers and project x_1z and x_2z . Suppose its first-order Taylor expansion around the function's arguments has negative coefficients on each interaction term $(x_1x_2, x_2z, x_1z, x_1x_2z)$ to ensure submodularity. In Appendix A.5, through a change of variables, we establish optimality of a mixed and countermonotonic assignment. This shows that our results extend in the first order sense to these functions of single, pairwise, and triple interactions.

4.3 Dual Problem

In this section, we develop the analysis of the dual problem to characterize wages and firm values.

Marginal Worker Product. The marginal worker product (4) is the marginal output loss induced by a worker, which is the negative product of the project value and the coworker skills.

¹⁸Theorem 1 does not imply uniqueness of an optimal assignment, only that any mixed and countermonotonic assignment solves the planning problem. We provide a further characterization of the mixed and countermonotonic assignment in Appendix A.3, and prove when it exists in Appendix A.4. In Section 4.3, we characterize equilibrium wages and firm values, which are uniquely determined.

Given an optimal assignment, the marginal worker product can be fully described.

First, consider a high-skill worker $x_i \in [0, I_i(\bar{p}_i)]$. Given team rank t , the high-skill worker is $x_i = I_i(\underline{t}_i(t))$, and their team is in a countermonotonic set. This worker is assigned to low-skill team members, given by $I_j(1 - \bar{t}_j(t))$ on a valuable project $I_n(1 - \bar{t}_n(t))$. The marginal product for high-skill workers is $m_i(x_i) = - \prod_{j \neq i} I_j(1 - \bar{t}_j(t))$.

Second, consider mediocre workers $x_i \in [I_i(\underline{p}_i), I_i(\bar{p}_i)]$. Mediocre workers work in mixed teams with identical output losses $\ell(x_1, \dots, x_m, z) = x_1 \dots x_m z = \mathcal{C}$ across all such teams. Given this constant output loss, the marginal product for mediocre workers is $m_i(x_i) = -\mathcal{C}/x_i$.

Finally, consider low-skill workers $x_i \in [I_i(\bar{p}_i), 1]$. The rationale for their marginal product is a direct consequence of the assignment of high-skill workers described above. There are exactly m teams containing this low-skill worker with the same loss. Let k be the index of the high-skill worker in one such team, so that $I_k(\underline{t}_k(t))$ is the high-skill worker. The value of team member j is given by $I_j(1 - \bar{t}_j(t))$ for all $j \neq k$. Thus, the marginal product of the worker is: $-I_k(\underline{t}_k(t)) \prod_{j \neq i, k} I_j(1 - \bar{t}_j(t))$. The description of the marginal product is summarized by Proposition 2.

Proposition 2. *Marginal Product.* The marginal product of worker $x_i = I_i(p_i)$ is:

$$m_i(x_i) = \begin{cases} - \prod_{j \neq i} I_j(1 - \bar{t}_j(t)) & \text{if } x_i \in [0, I_i(\bar{p}_i)] \\ -\mathcal{C}/x_i & \text{if } x_i \in [I_i(\underline{p}_i), I_i(\bar{p}_i)] \\ - I_k(\underline{t}_k(t)) \prod_{j \neq i, k} I_j(1 - \bar{t}_j(t)) & \text{if } x_i \in [I_i(\bar{p}_i), 1] \end{cases} \quad (14)$$

where the constant \mathcal{C} is given by (11).

Proposition 3 establishes properties of the marginal worker product.

Proposition 3. *Properties of the Marginal Product.* The marginal product $\{m_i\}$ is continuous and increasing.

We now give an outline of the proof to this proposition. First, continuity within all three segments directly follows as the distribution functions $\{F_i\}$, their inverse functions $\{I_i\}$, and $\{\underline{t}_i, \bar{t}_i\}$ are all continuous. In addition, at the point $x_i = I_i(\underline{p}_i)$ the values of the high-skill and mediocre workers align; at the point $x_i = I_i(\bar{p}_i)$ the values of the low-skill workers and the mediocre workers align. Hence, the marginal product $\{m_i\}$ is continuous across all skill levels.¹⁹

¹⁹The marginal worker product m_i is identical for all worker distributions, or $m_i = m$ as we show in Appendix A.4.

Second, the most insightful part of this proposition is that the marginal product is increasing. This fact follows from the necessary condition that total losses do not decrease by exchanging workers between their teams. For any two teams (x_1, \dots, x_m, z) and $(\hat{x}_1, \dots, \hat{x}_m, \hat{z})$ in the optimal assignment $\ell(x_1, \dots, x_m, z) + \ell(\hat{x}_1, \dots, \hat{x}_m, \hat{z}) \leq \ell(x_1, \dots, \hat{x}_i, \dots, x_m, z) + \ell(\hat{x}_1, \dots, x_i, \dots, \hat{x}_m, \hat{z})$. Given the loss function $\ell(x_1, \dots, x_m, z) = x_1 \dots x_m z$ and the marginal worker product (4):

$$(x_i - \hat{x}_i) \left(z \prod_{j \neq i} x_j - \hat{z} \prod_{j \neq i} \hat{x}_j \right) \leq 0 \quad \iff \quad (x_i - \hat{x}_i) (m_i(\hat{x}_i) - m_i(x_i)) \leq 0. \quad (15)$$

When a worker is more skilled, $\hat{x}_i < x_i$, (15) implies that they work with coworkers on a project that has a greater output loss, $\hat{z} \prod_{j \neq i} \hat{x}_j \geq z \prod_{j \neq i} x_j$. Equivalently, when a worker is more skilled, their marginal product (4) is more negative, and the marginal product is thus increasing.

Wages. Given the marginal product we formulate equilibrium wages. To simplify notation, define the surplus S as output minus payments to workers and firms:

$$S(x_1, \dots, x_m, z) = y(x_1, \dots, x_m, z) - \sum_{i=1}^m w(x_i) - v(z). \quad (16)$$

The constraint to the dual problem is that the surplus $S(x_1, \dots, x_m, z)$ is negative for any team $(x_1, \dots, x_m, z) \in X \times \dots \times X \times Z$. In Appendix A.6 we show that the dual solution exists in the class of continuous functions, hence the surplus function (16) is continuous also.

Using the duality result of Kellerer (1984), the optimal value to the planner problem equals the optimal value for the dual problem, $\int y(x_1, \dots, x_m, z) d\pi = \sum \int w(x_i) dF_x + \int v(z) dF_z$, where π solves the planning problem and w and v solve the dual problem. Given feasibility and the definition of the surplus, $\int S(x_1, \dots, x_m, z) d\pi = 0$. Since the surplus is non-positive for every team by the constraint to the dual problem (7), the surplus is zero almost everywhere with respect to the assignment π .

We next characterize the derivative of the wage schedule and the firm value function.

Proposition 4. Marginal Wages and Firm Values. Wages and firm values are both continuously differentiable, with derivatives $w'(x) = m(x)$ and $v'(z) = m_z(z) + 1$.

The proof is presented in Appendix A.7. Workers' marginal earnings equal their marginal product. The marginal increase in output must equal the marginal increase in the cost necessary to recruit a higher skill worker. While the marginal product depends only on peer effects, the equilibrium level of these peer effects is directly related to the worker's own skill through Proposition 2.

While a worker's marginal earnings reflect their marginal product, earnings levels reflect the marginal product of all workers that are more skilled. Specifically, workers incur an earnings penalty relative to the most skilled worker. Wages are characterized in Proposition 5.

Proposition 5. *Wages.* The wage schedule is, up to additive worker constant \mathcal{C}_w , given by:

$$w(x) = \mathcal{C}_w + \int_0^x m(s) ds. \quad (17)$$

To obtain intuition, observe the surplus is zero almost everywhere with respect to the assignment π , $S(x_1, \dots, x_m, z) = z(1 - \prod x_i) - \sum w(x_i) - v(z) = 0$. Moreover, since w is a differentiable function, fix the coworkers x_{-i} and the project z . Since the surplus function is negative $S(x_1, \dots, x_m, z) \leq 0$, the equilibrium team (x_1, \dots, x_m, z) is a local maximum with respect to the surplus function, and hence:

$$S_i(x_1, \dots, x_m, z) = 0 = -z \prod_{j \neq i} x_j - w'(x_i) \quad \implies \quad w'(x_i) = m(x_i), \quad (18)$$

where the final equality follows from the definition 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 (17).

The marginal cost of workers is equal to their marginal product $w'(x) = m(x)$ for all $x \in (0, 1)$. By Proposition 3, the negative marginal worker product is continuous and increasing, establishing that equilibrium wages are decreasing and convex.

Firm. We next characterize firm values. The intuition for the firm value is similar to the intuition for wages. If there are no excess resources almost everywhere for assignment π , then, by fixing a team of workers as well as by differentiability of the firm value v :

$$S_n(x_1, \dots, x_m, z) = 0 = 1 - \prod_j x_j - v'(z) \quad \implies \quad v'(z) = m_z(z) + 1, \quad (19)$$

where the final equality follows from the definition of the marginal product (4). A firm's marginal reward is its marginal product. When a firm undertakes a marginally more valuable project, its output is its expected probability of success, or team quality, $1 - \prod_j x_j$. In equilibrium, the firm's marginal product is similar to the worker product, up to the unit constant. Firm values are characterized in Proposition 6.

Proposition 6. *Firm Value.* The firm value v is, up to additive firm constant \mathcal{C}_v , given by:

$$v(z) = \mathcal{C}_v + \int_0^z k(s) ds, \tag{20}$$

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

By Proposition 6, the firm value shares properties with the wage function. By Proposition 3, the marginal firm product k is continuous and increasing. The marginal worker product m is continuous and increasing, and $m(z) \geq -1$ implying $v'(z) \geq 0$. The firm value function is thus increasing and convex in project value z .

Given the description of wages in Proposition 5 and the firm value function in Proposition 6, we observe that for any constants \mathcal{C}_w and \mathcal{C}_v that satisfy:

$$0 = \mathcal{C}_w + \mathcal{C}_v/m + \int_0^1 m_x(s) ds \tag{21}$$

the functions (w, v) are a dual solution. Equation (21) ensures the surplus is zero for the team $(x_1, \dots, x_m, z) = (1, \dots, 1, 0)$.

Summary. In Section 4.2 we show that the mixed and countermonotonic assignment solves the assignment problem. In Section 4.3 we construct the solution to the corresponding dual problem. Since the surplus is zero everywhere in the support of an optimal assignment, the value for the planning problem and the dual problem coincide. In Appendix A.1 we show that any mixed and countermonotonic assignment π together with dual functions (w, v) satisfying (21) are therefore an equilibrium.²⁰

5 Quantitative Analysis

We evaluate the ability of the model to quantitatively generate the observed wage dispersion, as well as its decomposition into within and between-firm components.

5.1 Data

We use data on worker earnings for 1981 and 2013 from the US Social Security Administration of Song, Price, Guvenen, Bloom, and Von Wachter (2019). The data considers employed individuals

²⁰Equation (21) 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 (21).

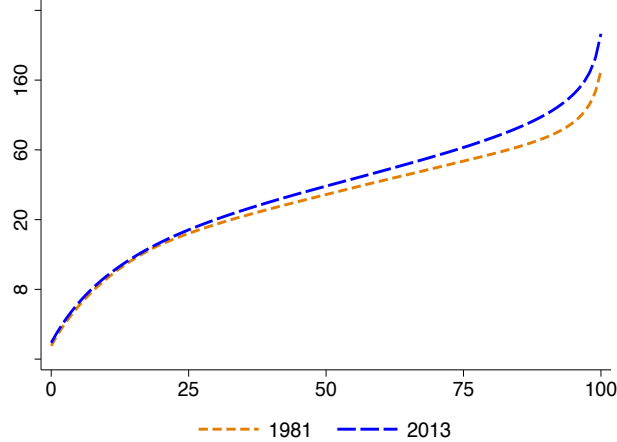


Figure 2: Distribution of Individual Earnings

Figure 2 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 20 and 60 years of age.²¹ The main reason for using these records is that the data matches the universe of individuals to their employer which allows us to analyze sorting patterns. Importantly, the matched employer-employee dataset provides information on the earnings distribution within every firm.

Figure 2 shows individual earnings by percentile of the earnings distribution for 1981 and 2013. In 1981, earnings vary between 18 thousand at the 25th percentile, to 31 thousand at the median, to 89 thousand at the 95th percentile. By 2013, earnings are 19 thousand at the 25th percentile, 35 thousand at the median, and 132 thousand at the 95th percentile. Figure 2 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 decompose the overall variance of log earnings into within- and between-firm components. Let w_{ij} be log earnings of worker i at firm j . Earnings can be written as $w_{ij} = \bar{w}_j + (w_{ij} - \bar{w}_j)$,

²¹Individual earnings are derived from W-2 forms, which capture compensation for labor services as defined by the Internal Revenue Service. An individual is considered employed when their earnings exceed the minimum wage earned full-time for one quarter, that is, for 520 hours. 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. Workers with multiple jobs in a year are linked to the firm that provides their largest share of labor earnings. All amounts are in 2013 dollars.

where \bar{w}_j is average log earnings within firm j . The total variance is then:

$$\text{Var}(w_{ij}) = \text{Var}(\bar{w}_j) + \sum_j \theta_j \text{Var}(w_{ij}|i \in j), \quad (22)$$

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

The 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 earnings dispersion as documented by [Song, Price, Guvenen, Bloom, and Von Wachter \(2019\)](#). From 1981 to 2013 the variance of log earnings increased by 0.20 log points to 0.85. Two-thirds of this increase is attributed to increased differences in mean earnings across firms, while a third is attributed to an increase in within-firm earnings dispersion.

5.2 Model

We assess the ability of our model to generate dispersion in earnings, its decomposition 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 earnings dispersion between changes on the worker side and changes on the firm side of the labor market.

In our model, the worker skill distribution F_x and the project value distribution F_z are both exogenous. We parameterize the distribution for workers as $\text{Beta}(\alpha_x, \beta_x)$ and the distribution for projects as $\text{Beta}(\alpha_z, \beta_z)$. We consider teams with two workers.

We use the cross-sectional distribution of earnings and the decomposition of the earnings variation into within-firm and between-firm variation to inform the worker and project distribution. We estimate the underlying distributions using earnings in the cross-section using the data underlying [Figure 2](#). The resulting model fit for cross-sectional earnings is summarized by [Figure 3](#). [Figure 3](#) shows that our framework generates the observed dispersion in earnings, as well as changes in the distribution of earnings over time while somewhat underestimating earnings at the very top percentiles in both 1981 and 2013.²²

[Table 1](#) displays the decomposition for the variance of log earnings between 1981 and 2013 in the data and the model. Our stylized model can account for the earnings decomposition in

²²The parameters which govern the Beta distribution for worker skills are (0.96, 2.85) in 1981 and (1.89, 1.48) in 2013. The estimated parameters for the project value distribution are (0.98, 2.83) in 1981 and (1.66, 1.63) in 2013.

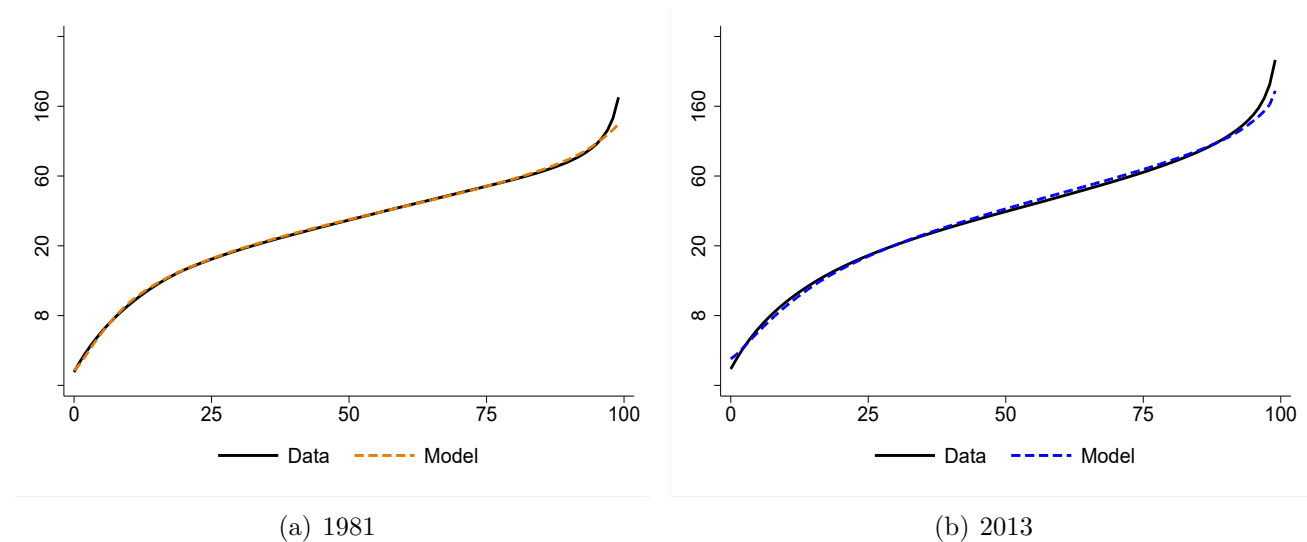


Figure 3: Data and Model Distribution of Individual Earnings

Figure 3 compares the empirical earnings distribution to the model earnings distribution. The empirical distributions, which we display by solid black lines, follow Figure 2, 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

Moment	Data			Model		
	1981	2013	change	1981	2013	change
Between	0.34	0.42	0.08	0.34	0.43	0.09
Within	0.66	0.58	-0.08	0.66	0.57	-0.09

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 (22), while the right panel shows the model analog.

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. The model generates an increase in the share of the between-firm variance from 0.34 to 0.43 percentage points between 1981 and 2013, while capturing the overall increase in earnings dispersion as shown in Figure 3.²³

In addition to studying the variance decomposition of log earnings, we evaluate non-targeted average log earnings of coworkers 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

²³Gavilan (2012) studies the change in between and within firm earnings dispersion in response to the change in the price of capital and in response to the changes in the skill distribution in the setup of Kremer and Maskin (1996) with endogenous capital.

Table 2: Model and Data Coworker Earnings

Percentile	Data			Model		
	1981	2013	change	1981	2013	change
25	10.01	10.07	0.06	10.19	10.30	0.11
50	10.30	10.45	0.15	10.44	10.63	0.19
75	10.56	10.75	0.19	10.56	10.83	0.27
90	10.67	10.98	0.31	10.46	10.70	0.24

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.

Table 3: Model and Data Earnings Decomposition

Moment	Model			Firm Effect		Worker Effect	
	1981	2013	change	2013	change	2013	change
Between	0.34	0.42	0.08	0.12	-0.22	0.72	0.38
Within	0.66	0.58	-0.08	0.88	0.22	0.28	-0.38

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.

deviations of 0.20 log points. At the top of the skill distribution, the data display a stronger positive correlation between worker earnings compared to our more negatively sorted model economy.

5.3 Counterfactual

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. 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, the share of within-firm earnings dispersion would have increased by 22 percentage points. The worker effect counterfactual similarly evaluates the model using the project distribution of 1981 and the worker distribution of 2013. The right panel shows that the share of the within-firm earnings dispersion would have decreased by 38 percentage points. That is, our counterfactual analysis shows that both the changes in the worker and project distributions between 1981 and 2013 are important in the model in generating the observed change in earnings dispersion.²⁴

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.

²⁴The quantitative analysis thus far assumes that teams are of size three – two workers and a project. In Appendix A.8, we demonstrate that our model can also account for the decomposition of earnings within and across firms with teams of sizes four, five, and six. We also find that the non-targeted moments of average coworker earnings are robust to team size. As the number of team members increases, the firm and worker effect become more muted.

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.
- ALKAN, A. (1988): “Nonexistence of Stable Threesome Matchings,” *Mathematical Social Sciences*, 16(2), 207–209.
- 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.
- BOERMA, J., A. TSYVINSKI, R. WANG, AND Z. ZHANG (2023): “Composite Sorting,” NBER Working Paper No. 31656.
- 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.
- 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.
- DUPUY, A., AND A. GALICHON (2014): “Personality Traits and the Marriage Market,” *Journal of Political Economy*, 122(6), 1271–1319.

- 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. PUCETTI, 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.
- 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.
- GAVILAN, A. (2012): “Wage Inequality, Segregation by Skill and the Price of Capital in an Assignment Model,” *European Economic Review*, 56(1), 116–137.
- 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.
- GRETSKY, N. E., J. M. OSTROY, AND W. R. ZAME (1992): “The Nonatomic Assignment Model,” *Economic Theory*, 2, 103–127.
- HARDY, G. H., J. E. LITTLEWOOD, AND G. PÓLYA (1929): “Some Simple Inequalities Satisfied by Convex Functions,” *Messenger of Mathematics*, 58, 145–152.

- HERKENHOFF, K., J. LISE, G. MENZIO, AND G. M. PHILLIPS (2023): “Production and Learning in Teams,” NBER Working Paper No. 25179.
- JAKOBSONS, E., X. HAN, AND R. WANG (2016): “General Convex Order on Risk Aggregation,” *Scandinavian Actuarial Journal*, 2016(8), 713–740.
- 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.
- KNOTT, M., AND C. SMITH (2006): “Choosing Joint Distributions so that the Variance of the Sum is Small,” *Journal of Multivariate Analysis*, 97(8), 1757–1765.
- KREMER, M., AND E. MASKIN (1996): “Wage Inequality and Segregation by Skill,” NBER Working Paper No. 5718.
- 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. (2022): “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.
- PUC CETTI, G., AND R. WANG (2015): “Extremal Dependence Concepts,” *Statistical Science*, 30(4), 485–517.
- QUINT, T. (1991): “The Core of An m -Sided Assignment Game,” *Games and Economic Behavior*, 3(4), 487–503.
- RACHEV, S. T., AND L. RÜSCHENDORF (1998): *Mass Transportation Problems: Volume I: Theory*. Springer Science and Business.
- RÜSCHENDORF, L., AND L. UCKELMANN (2002): “Variance Minimization and Random Variables with Constant Sum,” in *Distributions with Given Marginals and Statistical Modelling*, ed. by C. M. Cuadras, J. Fortiana, and J. A. Rodriguez-Lallena.
- 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.
- SHERSTYUK, K. (1999): “Multisided Matching Games with Complementarities,” *International Journal of Game Theory*, 28, 489–509.
- 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 Teams

Appendix

Job Boerma, Aleh Tsyvinski and Alexander Zimin

November 2023

A Proofs

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

A.1 Equilibrium, Planning Problem and Duality

We formally define an equilibrium for the economy.

A firm with project z chooses workers x_1, x_2, \dots, x_m to maximize profits taking the wage schedule for workers w as given. The firm problem is:

$$v(z) = \max_{\{x_i\}} y(x_1, \dots, x_m, z) - \sum_{i=1}^m w(x_i). \quad (\text{A.1})$$

Worker x_i chooses to work for firm z with coworkers $\{x_j\}$ for $j \neq i$ to maximize their wage income. The worker takes the wage schedule for their coworkers and the firm value v as given:

$$w(x_i) = \max_{\{x_j\}, z} y(x_1, x_2, \dots, x_m, z) - \sum_{j \neq i}^m w(x_j) - v(z). \quad (\text{A.2})$$

An equilibrium is a wage function w , firm value function v , and feasible assignment function π , such that firms solve their profit maximization problem (A.1), workers solve the worker problem (A.2), and satisfy a feasibility constraint:

$$\int y(x_1, x_2, \dots, x_m, z) d\pi = \sum_{i=1}^m \int w(x_i) dF_x + \int v(z) dF_z, \quad (\text{A.3})$$

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

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

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

Since the assignment in Lemma 1 is concentrated on set M , we refer to M as the set of potential matches. The proof to Lemma 1 only makes use of a notion of weak duality.

Weak Duality. Let $\pi \in \Pi(F_{x_1}, \dots, F_{x_m}, F_z)$ be a joint probability measure, and $f_i \in L^1(X, F_{x_i})$, $g \in L^1(Z, F_z)$ be integrable functions such that $y(x_1, \dots, x_m, z) \leq \sum_{i=1}^m f_i(x_i) + g(z)$ for all teams (x_1, \dots, x_m, z) . Then²⁵

$$\min_{\{f_i\}, g} \sum_{i=1}^m \int f_i(x_i) dF_{x_i} + \int g(z) dF_z \geq \max_{\pi \in \Pi} \int y(x_1, \dots, x_m, z) d\pi. \quad (\text{A.4})$$

Proof. For any functions $\{f_i\}$ and g so that $y(x_1, \dots, x_m, z) \leq \sum_{i=1}^m f_i(x_i) + g(z)$ we have:

$$\max_{\pi \in \Pi} \int y(x_1, \dots, x_m, z) d\pi \leq \int \left(\sum_{i=1}^m f_i(x_i) + g(z) \right) d\pi = \sum_{i=1}^m \int f_i(x_i) dF_{x_i} + \int g(z) dF_z,$$

where the equality follows as $\pi \in \Pi(F_{x_1}, \dots, F_{x_m}, F_z)$. Since the inequality holds for any $(\{f_i\}, g)$ and g it holds for $(\{f_i\}, g)$ that minimize the right-hand side. \square

We use weak duality to establish Lemma 1 by contradiction.

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

$$\begin{aligned} \max_{\pi \in \Pi} \int y(x_1, \dots, x_m, z) d\pi &> \int y(x_1, \dots, x_m, z) d\hat{\pi} = \sum_{i=1}^m \int \hat{f}_i(x_i) dF_{x_i} + \int \hat{g}(z) dF_z \\ &\geq \min_{\{f_i\}, g} \sum_{i=1}^m \int f_i(x_i) dF_{x_i} + \int g(z) dF_z, \end{aligned}$$

where the equality follows by the assumption. This contradicts weak duality (A.4).

²⁵The minimum and maximum are attained by strong duality. We discuss this notion of duality in Appendix A.6.

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

$$\begin{aligned} \min_{\{f_i\}, g} \sum_{i=1}^m \int f_i(x_i) dF_{x_i} + \int g(z) dF_z &< \sum_{i=1}^m \int \hat{f}_i(x_i) dF_{x_i} + \int \hat{g}(z) dF_z \\ &= \int y(x_1, \dots, x_m, z) d\hat{\pi} \leq \max_{\pi \in \Pi} \int y(x_1, \dots, x_m, z) d\pi, \end{aligned}$$

where the equality follows by the assumption. This inequality contradicts weak duality (A.4).

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 7.²⁶

Proposition 7. Let π solve the planning problem and let (w, v) solve the dual problem. Then, wage schedule w , firm value function v , and assignment π are an equilibrium.

Proof. The proof follows from the constraints on the dual problem, or $w(x_i) \geq y(x_1, \dots, x_m, z) - \sum_{j \neq i}^m w(x_j) - v(z)$, which imply:

$$w(x_i) \geq \max_{x_{-i}, z} \left(y(x_1, \dots, x_m, z) - \sum_{j \neq i}^m w(x_j) - v(z) \right). \quad (\text{A.5})$$

The equality follows since $y(x_1, \dots, x_m, z) = \sum_{i=1}^m w(x_i) + v(z)$ 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. \square

Proposition 7 implies that assignment function π , wage schedule w , and firm value function v are an equilibrium.²⁷ Given this connection, we call w the wage schedule and v the firm value.²⁸

²⁶Similar to [Gretsky, Ostroy, and Zame \(1992\)](#), we decentralize the solutions to the primal and dual problem as a competitive equilibrium. A similar argument for one-to-one assignment problems is in Proposition 2.3 of [Galichon \(2018\)](#) and discussed in [Chade, Eeckhout, and Smith \(2017\)](#).

²⁷We observe that an economy with m identical worker distributions F_x with mass equal to one for each position is equivalent to an economy with a single distribution of workers F_x with mass equal to m . Owing to the symmetry of worker skills in the team quality function (1), these are equivalent. Intuitively, any equilibrium worker assignment for a planning problem with distinct worker distributions can be made symmetric. By assigning $1/m$ of the mass of equilibrium worker pairings to one distribution and half to another, we obtain m identical worker distributions. The optimum value is unaffected as the same distribution of worker pairings can be made due to symmetry.

²⁸The 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 technology. Moreover, Proposition 7 shows that the decentralized equilibrium coincides with the solution to the planning problem so the competitive equilibrium is efficient.

A.2 Majorization Inequality

We verify the second assumption to the majorization inequality (13). The goal is to show that the mixed and countermonotonic assignment is optimal. We do not require all properties of the mixed and countermonotonic assignment described in Definition 1. Specifically, we prove Proposition 8.

Proposition 8. Suppose assignment $\pi \in \Pi$ satisfies two conditions. First, the loss of the $1 - q$ largest teams are the same. Second, consider the first $t \leq q$ teams. Worker $x_i \in F_i$ belongs to one of them if and only if $x_i \in [0, I_i(\underline{t}_i(t))]$ or $x_i \in [I_i(1 - \bar{t}_i(t)), 1]$. Then, $\pi \in \Pi$ solves the planning problem (5).

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

$$S_\gamma(t) = \int_0^t \log I_\gamma(p) dp. \quad (\text{A.6})$$

To verify (13) we compare S_{γ^*} to S_γ .

Suppose there exists a mixed and countermonotonic assignment π^* with an associated loss distribution γ^* . Next, we prove S_{γ^*} is the maximal cumulative logarithmic loss in line with the second assumption to the majorization inequality, that is, for any assignment π , $S_{\gamma^*}(t) \geq S_\gamma(t)$ for all $t \in [0, 1]$. To establish this, we first bound the cumulative logarithmic losses in the countermonotonic sets of π^* from below.

Lemma 2. For any feasible assignment π and for all $t \in [0, q]$, $S_\gamma(t) \leq S_{\gamma^*}(t)$.

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:

$$S_\gamma(t) \leq \int \log \ell(x_1, \dots, x_m, z) d\pi_U. \quad (\text{A.7})$$

where π_U restricts assignment π to the set U .

Define the set of teams with the $\underline{t}_k(t)$ highest skill workers in dimension k as:

$$V_k := \{(x_1, \dots, x_m, z) \in [0, 1]^n \mid 0 \leq x_k \leq I_k(\underline{t}_k(t))\} \quad (\text{A.8})$$

The measure of the set V_k with respect to assignment π is $\underline{t}_k(t)$. The union of sets $V = V_1 \cup \dots \cup V_n$ 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 additional set of teams to the set U .

Given that π_U represents the restriction of the assignment function π to the specified set U , we use (A.7) to write:

$$S_\gamma(t) \leq \int \log \ell(x_1, \dots, x_m, z) d\pi_U = \sum_{k=1}^m \int \log x_k d\text{Pr}_{x_k}(\pi_U) + \int \log z d\text{Pr}_z(\pi_U),$$

where Pr_{x_k} is the projection of measure π_U onto the k -th dimension. To characterize the upper bound, we bound $\int \log x_k d\text{Pr}_{x_k}(\pi_U)$. Since $V_k \subset U$, the density function that corresponds to measure $\text{Pr}_{x_k}(\pi_U)$ is equal to the skill density $f_k(x)$ for all $0 \leq x \leq I(\underline{t}_k(t))$ and is less than skill density f_k otherwise. Hence,

$$\begin{aligned} \int \log x_k d\text{Pr}_{x_k}(\pi_U) &\leq \int_0^{I_k(\underline{t}_k(t))} \log(x_k) f_k(x_k) dx_k + \int_{I_k(1-\bar{t}_k(t))}^1 \log(x_k) f_k(x_k) dx_k \\ &= \int_0^{\underline{t}_k(t)} \log I_k(p) dp + \int_{1-\bar{t}_k(t)}^1 \log I_k(p) dp. \end{aligned}$$

where the inequality follows as the second term on the right-hand side captures the contribution to logarithmic losses by the $\bar{t}_k(t)$ lowest skill workers. The equality follows by change of variables. Summing over all k dimensions, we obtain:

$$S_\gamma(t) \leq \sum_{k=1}^n \left(\int_0^{\underline{t}_k(t)} \log I_k(p) dp + \int_{1-\bar{t}_k(t)}^1 \log I_k(p) dp \right) = S_{\gamma^*}(t), \quad (\text{A.9})$$

which was what we wanted. □

Lemma 3. For any feasible assignments γ and for all $t \in [q, 1]$,

$$S_\gamma(t) \leq S_{\gamma^*}(t).$$

Proof. We first observe that cumulative logarithmic loss S_γ is a convex function as it is the integral over an increasing function. Since S_γ is convex, for all $t \in [q, 1]$:

$$S_\gamma(t) \leq \frac{1-t}{1-q} S_\gamma(q) + \frac{t-q}{1-q} S_\gamma(1) \leq \frac{1-t}{1-q} S_{\gamma^*}(q) + \frac{t-q}{1-q} S_{\gamma^*}(1) \quad (\text{A.10})$$

where the final inequality follows because $S_\gamma(q) \leq S_{\gamma^*}(q)$ by Lemma 2, and $S_\gamma(1) = S_{\gamma^*}(1)$ by the verification of first assumption to the majorization inequality.

The integral over logarithmic losses $S_{\gamma^*}(t)$ is linear for teams with the highest losses $t \in [q, 1]$ by the condition that the loss of the $1-q$ largest teams are the same. Hence, for teams $t \in [q, 1]$:

$$S_{\gamma^*}(t) = \frac{1-t}{1-q} S_{\gamma^*}(q) + \frac{t-q}{1-q} S_{\gamma^*}(1) \quad (\text{A.11})$$

Combining (A.10) and (A.11) the result follows. □

We combine Lemma 2 and Lemma 3 to verify the second assumption of the majorization inequality, which concludes the proof.

A.3 Characterization of Mixed and Countermonotonic Assignment

The mixed-and-countermonotonic assignment in Definition 1 may appear general, but in fact, the countermonotonic sets $b_i(t)$ fully determine the assignment π .

First, at each percentile $t \leq q$ in the distribution of output losses, there are n teams: one from each countermonotonic set. The logarithmic output loss at percentile $t \leq q$, which we denote $l(t)$, equals²⁹

$$l(t) = m\lambda(t) + \sum_{i=1}^n \log I_i(\underline{t}_i(t)). \quad (\text{A.12})$$

where $\lambda(t)$ is the skill gap between the high-skill and low-skill worker that are in teams at percentile $t \leq q$, which is identical across distributions i :

$$\lambda(t) = \log I_i(1 - \bar{t}_i(t)) - \log I_i(\underline{t}_i(t)). \quad (\text{A.13})$$

Since the functions I_i , \underline{t}_i , and \bar{t}_i are continuous, the functions λ and l are continuous too. It follows from the definition of l that the logarithmic loss function is increasing.

Next, we connect the functions \underline{t}_i and \bar{t}_i . Teams in the countermonotonic sets have low output losses and pair a single low value from one distribution with high values from the remaining distributions. Since the quality of the high-skill worker decreases with team loss percentile t , that is, $\underline{t}_i(t)$ increases from $\underline{t}_i(0) = 0$, it follows that the t teams with the lowest losses draw a single high-skill worker from each distribution, so that for $t \leq q$:

$$\underline{t}_1(t) + \dots + \underline{t}_n(t) = t. \quad (\text{A.14})$$

Equivalently, high values from the i -th distribution, starting from $\bar{t}_i(t) = 0$, are paired into teams with low values in the remaining countermonotonic sets. Thus, for $t \leq q$:

$$\bar{t}_i(t) = \sum_{j \neq i} \underline{t}_j(t) = t - \underline{t}_i(t). \quad (\text{A.15})$$

²⁹For each percentile $t \leq q$, consider all n teams at this percentile. The logarithmic loss of the i -th team equals $l_i(t) = \log I_i(\underline{t}_i(t)) - \log I_i(1 - \bar{t}_i(t)) + \sum \log I_j(1 - \bar{t}_j(t))$. Since the n teams are all at the t -th percentile in the distribution of output losses their losses are identical, and hence $\lambda(t)$ is independent of i . Using the skill gap λ , we express $l(t)$ as in equation (A.12). Summing the definition of the skill gap over $i \leq m$: $m\lambda(t) = \sum_{i \leq m} (\log I_i(1 - \bar{t}_i(t)) - \log I_i(\underline{t}_i(t)))$, and by adding and subtracting $\log I_n(\underline{t}_n(t))$: $m\lambda(t) = \log I_n(\underline{t}_n(t)) + \sum_{i \leq m} \log I_i(1 - \bar{t}_i(t)) - \sum \log I_i(\underline{t}_i(t))$ and thus equation (A.12).

where the final equality follows from (A.14).

Finally, we consider the part of the assignment π that is located on the mixed set. Given functions $\{\underline{t}_i\}^n$, we define for $x \in [0, q]$, the average logarithmic loss for mediocre workers and jobs $h(x) = \frac{1}{1-x} \sum \int_{\underline{t}_i(x)}^{\bar{t}_i(x)} \log I_i(t) dt$. By construction, the logarithmic loss of any team located on mixed set B is the same and equal to $h(q)$. We can compute it directly:

$$l(q) = \frac{1}{1-q} \int_B \log x_1 \dots x_m z d\pi = \frac{1}{1-q} \sum_{k=1}^n \int_{\underline{p}_k}^{\bar{p}_k} \log x_k dF_k(x_k) = h(q). \quad (\text{A.16})$$

This equation asserts continuity of the assignment: the output loss of teams at the end of countermonotonic sets, $l(q)$, equals the loss of teams located on the mixed set, $h(q)$.

If a mixed and countermonotonic assignment exists, then equations (A.12) to (A.16) are satisfied. Together with the ability to mix mediocre workers and projects, (A.12) to (A.16) are in fact sufficient to show that a mixed and countermonotonic assignment exists as we show in Proposition 9.³⁰

Proposition 9. Consider continuous non-negative increasing functions $\{\underline{t}_i\}_{i=1}^n$ and a mass $0 \leq q \leq 1$ which is located on the countermonotonic sets. There exists a mixed and countermonotonic assignment defined by these functions if and only if

1. For each $t \in [0, q]$, $\underline{t}_1(t) + \dots + \underline{t}_n(t) = t$ (A.14);
2. For each $t \in [0, q]$, $\bar{t}_i(t) = t - \underline{t}_i(t)$ are non-negative and increasing (A.15);
3. For each $t \in [0, q]$, the skill gap between the high-skill and low-skill worker $\lambda(t) = \log I_i(1 - \bar{t}_i(t)) - \log I_i(\underline{t}_i(t))$ is independent of i (A.13);
4. For each $t \in [0, q]$, the function $l(t) = m\lambda(t) + \sum_{i=1}^n \log I_i(\underline{t}_i(t))$ is increasing in t (A.12);
5. $h(q) = l(q)$ (A.16);
6. The restriction of $\{F_i\}_{i=1}^n$ to the intervals $[I_i(\underline{t}_i(q)), I_i(1 - \bar{t}_i(q))]$ is mixable.

Proof. To prove Proposition 9 it remains to be shown that if Conditions 1 to 6 apply, then there exists a mixed and countermonotonic assignment defined by the functions $\{\underline{t}_i\}$ and a mass q located on countermonotonic sets. The proof is split into two parts. First, we construct the assignment π associated with $\{\underline{t}_i\}$ and show that it is feasible. Second, we show that assignment π is a mixed and countermonotonic assignment. While the result is rather intuitive, this technical proof verifies this formally.

³⁰Formally, we say that the distributions $\{F_i\}_{i=1}^n$ restricted to the mediocre intervals $[I_i(\underline{t}_i(q)), I_i(1 - \bar{t}_i(q))]$ can be mixed if there exists an assignment between these distribution such that the output loss of each team is identical.

We construct the mixed and countermonotonic assignment completely using the functions $\{\underline{t}_i\}$ and the mass q located on the countermonotonic sets. To understand that the assignment is fully determined, first observe that we can construct the increasing functions $\{\bar{t}_i\}$ by (A.15). Given the functions $\{\underline{t}_i, \bar{t}_i\}$ and a mass q , we use the definition of countermonotonic sets in Definition 1 to construct countermonotonic sets.

We next identify the threshold percentiles \bar{p}_i and \underline{p}_i and show that these threshold percentiles are inside the unit interval. We denote $\underline{p}_i = \underline{t}_i(q)$ and $\bar{p}_i = 1 - \bar{t}_i(q) = \underline{p}_i + (1 - q)$, where the second equality follows from equation (A.15) evaluated at $t = q$. The threshold \underline{p}_i is non-negative since \underline{t}_i is a non-negative function, which also implies that $\bar{p}_i = \underline{p}_i + (1 - q)$ is non-negative. Next, since $\bar{t}_i(q) \geq 0$ by (A.15) it implies that $\bar{p}_i \leq 1$, and therefore $\underline{p}_i = \bar{p}_i - (1 - q) \leq 1$. Hence, both \underline{p}_i and \bar{p}_i are thresholds inside the unit interval.

Given the threshold percentiles $\{\underline{p}_i, \bar{p}_i\}$, we know the support of the mixed set. So far, this procedure only suggests a support of an assignment. A key step is to show there exists a measure concentrated on this support.

The first step in showing there exists a measure concentrated on the suggested support is to reparameterize the countermonotonic set $b_i(t)$ from percentiles in the distribution of team losses t to percentiles in the distribution of the i -th worker, which we here denoted by q_i . Since the function \underline{t}_i is increasing, its inverse \underline{t}_i^{-1} is well-defined. Function $\underline{t}_i(t)$ takes percentile t of team losses and returns the percentile of the worker in the i -th distribution that is part of this team. The inverse function takes the percentile of the worker in the i -th distribution and returns the percentile in the distribution of team losses. This percentile can thus directly be substituted into the definition of a countermonotonic set b_i . For each i , we formally consider the mapping

$$\tilde{b}_i(q_i) = b_i(\underline{t}_i^{-1}(q_i)),$$

where $q_i \in [0, \underline{p}_i]$ and b_i is the countermonotonic set defined in Definition 1.

Let λ_i be the Lebesgue measure restricted to the interval $[0, \underline{p}_i]$, that is, all high-skill workers in distribution i up to percentile \underline{p}_i . Define the corresponding countermonotonic set assignment π_i as the push-forward image of the measure λ_i under the mapping $q_i \rightarrow \tilde{b}_i(q_i)$. We take $\{\lambda_i\}_{i=1}^n$ and will then analyze the corresponding marginal distributions of these countermonotonic sets $\pi_b = \pi_1 + \dots + \pi_n$. First, we find $\tilde{b}_{ii}(q)$:

$$\tilde{b}_{ii}(q) = b_{ii}(\underline{t}_i^{-1}(q)) = I_i(\underline{t}_i(\underline{t}_i^{-1}(q))) = I_i(q), \tag{A.17}$$

where the second equality follows by definition of b_{ii} in Definition 1. In words, the i -th marginal distribution of assignment π_i is the push-forward image of the uniform measure on $[0, \underline{p}_i]$ under the mapping $x \rightarrow I_i(x)$, which is the distribution F_i restricted to the interval $[0, I_i(\underline{p}_i)]$.

How many low skill workers are taken by the countermonotonic sets? We assign low skill workers when we construct countermonotonic sets corresponding to the high skill workers in the remaining distributions. We next calculate how many low skill workers from distribution k are assigned according to the assignment function.

For each percentile $s_k \in [\bar{p}_k, 1]$ among the low skill workers in the marginal distribution of worker k , we find the measure of the set $S_k = \{x \in [0, 1]^n \mid x_k \geq I_k(s_k)\}$ with respect to the measure π_i for some $i \neq k$. In words, how many low-skilled workers from the k -th distribution $x_k \geq I_k(s_k)$ are taken by countermonotonic set π_i . First, since the function \bar{t}_k is increasing from 0 to $1 - \bar{p}_k$ by (A.15), its inverse \bar{t}_k^{-1} is well-defined on the interval $[0, 1 - \bar{p}_k]$. We can thus write:

$$\begin{aligned} x_k = \tilde{b}_{ik}(q_i) = b_{ik}(\underline{t}_i^{-1}(q_i)) \geq I_k(s_k) &\iff I_k(1 - \bar{t}_k(\underline{t}_i^{-1}(q_i))) \geq I_k(s_k) \iff \bar{t}_k(\underline{t}_i^{-1}(q_i)) \leq 1 - s_k \\ &\iff q_i \leq \underline{t}_i(\bar{t}_k^{-1}(1 - s_k)). \end{aligned}$$

where q_i is again the percentile of the high-skill workers from the i -th distribution, and the second equality follows (A.17). Worker x_k is in the set of lowest skill workers if we are looking at a percentile q_i that satisfies the final inequality. This means that the measure of S_k with respect to π_i is equal to $\underline{t}_i(\bar{t}_k^{-1}(1 - s_k))$. Thus, the measure of S_k with respect to π_b is:

$$\pi_b(S_k) = \sum_{i \neq k} \underline{t}_i(\bar{t}_k^{-1}(1 - s_k)) = \bar{t}_k(\bar{t}_k^{-1}(1 - s_k)) = 1 - s_k.$$

where the second equality follows from the definition of \bar{t}_k in equation (A.15). Thus, the k -th marginal distribution of $\pi_b(S_k)$ has the same cumulative distribution as F_k restricted to $[I_k(s_k), 1]$. For each s_k , the mass of low-skill workers above the s_k -th percentile that we assign is exactly $1 - s_k$. Together with equation (A.17) this implies

$$\text{pr}_i(\pi_b) = F_i \Big|_{[0, I_i(\underline{p}_i)] \cup [I_i(\bar{p}_i), 1]}. \tag{A.18}$$

Now, let π_c be an assignment that pairs mediocre workers in teams with the same total output loss, and denote $\pi = \pi_b + \pi_c$. Equation (A.18) and Condition 6 directly imply that $\pi \in \Pi(\{F_i\}_{i=1}^n)$.

The second part of the proof verifies that the constructed assignment is indeed a mixed and

countermonotonic assignment. Given the constructed assignment π , we check that all properties of the mixed and countermonotonic assignment in Definition 1 are satisfied.

The first property is directly verified since we discussed above that $\bar{p}_i = \underline{p}_i + (1 - q)$.

To verify the second property, we validate both the countermonotonic sets and the mixed set. It follows directly from their construction that the countermonotonic sets have the required form. Next, we verify the mixed set. For the verification of the mixed set there are two properties to validate. First, that the values x_i are bounded between $I_i(\underline{p}_i)$ and $I_i(\bar{p}_i)$ is immediate as these are the mediocre workers and jobs.

Second, we show the mixed teams attain a constant, in particular, $I_i(\underline{p}_i) \prod_{j \neq i} I_j(\bar{p}_j)$. To validate that the loss of any team on the mixed set, L_B , equals this constant, we use the definition of the average logarithmic loss for mediocre workers and jobs $h(x) = \frac{1}{1-x} \sum \int_{\underline{t}_i(x)}^{\bar{t}_i(x)} \log I_i(t) dt$, we have $\log L_B = h(q)$. By the fifth condition (A.16), $h(q) = l(q)$, and by the fourth and third condition, we write:

$$l(q) = \log I_i(\underline{t}_i(q)) + \sum_{j \neq i} \log I_j(1 - \bar{t}_j(q)) = \log I_i(\underline{p}_i) + \sum_{j \neq i} \log I_j(\bar{p}_j).$$

Hence we establish the loss on the mixed set attains $I_i(\underline{p}_i) \prod_{j \neq i} I_j(\bar{p}_j)$, verifying the second property.

The third property of Definition 1 need not be verified directly as it is implied by the statement of the countermonotonic sets in the second property.

Finally, for each percentile t , we find the mass of teams with output losses below that of the team located at point $b_i(t)$. The logarithmic loss of this team equals $l(t)$. Since the function $l(t)$ is increasing by (A.12) and $\log L_B = l(q) > l(t)$, the logarithmic loss of any team from the mixed set is larger than the logarithmic loss of the team $b_i(t)$.

We next find all the teams with a loss below the loss of team $b_i(t)$ located on the i -th countermonotonic set. The logarithm of the loss of team $b_i(x)$ is equal to $l(x)$ by Conditions 3 and 4 in Proposition 9. Since the loss function $l(t)$ is increasing, $l(x) \leq l(t) \iff x \leq t$. As a result, the mass on the countermonotonic set b_i such that $x \leq t$ is equal to $\underline{t}_i(t)$. In turn, this implies that the mass of teams with logarithmic losses below the logarithmic loss of $b_i(t)$ is $\sum_{i=1}^n \underline{t}_i(t) = t$, where the equality follows by the first condition in Proposition 9. This means that $b_i(t)$ is indeed located at the t -th quantile of the teams distribution. \square

A.4 Existence of the Mixed and Countermonotonic Assignment

In this appendix, we establish the existence of a mixed and countermonotonic assignment.³¹ We prove our results under Assumption 1.³²

Assumption 1. For all x_i , $x_i f_i(x_i)$ is non-decreasing.

To construct the optimal assignment, we find continuous non-negative increasing functions $\{\underline{t}_i\}$ and a threshold q satisfying all the conditions of Proposition 9. Conditions 1 and 3 in Proposition 9 can be considered a system of functional equations:

$$t = \underline{t}_1(t) + \cdots + \underline{t}_n(t), \quad (\text{A.14})$$

$$\lambda(t) = \log I_i(1 - \bar{t}_i(t)) - \log I_i(\underline{t}_i(t)) \quad (\text{A.13})$$

for all i . The structure of the proof is as follows. First, we show when the system of functional equations (A.13) and (A.14) has a unique solution. Second, we show this unique solution satisfies the remaining Conditions 4 to 6.³³ That is, Conditions 4 to 6 are then properties of the solution.

To show that the system of functional equations specified by Conditions 1 and 3 has a unique solution, we first introduce new notation to transform the system of equations.

Specifically, we introduce new functions $\tilde{y}_i(t)$, which capture the negative logarithmic skills of the high-skilled workers on the t -th team located on the i -th countermonotonic set:

$$\tilde{y}_i(t) = -\log I_i(\underline{t}_i(t)). \quad (\text{A.19})$$

Since $0 \leq I_i \leq 1$, we note $0 \leq \tilde{y}_i < \infty$. We also introduce the distribution function $\tilde{F}_i(x)$, which is defined as the cumulative distribution function of negative logarithmic skills of workers from the i -th distribution, or $\tilde{F}_i(t) := P(\tilde{y} \leq t) = P(x \geq \exp(-t)) = 1 - F_i(\exp(-t))$. Using this, we connect $\tilde{y}_i(t)$ to $\underline{t}_i(t)$:

$$\underline{t}_i(t) = 1 - \tilde{F}_i(\tilde{y}_i(t)). \quad (\text{A.20})$$

³¹For multi-sided assignment games Alkan (1988) and Quint (1991) show that stable assignments may not exist. Sherstyuk (1999) analyzes multi-sided assignment games with supermodularity and provides conditions under which stable assignments exists.

³²When the probability density function is differentiable, Assumption 1 is equivalent to bounding the elasticity of the density function from below by -1 . All distributions with non-decreasing density functions trivially satisfy this restriction. Assumption 1 is satisfied by Beta distribution, truncated exponential and lognormal distributions for a subset of parameters, amongst others. Naturally, Assumption 1 holds for linear combinations of density functions that satisfy it.

³³Condition 2 is implied by Condition 1 and the monotonicity of $\{\underline{t}_i\}$. Specifically, $\bar{t}_i(t) = \sum_{j \neq i} \underline{t}_j(t)$ is increasing because $\underline{t}_j(t)$ are increasing.

which is verified from $\tilde{F}_i(t) = 1 - F_i(\exp(-t))$, together with $\tilde{y}_i(t) = -\log I_i(\underline{t}_i(t))$.

With these new functions, we write the system of functional equations (A.13) and (A.14) as:

$$\begin{aligned} t &= \sum_{i=1}^n (1 - \tilde{F}_i(\tilde{y}_i(t))) \\ t &= \tilde{F}_i(\tilde{y}_i(t) - \lambda(t)) + (1 - \tilde{F}_i(\tilde{y}_i(t))) \end{aligned}$$

for all i .³⁴

Given a unique solution, we write the conditions of Proposition 9 in terms of $\tilde{y}_i(t)$. First, we check that $\{\underline{t}_i\}$ are increasing functions. First, the functions $\underline{t}_i(t) = 1 - \tilde{F}_i(\tilde{y}_i(t))$ is non-negative if and only if $\tilde{F}_i(\tilde{y}_i(t)) \leq 1$. Next, $\underline{t}_i(t)$ is increasing if and only if $\tilde{F}_i(\tilde{y}_i(t))$ is decreasing. By the monotonicity of \tilde{F}_i this is equivalent to $\tilde{y}_i(t)$ is decreasing. We verify that $\tilde{y}_i(t)$ decreases in Lemma 4.

The concavity of cumulative distribution functions \tilde{F}_i on $[0, \infty)$ is equivalent to $x_i f_i(x_i)$ is non-decreasing on $[0, 1]$, where f_i is the density of the skill distribution. Moreover, it is equivalent to \tilde{f}_i , the density function of the negative logarithmic skill distribution, is non-increasing on $[0, \infty)$. The distribution function for the negative logarithmic skill \tilde{x}_i is related to the distribution for x_i through $\tilde{F}_i(\tilde{x}_i) = 1 - F_i(\exp(-\tilde{x}_i))$. The distribution function \tilde{F}_i then has a corresponding density function:

$$\tilde{f}_i(\tilde{x}_i) = \frac{d}{d\tilde{x}_i} \tilde{F}_i(\tilde{x}_i) = -\frac{d}{d\tilde{x}_i} F_i(\exp(-\tilde{x}_i)) = f_i(\exp(-\tilde{x}_i)) \exp(-\tilde{x}_i). \quad (\text{A.21})$$

So, \tilde{f}_i is non-increasing is equivalent to $x_i f_i(x_i)$ is non-decreasing. Therefore, to prove our results, we use Assumption 1.

Conditions 1 to 3 directly follow since $\{\tilde{y}_i\}$ is a solution to the system of functional equations (A.13) and (A.14). To verify the fourth condition, we express the function $l(t)$ in terms of $\tilde{y}_i(t)$ using the definition of $\tilde{y}_i(t)$ in (A.19):

$$l(t) = m\lambda(t) - \sum_{i=1}^n \tilde{y}_i(t). \quad (\text{A.22})$$

³⁴The second equation is derived from equation (A.13) as follows $\lambda(t) = \log I_i(1 - \bar{t}_i(t)) - \log I_i(\underline{t}_i(t)) = \log I_i(1 - t + \underline{t}_i(t)) - \log I_i(\underline{t}_i(t))$ where the second equality follows by (A.15). We simplify the right-hand expression using (A.19) as $\tilde{y}_i(t) - \lambda(t) = -\log I_i(1 - t + \underline{t}_i(t))$ which is equivalent to $1 - t + \underline{t}_i(t) = 1 - \tilde{F}_i(\tilde{y}_i(t) - \lambda(t))$. Reorganizing this equality using (A.20), we obtain $t = \tilde{F}_i(\tilde{y}_i(t) - \lambda(t)) + \underline{t}_i(t) = \tilde{F}_i(\tilde{y}_i(t) - \lambda(t)) + (1 - \tilde{F}_i(\tilde{y}_i(t)))$. Suppose that all distribution functions \tilde{F}_i are concave on $[0, \infty)$, then we can use Lemma 3.1 in Jakobsons, Han, and Wang (2016) to show the system of equations has a unique solution. Since we have a symmetric system of functional equations with a unique solution the solution to the system of functional equations is itself symmetric. Uniqueness of the solution thus implies $(\underline{t}_i, \bar{t}_i)$ is identical for all workers i . As a result, the marginal worker product m_i in equation (14) is identical for all worker distributions, or $m_i = m$.

The fourth condition holds if and only if the right-hand side of equation (A.22) is increasing for $0 \leq t \leq q$. We verify this in Lemma 4.

To verify Condition 5, we rewrite the average logarithmic loss for mediocre workers and jobs $h(x)$ in terms of \tilde{y}_i :

$$h(x) = \frac{1}{1-x} \sum_{i=1}^n \int_{\underline{t}_i(x)}^{1-\bar{t}_i(x)} \log I_i(t) dt = -\frac{1}{1-x} \sum_{i=1}^n \int_{\tilde{y}_i(x)-\lambda(x)}^{\tilde{y}_i(x)} z_i d\tilde{F}_i(z_i).$$

The equality follows by a change of variables as $z_i = -\log I_i(t)$, which by definition is distributed according to \tilde{F}_i . By equation (A.19) it follows that $-\log I_i(\underline{t}_i(x)) = \tilde{y}_i(x)$ and by the third property of Proposition 9 it follows that $-\log I_i(1-\bar{t}_i(x)) = -\log I_i(\underline{t}_i(x)) - \lambda(x) = \tilde{y}_i(x) - \lambda(x)$, thereby giving the new bounds of integration. In sum, the fifth condition, $l(q) = h(q)$, holds if and only if

$$-m\lambda(q) + \sum_{i=1}^n \tilde{y}_i(q) = \frac{1}{1-q} \sum_{i=1}^n \int_{\tilde{y}_i(q)-\lambda(q)}^{\tilde{y}_i(q)} z d\tilde{F}_i(z). \quad (\text{A.23})$$

Lemma 3.2 in Jakobsons, Han, and Wang (2016) states the following:

Lemma 4. Suppose all cumulative distribution functions \tilde{F}_i are concave on $[0, \infty)$. Let $\{\tilde{y}_i\}$ and λ solve the system of functional equations (A.13) and (A.14). Denote

$$q = \inf\{q \in [0, 1] : l(q) \geq h(q)\} \quad (\text{A.24})$$

Then on the interval $(0, q)$ for each i :

- $0 < \lambda(t) < \tilde{y}_i(t)$
- $\lambda(t)$ and $\tilde{y}_i(t)$ are decreasing
- $\tilde{y}_i(t) - \lambda(t)$ is increasing
- $l(t)$ is increasing

We use Lemma 4 to establish that the fourth and fifth condition are satisfied. Condition 4 follows from the fourth bullet of Lemma 4. To verify the fifth condition, we consider three different cases. First, suppose the threshold q constructed in Lemma 4 is neither equal to 0 nor equal to 1. Then, it follows from continuity of l and h that $l(q) = h(q)$ by the definition of q , and hence the fifth condition holds automatically. The second case, where $q = 1$, is equivalent to the absence of the mixed set, and hence we do not need to verify Condition 5 and 6 of Proposition 9. Finally, the case $q = 0$ is equivalent to the absence of countermonotonic sets. In this case, we only need to check Condition 6.

To demonstrate Condition 6 we prove the existence of an assignment such that the output loss is constant across teams. This loss is represented by:

$$x_1 \dots x_n = \exp\left(-\sum_{i=1}^n y_i\right) \tag{A.25}$$

where $y_i = -\log x_i$ is the negative logarithmic skill of a worker. We can hence equivalently state the mixability condition as there exists an assignment of negative logarithmic skills such that the sum of logarithmic skills across teams remains constant.

To determine the distribution of negative logarithmic skills for mediocre workers, consider the following. The skill distribution is F_i confined to the interval $[I_i(\underline{t}_i(q)), I_i(1 - \bar{t}_i(q))]$. The distribution of negative logarithmic skills is \tilde{F}_i ; the distribution of negative logarithmic skills is thus \tilde{F}_i restricted to the interval $[-\log I_i(1 - \bar{t}_i(q)), -\log I_i(\underline{t}_i(q))]$. Using equations (A.13) and (A.19), $-\log I_i(\underline{t}_i(q)) = \tilde{y}_i(q)$ as well as $-\log I_i(1 - \bar{t}_i(x)) = \tilde{y}_i(x) - \lambda(x)$. Thus, the distribution of negative logarithmic skills \tilde{F}_i restricted to the interval $[\tilde{y}_i(q) - \lambda(q), \tilde{y}_i(q)]$. Finally, the total mass of mediocre workers is the same between all distributions and is equal to $1 - q$. We normalize all truncated distributions by this constant to make them probabilistic.

Given the distribution of negative logarithmic skills and the normalization of the truncated distribution, we can write the mixability condition in the following lemma.

Lemma 5. The mixability condition, Condition 6 of Proposition 9, is satisfied if and only if there exists an assignment, denoted $\tilde{\pi}$, such that:

- The marginal projections of assignment $\tilde{\pi}$ are $\frac{1}{1-q}\tilde{F}_i$, which are supported on the intervals $[\tilde{y}_i(q) - \lambda(q), \tilde{y}_i(q)]$.
- This assignment is supported on the hyperplane defined by $y_1 + \dots + y_n = C$, where C is a constant representing the sum of negative logarithmic skills across teams.

We show mixability under Assumption 1 as discussed above. This assumption implies that each cumulative distribution function \tilde{F}_i of negative logarithmic skills is concave over the interval $[0, \infty)$. Given that the density function of a distribution is the derivative of its cumulative distribution function, Assumption 1 means that the density function of negative log skills is non-increasing.

A direct result is that each distribution of negative log skills is supported on the interval $[\tilde{y}_i(q) - \lambda(q), \tilde{y}_i(q)]$, and that its density function decreases within this interval. Building on this,

we introduce Theorem 3.2 of Wang and Wang (2016) which provides necessary and sufficient conditions for mixability under these constraints.

Theorem 10 (Theorem 3.2 of Wang and Wang (2016)). Let $\{F_i\}$ be a collection of probability distributions. If each distribution F_i is supported on an interval $[l_i, r_i]$ and has a decreasing density function, then there exists an assignment $\tilde{\pi}$ supported on the hyperplane $\tilde{y}_1 + \dots + \tilde{y}_n = C$ with marginal distributions $\{F_i\}$ respectively if and only if for all i :

$$(r_i - l_i) + \sum_{j=1}^n l_j \leq \sum_{j=1}^n \int_{l_j}^{r_j} x_j dF_j.$$

For our problem, this result implies that the distributions of negative logarithmic skills of mediocre workers are mixable if and only if the following inequality is satisfied for each i :

$$\lambda(q) + \sum_{i=1}^n (\tilde{y}_i(q) - \lambda(q)) \leq \frac{1}{1-q} \sum_{i=1}^n \int_{\tilde{y}_i(q) - \lambda(q)}^{\tilde{y}_i(q)} z_i d\tilde{F}_i(z_i), \quad (\text{A.26})$$

since $l_k = \tilde{y}_k(q) - \lambda(q)$, $r_k = \tilde{y}_k(q)$, and $F_k = \frac{\tilde{F}_k}{1-q}$ in our environment. By equation (A.23) we find that this equation holds with equality if $q \in (0, 1)$. Under Condition 5 and Assumption 1, all conditions are thus met. When $q = 0$, $l(q) \geq h(q)$ by the definition of q in (A.24). Hence, the distributions of negative logarithmic skills of mediocre workers are mixable.

A.5 General Production Functions

We generalize the functional form of the production function.

Generalizing Technology for $n = 3$. In the case of two workers, the technology is generalized to additionally incorporate interaction terms between workers and coworkers x_1x_2 , workers and firms x_1z , and coworkers and firms x_2z . Let the technology be given by:

$$y(x_1, x_2, z) = \varphi_0 + \varphi_1x_1 + \varphi_2x_2 + \varphi_3z - \varphi_{12}x_1x_2 - \varphi_{13}x_1z - \varphi_{23}x_2z - x_1x_2z, \quad (\text{A.27})$$

where φ_{ij} are non-negative to ensure submodularity.

Proposition 11. Let the technology $y(x_1, x_2, z)$ take the form (A.27). Suppose the distributions of workers and projects, F_x and F_z respectively, satisfy Assumption 1 whereas $\tilde{x}f_{\tilde{x}}(\tilde{x})$ is non-decreasing for $\tilde{x} \in \{x + \varphi_{13}, x + \varphi_{23}, z + \varphi_{12}\}$. Then a mixed and countermonotonic assignment is optimal.

Proof. Without loss, we disregard the constant and linear terms in the technology $y(x_1, x_2, z)$, as their contribution to the output loss is independent of the assignment. Consequently, we express the technology as:

$$\hat{y}(x_1, x_2, z) = -(x_1 + \varphi_{23})(x_2 + \varphi_{13})(z + \varphi_{12}).$$

Let $\hat{z} = z + \varphi_{12}$, $\hat{x}_1 = x_1 + \varphi_{23}$, and $\hat{x}_2 = x_2 + \varphi_{13}$. With these substitutions, the technology function takes its original form $\hat{y}(\hat{x}_1, \hat{x}_2, \hat{z}) = -\hat{x}_1\hat{x}_2\hat{z}$.

Given that all $\varphi_{ij} \geq 0$, the shifted variables remain non-negative and are supported on the intervals $[\varphi_{ij}, 1 + \varphi_{ij}]$.³⁵ Denote the distributions corresponding to these variables respectively as \hat{F}_1 , \hat{F}_2 , and \hat{F}_z . These distributions satisfy Assumption 1 if and only if the density functions f_x and f_z of the original skills satisfy the proposition's stated assumptions. Therefore, invoking Appendix A.2 to Appendix A.4, a mixed and countermonotonic assignment $\hat{\pi} \in \Pi(\hat{F}_1, \hat{F}_2, \hat{F}_z)$ exists and solves the assignment problem in this transformed space. The push-forward image of this assignment π in the space of variables (x_1, x_2, z) provides a solution to the initial problem, completing the proof. \square

The argument in the proof also implies the following corollary.

Corollary. Consider technology $y(x_1, x_2, x_1x_2, x_2z, x_1z, x_1x_2z)$. Suppose its first-order Taylor expansion around the arguments of the function has negative coefficients on each interaction term $(x_1x_2, x_2z, x_1z, x_1x_2z)$ to ensure submodularity. Then a mixed and countermonotonic assignment is optimal.

In sum, our results apply to general functions of interactions (single, pairwise, triple) in the first-order sense. The main insight that allows this generalization is the change of variables.

Generalizing Technology with n Distributions. With n distributions, a somewhat different result applies.³⁶

Consider a general technology $y(x_1, \dots, x_m, z)$ with heterogeneous workers $\{x_1, \dots, x_m\}$. For ease of notation, let the project value z be denoted as $z = x_{m+1} = x_n$. Consider a second-order

³⁵Although the distributions of \hat{x}_1 , \hat{x}_2 , and \hat{z} are no longer confined to the unit interval, this does not impede our proof of existence and optimality. Our proofs neither leverage the specific scaling of these variables within the unit interval, nor depend on it in any essential way. We only require each variable to be non-negative, a condition that is preserved under the transformation.

³⁶The change of variables argument does not extend when there are multiple interactions of order higher than two (such as x_1x_2z or $x_1x_2x_3z$).

Taylor expansion of this technology, which is a second-order polynomial in x_1, \dots, x_n , that can be expressed as:

$$y(x_1, \dots, x_n) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j. \quad (\text{A.28})$$

Specifically, consider the case where the matrix $A = \{a_{ij}\}$ has rank one and is elementwise non-positive to ensure submodularity.³⁷

Proposition 12. Consider production technology (A.28) with $\lambda_i \geq 0$ and $f_i(x_i)$ is non-increasing. Define transformed skills $\hat{x}_i = \exp(-\lambda_i x_i)$, with \hat{F}_{x_i} denoting the corresponding distributions of \hat{x}_i . Then, there exists a mixed and countermonotonic assignment $\hat{\pi} \in \Pi(\hat{F}_{x_1}, \dots, \hat{F}_{x_n})$ such that its push-forward image in the space of original skills (x_1, \dots, x_n) is an optimal solution.

Proof. We disregard the constant and linear terms in technology (A.28) as their contribution to the loss is independent of the assignment. Thus, we focus on the technology:

$$\hat{y}(x_1, \dots, x_n) = - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j x_i x_j = - \left(\sum_{i=1}^n \lambda_i x_i \right)^2.$$

With $\hat{x}_i = \exp(-\lambda_i x_i)$, we write this technology as $\hat{y}(\hat{x}_1, \dots, \hat{x}_n) = - \left(\sum_{i=1}^n \log \hat{x}_i \right)^2$.

First, we verify that each distribution \hat{F}_{x_i} satisfies Assumption 1. Observe that:

$$\hat{f}_i(\hat{x}_i) = \frac{1}{\lambda_i \hat{x}_i} f_i(x_i),$$

which implies that the condition $\hat{x}_i \hat{f}_i(\hat{x}_i)$ is non-decreasing is equivalent to the density function $f_i(x_i)$ being non-increasing. Thus, by invoking Appendix A.3 and Appendix A.4, we deduce the existence of a mixed and countermonotonic assignment $\hat{\pi} \in \Pi(\hat{F}_{x_1}, \dots, \hat{F}_{x_n})$.

In the proof of optimality for a mixed and countermonotonic assignment, we show using the majorization inequality that the distribution of the sum of logarithmic skills under $\hat{\pi}$, that is, the distribution of $\sum \log \hat{x}_i$, is minimal in the sense of second-order stochastic dominance compared to any other assignment π . Formally, for any convex function h and alternative assignment π , the inequality

$$\int h \left(\sum_{i=1}^n \log \hat{x}_i \right) d\hat{\pi} \leq \int h \left(\sum_{i=1}^n \log \hat{x}_i \right) d\pi$$

holds. Taking $h(t) = t^2$, we find that $\hat{\pi}$ maximizes total output of the transformed technology $\hat{y}(\hat{x}_1, \dots, \hat{x}_n) = - \left(\sum \log \hat{x}_i \right)^2$, and the push-forward of assignment $\hat{\pi}$ to the space of original skills (x_1, \dots, x_n) solves the initial sorting problem. \square

³⁷Since the matrix has rank one, each element a_{ij} can be represented as $a_{ij} = -\lambda_i \lambda_j$, where $\lambda_i \geq 0$ for all i .

A.6 Dual Problem

To characterize the solution to the dual problem, we use a generalization of Kantorovich duality for multiple marginal distributions (Kellerer, 1984).³⁸ The duality result states that the optimal value for the planning problem coincides with the optimal value for the dual problem, and that the dual solution exists in the class of integrable functions.

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 $\ell(x_1, \dots, x_m, z)$ be continuous on $X \times \dots \times X \times Z$ and $\Pi(F_x, \dots, F_x, F_z)$ denote the set of measures π on $X \times \dots \times X \times Z$ with fixed projections $\text{pr}_{x_i}(\pi) = F_x$ and $\text{pr}_z(\pi) = F_z$. Then,

$$\min_{\pi \in \Pi} \int \ell(x_1, \dots, x_m, z) d\pi = \max \sum_{i=1}^m \int \hat{w}_i(x_i) dF_x + \int \hat{v}(z) dF_z, \quad (\text{A.29})$$

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 $\sum_{i=1}^m \hat{w}_i(x_i) + \hat{v}(z) \leq \ell(x_1, \dots, x_m, z)$ for any triplet $(x_1, \dots, x_m, z) \in X \times \dots \times X \times Z$.

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 establish that the solution exists in the class of continuous functions.

Lemma 6. For any dual solution (w_1, \dots, w_m, v) , there exist continuous functions $(\hat{w}_1, \dots, \hat{w}_m, \hat{v})$ such that $\hat{w}_i = w_i$ and $\hat{v} = v$ almost everywhere that solves the dual problem.

Proof. For any dual solution (w_1, \dots, w_m, v) we define:

$$\hat{w}_1(x_1) := \sup_{x_2, \dots, x_m, z} y(x_1, \dots, x_m, z) - \sum_{i=2}^m w_i(x_i) - v(z), \quad (\text{A.30})$$

where $\hat{w}_1(x_1) \leq w_1(x_1)$ since $w_1(x_1) \geq y(x_1, \dots, x_m, z) - \sum_{i=2}^m w_i(x_i) - v(z)$. By equation (A.30), we also have $\hat{w}_1(x_1) \geq y(x_1, \dots, x_m, z) - \sum_{i=2}^m w_i(x_i) - v(z)$, or equivalently, $\hat{w}_1(x_1) + \sum_{i=2}^m w_i(x_i) + v(z) \geq y(x_1, \dots, x_m, z)$ so that $(\hat{w}_1, w_2, \dots, w_n, v)$ is indeed a feasible solution to the dual problem (7). Since through $\hat{w}_1(x_1) \leq w_1(x_1)$, it follows that

$$\int \hat{w}_1(x_1) dF_x + \sum_{i=2}^m \int w_i(x_i) dF_x + \int v(z) dF_z \leq \int w_1(x_1) dF_x + \sum_{i=2}^m \int w_i(x_i) dF_x + \int v(z) dF_z.$$

³⁸An instructive exposition of Kellerer (1984) is presented in Rachev and Rüschendorf (1998).

Thus, $(\hat{w}_1, w_2, \dots, w_m, v)$ is a dual solution. Furthermore, $\hat{w}_1 = w_1$ almost everywhere. Suppose not, then instead $\hat{w}_1 < w_1$, which would imply that:

$$\int \hat{w}_1(x_1) dF_x + \sum_{i=2}^m w_i(x_i) dF_x + \int v(z) dF_z < \int w_1(x_1) dF_x + \sum_{i=2}^m w_i(x_i) dF_x + \int v(z) dF_z.$$

which would contradict that (w_1, \dots, w_m, v) is a dual solution.

We next establish that \hat{w}_1 is a continuous function. To show this, let $\hat{w}_1(x_1; x_2, \dots, x_m, z) := y(x_1, \dots, x_m, z) - \sum_{i=2}^m w_i(x_i) - v(z)$, that is, (A.30) holding constant all coworkers (x_2, \dots, x_m) and the firm z . Given the technology (3), and since $x_i \in [0, 1]$ and $z \in [0, 1]$, $\hat{w}_1(x_1; x_2, \dots, x_m, z)$ is a linear function in x_1 with slope less than one in absolute value. This implies $\hat{w}_1(x_1; x_2, \dots, x_m, z)$ is Lipschitz continuous. Since $\hat{w}_1(x_1)$ can be represented as $\hat{w}_1(x_1) = \sup_{x_2, \dots, x_m, z} \hat{w}_1(x_1; x_2, \dots, x_m, z)$, $\hat{w}_1(x_1)$ is Lipschitz continuous as the supremum of uniformly Lipschitz continuous functions is Lipschitz continuous. Hence, $\hat{w}_1(x_1)$ is continuous.

Similarly, we sequentially define $\{\hat{w}_i\}$ and \hat{v} :

$$\hat{w}_i(x_i) := \sup_{x_j \neq i, z} y(x_1, \dots, z) - \sum_{j=1}^{i-1} \hat{w}_j(x_j) - \sum_{j=i+1}^m w_j(x_j) - v(z),$$

$$\hat{v}(z) := \sup_{x_1, \dots, x_m} y(x_1, \dots, z) - \sum_{i=1}^m \hat{w}_i(x_i).$$

By sequentially applying the above argument, we obtain that $(\hat{w}_1, \dots, \hat{w}_m, \hat{v})$ is a dual solution with $\hat{w}_i = w_i$ for all i and $\hat{v} = v$ almost everywhere. Similarly, the same arguments shows that all \hat{w}_i and \hat{v} are continuous. \square

A.7 Proposition 4

For any worker $x_{1,0}$ there exist coworkers $\{x_{j,0}\}_{j=2}^m$ and a firm z_0 so that team $(x_{1,0}, \dots, x_{m,0}, z_0)$ is contained in the support of an optimal assignment, implying that:

$$w(x_{1,0}) + \dots + w(x_{m,0}) + v(z_0) = y(x_{1,0}, \dots, x_{m,0}, z_0) = z_0 + x_{1,0}m_x(x_{1,0}),$$

with the second equality following from the marginal worker product (4). For any other worker x_1 , it must by the constraint be that $w(x_1) + \dots + w(x_{m,0}) + v(z_0) \geq z_0 + x_1m(x_{1,0})$. Combining these expressions, it holds that for any x_0 and $x_1 \in [0, 1]$:

$$w(x_1) - w(x_{1,0}) \geq m(x_{1,0})(x_1 - x_{1,0}). \tag{A.31}$$

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

$$w(x_{1,0}) - w(x_1) \geq m(x_1)(x_{1,0} - x_1). \tag{A.32}$$

Combining (A.31) and (A.32), dividing by $(x_1 - x_{1,0})$ yields

$$m(x_1) \geq \frac{w(x_1) - w(x_{1,0})}{x_1 - x_{1,0}} \geq m(x_{1,0}). \tag{A.33}$$

Taking limits as $x_1 \rightarrow x_{1,0}$, 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 3, the derivative of w is continuous. The result is similarly established for $v'(z)$.

A.8 Quantitative Robustness and Extensions

The quantitative analysis in the main text assumes that teams are of size three – two workers and a project. In this appendix, we demonstrate that our model can also account for the decomposition of earnings variation within and across firms with teams of sizes four, five, and six. We also find that the non-targeted average coworker earnings are robust to team size. As the number of team members increases, the firm and worker effect become more muted. These results are summarized in Table A.1, Table A.2, and Table A.3.

Table A.1: Parametric Identification with Three Workers

Model and Data Earnings Decomposition						
Moment	Data			Model		
	1981	2013	change	1981	2013	change
Between	0.34	0.42	0.08	0.34	0.42	0.08
Within	0.66	0.58	-0.08	0.66	0.58	-0.08

Model and Data Coworker Earning						
Percentile	Data			Model		
	1981	2013	change	1981	2013	change
25	10.01	10.07	0.06	10.19	10.28	0.09
50	10.30	10.45	0.15	10.44	10.61	0.17
75	10.56	10.75	0.19	10.56	10.79	0.23
90	10.67	10.98	0.31	10.55	10.82	0.27

Model and Data Earnings Decomposition						
Moment	Model		Firm Effect		Worker Effect	
	1981	2013	2013	change	2013	change
Between	0.34	0.42	0.24	-0.10	0.52	0.18
Within	0.66	0.58	0.76	0.10	0.48	-0.18

Table A.2: Parametric Identification with Four Workers

Model and Data Earnings Decomposition						
Moment	Data			Model		
	1981	2013	change	1981	2013	change
Between	0.34	0.42	0.08	0.34	0.42	0.08
Within	0.66	0.58	-0.08	0.66	0.58	-0.08

Model and Data Coworker Earning						
Percentile	Data			Model		
	1981	2013	change	1981	2013	change
25	10.01	10.07	0.06	10.24	10.28	0.04
50	10.30	10.45	0.15	10.45	10.62	0.17
75	10.56	10.75	0.20	10.52	10.77	0.25
90	10.67	10.98	0.31	10.50	10.79	0.29

Model and Data Earnings Decomposition						
Moment	Model		Firm Effect		Worker Effect	
	1981	2013	2013	change	2013	change
Between	0.34	0.42	0.22	-0.12	0.49	0.15
Within	0.66	0.58	0.78	0.12	0.51	-0.15

Table A.3: Parametric Identification with Five Workers

Model and Data Earnings Decomposition						
Moment	Data			Model		
	1981	2013	change	1981	2013	change
Between	0.34	0.42	0.08	0.34	0.42	0.08
Within	0.66	0.58	-0.08	0.66	0.58	-0.08

Model and Data Coworker Earning						
Percentile	Data			Model		
	1981	2013	change	1981	2013	change
25	10.01	10.07	0.06	10.20	10.28	0.08
50	10.30	10.45	0.15	10.43	10.62	0.19
75	10.56	10.75	0.19	10.52	10.76	0.24
90	10.67	10.98	0.31	10.51	10.79	0.28

Model and Data Earnings Decomposition						
Moment	Model		Firm Effect		Worker Effect	
	1981	2013	2013	change	2013	change
Between	0.34	0.42	0.38	0.04	0.38	0.04
Within	0.66	0.58	0.62	-0.04	0.62	-0.04