Superstar Teams

Lukas B. Freund[§]

December 2024

Please click here for the latest version.

Abstract

In benchmark models of productivity and labor market inequality, workers are assumed perfectly substitutable, ignoring potential coworker interdependencies. I propose a novel and tractable theory in which production takes place in teams and requires many differentiated tasks; workers possess heterogeneous task-specific skills; and labor markets are frictional. I show analytically that skill specificity (dispersion in task-specific skills) endogenously generates coworker talent complementarities in production: talented workers gain disproportionately from talented colleagues. This promotes the concentration of talent into a few firms with "superstar teams," though search frictions prevent perfect sorting. Using German linked employer-employee panel data, I document industry-level patterns consistent with this mechanism and calibrate the model. Growing skill specificity since the mid-1980s has amplified coworker complementarities, explaining a significant share of the observed rise in between-firm wage inequality. Further, productivity gains from increased specialization hinge on enhanced matching of coworkers with complementary skills.

Keywords: firms, inequality, matching, productivity, specialization, teams

[§]Columbia University. *Email*: lukas.beat.freund@gmail.com. *Web*: www.lukasfreund.com.
I am grateful to Vasco Carvalho for his invaluable insights, extensive feedback and support. In addition, I thank Zsófia Báráy (discussant), Susanto Basu, Florin Bilbiie, Giancarlo Corsetti, Maarten De Ridder, Wouter Den Haan, Luis Garicano, John Grigsby, Kyle Herkenhoff, Tomer Ifergane, Lukas Mann, Simon Mongey, Christian Moser, Ezra Oberfield (discussant), Sergio Ocampo Díaz, Charles Parry, Tommaso Porzio, Richard Rogerson, Esteban Rossi-Hansberg, Bastian Schulz (discussant), Isaac Sorkin, and Gianluca Violante for helpful conversations and comments. I also thank seminar participants at BI Oslo, Bocconi, Boston College, Columbia, CREi/UPF, Edinburgh, Federal Reserve Bank of San Francisco, IIES, KU Leuven, LMU, Michigan, Notre Dame, OECD, Princeton, Toronto, UCL, UNC Chapel Hill, Uppsala, and Western for useful comments and suggestions; and the organizers and participants at the Annual Meeting of the CEPR Macroeconomics and Growth Programme 2023, BSE Summer Forum 2024, Bristol Macroeconomics Workshop 2024, CRC TR 224 Workshop on Labor Markets 2023, EES Workshop on 'New Developments in the Macroeconomics of Labor Markets' 2024, Minneapolis Fed Junior Scholar Conference 2023, NBER SI 2024 (Macro Perspectives), NBER Organizational Economics Workshop Fall 2024, and SED 2024. I gratefully acknowledge financial support from Gates Cambridge Trust (BMGF OPP1144), the Keynes Fund (JHVB) and the Procter Fund. This study uses the Sample of Integrated Employer-Employee Data (SIEED 7518) and the Linked Employer Employee Data longitudinal model (LIAB LM7519) from the German Institute for Employment Research (IAB). Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the IAB and subsequently remote data access under project numbers fdz2188/2491/2492.
First version: December 8, 2022.

1 Introduction

As production grows more complex, no individual can master all required tasks; hence, most people work in teams. Consider engineers designing a car engine. This example highlights two fundamental features of production. First, coworkers possess heterogeneous, task-specific skills: one might excel in materials design, another in electronics, and they may also differ in overall ability. Second, there is "team production," where joint output exceeds the sum of individuals' contributions: the engineer stronger in materials can better leverage her expertise when paired with an electronics specialist of similar caliber. These features underpin a classic view of the firm, according to which its central role is to organize the division of labor among heterogeneously skilled workers.[1]

This paper aims to understand how these fundamental features of production – multi-dimensional skill heterogeneity and teams – shape macroeconomic outcomes. For instance, how does growing task complexity and skill specialization (Jones, 2009; Deming, 2017) affect labor market inequality and aggregate productivity? These questions are challenging to address because benchmark macroeconomic models typically abstract from skill specialization and assume production functions additively separable in labor.[2]

I propose a task-based, aggregative theory of team production with heterogeneously skilled workers. The core insight is that skill specificity both makes division of labor advantageous and endogenously generates production complementarities across coworkers' talents, i.e., the marginal contribution of one employee's talent to output is greater when matched with other talented workers. This matters for macro-level outcomes for two reasons: It promotes the concentration of talent into select workplaces – for instance, top engineers collaborating alongside leading designers – and thereby fosters firm-level inequality; and productivity gains from specialization are realized only if workers are matched with coworkers possessing complementary skills.

The analysis proceeds in three main steps. First, I build and characterize an equilibrium theory of teams, centered on a task-based microfoundation for coworker production complementarities. Second, I confront this model with German longitudinal matched employer-employee data, testing key mechanisms and calibrating the model to gauge their relevance. Third, I use the model to show that growing skill specificity since the

[1] My description of "team production" follows Alchian and Demsetz's (1972) classic definition. Other important references include Becker and Murphy (1992), Kremer (1993) and Garicano (2000).

[2] Important advances focus on either coworker learning (Jarosch *et al.*, 2021; Herkenhoff *et al.*, 2024), task-based production (e.g. Acemoglu and Restrepo, 2018), or multi-dimensional skills (Lindenlaub, 2017).

mid-1980s has amplified complementarities, leading to an equilibrium featuring some firms with "superstar teams," who are jointly very productive and command high average wages, alongside other firms with less productive, lower-paid employees. The theory thus explains a significant share of the so-called "firming up of inequality," i.e., the increased role of firm-level wage inequality (Card *et al.*, 2013; Song *et al.*, 2019). Moreover, the interaction between labor market frictions and endogenous complementarities meaningfully limits aggregate productivity gains from increased specialization.

Theory. The theory has four core features. First, production involves a continuum of imperfectly substitutable tasks. Second, workers have limited time and possess heterogeneous, task-specific skills. Third, production occurs in teams. Ex-ante identical firms hire multiple workers – similar to Herkenhoff *et al.* (2024), I mostly study teams of size two – and assign them tasks to maximize output. Fourth, team assembly involves search frictions: firms sequentially and randomly encounter unmatched workers, observe their skills, and decide whether to hire them or continue searching. This generates a trade-off between team match quality and the opportunity cost of waiting.

The key analytical step is to microfound a tractable, aggregative production function, characterizing output for any combination of team members' skills under optimal task assignment. Tractability comes from modeling skills as realizations of an individual-specific Fréchet distribution, following trade models à la Eaton and Kortum (2002) and Lind and Ramondo (2023). I show that this assumption yields a low-dimensional representation of workers' joint skill distribution across tasks in terms of three objects: individual *talent* determines a worker's average skill across tasks; an economy-wide *skill specificity* parameter controls the dispersion of individual task-specific skills; and a scalar Copula term governs the *horizontal distance* between any two workers, i.e. how correlated their task-specific skills are. Conditional on optimal assignment – which accords with standard comparative-advantage logic – we can aggregate across the task continuum. This yields a reduced-form production function that is tractable, talents and horizontal distance being sufficient statistics for output, and has intuitive properties.

This microfoundation reveals that potential gains from team production increase with skill specificity, but realizing these gains becomes increasingly sensitive to team composition. Put succinctly, output is higher when team members are similar in talent level but diverse in the specific tasks they are best at. Formally, output is a constant-elasticity-of-substitution (CES) function of talents, featuring complementarities across coworkers' talent levels.[3] The strength of complementarities – measured by the CES elasticity term – endogenously increases with skill specificity, rather than being a "deep" structural parameter. The intuition? Suppose a worker must perform a larger share of tasks due to less capable colleagues. Absent skill specificity, this does not matter for realized productivity. But the more task-specific skills are, the greater the opportunity cost of handling tasks they are relatively slow at. This productivity loss is larger in absolute magnitude for more talented workers, whose potential output is greater by definition.

At the macro level, skill specificity then fosters an equilibrium with talent sorting: the workforce of some firms is composed of highly talented workers, while other firms mostly employ less talented ones. The reason is that, because of talent complementarities, firms who already have talented employees can always outbid competitors for talented candidates. As some degree of mismatch is accepted due to search frictions, stronger skill specificity, making it more profitable to be selective, generates more sorting.

Measurement. To evaluate these predictions, I confront the theory with administrative matched employer-employee panel data for Germany. I treat each establishment as one team and exploit the panel dimension to recover workers' time-invariant talent types from their wage histories. As the degree of skill specificity is not directly measurable, I develop a theory-guided method to indirectly infer this parameter from wage variation induced by coworker changes over time. A noteworthy aspect of my approach is that the model can be calibrated without targeting firm-level dispersion in worker talent or wages. Those instead serve as yardsticks to gauge the model's performance.

Given the estimated degree of skill specificity for the years 2010-2017, the calibrated model matches (untargeted) talent sorting patterns. The model also qualitatively reproduces quasi-experimental evidence on the heterogeneous wage effects of coworker deaths (Jäger and Heining, 2022). To test core model mechanisms more directly, I examine variation across industries. Industries which involve more complex tasks – a proxy for skill specificity – like architecture or the manufacturing of chemicals exhibit stronger coworker talent complementarities and more pronounced talent sorting than industries where routine tasks dominate, like accommodation services or furniture manufacturing.

Applications. Theory and data jointly reveal that growing skill specificity carries important macroeconomic implications. A rise in skill specialization has been discussed across several field, supported by suggestive evidence pointing to an economy-wide

[3] Technically, "coworker talent complementarity" (or supermodularity) refers to a positive cross-partial derivative of the production function with respect to own talent and coworker talent.

shift from routine to complex, cognitive non-routine tasks (Deming, 2017) – possibly driven by advances in IT (Bartel *et al.*, 2007) – and a growing "burden of knowledge" that necessitates a narrowing of individual expertise (Jones, 2009).

I find that growing skill specificity since the mid-1980s explains a significant share of the "firming up of inequality" (Card *et al.*, 2013; Song *et al.*, 2019). Structural estimation of the skill specificity parameter across time shows an increase, consistent with the suggestive evidence. Under my preferred specification, the model predicts a 21 percentage point increase in the between-establishment share of the log wage variance from 1985-1992 to 2010-2017, close to the empirically observed rise of 24 percentage points. Counterfactual exercises show that greater skill specificity can account for around 65% of the model-predicted increase. The same mechanism also helps explain growing firm-level productivity dispersion (Berlingieri *et al.*, 2017; Autor *et al.*, 2020).

Furthermore, I show how team production interacts with labor market frictions to shape aggregate productivity in the context of increasing skill specialization. While such specialization directly boosts productivity, this comes with a catch: as complementarities are endogenously amplified, productivity becomes more sensitive to coworker mismatch. I find that less than half of the gains from increased skill specificity may be realized unless accompanied by enhanced matching of coworkers with complementary skills.

Discussion. Overall, the paper offers a framework that links the organization of work around teams of workers with heterogeneous, specialized skills to macroeconomic outcomes. The model is parsimonious, so the quantitative analysis is geared toward illustrating the theory's usefulness rather than offering precise point estimates. I close by discussing how relaxing several simplifying assumptions would allow addressing further questions, such as the role of team quality for fast-growing firms (Sterk *et al.*, 2021).

Related literature. An emerging literature studies how coworker interactions shape macroeconomic outcomes. Notably, Jarosch *et al.* (2021) and Herkenhoff *et al.* (2024) examine coworker learning as a source of human capital accumulation and identify large aggregate consequences of structural shifts that affect the degree of talent sorting in teams.[4] My paper is complementary as it omits learning spillovers and instead provides a richer account of the determinants of team composition. In particular, relative to Herkenhoff *et al.* (2024), which the matching block of my model builds on, I make three contributions. First, I allow skills to be task-specific instead of one-dimensional. Second, in spite of this added complexity, I provide an analytical microfoundation for a team

[4]Also see Akcigit *et al.* (2018), Chade and Eeckhout (2020), Boerma *et al.* (2021), Pearce (2022).

production function. This reveals that absent skill specificity there is no gain from team production in the first place, and the strength of complementarities is endogenous. Third, I document empirical support for the proposed mechanism and use the model to study the effect of growing skill specificity on talent sorting and aggregate productivity.

The paper builds on seminal theories of firm production, notably Kremer (1993) and Garicano (2000).[5] While sharing Kremer's (1993) emphasis on coworker complementarities, this paper differs in three ways. First, rather than assuming complementarities to increase mechanically with team size, this paper microfounds a production function where complementarities increase in skill specificity for fixed team size. Second, it abstracts from increasing returns to team quality, which drives many key results in Kremer (1993, see discussion on p. 571), thus isolating complementarity effects. Third, introducing search enables study of changes in sorting patterns, whereas in Kremer (1993) sorting is perfect for any degree of coworker complementarity. In the spirit of Garicano (2000) and Garicano and Rossi-Hansberg (2006), I explicitly model the organizational problem underlying a production function. However, while the theory of knowledge hierarchies focuses on vertical relationships – how better workers enable larger spans of control – this paper examines "horizontal" complementarities with fixed span of control.

Technically, I use building blocks from the task-assignment literature, notably Acemoglu and Restrepo (2018) and Ocampo (2022), and international trade. The key contribution is to show how Eaton and Kortum's (2002) stochastic formulation of technological heterogeneity in terms of extreme-value distributions, generalized by Lind and Ramondo (2023), can be leveraged to aggregate across a continuum of tasks given multi-dimensional skill heterogeneity.[6] This yields a parsimoniously parametrized model bridging task assignment with macro models relying on a tractable production function.

The paper also contributes to a theoretical literature on worker-firm dynamics.[7] While this literature typically assumes a production function additively separable in labor, calls for richer, "empirically implementable theories of the firm" are common (Hagedorn *et al.* (2017, p.33); Lentz and Mortensen (2010, pp. 593, 598)). This paper contributes such

[5]More broadly, the paper relates to studies of firm organization, including Lucas (1978), Sattinger (1993), Kremer and Maskin (1996), Saint-Paul (2001) and Gabaix and Landier (2008); and, more recently, Porzio (2017), Adhvaryu *et al.* (2020), Caliendo *et al.* (2020), Tian (2021), Adenbaum (2022), Bloesch *et al.* (2022), Kohlhepp (2022), Minni (2022), Kuhn *et al.* (2023) and Bassi *et al.* (2023).

[6]Deming's (2017) study of social skills inspired the analogy to trade, but without offering a tractable team production function or analyzing coworker complementarities and equilibrium matching.

[7]Recent advances include Bilal and Lhuillier (2021); Bilal *et al.* (2022); Elsby and Gottfries (2022); Engbom and Moser (2022); Gouin-Bonenfant (2022); Haanwinckel (2023).

a theory. This theory emphasizes multi-dimensional skills, echoing Lindenlaub (2017), Lindenlaub and Postel-Vinay (2023), Baley *et al.* (2022) and Grigsby (2023), but introduces them into a team production setting. Substantively, my findings provide independent evidence for the importance of multi-dimensional skills.

Finally, the application to wage inequality connects two lines of research: the long-standing focus on how changes in work affect skill prices,[8] and more recent work that foregrounds the significant and growing role of firms in explaining wage dispersion.[9] However, structural models that explain this latter shift are lacking. This paper contributes a theory that offers an enriched conception of firm-level production, whereby deepening skill specificity induces stronger firm-level complementarities, which helps rationalize the growing importance of between-firm inequality through increased talent sorting.

Outline. Section 2 presents and characterizes the theory, Section 3 confronts it with data, Section 4 presents applications, and Section 5 concludes. Additional material, including proofs, is contained in an *Online Appendix* appended to this manuscript.

2 Theoretical model

This section develops the theoretical model. Section 2.1 lays out the environment, Section 2.2 derives and characterizes a tractable, reduced-form team production function, and Section 2.3 analyzes how teams are formed in equilibrium given this production function.

2.1 Environment

Time is continuous, runs forever, and we consider the economy in steady state.

Demographics. The economy is populated by workers and firms. All agents are infinitely-lived and have risk-neutral preferences over a single final good discounted at rate $\rho \in (0, 1)$. There is a unit measure of workers who are either employed or unemployed. Each is characterized by a continuum of time-invariant, task-specific skills, $\{z_i(\tau)\}_{\tau\in[0,1]}$, where i indexes workers and τ indexes tasks. It is convenient to denote, with some abuse of notation, $\{z_i(\tau)\}_{\tau\in[0,1]}$ as $\boldsymbol{z}_i$. These skills are i.i.d. draws from a worker-specific distribution, described below. When employed, a worker inelastically supplies a finite amount of time

[8] See, e.g., Autor *et al.* (2003), Spitz-Oener (2006), Atalay *et al.* (2020), Autor *et al.* (2022).

[9] See, e.g., Card *et al.* (2013), Barth *et al.* (2016), Helpman *et al.* (2017), Song *et al.* (2019), Haltiwanger and Spletzer (2020), Criscuolo *et al.* (2021), Håkanson *et al.* (2021), Engbom *et al.* (2023).

in exchange for a wage. While unemployed, a worker receives a flow utility that may vary with skills. There is a fixed unit measure of ex-ante homogeneous firms that are either idle or actively producing in a match with one or more workers.

Task-based production technology. Production of the final good requires that many different tasks be carried out. I have a granular notion of tasks in mind. Producing economics research, for instance, might involve coming up with an idea, writing down a model, solving it, collecting and analyzing data, writing a paper, hopefully publishing it, and so on. To produce, an active firm j employs a discrete number $n \in \{1, 2\}$ of workers – so a "team" consists of two workers. Workers perform a unit continuum of imperfectly substitutable tasks, $\mathcal{T} = [0, 1]$, that are combined into the final good according to

$$Y_j = \left(\int_{\mathcal{T}} q_j(\tau)^{\frac{\eta-1}{\eta}} d\tau \right)^{\frac{\eta}{\eta-1}}, \tag{1}$$

where Y_j is the quantity of the final good produced by j, $\eta \in (0, \infty)$ is the constant task elasticity of substitution, and $q_j(\tau)$ denotes the amount of task τ used, which is given by

$$q_j(\tau) = \sum_{i=1}^{n} y_i(\tau). \tag{2}$$

Here, $y_i(\tau)$ indicates how much of τ worker i has produced, which depends linearly on i's skill at task τ and the time dedicated by i to τ:

$$y_i(\tau) = l_i(\tau) z_i(\tau). \tag{3}$$

Equation (2) captures the idea that division of labor is possible, in that $q_j(\tau)$ can be positive even as $y_i(\tau) = 0$ as long as $y_k(\tau) > 0$ for $k \neq i$.

Labor market matching and information structure. Firms and workers meet through random search. Specifically, unemployed workers contact firms with a vacant position (i.e., $n < \bar{n}$) at an exogenous Poisson rate λ_u. After meeting– but not before – the firm and the worker can observe and contract on the task-specific skills. The parties decide whether to form a match or, instead, to keep on searching. Conditional on a match having been formed, the wage is determined through generalized Nash bargaining, the worker's bargaining power being $\omega \in [0, 1]$. Further details are relegated to Section 2.3.1; for now it suffices to say that this assumption ensures that the wage will be monotonically increasing in firm-level final output and all matching decisions are privately efficient. In an existing match, a worker i is separated from their employer at an exogenous Poisson rate δ_i that potentially varies with skills.

Distribution of task-specific skills. For any individual worker i, the task-specific draws $\{z_i(\tau)\}_{\tau\in[0,1]}$ are i.i.d realizations of a random variable, Z_i, that is Frechet distributed with a worker-specific scale parameter x_i and a common inverse shape parameter χ. Moreover, considering any arbitrary two workers, the correlation structure across their skills is governed by a single (Gumbel-Hougaard) copula parameter ξ.[10]

Assumption 1 (Multivariate Fréchet distribution)**.** *The distribution of worker-task specific skills satisfies*

$$Pr[Z_1(\tau) \leq z_1, Z_2(\tau) \leq z_2] = \exp\left[-\left(\sum_{i=1}^{n=2}\left(\left(\frac{z_i}{\iota x_i}\right)^{-\frac{1}{\chi}}\right)^{\frac{1}{\xi}}\right)^{\xi}\right],$$

where x_i is a worker-specific scale parameters, $\chi \in (0,\infty)$ is the common inverse shape parameter of the marginal Fréchet distributions, $\xi \in (0,1]$ controls dependencies in draws across workers, $\iota := \Gamma\left(1+\chi(1-\eta)\right)^{\frac{1}{1-\eta}}$ is a worker-invariant scaling term, with Γ denoting the Gamma function, and it is assumed throughout that $1+\chi(1-\eta) > 0$.

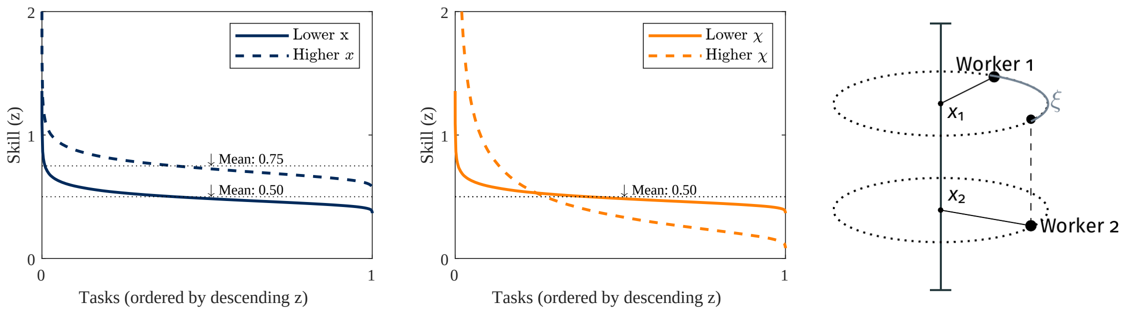
Under Assumption 1, a low-dimensional and easily interpretable set of parameters governs how workers vary in their *average* task-specific skill (x_i or "talent"); conditional on that average, the *dispersion* of skills across different tasks (χ or "skill specificity"); and, for any two workers, how (dis-)similar they are in what tasks they are relatively skilled at (ξ or "horizontal distance"). To illustrate the role of x and χ, Figures 1 (a) and (b) depict the realizations of z's for some i across the unit interval, ordering the otherwise unordered tasks by descending z.[11]

Remark 1 (Interpreting x_i and χ)**.**

- *The worker-specific scale term x_i denotes a worker's talent, or absolute advantage. Intuitively, a high-x worker will be more productive at the task they are best in than*

[10]As discussed in the introduction, this parametric assumption mirrors the Eaton and Kortum (2002) approach in international trade literature, and, in allowing for correlation, Lind and Ramondo (2023).

[11]A couple of technical remarks: First, throughout the paper it is assumed that $1+\chi(1-\eta) > 0$. Beyond requiring that tasks not be too substitutable, under the maintained assumptions the exact value of η will not influence worker-task assignment or team productivity. Second, the scaling term ι ensures that varying χ or η does not mechanically change skill levels.



(a) x_i: talent **(b)** χ: specificity **(c)** ξ: hor. distance

Figure 1: Illustration of parametrization of worker-task specific skills

Notes. This figure illustrates the properties of the multivariate Fréchet distribution. For this illustration, $\eta = 2$, so that $\iota = (\Gamma(1-\chi))^{-1}$, and hence $\mathbb{E}[Z_i] = x_i$ for $\chi < 1$.

a low-x will be at their best task respectively. Formally, $E[z_i(\tau)]$ is increasing in x_i. For ease of language, I will sometimes refer to a worker i by their talent type.

- *The parameter χ controls the degree of skill specificity, which is defined by the degree of within-worker dispersion in task-specific skills — is a given worker similarly skilled across all tasks (low χ) or good at some tasks but less so at others (high χ). In Figure 1(b), for a greater value of χ i's skill diminishes more rapidly as we move away from the task i is best at. For $\chi \longrightarrow 0$ we get closer to one-dimensional worker types.*

The description of the supply of skills in the economy is completed as follows.

Assumption 2 (Population distribution)**.** *In the population, x is distributed according to a cumulative distribution function $\Psi(x)$. Conditional on x, workers are uniformly distributed across a circle with unit circumference; the distance along the circle between any pair of workers corresponds to their horizontal distance ξ.*

Regarding part (i) of Assumption 2, instead of $x \in \mathbb{R}_+$, it will frequently be useful to work with the rank $\hat{x} = \Phi(x)$, i.e. $\hat{x} \in [0,1]$. Regarding part (ii), the interpretation and language implicitly presume that there is a continuum of every talent type. Associated to it is the following interpretation of ξ.

Remark 2 (Interpreting ξ)**.** *For any two workers i and j, ξ governs the horizontal distance across them – roughly, the degree of correlation in task-specific skills; strictly, the Kendall tau distance. If $\xi = 1$, draws are independent across workers; decreasing ξ means i and j become more similar in what tasks they are skilled in; and for $\xi \longrightarrow 0$ relative skill across any two tasks $\frac{z_i(\tau)/z_i(\tau')}{z_j(\tau)/z_j(\tau')} \longrightarrow c$ for some constant c.*

Looking ahead, χ will be treated as a structural parameter throughout, whereas worker talents $\{x_i\}_{i=1}^n$ and ξ will be endogenous to team formation.

Discussion. The theory has four central features: (i) workers have heterogeneous, multi-dimensional skills for tasks; (ii) production requires many imperfectly substitutable tasks to be carried out; (iii) production can involve multiple, but not infinitely many, workers; and (iv) hiring involves search, reflecting an information friction whereby skills are observable only upon meeting. The parametric structure imposed on (i) through Assumption 1 will be key for tractability, but the economics are more general; and the search structure (iv) introduces a panel dimension that will be important for measurement.

These features give rise to two sets of trade-offs. First, (i)-(iii) imply there is scope for gains from the division of labor, but each worker will also have to perform some tasks other than their best, so intuitively output should be sensitive to a team's composition. How much can different combinations of workers produce together – what is the production function? Second, who ends up working with whom in equilibrium when team composition matters for productivity but, due to the presence of search (iv), any productivity gains need to be traded off against the opportunity cost of time.

I analyze these trade-offs in two steps. First, I consider a single firm with an arbitrary set of workers. Optimal task assignment and Assumption 1 jointly yield an intuitively interpretable and tractable *reduced-form* production function that features a CES structure in coworker talents. The substantive point is that the parameters of that function are not "deep," instead the theory delivers testable predictions for how these parameters vary across sectors and time. Second, with this tractable representation at hand, I solve for the equilibrium assignment of workers to different firms.

Before moving forward, it is worth laying out important simplifications. First, there is no ex-ante heterogeneity across firms. All productive knowledge is embodied in workers, and a firm simply consists of an organized collection of workers; its value is tightly linked to their production capacity. While stark, this assumption allows showing transparently that observed ex-post firm heterogeneity can emerge solely due to coworker interdependencies. Second, there is no explicit span-of-control problem, such as a hierarchy à la Garicano (2000), nor multi-team firms or any agency problems. Third, tasks are symmetric – none is inherently more valuable or difficult than another – nor are different skills more or less scarce in the economy. Finally, a maximum team size of two is restrictive but sufficient to study team production and formation, and imparts significant tractability: The firm's state space expands combinatorically with team size when production features non-separabilities in labor even with single-dimensional skills; and under Assumption 1 we can summarize the correlation between any two workers' high-dimensional skill vectors in terms of the scalar ξ, i.e., the horizontal distance.

2.2 Team production

Consider a single, active firm that employs $n > 0$ workers (so I drop the j subscript here). Treating the skill composition as exogenous, how much final output can be produced?

2.2.1 Organizational problem

The firm solves a "mini-planner problem" by choosing total task usage $\{q(\tau)\}_{\tau\in\mathcal{T}}$, individual task production $\{\{y_i(\tau)\}_{\tau\in\mathcal{T}}\}_{i=1}^n$ and individual time allocation $\{\{l_i(\tau)\}_{\tau\in\mathcal{T}}\}_{i=1}^n$ to maximize total production Y, given $\{z_1, ..., z_n\}$ and subject to equations (1)-(3) as well as a time constraint for each worker i:[12]

$$1 = \int_{\mathcal{T}} l_i(\tau) d\tau. \tag{4}$$

Remark 3 (Reduced-form production function)**.** *The reduced-form production function f is defined as the solution to the problem of optimally assigning tasks,*[13]

$$\begin{aligned} f(\cdot) &= \max Y \\ & s.t. \ (1) - (4). \end{aligned} \tag{5}$$

The Lagrangean associated to the firm's problem is

$$\begin{aligned} \mathcal{L}(\cdot) = Y + \lambda\left[\left(\int_{\mathcal{T}} q(\tau)^{\frac{\eta-1}{\eta}} d\tau\right)^{\frac{\eta}{\eta-1}} - Y\right] + \int_{\mathcal{T}} \lambda(\tau)\left(\sum_{i=1}^{n} y_i(\tau) - q(\tau)\right)d\tau \\ + \sum_{i=1}^{n}\left\{\lambda_i^L\left(1 - \int_{\mathcal{T}} l_i(\tau)d\tau\right) + \int_{\mathcal{T}} \lambda_i(\tau)\left(z_i(\tau)l_i(\tau) - y_i(\tau)\right)d\tau + \int_{\mathcal{T}} \bar{\lambda}_i(\tau)y_i(\tau)d\tau.\right\}, \end{aligned}$$

Here, λ, λ_i^L, $\lambda_i(\tau)$ and $\lambda(\tau)$ denote the shadow values of, respectively, total production, i's time, a unit of task τ produced by i, and a unit of task τ used in final good production,

[12]Assuming each worker to be endowed with one unit of time is without loss of generality, even if technically we are in a continuous-time context, as this simply scales overall output.

[13]I am deliberately loose in specifying the domain of f for now, as Proposition 1 will clarify this point.

while $\bar{\lambda}_i(\tau)$ attaches to a non-negativity constraint in task production. Importantly, each worker's time is scarce, with multiplier λ_i^L capturing the opportunity cost of i's time.

First we derive the demand for task, treating the shadow prices as known. Taking the first-order condition (FOC) for $q(\tau)$, defining $Q(\tau) := \lambda(\tau)q(\tau)$, and standard CES algebra yields expressions for task demand and the shadow cost index λ:

$$Q(\tau) = \left(\frac{\lambda(\tau)}{\lambda}\right)^{1-\eta} \lambda Y, \tag{6}$$

$$\lambda = \left(\int_{\mathcal{T}} \lambda(\tau)^{1-\eta} d\tau\right)^{\frac{1}{1-\eta}}. \tag{7}$$

Optimal task usage thus depends on task-specific shadow costs, $\{\lambda(\tau)\}_{\tau\in\mathcal{T}}$, which in turn depend on how workers are assigned to tasks. The FOC with respect to $y_i(\tau)$ implies that $\lambda(\tau) = \lambda_i(\tau)$ if $y_i(\tau) > 0$. Since some worker will provide a given task τ, and with task production featuring constant returns to scale, cost-minimization requires that

$$\lambda(\tau) = \min_{i=1,\ldots,n} \left\{\frac{\lambda_i^L}{z_i(\tau)}\right\}, \tag{8}$$

where I used the FOC for $l_i(\tau)$ to substitute for $\lambda_i(\tau)$. Thus, any task τ will optimally be performed by the worker with the lowest ratio of the task-invariant shadow cost of their time over their task-specific skill; and the firm's shadow cost for that task, $\lambda(\tau)$, is equal to that minimum. The optimal assignment thus features complete division of labor – every worker performs the interval of tasks in which she has a comparative advantage, but no strictly positive mass of tasks is performed by more than one worker – and satisfies textbook comparative-advantage logic.

A second look at equation (8) clarifies that taking a step further to derive f is non-trivial. Doing so requires knowledge of the entire distribution of the minimum of the worker-task specific shadow costs. This object itself depends on each worker's shadow cost of time, λ_i^L, which in turn is an endogenous function of $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$.[14]

Assumption 1, and specifically the max-stability property of the Fréchet distribution, allows overcoming this challenge, as shown by Eaton and Kortum (2002) in a trade context.

[14] Observe that the first-order condition for $l_i(\tau)$ together with time constraint (4) and the condition that $\lambda(\tau) = \lambda_i(\tau)$ if $y_i(\tau) > 0$, allow writing $\lambda_i^L = \int \lambda(\tau)z_i(\tau)l_i(\tau)d\tau$. Thus, i's shadow cost of labor depends on which tasks i is assigned to and how much of each i produces, which in turn depends on i's task-specific skills as well as those of everyone else on the team, per equation (8).

Exploiting this property, Lemma A.3 in the Online Appendix summarizes the (optimal) share of tasks assigned to each worker and the shadow cost of the final good, λ. In the next section I characterize the reduced-form production f resulting from this assignment.

2.2.2 Characterizing the reduced-form production function

We are now in a position to analytically characterize the production function f, defined in equation (5), which summarizes how skills map onto final output Y in reduced form under the optimal task assignment. To preview the main insights: The more task-specific skills are, the greater the potential gains from team production, and the extent to which these gains are realized depends on the team's skill composition in two ways that neatly map onto 'vertical' (talent) and 'horizontal' composition. First, coworkers' *talents* are complements in production, so talented workers gain more from more talented colleagues; equivalently, and other things equals, dispersion in team members' talent reduces output. Second, production is greater when coworkers are good at different tasks, i.e. they are *horizontally distant.* Crucially, explicitly microfounding f shows the strength of these team effects to be endogenously increasing in χ, the skill specificity parameter.

Aggregating over tasks and team members under the optimal assignment yields:[15]

Proposition 1 (Reduced-form production function)**.** *Team output Y can be written as a function of members' talent types, $\boldsymbol{x}$, and horizontal distance, ξ, given parameter χ:*

$$f(\boldsymbol{x}, \xi; \chi) = n^{1+\chi\xi} \left(\frac{1}{n} \sum_{i=1}^{n} (x_i)^{\frac{1}{\chi\xi+1}} \right)^{\chi\xi+1} . \tag{9}$$

Proof. See Appendix A.1.3. The proof uses the normalization $\lambda = 1$ and Lemma A.3. □

Methodologically, f is a reduced-form, "aggregate" (firm-level) production function.[16] Given parameter χ, talents and horizontal distance, $(\boldsymbol{x}, \xi)$, are sufficient statistics for output Y, without knowledge of the underlying, micro-level task assignment being re-

[15] Output is deterministic dsepite Assumption 1, as we are integrating over a task continuum.

[16] This aggregation result relates to the seminal paper of Houthakker (1955) and, more recently, Acemoglu and Restrepo (2018). In IO, an important related reference is Anderson *et al.* (1989). In independent work, Dvorkin and Monge-Naranjo (2019) derive a similar microfounded CES function, though their analysis is less general by assuming $\xi = 1$ by construction.

quired.[17] Inspecting equation (9) implies:[18]

Corollary 1. *Let the inverse elasticity of substitution ("elasticity of complementarity") be defined as* $\gamma_{ij} := \frac{\partial \ln(f_j/f_i)}{\partial \ln(x_i/x_j)}$*, for any homothetic production function* $f(x_1, ..., x_n; \cdot)$*, where* $f_i = \partial f / \partial x_i$ *denotes a partial derivative. For the production function* (9)*, the elasticity of complementarity is symmetric and identical for all pairs of workers and equal to* $\gamma = \frac{\chi\xi}{1+\chi\xi}$.

Turning to an interpretation, in equation (9), output is written as the product of two terms that carry intuitive interpretations. Setting ξ, the (endogenous) horizontal coworker distance, to unity for a moment, the first term summarizes the potential efficiency gains from team production relative to a counterfactual in which every worker produces all tasks. It is straightforward to show that in the latter case, $Y = n \times \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$. That χ appears in the exponent of the first term of equation (9) indicates that these potential gains are increasing in skill specificity. Intuitively, when χ is low, each worker is similarly productive across tasks, hence little is gained from the division of labor, in contrast to the case of high χ, when a worker's realized productivity varies greatly depending on what tasks they perform. Now allowing ξ to vary makes clear that potential efficiency gains are realized if coworkers are skilled at different tasks, that is when ξ is high.

The second term has a CES structure, featuring complementarities across coworkers' talents. One way to interpret these complementarities is to consider output Y and note that by the power mean inequality, the stronger the degree of talent complementarity, the greater is the weight on the talent of the least-capable team member(s) in determining Y. Hence, output is greater when coworkers are similar in terms of talent:

Corollary 2. *Consider any two elements* x_i *and* x_j *of the vector* $\boldsymbol{x}$ *such that* $x_i > x_j$*. Let* $\boldsymbol{x}'$ *be the vector* $\boldsymbol{x}$ *such that* x_i *is replaced with* $x_i + \epsilon$ *and* x_j *with* $x_j - \epsilon$*, where* $\epsilon > 0$*. Then for* $\chi\xi > 0$*,* $f(\boldsymbol{x}) > f(\boldsymbol{x}')$*.*

Proof. For any arbitrarily small but strictly positive value of ϵ, $f(\boldsymbol{x}) > f(\boldsymbol{x}')$ if $\partial f(\boldsymbol{x}')/\partial\epsilon < 0$. Differentiating yields $\frac{\partial f(x')}{\partial \epsilon} = c\left[-(x_j - \epsilon)^{\frac{1}{\chi\xi+1}-1} + (x_i + \epsilon)^{\frac{1}{\chi\xi+1}-1}\right] < 0$, with c a positive constant. Evaluating at $\epsilon = 0$ and if $\chi\xi > 0$, $x_i > x_j$ implies $\partial f(\boldsymbol{x}')/\partial\epsilon < 0$. □

[17] The elasticity of substitution across tasks, η, does not show up in equation (9). Technically, as noted by Eaton and Kortum (2002, Footnote 18), this holds as long as $1 + \chi - \eta\chi > 0$, in which case η only appears in a constant term that cancels with the scaling term ι. The irrelevance of η in that sense is a tight implication of the Fréchet, and it serves to sharply bring into relief that coworker complementarities do not hinge on the assumption that tasks combine in a Leontief fashion (e.g., Kremer, 1993).

[18] The appropriate measure of substitutability in the case of more than two inputs is the Morishima (1967) elasticity (Blackorby and Russell, 1989). Under the functional form of equation (9), this elasticity is identical across worker pairs and identical to the Hicks and Allen definitions (Blackorby and Russell, 1981).

An equally valid interpretation foregrounds coworker productivity spillovers. Having more talented coworkers boosts anyone's productivity, but the marginal contribution of greater worker talent is higher when matched with other high-talent workers, i.e., f is supermodular. The magnitude of this effect is increasing in χ, i.e., $\frac{\partial^3 f(\cdot)}{\partial x_i \partial x_j \partial \chi} > 0$ for any $i \neq j$ provided $\xi > 0$.[19]

To gain intuition, it is instructive to study how task shares vary with team members' talents, and how this depends on χ.

Corollary 3 (Task shares)**.** *The share of tasks produced by worker i is equal to*

$$\pi_i = \left(x_i^{\frac{1}{1+\chi\xi}} \right) \left(\sum_{k=1}^{n} (x_k)^{\frac{1}{1+\chi\xi}} \right)^{-1}. \tag{10}$$

Hence: (i) $\frac{\partial \pi_i}{\partial x_i} > 0$ and $\frac{\partial \pi_i}{x_k} < 0$ for $k \neq i$; and (ii) considering two team members i and j, and supposing that $x_i > x_j$, it holds that $\pi_i > \pi_j$ for $\chi\xi < \infty$, $\pi_i/\pi_j \longrightarrow x_i/x_j$ as $\chi\xi \longrightarrow 0$, and $\pi_i/\pi_j \longrightarrow 1$ as $\chi\xi \longrightarrow \infty$.

Proof. Equation (10) follows directly from Lemma A.3 given $\lambda = 1$. □

Consider a team of a talented worker, i, and a less talented coworker, j. In general, per equation (10), it is optimal for i to perform a greater share of tasks than j. Else there would be an inefficiently high amount of those tasks produced in which i is most skilled. If skills are not task-specific, i's average usage-weighted skill (i.e. $\int_{\mathcal{T}} l_i(\tau) z_i(\tau) d\tau$) is unaffected. But if $\chi > 0$, $z_i(\tau)$ diminishes as we expand i's task share. This implies, firstly, that a greater share of tasks is performed by j when χ is larger – this underlies the power mean intuition offered above. And secondly, i's realized average skill is increasing in her coworker's talent, the absolute magnitude of this effect being increasing in x_i. Simply put, it does not matter much in absolute terms which tasks a low-talent worker performs; but it matters a great deal if the worker could potentially produce a lot of output.

Discussion. This section showed under the optimal organization of production, coworkers' realized productivities are inherently interdependent – hence, referring to the classic definition of Alchian and Demsetz (1972, Section II), I speak of "team production". The

[19]The cross-partial is very intuitively connected to the degree of supermodularity, measuring how much the marginal productivity of one worker's talent changes with a change in a coworker's talent. It is also directly linked to the elasticity of complementarity, as $\gamma = (f f_{ij})/(f_i f_j)$ for any $i \neq j$, where subscripts indicate partial derivatives. This observation goes back to Hicks (1932, 241-246).

theory formalizes and makes precise the intuitive notion of coworkers being "complementary" or having "complementary skills," which will be the case insofar as two workers are specialized in *different* tasks but are *similar* in talent. Crucially, while the model inherits the tractability of a CES production function, the theory reveals the CES elasticity parameter to be a reduced-form object, rather than being "deep" or structural. It predicts that the strength of complementarities is increasing with the degree of skill specificity χ.

2.3 Team formation

Equipped with reduced-form production function (9), we can now study how the skill composition of different teams – as summarized by individual talents x and coworker horizontal distance ξ – is determined endogenously in equilibrium.

2.3.1 Further assumptions and notation

Tractability. In a dynamic context, where a firm may have already filled a slot, the technical challenge arises that the decision to hire an unmatched worker should depend on that person's task-specific skills *and* those of the incumbent employee. Both are assumed to be observable to agents upon meeting, raising the prospect of having to carry around a very high- (infinite-) dimensional object as a state variable. Proposition 1 resolves this challenge, showing that $\boldsymbol{x}$ and ξ are sufficient statistics for joint production given parameter χ. Hence, even though to the econometrician, workers' talent types will be observable but not their task-specific skills (see Section 3), unobserved match-specific heterogeneity for any two workers is summarized in the scalar ξ, the horizontal distance. Essentially, ξ is treated as a match-specific shock that is drawn upon meeting from the offer distribution H, which given Assumption 2 is continuous uniform. Of course, the distribution of ξ among *realized* matches is endogenous to hiring decisions just as the joint distribution of talents in teams is.

To ensure that the full set of task-specific skills does not turn into a state variable through objects other than production, the following assumption is also maintained.

Assumption 3. *Separation rates and the flow utility from being unemployed are functions of talent but not the task-specific skills, and are denoted* $\delta(x)$ *and* $b(x)$*.*

Distributions. Let $d_{m.1}(x)$ denote the measure of producing matches consisting of a firm and one worker of talent x, while $d_{m.2}(x, x')$ is the corresponding measure of matches with an additional coworker with talent x' for any value of ξ.[20] The population measure of workers with talent x, $d_w(x) = \Psi'(x)$, is the sum of those with who are unemployed, in one-worker matches or in two-worker matches:

$$d_w(x) = d_u(x) + d_{m.1}(x) + \int d_{m.2}(x, \tilde{x}')d\tilde{x}', \tag{11}$$

The aggregate unemployment rate is obtained by integrating over x, i.e., $u = \int d_u(x)dx$. Similarly for firms, the total measure of firms with fewer than two employees is $v = d_{f.0} + \int d_{m.1}(x)dx$, where $d_{f.0}$ is the mass of idle firms, and adding-up requires

$$d_f = d_{f.0} + \int d_{m.1}(x)dx + \frac{1}{2}\int\int d_{m.2}(x, x')dxdx'. \tag{12}$$

Dividing by 2 in the last term avoids double-counting, $d_{m.2}(x, x')$ being symmetric.

Joint value and surplus. The joint value of a match between a firm and a worker of talent x is $\Omega_1(x) = V_{f.1}(x) + V_{e.1}(x)$, where $V_{e.1}(x)$ is x's value of being employed alone, whose value is $V_{f.1}(x)$. The surplus generated by such a match is

$$S(x) = \Omega_1(x) - V_{f.0} - V_u(x), \tag{13}$$

where $V_u(x)$ is the value of unemployment for x and $V_{f.0}$ is the value of an idle firm.

The joint value of a firm with a team characterized by (x, x', ξ) is $\Omega_2(x, x', \xi) = V_{f.2}(x, x', \xi) + V_{e.2}(x|x', \xi) + V_{e.2}(x'|x, \xi)$, with $V_{e.2}(x|x', \xi)$ denoting the value of x being employed together with a coworker of talent x' if the horizontal distance between them is ξ. Hence, the surplus generated when a firm that already has employee x' hires an additional worker is

$$S(x|x', \xi) = \Omega_2(x, x', \xi) - \Omega_1(x') - V_u(x). \tag{14}$$

This surplus is increasing in the joint value generated by the team and declines with both sides' outside values. Note $S(x|x', \xi)$ is generically not symmetric in x and x' even if $\Omega_2(x, x', \xi)$ is, because it matters for outside options who is the incumbent employee.

[20] For notational simplicity, and with some abuse of language, I use the term "measure" throughout to refer to these and analogous functions. In the context of the continuous-type model, these should be interpreted as densities, while in the numerical implementation, which is based on a discretized type space, they represent point measures or masses.

Surplus sharing. Wage bargaining takes the generalized Nash form, concerns the entire surplus, and the firm treats each employee as marginal, so that their outside option is unemployment. This protocol ensures that all matching decisions are privately efficient and can be characterized by parties' joint surplus (cf. Bilal *et al.*, 2022). Moreover, the wage is unaffected by the order with which workers join a team.

In particular, the wage $w(x|x', \xi)$ of a worker of talent x employed with a coworker of talent x' who is horizontal distance ξ apart is

$$(1-\omega)\big(V_{e.2}(x|x',\xi) - V_u(x)\big) = \omega\big(V_{e.2}(x'|x,\xi) + V_{f.2}(x,x',\xi) - V_{e.1}(x') - V_{f.1}(x')\big). \tag{15}$$

2.3.2 Equilibrium conditions and definition

A set of Hamilton-Jacobi-Bellman (HJB) equations characterizes optimal matching decisions and associated values, taking the distribution of agents over type and employment states as given. The stationary distribution satisfies a system of Kolmogorov-Forward equations (KFEs) given optimal matching decisions.

These matching decisions are related to surplus values as follows:

$$h(x) = \mathbf{1}\{S(x) > 0\}, \quad h(x|x',\xi) = \mathbf{1}\{S(x|x',\xi) > 0\}. \tag{16}$$

These functions describe, respectively, whether a match between an unmatched firm and a type-x worker will (optimally) be consummated; and whether a firm that already employs worker x' is willing to hire a worker x given shock ξ. To track the evolution of the distribution, it is furthermore useful to define conditional matching probabilities,

$$h(x|x') = \mathrm{P}\{S(x|x',\xi) > 0\}. \tag{17}$$

Value functions. The asset value of an idle firm, $V_{f.0}$, satisfies

$$\rho V_{f.0} = (1-\omega)\lambda_{v.u} \int \frac{d_u(x)}{u} S(x)^+ dx, \tag{18}$$

where for any r, I let $r^+ = \max\{r, 0\}$ indicate the optimal decision, to ease notation. The discounted value thus corresponds to the weighted conditional expectation of the firm's share of the match surplus generated with an unemployed worker times the unconditional probability of meeting any unmatched worker.

The value of such an unemployed worker x is given by

$$\rho V_u(x) = b(x) + \lambda_u \omega \left[\frac{d_{f.0}}{v} S(x)^+ + \int \int \frac{d_{m.1}(\tilde{x}')}{v} S(x|\tilde{x}', \tilde{\xi})^+ dH(\tilde{\xi}) d\tilde{x}' \right] \tag{19}$$

Notice that we need to account for the worker's flow value from being unmatched, $b(x)$, and distinguish between the worker meeting an unmatched firm or a one-worker firm.

Turning to matched agents, the joint value of a firm employing a team (x, x', ξ) is

$$\rho \Omega_2(x, x', \xi) = f(x, x', \xi) - \delta(x) S(x|x', \xi) - \delta(x') S(x'|x, \xi), \tag{20}$$

and – importantly – the joint value of a firm with a talent-x employee satisfies

$$\begin{aligned} \rho \Omega_1(x) = {} & f(x) + \delta(x)\big[-\Omega_1(x) + V_u(x) + V_{f.0} \big] \\ & + \lambda_{v.u} \int \int \frac{d_u(\tilde{x}')}{u} \big(\underbrace{-\Omega_1(x) + V_{e.2}(x|\tilde{x}', \tilde{\xi}) + V_{f.2}(x, \tilde{x}', \tilde{\xi})}_{(1-\omega)S(\tilde{x}'|x,\tilde{\xi})} \big)^+ dH(\tilde{\xi}) d\tilde{x}'. \end{aligned} \tag{21}$$

The discounted value thus includes the flow value of production. At rate $\delta(x)$ the match is destroyed. At rate $\lambda_{v.u}$, the firm is contacted by an unmatched worker with talent $\tilde{x}'$, they draw a shock ξ, and the worker is hired if the sum of values accruing to the firm and x from teaming up exceeds their joint value if $\tilde{x}'$ is not hired, that is, if the joint surplus from a match is positive.

Online appendix A.2.1 derives the recursions for the surplus values, see specifically equations (A.8) and (A.9), so that the optimality conditions are fully described in terms of the values of unmatched agents and surpluses.

Population dynamics. In stationary equilibrium, the inflows and outflows into different states balance. The KFEs summarize these flows for each state, with the matching decisions modulating the flow intensities implied by exogenous meeting rates and the distributions themselves. As the equations are straightforward but lengthy, they are collected in Appendix A.2.2. With that, we are in a position to define the equilibrium.

Definition 1. *A stationary equilibrium consists of a production function $f(\boldsymbol{x}, \xi; \chi)$, a tuple of value functions, $\big(V_u(x), V_{f.0}, S(x), S(x|x', \xi)\big)$, together with a distribution of agents across states, $\big(d_{m.1}(x), d_{m.2}(x, x')\big)$, such that (i) the production function f is consistent with optimal task assignment per equation* (5)*, (ii) the value functions satisfy the HJB* *equations* (18), (19), (A.8) *and* (A.9) *given the distributions; and (iii) the stationary distributions satisfy the KFEs* (A.11)-(A.12) *given the policy functions implied by the value functions following equation* (16).

The equilibrium needs to be computed numerically because of a non-trivial equilibrium interaction: Agents' expectations and matching decisions must conform with the distribution to which they give rise, yet as that distribution evolves, so do agents' expectations over future meeting probabilities and, hence, their optimal actions.

2.3.3 Key trade-off and qualitative properties of equilibrium

I next discuss the key trade-off that shapes team composition and illustrate how, given this trade-off, shifts in skill specificity χ affect sorting.

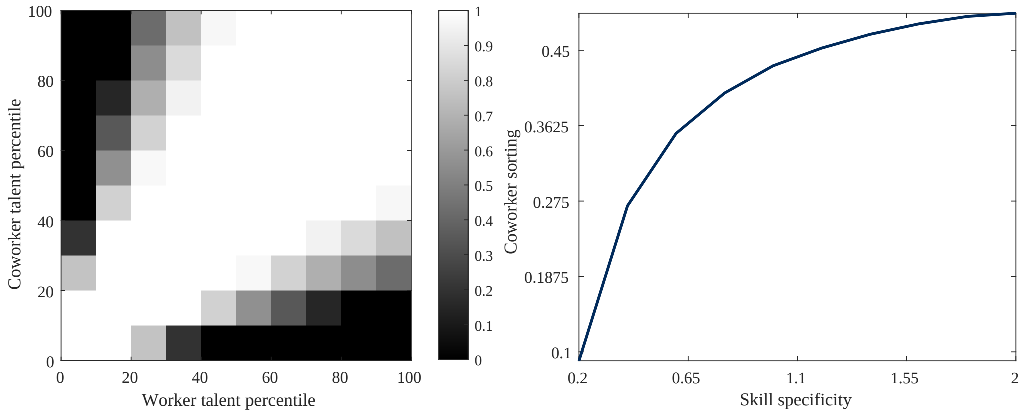
Trade-off. Suppose a firm with an employee of talent x meets an unmatched worker of talent x' and their horizontal distance is ξ. The matching decision, encapsulated in the policy function $h(x'|x,\xi)$, balances match quality considerations and search costs. If the hire is made, output $f(x,x',\xi)$ is produced and shared. Else, the firm produces $f(x)$, the unmatched worker receives the flow value $b(x')$, and both sides search for another production partner with whom they can generate a positive surplus.

The production function implies that team formation is more likely between workers of similar in talent who are horizontally distant. Figure 2a illustrates the matching probabilities $h(x'|x)$ under an illustrative parametrization, plotting incumbent worker talent x against candidate coworker talent x'. A low-talent candidate will never be matched with a high-talent incumbent, even if ξ is high, since the incumbent's employer can always find a coworker who is at similar horizontal distance but closer in talent. In contrast, teams of workers with similar talent levels are readily formed. For intermediate talent differences, matching occurs only when workers are sufficiently horizontally distant to offset the productivity loss from talent mismatch.

This is a good place to remark that, since firms are ex-ante homogeneous, in this model it is completely random *which* firm ends up with the most productive teams. The focus lies on systematic sorting patterns given random draws of the first worker.

Coworker sorting statistic. A useful statistic to quantify talent sorting is the correlation between coworkers' talent ranks (Lopes de Melo, 2018).[21] To derive this statistic from

[21] Measuring correlation across ranks aligns with my empirical approach of recovering discrete worker types. It also avoids confounding shifts in assortativeness and in the marginal talent distribution.



(a) Matching probabilities

(b) Skill specificity and sorting

Figure 2: Equilibrium matching and comparative statics: skill specificity fosters sorting

Notes. This figure summarizes matching patterns in the stationary equilibrium. Figure 2a visualizes the conditional matching probabilities defined in equation (17) under an illustrative calibration. Figure plots moments like coworker sorting for different values of χ.

the equilibrium objects specified above, let $d_{m.2}(\hat{x},\hat{x}')$ denote the density of matches in terms of talent ranks. (I use this notation for simplicity, with the understanding that appropriate transformations due to the change of variables are applied.) Then the equilibrium joint density in teams is $\phi(\hat{x},\hat{x}')\frac{d_{m.2}(\hat{x},\hat{x}')}{\int\int d_{m.2}(\hat{x},\hat{x}')d\hat{x}d\hat{x}'}$; the unconditional density in teams is $\phi(\hat{x}) = \int \phi(\hat{x},\hat{x}')d\hat{x}'$; and the distribution of coworker types conditional on type is $\phi(\hat{x}'|\hat{x}) = \frac{\phi(\hat{x},\hat{x}')}{\int \phi(\hat{x},\hat{x}')d\hat{x}'}$, the corresponding CDFs being $\hat{\Phi}(\hat{x})$ and $\hat{\Phi}(\hat{x}'|\hat{x})$. Then

$$\rho_{xx} = \frac{\int\int(\hat{x}-\bar{x})(\hat{x}'-\bar{x})d\Phi(\hat{x}'|\hat{x})d\hat{\Phi}(\hat{x})}{\int\left(\hat{x}-\bar{x}\right)^2 d\hat{\Phi}(\hat{x})}, \tag{22}$$

where $\bar{x} = \int \hat{x}d\hat{\Phi}(\hat{x})$ is the average worker type among those in teams. "Sorting" in this paper thus exclusively refers to *coworker* sorting, as opposed to sorting between ex-ante heterogeneous firms and workers.

Comparative statics for χ. As greater skill specificity amplifies coworker talent complementarities, and holding fixed search costs, this fosters an equilibrium featuring more positively assortative coworker matching in equilibrium. Figure 2b demonstrates this positive relationship between χ and ρ_{xx}. Succinctly put, under strong skill specificity the equilibrium features some firms with "superstar teams" composed of the most talented workers and other firms with "laggard teams" (cf. Andrews *et al.*, 2019; Autor *et al.*, 2020).

3 Model Meets Data

In this section, I bring the model to the data. After introducing the dataset in Section 3.1, Section 3.2 maps key model objects to those data. I next calibrate the structural parameters and explore its quantitative properties (Section 3.3), and finally validate core model mechanisms using industry-level variation (Section 3.4).

3.1 Data

The empirical analysis draws on the Sample of Integrated Employer-Employee Data (SIEED), an administrative matched employer-employee dataset for Germany provided by the Institute for Employment Research (IAB). The dataset covers the entire workforce of a 1.5% sample of all establishments in Germany as well as the complete employment biographies of these workers, including spells in which they are not employed at the sampled establishments. For each worker, the data include demographic characteristics (e.g. age, gender, and education) and job information (e.g. employer, wages, occupation, and industry). The production unit is the establishment, defined by ownership, industry and location; with that proviso, I use "establishment" and "firm" interchangeably. This dataset enables the analysis of workers' employment trajectories, including their coworker relationships and earnings, over a long period.

To construct the main dataset used in the analysis, I proceed as follows. First, I convert the spell-level data into an annual panel. Second, I restrict the sample to workers aged 20–60 who are employed full-time at West German establishments in the non-financial, non-agricultural private sector, requiring each establishment to have at least ten worker observations per year. Third, the earnings variable is the residual (daily) wage, constructed in line with previous studies (Card *et al.*, 2013; Hagedorn *et al.*, 2017) by controlling for an unrestricted set of year dummies, education-specific age profiles, and job tenure. Appendix B.1.1 provides further details on the sample construction and descriptive statistics. The analysis in this section uses the subsample spanning 2010-2017 (3,605,972 person-year observations), unless indicated otherwise, while the analysis of time-series trends in Section 4 uses the full sample (14,429,61 person-year observations).

3.2 Mapping theory to data

This section describes how theoretical objects are mapped to the data. Specifically, more elaborate measurement strategies are needed to pin down three objects: worker types; teams and coworkers; and skill specificity.

3.2.1 Worker talent types

In the theory, a worker's type $\hat{x}$ indexes their time-invariant talent. Empirically, I measure $\hat{x}$ by exploiting the panel structure of data and model. As the wage is monotonically increasing in talent – a standard propery of matching models with individual-level heterogeneity in absolute advantage – more talented workers have higher lifetime earnings, even if in a single cross-section unemployment or poor match quality may imply a lower wage compared to a less talented worker. I therefore estimate individual fixed effects (FE) from panel wage regressions, controlling for observable characteristics and time-invariant employer effects. Details are relegated to Online Appendix B.1.2. Ranking workers by their FEs and binning them into deciles – aligned with the numerical solution of the model on a ten-point talent grid – yields $\hat{x}_i$ for each worker i. The mapping from these ordinal rankings to cardinal measures of output and wages is discussed in Section 3.3.1.

In addition to ranking workers economy-wide, as my baseline, I also consider ranking workers within occupations as an interesting alternative. Under this approach, the coworker correlation coefficient ρ_{xx}, for example, captures whether top performers in each occupation tend to cluster together. Additionally, Online Appendix B.2 examines robustness to alternative measures of worker productivity, notably years of schooling as a non-wage metric, and addresses potential concerns about time-variation in talent due to about human capital accumulation.

3.2.2 Defining teams and coworker(s)

Real-world production units typically have more than two workers, so I construct a "representative coworker" type for each worker-year observation to map the two-worker model to the data.[22] Concretely, I take i's coworkers in year t to be given by the set of employees in the same establishment-year cell, denoted $S_{-it} = \{k : j(kt) = j(it), k \neq i\}$, $j(it)$ being the identifier of i's period-t employer. Then i's representative coworker in

[22]This approach is similar to Jarosch *et al.* (2021) and Herkenhoff *et al.* (2024), but aggregates worker types rather than wages.

year t is the leave-out mean, $\hat{x}_{-it} = \frac{1}{|S_{-it}|} \sum_{k \in S_{-it}} \hat{x}_k$.[23] Importantly, this step aligns the theoretical notion of teams with establishments.

This empirical aggregation step is immensely useful but also amounts to a significant simplification, so it is worth considering what could go wrong. First, should only same-occupation employees be treated as coworkers, rather than everyone in the same establishment?[24] The theory suggests otherwise – coworker complementarities arise from skill differentiation, which is more likely across different occupations contributing to joint output. Second, the theory strictly speaking implies that lower-talent coworkers should receive higher weights in aggregation insofar as $\chi > 0$ (cf. Proposition 1). Taking an unweighted average is in line with existing studies but ignores these complementarities across an individual's coworkers. Online Appendix B.2.5) explains why the resulting bias is minor. Third, ignoring team size variation may seem problematic. However, the averaging naturally captures that a single coworker change matters less in larger teams.

3.2.3 Skill specificity: approach & auxiliary regression results

Measuring skill specificity (χ) presents the greatest challenge. Direct measurement would involve detailed individual-level data on task-specific skills, computing the average dispersion across individuals, but those data do not exist.[25] Instead, I exploit the model's structure to infer χ indirectly; then show that the resulting estimate aligns with an intuitive, albeit ordinal, task-based proxy for skill specificity across industries (Section 3.4) and over time (Section 4.2). In this section, I first present the key theoretical result that guides measurement, then discuss the empirical implementation, and finally compare my approach to the alternative of inferring χ from observed sorting patterns.

A useful identification result. Proposition 1 and Corollary 1 showed that χ is systematically related to $\partial^2 f(\cdot)/\partial x \partial x'$. How can we quantify this moment given our data on worker types and wages? The central insight here is that the *production* cross-partial is directly proportional to the *wage* cross-partial.

Corollary 4 (Measuring the production cross-partial)**.** *The cross-partial derivative of the production function with respect to x and x', conditional on ξ, is proportional to that of*

[23] By computing an arithmetic average, it is implicitly assumed that the intervals between (ordinal) *types* are similar. This is consistent with the calibration of the model in Section 3.3. An alternative would be to work with cardinal worker types throughout.

[24] Jarosch *et al.* (2021) refer to this distinction as "Team Definition 1" vs. "Team Definition 2."

[25] Some datasets such as the NLSY and administrative data for Sweden and Denmark contain skill-specific test scores, but these are typically limited to broad categories like verbal or technical ability.

the wage function:

$$\frac{\partial^2 f(x, x', \xi)}{\partial x \partial x'} \propto \frac{\partial^2 w(x|x', \xi)}{\partial x \partial x'}. \tag{23}$$

To see the reasoning behind this result, note first that can write the wage of x employed alongside x' given ξ as (see Appendix A.2.3)

$$\begin{aligned} w(x|x', \xi) = {} & \omega\big(f(x, x', \xi) - f(x')\big) + (1-\omega)\rho V_u(x) \\ & - \omega(1-\omega)\lambda_{v.u} \int \int \frac{d_u(\tilde{x}'')}{u} S(\tilde{x}''|x', \tilde{\xi})^+ dH(\tilde{\xi}) d\tilde{x}''. \end{aligned} \tag{24}$$

Equation (23) follows from differentiating with respect to x and x'. Intuitively, the worker receives an ω-share of the increase in production from being match with a coworker, but the wage level also reflects outside options. While this complicates identification of production functions using wage data (Eeckhout and Kircher, 2011; Hagedorn *et al.*, 2017), the cross-partial derivative is unaffected as the parties' outside options are separable. Intuitively, the new hire's option is foregone regardless of coworker quality, while the firm and incumbent's options are foregone regardless of who is hired.

To operationalize Corollary 4, three more steps are needed. First, as ξ is unobserved, I construct the average wage of a type-x worker employed together with a coworker of talent x', integrating over ξ, which is denoted by $\bar{w}(x|x')$.[26] Second, I construct a first-order approximation of $\frac{\partial \bar{w}(x|x')}{\partial x \partial x'}$ in the data, as discussed next. Third, this approximation serves as a targeted moment informative about χ when calibrating the model in a simulated-method-of-moments procedure.

Empirical operationalization. To empirically approximate $\frac{\partial \bar{w}(x|x')}{\partial x \partial x'}$, I estimate the following regression of individual-year level wages using OLS:

$$\begin{aligned} \frac{w_{it}}{\overline{w}_t} = {} & \beta_0 + \sum_{d=2}^{10} \beta_{1d} \mathbf{1}\{\hat{x}_i = d\} + \sum_{d=2}^{10} \beta_{2d} \mathbf{1}\{\hat{x}_{-it} = d\} + \beta_c(\hat{x}_i \times \hat{x}_{-it}) \\ & + \psi_{j(i,t)} + \nu_{0(i,t)t} + \xi_{S(i,t)t} + \epsilon_{it} \end{aligned} \tag{25}$$

[26]This step is not trivial, as the distribution of ξ among matches reflects selection processes (Borovickova and Shimer, 2024). Specifically, for $\chi > 0$ the horizontal distance ξ tends to be lower for pairs whose distance in talent space is high. Importantly, the direction of the implied bias in my estimate of the production cross-partial, and hence χ, is downward, so overall the results should be interpreted as a lower bound for the importance of skill specificity and, hence, team production.

where $\mathbf{1}\{\hat{x}_i = d\}$ and $\mathbf{1}\{\hat{x}_{-it} = d\}$ are dummies for worker and coworker types, respectively, $\psi_{j(i,t)}$ denotes employer fixed effects (FE), $\nu_{o(i,t)t}$ are occupation-year FEs, $\xi_{s(i,t)t}$ are industry-year FEs. I weight each observation by the inverse empirical frequency of the associated $(\hat{x}_i, \hat{x}_{-it})$ match to ensure equal coverage of the state space. If the true data-generating process for wages follows equation (25), then β_c measures the cross-partial of interest. Specifically, β_c indicates how much more the wage of an individual i rises, as a percentage of the average wage $\bar{w}_t$, with a one-decile increase in coworker talent compared to an individual i' whose rank is one decile lower than that of i. The greater is the degree of skill specificity χ, the larger β_c will be.

Identification comes from variation in coworker quality over time – which in the theory naturally occurs due to search frictions. Variation includes both movers (i.e. changes in coworker quality for individuals who switch employer) and stayers (i.e. changes in coworker quality induced by other employees joining or leaving the coworker group). Using pre-determined type measures – as opposed to estimating types and coworker effects jointly – avoids the reflection problem (Manski, 1993), while controlling non-parametrically for individual types addresses coworker sorting. Additionally, a rich set of fixed effects absorb unobserved time-invariant employer heterogeneity and shocks at the occupation-year or industry-year level.[27]

Auxiliary regression results. Table 1 reports the estimates of β_c for the sample period 2010-2017. The point estimate under the baseline specification, which corresponds to equation (25) and is reported in column (4), is 0.0063 and statistically significant at the 1% level. To provide a sense of magnitude, this implies that the wage increase from a one decile improvement in the average coworker quality is 3.15% greater, as a percentage of the average wage, for a worker who is in the top decile compared to one in the fifth decile (0.0063·(10-5)·1=0.0315).

This finding holds up in a large battery of robustness analyses, some of which are also reported in Table 1. First, it makes little difference to $\hat{\beta}_c$ what combination of employer/industry/occupation FEs is included. Second, when workers are ranked within occupation, instead, $\hat{\beta}_c$ is only slightly lower, at 0.0059 (column (5)). Third, when I repeat the exercise of ranking workers and coworkers and estimating equation (25) separately for each 2-digit industry, the average value of $\hat{\beta}_c$ is likewise similar at 0.0061 (see Section 3.4).

[27] Different from typical peer effects studies (e.g., Cornelissen *et al.*, 2017; Nix, 2020), I focus on the interaction term as opposed to an average treatment effect, following Corollary 4, and I leverage the model's structure to separate type measurement from spillover estimation.

	(1)	(2)	(3)	(4)	(5)
Interaction coefficient ($\hat{\beta}_c$)	0.0067***	0.0067***	0.0063***	0.0063***	0.0059***
	(0.0005)	(0.0004)	(0.0005)	(0.0005)	(0.0008)
Employer FEs	No	No	Yes	Yes	Yes
Industry-year FEs	No	Yes	No	Yes	Yes
Occupation-year FEs	No	No	Yes	Yes	Yes
Type ranking	Economy	Economy	Economy	Economy	Occupation
Obs. (1000s)	3,606	3,606	3,606	3,606	3,606
Adj. R^2	0.788	0.800	0.801	0.813	0.769

Table 1: Auxiliary regression results (2010-2017 period)

Notes. Regression results based on equation (25), reporting the estimates of the coefficient on the interaction term, $\hat{\beta}_c$. The dependent variable is the (residualized) wage, in levels and divided by the year-specific average wage. Employer-clustered standard errors are given in parentheses. Observations are weighted by the inverse employment share of the respective type and (rounded) coworker type cell. Observation count rounded to 1000s. * p<0.1; ** p<0.05; *** p<0.01.

Appendix B.2 reports a variety of other robustness checks, e.g. using years of schooling as a non-wage measure of worker types.

Discussion. The coefficient β_c serves as the targeted moment disciplining χ, the skill specificity parameter, when estimating the model using indirect inference (as described in the next section); equilibrium moments like the degree of coworker sorting remain untargeted. An alternative, and in some ways more conventional, approach would have been to infer χ by targeting, say, the coworker correlation coefficient ρ_{xx}, given the monotone relationship shown in Figure 2b.

The strategy adopted here, facilitated by the identification result in Corollary 4, has three advantages. First, the model is very parsimonious and omits other mechanisms that could generate positive coworker sorting, such as preference-based homophily (we might just *enjoy* working together with others who are similar in talent) or sorting due to ex-ante firm heterogeneity. Hence, targeting ρ_{xx} risks attributing "too much" to χ. Second, my calibration strategy is aimed at answering the original, motivating question whether team production 'matters' for macro-level outcomes. In principle, given the estimated value of $\hat{\beta}_c$ and the implied value of χ, the model might generate very little or a lot of equilibrium dispersion across firms. Third, the approach allows testing the postulated link between skill specificity and sorting, which is done in Section 3.4.

3.3 Model calibration and quantitative properties

I now turn to the quantitative analysis of the model. I start by describing how structural parameters are disciplined, then discuss the estimation results.

3.3.1 Methodology

I proceed in four steps. First, I impose some functional form assumptions to reduce the complexity of the problem. Second, I preset two parameters based on the literature. Third, I directly infer parameters which the model maps one-to-one to empirical moments. Fourth, the remaining parameters are jointly estimated using indirect inference.

Functional form assumptions. To reduce the dimensionality of the estimation problem, I make some simplifying assumptions. In particular, the model features a continuum of talent types, $\hat{x}$, which in the numerical analysis is discretized using ten grid points. Moreover, the flow value of unemployment is assumed to be proportional to output produced alone, with proportionality factor $\bar{b}$. I also assume that the separation rate is a linear function of talent, indexed by parameters δ_0 and δ_1.

Importantly, the functional form of the production function f is directly informed by the microfoundation, but for the quantitative exploration I introduce a few amendments. First, Section 3.2.1 estimated worker *types* $\hat{x}_i$, whereas production function (9) took cardinal talent x_i as its input. To flexibly map between ordinal and cardinal objects I introduce two parameters, a_0 and a_1, into the production function. (Alternatively, one could directly estimate a mapping between $\hat{x}$ and x.) Thus, a firm with a single worker produces $\hat{f}(\hat{x}) = a_0 + a_1\hat{x}$. Second, I introduce a size adjustment to account for the fact that real-world teams have more than two members. This is important because team size affects how variation in skill specificity χ influences the efficiency gains from working in teams relative to producing alone. Practically, I suppose that what the model treats as the second hire shows up, in the production function, as the $\bar{n}^{\text{th}}$ hire. Third, according to Proposition 1, a greater value of ξ, the horizontal distance, raises both productivity *and* the strength of coworker talent complementarities; the latter effect (weakly) lowers productivity. In the current analysis, I omit the latter effect in order to isolate the mechanism that greater *horizontal* coworker distance *raises* output whereas greater *vertical* (talent) coworker distance *lowers* it. Thus, the team production function in the quantitative analysis is:

$$\hat{f}(\hat{x}, \hat{x}', \xi) = 2 \times \left(\frac{\bar{n}}{\bar{n}-1}\right)^{\chi\xi} \times \left(a_0 + a_1 \left(\frac{1}{2}\left(\hat{x}\right)^{\frac{1}{\chi+1}} + \frac{1}{2}\left(\hat{x}'\right)^{\frac{1}{\chi+1}}\right)^{\chi+1}\right). \qquad (26)$$

Preset parameters. Regarding preferences and bargaining, I follow Herkenhoff *et al.* (2024). The discount rate ρ is set to 0.008, consistent with an annual interest rate of 10%. Such a high rate is common in this type of model. A high discount factor effectively proxies for concavity in the utility function, which tractability requires us to abstract from. The bargaining parameter $\omega = 0.50$ implies equal sharing of surplus.

Directly inferred parameters. I set $\bar{n}$ equal to 14, which is the median unweighted establishment size in the data. The parameters indexing the separation rate function, δ_0 and δ_1 are estimated offline to match type-specific monthly job losing rates, which I compute from an auxiliary dataset, as described in Online Appendix C.1.1.

Internally estimated parameters. The remaining five parameters, collected in $\psi = \{\chi, a_0, a_1, \bar{b}, \lambda_u\}$, are estimated using standard indirect inference methods by matching moments. The estimated values of these parameters minimize the objective function

$$\mathcal{G}(\psi) = \sum_{j=1}^{5} \left(\frac{\hat{m}_j - m_j(\psi)}{\frac{1}{2}|\hat{m}_j| + \frac{1}{2}|m_j(\psi)|}\right)^2,$$

where $\hat{m}_j$ refers to the empirical moment and $m_j(\psi)$ denotes its model counterpart.

While the elements of ψ are jointly estimated, each is closely informed by one of these moments, as explained next. Online Appendix Figure C.1 provides an illustration. Most importantly, $\hat{\beta}_c$ directly informs the strength of χ, as discussed in Section 3.2.3. In practice, I estimate an auxiliary regression analogous to equation (25) on model-generated data. As the target moment I use β_c, scaled by the standard deviation of coworker types, denoted $\sigma_{x'}$; including the latter takes care of the concern that constructing the representative coworker variable lowers the variability of coworker talent relative to the model.

The values of a_0 and a_1 are guided, respectively, by the average wage, which is normalized to unity, and the total variance of log wages, which is equal to 0.23. Note that both parameters raise the average wage but the dispersion of wages is decreasing in a_0 yet increasing in a_1. The parameter b_1 is informed by the ratio at which the flow value of unemployment replaces the (type-specific) average wage. Based on official administrative replacement rates (post-Hartz reforms of the labor market undertaken in the 2000s)

Parameter	Description	Value	Source	m	$\hat{m}$
ρ	Discount rate	0.008	External		
ω	Worker barg. weight	0.50	External		
$\bar{n}$	Team size	14	Offline est.		
δ_0	Sep. rate, constant	0.0147	Offline est.		
δ_1	Sep. rate, scale	-0.84	Offline est.		
χ	Skill specificity	1.17	β_c	0.0063	0.0063
a_0	Production, constant	0.26	Normalized wage	1	1
a_1	Production, scale	1.49	Var. log wages	0.23	0.23
$\bar{b}$	Unemp. flow utility, scale	0.64	Replacement rate	0.63	0.63
λ_u	Meeting rate	0.23	Job finding rate	0.16	0.16

Table 2: Model parameters (2010-2017 period)

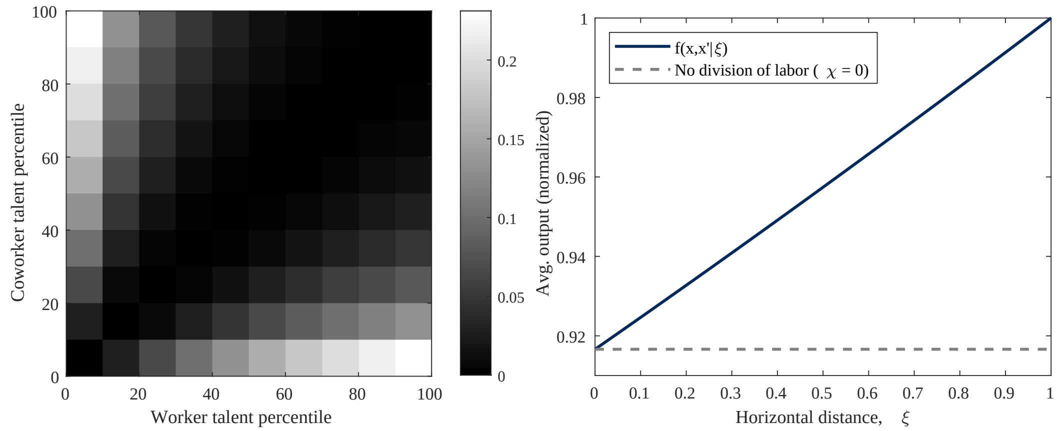
Notes. This table lists for each of the parameters the value. For internally estimated parameters, it also indicates the targeted moment, and moment values in data ($\hat{m}$) and model (m).

and recent work by Koenig *et al.* (2021) that also accounts for non-monetary opportunity costs of employment, I target a ratio of 0.63. Finally, λ_u targets a monthly job finding rate of unemployed workers equal to 16.2%, again computed from the LIAB.

3.3.2 Estimated parameters and model properties

Table 2 summarizes the model parameters and evaluates the model fit. The model is capable of matching the targeted moments perfectly, ψ being exact-identified. I next discuss the properties of the model, the guiding questions being: Given the estimated value of χ equal to 1.17, how does output vary with team composition; what are the implied sorting patterns; and what patterns of firm-level dispersion does this give rise to?

Properties of the estimated production function. Figure 3 illustrates key properties of the production function. The left panel quantifies *coworker talent complementarities*. Each cell value can be interpreted as the productivity gain from reallocating workers from two talent-mixed teams into two talent-homogeneous teams. Whereas absent skill specificity this would be zero, given the estimated parameters, for workers at opposite ends of the talent distribution, output is 23% higher under assortative matching. Tying this back to the underlying task assignment, when skills are specific, a talented worker's expands more slowly as coworker talent falls, hence a larger share of tasks is performed by the less talented team member. The right panel shows how output varies with *horizontal*



(a) Talent complementarity **(b)** Horizontal distance

Figure 3: Properties of the production function

Notes. In panel 3a, each cell indicates for every pair (x, x'), the value of $\frac{f(x,x|\xi=1)+f(x',x'|\xi=1)-2f(x,x',\xi=1)}{f(x,x|\xi=1)+f(x',x'|\xi=1)}$. Panel 3b plots $\int\int \frac{f(x,x',\xi)}{f(x,x',1)}dxdx'$ as a function of ξ.

distance, ξ. Averaging over all possible talent combinations, output is 8% lower when $\xi = 0$ – which corresponds to no division of labor – than when $\xi = 1$.

Effects of coworker separation by horizontal distance. As a validation exercise, I show that the model is consistent with and offers a structural interpretation of Jäger and Heining's (2022) findings, based on a quasi-experimental research design, regarding the heterogeneous wage effects from unexpected coworker deaths. A core finding is that while the average wage effect for remaining workers is positive, this masks substantial heterogeneity. For example, the death of a worker with highly specialized human capital (proxied by occupation-specific returns to experience) causes the wage of incumbents in other occupations to decrease by 0.42% in the short-run and 1.99% in the long-run.[28] Through the lens of the model, these effects are negative when the separated coworker had very complementary skills. Figure 4 shows that the percentage change in the remaining worker's wage due to a coworker separation (δ) shock becomes more negative as ξ increases. While the model predicts larger effect sizes than reported by Jäger and Heining's (2022) for highly differentiated tasks, it captures the key pattern that separations reduce wages only when coworkers had complementary specialized skills.

Employment outcomes. Turning to macro-level outcomes, Appendix Figure C.3 show that in the model – and in line with empirical evidence (e.g. Cairó and Cajner, 2018) –

[28] I thank the authors for sharing these statistics in percentage terms with me.

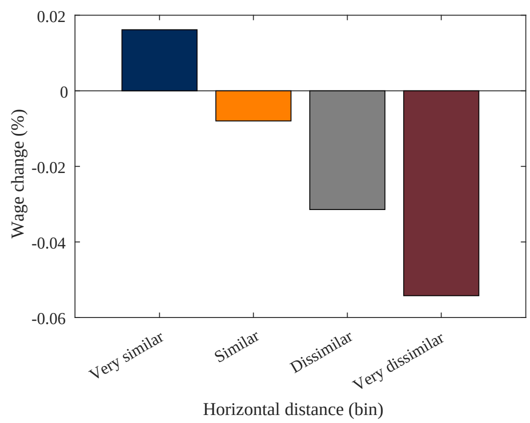


Figure 4: Theory predicts heterogeneous treatment effects of coworker separation

Notes. This figure shows the statistic $\int \frac{w(\hat{x})-w(\hat{x}|\hat{x}',\xi)}{w(\hat{x}|\hat{x}',\xi)} d\hat{\Phi}(\hat{x},\hat{x}')$ for 4 differents bins of horizontal distance ξ.

more talented workers face a lower unemployment risk. This is due to lower separation rates, not higher finding rates, talented workers being selective in accepting offers.

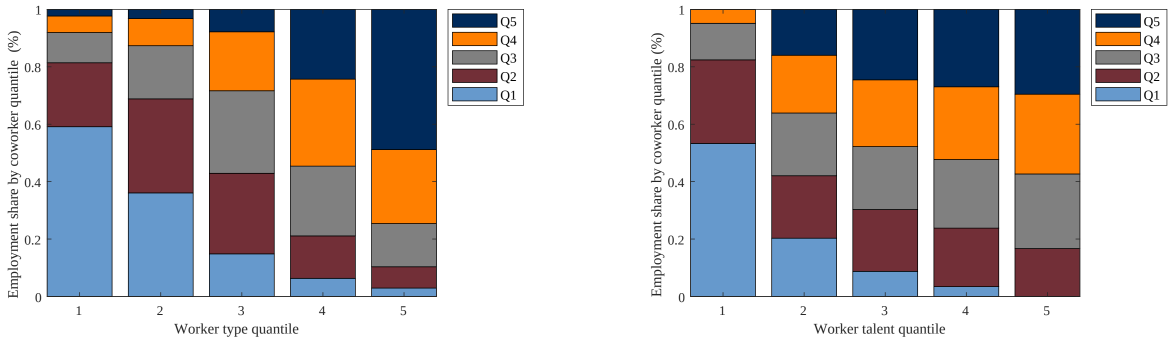
Labor market sorting and wage dispersion. Importantly, the model successfully reproduces the empirically observed, but untargeted, sorting patterns. The theoretical coworker correlation, ρ_{xx}, is strongly positive (0.45), approaching the empirical value (0.64). Providing a more disaggregated picture, Figure 5 compares predicted and actual conditional distributions of coworker talent in terms of quantiles (for ease of visualization). The model's predictions (panel 5b) align fairly well with the data (panel 5a). This fit is notable since only $\hat{\beta}_c$ informs sorting in the estimation.

The model-predicted decomposition of wage dispersion into between- and within-firm components likewise fits the data well, as will be discussed in detail in Section 4.2. Thus, even though firms are assumed to be ex-ante identical, the model endogenously generates large ex-post differences in terms of observables – some firms are more productive and pay more simply due to the quality of their teams..

3.4 Validation of model mechanisms: industry-level variation

This section reports correlational evidence in support of the core model mechanisms, exploiting variation across industries and within-industries over time.[29]

[29]An earlier version of this paper (Freund, 2023) also reported supportive cross-sectional evidence based on Portuguese micro data. Also see the findings in a companion note, Criscuolo *et al.* (2024).



(a) Data

(b) Theory

Figure 5: Model generates empirically an realistic coworker talent type distribution
Notes. For each quantile of the talent distribution, these plots – data on the left, model on the right – show the employment share of coworker talent types, where those are binned into quantiles, too.

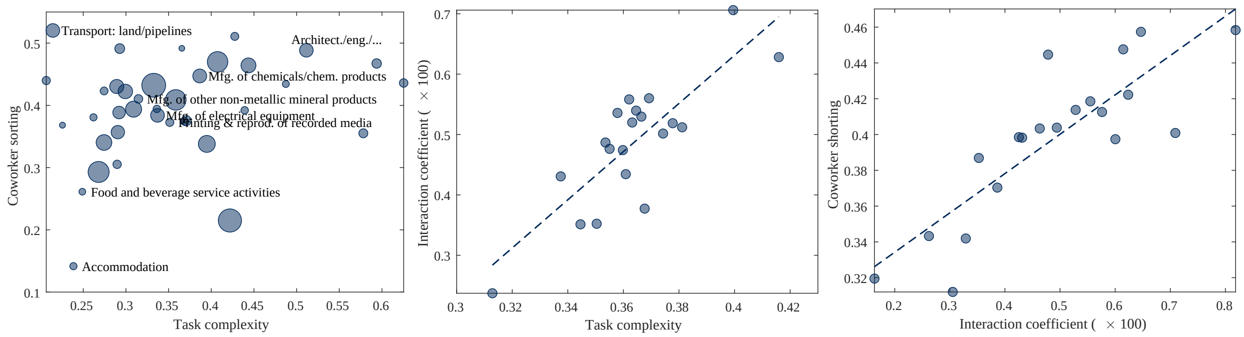
For these analyses, and foreshadowing the analysis of time trends in Section 4.2, I expand the sample to span 1985-2017 and divide these years into 5 periods indexed by p (1985-1992, 1993-1997, 1998-2003, 2004-2009, 2010-2010). Then I measure worker and coworker types, construct the coworker sorting coefficient ρ_{xx} separately for each (2-digit) (industry s, sample period p) cell, and obtain a point estimate for the interaction coefficient β_c by estimating regression (25) within each cell (s,p).

Task complexity proxy. In addition to structurally estimating χ, the degree of skill specificity, I use a task-based proxy: the share of tasks that are abstract and non-routine, henceforth termed "task complexity."[30] Intuitively, a worker is likely to perform similarly across different routine tasks, since almost definitionally no specific skills are required (Martellini and Menzio, 2021, p. 340). By contrast complex tasks like investigating, teaching or coordinating work offer greater scope for productivity variation across tasks, as they demand cognitive skills that typically require specific training and which cannot be executed by following pre-specified rules (cf. Autor *et al.*, 2003).[31] Thus, the share of complex tasks involved in production can be viewed as a proxy for the extensive margin of skill specificity (Deming, 2017). The proxy is useful because we can construct it, model-free, from data on the task content of production.

Concretely, building on Spitz-Oener (2006), I use multiple waves from the BIBB/IAB

[30]To be clear, this proxy does not map one-to-one to the theoretical task-based setup. Unfortunately, analytical tractability is lost with two types of tasks, essentially because a mixture of two Fréchets with different shape parameters is not Fréchet.

[31]Consistent with this idea, Caplin *et al.* (2022) find the time needed to reach maximal productivity to be highest for management roles or knowledge-intensive occupations, which tend to involve complex tasks.



(a) Task complexity and sorting **(b)** Task complexity and β_c **(c)** β_c and sorting

Figure 6: Industry-level evidence is consistent with the model mechanisms

Notes. Panel (a) is a scatter plot; the data are collapsed across periods. Panel (b) and (c) are binned scatter plots, controlling for industry FEs.

and BIBB/BAuA Employment Surveys to construct an index capturing the share of complex tasks among a worker's activities. Online Appendix B.1.3 details the survey question, how I classify tasks, and how the individual level measures are merged into the SIEED-based main dataset. In brief, the industry-level variation in task complexity in this section reflects differences in occupational employment shares.

Results. Figure 6a demonstrates that industries involving more complex tasks, such as architectural and engineering activities, exhibit more pronounced talent sorting than industries where routine tasks dominate, like accommodation services or furniture manufacturing. Panels 6b and 6c unpack this positive association through the lens of the theory. They demonstrate that, empirically, greater task complexity is associated with a higher value of β_c, which in turn predicts more positive assortative matching of coworkers based on talent. While these two panels are based on *within-industry* variation, Online Appendix B.2.3 shows that the picture looks similar using within-period variation across industries. Finally, we can exploit the industry-level variation to validate the task complexity proxy for skill specificity. Online Appendix Figure C.4 shows that, at the industry-level, χ_s is positively correlated with task complexity. In summary, industry-level variation strongly corroborates the model's core mechanisms as well as the use of task complexity as a proxy for skill specificity χ.

4 Applications

I now proceed to show, in two applications, how the theory sheds light on the implications of a rise in skill specificity on the evolution of wage inequality and what it tells us about the determinants of aggregate productivity. I conclude with a discussion of the model's shortcomings and future research directions.

4.1 Growing skill specificity

Both applications focus on the implications of deepening skill specificity (χ). The idea that skills have become more specialized over recent decades has been articulated across multiple strands of research, though thus far it has largely been treated qualitatively.[32] I briefly discuss three lines of reasoning in support of the idea of rising skill specificity and document supportive evidence before turning to a model-based analysis of the implications of such a shift.

First, Deming (2017) documents an increased return to social skills, which he traces back to an increase in skill specificity. Deming (2017) argues that this is due to jobs becoming less routine over time, appealing to the link between task complexity and skill specificity discussed in Section 3.4. Aligned with this idea, and using the same BIBB Data as before, Figure 7a depicts an upward trend in the share of complex tasks reported by individuals, increasing from 0.252 in 1986 to 0.647 in 2018, with an especially pronounced rise between the 1990s and early 2000s, manifesting across all education groups.

Second, Jones (2009) argues that the "burden of knowledge" – the growing cost of reaching the frontier – necessitates increasingly narrow individual expertise in science. This logic extends to much of knowledge work more broadly (Neffke, 2019). For instance, returning to the example of medical work introduced in the opening paragraph, Figure 7b shows that the number of distinct specialty certificates issued by the American Board of Medical Specialties nearly doubled from 1980 to 2020.

A third reason to believe that skill specificity has increased is the sharp acceleration in secondary education since the 1980s. According to the Barro and Lee (2013) dataset, the average years of secondary schooling in the German population aged 25+ increased from 1.32 years in 1985 to 6.9 years in 2010 after being relatively flat previously. Through

[32] Grigsby (2023) is the primary exception, showing that an increase in skill specificity, which in his model mutes cross-occupation labor supply spillovers, is important to understand why real wages have shifted from being countercyclical to mildly procyclical since the 2000s.

the lens of my theory, if more education augments task-specific skills randomly, this rise in schooling translates into more dispersion in task-specific skills. Indeed, under suitable functional form assumptions, this general reasoning can be shown to imply a rise in the skill specificity parameter χ.[33]

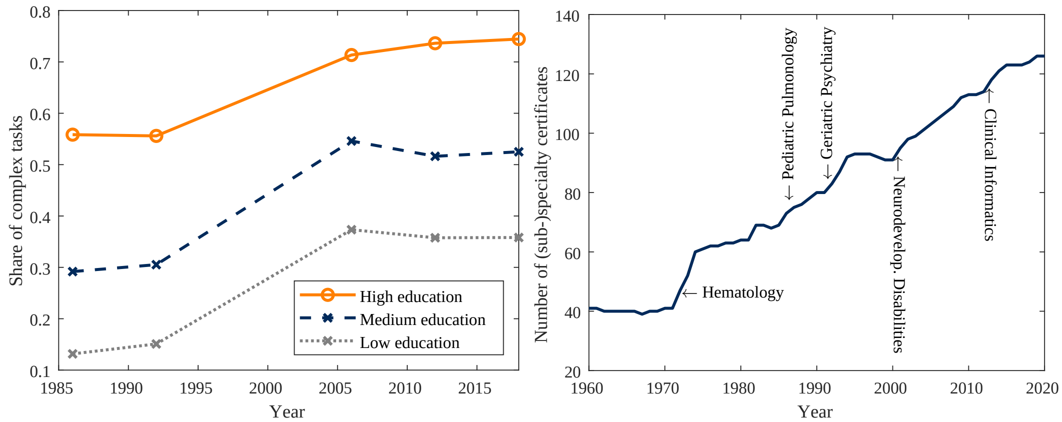
Discussion. Before moving on, it is worth stepping back to address two issues relating to the measurement and interpretation of trends in skill specificity χ. First, as we have no direct, cardinal measure of χ due to data constraints, I combine suggestive evidence with a structural approach to estimating χ. Second, I do not attempt disentangling disentangle technological shifts in task composition and human capital acquisition decisions as alternative drivers of rising χ; I view them as intricately intertwined. For example, one potentially important factor is automation, displacing humans in routine tasks and leaving us to handle the more complex problems (Acemoglu and Restrepo, 2018). Handling these complex problems plausibly involves larger fixed learning costs, incentivizing the acquisition of specialized skills (Rosen, 1983; Alon, 2018), and affords more ample opportunities for learning-by-doing.[34] Bartel *et al.* (2007, esp. pp. 1755-1756) instructively describe how the introduction of computer numerically controlled (CNC) machines shifted requirements from routine to specialized problem-solving skills.

4.2 A model-based explanation for the "firming up" of inequality

Firms are increasingly viewed as playing a central role in the evolution of wage inequality. An extensive empirical literature documents that, across many advanced economies, a prominent feature of wage dispersion and its rise since the 1980s is the large and increased share attributable to between-firm differences in pay. Despite figuring prominently in academic and policy discussions, what we should infer from this so-called "firming up" of

[33]Let Z be a Fréchet random variable (r.v.) with shape parameter $\theta > 0$ and scale parameter $x > 0$, and let $\{B_n\}_{n\geq 1}$ be a sequence of independent r.v.'s defined recursively as $B_n = \exp(-b_n/(\alpha\theta_{n-1}))$ where $\alpha \in (0,1)$, $\theta_0 = \theta$, $\theta_n = \theta_{n-1}\alpha = \theta\alpha^n$ for $n \geq 1$, $\{b_n\}_{n\geq 1}$ are independent r.v.'s such that $\exp(b_n/\alpha)$ are i.i.d. positive α-stable r.v.'s. Assume Z and $\{B_n\}$ are independent. Define the r.v.'s $\{Z^{(n)}\}_{n\geq 1}$ recursively as $Z^{(0)} = Z, \quad Z^{(n)} = Z^{(n-1)} \times B_n, \quad n \geq 1$. Then for each $n \geq 1$, $Z^{(n)}$ is a Fréchet r.v. with scale x and shape $\theta_n = \theta\alpha^n$. This result is a straightforward extension of Theorem 1 in Shanbhag and Sreehari (1977) and Boehm and Oberfield (2022, Footnote 11). Intuitively, if an individual starts a school year with a certain distribution of task-specific skills and leaves with a distribution where each task-specific skill is multiplied by an independent draw from a fat-tailed distribution, with α controlling the dispersion in the multiplying draws, the more years of schooling they have, the more specific their skills will be, χ being the inverse of the Fréchet shape parameter. I thank Ezra Oberfield for suggesting this idea.

[34]The broader point is that "skill" and "technology" are not neatly separable in a task-based setting, different from theories where education or experience are sufficient statistics for a worker's human capital.



(a) Rising task complexity

(b) Deepening medical specialization

Figure 7: Rising economy-wide task complexity & deepening medical specialization

Notes. Panel 7a is based on the BIBB data and depicts the average share of complex, or abstract non-routine, tasks in individual workers' set of activities for four points in time and distinguishing between three education groups. Panel 7b is sourced from the American Board of Medical Specialities. For each year, it shows the number of unique speciality or sub-speciality certificates that have been approved and issued at least once by that year and which are are still being issued.

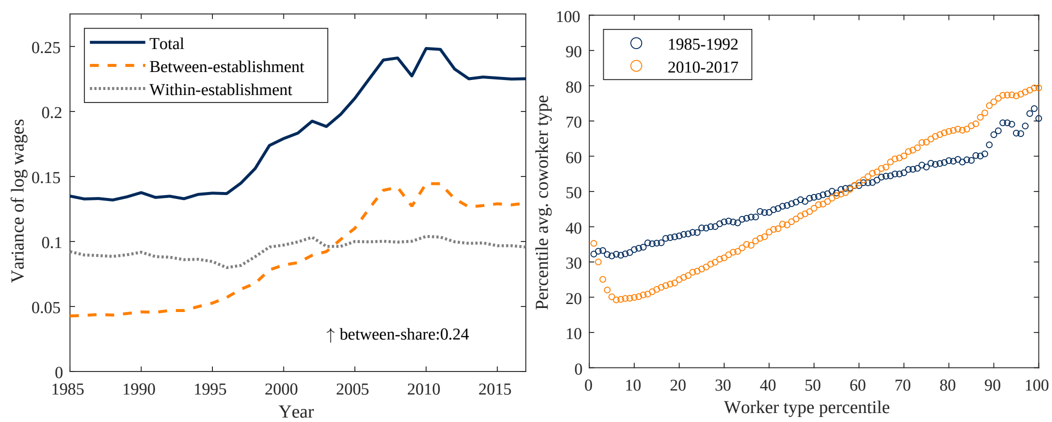
inequality (Song *et al.*, 2019) for our understanding of the rise in inequality and potential policy responses depends on the underlying structural drivers. Those remain obscure.

My theory offers a parsimonious explanation for these trends in the firm-level structure of wage inequality: Growing skill specificity has amplified coworker talent complementarities, leading to more sorting and, hence, greater firm-level wage inequality.

4.2.1 Empirical evolution of labor market inequality in Germany

I start by revisiting evidence on the evolution of wage inequality in the context of Germany and document an upward trend in coworker talent sorting.

Wage distribution. Wage inequality in Germany has risen substantially since the mid-1980s, and this increase is primarily accounted for by widening pay gaps between firms. Figure 8a presents the yearly total variance of log wages (solid line) and decomposes it into between-employer (dashed) and within-employer (dotted) component. By the law of total variance, the total variance equals the sum of the two components. It rose from an average of 0.13 in the interval 1985-1992 to 0.23 during 2010-2017. Of this 0.1 point increase, around 90% are accounted for by the between-employer component. Put differently, the *share* of the total log wage variance due to between-employer differences rose increased from 33% in 1985-1992 to 57% during 2010-2017. This evidence is consistent with the results documented by Card *et al.* (2013) and Song *et al.* (2019), among others.



(a) Wage inequality

(b) Coworker sorting

Figure 8: The evolution of wage inequality and coworker sorting

Notes. The left panel shows the variance of log (residual) wages, decomposed into between-employer (the person-weighted variance of firm-level average log wages) and within-employer components. The change in the between-share indicated compares the average over 1985-1992 and 2010-2017. The right panel plots, for any percentile of the worker type distribution the percentile rank of the average coworker type. Workers are ranked economy-wide. For visual clarity, types are grouped into 50 cells, then coworker quality is computed for each cell.

Coworker sorting. Furthermore, highly productive workers increasingly cluster in the same establishments, segregated from other, less productive workers, who cluster in different workplaces. To establish this point, I construct worker and (representative) coworker types separately for 5 sample periods. Table 3 reports the period-specific, empirical coworker correlation. Column (1) reveals that coworker sorting has risen from an average of 0.38 during 1985-1992 to 0.64 during 2010-2017, a 68% increase. Column (2) displays same statistic when workers are ranked within occupation, thus controlling for occupational composition effects. The increase is smaller (+0.16 instead of +0.26), suggesting that the baseline measure partly picks up enhanced segregation along occupational lines, but still very notable – a 50% rise – indicating that top-performing workers *within* each occupation are increasingly likely to work together. Figure 8b visualizes this trend as a binscatter plot.

4.2.2 Model-based analysis

I next use the model to show that rising skill specificity can account for these trends.

Calibration. I re-calibrate the model for each of the 4 sample periods prior to 2010-2017, reaching back to 1985. This is possible thanks to the SIEED's long coverage.

	Coworker sorting	
Period	Within-economy type ranking	Within-occupation type ranking
1985-1992	0.38	0.32
1993-1997	0.44	0.37
1998-2003	0.51	0.42
2004-2009	0.58	0.45
2010-2017	0.64	0.48

Table 3: Coworker sorting over time

Notes. This table indicates the correlation coefficient between a worker's estimated type and that of their average coworker, computed separately for five sample periods.

Concretely, I first estimate equation (25) period-by-period to recover the interaction coefficient β_c. I use these moments to discipline the skill specificity parameter χ when re-estimating the parameter vector ψ by targeting the same empirical moments as for 2010-2017 but measured for earlier time periods. Two comments about the targeted moments are warranted. First, considering the replacement rate, I summarize the multiple dimensions of the Hartz reforms as a reduction of the monetary replacement rate by around 10%, similar to Jung *et al.* (2023). This implies a target equal to 0.69 for the first three intervals, compared to 0.63 for the latter two. Second, to allow for the possibility of a decline in the strength of search frictions, as argued e.g. by Martellini and Menzio (2020), I allow the labor market transition rates to change across sample periods, too.

Turning to estimation results, I find an increase over time in the strength of skill specificity. Figure 9 displays the evolution of the point estimates for β_c and the indirectly inferred value of χ alongside the task complexity proxy. It can be seen that the estimates for β_c, as a reduced-form moment, and hence χ, as the underlying structural parameter, increase over time and co-move with the task-complexity proxy.[35] The full set of parameter estimates is reported in Online Appendix Table C.1. Of note, the production parameter a_0 has decreased and a_1 has increased over time, which is consistent with technological change that has made output more sensitive to skill. In addition, increases over time in both job arrival and job separation rates may reflect the emergence of online job portals (Bhuller *et al.*, 2023).

Model-implied trends in wage dispersion. The structural model qualitatively replicates the empirically recorded evolution of between- and within-firm inequality. Figure 10a

[35]The finding that β_c has increased over item is robust to the same checks discussed in Section 3.2.3.

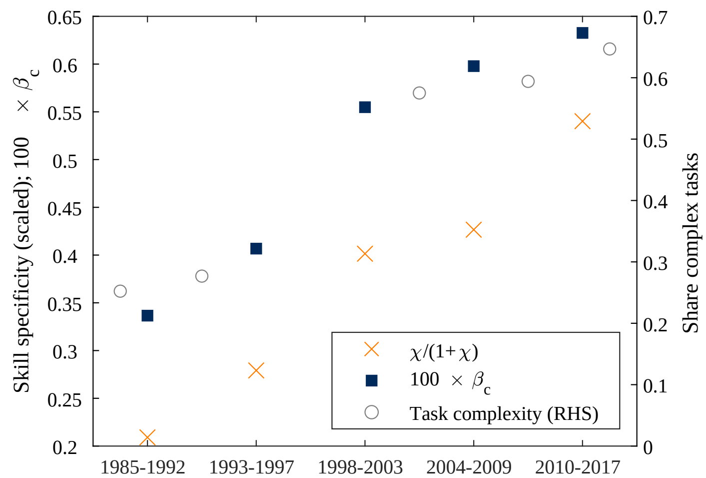


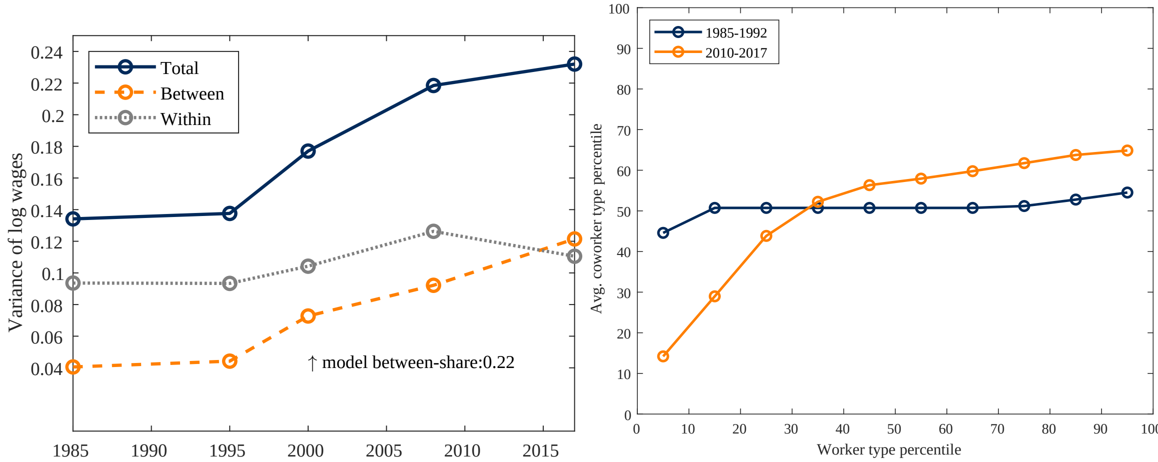
Figure 9: Skill specificity has increased alongside task complexity

Notes. This figure reports the estimated value of skill specificity χ (shown as $\frac{\chi}{\chi+1}$) alongside the point estimate for the coefficient β_c ($\times$ 100) estimated from regression (25), separately for five sample periods. In addition, the circles reproduce the share of complex tasks in workers' activities from Figure 7a. (The years of the survey waves and the sample split in the matched employer-employee data do not align perfectly, so the task measures are placed approximately at the mid-points of the closest sample period.

depicts the model-implied decomposition of the variance of log wages for each of the sample periods. In spite of only the total variance of log wages having been targeted, the model captures almost the entire empirically recorded increase in the between-firm share (+0.22 vs. 0.24). Furthermore, aligned with the evidence, the increased between-firm share goes along with more positively assortative matching, as illustrated in Figure 10b which compare 2010-2017 with 1985-1992. The upward tilt in the mapping between worker talent and coworker talent echoes the data patterns presented in Figure 8b: High types increasingly pair up among themselves, as do lower types. A limitation of the model is that it understates the degree of coworker sorting in the earlier years.

Beyond wage disparities across workplaces, the model indicates that enhanced talent sorting also contributed to widening labor productivity gaps between firms – a trend widely discussed in the firm dynamics literature (e.g., Decker *et al.*, 2020; Sorkin and Wallskog, 2021). Online Appendix Figure C.5 shows that the model-implied cumulative distribution of log labor productivity for 2010-2017 exhibits more density at both the lower and the upper end of the distribution compared to 1985-1992. We can infer, therefore, that increased coworker complementarities also reinforced firm-level productivity dispersion,

Counterfactual analysis. What portion of the model predicted rise in the between-firm share of wage inequality can be attributed to stronger skill specificity? The answer



(a) Variance decomposition **(b)** Coworker sorting

Figure 10: Evolution of wage inequality and sorting in the model

Notes. In the left panel, the solid lines indicate the model-predicted variance of log wages, decomposed into between- and within-employer components. The model-generated between-within decomposition is corrected for a mechanical bias, as described in Appendix C.1.3. The right panel plots the average coworker talent type (vertical axis) for each worker talent type (horizontal axis), for two separate sample periods.

is not obvious since each of the internally estimated parameters changed, potentially influencing the model implied decomposition. The theory implies, for instance, that a higher job arrival rate amplifies sorting, as workers can accept offers more selectively. Additionally, technological change that amplifies the return to talent can mechanically lead to greater between-firm inequality, even holding the distribution of workers constant, provided it exhibits positive sorting: It would elevate the relative remuneration of high-earning individuals, who cluster together.[36]

Counterfactual model exercises indicate that the increase in χ is an important driver of elevated firm-level wage inequality. To quantify the mechanism, I consider the following counterfactual scenario: What would the between-firm share of the variance of log wages have been in the 2010s had χ remained constant at its 1985-1992 level? I then compute the implied difference in the between-share across the factual and the counterfactual scenarios. Row 2 in Table 4 summarizes the analysis. Absent a rise in χ, the model predicts, the between-share would have risen only by 0.077 percentage points rather than 0.221. Thus, the rise in χ can account for about 65% of the model-predicted

[36]Within the AKM-framework, increased return to talent could mechanically magnify both the worker-firm sorting and the worker segregation components. Song *et al.* (2019) attribute 9% and 35%, respectively, of the increase in worker-firm sorting and worker segregation in the U.S. to these effects.

	Δ model	Implied % Δ model due to Δ par
Baseline	0.221	–
Cf.: χ fixed	0.077	65.2%
Cf.: labor mkt. transition rates fixed	0.198	10.6%
Within-occ. ranking	0.176	-
Cf. χ fixed	0.005	98.0%
Within-industry analysis	0.19	-
Cf. χ fixed	0.004	98.0%

Table 4: Counterfactuals: explaining the rise in the between-firm share of wage inequality

Notes. This table summarizes the counterfactual ("Cf.") exercises evaluating changes in the between-firm share of wage inequality. The second column indicates the change in the between-share from 1985-1992 to 2010-2017 for the model specified in the first column. The final column is computed as follows, using the example of "Cf. a:". Denoting by m_j the model-implied between-share, the value is equal to $100 \times \left(1 - \frac{m_j(\chi^{p1}, \psi_{-1}^{p2}) - m_j(\psi^{p1})}{m_j(\psi^{p2}) - m_j(\psi^{p1})}\right)$, where χ^{pt} is the value of χ estimated for period pt, $t \in \{1, 2\}$, and ψ_{-1}^{pt} collects all other parameters.

rise in the between-firm share of wage inequality.

Shifts in the production technology represent a more important factor than changes in search technology, though both are consequential. For this comparison, I consider a further counterfactual, imposing that both job arrival and separation rates, λ_u and δ, had remained at their 1985-1992 levels. This exercise suggests that their joint increase can explain 10.6% of the model-predicted rise in the between-share.[37]

Sensitivity analysis. Robustness exercises suggests that the results are robust to two important points of concern, relating to occupational and industry dynamics. Table 4 summarizes the results from these counterfactual exercises.

One concern is that the measurement underlying the baseline figures is confounded by outsourcing dynamics or related shifts in the boundary of the firm (e.g. Song *et al.*, 2019, Section V.B.). To address this possibility, I consider the aforementioned alternative of ranking workers within-occupation as opposed to economy-wide. Under this specification, an increasing in sorting, for instance, could not be driven by greater sorting of high- and low-wage occupations into different establishments. I estimate the model parameters targeting β_c estimates obtained under this approach (see Online Appendix

[37]The relatively limited contribution made by declining search frictions is consistent with Kantenga and Law's (2016) study. A caveat regarding this exercise is that the transition rates are treated as exogenous in the structural model, despite being plausibly endogenous to technological change. Intuitively, an increase in χ, by rendering mismatch more costly, would itself incentivize greater search effort. The increase in meeting rates is consistent with this mechanism, but examining it further is beyond the scope of this paper.

Table C.2). The analysis yields a model-predicted increase in the between-firm share equal to +0.176, of which the great majority is attributed to increased skill specificity.

A second worry is that the analysis so far has abstracted from industry-level differences in the production function (e.g. Haltiwanger and Spletzer, 2020). I therefore repeat the calibration and counterfactual exercise, this time targeting the *average within-industry average* of the variance of log wages and β_c estimated at the two-digit industry level (see Online Appendix Table C.3 for parameter estimates). As summarized in Table 4, under this alternative specification, the rise in skill specificity explains almost the entirety of the empirically observed rise in the between-firm within-industry share of wage dispersion.

Person-level inequality and comparison to Kremer (1993). The theory also clarifies that rising skill specificity, talent complementarities and coworker sorting do not *by themselves* exacerbate person-level inequality. This may seem surprising in light of Kremer (1993), which is sometimes understood as saying that amplified complementarities lead to a convexification of the wage function. However, this prediction hinges on the assumption that production exhibits increasing returns to team quality (cf. Kremer, 1993, III). Here I instead consider the constant-returns case, thus isolating the role of complementarities. Nor is it the case that rising coworker sorting by itself leads to higher person-level inequality. In Kremer (1993) changes in sorting play no role at all – sorting is *always* perfect (i.e. $\rho_{xx} = 1$) – and in this model, increasing talent sorting leads to a rise in between-firm wage inequality that is offset by a fall in within-firm inequality. The takeaway from this model is, therefore, that a rise in χ together with talent-biased technological change – captured by lower a_0 and higher a_1 – *jointly* lead to a rise in person-level inequality driven by heightened firm-level wage dispersion.

Several extensions of the model could introduce channels through which greater coworker sorting amplifies person-level inequality. First, coworker sorting might weaken the effectiveness of institutions, both formal or informal (Akerlof and Yellen, 1990), that compress the within-firm pay distribution for heterogeneous workers. Second, greater sorting may dynamically foster lifetime inequality if coworker learning effects are important (Jarosch *et al.*, 2021; Herkenhoff *et al.*, 2024).

4.3 Labor market frictions impede gains from specialization

This section quantitatively shows how the aggregate productivity gains from skill specialization hinge on the 'right' teams being formed.

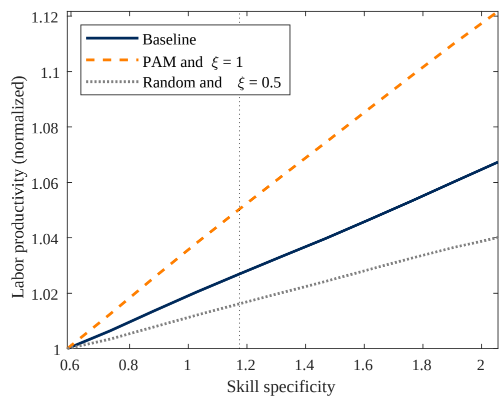


Figure 11: Productivity gains from growing specialization under alternative matching regimes

Quantitative analysis. To clarify the mechanism, I first consider economy parameterized for the 2010s and quantify the output costs of coworker mismatch. These costs are measured as the difference in average labor productivity between the equilibrium allocation and counterfactual scenarios with alternative coworker distributions, holding constant the marginal distribution of workers in teams $\phi(x)$. If sorting were perfectly assortative and horizontal distance $\xi = 1$ for every match, output would be 6% greater; keeping the joint talent distribution $\phi(x, x')$ at its equilibrium value but setting $\xi = 1$ raises output by 4.2%.

While this exercise held the skill specificity parameter χ fixed, the theory's most distinctive implications concern the effects of an *increase* in χ when accounting for both endogenous changes in production complementarities and labor market frictions. Intuitively, while increased skill specificity delivers a direct boost to productivity – under the division of labor, dispersion is desirable – the reduced-form production function (9) clarified that, at the same time, it renders output more vulnerable to mismatch.

Figure 11 shows how labor productivity varies with χ for different labor market allocations. To construct this figure, all other parameters are held constant at their baseline values and labor productivity is normalized at the lowest value of χ. The solid line represents the estimated German economy, the dashed line captures potential productivity and the dotted gray line represents a very frictional economy. The productivity frontier is calculated as above (PAM + $\xi = 1$), while the very frictional economy corresponds to a scenario where talent sorting is random and $\xi = 0.5$.

The figure reveals that labor market frictions can severely curtail the realized productivity benefits from greater skill specialization. In the baseline economy, roughly half of the potential gains are realized, and this number falls to about a third in the very frictional economy. The key to understand this result is that as χ increases, the team's joint output becomes increasingly vulnerable to coworker mismatch.

Discussion. Economic development tends to involve a trend toward greater skill specialization. This section implies to understand the implications of this trend it is important to account for the endogeneity of coworker complementarities and their interaction with labor market frictions. This finding aligns well with – and offers a microfoundation for – the cross-country evidence in Bandiera *et al.* (2024). Moreover, if labor market frictions limit the gains from specialization, this may help explain recent findings of limited specialization in poorer countries (Bassi *et al.*, 2023; Atencio-De-Leon *et al.*, 2024).

4.4 Discussion: shortcomings and directions for future work

The model imposes several simplifying assumptions. I briefly discuss the resulting shortcomings and highlight avenues for future research.

Firm size. Most conspicuously, firm size was limited to two workers, which precludes analyzing firm growth patterns. Relaxing this assumption would allow assessing the role of team quality in explaining large growth rate differences among young firms in particular (Sterk *et al.*, 2021). In addition, adding ex-ante firm heterogeneity in product quality and physical capital would facilitate a more comprehensive quantitative assessment of the relative importance of team quality in explaining firm-level productivity dispersion. A companion paper (Criscuolo *et al.*, 2024) documents reduced-form empirical evidence which suggests both avenues are promising.

On-the-job search. The model did not allow for on-the-job search (OJS). To incorporate this empirically salient feature is feasible, though it requires introducing ξ as an additional state variable. An enriched model would likely be better able to capture empirical sorting patterns, which, as mentioned above, the current model does far from perfectly. Introducing both "large firms" and OJS would also facilitate assessing structurally and quantitatively to what extent reduced-form AKM (Abowd *et al.*, 1999) estimates of "firm pay premia" pick up coworker effects.

Within-organization frictions. Enriching the organizational core of the model would open the door to studying additional implications of skill specialization and team production. For instance, the model in its current world abstracted from real-world coordination frictions, yet empirically some managers and some firms appear to be better in coordinating their workforce than others (Coraggio *et al.*, 2022; Minni, 2022; Kuhn *et al.*, 2023). An extension of my framework that introduces an explicit managerial role has the potential to capture this important feature of the data.

5 Conclusion

This paper developed a task-based theory of team production and formation – formalizing the idea of the firm as a "team assembly" – which is both rich enough to be confronted with micro data and remains sufficiently tractable for quantitative explorations.

The theory delivers four key insights. First, when production involves the division of labor among workers with specialized skills, coworkers' productivity is inherently non-separable. Production features coworker talent complementarities whose strength is endogenously increasing in the degree of skill specificity. Second, in frictional labor markets, skill specificity thus shapes the degree of dispersion in talent, productivity and wages across firms. Third, growing skill specificity since the mid-1980s has amplified complementarities, which helps explain the "firming up" of inequality. Fourth, while greater specialization potentially boosts productivity, frictions that impede the matching of coworkers with complementary skills can significantly limit realized gains.

I hope that the framework developed in this paper encourages and facilitates further empirical and theoretical research taking a team-centric approach to firms. Relevant directions include, among others, the role of managers as coordinators and that of of teams in explaining heterogeneous firm growth.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, **67**(2), 251–333.

Acemoglu, D. and Restrepo, P. (2018). The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review*, **108**(6), 1488–1542.

Adenbaum, J. (2022). Endogenous Firm Structure and Worker Specialization. *mimeo*.

Adhvaryu, A., Bassi, V., Nyshadham, A., and Tamayo, J. A. (2020). No Line Left Behind: Assortative Matching Inside the Firm.

Akcigit, U., Caicedo, S., Miguelez, E., Stantcheva, S., and Sterzi, V. (2018). Dancing with the Stars: Innovation Through Interactions.

Akerlof, G. A. and Yellen, J. L. (1990). The Fair Wage-Effort Hypothesis and Unemployment. *The Quarterly Journal of Economics*, **105**(2), 255–283.

Alchian, A. A. and Demsetz, H. (1972). Production, Information Costs, and Economic Organization. *The American Economic Review*, **62**(5), 777–795.

Alon, T. (2018). Earning more by doing less: Human capital specialization and the college wage premium. *manuscript, University of California at San Diego*.

Anderson, S. P., De Palma, A., and Thisse, J.-F. (1989). Demand for Differentiated Products, Discrete Choice Models, and the Characteristics Approach. *The Review of Economic Studies*, **56**(1), 21–35.

Andrews, D., Criscuolo, C., and Gal, P. N. (2019). The Best versus the Rest: Divergence across Firms During the Global Productivity Slowdown. *Working Paper*.

Antonczyk, D., Fitzenberger, B., and Leuschner, U. (2009). Can a Task-Based Approach Explain the Recent Changes in the German Wage Structure? *Jahrbücher für Nationalökonomie und Statistik*, **229**(2-3), 214–238.

Atalay, E., Phongthiengtham, P., Sotelo, S., and Tannenbaum, D. (2020). The Evolution of Work in the United States. *American Economic Journal: Applied Economics*, **12**(2), 1–34.

Atencio-De-Leon, A., Lee, M., and Macaluso, C. (2024). Does Turnover Inhibit Specialization? Evidence from a Skill Survey in Peru. *American Economic Review: Insights*.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, **135**(2), 645–709.

Autor, D., Chin, C., Salomons, A. M., and Seegmiller, B. (2022). New Frontiers: The Origins and Content of New Work, 1940–2018.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, **118**(4), 1279–1333.

Baley, I., Figueiredo, A., and Ulbricht, R. (2022). Mismatch Cycles. *Journal of Political Economy*, **130**(11), 2943–2984.

Bandiera, O., Kotia, A., Lindenlaub, I., Moser, C., and Prat, A. (2024). Meritocracy across Countries.

Barro, R. J. and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, **104**, 184–198.

Bartel, A., Ichniowski, C., and Shaw, K. (2007). How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *The Quarterly Journal of Economics*, **122**(4), 1721–1758.

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States. *Journal of Labor Economics*, **34**(S2), S67–S97.

Bassi, V., Lee, J. H., Peter, A., Porzio, T., Sen, R., and Tugume, E. (2023). Self-Employment within the Firm. *mimeo*.

Becker, G. S. and Murphy, K. M. (1992). The Division of Labor, Coordination Costs, and Knowledge. *The Quarterly Journal of Economics*, **107**(4), 1137–1160.

Berlingieri, G., Blanchenay, P., and Criscuolo, C. (2017). The great divergence(s). *OECD Discussion Paper*.

Bhuller, M., Ferraro, D., Kostøl, A. R., and Vigtel, T. C. (2023). The Internet, Search Frictions and Aggregate Unemployment.

Bilal, A. and Lhuillier, H. (2021). Outsourcing, Inequality and Aggregate Output.

Bilal, A., Engbom, N., Mongey, S., and Violante, G. L. (2022). Firm and Worker Dynamics in a Frictional Labor Market. *Econometrica*, **90**(4), 1425–1462.

Blackorby, C. and Russell, R. R. (1981). The Morishima Elasticity of Substitution; Symmetry, Constancy, Separability, and its Relationship to the Hicks and Allen Elasticities. *The Review of Economic Studies*, **48**(1), 147–158.

Blackorby, C. and Russell, R. R. (1989). Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities). *The American Economic Review*, **79**(4), 882–888.

Bloesch, J., Larsen, B., and Taska, B. (2022). Which Workers Earn More at Productive Firms? Position Specific Skills and Individual Worker Hold-up Power. *mimeo*.

Boehm, J. and Oberfield, E. (2022). Growth and the Fragmentation of Production. *mimeo*.

Boerma, J., Tsyvinski, A., and Zimin, A. (2021). Sorting with Team Formation. Technical Report w29290, National Bureau of Economic Research, Cambridge, MA.

Bonhomme, S., Lamadon, T., and Manresa, E. (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica*, **87**(3), 699–739.

Borovickova, K. and Shimer, R. (2024). Assortative Matching and Wages: The Role of Selection. *mimeo*.

Cairó, I. and Cajner, T. (2018). Human Capital and Unemployment Dynamics: Why More Educated Workers Enjoy Greater Employment Stability. *The Economic Journal*, **128**(609), 652–682.

Caliendo, L., Mion, G., Opromolla, L. D., and Rossi-Hansberg, E. (2020). Productivity and Organization in Portuguese Firms. *Journal of Political Economy*.

Caplin, A., Lee, M., Leth-Petersen, S., Saeverud, J., and Shapiro, M. D. (2022). How Worker Productivity and Wages Grow with Tenure and Experience: The Firm Perspective.

Card, D., Heining, J., and Kline, P. (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *The Quarterly Journal of Economics*, **128**(3), 967–1015.

Chade, H. and Eeckhout, J. (2020). Competing Teams. *The Review of Economic Studies*, **87**(3), 1134–1173.

Coraggio, L., Pagano, M., Scognamiglio, A., and Tåg, J. (2022). JAQ of All Trades: Job Mismatch, Firm Productivity and Managerial Quality. *SSRN Electronic Journal*.

Cornelissen, T., Dustmann, C., and Schönberg, U. (2017). Peer Effects in the Workplace. *American Economic Review*, **107**(2), 425–456.

Correia, S. (2017). Linear models with high-dimensional fixed effects: An efficient and feasible estimator. *Unpublished manuscript, http://scorreia. com/research/hdfe. pdf*.

Criscuolo, C., Hijzen, A., Schwellnus, C., Barth, E., Chen, W.-H., Fabling, R., Fialho, P., Stadler, B., Upward, R., Zwysen, W., Grabska-Romagosa, K., Kambayashi, R., Leidecker, T., Skans, O. N., Riom, C., and Roth, D. (2021). Workforce composition, productivity and pay: The role of firms in wage inequality. *OECD Discussion Paper*.

Criscuolo, C., Freund, L. B., and Gal, P. N. (2024). Human Capital at Work: Five Facts about the Role of Skills for Firm Productivity, Growth, and Wage Inequality. *OECD Productivity Working Papers*, **No. 36**.

Dauth, W. and Eppelsheimer, J. (2020). Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: A guide. *Journal for Labour Market Research*, **54**(1), 1–14.

Decker, R. A., Haltiwanger, J., Jarmin, R. S., and Miranda, J. (2020). Changing Business Dynamism and Productivity: Shocks versus Responsiveness. *American Economic Review*, **110**(12), 3952–3990.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics*, **132**(4), 1593–1640.

Dvorkin, M. and Monge-Naranjo, A. (2019). Occupation Mobility, Human Capital and the Aggregate Consequences of Task-Biased Innovations. *St. Louis Federal Reserve Bank Working Paper 2019-013C*.

Eaton, J. and Kortum, S. (2002). Technology, Geography, and Trade. *Econometrica*, **70**(5), 1741–1779.

Eeckhout, J. and Kircher, P. (2011). Identifying Sorting—In Theory. *The Review of Economic Studies*, **78**(3), 872–906.

Elsby, M. W. L. and Gottfries, A. (2022). Firm Dynamics, On-the-Job Search, and Labor Market Fluctuations. *The Review of Economic Studies*, **89**(3), 1370–1419.

Engbom, N. and Moser, C. (2022). Earnings Inequality and the Minimum Wage: Evidence from Brazil. *American Economic Review*, **112**(12), 3803–3847.

Engbom, N., Moser, C., and Sauermann, J. (2023). Firm pay dynamics. *Journal of Econometrics*, **233**(2), 396–423.

Freund, L. B. (2023). Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities. *Available at SSRN 4312245*.

Gabaix, X. and Landier, A. (2008). Why has CEO Pay Increased So Much? *The Quarterly Journal of Economics*, **123**(1), 49–100.

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

Garicano, L. and Rossi-Hansberg, E. (2006). Organization and Inequality in a Knowledge Economy. *The Quarterly Journal of Economics*, **121**(4), 1383–1435.

Gouin-Bonenfant, É. (2022). Productivity Dispersion, Between-Firm Competition, and the Labor Share. *Econometrica*, **90**(6), 2755–2793.

Grigsby, J. (2023). Skill Heterogeneity and Aggregate Labor Market Dynamics. *mimeo*.

Haanwinckel, D. (2023). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution.

Hagedorn, M., Law, T. H., and Manovskii, I. (2017). Identifying Equilibrium Models of Labor Market Sorting. *Econometrica*, **85**(1), 29–65.

Håkanson, C., Lindqvist, E., and Vlachos, J. (2021). Firms and Skills The Evolution of Worker Sorting. *Journal of Human Resources*, **56**(2), 512–538.

Hall, A. and et al. (2018). BIBB/IAB / BIBB/BAuA Employment Survey 2018 (2012, 2006, 1998/99, 1991/92, 1985/86). Technical Report doi:10.7803/501.18.1.1.10, doi:10.7803/501.12.1.1.60, doi:10.7803/501.06.1.1.20, doi:10.4232/1.12247, doi:10.4232/1.2565, 10.4232/1.12563, Research Data Centre at BIBB.

Haltiwanger, J. C. and Spletzer, J. R. (2020). Between Firm Changes in Earnings Inequality: The Dominant Role of Industry Effects. Working Paper 26786, National Bureau of Economic Research.

Helpman, E., Itskhoki, O., Muendler, M.-A., and Redding, S. J. (2017). Trade and Inequality: From Theory to Estimation. *The Review of Economic Studies*, **84**(1 (298)), 357–405.

Herkenhoff, K., Lise, J., Menzio, G., and Phillips, G. M. (2024). Production and Learning in Teams. *Econometrica*, **92**(2), 467–504.

Hicks, J. (1932). *The Theory of Wages*. Macmillan, London.

Houthakker, H. S. (1955). The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis. *The Review of Economic Studies*, **23**(1), 27–31.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation.

Jäger, S. and Heining, J. (2022). How Substitutable Are Workers? Evidence from Worker Deaths. *mimeo*.

Jarosch, G. (2023). Searching for Job Security and the Consequences of Job Loss. *Econometrica*, **91**(3), 903–942.

Jarosch, G., Oberfield, E., and Rossi-Hansberg, E. (2021). Learning From Coworkers. *Econometrica*, **89**(2), 647–676.

Jones, B. F. (2009). The Burden of Knowledge and the “Death of the Renaissance Man”: Is Innovation Getting Harder? *The Review of Economic Studies*, **76**(1), 283–317.

Jung, P., Korfmann, P., and Preugschat, E. (2023). Optimal regional labor market policies. *European Economic Review*, **152**, 104318.

Kantenga, K. and Law, T.-H. (2016). Sorting and Wage Inequality. *mimeo*.

Koenig, F., Manning, A., and Petrongolo, B. (2021). Reservation Wages and the Wage Flexibility Puzzle. *SSRN Electronic Journal*.

Kohlhepp, J. (2022). The Inner Beauty of Firms. *mimeo*.

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

Kremer, M. and Maskin, E. (1996). Wage Inequality and Segregation by Skill. Working Paper 5718, National Bureau of Economic Research.

Kuhn, M., Luo, J., Manovskii, I., and Qiu, X. (2023). Coordinated Firm-Level Work Processes and Macroeconomic Resilience. *Journal of Monetary Economics*.

Lentz, R. and Mortensen, D. T. (2010). Labor Market Models of Worker and Firm Heterogeneity. *Annual Review of Economics*, **2**(1), 577–602.

Lind, N. and Ramondo, N. (2023). Trade with Correlation. *American Economic Review*, **113**(2), 317–353.

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

Lindenlaub, I. and Postel-Vinay, F. (2023). Multidimensional Sorting under Random Search. *Journal of Political Economy*, **131**(12), 3497–3539.

Lopes de Melo, R. (2018). Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence. *Journal of Political Economy*, **126**(1), 313–346.

Lucas, R. (1978). On the Size Distribution of Business Firms. *Bell Journal of Economics*, **9**(2), 508–523.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, **60**(3), 531–542.

Martellini, P. and Menzio, G. (2020). Declining Search Frictions, Unemployment, and Growth. *Journal of Political Economy*, **128**(12), 4387–4437.

Martellini, P. and Menzio, G. (2021). Jacks of All Trades and Masters of One: Declining Search Frictions and Unequal Growth. *American Economic Review: Insights*, **3**(3), 339–352.

Minni, V. (2022). Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs. *Job Market Paper*.

Morishima, M. (1967). “A few Suggestions on the Theory of Elasticity". *Keizai Hyoron (Economic Review)*, **16**, 144–150.

Neffke, F. M. H. (2019). The Value of Complementary Co-Workers. *Science Advances*, **5**(12), eaax3370.

Nix, E. (2020). Learning spillovers in the firm. Technical report, Working Paper.

Ocampo, S. (2022). A task-based theory of occupations with multidimensional heterogeneity. Technical report, mimeo.

Pearce, J. (2022). Idea Production and Team Structure. *Working paper*.

Porzio, T. (2017). Cross-Country Differences in the Optimal Allocation of Talent and Technology. *mimeo*.

Rohrbach-Schmidt, D. and Tiemann, M. (2013). Changes in workplace tasks in Germany—evaluating skill and task measures. *Journal for Labour Market Research*, **46**(3), 215–237.

Rosen, S. (1983). Specialization and Human Capital. *Journal of Labor Economics*, **1**(1), 43–49.

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.

Shanbhag, D. N. and Sreehari, M. (1977). On certain self-decomposable distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **38**(3), 217–222.

Song, J., Price, D. J., Guvenen, F., Bloom, N., and von Wachter, T. (2019). Firming Up Inequality. *The Quarterly Journal of Economics*, **134**(1), 1–50.

Sorkin, I. and Wallskog, M. (2021). The Slow Diffusion of Earnings Inequality. *Working Paper*.

Spitz-Oener, A. (2006). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics*, **24**(2), 235–270.

Sterk, V., Sedláček, P., and Pugsley, B. (2021). The Nature of Firm Growth. *American Economic Review*, **111**(2), 547–579.

Tian, L. (2021). Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil. *mimeo*.

vom Berge, P., Schmidtlein, L., Seth, S., Graf, T., Griessemer, S., Kaimer, S., Koehler, S., Lehner, C., Oertel, M., and Seysen, C. (2020). The Sample of Integrated Employer-Employee Data (SIEED): SIEED 7518, Version 1. *Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB)*.

Superstar Teams

Appendix - For Online Publication

Lukas B. Freund§

This online appendix contains supplemental material for the paper "Superstar Teams." Any references to sections, equations, figures, tables, assumptions, propositions, lemmas or corollaries that are not preceded by a capital letter refer to the main article.

A Theoretical appendix

A.1 Proofs and derivations for team-production model

Following the introduction of auxiliary definitions and lemmas (Section A.1.1), this section states Lemma A.3 and proves Proposition 1. The proofs exploit that the team production problem is cast so as parallel the structure of an Eaton and Kortum (2002) style trade model, allowing for correlation across producers' task-specific skills (Lind and Ramondo, 2023), and formulated as a planner problem.

A.1.1 Auxiliary definitions and lemmas

Definition A.1 (Fréchet distribution)**.** *A random variable Z has a Fréchet distribution with scale parameter $s > 0$ and shape parameter $\alpha > 0$ if $P[Z \leq z] = e^{-\left(\frac{z}{s}\right)^{-\alpha}}$.*

Lemma A.1. *Let the random variable Z be Fréchet-distributed with scale parameter s and shape parameter $\alpha > 1$. Then $\mathbb{E}[Z] = x\Gamma\left(1-\frac{1}{\alpha}\right)$, where $\Gamma(r) = \int_0^\infty t^{r-1}e^{-t}dt$ is the Gamma function.*

Proof.

$$\begin{aligned}\mathbb{E}[Z] &= \int_0^\infty z\frac{\partial}{\partial z}\mathrm{P}[Z \leq z]\,dz = \int_0^\infty z\left(\alpha\frac{1}{s}\left(\frac{z}{s}\right)^{-\alpha-1}e^{-\left(\frac{z}{s}\right)^{-\alpha}}\right)dz \\ &= \int_0^\infty ze^{-\left(\frac{z}{s}\right)^{-\alpha}}\left(\alpha s^{\alpha}(z)^{-\alpha-1}\right)dz = s\int_0^\infty t^{-1/\alpha}e^{-t}dt = s\Gamma(1-\frac{1}{\alpha}).\end{aligned}$$

□

§*Email*: lukas.beat.freund@gmail.com. *Web*: www.lukasfreund.com.

Definition A.2 (Multivariate Fréchet distribution)**.** *A random vector* $(Z_1,...,Z_n)$ *has a multivariate Fréchet distribution with a vector of scale parameters* $(s_1,...,s_n) \in \mathbb{R}_{++}$*, shape parameter* $\alpha > 0$ *and correlation parameter* $\xi \in (0,1]$*, if*

$$P[Z_1 \leq z_1, \ldots, Z_n \leq z_n] = \exp\left[-\left(\sum_{i=1}^{n}\left(\frac{z_i}{s_i}\right)^{-\frac{\alpha}{\xi}}\right)^{\xi}\right].$$

The term "multivariate Fréchet distribution" is used here to refer to the specific, symmetric dependence structure above; in principle a different or more flexible dependence structure could be considered.

Lemma A.2. *Let the random vector* $(Z_1,...,Z_n)$ *be distributed multivariate-Fréchet with scale parameters* $(s_1,...,s_n)$*, shape parameter* α *and correlation parameter* ξ*. Then for any* $B_i > 0$*,* $i = 1,...,n$ *and* $\beta > 0$*, the random variable* $\max_{i=1,...n} B_i Z_i^{\beta}$ *is Fréchet distributed with scale* $\left(\sum_{i=1}^{n}\left(s_i^{\alpha}B_i^{\alpha/\beta}\right)^{1/\xi}\right)^{\xi\beta/\alpha}$ *and shape* α/β*.*

Proof.

$$\begin{aligned}
\mathrm{P}\left[\max_{i=1,\ldots,n} B_i Z_i^{\beta} \leq p\right] &= \mathrm{P}\left[Z_1 \leq \left(p/B_1\right)^{1/\beta}, ..., Z_n \leq \left(p/B_n\right)^{1/\beta}\right] \\
&= \exp\left[-\left(\sum_{i=1}^{n}\left(s_i^{\alpha}B_i^{\alpha/\beta}\right)^{\frac{1}{\xi}}\right)^{\xi} p^{-\alpha/\beta}\right].
\end{aligned}$$

□

A.1.2 Lemma: task shares and shadow cost index

Lemma A.3. *Under the optimal task assignment, assuming* $\eta > 1$ *and a symmetric horizontal distance across coworker pairs,* ξ*:*

(i) *the shadow cost index is*

$$\lambda = \left(\sum_{i=1}^{n}\left(\frac{x_i}{\lambda_i^L}\right)^{\frac{1}{\chi\xi}}\right)^{-\chi\xi}; \tag{A.1}$$

(ii) *and the share of tasks for which worker i is the least-cost provider is*

$$\pi_i := P\left[\lambda_i(\tau) = \min_{k=1,..,n} \lambda_k(\tau)\right] = \frac{\left(x_i/\lambda_i^L\right)^{\frac{1}{\chi\xi}}}{\sum_{k=1}^n \left(x_k/\lambda_k^L\right)^{\frac{1}{\chi\xi}}}, \tag{A.2}$$

which furthermore corresponds to the fraction of the shadow value of all tasks used in final good production accounted for by i:

$$\pi_i = Q_i/Q \tag{A.3}$$

where $Q_i := \int_{\mathcal{T}} \lambda(\tau) y_i(\tau) d\tau$.

Proof. For Part (i), start with the expression for the shadow-price index in equation (7) and substitute for $\lambda(\tau)$ using equation (8):

$$\lambda = \left[\int_0^\infty (\lambda(\tau))^{1-\eta} \, d\tau\right]^{\frac{1}{1-\eta}} = \left[\int_0^\infty \left(\min_{i=1,..,n} \frac{\lambda_i^L}{z_i(\tau)}\right)^{1-\eta} d\tau\right]^{\frac{1}{1-\eta}} = \left[\mathbb{E}\left[\max_{i=1,...,n} \left(\frac{Z_i}{\lambda_i^L}\right)^{\eta-1}\right]\right]^{\frac{1}{1-\eta}}$$

$$= \left[\left(\sum_{i=1}^n \left(\frac{x_i}{\lambda_i^L}\right)^{\frac{1}{\chi\xi}}\right)^{-\xi\chi(1-\eta)} \iota^{\eta-1}\Gamma(1+\chi(1-\eta))\right]^{\frac{1}{1-\eta}}$$

$$= \left(\sum_{i=1}^n \left(\frac{x_i}{\lambda_i^L}\right)^{\frac{1}{\chi\xi}}\right)^{-\chi\xi},$$

where $\Gamma(\cdot)$ is the Gamma function, evaluated at the argument $1 + \chi(1-\eta) > 0$. The third equality uses that λ_i^L and $z_i(\tau)$ are strictly positive and that $\eta > 1$, and the fourth equality follows from Lemma A.2, used to derive the distribution of the random variable $\max_{i=1,...,n} \left(Z_i/\lambda_i^L\right)^{\eta-1}$, followed by application of Lemma A.1.

For Part (ii), we start with the definition of task shares

$$\begin{aligned}\pi_i &:= \mathrm{P}\left[\lambda_i(\tau) = \min_{k=1,..,n} \lambda_k(\tau)\right] \\ &= \mathrm{P}\left[\left(Z_i/\lambda_i^L\right) = \max_{k=1,...,n} \left(Z_k/\lambda_l^L\right)\right]\end{aligned}$$

To ease notation, let $G(r_1,...,r_n) = \left(\sum_{i=1}^n r_i^{1/\xi}\right)^{\frac{1}{\xi}}$ so that the distribution defined in Assumption 1 can be written as $\exp\left(-G\left(\left(\frac{z_1}{\iota x_1}\right)^{-1/\chi},...,\left(\frac{z_n}{\iota x_n}\right)^{-1/\chi}\right)\right)$. Note that since $G(\cdot)$ is homogeneous of degree one, the partial derivatives $G_i(\cdot) = \partial G(\cdot)/\partial r_i = r_i^{\frac{1}{\xi}-1}\left(\sum_{i=1}^n r_i^{1/\xi}\right)^{\frac{1}{\xi}-1}$ are homogeneous of degree 0. Additionally, let $\tilde{x}_i = \iota x_i/\lambda_i^L$ for any i.

Consider initially the following expression for any z,

$$
\begin{aligned}
&\mathrm{P}\left[\max_{k=1,...,n}\frac{Z_k}{\lambda_k^L}\leq z \text{ and } \frac{Z_i}{\lambda_i^L} = \max_{k=1,...,n}\frac{Z_k}{\lambda_k}\right] = \mathrm{P}\left[\frac{Z_i}{\lambda_i^L}\leq z \text{ and } \frac{Z_k}{\lambda_k}\leq\frac{Z_i}{\lambda_i^L}, \forall k=1,...,n\right]\\
&= \int_0^z \frac{\partial}{\partial z_i}\exp\left(-G\left(\left(\frac{z_1}{\tilde{x}_1}\right)^{-\frac{1}{\chi}},...,\left(\frac{z_n}{\tilde{x}_n}\right)^{-\frac{1}{\chi}}\right)\right)\Bigg|_{z_1=t,...,z_n=t} dt\\
&= \int_0^z \frac{1}{\chi}z_i^{-\frac{1}{\chi}-1}\tilde{x}_1^{\frac{1}{\chi}}G_i\left(\left(\frac{z_1}{\tilde{x}_1}\right)^{-\frac{1}{\chi}},...,\left(\frac{z_n}{\tilde{x}_n}\right)^{-\frac{1}{\chi}}\right)\times\exp\left(-G\left(\left(\frac{z_1}{\tilde{x}_1}\right)^{-\frac{1}{\chi}},...,\left(\frac{z_n}{\tilde{x}_n}\right)^{-\frac{1}{\chi}}\right)\right)\Bigg|_{z_1=t,...,z_n=t} dt\\
&= \int_0^z \frac{1}{\chi}t^{-\frac{1}{\chi}-1}\tilde{x}_i^{\frac{1}{\chi}}G_i\left(\left(\frac{t}{\tilde{x}_1}\right)^{-\frac{1}{\chi}},...,\left(\frac{t}{\tilde{x}_n}\right)^{-\frac{1}{\chi}}\right)\times\exp\left(-G\left(\left(\frac{t}{\tilde{x}_1}\right)^{-\frac{1}{\chi}},...,\left(\frac{t}{\tilde{x}_n}\right)^{-\frac{1}{\chi}}\right)\right)dt\\
&= \tilde{x}_i^{\frac{1}{\chi}}G_i(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}})\int_0^z\frac{1}{\chi}t^{-\frac{1}{\chi}-1}\exp\left(-G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)t^{-\frac{1}{\chi}}\right)dt\\
&= \frac{\tilde{x}_i^{\frac{1}{\chi}}G_i(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}})}{G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)}\int_0^z\frac{1}{\chi}t^{-\frac{1}{\chi}-1}G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)\exp\left(-G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)t^{-\frac{1}{\chi}}\right)dt\\
&= \frac{(x_i/\lambda_i^L)^{\frac{1}{\chi\xi}}}{\sum_{k=1}^n(x_k/\lambda_k^L)^{\frac{1}{\chi\xi}}}\times\left[\exp\left(-G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)t^{-\frac{1}{\chi}}\right)\right]_0^z\\
&= \frac{(x_i/\lambda_i^L)^{\frac{1}{\chi\xi}}}{\sum_{k=1}^n(x_k/\lambda_k^L)^{\frac{1}{\chi\xi}}}\times\exp\left(-G\left(\tilde{x}_1^{\frac{1}{\chi}},...,\tilde{x}_n^{\frac{1}{\chi}}\right)z^{-\frac{1}{\chi}}\right),
\end{aligned}
$$

where line five exploits $G(\cdot)$ being h.o.d. one and $G_i(\cdot)$ being h.o.d. zero; the last equality follows because $\exp(-\infty) = 0$. Now letting $z\longrightarrow\infty$ yields the desired result.

Note, furthermore, that the max-stability property of the Fréchet distribution ensures that the conditional distribution of the maximum equals the unconditional distribution of the maximum. This ensures that $Q_i = \pi_i Q$, so that π_i also corresponds to fraction of the shadow value of all tasks used in final good production attributable to i, analogous to the reasoning in Eaton and Kortum (2002). □

A.1.3 Proof of Proposition 1

Letting $\tilde{\chi} = \chi\xi$, part (i) of Lemma A.3 implies, given the normalization $\lambda = 1$, that

$$1 = \sum_{i=1}^{n} \left(\frac{x_i}{\lambda_i^L} \right)^{\frac{1}{\tilde{\chi}}}. \tag{A.4}$$

Substituting this into the expression for π_i in part (ii) of Lemma A.3 yields

$$\pi_i = (x_i/\lambda_i^L)^{\frac{1}{\tilde{\chi}}}, \tag{A.5}$$

which says that the share of tasks performed by i is increasing in her talent and decreasing in the shadow cost of her labor.

Next we derive an expression for λ_i^L as a function of π_i and Y. The FOC for $l_i(\tau)$ given $y_i(\tau) > 0$ implies that

$$l_i(\tau) = (\lambda_i(\tau)\lambda_i^L) y_i(\tau).$$

By integrating over tasks and using the time constraint (4) we find that

$$\lambda_i^L = \int \lambda_i(\tau) y_i(\tau) d\tau.$$

Since $\lambda(\tau) = \lambda_i(\tau)$ if $y_i(\tau) > 0$, and given the definition of Q_i, it follows that

$$\lambda_i^L = Q_i,$$

which says that the shadow value of worker i's time is equal to the shadow value of all tasks produced by that worker. Combining this $Q_i = \pi_i Q$ (part (ii) of Lemma A.3), we find

$$\lambda_i^L = \pi_i Q. \tag{A.6}$$

Moreover, Q is related to Y as follows. Starting with equation (1), multiply both sides by $\lambda^{\frac{1}{\eta-1}}$, and bring this term inside the integral on the left-hand side. Substituting for

$\lambda^{\frac{1}{\eta-1}}$ on that left-hand side using equation (6), rearranged as

$$\lambda^{\frac{1}{\eta-1}} = \left(\frac{Q(\tau)}{Y}\right)^{\frac{1}{\eta-1}} \frac{\lambda(\tau)}{\lambda},$$

and simplifying algebra yields

$$\int_{\mathcal{T}} Q(\tau) d\tau = \lambda Y.$$

As the left-hand side is the definition of Q and given $\lambda = 1$, it follows that $Q = Y$. Combining this with equation (A.6) we find that

$$\lambda_i^L = \pi_i Y. \tag{A.7}$$

In a last step, substitute for π_i in equation (A.7) using equation (A.5). Solving the resulting expression for λ_i^L, substituting into equation (A.4) and rearranging for Y yields equation (9). □

A.2 Matching block

A.2.1 Surplus recursions

Combining the definition of $S(x)$ with the expression for $\Omega(x)$ in equation (21) and the surplus sharing rules yields

$$\big(\rho+\delta\big)S(x) = f_1(x) - \rho\big(V_u(x) + V_{f.0}\big) + \lambda_{v.u}(1-\omega)\int \frac{d_u(\tilde{x}')}{u} S(\tilde{x}'|x, \tilde{\xi})^+ dH(\tilde{\xi})\tilde{x}'. \tag{A.8}$$

Similarly, for $S(x|x', \xi)$, and using the expression for $\Omega(x, x', \xi)$ in equation (20) gives

$$S(x|x', \xi)\big(\rho + 2\delta\big) = f_2(x, x', \xi) - \rho\big(V_u(x) + V_u(x') + V_{f.0}\big) + \delta S(x) - (\rho+\delta)S(x'). \tag{A.9}$$

A.2.2 Population dynamics

For any type x, the measure of unemployment satisfies

$$\delta(x)\left(d_{m.1}(x) + \int d_{m.2}(x,\tilde{x}')d\tilde{x}'\right) = d_u(x)\lambda_u\left(\int \frac{d_{f.0}}{v}h(x,\tilde{y}) + \int \frac{d_{m.2}(\tilde{x}')}{v}h(x|\tilde{x}')d\tilde{x}'\right). \tag{A.10}$$

The measure of exogenously separated workers of any type is equal to the measure of unemployed workers of that type finding new employment at either one-worker or two-worker firms.

For all x, the measure of one-worker matches follows

$$d_{m.1}(x)\left(\delta(x) + \lambda_{v.u}\int \frac{d_u(\tilde{x}')}{u}h(\tilde{x}'|x)d\tilde{x}'\right) = d_u(x)\lambda_u\frac{d_{f.0}}{v}h(x) + \delta(x)\int d_{m.2}(x,\tilde{x}')d\tilde{x}'. \tag{A.11}$$

Outflows from this state occur due to exogenous separation or because the one-worker firm meets and decides to hire a coworker of some type. Inflows occur when an unemployed worker of type x meets and gets hired by an unmatched firm or because a two-worker firm that has a type x as one of its employees loses the coworker.

Finally, for all (x,x'),

$$(\delta(x)+\delta(x'))d_{m.2}(x,x') = d_u(x)\lambda_u\frac{d_{m.1}(x')}{v}h(x|x') + d_u(x')\lambda_u\frac{d_{m.1}(x)}{v}h(x'|x). \tag{A.12}$$

The economic intuition parallels the aforementioned reasoning.

A.2.3 Wage function

The value of employment for worker x given a coworker of type x' and match shock ξ is

$$\rho V_{e.2}(x|x',\xi) = w(x|x',\xi) - 2\delta\omega S(x|x') + \delta\omega S(x).$$

Combining with the surplus sharing rule (15) yields

$$w(x|x',\xi) = \rho V_u(x) + (\rho+2\delta)\omega S(x|x',\xi) - \delta\omega S(x) \tag{A.13}$$

To simplify this expression further, substitute for $(\rho+2\delta)S(x|x',\xi)$ from equation

(A.9)

$$w(x|x',\xi) = \rho V_u(x) + \omega\Big[f_2(x,x',\xi) - \rho\big(V_u(x) + V_u(x') + V_{f.0}\big) + \delta S(x) - (\rho+\delta)S(x')\Big] - \delta\omega S(x),$$
$$= \omega f_2(x,x',\xi) + (1-\omega)\rho V_u(x) - \omega\rho\big(V_u(x') + V_{f.0}\big) - \omega(\rho+\delta)S(x'),$$

and then substitute for $S(x')$ from equation (A.8) to obtain

$$w(x|x',\xi) = \omega\big(f_2(x,x',\xi) - f_1(x')\big) \qquad \text{(A.14)}$$
$$+ (1-\omega)\rho V_u(x) - \omega(1-\omega)\lambda_{v.u}\int\int \frac{d_u(\tilde{x}'')}{u} S(\tilde{x}''|x',\tilde{\xi})^+ dH(\tilde{\xi})d\tilde{x}''.$$

B Empirical appendix

B.1 Data sources and construction

B.1.1 SIEED

This section provides further details on the Sample of Integrated Employer-Employee Data (SIEED 7518) and how I process the data. Access is provided by the Research Data Center of the German Federal Employment Agency at the Institute for Employment Research (IAB). Compared to the more well-known LIAB, the SIEED dispenses with survey information, but comprises a larger sample and represents a broader period. This is especially advantageous in the analysis of time-series patterns in Section 4.2. A detailed description can be found in vom Berge *et al.* (2020).

The SIEED is based on administrative employer notifications and covers every worker at a random sample of ("panel") establishments as well as, crucially, the complete employment biographies of each of these workers, even when not employed at the establishments in the sample. To maximize sample coverage, I do not restrict myself to panel establishments, but instead require a minimum number of persons in every establishment-year cell (see below). Variables available describe characteristics of individuals (e.g. age, gender, education) and their job as well as employer (e.g. establishment, average daily wage, occupation, industry). Throughout, I use the KldB-1988 2-digit classification for occupations and the WZ08 2-digit industry classification, relying on the harmonization over time provided by the IAB. The employment biographies come in spell format. I transform the dataset into an annual panel. Where a worker holds multiple jobs in a year, I define the job with the highest daily wage as the main episode. Nominal values are deflated using the Consumer Price Index (2015 = 100).

The main dataset is constructed based on the following selection criteria and steps. (1) I retain observations for individuals between 20-60 with a workplace in former West Germany (excl. Berlin), employed full-time and subject to social security. Among other things, spells associated with marginal part-time employment (recorded only from 1999 onwards) are excluded. (2) I drop observations in industries classified as agriculture and mining, utilities, finance and insurance, households as employers, as well as semi- or fully public industries such as social security and education (codes 31-9, 35-39, 64-68, 84-85, 87-88, 91, 94, 96-99). (3) I require non-missing observations for identifiers, wages, industry, occupation, age, tenure, gender and education. (4) I drop singleton person observations and focus on the largest connected set for each of 5 sample periods (1985-1992, 1993-1997, 1998-2002, 2003-2009, 2010-2017). (5) Using the resulting sample, I estimate residual wages and types, as described in detail below, and retain only observations for which these variables are not missing. (6) The final analysis sample is based on the additional restrictions that the establishment-year cell contains no fewer than ten observations; and that the industry cell contains at least 500 person observations and at least 25 unique employers in any of the years. These restrictions are important for analyses with employer and industry fixed effects as well as the industry-level analysis in Section 3.4.

Turning to wages, the earnings variable is top-coded at the so-called "contribution assessment limit" (or, in a single German word, "Beitragsbemessungsgrenze") of the social security system. To impute right-censored wages, which is done after step (3), I follow standard practices, notably Card *et al.* (2013) and Dauth and Eppelsheimer (2020), by fitting a series of Tobit models to log daily wages, then imputing an uncensored value for each censored observation using the estimated parameters of these models and a random draw from the associated (censored) distribution. I fit 16 Tobit models (4 age groups, 4 education groups), after having restricted the sample per the above. I follow Card *et al.* (2013) in the specification of controls by including not only age, firm size, firm size squared and a dummy for firms with more than ten employees, but also the mean log wage of co-workers and fraction of co-workers with censored wages. Finally, imputed wages are limited to 10 times the 99th percentile.

The empirical analyses use residualized wages as an input. To construct those, I proceed similarly to Card *et al.* (2013) and Hagedorn *et al.* (2017). Specifically, in the pooled sample I regress the (raw) log real daily wage of worker i in year t, $\tilde{w}_{it}$, on a person fixed effect, α_i and a time-varying characteristics index, X_{it} that comprises an unrestricted set of year dummies, quadratic and cubic terms in age interacted with educational attainment — education itself is deliberately not included as a covariate, as it arguably represents time-invariant measure of worker quality — and a quadratic in job tenure.[B.1] The choice of covariates is informed by the observation that the theoretical model does not incorporate life-cycle factors or on-the-job learning nor aggregate productivity growth. As an input into the analysis, I then use $w_{it} = \exp\left(\tilde{w}_{it} - X'_{it}\hat{\beta}\right)$.

The final sample (1985-2017) includes 14,429,61 person-year observations for 1,942,406 unique persons, whose average age is 38 years. The mean unweighted establishment size is 29, the median being 14. The log average real (raw) daily wage in 2010 is 4.65.

B.1.2 Constructing worker types

My baseline measure of worker talent types is based on their position in the economy-wide lifetime earnings distribution, which I recover from the individual fixed effect (FE) in a two-way fixed-effects wage regression à la Abowd *et al.* (1999, "AKM" henceforth). It bears emphasis that in this paper I am not concerned with the estimation of employer FEs or an AKM-based variance decomposition, so prominent debates in the literature about the (structural) interpretation of these terms can be side-stepped. Instead, I simply adopt this approach as a statistical tool to recover the time-invariant component of an individual's earnings ability that conveniently allows controlling for the effects of unmodelled person-level observable characteristics and employer heterogeneity.

Likewise, estimation concerns that relate to limited mobility bias affecting pertain to the estimation of *employer* FEs, and specifically their variance (as well as a the covariance term) are less relevant for person FEs. Nonetheless, to follow best practices, I initially cluster similar firms through a weighted k-means problem, similar to Bonhomme *et al.* (2019),

$$\min_{k(1),\dots,k(J),H_1,\dots,H_K} \sum_{j=1}^{J} n_j \int \left(\hat{F}_j(w) - H_{K_j}(w)\right)^2 d\mu(w), \tag{B.1}$$

[B.1]As the regression includes person and year fixed effects, I exclude the linear age term in light of the age-year-cohort identification problem. As in Card *et al.* (2013), age is normalized around 40.

	α	α lag 1	α lag 2	α lag 3	α lag 4
α	1.0000				
α lag 1	0.8717	1.0000			
α lag 2	0.7103	0.8629	1.0000		
α lag 3	0.5774	0.6820	0.8560	1.0000	
α lag 4	0.4550	0.5270	0.6405	0.8624	1.0000

Table B.1: Auto-correlation structure of individual FEs

where $k(1), ..., k(J)$ constitutes a partition of firms into K known classes.[B.2]

After imputing a cluster to each worker-year observation, I estimate the regression

$$\tilde{w}_{it} = \alpha_i + \sum_{k=1}^{K} \psi_k \mathbf{1}(j(i,t) = k) + X'_{it}\beta + \epsilon_{it} \tag{B.2}$$

where α_i is the individual fixed effect, $\mathbf{1}(J(i,t) = k)$ are dummies indicating which cluster k firm the employer of i in period t, $j(i,t)$ has been assigned to, and X_{it} is the same vector of time-varying controls as in the preceding section B.1.1.[B.3] The estimation is implemented in Stata using the *reghdfe* package (Correia, 2017).

I estimate this regression separately for each of the 5 sample periods, then equate worker i's type $\hat{x}_i$ with their decile rank in the distribution of α_i. Note that I thus allow for the possibility of human capital accumulation across but not within sample periods. Table B.1 reports the auto-correlation structure for period-specific α_i, indicating a strongly positive but less than unit correlation.

[B.2] Here, $\hat{F}_j$ is the empirical cdf of log-wages in firm j; n_j is the average number of workers of firm j over the sample period; and $H_1, ..., H_K$ are generic cdf's. I use a baseline value of $K = 20$ but have experimented with $K = 10$ and $K = 100$ as well; the choice makes little practical difference. I use firms' wage distributions over the entire sample period on a grid of 20 percentiles for clustering.

[B.3] While residualizing wages for observables aligns with the model's exclusion of life-cycle and on-the-job learning effects, it is not self-evident that worker types should likewise be computed from residualized wages. It could be argued, for instance, that for the interpretation of the production function it matters, for example, whether a worker is good," not whether they are good for their age." I include controls to maintain maximum consistency, both internally and with respect to the existing literature. A previous version (Freund, 2023) reported results when worker types are constructed without controlling for observables; differences to the baseline were minimal.

B.1.3 Task complexity measure

This section briefly summarizes how the task-complexity proxy is constructed. A more detailed analysis can be found in the unpublished *Supplemental Appendix*.

I draw on 5 waves of the Employment Surveys (ES) carried out by the German Federal Institute for Vocational Training (Bundesinstitut fuer Berufsbildung, BIBB; Hall and et al. (2018)), for the years 1985/86, 1991/92, 2006, 2012, and 2018. The survey asks individual workers, among other things, a recurring question about the tasks performed at work, e.g. in 2012: "Please think of your occupational activity as [...]. I will now give you a number of specific job tasks. Please tell me how often these job tasks occur in your work, whether they occur often, sometimes or never." While the exact question and the list of tasks varies across some of the waves, I follow the BIBB's guidance (Rohrbach-Schmidt and Tiemann, 2013) in standardizing the sample basis, selecting tasks that were repeatedly queried over time; and focus on the share of "complex/abstract" (cognitive non-routine) tasks compared to other tasks, without having to distinguish between routine-cognitive or routine-manual tasks, which can be difficult.

Task classification	**Task name**	**Description**
Complex	investigating	Gathering information, investigating, documenting
	organizing	Organizing, making plans, working out operations, decision making
	researching	Researching, evaluating, developing, constructing
	programming	Working with computers, programming
	teaching	Teaching, training, educating
	consulting	Consulting, advising
	promoting	Promoting, marketing, public relations
Other	repairing, buying, accommodating, caring, cleaning, protecting, measuring, operating, manufacturing, storing, writing, calculating	

Table B.2: Classification of tasks in the BIBB Employment Surveys

Notes. This table summarizes the classification of tasks into two groups: "Complex" and "Other."

Table B.2 summarizes how I classify tasks, which is guided by Spitz-Oener (2006) and Rohrbach-Schmidt and Tiemann (2013). Next, we can define an index capturing the importance of complex tasks for worker i in period p, following Antonczyk *et al.* (2009):

$$T_{ip}^{\text{complex}} = \frac{\text{number of activities performed by } i \text{ in task category “complex” in sample year } p}{\text{total number of activities performed by } i \text{ in sample period } p}$$

For example, if worker i performs five distinct activities and two of those belong to the category of complex tasks, then $T_{ip}^{\text{complex}} = 0.4$.

When considering trends over time, I report T_{ip}^{complex} at the aggregate level, or disaggregated by three education groups. To merge the task-complexity measure into the SIEED I use occupational averages, computed using the German Classification of Occupations 1988 (KldB88). To link the KldB88 with waves using the KldB92, I rely on a crosswalk, which is of high quality as the two classifications are very similar.[B.4]

B.2 Additional empirical results

B.2.1 Alternative worker type measures

Years of schooling. One concern with the auxiliary regression estimated in Section 3.2.3 is that the dependent variable in regression (25) is the period-t wage and the independent variables of interest, own and coworker type, are likewise a function of wages (in all years t belonging to the sample period in question). While this approach is theory-consistent, with identification coming from variation in wages over time while the regressors are invariant across years within sample periods, we may still be worried about a confounding effect. This section shows that the results are robust to using education as a proxy for worker talent types instead, similar to Nix (2020),

Table B.3 shows the coworker sorting correlation coefficient and the wage regression interaction coefficient β_c when using years of schooling a a worker type measure. While the magnitudes for the latter are not directly comparable to those obtained under the FE-based type measure, the time trends for both moments align with those reported in the main text.

Period-invariant worker types. I also considered an alternative where individual FEs and hence types $\hat{x}_i$ are constructed that are invariant across sample periods. The implied time trends in sorting and the reduced-form interaction coefficient $\hat{\beta}_c$ are very similar.

[B.4] I thank Anett Friedrich for sharing the crosswalk.

Period	Coworker sorting	Interaction coefficient ($\hat{\beta}_c$)
1985-1992	0.45	0.007***
1993-1997	0.48	0.005***
1998-2003	0.52	0.008***
2004-2009	0.55	0.009***
2010-2017	0.56	0.010***

Table B.3: Sorting and interaction coefficient: years of schooling

Notes. The entry in the first main column shows the correlation between own and average coworker years of schooling. The entry in the second main column is the point estimate for the interaction coefficient β_c in a version of regression (25) adapted to the use of years-of-schooling as a type measure. Years of schooling are imputed from completed education following Card *et al.* (2013). * p<0.1; ** p<0.05; *** p<0.01.

	1985-1992	1993-1997	1998-2003	2004-2009	2010-2017
Type	0.0792***	0.0776***	0.0822***	0.0860***	0.0829***
	(0.0003)	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Coworker Type	0.0019***	0.0018**	0.0025***	0.0048***	0.0052***
	(0.0005)	(0.0007)	(0.0006)	(0.0009)	(0.0006)
Observations (1000s)	3,233	2,174	2,263	3,153	3,606
Adj. R^2	0.758	0.790	0.804	0.847	0.834

Table B.4: Auxiliary regression results by sample period - log specification

Notes. All regressions include industry-year, occupation-year, and establishment FEs. Employer-clustered standard errors are given in parentheses. Observations are unweighted and rounded to the nearest thousand. Significance levels: * p<0.1; ** p<0.05; *** p<0.01.

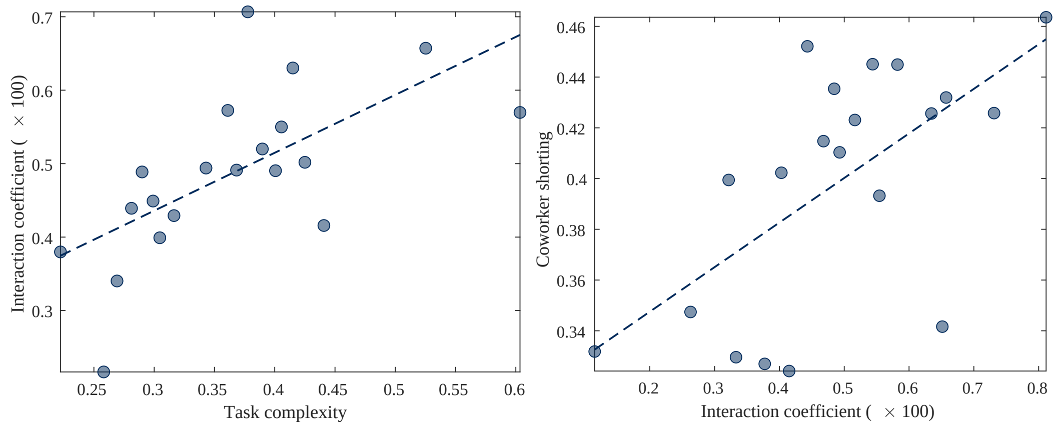
B.2.2 Log specification

The main text reports an auxiliary regression with the (normalized) wage level as the dependent variable, consistent with the theoretical model. This section reports results for a complementary regression that uses the log wage on the left-hand side:

$$\ln w_{it} = \beta_0 + \beta_1 \hat{x}_i + \beta_2 \hat{x}_{-it} + \psi_{j(i,t)} + \nu_{0(i,t)t} + \xi_{S(i,t)t} + \epsilon_{it}, \tag{B.3}$$

where types are treated as continuous variables.

Table B.4 reports the coefficient estimates for each of the 5 sample periods. The coefficient on the coworker type has increased over time.



(a) Task complexity vs. β_c **(b)** β_c vs. ρ_{xx}

Figure B.1: Industry-level evidence: within-period moments

Notes. Binscatter plots controlling for 5 sample-period fixed effects.

B.2.3 Industry-level evidence

To complement the binscatter plots in Figures 6b 6c, which are constructed controlling for industry FEs, Figure B.1 reports the relationship between the task-complexity proxy for χ and β_c and that between β_c and coworker sorting ρ_{xx} when using cross-industry within-period variation instead, i.e., controlling for period FEs. This yields the same positive relationships reported in the main text.

B.2.4 Moments by establishment size groups

A potential concern with the estimation of β_c in the main text is that the variation in coworker quality exploited for identification in regression (25) is not exogenous with respect to the error term ϵ_{it}. In the structural model, this variation naturally occurs due to search frictions and consequent random variation in what type of employee a given employer has the opportunity to make an offer to. But it could be argued that for larger firms idiosyncratic variation averages out when constructing a representative coworker type (cf. Hoxby, 2000).

To evaluate these concerns, I compute the coworker correlation measure of sorting, ρ_{xx}, and interaction coefficient, β_c, separately for 4 different establishment size groups and, to study time trends, for 5 different sample periods. I perform this exercise both for the full sample and for the subsample of panel establishments for which the SIEED contains information on the entire workforce. Figure B.2 visualizes the results.

The results from this exercise are reassuring. Across establishment size groups, there is no evident sign of a bias, in that the magnitude of the estimates for β_c and sorting are fairly similar across size bins,ts. The one potential exception is the estimated value of β_c for the 2010s, which is notably lower in the two small-size groups. The same pattern is not observed, however, when focusing on the panel establishments, for which the representative coworker types can be constructed in a more reliable manner.

B.2.5 Non-linear coworker aggregation

As discussed in the main text, to construct a representative coworker type I compute, for each person-year observation, the unweighted arithmetic mean of all coworkers' types, i.e. $\hat{x}_{-it} = \left(\frac{1}{|S_{-it}|}\sum_{k\in S_{-it}}\hat{x}_k\right)$. This approach ignores a non-linearity in the aggregation *across coworkers* that is implied by the reduced-form production function (9), i.e., we ought to aggregate using a power mean that assigns disproportionate weight to low-type coworkers insofar as χ and hence talent complementarities are large.

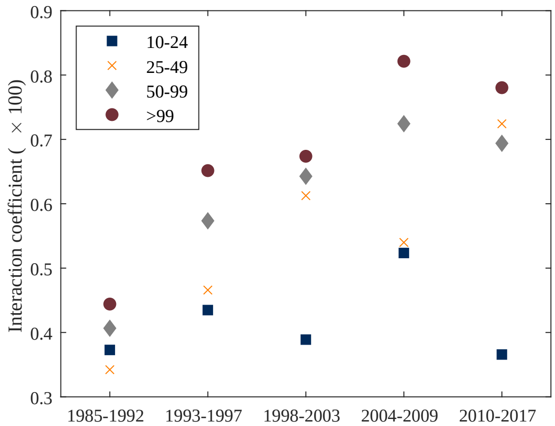
However, the bias from ignoring this non-linearity is small for the following reason. Define the correctly aggregated average coworker by $\tilde{x}_{-it} = \left(\frac{1}{|S_{-it}|}(\hat{x}_k)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$, where γ is the reduced-form elasticity of complementarity defined in Corollary 1. Then taking a second-order Taylor approximation around $\hat{x}_{-it}$ shows that $\tilde{x}_{-it} - \hat{x}_{-it} \approx -\frac{1}{2}\gamma\frac{\sigma_x^2}{\hat{x}_{-it}}$, where $\hat{\sigma}_{x'}^2$ is the variance of coworker types. This shows that the unweighted average is upward biased in proportion to the product of γ and the dispersion among coworkers (within a match). The reason the bias is small is that precisely when γ is high, positive assortative matching based on talent means that dispersion in coworker talent is low.

C Quantitative appendix

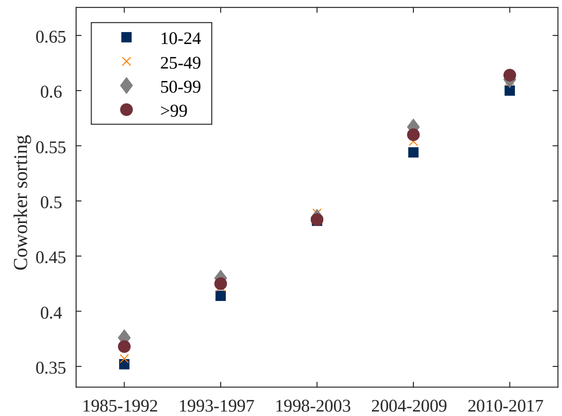
C.1 Methodology

C.1.1 Labor market transition rates from the LIAB

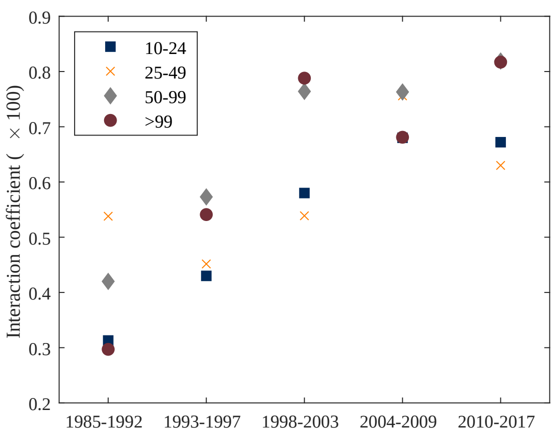
This section describes how the empirical labor market transition rates that discipline the job arrival and destruction rates in the quantitative model are computed. As a data source, I supplement the SIEED with the Linked Employer Employee Data longitudinal model (LIAB LM7519), which contains information also on non-employment spells, unlike the SIEED. I proceed in four main steps. First, I convert the spell-level data into a monthly



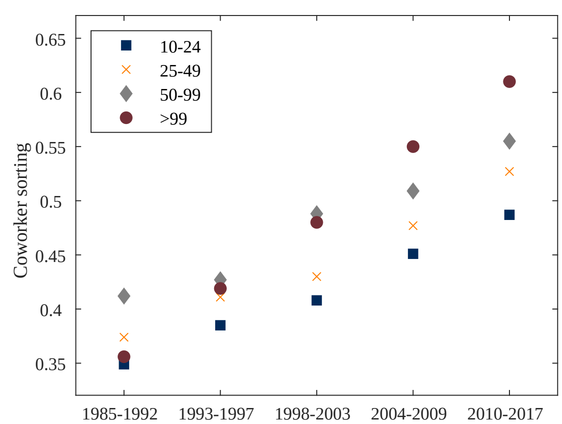
(a) Interaction coefficient – all establishments



(b) Sorting – all establishments



(c) Interaction coefficient - panel est.



(d) Sorting - panel est.

Figure B.2: Reduced-form complementarity and sorting estimates by establishment size

Notes. This figure indicates empirical estimates of β_c (panels (a) and (c)) and the coworker correlation (panels (b) and (d)) and when estimated separately for 5 sample periods and several establishment size groups. "Coworker complementarity" corresponds to the point estimate of the coefficient on the interaction term, $\hat{\beta}_c$, in regression (25). The underlying worker types are estimated from a pooled sample across all establishment size groups, as in the remainder of the paper. In panels (a) and (b), the estimation sample comprises all establishments, whereas panels (c) and (d) are based on the subsample of 'panel establishments'.

panel. Second, I restrict the sample to approximate the selection criteria used in the other empirical analysis but without being limited to employed persons, i.e., I select individuals aged 20-60 who only ever worked for establishments in West Germany. Third, in the construction of transition rates I largely follow Jarosch (2023). Employment refers to full-time employment subject to Social Security. The job finding rate is computed as the rate at which currently non-employed workers who are receiving unemployment insurance (UI) transition into employment. For the job destruction rate, I compute the frequency with which a worker is employed in one month but not in the month thereafter. Note that here I do not condition on receiving UI after separation, as the model does not distinguish between unemployment and non-employment. I instead define the job finding rate based on unemployment to employment transitions, since the model does assume search effort conditional on non-employment. Finally, for job-to-job transitions I compute the rate at which currently employed workers are employed at another establishment the following month. In step four, I compute averages of these different transition rates across months and for different sample periods.

C.1.2 Validation of identification approach

To support the argument laid out in the main text that each element of the parameter vector ψ is closely linked to a particular moment, Figure C.1 plots the relevant moment against the respective parameter. As required for local identification, the relationships are monotonic and exhibit significant amount of variation.

C.1.3 Between-share adjustment procedure

This section describes an adjustment method I propose and adopt to correct the model-implied between-within firm variance decomposition for a statistical bias.

The challenge is as follows. Since in the structural model the team size is two, the level of the between-share will be upward biased for any degree of coworker sorting less than unity, i.e. $\rho_{xx} < 1$. Underlying this is a mechanical statistical effect, i.e. the law of large numbers does not apply within production units. For example, even when $\rho_{xx} = 0$ there will be between-firm average wage differences solely due to chance, which would not be the case with large firms. Note that this bias is larger when the coworker correlation is lower: for $\rho_{xx} = 1$, *all* dispersion is *across* and *none* within units, regardless of team size. Figure C.2 illustrates these ideas graphically; its construction is described below.

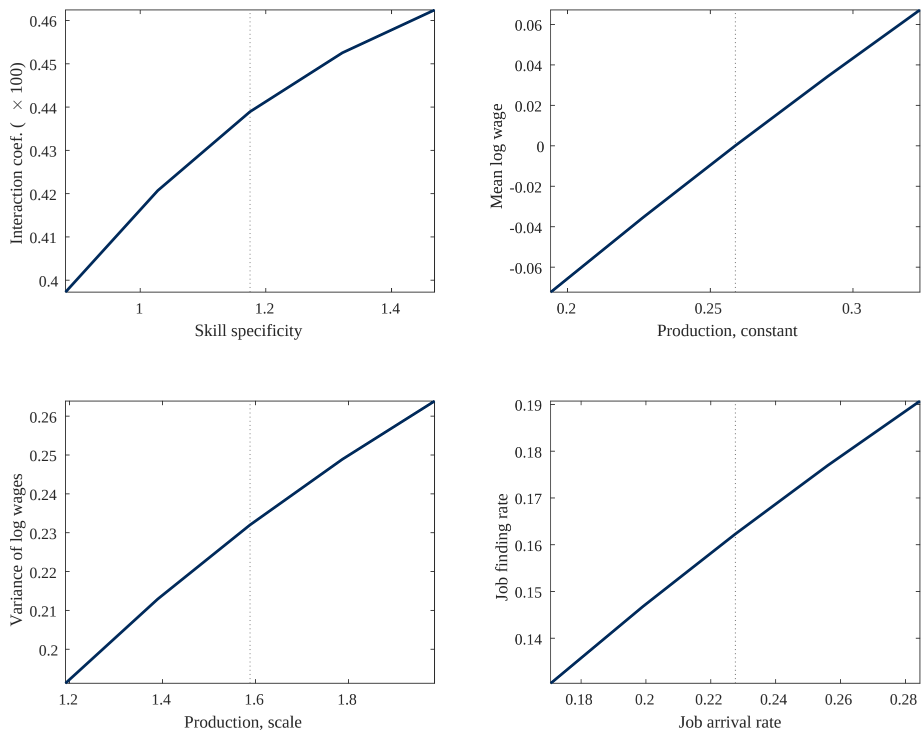
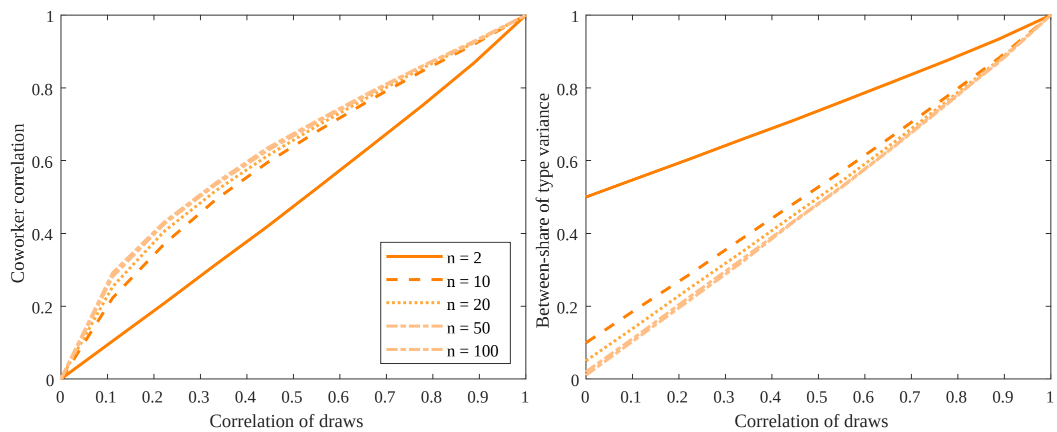


Figure C.1: Validation of identification method: moment against parameter

Notes. This figure plots the targeted moment against the relevant parameter, holding constant all other parameters.



(a) Coworker sorting

(b) Between-share of type var.

Figure C.2: Graphical illustration of the adjustment method

Notes. This figure illustrates how the coworker correlation coefficient (left panel) and the between-firm share of the variance of types (right panel) vary with the correlation parameter in the Gaussian copula. The different lines represent different "team sizes," that is, varying lengths n of the vector X. Workers are binned into ten deciles. The results are based on one million draws.

The adjustment method I propose is based on a *statistical* model that can flexibly accommodate different degrees of coworker sorting as well as team sizes. Consider a random vector $X = (X_1, X_2, \ldots, X_n)'$ whose distribution is described by a Gaussian copula over the unit hypercube $[0,1]^n$, with an $n \times n$ correlation matrix $\Sigma(\rho^c)$, which contains ones on the diagonal, while the off-diagonal elements are all equal to a parameter ρ_c. Formally, the Gaussian copula with parameter matrix $\Sigma(\rho_c)$ is $C_{\Sigma}^{\text{Gauss}}(x) = \Phi_R(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_n))$, where Φ^{-1} is the inverse cdf of a standard normal and Φ_R is the joint cdf of a multivariate normal distribution with mean vector zero and covariance matrix equal to $\Sigma(\rho_c)$. In our context, n may be interpreted as the average team size. Each vector of observations drawn from the distributions of X, $x_j = (x_{1j}, x_{2j}, \ldots, x_{nj})'$, describes the types of workers in that team, indexed by j.

In this setup, we can derive an analytical formula for the population between-team share of the variance of types as a function of n and ρ_c. As the marginals of the Gaussian copula are simply continuous uniforms defined over $[0,1]$, the variance of the union of all draws is just $\frac{1}{12}$. Furthermore, the mean of the elements of X is itself a random variable, $\bar{X}$, so for some realization x_j, we can define $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$. Since the elements of X all have the same variance and we specified their correlation profile, the variance of $\bar{X}$ will be $\frac{1}{n^2}\left(\frac{n}{12} + n(n-1)(\frac{\rho_c}{12})\right)$. Taking the ratio, we find that the between share, as a function of ρ_c and n, is equal to $\sigma^2_{x,\text{between-share}}(\rho_c, m) = \frac{1}{m}\left(1 + (m-1)\rho_c\right)$. Denoting by the empirical average size be $\hat{n}$, the adjusted results for the between-share are therefore obtained by subtracting the following correction factor: correction-factor $= \frac{1}{2}\left(1+\rho_c\right) - \frac{1}{\hat{n}}\left(1+(\hat{n}-1)\rho_c\right)$. I set $\hat{n} = 15$ and for ρ_c I use the average coworker correlation value in the period-1 sample.[C.1]

Finally, I briefly discuss two potential concerns with this approach. One is that rho_c is not the same measure as the coworker correlation, ρ_{xx}. To compare the two measures, suppose we draw M samples (i.e., distinct teams) from X, so that the total number of observations is $M \times n$. Each individual observation is indexed by $i = 1, \ldots, M \times n$, and the sample to which i belongs is $j(i)$. Then we can define the leave-out-mean $\bar{x}_{-i,j} = \frac{1}{n-1}\sum_{k\neq i} x_{k,j(i)}$. As in the main analysis, the coworker correlation is $\rho_{xx} = \text{corr}(x_i, \bar{x}_{-i})$, where the j indexed is suppressed to emphasize that we are considering a worker-weighted statistic. Figure C.2 confirms that ρ_{xx} and ρ_{cc} track each other quite closely, even though for larger values of n and intermediate values of ρ_c we find that $\rho_{xx} > \rho_{cc}$. The second concern is that the adjustment approach concerns the between-unit share of the variance of *types*, as opposed to that of *wages*. However, given a distribution of workers across production units derived from the statistical model, we can impute wages based on the wage function derived from the structural model, and then repeat the variance decomposition for wages. Simulations confirm that the two adjustment factors obtained when looking at types and wages, respectively, are very similar to one another. Overall, the proposed adjustment approach therefore accomplishes the desired goal.

C.2 Additional figures and tables

[C.1] Basing the correction on the earlier sample yields a bigger downward adjustment, which avoids overstating the degree of between-firm inequality that the model can generate without assuming ex-ante firm heterogeneity. Note also that the exact value of $\hat{n}$ does not matter much, since for reasonable values of $\hat{n}$ the implied correction factors are very close to each other. The magnitude of the bias rapidly diminishes as $\hat{n}$ grows, as is evident from the above formula.)

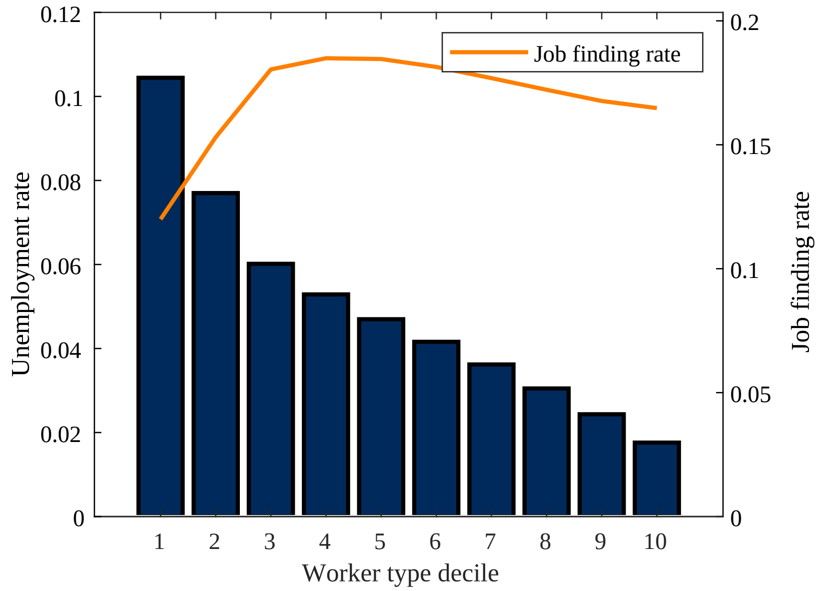


Figure C.3: Labor market finding and unemployment rates by worker talent type

Notes. This figure shows unemployment and job finding rates by talent type in the estimated model.

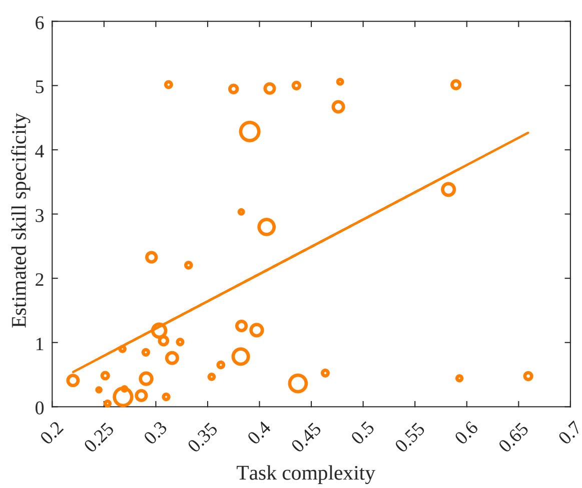


Figure C.4: Task complexity and estimated skill specificity χ χ

Notes. This figure shows the estimated value of χ for each industry, plotted against task complexity. In a first step, I repeat the online estimation of the parameter vector ψ, targeting average within-industry moments for 2010-2017. In a second step, I estimate industry-specific values of χ, denoted χ_s, by keeping all other parameters fixed and letting only the targeted moment β_c vary.

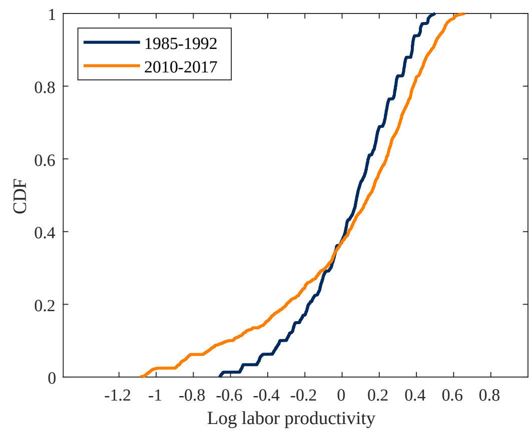


Figure C.5: Model predicts increased firm-level labor productivity dispersion

Notes. This figure plots the cumulative distribution of labor productivity implied by the calibrated model(s).

Parameter	Description	'85-'92	'93-'97	'98-'03	'04-'09	'10-'17
δ_0	Sep. rate, constant	0.014	0.016	0.017	0.019	0.015
δ_1	Sep. rate, scale	-0.981	-1.009	-1.014	-0.941	-0.841
χ	Skill specificity	0.264	0.387	0.671	0.744	1.175
a_0	Production, constant	0.456	0.451	0.365	0.308	0.259
a_1	Production, scale	1.242	1.251	1.409	1.557	1.589
$\bar{b}$	Unemp. flow utility, scale	0.674	0.671	0.676	0.617	0.637
λ_u	Meeting rate	0.155	0.112	0.119	0.145	0.228

Table C.1: Estimated parameters – baseline

Notes. This table summarizes the parameters estimated under the baseline specification for each of the 5 sample periods.

Parameter	Description	'85-'92	'93-'97	'98-'03	'04-'09	'10-'17
δ_0	Sep. rate, constant	0.014	0.015	0.016	0.017	0.013
δ_1	Sep. rate, scale	-0.946	-0.956	-0.936	-0.834	-0.682
χ	Skill specificity	0.140	0.240	0.335	0.461	0.561
a_0	Production, constant	0.447	0.445	0.386	0.328	0.274
a_1	Production, scale	1.255	1.266	1.416	1.570	1.634
$\bar{b}$	Unemp. flow utility, scale	0.678	0.673	0.671	0.612	0.621
λ_u	Meeting rate	0.152	0.110	0.109	0.134	0.206

Table C.2: Estimated parameters – within-occupation ranking of worker types

Notes. This table summarizes the parameters estimated for each of the 5 sample periods when workers are ranked within two-digit occupations.

Parameter	Description	'85-'92	'93-'97	'98-'03	'04-'09	'10-'17
δ_0	Sep. rate, constant	0.014	0.016	0.017	0.019	0.015
δ_1	Sep. rate, scale	-0.981	-1.009	-1.014	-0.941	-0.841
χ	Skill specificity	0.204	0.303	0.434	0.468	0.830
a_0	Production, constant	0.484	0.488	0.441	0.419	0.371
a_1	Production, scale	1.172	1.165	1.270	1.340	1.370
$\bar{b}$	Unemp. flow utility, scale	0.675	0.671	0.670	0.610	0.625
λ_u	Meeting rate	0.152	0.110	0.111	0.133	0.211

Table C.3: Estimated parameters – within-industry analysis

Notes. This table summarizes the parameters estimated for each of the 5 sample periods when the targeted variance of wages and β_c are constructed as the averages across two-digit industries.