Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities*

Lukas B. Freund§

University of Cambridge

December 3, 2023

Please click here for the latest version.

Abstract

Recent studies attribute the rise in wage inequality primarily to widening pay disparities between rather than within firms. I develop a novel theory to quantitatively explain this fact. The theory has three core features: production takes place in teams; workers are heterogeneous in talent and are specialized in specific tasks; and labor markets are frictional. Specialization endogenously generates coworker complementarity: talented workers gain more from more talented colleagues. This creates an incentive for assortative matching, fostering dispersion in average wages across firms. Search frictions prevent perfect sorting in equilibrium, though. Using administrative panel data for Germany, I measure complementarities, validate key mechanisms, and structurally estimate the model. I argue that specialization has intensified since the mid-1980s, and show that coworker complementarities and talent sorting have strengthened concurrently, aligned with the theory's predictions. According to model exercises, this explains ∼40% of the observed increase in the between-firm share of wage inequality, and contributed to elevated firm-level productivity dispersion. Rising complementarities also worsened aggregate output costs from coworker mismatch, but endogenously increased sorting partly mitigated this effect.

Keywords: coworker complementarity, firms, inequality, matching, specialization, teams

*I am immensely grateful to Vasco Carvalho for his guidance, support and feedback. I thank Wouter Den Haan, Luis Garicano, Ezra Oberfield, Pontus Rendahl, Richard Rogerson and Gianluca Violante for many valuable discussions. For helpful comments I furthermore thank Elena Aguilar, Zsófia Bárány (discussant), Martin Beraja, Florin Bilbiie, Leah Boustan, Giancarlo Corsetti, Tyler Cowen, Maarten De Ridder, Mayara Felix, Chryssi Giannitsarou, John Grigsby, Joe Hazell, Kyle Herkenhoff, Tomer Ifergane, Bill Janeway, Ricardo Lagos, Hanbaek Lee, Lukas Mann, Kristina Manysheva, Ben Moll, Simon Mongey, Charles Parry, Adrian Ochs, Jane Olmstead-Rumsey, Jeremy Pearce, Tommaso Porzio, Steve Redding, Ricardo Reis, Esteban Rossi-Hansberg, Isaac Sorkin and Christian Wolf; as well as audiences at Cambridge, OECD, Notre Dame, Penn "Labor, Firms, and Macro" reading group, Princeton PhD macro workshop, 3rd CRC TR 224 Workshop on Labor Markets, and the 2023 Annual Meeting of the CEPR Macroeconomics and Growth Programme. I am indebted to Chiara Criscuolo and Peter Gal for facilitating the analysis of Portuguese micro data. 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.

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

1 Introduction

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. A key reason is that gaps across firms in the quality of their workforces have widened. While these facts figure prominently in academic and policy discussions, their implications for our understanding of the rise in inequality and potential policy responses depend on the underlying structural drivers. Those remain obscure.[1]

In this paper, I propose a novel framework that rationalizes these empirical patterns of firm-level inequality, linking them to the nature of production. The framework centers on a conception of the firm as a "team assembly technology." Firms hire multiple workers of heterogeneous talent and production features coworker complementarity: Workers' realized productivity is interdependent such that the marginal productivity of one employee's talent increases with that of colleagues. Labor market sorting is interpreted as describing how workers of varying talent are matched together in the workplace. I argue that shifts in the nature of work have reinforced coworker complementarity, making it increasingly advantageous for talented workers to collaborate among themselves, and thereby leading to heightened labor market sorting and between-firm wage dispersion.

The analysis proceeds in three main steps. First, I build and characterize an equilibrium model of production with teams of heterogeneous workers that are hired in a frictional labor market. This model provides an analytical, task-based microfoundation for coworker complementarities and describes their influence on the distribution of talent and wages within and between firms. Second, I take the model to the data. I develop a method to estimate the strength of coworker complementarity, and implement it using German matched employer-employee data. This novel micro evidence facilitates a structural estimation of the model. Third, I combine theory and data to quantitatively explain the evolution of firm-level inequality. Empirically, complementarity has doubled in magnitude since the mid-1980s, mirroring a shift in the nature of production away from routine tasks toward more complex requirements. Through the lens of the model, this rise in complementarity accounts for 40% of the empirically observed growth in the between-firm share of wage inequality. Overall, the paper thus contributes theory, tools, and evidence that jointly facilitate a parsimonious, quantitative interpretation of changes in the nature of work and in the distribution of talent and wages within and across firms.

Elaborating on the first step, the theory has three core features. First, workers are heteroge-

[1] Key references on firm-level inequality include Card *et al.* (2013) for Germany and Song *et al.* (2019) for the U.S., while Criscuolo *et al.* (2021) provide a cross-country analysis and discuss policy implications.

neous in overall talent, modeled in terms of absolute advantage, and their skills are specialized. I take "specialization" to describe the degree to which an individual's productivity varies across different tasks. Specialization contrasts with a scenario where a talented employee excels – and is thus more productive than a less talented peer – equally across all production tasks. Second, in the spirit of Garicano (2000), production takes place in teams. Firms are ex-ante identical and coordinate the division of labor: each hires multiple workers, then assigns them tasks to maximize output. Third, labor services are traded in markets with frictions, specifically a random-search setting à la Herkenhoff *et al.* (2022), with wages determined through bargaining. Team assembly consequently involves a trade-off between the productivity gain from a well-matched team and search costs. These elements jointly give rise to a model that can speak to dispersion in worker talent and pay within and across firms.

In equilibrium, the workforce of some firms is composed of highly talented workers, while other firms mostly employ less productive ones. Specialization endogenously generates coworker complementarity in talent. As a worker's efficiency varies across tasks, their realized productivity is lowered if, due to less capable colleagues, they have to dedicate more time to tasks in which they are relatively inefficient. The loss in productivity is greater in absolute magnitude for workers with higher potential output. Thus the production function is supermodular in talent (i.e., it features complementarity). If labor market frictions are absent, *any* degree of complementarity leads to pure positive assortative matching in equilibrium (Kremer, 1993). As the best candidates for hire are more productive in firms employing other talented employees, those firms can always outbid competitors for their services. In the presence of search costs, however, a range of matches is accepted. The strength of coworker complementarity influences how wide this range is, thus shaping equilibrium matching outcomes. If specialization is strong, the resulting complementarities make it relatively more profitable for agents to expend time on search, and talent becomes more concentrated in select workplaces.

In this environment, exogenous technological shifts that reinforce specialization foster positive assortative matching and raise between-firm inequality at the macro level. I argue that this theoretical mechanism is empirically relevant. In some domains, specialization is widely understood to have intensified. In science, for instance, the increasing cost of reaching the frontier – the "burden of knowledge" (Jones, 2009) – necessitates increasingly narrow individual expertise, with team production gaining in prevalence concomitantly. Considering the economy more broadly, employees are performing fewer routine tasks, where everyone is similarly productive, instead confronting more complex, cognitive non-routine problems that require specific skills.[2] I interpret these empirical trends as indicating that deeper specialization is one of several

[2]The literature section below discusses studies of the changes in task requirements. Jones (2009), Bloom *et al.* (2014), and Neffke (2019) highlight shifts in knowledge and communication costs.

dimensions along which the nature of production has changed, alongside, say, skill bias. The model reveals that intensified specialization strengthens coworker complementarities, thereby widening gaps in workforce quality and pay between firms.

I analytically characterize the two key relationships underpinning this mechanism: deeper specialization leads to stronger coworker complementarity; and complementarity fosters between-firm inequality. The firm's organizational problem is to optimally assign tasks when each employee has limited time and varies in productivity across tasks. The probabilistic, extreme-value formulation of productivity heterogeneity popular in trade and spatial economics can be leveraged to keep this high-dimensional problem tractable. Specifically, I derive a team production function which is aggregative, in that the vector of employees' talent types is a sufficient statistic for team output, and which has a constant-elasticity-of-substitution (CES) structure. The degree of substitutability endogenously varies with a parameter that determines the strength of specialization, i.e., how dispersed any worker's efficiencies are across tasks. The stronger specialization is, the more advantageous is the division of labor compared to each employee performing all tasks. Crucially, coworker complementarity is also stronger, so output is more sensitive to the talent of the least capable team member(s). Furthermore, in a stylized version of the matching model, I derive in closed-form that stronger complementarity leads to more positive assortative matching into teams and, hence, increased between-firm wage inequality in equilibrium.

Next, I bring the model to the data. To this end, I develop an empirical strategy to estimate coworker complementarities. Since production complementarities cannot be measured directly in general data environments, I use the theoretical model to derive an identification result which reveals an empirically observable moment that is highly informative about production complementarity: the cross-partial derivative of the empirical, non-parametric wage function with respect to own type and coworker type. I implement this measurement strategy using administrative matched employer-employee data for Germany, interpreting each establishment as one team.[3] Coworker complementarity is robustly found to be positive. In the cross-section, complementarities are stronger in occupations involving more complex tasks, which in turn is associated with more pronounced coworker sorting – consistent with the theoretical predictions.[4]

Empirical estimates of the strength of coworker complementarity allow calibrating the model without relying on between- or within-firm dispersion in talent or wages as targets. I fit the model to a parsimonious set of moments characterizing the German economy for 2010-2017. The micro estimates of the wage cross-partial derivative serve a key source of empirical discipline. The model reproduces key untargeted moments, notably coworker sorting patterns and the empirical decomposition of wage inequality into a between- and a within-employer component.

[3] I use "firm" and "establishment" interchangeably; in the data analysis the latter is the observed production unit.

[4] For the cross-sectional analyses and robustness checks, I also consider comprehensive micro data for Portugal.

Finally, I use micro data and model to provide a quantitative assessment of historical trends. Longitudinal surveys indicate that the complexity of tasks performed at work has risen since the mid-1980s. Over the same time, coworker wage complementarity has more than doubled and coworker talent sorting has intensified, consistent with the predictions of the theory. To evaluate the macro-distributional implications of this rise, I calibrate the model for 1985-1992. Comparing between 1985-1992 and 2010-2017, the model predicts a 16 percentage point increase in the between-establishment share of the variance of log wages. Counterfactual exercises imply that the rise in coworker complementarity estimated in the micro data can account for 40% of the empirically observed change in the between-share.[5] Beyond wages, complementarities also contributed to a rise in firm-level productivity dispersion.

Lastly, I discuss the model's implications for aggregate labor productivity. The microfounded production function reveals that while greater specialization yields direct productivity gains, this comes with a catch: as complementarities are amplified, productivity becomes more sensitive to how positively assortative coworkers are matched. The potential output costs of talent misallocation due to search frictions are consequently greater. For the period 1985-1992, equilibrium per-capita output is estimated to be 1.8% below potential (i.e., output under perfectly assortative matching). The gap rises to 2.2% in the 2010s. The increase in mismatch costs is limited, despite the doubling of complementarities, precisely because coworker sorting has increased. Without the rise in sorting, the gap between efficient and realized labor productivity would have been more than twice as large in 2010-2017.

LITERATURE. This paper relates to five strands of literature. First, I connect two lines of research on wage inequality. One line describes how trends in the nature of work – notably, a shift from routine to non-routine tasks – affect skill prices.[6] The other, more recent line of research employs descriptive and reduced-form approaches to document firms' significant and growing role in explaining wage inequality. Notable examples such as Card *et al.* (2013) and Song *et al.* (2019) find that changes in workforce composition play a central role.[7] However, models that explain these empirical patterns are lacking. Such a model needs to incorporate both within- and between-firm worker heterogeneity and be capable of explaining changes over time. Yet, most influential theories of trends in inequality center on the aggregate production function and wage distribution (e.g., Krusell *et al.*, 2000). Meanwhile, canonical models of labor market sorting lack a well-defined notion of within-firm worker heterogeneity (e.g., Shimer and Smith, 2000).

[5] Beyond complementarities, changes can be attributed to mechanical effects of skill-biased technological change and to declining search frictions. I also conduct robustness check that consider the influence of occupational composition effects (e.g., due to outsourcing) and on-the-job search.

[6] See, e.g., Autor *et al.* (2003), Spitz-Oener (2006), Deming (2017), Atalay *et al.* (2020), Autor *et al.* (2022).

[7] See, among others, Kremer and Maskin (1996), Barth *et al.* (2016), Helpman *et al.* (2017), Haltiwanger and Spletzer (2020), Criscuolo *et al.* (2021), Håkanson *et al.* (2021), Sorkin and Wallskog (2021), Wilmers and Aeppli (2021), Kleinman (2022) and Engbom *et al.* (2023).

This paper presents a model that integrates a team production function with heterogeneous workers and labor market search. By accounting for endogenous coworker complementarities, the models reveals that the transformation of work helps to rationalize the growing importance of between-firm inequality, thereby bridging both literatures.

Second, I add to the literature on firm organization, notably studies that examine how the within-firm division of labor shapes productivity and inequality.[8] My model is especially linked to and builds on the theory of knowledge-based hierarchies of Garicano (2000) and Garicano and Rossi-Hansberg (2006), as the organization of production and the distribution of wages is jointly determined by the equilibrium assignment of heterogeneous individuals to firms and to tasks within firms. While hierarchical models have a "vertical" focus – complementarities between managers and their endogenous number of supervisees – mine considers complementarities more generally and lends itself to quantification, at the cost of treating the span of control as exogenous. I also build on Kremer (1993) who argues that what I call coworker complementarity should be pronounced in more "complex" environments, though an empirical or quantitative analysis is missing.[9] All of these papers assume frictionless labor markets. I contribute to this literature by providing a micro-foundation for coworker complementarities, showing that their macro implications crucially depend on the interaction with labor market frictions, taking the theory to micro data, and performing a quantitative assessment.

Third, the paper relates to theoretical models of task assignment, including Costinot and Vogel (2010), Acemoglu and Restrepo (2018), Martinez (2021) and Ocampo (2022). I contribute to this literature by showing how Eaton and Kortum's (2002) probabilistic formulation of technological heterogeneity in terms of extreme-value distributions facilitates aggregation across a continuum of tasks in the context of a discrete number of producers characterized by multidimensional skill heterogeneity.[10] The resulting CES production function makes it possible to connect what are typically qualitative task assignment theories with quantitatively oriented macro models that rely on a tractable production function.

Fourth, I relate to the literature that develops structural models of worker and firm dynamics in labor markets, most notably those that examine matching between heterogeneous agents under random search.[11] Models in the latter strand of research typically share the Beckerian

[8]See, for instance, Lucas (1978), Sattinger (1993), and Gabaix and Landier (2008); and, for more recent contributions, 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).

[9]In Kremer's (1993) model, the strength of complementarities is increasing in the number of tasks comprised in the production process ("complexity"), which by construction equals team size. The model developed here severs a mechanical link between team size and complementarities by endogenizing the number of tasks assigned to each team member, and disentangles the effects of complementarities and increasing returns to firm-level labor quality.

[10]Deming's (2017) study of social skills provides inspiration for an analogy to trade, but the setup does not yield an aggregative team production function, nor does it focus on coworker complementarities or equilibrium matching.

[11]Regarding the matching literature specifically, see for instance, Hagedorn *et al.* (2017), Lopes de Melo (2018),

premise that labor market sorting is shaped by complementarities across productivity types. I examine complementarities across coworkers' attributes in multi-worker firms, instead of the typical setup of one-worker-one-firm matches. In that respect, the matching block of my model largely follows Herkenhoff *et al.* (2022), which is the first model where frictional sorting is between workers within teams rather than between workers and firms.[12] While Herkenhoff *et al.* (2022) primarily focus on coworker learning, they also conjecture that stronger complementarities can explain increased labor market segregation. I offer an explanation for such a change in production technology by deriving a task-based microfoundation for complementarity, documenting empirical support for this explanation, and quantitatively evaluating it.[13] More generally, the literature usually treats complementarity as exogenous. This limits our understanding of the economic forces that drive changes in sorting. This paper instead endogenizes (coworker) complementarities by modelling the organizational task assignment problem, which ties complementarities to specialization. I also develop an approach to directly measure complementarities in widely available micro data and provide evidence that sorting is tied to complementarities, which supports a key premise of this literature.

Fifth, the paper relates to a small but growing literature that highlights how team production shapes macroeconomic outcomes. In addition to studies of innovation (Akcigit *et al.*, 2018; Ahmadpoor and Jones, 2019; Pearce, 2022), recent papers on coworker learning underscore the dynamic importance of coworker sorting for human capital growth, as stronger sorting deprives less productive workers of opportunities to learn from more knowledgeable team members (Jarosch *et al.*, 2021; Herkenhoff *et al.*, 2022; Hong, 2022). While my model omits such spillovers, these studies underline the importance of understanding the determinants of coworker sorting. I contribute a quantitative framework that explicates how specialization among team members shapes sorting and allocative efficiency.

OUTLINE. Section 2 lays out motivating empirical evidence. Section 3 presents and characterizes the theoretical model. Section 4 maps the model to micro data, develops a theory-informed strategy to measure complementarities, calibrates and validates the model. Section 5 uses the structural model, along with time series evidence on complementarities, to study the evolution of wage inequality, sorting, and coworker mismatch. Section 6 offers a concluding discussion.

---

and Lindenlaub and Postel-Vinay (2023). For recent advances in the structural modeling of firms in labor markets more generally, see also Bilal and Lhuillier (2021); Bilal *et al.* (2022); Elsby and Gottfries (2022); Engbom and Moser (2022); Gouin-Bonenfant (2022); Haanwinckel (2023).

[12]Boerma *et al.* (2021) resolve the major challenge of solving a (frictionless) model of sorting by heterogeneous firms with teams of heterogeneous workers when technology is submodular, instead of supermodular as in this paper. Their quantitative analysis does not match the empirically observed rise in the between-firm share of wage inequality, which is a key contribution of the present paper.

[13]Chade and Eeckhout (2020) provide an alternative theory of assortative but stochastic matching with teams. A version of their model in which knowledge spillovers in downstream markets affect firms' hiring choices likewise predicts, qualitatively, that increased complementarities tilt wage inequality toward the between-firm component.

2 Descriptive evidence

This section describes key trends in the data that inform the quantitative model. Section 2.1 introduces the underlying data, which are also used in the remainder of the paper. Section 2.2 revisits evidence on the evolution of wage inequality in Germany and documents patterns of positively assortative matching between workers. Section 2.3 examines changes in the nature of tasks which speak to an increasing specialization of worker skills.

2.1 Data

The primary dataset is the Sample of Integrated Employer-Employee Data (SIEED) from the German Institute for Employment Research (IAB). The SIEED covers the entire workforce of a 1.5% sample of all establishments in Germany and tracks the complete employment biographies of these individuals, including when they are not employed at the sampled establishments. Establishments are distinguished by ownership as well as industry and location. For ease of language, in the remainder of the text I use the terms "establishment" and "firm" interchangeably.

For most analyses, I rely on an annual panel, created from the original spell-level data. When parallel spells exist, the highest-paying job is retained as the main episode. Wages, measured at the daily level, are deflated to 2015 euros, and top-coded observations are imputed using standard practices. The sample is restricted in several steps. I initially select full-time employed individuals at establishments in West Germany, aged 20-60, with real daily wages of at least 10 Euros. Then, observations in establishment-year cells containing fewer than ten worker observations are dropped. Subsequently, I restrict attention to the largest connected set, as we will rely on worker mobility to identify time-invariant worker types, as well as dropping singletons. The final sample, spanning 1985-2017, comprises 17,126,027 observations for 1,982,239 unique individuals. More details, including summary statistics, are in Appendix B.1.1.

As a baseline, I use wages that are residualized with respect to age and job tenure as well as years. The goal is to maintain consistency across empirical and theoretical analyses, anticipating that the structural model will not incorporate life-cycle, on-the-job learning, or aggregate productivity growth. To calculate residualized wages, I regress the (raw) log real daily wage of worker i in year t, $\tilde{w}_{it}$, on a person fixed effect and a time-varying characteristics index, X_{it}, which includes year dummies, a cubic in age and a quadratic in job tenure.[14] As an input into the analysis, I then use $\ln w_{it} = \tilde{w}_{it} - X'_{it}\hat{\beta}$, which includes the individual fixed effects. The variance of log residualized wages accounts for around 81% of the variance in raw log wages.

To examine shift in the nature of work over time, I supplement the SIEED with micro data

[14] 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), I normalize age around 40.

from the BIBB/IAB and BIBB/BAuA Employment Surveys. These large-scale, representative surveys have been conducted roughly every seven years since 1979. Each survey asks employees to indicate which tasks they perform on the job. Occupation codes are harmonized across waves. Hence, the survey is ideal for monitoring how the task content of production evolves *within* occupations, an aspect which has been shown to be at least as significant as shifts in occupational employment shares (Spitz-Oener, 2006; Atalay *et al.*, 2020). Furthermore, the occupation-level data from these survey can be linked to the SIEED. Appendix D contains further details.

2.2 The evolution of labor market inequality

THE ROLE OF FIRMS IN SHAPING WAGE INEQUALITY. I begin the descriptive analysis by revisiting – through the lens of the SIEED – the evidence that widening differences between firms are an important feature of the rise in wage inequality.

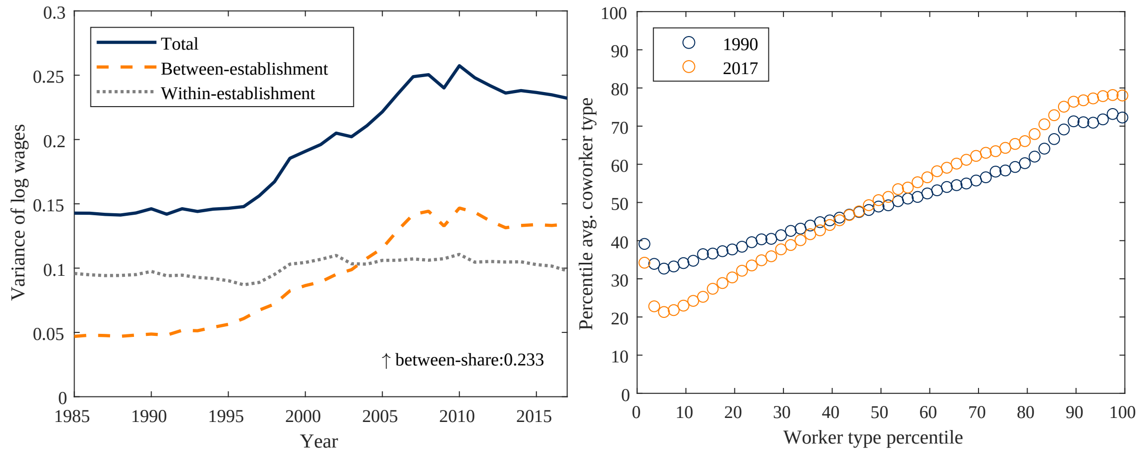
Figure 1a indicates that wage inequality in Germany has risen substantially since the mid-1980s, and that this increase is primarily accounted for by widening pay gaps between firms. The figure presents the yearly total variance of log wages (solid line) and decomposes it into between-employer (dashed line) and within-employer (dotted line) component. By the law of total variance, the total variance equals the sum of the two components. It rose from an average of 0.143 in the interval 1985-1992 to 0.241 during 2010-2017. Of this 0.097 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 33.6% in 1985-1992 to 56.8% during 2010-2017. This trend is consistent with what has been documented by Card *et al.* (2013) and Song *et al.* (2019), among others.

These trends are robust to a battery of robustness checks. Appendix B.3 confirms that between-firm differences play an increasingly important role in accounting for overall person-level wage inequality when considering different measures of wages, within-occupation, and within-industry. Moreover, the trend is also evident in countries other than Germany, albeit the rise in the between-employer share is indeed particularly marked there. To summarize:

Fact 1 (Wage Inequality). *The overall increase in German wage inequality between 1985 and 2017 is primarily accounted for by widening pay gaps between firms rather than within them.*

COWORKER SORTING. That wage inequality stems from both differences in pay among employees *within* firms and *between* firms underscores that each firm is essentially a collection of workers. Following this line of thought, I next examine to what extent coworkers' pay-relevant characteristics are systematically related.

This analysis necessitates measuring worker types. I construct my baseline measure using a



(a) Wage inequality

(b) Coworker sorting

Figure 1: The evolution of wage inequality and coworker sorting

Notes. The left panel shows the variance of log (residual) wages, decomposed into between- and within-employer components. The latter is given by the person-weighted variance of firm-level average log wages. 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.

standard approach in the empirical literature by estimating worker fixed effects, denoted α_i for worker i, from two-way fixed effect wage regression in the spirit of Abowd *et al.* (1999, AKM). In a pre-estimation step I cluster firms based on their empirical wage distribution using a k-means algorithm, following Bonhomme *et al.* (2019), to mitigate a well-known incidental parameter bias problem (Andrews *et al.*, 2008). Letting $\mathbf{1}(J(i,t) = k)$ denote dummies indicating which cluster k firm the employer of i in period t, $j(i,t)$ has been assigned to, I estimate

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

where X'_{it} contains the same time-varying observables as in the previous section, and ϵ_{it} is a residual.[15] I estimate equation (1) separately for five periods that split the sample into intervals of roughly equal duration. Details are in Appendix B.2.1 The average coworker fixed effect for individual i in period t is then defined as $\alpha_{-it} = \frac{1}{|S_{-it}|}\sum_{k \in S_{-it}} \alpha_k$, where $S_{-it} = \{k : j(kt) = j(it), k \neq i\}$ is the set of coworkers with the same employer as i in year t.

Table 1 reports a measure of *coworker sorting* for each of the five sample periods. This measure is the correlation between a worker's own type and that of her coworkers (Lopes de Melo, 2018). Column 1 reveals that coworker sorting is both positive and increasing. It has risen

[15]Appendix B.4.7 reports results when time-varying observables are not controlled for.

	Coworker sorting	
Period	Within-economy type ranking	Within-occupation type ranking
1985-1992	0.427	0.423
1993-1997	0.458	0.443
1998-2003	0.495	0.452
2004-2009	0.547	0.470
2010-2017	0.617	0.519

Table 1: 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.

from an average of 0.427 during 1985-1992 to 0.617 during 2010-2017 – a 44.5% increase.[16]

The second column of Table 1 examines to what extent this increase in coworker sorting reflects greater segregation along occupational lines, whereby higher-pay and lower-pay occupations cluster in distinct firms. Specifically, it displays the correlation between worker and coworker types when the type measures are standardized within (2-digit) occupation-years. The increase is smaller (+0.096 instead of +0.19) but still very notable – a 22.7% rise – indicating that top-performing workers *within* each occupation are increasingly likely to work together.

Figure 1b offers a more disaggregated perspective on coworker sorting through a binscatter plot. The plot shows average coworker type for each worker type, comparing data from 2017 (orange line) and 1990 (blue line). The graph reveals an upward-tilted slope in 2017, indicating great labor market polarization: high types increasingly work with other high types, while the opposite is true for low types, with less change in the middle. The figure also confirms that the rise correlation is not due to isolated, idiosyncratic changes in the type distribution.

A limitation of the preceding analysis is that worker types are inferred from wage data. While the resulting worker type measure is cardinal, granular and comparable across countries, it risks conflating shifts in skill quantities and prices. Two points alleviate this concern. First, a similar rise in labor market sorting is observed when using years of schooling as an alternative proxy for productive capacity (see Appendix B.3.5). Second, other research using non-wage skill measures,

[16] Appendix B.3.4 supplements the analyses here with a series of AKM-based wage variance decompositions. This popular diagnostic tool ties wage dispersion to the variances and covariance of worker and firm FEs. Consistent with the existing literature, I find an important role for changes in workforce composition across firms in explaining the rise in between-firm inequality. Increased variation in the firm-level average worker FEs and the covariance of firm and worker FEs jointly explain over fourth fifths of the increase in the between-employer component, with the remainder accounted for by a rise in the variance of firm FE. In the absence of a structural model that ties the statistical terms to primitives, caution needs to be exercised, though, when interpreting these wage variance decompositions in terms of labels such as "firm-specific pay premia" or "worker-firm sorting". The interpretation of *firm* fixed effects is particularly difficult (e.g., Eeckhout and Kircher, 2011; Lochner and Schulz, 2022).

such as Håkanson *et al.*'s (2021) analysis of Swedish administrative data that includes test scores from military enlistment tests, support the finding of increased coworker sorting. In summary:

Fact 2 (Coworker sorting). *Highly productive workers increasingly cluster in the same firms, segregated from other, less productive workers, who cluster in different firms.*

2.3 Worker specialization and the evolving task content of production

This paper argues that to explain the labor market patterns collected in Facts 1 and 2, we should consider that production typically involves *teams* of workers with *specialized* skills. By "team production" I refer to the ubiquitous economic phenomenon that production involves multiple workers whose productivity may be interdependent.[17] I take "specialization" to describe the degree to which an individual's productivity varies across different tasks required for production.[18]

This section provides evidence in support of the idea that – alongside growing *between*-firm inequality in workforce talent and wages – the nature of work being performed *inside* of firms has likewise meaningfully changed; specifically, the importance of specialization has grown over the past four decades. Despite its prominent place in economic theory, directly measuring specialization is challenging. I therefore draw on multiple sources of evidence: micro data on tasks, both by itself and linked to the SIEED; case studies; and the broader literature.

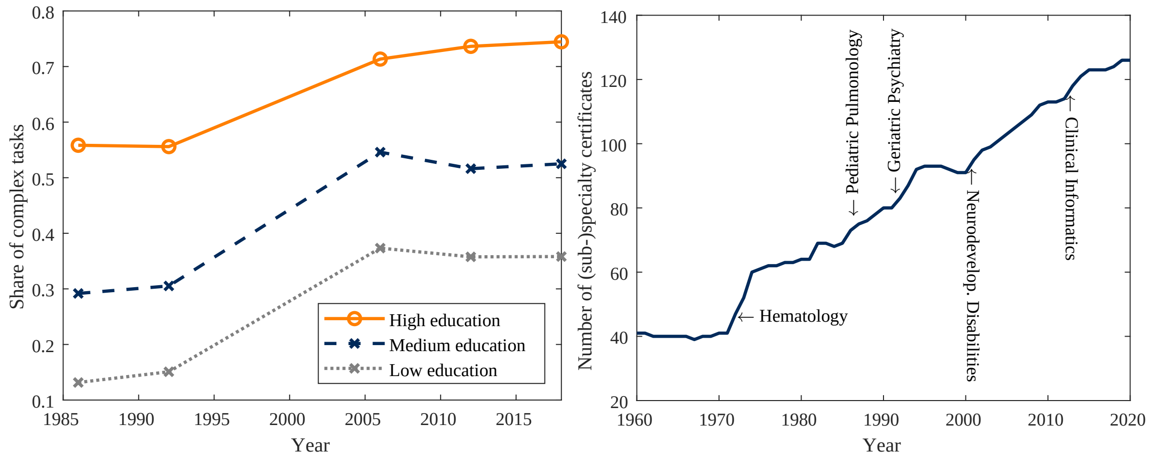
Data on the task content of production can be used to construct proxy measures of specialization that cover the economy as a whole and offer longitudinal insights. My primary proxy is the share of tasks that are abstract and non-routine, henceforth termed "task complexity." Complex tasks – like teaching, investigating or organizing – demand cognitive skills and cannot be executed by following pre-specified rules (cf. Autor *et al.*, 2003). Workers are likely to perform similarly across different routine tasks, since almost definitionally no specific skills are required (Martellini and Menzio, 2021, p. 340). In contrast, non-routine tasks offer greater scope for productivity variation across tasks.[19] Thus, task complexity can be viewed as a proxy for the extensive margin of specialization.

To chart the temporal evolution of task complexity in Germany, I use the BIBB micro data introduced in section 2.1. I calculate the share of complex tasks in each respondent's activities, then average across four time points and different educational or occupational groups, following Spitz-Oener (2006). Appendix D provides details. To corroborate the reliability of task complexity

[17]This definition is similar to that offered in the seminal work of Alchian and Demsetz (1972, Section II). While richer conceptualizations of team production as involving collaborative problem solving or task-oriented, flexible team rotation exist – and are intuitively related – they are not essential to this paper's argument.

[18]As used here, the concept should thus be understood in terms of the task-specificity of human capital, as opposed to time allocation – what tasks you concentrate on doing. In the theoretical model, those two distinct notions will be tied together endogenously.

[19]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 are generally associated with complex tasks.



(a) Rising task complexity

(b) Deepening medical specialization

Figure 2: Rising economy-wide task complexity & deepening medical specialization

Notes. Panel 2a 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 2b 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.

as a proxy for specialization, I compare it, in the cross-section, to an alternative measure that gauges worker specialization more directly but which is available only for recent years. Specifically, Bloesch *et al.* (2022) use U.S. online vacancy data to construct an occupation-level measure of within-firm task differentiation across positions. The two proxies yield very similar rankings of occupations: managerial and professional jobs score the highest, followed by technicians, while routine service and manual jobs exhibit the lowest levels of task complexity and within-firm differentiation. [20] This suggests that both proxies capture similar latent features.

Figure 2a depicts an upward trend in the share of complex tasks reported, increasing from 0.252 in 1986 to 0.647 in 2018. This increase was especially pronounced between the 1990s and early 2000s, and manifested across among all education groups. Moreover, it primarily resulted from within-occupation changes, not employment shifts across occupations (see Appendix D), echoing other studies that employ different methods (Atalay *et al.*, 2020). The literature has highlighted automation as an impactful technological driver, displacing humans in routine tasks and leaving us to handle the more complex problems (Acemoglu and Restrepo, 2018).[21]

In Appendix B.3.6 I document additional evidence suggestive of a rise in specialization.

[20]At the 1-digit level of ISCO-08 occupations, the correlation between Bloesch *et al.*'s (2022) differentiation measure and the non-routine abstract score from Mihaylov and Tijdens (2019), which is likewise available for the ISCO classification, is 0.57. Many thanks to Justin Bloesch for kindly sharing their data.

[21]As this point and the literature review in Section 1 make clear, there is a very extensive literature documenting the evolving nature of work. These studies typically focus on the resulting labor displacement effects or changing skill prices, whereas here I highlight implications for human capital specificity.

Notably, by linking the administrative SIEED data and the BIBB survey data, I find that, over time, individuals have come to perform more similar tasks across different jobs – consistent with task-specific human capital being increasingly important (cf. Gathmann and Schönberg, 2010). Moreover, existing literature, notably Grigsby (2023), points in the same direction.

Complementing the evidence from task data, I next consider three examples that illustrate changes in specialization more tangibly. The first example focuses on science. Research indicates that a growing "burden of knowledge" has compelled individual researchers toward increasingly narrow expertise.[22] Concurrently, teamwork has grown in prevalence (e.g., Wuchty *et al.*, 2007), facilitating the integration of distributed, embodied knowledge. This example highlights a connection between specialization and efficiency gains achievable through team work. While a single scientist may still be capable of performing all tasks required for a project, collaboration between individuals with specialized skills enables each to pursue their comparative advantage.[23]

The second example considers the medical sector. Over recent decades, the composition of work has shifted away from routine tasks manageable by most healthcare professionals, like patient data retrieval, toward others that require advanced training, for instance operating novel equipment like advanced ventilators or smart IV pumps. Thus, Figure 2b shows that the number of distinct speciality certificates issued by the American Board of Medical Specialities nearly doubled from 1980 to 2020. This example illustrates how specialization results from the interplay between technological shifts in task composition and human capital acquisition decisions. The model will treat specialization in reduced-form, without explicitly disentangling these two dimensions.[24]

A third example helps clarify that the trend toward increased specialization may not be inexorable. Early studies of novel large-language artificial intelligence (AI) models suggest that these technologies enable individuals to perform tasks outside their previous areas of expertise (e.g., Brynjolfsson *et al.*, 2023). Consequently, tools like chatGPT may actually reverse the historical trajectory toward greater specialization, a theme I revisit in Section 6.

Summarizing this section:

[22] See, for instance, Jones (2009); Akcigit *et al.* (2018); Ahmadpoor and Jones (2019); Pearce (2022).

[23] This example foregrounds the role of increasing knowledge costs in driving specialization. Alternatively, theories of knowledge-based hierarchies highlight that declining communication costs can also lead to narrower individual expertise (Garicano, 2000). Bloom *et al.* (2014) corroborate this mechanism using data from U.S. and European manufacturing firms, with intranets exemplifying a communication improving technology. An example from software development further illustrates the variety of real-world technological advances that can underpin a rise in specialization. Version control systems (e.g., Git) allow multiple developers to simultaneously contribute to a shared codebase, which encourages specialization as individuals can rely on teammates' expertise for other parts.

[24] One could plausibly argue that the nature of production-related knowledge is the more fundamental factor. If task-specific productivities reflect human capital investment, dispersion in productivity will be significant when the returns to specialization are high. This is likely if knowledge acquisition for any single task involves increasing returns due to fixed costs like time or ample opportunities for learning-by-doing (Rosen, 1983; Alon, 2018).

Fact 3 (Specialization). *Worker specialization has strengthened since the mid-1980s.*

To recap, a prominent feature of the overall rise in wage inequality since the mid-1980s is the increased between-firm share (Fact 1); workers increasingly tend to be employed together with coworkers of similar productivity (Fact 2); and specialization has strengthened (Fact 3). Next, I develop a theoretical model in which a rise in specialization fosters more positively assortative matching among coworkers, which generates widening wage gaps between but not within firms.

3 Theoretical model: team production and matching

This section presents and characterizes the theoretical model. Section 3.1 sets out a static theory of firm organization and analytically derives a task-based microfoundation for coworker talent complementarities, revealing their origin in worker specialization. Section 3.2 embeds the associated firm-level production function into a dynamic equilibrium matching model, wherein complementarities shape labor market sorting and the firm-level structure of wage inequality.

3.1 Team production & the origins of coworker complementarity

I start by considering a single production unit and ask what the optimal assignment of tasks to a team of ex-ante heterogeneous workers is, treating their number and composition as exogenous.

3.1.1 Environment

A team consists of a discrete number n of workers who perform tasks required for the production of a single good. The production technology is owned by the employer, who allocates workers' time across tasks in order to maximize production. Each worker has a type $x_i \in [0,1]$, $i = 1, \ldots, n$, which denotes their rank in the worker talent distribution, and is endowed with one unit of time that they supply inelastically. For convenience, denote the set of team members by $\mathcal{S} = \{1, \ldots, n\}$.

Final good production. The final good, or service, is produced from a unit mass of tasks $\tau \in \mathcal{T} = [0,1]$ according to a constant elasticity of substitution (CES) aggregator,

$$Y = \left(\int_{\mathcal{T}} q(\tau)^{\frac{\eta-1}{\eta}} d\tau \right)^{\frac{\eta}{\eta-1}}, \tag{2}$$

where $q(\tau)$ denotes the amount of task τ used and $\eta > 0$ is the elasticity of substitution across tasks. I next describe how $q(\tau)$ is related to task production by individual workers.

TASK PRODUCTION. Tasks are fulfilled by workers. Denote the amount of task τ produced by worker i as $y_i(\tau)$. In classical Ricardian fashion, production is linear in efficiency units of labor:

$$y_i(\tau) = a_1 z_i(\tau) l_i(\tau), \tag{3}$$

where $z_i(\tau)$ is i's efficiency in producing task $\tau \in \mathcal{T}$, the parameter a_1 is strictly positive and controls the sensitivity of task output to worker efficiency, and $l_i(\tau)$ is the time dedicated by i to task τ. With this notation, the time constraint for worker i is

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

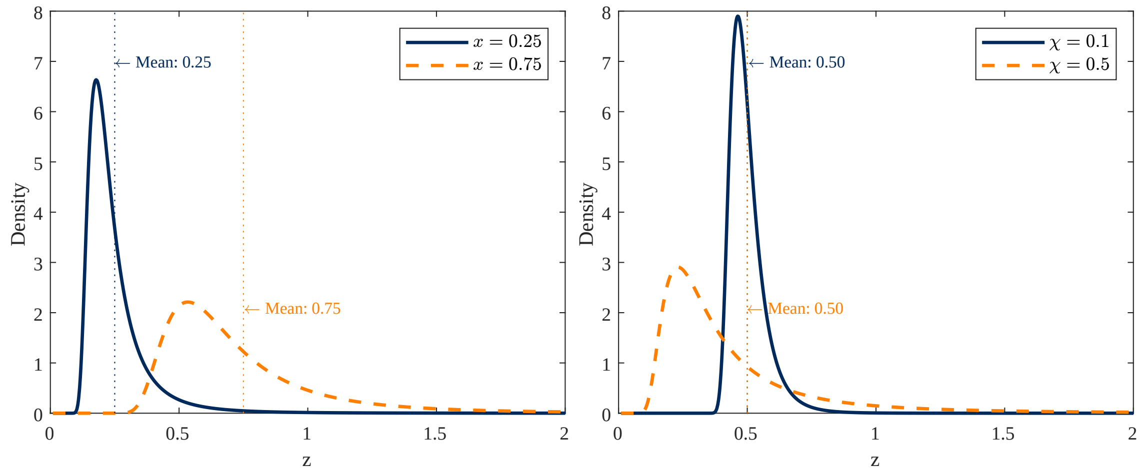
TASK-SPECIFIC EFFICIENCIES. Workers are heterogeneous along two dimensions: vertically, as captured by their type x_i, and horizontally, meaning that conditional on x_i, a given worker may be better in some tasks than in others. I treat the task-specific efficiencies for worker i, $\{z_i(\tau)\}_{\tau \in \mathcal{T}}$, as the realizations of a Fréchet-distributed random variable, drawn independently for each worker i. This assumption parallels trade models in the tradition of Eaton and Kortum (2002) and parsimoniously captures both vertical and horizontal differentiation among workers. Thus, for all $z \geq 0$, the distribution of efficiencies for worker i is

$$G_i(z) := Pr\{z_i(\tau) \leq z\} = \exp\left(- \left(\frac{z}{\iota x_i} \right)^{-1/\chi} \right). \tag{5}$$

A worker's type, x_i, determines the scale of the worker-specific distribution. The (inverse) shape parameter, $\chi \in (0, \infty)$, determines the degree of dispersion. It is identical for all workers. Lastly, $\iota := \Gamma(1 + \chi - \eta\chi)^{\frac{1}{1-\eta}}$ is a scaling term, with Γ denoting the Gamma function.[25]

A few remarks are in order to unpack this description, which is at the heart of the model. Equation (5) implies that workers may differ along both *absolute* and *comparative* advantage lines. Each worker can produce any task, but they are potentially heterogeneous in both their *average* task-specific production efficiency and, conditional on that average, in the distribution of efficiencies *across* different tasks. A worker's type x connotes absolute advantage or talent. Formally, the expectation of $z_i(\tau)$ is greater when x_i is higher. Intuitively, *if* a high-x worker is specialized in a particular task, they will be very productive at it. In addition, there is dispersion in worker-task specific efficiencies; for a given value of x_i, i is better in some tasks than in others. The degree of dispersion is increasing in χ. Figure 3 illustrates how the distribution (density) of

[25]A couple of technical remarks: First, throughout the paper it is assumed that $1 + \chi(1 - \eta) > 0$, for reasons discussed when deriving the optimal organization. Beyond requiring that tasks not be too substitutable, under the maintained assumptions the precise 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 production levels.



(a) Worker-specific scale parameter

(b) Technological (inverse) shape parameter

Figure 3: Illustration of the distribution of worker-task specific efficiencies

Notes. This figure illustrates how the probability density function associated with the Fréchet distribution in equation (5) varies with the scale parameter x and (inverse) shape parameter χ. The left panel assumes $\chi = 0.3$, the right panel assumes $x = 0.5$. For this illustration, $\iota = \Gamma(1 + \chi)$.

task-specific productivities varies with x and χ, respectively.

The key parameter for comparative statics exercises is χ. Labelled "specialization," it formally controls the degree of dispersion in an individual worker's task-specific productivities and thus – from a team perspective – the importance of comparative advantage across coworkers.[26] Recurring to the discussion in Section 2.3, χ captures in reduced-form the nature of both task requirements and worker skills, which jointly determine the importance of specialization.

ROLE OF FIRMS. Completing the model description is an account of how team production differs from multiple individuals producing the final good separately. Inside a team, collaboration in the sense of a division of labor is possible. Formally, the amount of any task that is available for final good production, $q(\tau)$, is the sum over production of that task by *all* team members,

$$q(\tau) = \sum_{i=1}^{n} y_i(\tau). \tag{6}$$

This account rests on the idea that a central role of firms is to coordinate the collaboration of workers with specialized knowledge (cf. Becker and Murphy, 1992; Garicano, 2000). On this view, a firm not merely incorporates the intellectual property rights, or a "recipe", for a particular product, nor is it just the sum of machines and tools with which workers are equipped. Much of the knowledge required for production is intangible and embedded in individuals with limited

[26] Section 3.1.4 offers a more careful treatment of the distinction between within-worker productivity dispersion and across-coworker comparative advantage.

time to learn and work, as opposed to that knowledge being codified and tradeable on markets. Hence, its mobilization and efficient use requires individuals to specialize and collaborate. An important role of firms is to coordinate this process.[27]

ORGANIZATIONAL OPTIMIZATION PROBLEM. This role of the firm is reflected in its optimization problem. Solving a 'mini-planner' problem, it chooses 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, subject to equations (2)-(6). The associated Lagrangean is

$$\begin{aligned}\mathcal{L}(\cdot) = Y + \lambda\left[\left(\int_{\mathcal{T}} q(\tau)^{\frac{\eta-1}{\eta}} d\tau\right)^{\frac{\eta}{\eta-1}} - Y\right] + \int_{\mathcal{T}} \tilde{\lambda}(\tau)\left(\sum_{i=1}^{n} y_i(\tau) - q(\tau)\right) d\tau \\ + \sum_{i=1}^{n}\left\{\lambda_i^L\left(1 - \int_{\mathcal{T}} l_i(\tau)d\tau\right) + \int_{\mathcal{T}} \lambda_i(\tau)\left(a_1 z_i(\tau) l_i(\tau) - y_i(\tau)\right) d\tau + \int_{\mathcal{T}} \bar{\lambda}_i(\tau) y_i(\tau) d\tau.\right\},\end{aligned}$$

Here, λ, λ_i^L, $\lambda_i(\tau)$, $\tilde{\lambda}(\tau)$ respectively 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. Lastly, $\bar{\lambda}_i(\tau)$ relates to a non-negativity constraint in task production.

3.1.2 Solving the organizational problem

Solving for the optimal production plan involves two main steps. First, we derive the demand for tasks for a given set of shadow prices, treating those as known. Then, secondly, we determine these shadow prices given the distribution of task-specific productivities across workers. To begin, taking the first-order condition (FOC) with respect to $q(\tau)$, substituting for Y from equation (2) and defining $Q(\tau) := \tilde{\lambda}(\tau)q(\tau)$ yields

$$\frac{Q(\tau)}{\lambda Y} = \left(\frac{\tilde{\lambda}(\tau)}{\lambda}\right)^{1-\eta}. \tag{7}$$

Now integrate on both sides and use that the shadow value of all tasks used is related to total production (as derived in Appendix A.1.1),

$$Q := \int_{\mathcal{T}} Q(\tau) d\tau = \lambda Y. \tag{8}$$

[27] Rivkin and Siggelkow (2003, 292), quoted in Dessein and Santos (2006) write: "The [qualitative management] literature is unified in what it perceives as the central challenge of organizational design: to divide the tasks of a firm into manageable, specialized jobs, yet coordinate the tasks so that the firm reaps the benefits of harmonious action."

This gives an expression for the shadow cost index:

$$\lambda = \left(\int_{\mathcal{T}} \tilde{\lambda}(\tau)^{1-\eta} d\tau \right)^{\frac{1}{1-\eta}}. \tag{9}$$

Next, who should produce which tasks, and what does that imply for the shadow costs faced by the firm, i.e., for $\{\tilde{\lambda}(\tau)\}_{\tau \in \mathcal{T}}$? To answer this question, observe first that the FOC with respect to $y_i(\tau)$ implies that $\tilde{\lambda}(\tau) = \lambda_i(\tau)$ if $y_i(\tau) > 0$. Since some worker will provide a given task τ, and with task production (equation (3)) featuring constant returns to scale, cost-minimization requires that the shadow value of a task be equal to the minimum shadow cost of producing it. When combined with the FOC w.r.t. $l_i(\tau)$, this implies that

$$\tilde{\lambda}(\tau) = \min_{i \in \mathcal{S}} \Bigg\{ \underbrace{\frac{\lambda_i^L}{a_1 z_i(\tau)}}_{\lambda_i(\tau)} \Bigg\}, \tag{10}$$

so that worker i produces task τ if i has the lowest ratio of any worker's shadow value of their time over their efficiency in producing that task.

The following lemma summarizes what share of tasks is performed by each worker, and at what (shadow) cost the final good can be produced.

Lemma 1. *If workers' task-specific efficiencies are independently Fréchet-distributed, then:*

(i) *The shadow cost index is*

$$\lambda = \left(\sum_{i \in \mathcal{S}} \left(\frac{a_1 x_i}{\lambda_i^L} \right)^{1/\chi} \right)^{-\chi}. \tag{11}$$

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

$$\pi_i := Pr\{\lambda_i(\tau) \leq \min_{k \in \mathcal{S} \setminus i} \lambda_k(\tau)\} = \frac{(x_i/\lambda_i^L)^{1/\chi}}{\sum_{k \in \mathcal{S}} \left(x_k/\lambda_k^L\right)^{1/\chi}}. \tag{12}$$

(iii) *The shadow value of all tasks used in final goods production that were produced by worker i, defined as $Q_i := \int_{\mathcal{T}} \tilde{\lambda}(\tau) y_i(\tau) d\tau$, is a fraction π_i of the total shadow value of tasks used:*

$$Q_i = \pi_i Q. \tag{13}$$

Proof. Appendix section A.1.2. □

That this type of problem – with an arbitrarily large, discrete number of potential producers (i.e., workers) of a continuum of tasks – is tractable under the assumptions made is a key insight of Eaton and Kortum (2002). Crucially, the max-stability property of the Fréchet distribution means that the maximum of Fréchet-distributed random variables is also Fréchet. Hence, the distribution of shadow costs conditional on any task being performed by the worker with a comparative advantage is analytically tractable. Using Lemma 1, I next characterize the optimal organization of production.

3.1.3 Characterizing the optimal organization of production

Intuition for the properties of the optimal organization of production can be gained in three steps. First, the optimal organization features complete division of labor, with each worker assigned those tasks in which she has a comparative advantage. Thus, per equation (10), any task τ is produced by whichever worker is most efficient in producing that task, as captured by $z_i(\tau)$, relative to the task-invariant, endogenous shadow cost of their time, λ_i^L.[28] Second, better workers perform more task. The share of tasks produced by worker i can be shown to equal $\pi_i = \left(x_i^{\frac{1}{1+\chi}}\right)\left(\sum_{k\in\mathcal{S}}(x_k)^{\frac{1}{1+\chi}}\right)^{-1}$, hence $\frac{\partial \pi_i}{\partial x_i} > 0$. Third, and most importantly, if $\chi > 0$, then the value of any worker's time is a function not only of her own endowments but also her coworkers. The shadow value of all tasks produced by worker i is

$$Q_i = n^{1+\chi} a_1 x_i^{\frac{1}{1+\chi}} \tilde{x}_i^{\chi}, \tag{14}$$

where $\tilde{x}_i = \frac{1}{n}\sum_{k\in\mathcal{S}} x_k^{\frac{1}{1+\chi}}$. Other things equal, the division of labor enables a worker to focus on those tasks in which she is particularly adept, raising the total value of tasks she can produce. For a given team size, better coworkers facilitate this concentration on high-efficiency tasks. These coworker interdependencies are increasing in χ; it is when individuals are specialized that their productivity will depend on colleagues' capability. (In contrast, if workers are equally productive in all tasks ($\chi \to 0$), their overall productivity is independent of the task assignment.) Crucially, as equation (14) underlines, the absolute gain from better coworkers is greater for more talented workers. Simply put, more output is at stake.

The following result concisely summarizes how team members' talents jointly pin down team output.[29] Exploiting that the Fréchet assumption renders the firm's 'demand' for a worker's time

[28] Strictly speaking, we need to specify a tie-breaking rule for tasks where multiple workers have the same shadow cost. But those tasks have mass zero anyway.

[29] This aggregation result relates to the seminal paper of Houthakker (1955) and, more recently, Acemoglu and Restrepo (2018), Martinez (2021) and Ocampo (2022). In the IO literature, an important reference is Anderson *et al.* (1989). In independent work, Dvorkin and Monge-Naranjo (2019) likewise exploit properties of the Fréchet distribution to derive a microfoundation for an aggregate CES production function.

an iso-elastic function of its shadow price, one can show that under the optimal organizational plan, the number and quality types of team members are sufficient statistics for team output.[30]

Proposition 1 (Aggregation result)**.** *Team output Y can be written as a function of members' talent types, $f : [0,1]^n \to \mathbb{R}_+$,*

$$f(x_1, \cdots, x_n) = \underbrace{n^{1+\chi}}_{\textit{efficiency gains}} \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}(a_1 x_i)^{\frac{1}{1+\chi}}\right)^{1+\chi}}_{\textit{complementarity}}. \tag{15}$$

Proof. See Appendix Section A.1.3. □

The first term summarizes the benefits from team production relative to each worker producing all tasks themselves, which are increasing in χ. Suppose that there is no division of labor, perhaps because workers are not coordinated by a firm. Then each worker produces all tasks (in proportions minimizing their individual shadow costs of producing the final good), and it is straightforward to show that total output is simply $\sum_{i=1}^{n} a_1 x_i$. The "Smithian" efficiency gains from team production relative to this scenario, which is captured by the first term of equation (15), echo a long-standing theoretical literature (e.g., Becker and Murphy, 1992; Garicano, 2000), as well as cohering with recent empirical papers on firm organization (cf. Section 1).

The second term takes the form of a CES function. Intuitively, the more *horizontally* differentiated team members are at the task level, the more limited is the scope for substitution in terms of workers' *absolute advantage*. Simply put, even the most talented worker may lose much time when performing tasks they are not specialized in.[31] The following corollary summarizes these coworker interdependencies in terms of the elasticity of substitution.[32]

Corollary 1 (Coworker complementarity)**.** *For any homothetic production function $f(x_1, ..., x_n)$, the inverse elasticity of substitution ("elasticity of complementarity") is defined as $\gamma_{ij} := \frac{\partial \ln(f_j/f_i)}{\partial \ln(x_i/x_j)}$,*

[30] Notice that the elasticity of substitution across tasks, η, does not show up in equation (15). Technically, as noted by Eaton and Kortum (2002, Footnote 18), this holds as long as we maintain that $1 + \chi - \eta\chi > 0$, 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). In a more general setting, the strength of coworker complementarities would also be influenced by the value of η.

[31] Proposition 1 provides a microfoundation for the reduced-form model assumed by Ahmadpoor and Jones (2019) in their study of team production and matching in science and invention. They postulate that team output corresponds to $\beta_n \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{\zeta}\right]^{1/\zeta}$, where β_n "captures impact benefits associated with teamwork" while the second term measures by how much these benefits diminish as the gap between team members' abilities grows.

[32] 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 (15), the Morishima elasticity is identical across worker pairs and identical to the Hicks and Allen definitions (Blackorby and Russell, 1981). For recent applications of the MES, see Baqaee and Farhi (2019) and Ocampo (2022).

where $f_i = \partial f / \partial x_i$ *is a partial derivative. For the production function in equation* (15)*, it is symmetric and identical for all pairs of workers and equal to* $\gamma = \frac{\chi}{1+\chi}$.

Coworker complementarity signifies that team productivity is lowered if, for a given (arithmetic) *average* value of x_i, members vary in their talent. In statistical terms, the "complementarity" term in equation (15) corresponds to the power mean of the underlying x_i values. When $\gamma = 0$, it reduces to the arithmetic mean, consistent with an efficiency-unit treatment of worker types. Otherwise, by the power mean inequality, the stronger the degree of complementarity, the greater is the weight on quality of the least-capable team member(s) in determining total output.

A key take-away from this task-based microfoundation of team production is, therefore, that greater worker-task specialization not only enhances the scope for efficiency gains from collaboration. Output also becomes more vulnerable to relatively lower-quality team members. Specifically, larger values of χ render *coworker mismatch* – dispersion in team members' quality – more costly in terms of output. In short, the composition of a team becomes more important for productivity. The next section therefore examines how firms choose to assembly teams. Before doing so, in the next sub-section I briefly introduce an extension to the team-production model that helps tightening the interpretation of the key structural parameter χ.

3.1.4 Extension: specialization and comparative advantage

So far, we adopted a very low-dimensional parameterization of workers' characteristics, describing n team members' productivity distributions over tasks in terms of an n-dimensional vector of absolute advantage types, $\mathbf{x} = (x_1, \ldots, x_n)'$, and a common shape parameter χ. This approach is rich enough to capture both absolute and comparative advantage considerations while remaining tractable, including in the dynamic setting introduced in the following section. However, it also conflates two distinct concepts, namely within-worker productivity dispersion across task and comparative advantage across coworkers. This brief section sketches an extension that formalizes this distinction and offers an interpretation in the context of team production.

In Section 3.1.1, and specifically equation (5), task-specific productivity draws were assumed to be independent across coworkers. As a result, the common (inverse) shape parameter χ effectively controls two conceptually distinct moments: the dispersion of productivity across tasks for any a given worker, or *specialization*; and the joint distribution of relative productivities for any given tasks across individuals, and therefore the strength of *comparative advantage*.

We can relax the independence assumption by supposing instead that the joint distribution of task-specific productivities across coworkers satisfies

$$\Pr\left[z_i(\tau) \leq z_1, \ldots, z_n(\tau) \leq z_n\right] = \exp\left[-\left(\sum_{i=1}^{n}\left(\left(\frac{z}{\iota x_i}\right)^{-\frac{1}{\tilde{\chi}}}\right)^{\frac{1}{\xi}}\right)^{\xi}\right], \tag{16}$$

where $\tilde{\chi}$ is the common shape parameter of the the marginal Fréchet distributions, while $\xi \in (0, 1]$ controls correlation in draws across workers. The baseline setup featuring independence across coworkers is nested for $\xi = 1$. Yet, with decreasing ξ, coworkers are increasingly similar in *what* they are specialized in; and for $\xi \to 0$, the strength of comparative advantage vanishes despite each worker having specialized skills. Importantly, in the extension $\tilde{\chi}$ unambiguously captures within-worker productivity dispersion and, thus, specialization.

Since the symmetric, multivariate Fréchet distribution (equation (16)) retains the max-stability property, an aggregative team production function analogous to equation (15) can be derived:[33]

$$Y = f(\mathbf{x}; \tilde{\chi}, \xi) = n^{1+\tilde{\chi}\xi}\left(\frac{1}{n}\sum_{i=1}^{n}(a_1 x_i)^{\frac{1}{\tilde{\chi}\xi+1}}\right)^{\tilde{\chi}\xi+1}. \tag{17}$$

Intuitively, this representation tells us that output Y is greater, other things equal, if coworkers are *similar* to each other in terms of *how* talented they are (for a given arithmetic mean of $\mathbf{x}$); and if they are more *dissimilar* in terms of *what* tasks they are good at, that is, if $\xi \to 1$. The latter consideration is abstracted from in the baseline version.

This extension makes clear that all the results discussed thus far go through with $\chi = \tilde{\chi}\xi$. For the remainder of the paper, I will retain the simplifying assumption of independent draws across coworkersm but where there is ambiguity, χ ought to be interpreted in terms of within-worker productivity dispersion, i.e., specialization.[34]

[33] See McFadden (1977, 1980) and Train (2009) regarding the development of these tools in the context of probabilistic choice, and Lind and Ramondo (2023) for a recent application to international trade. The aggregation result developed here is, to the best of my knowledge, novel.

[34] The extension sketched in this sub-section can be embedded into the dynamic matching model as well, in which case ξ is treated as being endogenous to matching decisions, while $\tilde{\chi}$ continues to be a structural parameter. This extended micro-foundation motivates the introduction of meeting-specific productivity shocks that generate random variation in match outcomes conditional on worker types and affords them a natural structural interpretation in terms of *horizontal* match quality: how dissimilar are team members in *what* they are good at, conditional on χ. It also facilitates an alternative strategy to identify χ that directly exploits the production function microfoundation. A future revision of the paper will fully develop these results.

3.2 Team matching & the macro implications of coworker complementarity

Having previously considered one team of given composition, suppose now that there are many workers and many potential employers. What teams of workers would different multi-worker firms hire – who does, in fact, work with whom – and how do matching patterns and wage distribution respond to changes in χ? To answer this set of questions, in this section I embed the aggregative team production model into a general-equilibrium environment where firms and employees meet in a search-frictional labor market.

To motivate the inclusion of search frictions, it is instructive to take a step back and ask, what matching patterns would emerge in a frictionless labor market? The answer is that for *any* value of χ greater than zero, the equilibrium allocation features pure positive assortative matching (PAM), and thus no within-firm worker talent dispersion. This is the efficient outcome. Since production is supermodular in talent for $\chi > 0$, this result immediately follows from Becker (1973), extended to the many-to-one case by Kremer (1993). Specifically, when a worker's marginal productivity is increasing in coworkers' quality, a feasible Pareto improvement exists unless all team mates of a worker of type x are likewise of type x. Although both a low-x and a high-x worker's productivity is increasing in coworker talent, complementarities mean that the benefit from increased coworker talent is differentially larger for the high type. Indeed, for any initial value $\chi_0 > 0$, an increase to $\chi_1 > \chi_0$ does not alter equilibrium matching patterns. Intuitively, in the absence of frictions even weak complementarities imply that firms with other talented employees pn their payroll can always outbid other firms with less talented employees when competing for talented new hires. The equilibrium joint distribution of worker talent is, thus, degenerate; matching is characterized by a single-valued function $\mu(x) = x$ that deterministically relates worker and coworker types.

This strong prediction points to two distinct reasons for incorporating search.[35] First, the prediction of pure positive sorting is counterfactual. The correlation between coworkers' types would need to equal one, which is inconsistent with the data. Second, variation in coworker types conditional on a worker's own type is essential to empirically measure worker types and coworker complementarities. Search frictions align the structural model more closely with the empirical analysis by generating stochastic matching. In their presence, some degree of "mismatch" between team members' talent types is tolerated, since waiting is costly. The equilibrium amount of mismatch in crucially depends on the strength of complementarities.

[35] In principle, other frictions could be considered, including information and screening costs or spatial restrictions. Search frictions, besides being the most common route toward stochastic sorting and they are particularly useful for quantitative analyses, as we can bring to bear micro data on labor market transitions to discipline search costs.

3.2.1 Environment

I embed the team production function into a continuous-time version of Herkenhoff *et al.*'s (2022) equilibrium search model of the economy in steady state. To focus on the key mechanism of changing production complementarities, I abstract from learning and, in the baseline, from on-the-job search (considered in Section 5.4). These simplifications afford a tight characterization of the equilibrium properties in a stylized version of the model, and permit a transparent derivation of a novel identification result that indicates how the strength of complementarities can be measured using micro data.[36]

DEMOGRAPHICS, PREFERENCES AND PRODUCTION. Time is continuous. All agents are infinitely-lived and have risk-neutral preferences over consumption discounted at rate $\rho \in (0, 1)$. There is a unit mass of workers, denoted $d_w = 1$, who are either employed (e) or unemployed (u). Workers are ex-ante heterogeneous, as just described. For ease of exposition, I often refer to a "worker x" instead of a "worker of type x." Noting that the production function (15) is increasing in each argument, I follow Hagedorn *et al.* (2017) in treating x as a worker's rank in the underlying productivity distribution, so the distribution of worker types is uniform. The production function thus maps from an ordinal to a cardinal space.[37] There is a fixed unit mass of firms, $d_f = 1$, that are either idle, with mass $d_{f,0}$, or actively producing in a match. I write down the model assuming that all firms are ex-ante homogeneous, the idea being that the firm matters, namely through team assembly, but its value is tightly linked to the team it employs.[38]

Production requires a firm and either one or two employees. That is, the maximum team size is $n = 2$. This assumption of sharply decreasing returns is evidently unrealistic and made for computational reasons; the state space combinatorically expands with team size, as complementarities mean that we need to keep track of every team member's type. A team size of two is sufficient, however, to study how coworker complementarities shape between-firm inequality. Section 4.1 discusses the mapping between theory and data in light of this assumption.[39]

Denote the production function of a single-worker firm $f_1(x) : [0, 1] \longrightarrow \mathbb{R}_+$, and that of a two-worker firm as $f_2(x, x') : [0, 1]^2 \longrightarrow \mathbb{R}_+$. Note that I indicate the team size state as a subscript to aid orientation when there is scope for confusion, but suppress it otherwise, and I refer to the coworker type by x'. A key benefit from Proposition 1 is that we do not need to keep track of

[36]My notation differs from Herkenhoff *et al.*'s (2022), seeking to maximize comparability with the standard model of matching between workers and one-worker firms (e.g., Hagedorn *et al.*, 2017).

[37]As noted by Hagedorn *et al.* (2017, 2.1.1), if the non-rank distribution of workers is Φ and the production function which takes the non-rank type of worker as an input is $\hat{f}(\hat{x})$, then $f(x) = \hat{f}(\Phi^{-1}(x))$. The logic straightforwardly extends to multiple workers.

[38]A Working Paper version incorporated ex-ante heterogeneity in firm productivity, underscoring that results proved here do not hinge on assuming otherwise, but the gain in generality was outweighed by a loss in clarity.

[39]Appendix section C.2.3 shows that allowing for a maximum team size equal to 3 leads to qualitatively identical and quantitatively very similar predictions as the baseline model.

worker-task specific efficiencies or worker-task assignment, which happens 'in the background'; instead we only need to keep track of workers' quality types.

POPULATION COMPOSITION. Let $d_{m.1}(x)$ denote the measure of producing matches consisting of a firm and one worker of type x, while $d_{m.2}(x, x')$ is the corresponding measure of matches with an additional coworker x'. The following adding-up property holds for workers of type x:

$$d_w(x) = d_u(x) + d_{m.1}(x) + \int d_{m.2}(x, \tilde{x}')d\tilde{x}', \tag{18}$$

so that the total measure of type x, $d_w(x)$, is the sum of type-x workers who are unemployed, $d_u(x)$, in one-worker matches or in two-worker matches. The aggregate unemployment rate is obtained by integrating over the first of these three terms, $u = \int d_u(x)dx$. Similarly for firms,

$$d_f = d_{f.0} + \int d_{m.1}(x)dx + \frac{1}{2}\int\int d_{m.2}(x, x')dxdx'. \tag{19}$$

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

TIMING. Within any increment of time, each matched worker may be exogenously separated from their employer at an exogenous Poisson rate δ. There are no endogenous separations. Next, unmatched workers and all firms engage in random search and matching decisions. Upon meeting, the types x are perfectly observed, but the task-specific productivities are revealed only after the hiring decision is made. Then production and surplus sharing happen.

SEARCH AND BARGAINING. Every unemployed worker engages in search. So do all firms with empty spots, the total number of which is $v = d_{f.0} + \int d_{m.1}(x)dx$. It is not permissible to replace a worker with a better matched candidate, so two-worker firms do not search. Apart from being hard to reconcile with labor market policies in Germany – the economy on which I will estimate the model – such replacement hiring would distract from the central trade-off between match quality and search costs. Unemployed workers contact a firm at an exogenous Poisson rate λ_u.[40] Hence, for any searching firm the rate of meeting an unemployed worker is $\lambda_{v.u} = \frac{\lambda_u u}{v}$. Wages are continuously renegotiated and workers' bargaining power is $\omega \in [0, 1]$. Negotiation 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, which in turn only depends on individual types (cf. Bilal *et al.*, 2022). Moreover, the wage is unaffected by the order with which workers join a team.

[40]This meeting rate could be endogenized by introducing vacancy-posting decisions and a matching function.

3.2.2 Definition and sharing of joint surplus

I now turn to joint value and match surplus functions and describe the protocol by which this surplus is shared. Denote the joint value of a match between a firm and a type-x worker as

$$\Omega_1(x) = V_{f.1}(x) + V_{e.1}(x), \tag{20}$$

where $V_{e.1}(x)$ is x's value of being employed alone at a firm, whose value in turn is $V_{f.1}(x)$. Then the surplus generated by such a match is

$$S(x) = \Omega_1(x) - V_{f.0} - V_u(x), \tag{21}$$

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

Similarly, the joint value of a firm with team (x, x') is

$$\Omega_2(x, x') = V_{f.2}(x, x') + V_{e.2}(x|x') + V_{e.2}(x'|x), \tag{22}$$

with $V_{e.2}(x|x')$ denoting the value of x being employed in a team if the coworker is of type x'. Hence, the surplus generated when a firm that already has employee x' hires a type x worker is:

$$S(x|x') = \Omega_2(x, x') - \Omega_1(x') - V_u(x). \tag{23}$$

This surplus will rise in the output generated by the team, yet fall in both sides' outside values. $S(x|x')$ is not, in general, symmetric in the two arguments even if $\Omega_2(x, x')$ is, because it matters for both outside options whether type x is the potential employee or type x' is (unless $x = x'$).

Turning to surplus sharing, the wage of a worker of type x employed alone satisfies

$$(1-\omega)\big(V_{e.1}(x) - V_u(x)\big) = \omega\big(V_{f.1}(x) - V_{f.0}\big), \tag{24}$$

In a two-worker firm, the wage $w(x|x')$ of a type-x worker with a coworker of type x' satisfies

$$(1-\omega)\big(V_{e.2}(x|x') - V_u(x)\big) = \omega\big(V_{e.2}(x'|x) + V_{f.2}(x, x') - V_{e.1}(x') - V_{f.1}(x')\big). \tag{25}$$

3.2.3 Equilibrium conditions

I focus on a stationary equilibrium. Formally, this equilibrium is characterized by a set of Hamilton-Jacobi-Bellman (HJB) equations, which characterize agents' optimal matching decisions and associated values, coupled with the stationary version of a system of Kolmogorov-Forward equations (KFEs), pinning down the distributions. The optimality conditions influence the distributions through a set of control functions that determine whether random meetings translate into matches. As the bargaining protocol ensures that matching decisions are privately efficient, those are related to surplus values as follows:

$$h(x) = \mathbf{1}\{S(x) > 0\}, \tag{26}$$

$$h(x|x') = \mathbf{1}\{S(x|x') > 0\}. \tag{27}$$

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.[41]

OPTIMALITY. Starting with unmatched agents, the asset value of an idle firm, $V_{f.0}$, satisfies the following HJB equation:

$$\rho V_{f.0} = (1-\omega)\lambda_{v.u} \int \frac{d_u(x)}{u} S(x)^+ dx, \tag{28}$$

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

$$\rho V_u(x) = b(x) + \lambda_u \omega \left[\int \frac{d_{f.0}}{v} S(x)^+ + \int \frac{d_{m.1}(\tilde{x}')}{v} S(x|\tilde{x}')^+ d\tilde{x}' \right]. \tag{29}$$

Notice that we need to take into account the worker's flow value from being unmatched, $b(x)$, as well as differentiating between the worker meeting an unmatched firm or a one-worker firm.

What about the matched agents? To understand the model mechanics, it is instructive to consider the joint value of firm and worker(s) – instead of jumping straight to surplus values – starting with that of a firm employing a pair of workers x and x':

$$\rho\Omega_2(x,x') = f_2(x,x') + \delta\Big[\underbrace{\big(-\Omega_2(x,x') + \Omega_1(x) + V_u(x')\big)}_{S(x|x')} + \underbrace{\big(-\Omega_2(x,x') + \Omega_1(x') + V_u(x)\big)}_{S(x'|x)} \Big].$$

The discounted value contains the flow value of production, and at rate δ either x or x' leaves.

[41] I impose that matching only takes place when the relevant surplus is strictly greater than zero; since the indifference case occurs for a measure zero of agents, this assumption has no bearing on the result.

Lastly, and importantly, the joint value of a firm with a type-x employee satisfies

$$\rho\Omega_1(x) = f_1(x) + \delta\big[-\Omega_1(x) + V_u(x) + V_{f.0}\big] \\ + \lambda_{v.u}\int \frac{d_u(\tilde{x}')}{u}\big(\underbrace{-\Omega_1(x) + V_{e.2}(x|\tilde{x}') + V_{f.2}(x,\tilde{x}')}_{(1-\omega)S(x'|x)}\big)^{+}d\tilde{x}'. \tag{30}$$

The discounted value thus includes the flow value of production. At rate δ the match is exogenously destroyed. But at rate $\lambda_{v.u}$, the firm is contacted by an unmatched worker of type $\tilde{x}'$, who is hired if the sum of values accruing to the firm and x from teaming up exceed their joint value if $\tilde{x}'$ is not hired, that is, if the joint surplus from a match is positive.

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.2.1.

3.2.4 Equilibrium definition and objects

In stationary equilibrium, agents' values must be consistent with the distributions to which they give rise, and vice versa. Appendix A.2.2 derives the recursions for the surplus values, see specifically equations (A.16) and (A.19), so that the optimality conditions are fully described in terms of the values of unmatched agents and surpluses. A formal definition follows.

Definition 1. *A stationary search equilibrium is a tuple of value functions, $\big(V_u(x), V_{f.0}, S(x), S(x|x')\big)$, together with a distribution of agents across states, $\big(d_{m.1}(x), d_{m.2}(x,x')\big)$, such that (i) the value functions satisfy the HJB equations* (28), (29), (A.16) *and* (A.19) *given the distributions; and (ii) the stationary distributions satisfy the KFEs* (A.13)-(A.14) *given the policy functions implied by the value functions according to equations* (26)-(27).

The equilibrium needs to be computed numerically because of a non-trivial general-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.

EQUILIBRIUM OBJECTS. The model implies specific coworker sorting matching patterns as well as a decomposition of wage inequality into a between- and within-employer component. Consistent with the focus on multi-worker firms, and for the remainder of the paper, I consider only workers employed in a team to compute these moments. As a preliminary step, note that the joint distribution of worker types in teams is characterized by the density $\phi(x,x') = \frac{1}{e_2}d_{m.2}(x,x')$, where $e_2 = \int\int d_{m.2}(x,x')dxdx'$ is total employment in teams. Hence, the unconditional density of types in teams is $\phi(x) = \int \phi(x, \tilde{x}')d\tilde{x}'$, while the distribution of coworker types conditional on type is $\phi(x'|x) = \frac{\phi(x,x')}{\int \phi(x,\tilde{x}')d\tilde{x}'}$. The corresponding CDFs are denoted $\Phi(x)$ and $\Phi(x'|x)$.

One measure of sorting is the coworker correlation coefficient,

$$\rho_{xx} = \frac{\int \int (x - \bar{x})(x' - \bar{x})d\Phi(x'|x)d\Phi(x)}{\int \left(x - \bar{x}\right)^2 d\Phi(x)}, \tag{31}$$

where $\bar{x} = \int x d\Phi(x)$ is the average worker type among those in teams. Alternatively, providing a more disaggregated picture, the mapping between worker type and average coworker type is

$$\hat{\mu}(x) = \int \tilde{x}' d\Phi(\tilde{x}'|x), \tag{32}$$

which is the natural extension of a deterministic matching function to a stochastic setting.

Next, the average log wage and the variance of individual or mean firm-level (log) wages are

$$\ln \bar{w} = \int \int \ln w(x|x')d\Phi(x'|x)d\Phi(x), \tag{33}$$

$$\sigma_w^2 = \int \int \left(\ln w(x|x') - \ln \bar{w}\right)^2 d\Phi(x'|x)d\Phi(x), \tag{34}$$

$$\sigma_{\bar{w}}^2 = \int \int \left(\ln \bar{w}(x, x') - \ln \bar{w}\right)^2 \phi(x, x')dxdx', \tag{35}$$

where $\ln \bar{w}(x, x') = \frac{1}{2}\big(\ln w(x|x') + \ln w(x'|x)\big)$ denotes the average log wage in a firm with team (x, x'). The between-employer share of wage inequality is $\sigma_{\bar{w}}^2/\sigma_w^2$.

3.2.5 Elucidating the mechanism in a simplified environment

The key trade-off shaping within- and between-firm worker heterogeneity relates to the matching decision to be taken when a firm with a type-x employee meets an unmatched worker of type x'. This decision, encapsulated in policy function $h(x'|x)$, balances match quality considerations and search costs.[42] If the hire is made, output $f_2(x, x')$ is produced and shared. Else, the firm with its employee produces $f_1(x)$, 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 strength of coworker production complementarity determines to what extent team output is influenced by the match quality between x and x'. Suppose that specialization (χ)

[42]The decision of an unmatched firm and an unmatched worker is quite trivial, as the surplus is generally positive across potential matching combinations for plausible parameterization.

and, hence, coworker complementarity is pronounced, and consider a firm with a talented worker x and a less capable potential hire x'. The match surplus is low, because x's potentially high productivity would be dragged down by collaborating with x'. It is preferable for both sides to keep searching. Consequently, in a high-χ economy, only workers of similar talent are matched together. If, however, χ is low, then x can perform a greater share of tasks without this substantially lowering the average efficiency with which she performs these tasks, mismatch is less costly, and consequently the degree of coworker sorting is low.

To further clarify this mechanism, I briefly sketch and describe key findings from a stylized version of the model, set out in Appendix A.3. While not suited for quantitative analysis, it qualitatively preserves the same mechanism operative in the full model and the equilibrium can be characterized analytically. I adapt the setup of Eeckhout and Kircher (2011) to the context of team formation. In brief, this is a finite-horizon model and instead of search costs arising because agents discount the future, workers incur a fixed search cost if they reject a match and get paired with their optimal coworker type in return. In this environment, the matching decision obeys a simple threshold rule: Types x and x' will be teamed up if and only if the absolute type distance, $|x - x'|$, is lower than a threshold value s, which is decreasing in the strength of complementarity and increasing in the cost of search. I derive closed-form expressions for the conditional match distribution and firm-level average wages, yielding the following predictions.

Corollary 2 (Coworker complementarity and distributional outcomes)**.** *In the simplified model, a marginal increase in coworker complementarity lowers the equilibrium threshold s^*, other things equal, which leads to the following outcomes.*

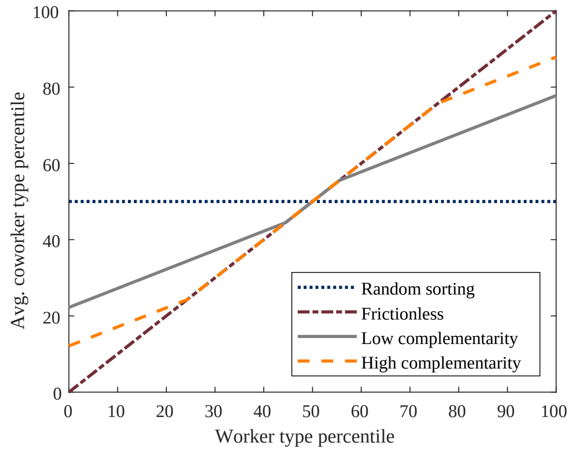
(i) *The correlation between coworkers' types is higher, specifically* $\rho_{xx} = (2s^* + 1)((s^*)^2 - 1)^2$.

(ii) *The average coworker type is lower for types of quality below s^*, greater for types above $1 - s^*$, and unchanged for intermediate types. Specifically, the expected coworker type is*

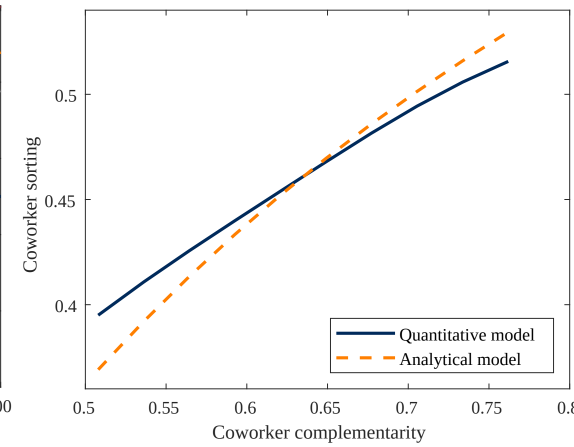
$$\hat{\mu}(x) = \begin{cases} \frac{x+s^*}{2} & \textit{for} \quad x \in [0, s^*) \\ x & \textit{for} \quad x \in [s^*, 1 - s^*] \\ \frac{1+x-s^*}{2} & \textit{for} \quad x \in (1 - s^*, 1]. \end{cases}$$

(iii) *The between-firm share of the variance of wages, $\sigma^2_{\bar{w}}/\sigma^2_w$, is higher (for a precise expression, see Appendix equation* (A.23)*).*

Figure 4a graphically illustrates the equilibrium matching outcomes under four alternative parameterizations. For any worker type, it plots the average coworker type, $\hat{\mu}(x)$. Consider first the dotted line, which represents the case of perfect substitutability between coworker



(a) Matching patterns in the stylized model



(b) Comparison to quantitative model

Figure 4: The strength of coworker complementarity shapes assortative matching

Notes. The left panel is based on the stylized model and plots, for each worker type x, the expected coworker type, $\hat{\mu}(x)$. Each line represents an alternative equilibrium assignment, which is associated with a threshold s^*. Specifically, "Random sorting" corresponds to $s^* = 1$; "frictionless" to $s^* = 0$; "Low complementarity" to $s^* = 0.45$; and "High complementarity" to $s^* = 0.25$. The right panel plots coworker sorting, as measured by ρ_{xx} against the elasticity of complementarity, γ. The solid line is derived from the quantitative model; the calibration of parameters other than γ is taken from Section 4.4; I allow for small match-specific shocks to smooth the plot (details are available upon request). The dashed line is again based on the stylized model, which is parametrized so that the values of ρ_{xx} implied by analytical and quantitative model coincide for the mid-point value of γ.

types. Then $\hat{\mu}(x)$ is the same for any x. Next, the dashed-dotted line illustrates that when there are complementarities and no search frictions, pure assortative matching obtains. The remaining two lines represent cases with both complementarities and search frictions. Under weak complementarities, search frictions mean that worker types below the threshold s^* are paired up with coworkers that are better than themselves by a margin of up to s^* (solid line). The opposite applies to high types, that is if $x > 1 - s^*$. For types in between those two kink points, the average coworker type corresponds to their own type, as the matching set is symmetric. Now consider the case with stronger complementarities and, hence, a lower value of s^* (dashed line). For an intermediate worker, the matching set shrinks symmetrically in both directions, but for the highest (lowest) types, stronger complementarities can only raise (lower) the minimum (maximum) type with whom they get matched in equilibrium, so that the *average* coworker type is higher (lower) than under weak complementarities.

The same mechanism that links complementarities and labor market sorting in the tractable, simplified model also operates in the quantitative model, as Figure 4b illustrates.

In summary, stronger complementarities imply an equilibrium which features firms with "superstar teams" composed of the best workers, on the one hand, and other firms with "laggard teams," on the other hand (cf. Andrews *et al.*, 2019; Autor *et al.*, 2020).[43]

[43]This result underscores the importance of analyzing input complementarities and input choice jointly. Following

4 Model Meets Data

In this section, I bring the model to the data, considering a single time period (2010-2017).[44] This process involves four steps. Section 4.1 maps key model objects to matched employer-employee panel data. Notably, I present a theoretical identification result that guides the quantification of coworker complementarity specifically. Section 4.2 uses this result to measure complementarities in the data and discusses estimates. Section 4.3 tests model predictions using cross-sectional variation (and relying on additional, Portuguese data). Lastly, Section 4.4 calibrates the structural model, with micro estimates of complementarity serving as a key source of empirical discipline.

4.1 Mapping model objects to data

For the model to deliver a meaningful, quantitative interpretation of the motivating evidence in Section 2, we need to take a stand on how theoretical model objects are connected to data. While straightforward for certain primitives (e.g., job separation rates) and outcomes (e.g., wage dispersion), measuring others is less straightforward. I next outline how the following objects are recovered: worker types; firms, teams and coworkers; specialization and complementarity.

4.1.1 Worker types

In the model, a worker's type x represents their time-invariant productivity (or absolute advantage) type. I adopt two established approaches to empirically approximate this concept.[45]

The first approach is the AKM-based method introduced in Section 2.2. Worker are ranked by their fixed effect, which yields an ordinal measure that corresponds to $x \in [0, 1]$. Furthermore, workers are binned by decile rank, denoted $\hat{x}_i$ for worker i. Binning aligns the empirical with the numerical analysis of the structural model, which is solved by discretizing the (continuous) type space using ten equidistant grid points, as well as reducing noise.

While the AKM approach is straightforward and popular, several of the underlying assumptions – including the log-linear functional form for wages – are inconsistent with the structural

Rosen (1981), greater *sub*-modularity may be associated with a "superstar" phenomenon, as it implies extra weight on, and hence reward for, the highest-quality input *within* a production unit of given input composition. The analysis here underlines that when inputs are chosen endogenously and subject to frictions, stronger *super*-modularity leads to some firms matching together the highest-quality inputs available, resulting in greater differences *between* firms.

[44] I turn to trends over time in the following Section 5.

[45] Throughout, the model is brought to the data using residualized wages, the construction of which was described in Section 2.1. While residualizing wages for observables aligns with the model's exclusion of life-cycle and on-the-job learning effects, it is not straightforward whether worker types should likewise be computed from residualized wages. In the baseline, I do so to maintain maximum consistency, both internally and with respect to the existing literature. It could be argued, though, 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." See Appendix B.4.7 for a robustness check.

model (e.g., Eeckhout and Kircher, 2011; Bonhomme *et al.*, 2019). Therefore, I also implement the non-parametric ranking algorithm proposed in Hagedorn *et al.* (2017), which imposes weaker assumptions that are satisfied in the structural model.[46] The correlation between the resulting two rankings is high (0.86). Implementation details are relegated to Appendices B.2.1 I treat the AKM-based measure as a baseline, following the majority of the literature, and report robustness checks in Appendix B.4.3.

One concern with either approach is that it any worker in a high-wage occupation will tend to be interpreted as having high absolute advantage, even if conditional on their occupation they are not particularly productive. Paralleling the approach taken in Section 2.2, as a robustness check I also consider an alternative ranking of workers within their respective occupation.

4.1.2 Firms, teams and coworker(s)

In the dynamic model, workers are sorted into teams with two members, while real-world production units typically have more than that. To align theory and data, I follow Herkenhoff *et al.* (2022) and construct a "representative coworker" type for each empirical worker-year observation, which yields a counterpart to x' in the model. This step implicitly interprets the theoretical notion of teams as actual firms. Though useful for tractability, this represents a significant simplification, so I discuss the method and robustness checks in some detail.

Worker i's representative coworker in year t is constructed as the unweighted leave-out mean type among employees in the same establishment-year cell. Recalling that $S_{-it} = \{k : j(kt) = j(it), k \neq i\}$ denotes the set of i's coworkers in year t, $j(it)$ being the identifier of i's period-t employer, the average type of i's coworkers in year t is $\hat{x}_{-it} = \frac{1}{|S_{-it}|} \sum_{k \in S_{-it}} \hat{x}_k$. Aligned with the discussion in the preceding section, I construct this object based on an economy-wide ranking of workers, as a baseline, but also consider a within-occupation ranking as a robustness check.

Three potential concerns with this approach relate to the definition of the coworker group; the averaging step; and firm size variation. Regarding the first of these, the baseline specification corresponds to what Jarosch *et al.* (2021) refer to as "Team Definition 1". One could alternatively argue for including only coworkers in the same occupation. I deliberately do not adopt this specification, as Corollary 1 implies that coworker complementarities arise precisely when workers are *differentiated* in their task-specific productivities, which is more likely across varying occupations contributing to the same output.

A second point involves the non-linearity in coworker aggregation implied by the structural model, which the equally-weighted average method neglects.[47] According to Proposition 1,

[46]Appendix A.2.4 derives a Lemma confirming that $w(x|x')$ is indeed monotonically increasing in x and, hence, Hagedorn *et al.*'s (2017) approach extends to the present environment.

[47]While aligning with existing studies of coworker interactions (including Cornelissen *et al.*, 2017; Jarosch *et al.*,

lower-type coworkers should have a higher aggregation weight insofar as $\chi > 0$. By taking an unweighted average, the baseline approach thus effectively ignores complementarities across an individuals' multiple coworkers (the focus here being on complementarities between that individual and her coworkers). For instance, having two "middle" level coworkers appears equivalent to having one "high" and one "low" level coworker. However, robustness checks using a non-linear averaging method indicate that the resulting bias is minor in magnitude (see Appendix B.4.1). Intuitively, bias is small because dispersion among coworkers is limited in the data, which itself is a manifestation of coworker sorting, which the structural model predicts to be high precisely when the non-linearity in aggregation would be significant. Hence, I adopt the simpler, linear aggregation method as a baseline, aligned with the existing literature.

A third worry is that the model's fixed team size could distort the empirical interpretation. Intuitively, a single colleague's exit or entry impacts a worker's productivity (and, hence, wage) less in larger teams. The collapsing step is consistent with this intuition: a single move will induce a smaller change in the *average* coworker quality in a large team than in a small one.[48]

Having ranked workers and collapsed multiple coworkers in the data into a representative coworker, we can link types and matching patterns in model and data. For example, using $\hat{x}_i$ and $\hat{x}_{-it}$, we can compute direct counterparts to the theoretical measures of labor market sorting defined in equations (31) and (32).

4.1.3 Specialization and coworker complementarity: a useful identification result

Next, I explain how the specialization parameter χ – respectively the elasticity of complementarity γ, per Corollary 1 – can be recovered from the micro data. I present and discuss a theoretical identification result; the empirical implementation and results follow in the next sub-section.

Considering the theoretical model, specifically Section 3.2, and treating each worker's type as known for present purposes, how can we measure the degree of coworker complementarity in production, and specifically $\frac{\partial^2 f(x,x')}{\partial x \partial x'}$?[49] Except in circumscribed contexts (e.g., Mas and Moretti, 2009; Adhvaryu *et al.*, 2020) we do not have measures of individual output, while firm output may be affected by many variables other than workforce quality.

2021; Herkenhoff *et al.*, 2022), this neglect is more concerning here given the distinctive focus on complementarities.

[48] A distinct issue that is related to firm size concerns the decomposition of the variance of log wages. The empirical computations are performed in the uncollapsed sample (cf. Section 2.2). In the model, where firm size is two, the between-component is mechanically overstated; it would be greater than zero even under random matching. This is simply an outgrowth of the law of large numbers not applying within production units (cf. Carrington and Troske, 1997). Throughout, I report a corrected decomposition, using a method explained in Appendix C.1.3.

[49] The cross-partial is very intuitively connected to the 'strength 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 direct determinant of the elasticity of complementarity. Considering equation (15), observe that $\gamma = (f f_{ij})/(f_i f_j)$ for any $i \neq j$, where subscripts indicate partial derivatives. This observation goes back, in fact, to Hicks (1932, 241-246).

The central, and novel, insight underpinning my quantification strategy is that coworker complementarity in *wages* is directly proportional to complementarity in *production*. To see this, note that using the surplus sharing rule (25) and after some algebra (cf. Appendix A.2.3), we can write the implied wage of x employed alongside x' as

$$w(x|x') = \omega\big(f(x,x') - f(x')\big) + (1-\omega)\rho V_u(x) - \omega(1-\omega)\lambda_{v.u}\int \frac{d_u(\tilde{x}'')}{u} S(\tilde{x}''|x')^+ d\tilde{x}''.$$

Intuitively, the worker is rewarded with an ω-share of the increase in production from x being added to x', plus a fraction of x's outside option, minus compensation to the firm and its other employee for their foregone share of a surplus from eventually hiring a different worker. Differentiating twice with respect to x and x', respectively, yields:

Corollary 3 (Measuring complementarities in theory)**.** *The strength of coworker complementarities in production is proportional to the strength of coworker complementarities in wages:*

$$\frac{\partial^2 f(x,x')}{\partial x \partial x'} = \frac{1}{\omega}\frac{\partial^2 w(x|x')}{\partial x \partial x'}. \tag{36}$$

The intuition is as follows. In a situation of bilateral monopoly, as it arises in the presence of search frictions, a fraction of the surplus accrues to the worker. The *level* of the wage is influenced by the bargaining partners' respective outside options. When we consider the cross-partial derivative – focusing on differential wage responsiveness to changes in coworker talent – these outside options drop out: that of employee x is foregone irrespective of whether she is matched with a bad or a good coworker; those of the employer and coworker are foregone irrespective of whether x or someone else is hired.[50] The curvature of surplus is shaped by the production function. If the surplus increases by some increment Δ, the worker x appropriates a fraction ω of that Δ-increase, and the magnitude of the associated wage change is informative about output changes due to the interaction between x and x'.

Corollary 3 indicates that we can use measurements of $\frac{\partial^2 w(x|x')}{\partial x \partial x'}$ to quantitatively discipline $\frac{\partial^2 f(x,x')}{\partial x \partial x'}$; χ will then be indirectly recovered using the model's full structure.[51] One attractive feature of this approach is that this moment can straightforwardly be estimated for different

[50] This argument extends to a setting with on-the-job search (OJS), where a worker's outside option in negotiation with a "poaching" employer may be tied to the surplus obtained with the current employer, as opposed to the value of unemployment. Appendix C.2.2 elaborates on the OJS case.

[51] A different avenue is to directly measure the within-worker dispersion in productivity across tasks – consistent with the idea that complementarity is a reduced-form object and ultimately pinned down by specialization. As argued in Section 2.3, however, cardinal measures of specialization are extremely rate, especially longitudinal ones covering the economy as a whole. Alternatively, one may wish to consider a model version that features two types of tasks, one associated with a low value of χ ("routine") and the other with a high value of χ ("complex"). Rising task complexity, as in Figure 2a, would then be represented by an increase in the share of the latter type, leading to a rise in the 'average' χ. Unfortunately, analytical tractability is lost with two types of tasks. Numerically, I have verified

periods, enabling a quantitative analysis of trends over time, as well as occupations or industries, facilitating tests of model mechanisms in the cross-section.[52]

4.2 Coworker complementarity in the data

This section empirically quantifies coworker complementarity.

METHODOLOGY. Conceptually, to measure how the slope of the wage function with respect to the coworker type varies with the own type, worker and coworker type measures are combined with the joint wage distribution. Practically, two routes are available. One is to construct a non-parametric wage function and compute the average cross-partial derivative using finite-difference methods (see Appendix B.2.2). This method is straightforward and imposes minimal assumptions on the shape of the wage function. A concern is that threats to identification are ignored that stem from features of the data that are absent from the theoretical model.[53]

An alternative method is regression based. It imposes more structure but allows addressing identification concerns more transparently. Specifically, I estimate

$$\frac{w_{it}}{\bar{w}_t} = \beta_0 + \beta_x \hat{x}_i + \beta_{x'} \hat{x}_{-it} + \beta_c \left(\hat{x}_i \times \hat{x}_{-it}\right) + \psi_{j(it)} + \nu_{o(i)t} + \xi_{s(i)t} + \epsilon_{it}, \tag{37}$$

using OLS, where $\psi_{j(it)}$ denotes employer fixed effects (FE), $\nu_{o(i)t}$ are occupation-year FEs, $\xi_{s(i)t}$ are industry-year FEs.[54] Weighting each observation by the inverse empirical frequency of the associated $(\hat{x}_i, \hat{x}_{-it})$ match ensures equal exploration of each part of the state space, including those under-sampled in equilibrium. If the true data-generating process for wages follows equation (37), then β_c measures the coworker wage complementarities of interest; differentiating the wage with respect to $\hat{x}_i$ and $\hat{x}_{-it}$ yields β_c. Thus, the coefficient of interest, β_c, answers the following question: How does the effect of having a better coworker vary with your own type? As $\hat{x}$ is treated as continuous for estimation, β_c specifically 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.

---

that the maximum likelihood estimate of the χ parameter of a univariate Fréchet when fitted to data generated from a mixture of a low-χ and a high-χ Fréchet is increasing in the weight of the latter. To further pursue this approach, though, one would require estimates of the (cardinal) χ values associated with routine and complex tasks.

[52]One limitation of Corollary 3 that ought to be kept in mind is that the mapping between wage and production complementarities is contingent on a particular value of the bargaining power parameter ω. I will externally calibrate ω and treat it as time-invariant going forward. Foreshadowing the discussion of time trends in Section 5, if workers' bargaining has declined over recent decades, however, as some studies argue is the case at least for the U.S. and France (Stansbury and Summers, 2020; Mengano, 2023), then any increase in wage complementarities would understate rising production complementarities.

[53]Appendix B.2.2 shows that both the non-parametric and the regression-based method are capable of approximating the true average cross-partial well when applied to data generated from the structural model.

[54]I also include squared terms in $\hat{x}_i$ and $\hat{x}_{-it}$ to address the concern that the interaction term picks up convexity in the return to own or coworker talent. Equation (37) omits these terms for sake of readability.

This regression method builds on the peer effects literature and existing analyses of coworker wage effects in three respects.[55] First, identifying variation comes from both "movers" (changes in coworker quality for individuals who switch employer) and "stayers" (changes in coworker quality induced by other employees joining or leaving the coworker group). Second, to sidestep the reflection problem (Manski, 1993), I use a pre-determined measure of coworker quality to isolate "exogenous peer effects," rather than estimating types and coworker effects in one step. Third, a rich set of fixed effects controls for unobserved time-invariant employer heterogeneity and shocks at the occupation-year or industry-year level. Controlling for the worker's own type furthermore accounts for selection of high types into coworker groups with a high average type.

Different from the usual peer effects literature, though, the objective here is to capture coworker wage *complementarity*, not simply wage *effects*. In regression (37), this is manifested in two ways. First, the focal point is the interaction term – and thus *heterogeneity* in coworker effects – contrary to peer-effect studies that usually target an average treatment effect.[56] Second, the dependent variable is the wage *level* (scaled by the average wage to aid interpretation and comparability), rather than the *log* wage.[57] This choice is significant as theory links complementarity (i.e., supermodularity) and sorting. If having a better coworker confers a positive but uniform to any individual, then this creates no incentive for positively assortative matching. What drives positive sorting is a differentially greater benefit from greater coworker quality for high types. Moreover, it is the absolute gain that matters, not the gain relative to one's own wage.[58]

RESULTS. Table 2 reports the estimates of β_c for the sample period 2010-2017 and under a variety of specifications. In column (1), no fixed effects are included; column (2) introduces employer FEs, absorbing variation due to time-invariant differences in pay-levels across different firms; and column (3) adds industry-year and occupation-year FEs, controlling for time-varying shocks at the industry- or occupation level that correlate with coworker types and wages. One can observe, however, that these additional controls turn out have only a limited influence on the magnitude of the point estimate of β_c.

Treating column (3) – which corresponds to the specification in equation (37) – as the baseline, we find that the point estimate of the coefficient β_c on the interaction term of interest is equal to

[55] For example, Cornelissen *et al.* (2017); Barth *et al.* (2018); Cardoso *et al.* (2018); Hong and Lattanzio (2022).

[56] Relative to the literature on coworker *learning* (Jarosch *et al.*, 2021; Herkenhoff *et al.*, 2022), there is an additional, salient difference: rather than considering the dynamic influence of coworkers, here I concentrate on contemporaneous spillovers.

[57] A common peer effect specification where the dependent variable is in logs imposes a constant (semi)-elasticity. If having better colleagues is beneficial, this is true for all types, with the absolute gain constrained to be greater for high types.

[58] For example, consider a situation with low and high types. Under matching pairings of (low, low) each produces 1, while under (high, high) each produces 10, while (low, high) yields 3 and 5. Here, the absolute gain from a high-type coworker is greater for the high type, even though the proportional gain is be larger for the low type. Production is maximized under PAM with an output equal to 22, compared to 16 under mixing.

0.0091, which is statistically significant at the 1% level. (The non-parametric, finite-difference based method yields a very similar estimate, equal to 0.0097.) To provide a sense of magnitude, this implies that the wage increase from a one decile improvement in the average coworker quality is 4.55% greater, as a percentage of the average wage, for a worker who is in the top decile compared to one in the fifth decile (0.0091·(10-5)·1=0.0455).

	(1)	(2)	(3)	(4)
$\hat{\beta}_c$	0.0095***	0.0094***	0.0091***	0.0058***
	(0.00051)	(0.00039)	(0.00035)	(0.00278)
Own type controls	Yes	Yes	Yes	Yes
Coworker type controls	Yes	Yes	Yes	Yes
Employer FEs	No	Yes	Yes	Yes
Industry-year FEs	No	No	Yes	Yes
Occupation-year FEs	No	No	Yes	Yes
Type ranking	Economy	Economy	Economy	Occupation
Obs. (100,000s)	4,410	4,410	4,410	4,410
Adj. R^2	0.736	0.803	0.811	0.800

Table 2: Regression evidence on coworker complementarity (2010-2017)

Notes. Regression results based on equation (37), 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. All regressions include a constant. "Own type controls" and "Coworker type controls" are quadratics in $\hat{x}_i$ and $\hat{x}_{-it}$. 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 100,000s. * p<0.1; ** p<0.05; *** p<0.01.

This finding is robust to the worry that $\hat{\beta}_c$ may be biased because the worker and coworker type measures appearing on the right-hand side of equation (37) are ultimately still recovered from wages, including the contemporaneous wage used here as the dependent variable. I address this concern with two different checks. First, Appendix B.4.5 considers education as a non-wage measure of worker types. In a regression of wages on fully interacted years of schooling and coworker schooling, the coefficient on the interaction term likewise indicates that complementarity is positive (this is true not only for 2010-2017 but for the entire sample). Second, I re-run regression (37) using worker and coworker types constructed from lagged periods, which rules out endogeneity due simultaneity. This approach, too, yields comparable results, with details reported in Appendix B.4.6.

Finally, column (4) indicates that strength of complementarities when workers are ranked within occupation rather than economy-wide. Under this alternative specification, $\hat{\beta}_c$ is equal to 0.0058, suggesting that part of the baseline coworker complementarity estimate may reflect a dimension of worker and coworker quality that is picked up by occupational choice. I quantitatively evaluate this alternative mapping between model types and data in Section 5.4.

4.3 Validation of model mechanisms in the cross-section

Prior to quantifying the model, I use cross-sectional variation to scrutinize two central, model-predicted relationships: that coworker complementarities are stronger when specialization is more marked; and that complementarities promote positively assortative coworker matching.

For these analyses, I supplement the SIEED data with a second micro dataset, the Portuguese Quadros de Pessoal/Relatório Único (QP) matched employer-employee panel. All results reported for Portugal are drawn from joint work with Criscuolo and Gal (Criscuolo *et al.*, 2023).[59] While the SIEED could also be used here, there are non-trivial limitations for present purposes, notably a loss of statistical power when parsing the data more granularly and top-coding of wages.[60] The QP, meanwhile, is an annual mandatory census of all employers in Portugal and wages are not top-coded (see Appendix B.1.2 for a more detailed description). I again use the 2010-2017 subsample, which also includes standardized ISCO occupation codes and detailed, NAICS 4-digit industry codes. I examine variation across occupations and industries, while acknowledging two caveats. First, the analysis establishes correlations that are consistent with the structural model but cannot speak to causality. Second, as the model does not explicitly incorporate occupations or industries, each data cell is effectively interpreted as a separate instantiation of the model.

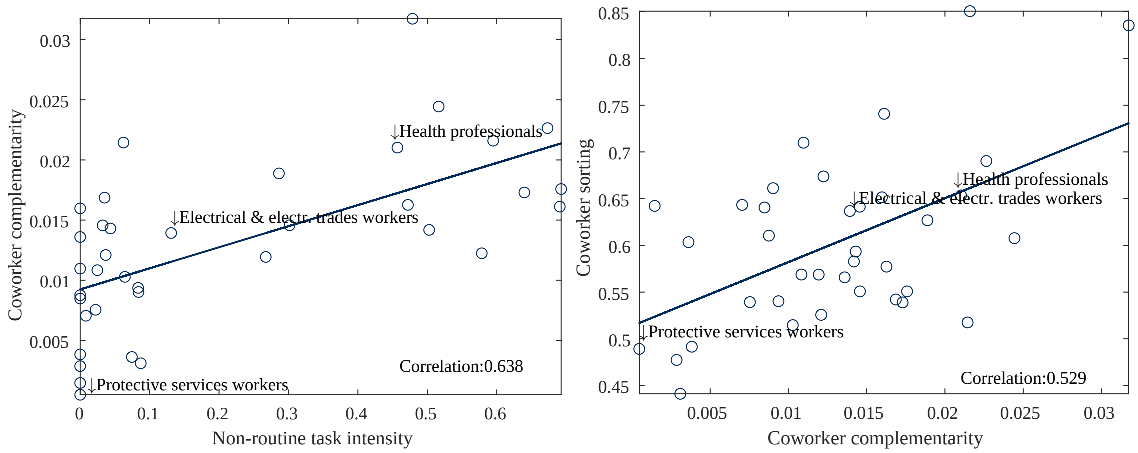
OCCUPATIONS. I begin by analyzing tasks, complementarities, and sorting at the occupation level. Each ISCO-08 2-digit occupation is classified by its reliance on complex (i.e., non-routine, abstract) tasks, relying on task indices from Mihaylov and Tijdens (2019).[61] Sorting and complementarity measures are then constructed for each occupation. For these measures to be independent across occupations, I use the within-occupation ranking of workers and, additionally, restrict the set of coworkers to those in the same employer-occupation-year cell ("Team definition 2" in Jarosch *et al.* (2021)). Coworker complementarity is computed by estimating regression (37) separately for each occupation.

The left-hand panel in Figure 5 plots the estimated, occupation-specific estimate of coworker

[59] I thank Chiara Criscuolo and Peter Gal, and the OECD's Directorate for Science, Technology, and Innovation (STI) more broadly, for facilitating these analyses and the usage of the results in this paper.

[60] Why not use Portugal as the primary data source? One reason is that I want to connect my analyses to previous results in the literature, which have often treated Germany, the largest European economy, as a reference case. Portugal has, furthermore, experienced several idiosyncratic macroeconomic shocks over the past three decades.

[61] Mihaylov and Tijdens's (2019) index attractive, as it is constructed from occupation-specific lists of tasks provided by ISCO-08. Therefore, no selection among many alternative task scales is involved, which can be controversial, as discussed by Acemoglu and Autor (2011). Moreover, the ISCO-08 codes are consistent with the Portuguese classification for the years in question, whereas the task complexity measure discussed in Section 2.3 is based on a Germany-specific classification, which is difficult to map to ISCO codes.



(a) Tasks and complementarities

(b) Complementarities and sorting

Figure 5: Occupation-level relationships between tasks, complementarities, and coworker sorting

Notes. The left panel plots, for each ISCO-08 2-digit occupation, the occupation-specific point estimate for the coefficient on the interaction term, β_c – obtained from regression (37) when estimated separately for each occupation – against the non-routine task intensity of that occupation. The right panel plots the occupation-level coworker correlation coefficient, i.e., the correlation between a worker's type and the average coworker type against β_c, which is now shown on the horizontal axis. The linear regression line is fitted based on unweighted observations.

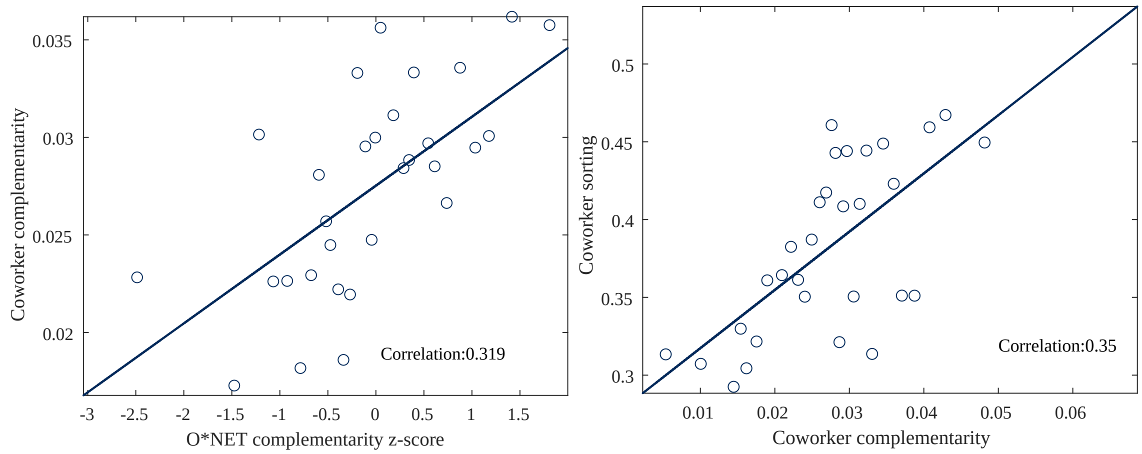
wage complementarity against the NRA score. We observe a positive relationship, consistent with the theory. The estimated interaction coefficient is close to zero for those occupations with the lowest NRA score, whereas it is above 0.02 in occupations that primarily perform non-routine abstract tasks. Next, the right panel shows that workers in occupations featuring greater complementarity are also matched together in a more positively assortative pattern.

These findings are also consistent with previous literature studying broadly related phenomena. Neffke (2019) studies Swedish micro data and finds that coworker effects are important in many knowledge-intensive jobs (e.g., health care, engineering) and professional occupations (e.g., lawyers); and in skill-intensive (e.g., R&D) an crafts-based industries (e.g., construction). Jäger and Heining (2022) find that when a high-skilled or specialized worker dies, their coworkers in other occupations experience wage decreases.[62]

INDUSTRIES. Turning to an industry-level analysis, I classify each 4-digit NACE industry according to a proxy for complementarity proposed by Bombardini *et al.* (2012). The basic idea is that an industry is more likely to feature complementarity if many workers are in occupations for which the following four characteristics are important: teamwork, impact on coworker output, communication, and contact. I operationalize this measure using O*NET data alongside industry-level occupational employment weights from QP.[63] Similar to the approach adopted for

[62]Appendix B.4.2 provides supplementary evidence using variation across seven different, vertical layers of the firm, grouped by similarity in the complexity of tasks and skills required, in the spirit of Caliendo *et al.* (2020).

[63]Clearly, we could characterize industries also on the basis of occupational task indices, as considered before. I



(a) Tasks and complementarities **(b)** Complementarities and sorting

Figure 6: Industry-level relationships between tasks, complementarities, and coworker sorting

Notes. The left panel depicts a binscatter plot of the point estimate for the coefficient on the interaction term, β_c, estimated at the industry-level, against an industry-level proxy for complementarity. This proxy is constructed from O*NET data following Bombardini *et al.* (2012). The right panel is an industry-level binscatter plot of the coworker correlation coefficient against β_c, which is now shown on the horizontal axis. The underlying unit of observation is a 4-digit NACE industry, with at least 5,000 person-year observations and where the industry-level proxy for complementarity is within two standard deviations of the mean. Observations are grouped into 30 bins. The linear regression line is fitted based on unweighted observations and calculated prior to binning.

occupations, I also compute coworker complementarity and sorting separately for each industry.

The results cohere with the model's predictions. The binscatter plots in Figure 6 shows that the higher an industry scores on the proxy measure, the greater tends to be the estimated value of complementarity, which in turn is positively associated with coworker sorting. Appendix B.4.2 furthermore shows that measures of between-firm inequality in productivity and wages are increasing in the degree of coworker sorting.[64]

In summary, cross-sectional variation corroborates the model's core mechanisms.

4.4 Model calibration & quantitative validation

I now turn to the calibration of the quantitative model. I start by describing how structural parameters are disciplined through a combination of micro- and macro-moments, including the complementarity estimates obtained in Section 4.2. I then discuss the estimation results.

I emphasize that in this approach, the key distributional moments of interest – the degree of

adopt Bombardini *et al.*'s (2012) approach as a baseline in order to test the theory against a variety of proxies for χ, including ones used in the extant literature for related purposes.

[64] Reassuringly, the relationship between coworker sorting and the between-firm wage of the variance of log wages is almost linear with a slope equal to one. That is, the industries with the highest measure of coworker sorting (a correlation of around 0.55) have, on average, a between-firm share of the variance of log wages equal to 0.55; industries featuring a coworker sorting coefficient around 0.15 have a between-share of close to 0.15.

coworker sorting and the between-firm share of wage inequality – are left untargeted. Instead, the point estimates of coworker wage complementarity at the micro-level serve to discipline the elasticity of complementarity, γ, and thus indirectly the specialization parameter χ. The macro-distributional moments then serve as yardsticks to assess the model's performance to explain within- and between-firm heterogeneity in workforce quality composition and pay.

4.4.1 Methodology

Four parameters are externally calibrated, the job separation rate is set to directly target an empirical moment, the remaining five are structurally estimated by indirect inference. I set a unit interval of time to be one month, and since the model is solved in continuous time I can construct correctly time-aggregated measures at any desired frequency.[65] Empirical moments are constructed as averages across the years 2010-2017.

DISTRIBUTIONAL AND FUNCTIONAL FORM ASSUMPTIONS. The functional form of the production function is directly informed by the microfoundations set out in Section 3.1. For quantitative purposes, I introduce two generalizations. First, I introduce a constant factor, a_0, such that even the lowest-ranked employee produces a strictly positive amount of output. Including a_0 is motivated by the observation that in the model even the least productive worker searches for employment, which is likely not true in the data, where those individuals would be out of the labor force. One can interpret the intercept term a_0 as standing in for this selection margin. Hence, a firm with a single worker x produces $f(x) = a_0 + a_1 x$. Second, the degree to which labor productivity is greater when working in teams than working alone is controlled by a parameter a_2, so that team output is[66]

$$f(x, x') = 2a_2\left[a_0 + a_1\left(\frac{1}{2}(x)^{1-\gamma} + \frac{1}{2}(x')^{1-\gamma}\right)^{\frac{1}{1-\gamma}}\right].$$

Next, the flow value of unemployment is proportional to output produced alone, $b(x) = \bar{b} \times f(x)$.

PRESET SET PARAMETERS OR ESTIMATED OFFLINE. The following three parameters are externally calibrated. Regarding preferences and bargaining, I follow Herkenhoff *et al.* (2022). 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

[65]For example, if the empirical average probability that an employed worker loses their job during a month is d, then I recover the Poisson rate δ at which such a shock arrives as $d = 1 - e^{-\delta}$.

[66]I have experimented with a specification that allows a_2 to endogenously vary with χ, as implied by Proposition 1. However, as the returns to scale are decreasing in team size, the assumption that teams are of maximum size two which was made for tractability reasons, could lead to an overstatement of this benefit. I therefore opt to instead treat a_2 as exogenous in order to cleanly isolate the effect of χ via complementarities.

$\omega = 0.50$ implies equal sharing of surplus. Third, as explained in Section 3.1, the production parameter a_2 has to be greater than unity, else it would never be beneficial to work in teams, as complementarities mean that the least-capable team member disproportionately determines output. I set a_2 equal to 1.1, though the exact value turns out not to be crucial.[67]

The exogenous separation rate, δ, can be directly chosen to target the empirically monthly job losing rate, which across 2010-2017 equals 0.0085 on average. The latter moment is computed by supplementing the SIEED with the Linked Employer Employee Data (LIAB), which unlike the SIEED contains information on non-employment spells (see 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.

Before describing the moments used, two comments are in order. First, the model moments are obtained from the numerically computed stationary equilibrium, found using a fixed point algorithm after discretizing worker types into ten equally spaced grid points. For example, the variance of log wages is recovered using equation (34). Second, while the parameters in ψ are jointly estimated, each is closely informed by one of these moments, as explained next. To validate my approach, and following Bilal *et al.* (2022), Appendix C.1.2 conducts two exercises that jointly support the notion that the vector ψ and each of its elements are well-identified.

The most important moment is the coworker wage complementarity, which following Corollary 3 directly informs the strength of production complementarity. I estimate an analogous regression to equation (37) inside the model and target the empirically estimated point estimate of the interaction coefficient β_c, which for 2010-2017 is equal to 0.0091 (see column (3) of Table 2). Next, the values of a_0 and a_1 are guided, respectively, by the average wage, which I normalize to unity, and the total variance of log wages, which is equal to 0.241. 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 in Germany 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. This value is somewhat lower than the replacement rate used

[67] This means that the average labor productivity in a team with two workers x and x' is 10% greater than that of a worker producing alone if that worker's rank equalled the χ-weighted power mean of x and x'.

Parameter	Description	Value	Source
ρ	Discount rate	0.008	External
ω	Worker bargaining weight	0.50	External
a_2	Team benefit	1.1	External
δ	Separation rate	0.0085	Offline estimation
γ	Elasticity of complementarity	0.837	Internal estimation
a_0	Production constant	0.239	Internal estimation
a_1	Production quality scale	1.557	Internal estimation
b_1	Unemployment flow scale	0.664	Internal estimation
λ_u	Meeting rate	0.230	Internal estimation

Table 3: Model parameters - baseline (2010-2017)

Notes. This table summarizes the calibration of the model for the period 2010-2017. Parameter estimates are expressed at a monthly frequency, when applicable. Note that $\gamma = \frac{\chi}{1+\chi}$ given Corollary 1.

by Jarosch (2023), for instance, since my calibration targets moments after the Hartz reforms of the labor market undertaken in the early to mid-2000s, which reduced the generosity of the unemployment insurance system. Finally, λ_u targets a monthly job finding rate of unemployed workers equal to 16.2%, again computed from the LIAB.[68]

4.4.2 Results

Table 3 summarizes the baseline parameterization of the model, including the internally estimated parameters. The key parameter γ is inferred to be 0.84. This corresponds to χ equal to 5.15.[69] Table 4 summarizes the model fit. It indicates that the model is capable of matching the targeted moments perfectly, ψ being exact-identified.

I now turn to an evaluation of the model in terms of untargeted moments. In terms of matching patterns, the theoretical coworker correlation indicates substantial, positive coworker sorting, at a magnitude only slightly below what is observed in the data (0.53 vs. 0.62). Noting that the only moment which informs the degree of coworker sorting in the estimation is the micro estimate of coworker wage complementarity, I view this as a success for the model. Providing a more disaggregated picture, Figure 7 plots, for each type, the decile of the average coworker

[68]This value is higher than that reported by, e.g., Jarosch (2023). This is unsurprising, as I compute moments in matched employer-employee data rather than a sample representative of both the employed and the non-employed. As the other moments likewise derive from employer-employee data, the estimation is internally consistent.

[69]Recurring to Footnote 51, which discussed an alternative calibration strategy using direct measurements of χ, I have experimented with fitting a (truncated) Fréchet to Grigsby's (2023) non-parametric estimates of the multi-dimensional skill distribution estimated from CPS data for the U.S. in the 2000s. With the caveat that adopting this approach leads to extremely noisy estimates of χ, that the data are for a different country, and that the concept of task-specific skills is coarser than may be appropriate for present purposes, it is reassuring that the recovered value of χ is in the same ballpark as recovered using my baseline calibration strategy.

Parameter	Targeted moment	2010-2017 Value	m	$\hat{m}$	1985-1992 Value	m	$\hat{m}$
γ	$\hat{\beta}_c$	0.837	0.0091	0.0091	0.434	0.0035	0.0035
a_0	Avg. wage (norm.)	0.239	1	1	0.378	1	1
a_1	Var. log wage	1.557	0.241	0.241	1.216	0.143	0.143
b_1	Replacement rate	0.664	0.63	0.63	0.740	0.69	0.69
λ_u	Job finding rate	0.230	0.162	0.162	0.168	0.141	0.141
δ	Job loss rate	0.008	0.008	0.008	0.007	0.007	0.007

Table 4: Estimated parameters and targeted moments

Notes. This table lists for each of the estimated parameters, the targeted moment, the estimated value, and moment values in data ($\hat{m}$) and model (m). Each is shown separately for the years 2010-2017 and 1985-1992. The calibration for the latter sample period is discussed in Section 5.2.1. Parameter estimates are expressed at a monthly frequency, when applicable. Note that $\gamma = \frac{\chi}{1+\chi}$ given Corollary 1.

type, according to both model and data. The model successfully replicates the upward slope, but slightly underestimates the quality of coworkers at both bottom and top. Section 5.4 will show that incorporating on-the-job search improves the model fit in this respect.

What about between- and within-firm wage inequality, the key untargeted moment(s) of interest? In the next section, we dissect the model-implied decomposition of wage dispersion into between- and within-firm components. In brief, this analysis reveals that the stationary equilibrium of the theoretical model fits the data very well along this dimension. That is despite assuming that there is no permanent, ex-ante heterogeneity across firms. Insofar as there are differences across firms in the stationary equilibrium, they emerge purely due to ex-post heterogeneity in firms' workforce composition; they do not reflect any intrinsic feature of firms.

The quantitative model also generates substantial ex-post dispersion in firm productivity. As a reference point in terms of broad magnitudes, Syverson (2004) reports an average 90:10 labor productivity ratio of 4:1. The model-implied counterpart is 3.14:1, going quite some way toward the data even without assuming ex-ante heterogeneity among firms.

As an additional validation exercise, in Appendix C.2.4 I empirically test the prediction of the model when extended to allow for on-the-job search (and re-calibrated accordingly). In the data, job-to-job transitions reallocate workers from less to more positively assortative matches, which is consistent with the theoretical model in the presence of complementarities.

Overall, the model's predictions are not only qualitatively born out in the cross-section; the calibrated version also quantitatively reproduces key features of the German economy in the 2010s quite well, in spite of its parsimonious structure.

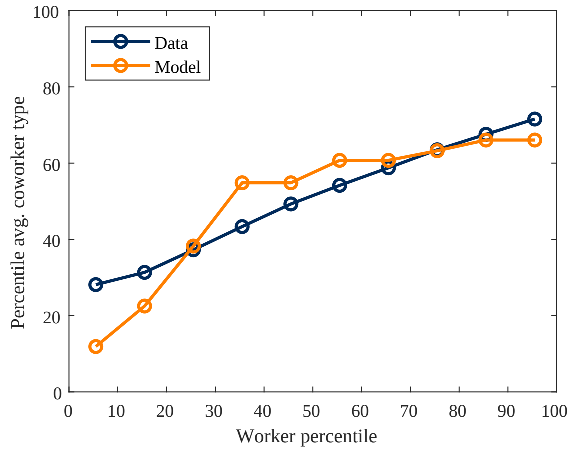


Figure 7: Average coworker type by worker type: model vs. data

Notes. This figure plots, for each type, the average coworker type, in both model and data. The data plot is for the year 2017, repeated from Figure 1b but using ten bins for comparability to the model. The theoretical counterpart is computed using equation (32).

5 Model-based Analysis of Trends Over Time

I now proceed to the main exercises in the paper, combining theory and micro data to examine the evolution of labor market inequality over time. Section 5.1 empirically assesses the model's prediction that the rise in specialization, outlined in Section 2.3, implies a strengthening of complementarities. Section 5.2 uses the time-series data on complementaries together with the structural model to examine trends in talent sorting and wage dispersion. Section 5.3 explores the implications of these trends for allocative efficiency. Section 5.4 offers a discussion and presents robustness checks and extensions.

5.1 The empirical evolution of coworker complementarity

A central trend motivating this paper is the intensification of specialization since the mid-1980s. Based on this trend, Proposition 1 predicts a strengthening of coworker complementarity: the gains from better coworkers should have become differentially greater for more talented individuals. The empirical strategy for measuring these complementarities, outlined in Section 4.2, allows for a test of this prediction.

To investigate this, I separately estimate equation (37) for five sample periods. Figure 8 displays the resulting point estimate for the strength of complementarity – that is, the interaction coefficient, $\hat{\beta}_c$ – alongside 95% confidence intervals. The task-complexity proxy for specialization, described in Section 2.3, is also overlaid and represented as circles.

This analysis reveals that coworker complementarity has indeed strengthened over time. Specifically, $\hat{\beta}_c$ more than doubled between the years 1985-1992 and 2010-2017, rising from

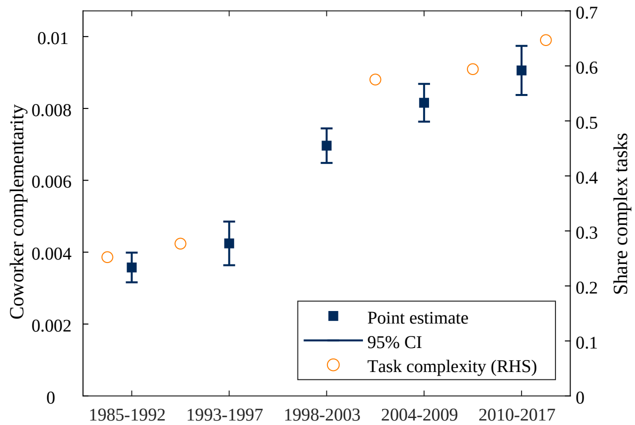


Figure 8: Coworker complementarity has increased alongside task complexity

Notes. This figure reports the point estimate and confidence intervals for the coefficient β_c in regression (37), separately for five sample periods. Standard errors are clustered at the employer level. In addition, the circles reproduce the share of complex tasks in workers' activities from Figure 2a. 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.

0.0036 to 0.0091. The estimate of coworker complementarity and the task-complexity for χ co-move closely, providing suggestive evidence in support of the model's predictions.

This finding is robust to several checks, which were introduced in Section 4.2. Appendix B.4.4 provides estimates for complementarity when workers are ranked within occupation. Under this specification, the strength of complementarities is lower across all periods, but shows a similar proportionate increase, with $\hat{\beta}_c$ rising from 0.0022 to 0.0058. Similar patterns emerge also when worker types are (i) measured by years of schooling, as a non-wage based measure of worker quality; (ii) recovered non-parametrically using the ranking algorithm of Hagedorn *et al.* (2017); or (iii) estimated from information on *past* wages.

5.2 A model-based perspective on the "firming up" of inequality

To assess the macroeconomic impact of strengthening complementarities, I first re-calibrate the model for the period 1985-1992, which facilitates a study of the model's predictions for changes over time.

5.2.1 Model calibration for 1985-1992

Using the time-series evidence on coworker complementarities, I re-estimate the vector of parameter ψ and the separation rate by targeting the same empirical moments as for 2010-2017 but measured over the years 1985-1992.

Regarding the targeted moments for 1985-1992, two aspects warrant comments. First, con-

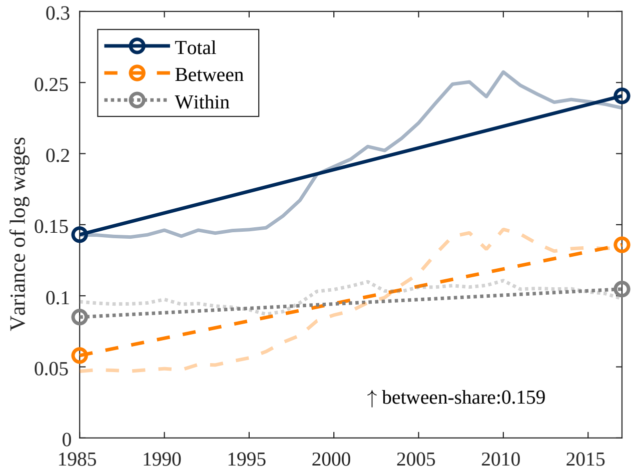


Figure 9: Decomposition of the variance of log wages: model vs. data

Notes. The solid lines indicate the model-predicted variance of log wages, decomposed into between- and within-employer components and computed using equations (34) and (35). The model-generated between-within decomposition is corrected for a mechanical bias, as described in Appendix C.1.3. The semi-transparent lines reflect the data moments, computed from the same residualized wages as used in the calibration of the model.

sidering 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 0.72. Second, labor market transition rates are also re-estimated for 1985-1992, to gauge the impact of changes in search frictions on the evolution of sorting and wage inequality.

The final main column in Table 4 presents the estimated parameter values. The elasticity of complementarity is lower than the estimate obtained for the 2010s (0.43 vs. 0.84). In addition, the uptick in a_1 and the dip in a_0 over time are consistent with technological change that made output more sensitive to talent. Increases over time in both job arrival and job separation rates may signal the emergence of online job portals (Bhuller *et al.*, 2023).

With the model calibrated for two time spans, we can assess its predictions for the evolution of labor market sorting and firm-level wage inequality. This sets the stage to quantify the role of amplified coworker complementarities to the "firming up" of inequality in Germany.

5.2.2 Analysis

Figure 9 depicts the model-implied decomposition of the variance of log wages, computed at the parameters for 1985-1992 and 2010-2017, respectively. The structural model qualitatively replicates the empirically recorded evolution of between- and within-firm inequality. In quantitative terms – and even though only the *total* variance of log wages had been targeted – the model captures 68% of the empirically recorded increase in the between-firm share (+0.159 vs. +0.233).

What portion of the model predicted rise in the between-firm share of wage inequality

	Δ model	Implied % Δ model due to Δ parameter
Model 1: baseline	0.159	-
Cf. a: fix period-1 complementarity	0.065	59.23
Cf. b: fix period-1 search frictions	0.150	5.88
Model 2: within-occ. ranking	0.198	-
Cf. a: fix period-1 complementarity	0.076	61.47
Cf. b: fix period-1 search frictions	0.189	4.43
Model 3: OJS	0.096	-
Cf. a: OJS, fix period-1 complementarity	0.067	30.54
Cf. b: OJS, fix period-1 search frictions	0.067	30.45

Table 5: 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 $100 \times \left(1 - \frac{m_j(\gamma^{p1}, \psi_{-1}^{p2}) - m_j(\psi^{p1})}{m_j(\psi^{p2}) - m_j(\psi^{p1})}\right)$, where γ^{pt} is the value of γ estimated for period pt, $t \in \{1, 2\}$, and ψ_{-1}^{pt} collects all other parameters.

can be attributed to stronger coworker complementarities? The answer is not obvious ex-ante 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, as Song *et al.* (2019) point out, 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 already high-earning individuals, who cluster together.[70]

To quantify the contribution of coworker complementarities, I use the structural model to conduct counterfactual exercises. The question posed is: 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.

The model indicates that the increase in coworker complementarity is an important driver of elevated firm-level wage inequality. Row 2 in Table 5 summarizes the analysis. Absent a rise in χ, the model predicts, the between-share would have risen only by 7 percentage points rather than 15.9. Thus, heightened coworker complementarities can account for about 59% of the model-predicted rise in the between-firm share of wage inequality, translating into 40% of the

[70]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.

empirically observed change.

Shifts in the production technology represent a more important factor than changes in labor market 'search technology', though both are consequential. I gauge the relative importance of changed complementarities by comparing their effect to that of the estimated changes in labor market transition rates. Specifically, I consider the 2010-2017 parametrization except counterfactually 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 6% of the model-predicted rise in the between-share.[71]

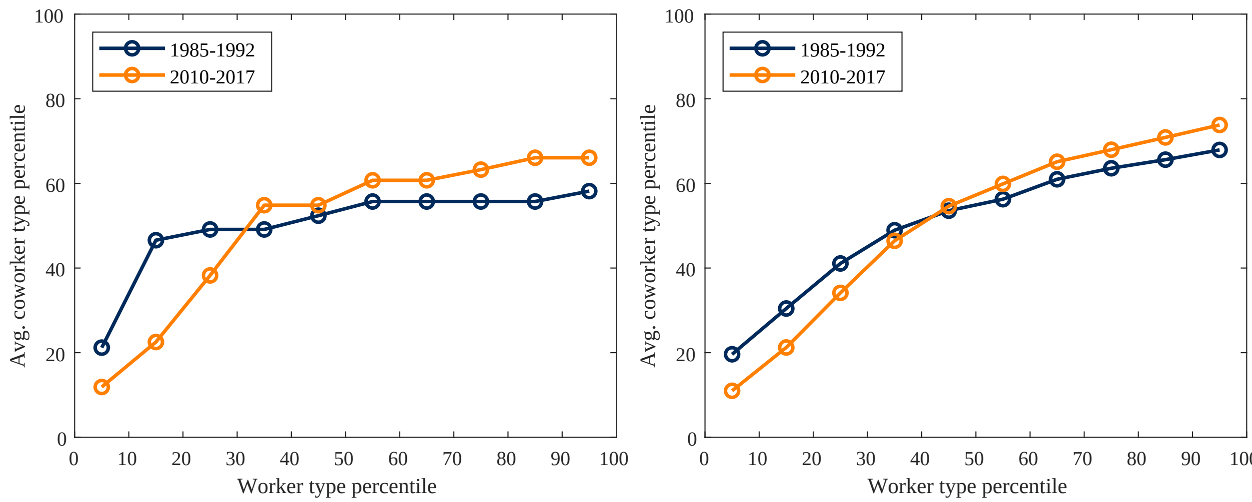
The pivotal driver of the increased between-firm share of wage inequality in the model is more positively assortative matching, aligning with the empirical evidence documented in Section 2.2.[72] Thus, the coworker correlation coefficient rises from 0.24 to 0.522. In the model, stronger complementarity signifies that mismatch is more costly. Hence, the market price mechanism pushes towards greater sorting; the wage that a firm with a less talented employee can offer to a talented potential hire lies farther below that affordable to a firm with another talented worker, since the new hire will be considerably more productive when paired with a peer of similar talent. Figure 10a depicts this shift in sorting in the model. The upward twist in the mapping between worker type and average coworker type echoes the data patterns presented in Figure 1b: High types increasingly pair up amongst themselves, as do lower types.

5.3 Coworker complementarities, sorting, and TFP

How does the interaction between specialization, complementarities, and labor market sorting affect aggregate productivity? While specialization may directly boosts productivity via the division of labor, it also enhances coworker complementarity, making team productivity more sensitive to the least capable team member's ability. Therefore, search frictions that generate within-firm heterogeneity in worker talent ("coworker mismatch") lower allocative efficiency more when specialization, and hence complementarity, is strong. This section uses the estimated model to quantify this mechanism: How large is the deviation from potential output due to coworker mismatch in the German economy in the 2010s, and how has it changed over time?

[71]The limited importance of 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 production complementarities, by rendering mismatch more costly, itself would 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.

[72]Although the model does not admit a straightforward decomposition of inequality into prices and distributions given their complex equilibrium interplay, we can compute total dispersion and the between-firm share from equations (34) and (35), imposing the wage function $w(x|x')$ implied by the 1985-1992 parameterization but the distribution $\phi(x, x')$ associated with 2010-2017. The implied difference from the 1985-1992 between-share accounts for almost the total change over time. Moreover, it is the change in the conditional coworker distribution, i.e., sorting patterns, that drives the rise in the between-share, as opposed to composition changes among the employed.



(a) Baseline

(b) OJS

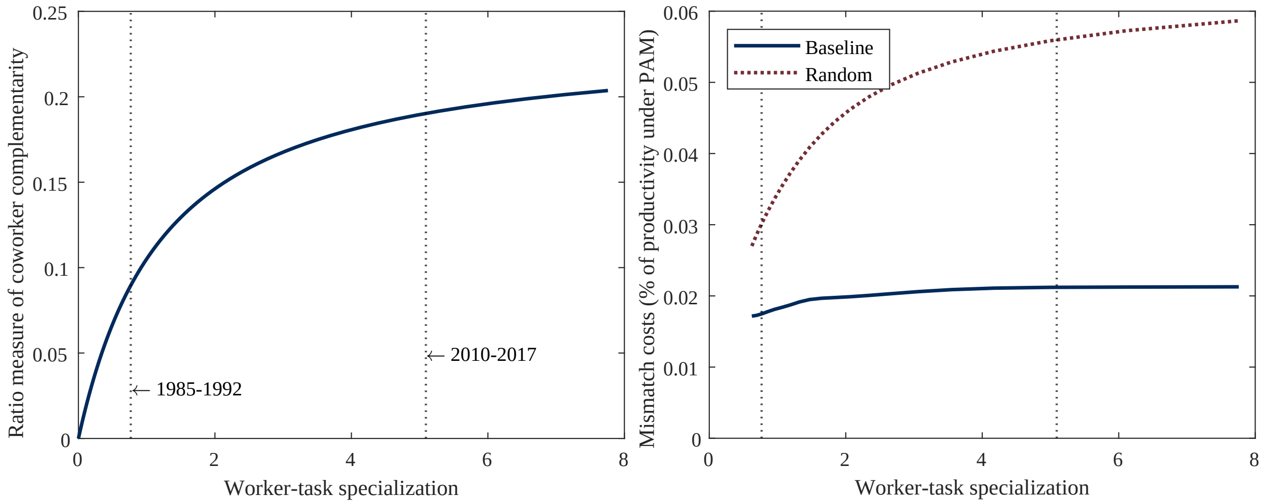
Figure 10: Model-implied coworker sorting across time

Notes. Both panels plot the average coworker type (vertical axis) for each worker type (horizontal axis), separately for the period-1 and period-2 parameterization, computed using equation (32). The left panel is computed from the baseline model, the right panel considers the model extended to incorporate on-the-job search.

To quantify equilibrium mismatch costs, I calculate the difference in average labor productivity gap between the equilibrium allocation and a counterfactual scenario of perfectly assortative matching (PAM), which maximizes productivity for any value $\chi > 0$. Changes in coworker mismatch are isolated by holding the marginal distribution of worker types employed in teams constant. Counterfactual output per worker under PAM is then equal to $\frac{1}{2}\int f(x,x)\phi(x)dx$, where $\phi(x)$ represents the unconditional density of types in teams.

Eliminating coworker mismatch would raise per capita output by 2.17% in 2010-2017. In 1985-1992, when complementarities were weaker, the equilibrium costs of mismatch are slightly lower at 1.78%. Through the lens of the team production function, this increase in mismatch costs reflects more inefficient task assignment. A talented worker ends up performing tasks at which she is relatively unproductive when paired up with a less capable colleague. Deeper specialization amplifies this inefficiency. Figure 11a illustrates this point by plotting a simple measure of the strength of production complementarities against χ. This measure compares the output produced by two teams under perfect assortative matching ("high-high" and "low-low") to the output produced by the same types under mixing of types (twice "high-low"). Keeping all other parameters fixed, this ratio approximately doubles from 1.09 in 1985-1992 to 1.19 in 2010-2017.

Why have equilibrium coworker mismatch costs not risen to a similar extent? The answer lies in the rise in coworker sorting, which signifies a diminishing of the gap between factual and efficient matching distribution. Using the estimated production function and the unconditional



(a) Specialization and coworker mismatch costs

(b) The role of labor market frictions

Figure 11: Worker-task specialization, complementarity, and mismatch costs

Notes. This figure depicts the productivity implications of worker-task specialization (χ), measured on the horizontal axis, through coworker complementarity. In the left-hand plot, the vertical axis indicates $\frac{f(x^{p80},x^{p80})+f(x^{p20},x^{p20})}{f(x^{p80},x^{p20})+f(x^{p80},x^{p20})} - 1$. The right hand panel plots, for any value of χ, the cost of mismatch as a fraction of productivity under perfect sorting, under the realized allocation (solid line) and under random sorting (dotted line).

distribution of types in teams for 2010-2017 but applying the 1985-1992 conditional coworker type distribution, productivity is 2.58% lower than with the actual conditional type distribution for 2010-2017.[73] Thus, the gap between efficient and realized labor productivity would have been more than twice as large absent the rise in coworker sorting. In short, the model highlights that widened between-firm gaps in workforce talent and ensuing differences in average pay need not reflect worsening frictions such as product market power.[74] Here, they are equilibrium outcomes that limit productivity losses due to greater complementarity, in effect keeping the economy closer to the productivity frontier.

The model indicates that a labor market which efficiently allocates workers of heterogeneous quality into teams becomes more critical for the economy to perform at its productive potential as specialization intensifies. Figure 11b illustrates this relationship by plotting the percentage gap between realized and efficient allocation in terms of labor productivity against varying levels of χ. The solid line indicates the gap under the baseline parameterization, where all parameters other than χ are kept at their 2010s levels. Consistent with the preceding analysis, this gap widens only marginally for higher values of χ. In contrast, the dotted line indicates

[73] Counterfactual average output per worker is $\frac{1}{2}\int f(x,x')d\Phi^{p1}(x'|x)d\Phi^{p2}(x)$. This approach avoids confounding changes in the unconditional distribution of types with shifts in conditional matching patterns.

[74] The model does not incorporate product market frictions, nor other features that render increased coworker sorting inefficient. The model shows, though, that such features are not necessarily required to explain the empirically observed trends. Hence, any policy response needs to carefully evaluate the underlying mechanism(s).

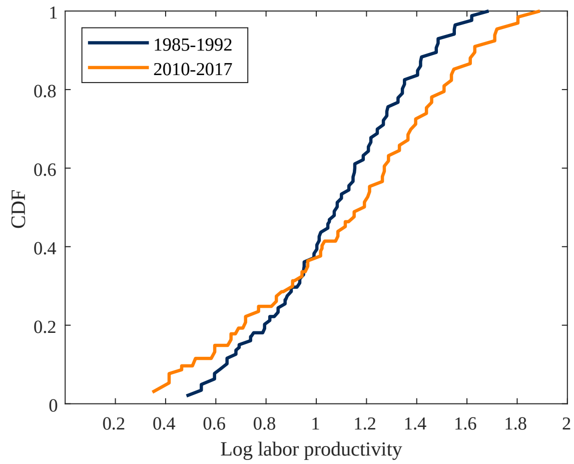


Figure 12: Increasing firm-level labor productivity dispersion in the model

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

the gap under a random allocation into teams, that is, when the joint distribution is simply the product of the marginals. Under this scenario, misallocation costs escalate more rapidly since stronger coworker complementarities induce a greater gap to potential productivity without the mitigating effect resulting from an endogenous rise in sorting. At the 1985-1992 level of χ, the gap is 3%, compared to 1.78% in the equilibrium allocation. The gap rises to 5.42% at the 2010-2017 level, compared to the baseline value of 2.17%. Simply put, the benefits accruing from specialization under the division of labor are limited not merely by the extent of the market or the cost of coordination (Becker and Murphy, 1992), but also by labor market frictions.

5.4 Discussion: robustness & extensions

This section discusses additional moments of interest, robustness exercises and extensions.

PERSON-LEVEL INEQUALITY. While the model is chiefly designed to explain the firm-level structure of wage inequality, it is instructive to consider how coworker complementarities relate to individual-level inequality. Perhaps counterintuitively, an increase in χ does not necessarily translate into a higher overall variance of log wages. Although Kremer's (1993) influential O-ring theory may lead to the prior that amplified complementarities lead to a "convexification" of the wage function, this is not the case in the model presented here. Notice first that in Kremer (1993) input markets are frictionless and, hence, strictly positive complementarity of *any* degree suffices to yield pure PAM in equilibrium; it is not the case that variation in complementaries influence the wage distribution through changes in labor market sorting. Instead, the predictions in Kremer (1993) relating to a right-skewed earnings distribution critically hinge on the assumption that production exhibits increasing returns to team quality (cf. Kremer, 1993, Section

III). Here, I deliberately consider the constant-returns case, instead, thus isolating the role of complementarities. As a result, the wage function is linear in a worker's own type only if there are no search frictions, and concave otherwise for $\chi > 0$.[75]

Extensions of the baseline model in several directions could introduce channels through which stronger complementarities raise overall inequality. One possibility is that fairness concerns (Akerlof and Yellen, 1990) or formal institutions generate within-firm pay compression for heterogeneous workers. Increased sorting due to complementarity may then weaken the degree to which these forces dampen overall wage inequality. [76] Second, coworker learning in interaction with sorting can dynamically foster lifetime inequality (Herkenhoff *et al.*, 2022; Jarosch *et al.*, 2021). Third, increasing returns to workforce quality, as in Kremer (1993), would generate a positive relationship between complementarities and person-level inequality.

PRODUCTIVITY DISPERSION. Beyond wage disparities, the model indicates that more positively assortative labor market sorting also contributed to widening labor productivity gaps between firms. Illustrating this effect, Figure 12 plots the model-implied cumulative distribution of labor productivity. The distribution 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, consistent with trends discussed in the firm dynamics literature.[77]

OUTSOURCING AND WITHIN-OCCUPATION WORKER RANKING. A concern with the baseline analysis is that the measurement of complementarities and sorting may be confounded by outsourcing dynamics or related shifts in the boundary of the firm. For instance, Song *et al.* (2019, Section V.B.) hypothesize that outsourcing may be a factor behind increased between-firm inequality. To address this possibility, I consider an alternative specification in which workers are ranked within-occupation as opposed to economy-wide. An increase in sorting or complementarities under this specification could not be driven by greater sorting of high- and low-wage occupations into different establishments. As discussed in Section 5.1, this empirical specification yields slightly lower values of coworker complementarity in both 1985-1992 and 2010-2017, but the increase is similar to the baseline in proportional terms (see Table B.3).

[75] Indeed, in the counterfactual scenario with 2010-2017 parameters except lower complementarities the overall variance of low wages is actually slightly higher than in the baseline. The primary reason is that the increase in between-firm inequality is offset by a reduction in within-firm inequality. The model-predicted rise in overall wage dispersion is due to a lower value of a_0 and a higher value of a_1. In addition, within firms stronger complementarities reduce the contribution of higher-quality types to joint surplus (cf. Bombardini *et al.*, 2012).

[76] Beyond pay, Ouimet and Tate (2023) document that the presence of high-wage workers positively predicts colleagues' non-wage benefits, likely due to regulations, fairness concerns, or administrative constraints. Complementarity induced coworker sorting may, thus, also have consequences for unequal access to non-wage benefits.

[77] While the SIEED is not suited for examining firm-level productivity, evidence from other countries suggests that rising firm-level wage inequality has indeed been paralleled by a growth in firm-level productivity dispersion. See, for instance, Decker *et al.* (2020); Sorkin and Wallskog (2021); De Ridder (2023).

I re-estimate the model for both periods by targeting coworker complementarity moments based on within-occupation worker ranking (see Appendix C.2.1), then repeat the counterfactual exercises. The important role of complementarities remains intact when controlling for changes in occupational composition across firms in this way, as the second bloc of rows in Table 5 confirms. The increase in complementarity now explains 61% of the model-predicted increase in the between-employer share of wage inequality. This corresponds to 52% of the empirically observed rise, as this specification of the model predicts a somewhat larger increase in the overall between-share.

ON-THE-JOB SEARCH. Lastly, I study the implications of on-the-job search (OJS). There are two main reasons for incorporating OJS. First, E-E transitions represent an empirically salient feature of the labor market. Second, the relationship between OJS and coworker sorting is ex-ante ambiguous. On the one hand, when workers receive outside offers while employed, they will be less discriminating among offers they receive when unemployed, since accepting a job characterized by low match quality does not imply foregoing future job opportunities with better match quality. On the other hand, precisely this possibility of on-the-job switching could enhance coworker sorting.

I model OJS by supposing that at an exogenous Poisson rate λ_e, employed workers receive outside offers from unmatched and one-worker firms. Wages both off and on the job are continuously renegotiated under Nash bargaining, with unemployment serving as the outside option.[78] In the estimation, λ_e is disciplined by EE rates taken from the matched employer-employee micro data. These indicate an increase in on-the-job mobility over time: the monthly EE rate is 0.0076 in 1985-1992 and 0.0106 in 2010-2017. Details are relegated to Appendix C.2.2.

The estimated model with OJS generates a profile of coworker sorting that is remarkably close to what we observe in the data, including in terms of changes over time, as a comparison of Figures 1b and 10b reveals. Meanwhile, the predicted increase in the between-firm share over time is smaller, reflecting a higher initial level of between-firm wage inequality. Turning to counterfactuals, when we repeat the previous exercises using the model with OJS, the rise in coworker complementarities explains about 30.5% of the overall model-predicted change. This contribution should be viewed as a lower bound, as Appendix C.2.2 explains.

Finally, in both theory and data, EE transitions reinforce positively assortative matching, and increasingly so over time. In the quantitative model, sorting is stronger with OJS than without for any given degree of coworker complementarity. Allowing for the possibility of on-the-job switching toward matches characterized by higher match quality is, thus, more than sufficient

[78]This wage specification is motivated by the empirical evidence in Di Addario *et al.* (2021), who document that where a worker is hired from tends to be relatively inconsequential for their wages in comparison to where they are currently employed, contrary to sequential auction models of labor market competition.

to offset any effect due to more muted sorting out of unemployment. The data confirm that changes in coworker quality associated with EE moves are positively related to a worker's own type, as Appendix C.2.4 verifies. Moreover, this tendency of EE transitions to reinforce positively assortative matching has strengthened over time, consistent with rising complementarities. However, the model overstates the positive correlation between own type coworker quality changes. Future work that incorporates additional determinants of the directionality of job-to-job moves – including productivity-related considerations but also amenities or idiosyncratic preferences – could improve the model's fit to the data.

6 Concluding Remarks

This paper developed, empirically tested, and quantitatively analyzed a model of the firm as a "team assembly technology." The model delivers three key insights. First, when production involves the division of labor among workers with specialized skills – team production – their talents are naturally complements; the more differentiated workers are in terms of task-specific skills, the stronger is this complementarity. Second, in labor markets characterized by frictions, coworker complementarity shapes between- and within-firm heterogeneity in worker talent and pay, and introduces a wedge between equilibrium and efficient output. Third, complementarities have strengthened over time, paralleling an increase in specialization, which accounts for a quantitatively significant part of the "firming up" of wage inequality.

In terms of broader framework development, the paper points in two useful directions. First, concerning technological change, the theory developed is tractable but encompassing enough to explore specialization as one dimension along which the nature of work varies across time and jobs. This points to a richer way of thinking about technological change, as can be illustrated by considering the example of novel Artificial Intelligence (AI) tools. By diffusing both codified and tacit knowledge, these tools appear to not only elevate the average productivity of previously less productive individuals – a leveling effect in terms of absolute advantage – but also enables individuals to perform unfamiliar tasks, thereby potentially reducing the depth of specialization.[79] Beyond labor markets, this likely carries implications for firm organizations, management, and product market barriers to entry.

Second, regarding the theory and empirics of firms, the model developed here conceives of a firm as a collection of heterogeneous individuals, rather than an abstract productivity term that workers interact with.[80] It opens a small window into the black box of firms, shedding

[79] For early studies, see for instance, Brynjolfsson *et al.* (2023), Peng *et al.* (2023) and Noy and Zhang (2023).

[80] In Caicedo *et al.*'s (2019) pithy formulation: "[...] all knowledge in the economy is held by the individual people who comprise it: there is no abstract technology hovering above them in the ether."

light on the complementarities between heterogeneous coworkers' talents. However, the model omits many other important considerations, such as the empirical firm size distribution or ex-ante firm heterogeneity in dimensions such as product quality. In ongoing work, I explore these dimensions by integrating heterogeneous firm dynamics, as in Bilal *et al.* (2022), with team production. Moreover, inside the firm, the current model assumes perfect division of labor, ignoring real-world coordination frictions (Becker and Murphy, 1992). Indeed, some firms appear to be better in coordinating their workforce than others (Coraggio *et al.*, 2022; Kuhn *et al.*, 2022). An extension of the team production model suggests that this organizational capacity is a more significant determinant of productivity when specialized skills are important (Freund, 2023). This theoretical conjecture invites empirical and quantitative evaluation.

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 Autor, D. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

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. *Job Market Paper*.

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

Ahmadpoor, M. and Jones, B. F. (2019). Decoding team and individual impact in science and invention. *Proceedings of the National Academy of Sciences*, **116**(28), 13885–13890.

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.

Allen, T. and Arko (2019). Modern Spatial Economics. *Graduate Lecture Notes*.

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

Alvarez, F. and Lucas, R. E. (2007). General equilibrium analysis of the Eaton–Kortum model of international trade. *Journal of Monetary Economics*, **54**(6), 1726–1768.

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*.

Andrews, M. J., Gill, L., Schank, T., and Upward, R. (2008). High wage workers and low wage firms: Negative assortative matching or limited mobility bias? *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **171**(3), 673–697.

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.

Atakan, A. E. (2006). Assortative Matching with Explicit Search Costs. *Econometrica*, **74**(3), 667–680.

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.

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. and Handel, M. J. (2013). Putting Tasks to the Test: Human Capital, Job Tasks, and Wages. *Journal of Labor Economics*, **31**(S1), S59–S96.

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.

Babet, D., Godechot, O., and Palladino, M. G. (2022). In the Land of AKM: Explaining the Dynamics of Wage Inequality in France. *mimeo*.

Baqaee, D. R. and Farhi, E. (2019). The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem. *Econometrica*, **87**(4), 1155–1203.

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.

Barth, E., Davis, J., and Freeman, R. B. (2018). Augmenting the Human Capital Earnings Equation with Measures of Where People Work. *Journal of Labor Economics*, **36**(S1), S71–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. (1973). A Theory of Marriage: Part I. *Journal of Political Economy*, **81**(4), 813–846.

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*.

Bloom, N. and Reenen, J. V. (2011). Chapter 19 - Human Resource Management and Productivity. In D. Card and O. Ashenfelter, editors, *Handbook of Labor Economics*, volume 4, pages 1697–1767. Elsevier.

Bloom, N., Garicano, L., Sadun, R., and Reenen, J. V. (2014). The Distinct Effects of Information Technology and Communication Technology on Firm Organization. *Management Science*.

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

Bombardini, M., Gallipoli, G., and Pupato, G. (2012). Skill Dispersion and Trade Flows. *American Economic Review*, **102**(5), 2327–2348.

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

Borovičková, K. and Shimer, R. (2020). High Wage Workers Work for High Wage Firms. Technical Report w24074, National Bureau of Economic Research, Cambridge, MA.

Braxton, J. C., Herkenhoff, K. F., Rothbaum, J. L., and Schmidt, L. (2021). Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. M. (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, **117**(1), 339–376.

Brynjolfsson, E., Li, D., and Raymond, L. R. (2023). Generative AI at Work.

Burdett, K. and Coles, M. G. (1999). Long-Term Partnership Formation: Marriage and Employment. *The Economic Journal*, **109**(456), F307–F334.

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

Caicedo, S., Lucas, R. E., and Rossi-Hansberg, E. (2019). Learning, Career Paths, and the Distribution of Wages. *American Economic Journal: Macroeconomics*, **11**(1), 49–88.

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.

Card, D., Cardoso, A. R., and Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *The Quarterly Journal of Economics*, **131**(2), 633–686.

Cardoso, A. R., Guimaraes, P., Portugal, P., and Reis, H. (2018). The Returns to Schooling Unveiled. Technical Report 11419, Institute of Labor Economics (IZA).

Carrington, W. J. and Troske, K. R. (1997). On Measuring Segregation in Samples with Small Units. *Journal of Business & Economic Statistics*, **15**(4), 402–409.

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

Coles, M. G. and Francesconi, M. (2019). Equilibrium Search with Multiple Attributes and the Impact of Equal Opportunities for Women. *Journal of Political Economy*, **127**(1), 138–162.

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*.

Costinot, A. and Vogel, J. (2010). Matching and Inequality in the World Economy. *Journal of Political Economy*, **118**(4), 747–786.

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. (2023). Human Capital at Work: Skill Composition, Productivity, and Wages. *OECD Discussion Paper*.

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.

De Ridder, M. (2023). Market Power and Innovation in the Intangible Economy. *American Economic Review*, **forthcoming**.

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.

Dessein, W. and Santos, T. (2006). Adaptive Organizations. *Journal of Political Economy*, **114**(5), 956–995.

Di Addario, S. L., Kline, P. M., Saggio, R., and Sølvsten, M. (2021). It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German Wage Structure. *The Quarterly Journal of Economics*, **124**(2), 843–881.

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.

Eaton, J., Kortum, S., Neiman, B., and Romalis, J. (2016). Trade and the Global Recession. *American Economic Review*, **106**(11), 3401–3438.

Eberle, J. and Schmucker, A. (2017). Creating cross-sectional data and biographical variables with the Sample of Integrated Labour Market Biographies 1975-2014 : Programming examples for Stata. *FDZ-Methodenreport*, (201706 (en)).

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.

Fitzenberger, B., Osikominu, A., and Völter, R. (2006). Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift für Wirtschafts- und Sozialwissenschaften*, **126**(3), 405–436.

Freund, L. B. (2023). Coordination is Key: Task Assignment and Firm Productivity. *mimeo*.

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.

Gartner, H. (2005). The imputation of wages above the contribution limit with the German IAB employment sample. *FDZ Methodenreport 2/2005*.

Gathmann, C. and Schönberg, U. (2010). How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, **28**(1), 1–49.

Goldschmidt, D. and Schmieder, J. F. (2017). The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure. *The Quarterly Journal of Economics*, **132**(3), 1165–1217.

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.

Hall, A. and Rohrbach-Schmidt (2020). BIBB/BAuA Employment Survey 2018. BIBB-FDZ Data and Methodological Report 1/2020. *BIBB Report*.

Hall, R. E. and Milgrom, P. R. (2008). The Limited Influence of Unemployment on the Wage Bargain. *American Economic Review*, **98**(4), 1653–1674.

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. (2022). Production and Learning in Teams. Working Paper 25179, National Bureau of Economic Research.

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

Hong, L. (2022). Coworker Sorting, Learning, and Inequality. *mimeo*.

Hong, L. and Lattanzio, S. (2022). The Peer Effect on Future Wages in the Workplace. *Available at SSRN*.

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.

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.

Kambourov, G. and Manovskii, I. (2009). Occupational Mobility and Wage Inequality. *The Review of Economic Studies*, **76**(2), 731–759.

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

Kleinman, B. (2022). Wage Inequality and the Spatial Expansion of Firms. *mimeo*.

Kline, P., Saggio, R., and Sølvsten, M. (2020). Leave-Out Estimation of Variance Components. *Econometrica*, **88**(5), 1859–1898.

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.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V., and Violante, G. L. (2000). Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis. *Econometrica*, **68**(5), 1029–1053.

Kuhn, M., Luo, J., Manovskii, I., and Qiu, X. (2022). Coordinated Firm-Level Work Processes and Macroeconomic Resilience. *mimeo*.

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

Lazear, E. P. and Shaw, K. L. (2007). Personnel Economics: The Economist's View of Human Resources. *Journal of Economic Perspectives*, **21**(4), 91–114.

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

Lindenlaub, I. and Postel-Vinay, F. (2023). Multi-Dimensional Sorting under Random Search. *Journal of Political Economy*.

Lochner, B. and Schulz, B. (2022). Firm Productivity, Wages, and Sorting. *Journal of Labor Economics*.

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.

Marimon, R. and Zilibotti, F. (1999). Unemployment vs. Mismatch of Talents: Reconsidering Unemployment Benefits. *The Economic Journal*, **109**(455), 266–291.

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.

Martinez, J. (2021). Putty-Clay Automation. *CEPR Discussion Paper DP16022*.

Mas, A. and Moretti, E. (2009). Peers at Work. *American Economic Review*, **99**(1), 112–145.

McFadden, D. (1977). Modelling the Choice of Residential Location. Cowles Foundation Discussion Paper 477, Cowles Foundation for Research in Economics, Yale University.

McFadden, D. (1980). Econometric Models for Probabilistic Choice Among Products. *The Journal of Business*, **53**(3), S13–S29.

Mengano, P. (2023). Trends in Worker Bargaining Power. *Job Market Paper*.

Mihaylov, E. and Tijdens, K. G. (2019). Measuring the Routine and Non-Routine Task Content of 427 Four-Digit ISCO-08 Occupations.

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

Mion, G. and Opromolla, L. D. (2014). Managers' mobility, trade performance, and wages. *Journal of International Economics*, **94**(1), 85–101.

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.

Noy, S. and Zhang, W. (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *mimeo*.

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

Ouimet, P. and Tate, G. (2023). Firms with Benefits? Nonwage Compensation and Implications for Firms and Labor Markets.

Palladino, M. G., Roulet, A., and Stabile, M. (2021). Understanding the Role of Firms in the Gender Wage Gap Over Time, Over the Life Cycle, and Across Worker Types.

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

Peng, S., Kalliamvakou, E., Cihon, P., and Demirer, M. (2023). The Impact of AI on Developer Productivity: Evidence from GitHub Copilot.

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

Rivkin, J. and Siggelkow, N. (2003). Balancing Search and Stability: Interdependencies Among Elements of Organizational Design. *Management Science*, **49**(3), 290–311.

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. (1981). The Economics of Superstars. *The American Economic Review*, **71**(5), 845–858.

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

Sandmann, C. and Bonneton, N. (2022). Existence of a non-stationary equilibrium in search-and-matching models: TU and NTU. *mimeo*.

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

Shimer, R. and Smith, L. (2000). Assortative Matching and Search. *Econometrica*, **68**(2), 343–369.

Silva, J., Leitao, M., and Montana, J. (2022). The Role of Firms in Wage Inequality Dynamics.

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.

Stansbury, A. and Summers, L. (2020). The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy. *Brookings Papers on Economic Activity*, **Spring 2020**.

Syverson, C. (2004). Product Substitutability and Productivity Dispersion. *The Review of Economics and Statistics*, **86**(2), 534–550.

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

Tomaskovic-Devey, D., Rainey, A., Avent-Holt, D., Bandelj, N., Boza, I., Cort, D., Godechot, O., Hajdu, G., Hällsten, M., Henriksen, L. F., Hermansen, A. S., Hou, F., Jung, J., Kanjuo-Mrčela, A., King, J., Kodama, N., Kristal, T., Křížková, A., Lippényi, Z., Melzer, S. M., Mun, E., Penner, A., Petersen, T., Poje, A., Safi, M., Thaning, M., and Tufail, Z. (2020). Rising between-workplace inequalities in high-income countries. *Proceedings of the National Academy of Sciences*, **117**(17), 9277–9283.

Train, K. E. (2009). *Discrete Choice Methods with Simulation, Second Edition*, volume 9780521766555.

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)*.

Wilmers, N. and Aeppli, C. (2021). Consolidated Advantage: New Organizational Dynamics of Wage Inequality. *American Sociological Review*, **86**(6), 1100–1130.

Wood, S. and Bryson, A. (2009). High Involvement Management. In W. Brown, A. Bryson, J. Forth, and K. Whitfield, editors, *The Evolution of the Modern Workplace*, pages 151–175. Cambridge University Press, Cambridge, UK.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*, **316**(5827), 1036–1039.

Online Appendices

A Theory

A.1 Team production

The derivations are lengthy but standard in the trade literature (e.g., Eaton *et al.*, 2016; Allen and Arko, 2019).

A.1.1 Derivation of equation (8)

Start with equation (2), repeated here for convenience:

$$Y = \left(\int_0^1 q(\tau)^{\frac{\eta-1}{\eta}} d\tau \right)^{\frac{\eta}{\eta-1}},$$

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 (7), rearranged as

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

and simplifying, we obtain

$$\left(\int_0^1 \left(q(\tau)\tilde{\lambda}\right)^{\frac{\eta-1}{\eta}} dQ(\tau)^{\frac{1}{\eta}} \left(\frac{1}{Y}\right)^{\frac{1}{\eta-1}} \left(\frac{1}{\lambda}\right)^{\frac{\eta-1}{\eta}} d\tau \right)^{\frac{\eta}{\eta-1}} = Y\lambda^{\frac{1}{\eta-1}}. \tag{A.1}$$

Use $Q(\tau) = \tilde{\lambda}(\tau)q(\tau)$, bring the terms independent of τ outside the integral, and cancel exponents. Then

$$\int_{\mathcal{T}} Q(\tau) d\tau = \lambda Y. \tag{A.2}$$

A.1.2 Proof of Lemma 1

PART (I). Step 1 is to derive the distribution of shadow costs of worker i providing task τ, $\lambda_i(\tau)$. Since the efficiency draws are independently and identically distributed, the probability of i producing that task at a shadow price less than p is the same for all $\tau \in \mathcal{T}$. Given the definition

$$G_i(p) := \Pr\{\lambda_i(\tau) \le p\}, \tag{A.3}$$

the properties of the Fréchet distribution together with the FOC w.r.t. $l_i(\tau)$

$$\lambda_i(\tau) = \frac{\lambda_i^L}{a_1 z_i(\tau)}. \tag{A.4}$$

imply

$$\begin{aligned} G_i(p) &= \Pr\left\{\frac{\lambda_i^L}{a_1 z_i(\tau)} \leq p\right\} \\ &= \Pr\left\{\frac{\lambda_i^L}{a_1 p} \leq z_i(\tau)\right\} \\ &= 1 - \Pr\left\{z_i(\tau) \leq \frac{\lambda_i^L}{a_1 p}\right\} \qquad (A.5) \\ &= 1 - \exp\left(-\left(\frac{\lambda_i^L}{\iota a_1 p x_i}\right)^{-\frac{1}{\chi}}\right). \qquad (A.6) \end{aligned}$$

For step 2, consider the probability that a task $\tau \in \mathcal{T}$ can be obtained for a shadow cost of less than p,

$$G(p) := \Pr\{\tilde{\lambda}(\tau) \leq p\}. \tag{A.7}$$

Now using

$$\tilde{\lambda}(\tau) = \min_{i \in \mathcal{S}} \left\{\lambda_i(\tau)\right\}. \tag{A.8}$$

together with tools in probability, we obtain:

$$\begin{aligned} G(p) &= \Pr\left\{\min_{i \in \mathcal{S}} \tilde{\lambda}_i(\tau) \leq p\right\} \\ &= 1 - \Pr\left\{\cap_{i \in \mathcal{S}} \left(\tilde{\lambda}_i(\tau) \geq p\right)\right\} \\ &= 1 - \Pi_{i \in \mathcal{S}}\left(1 - G_i(p)\right) \end{aligned}$$

Intuitively, the lowest shadow price is weakly lower than p unless the shadow price of producing that task is greater than p for each worker, so that the distribution $G(p)$ is the complement of the probability that for every $i \in \mathcal{S}$ the shadow cost of providing the task is greater than p.

In step 3, substitute for $G_i(p)$ using equation (A.6)

$$
\begin{aligned}
G(p) &= 1 - \Pi_{i\in\mathcal{S}} \exp\left(- \left(\frac{\lambda_i^L}{\iota a_1 p x_i} \right)^{-\frac{1}{\chi}} \right) \\
&= 1 - \exp\left(- (\iota p)^{\frac{1}{\chi}} \sum_{i\in\mathcal{S}} \left(\frac{\lambda_i^L}{a_1 x_i} \right)^{-\frac{1}{\chi}} \right) \qquad \text{(A.9)} \\
&= 1 - \exp(-(\iota p)^{\frac{1}{\chi}} \Phi), \qquad \text{(A.10)}
\end{aligned}
$$

where $\Phi = \sum_{i\in\mathcal{S}} \left(\frac{\lambda_i^L}{a_1 x_i} \right)^{-\frac{1}{\chi}}$

For the final step, consider equation (9), and substitute using equation (A.10):

$$
\begin{aligned}
\lambda^{1-\eta} &= \int_0^\infty p^{1-\eta} dG(p) \\
&= \int_0^\infty p^{1-\eta} \left(\frac{d}{dp} \left(1 - \exp(-(\iota p)^{\frac{1}{\chi}} \Phi) \right) \right) dp, \\
&= \frac{1}{\chi} \iota^{\frac{1}{\chi}} \Phi \int_0^\infty p^{\frac{1}{\chi} - \eta} \left(\exp(-(\iota p)^{\frac{1}{\chi}} \Phi) \right) dp.
\end{aligned}
$$

Now use a change of variables, with $m = (\iota p)^{1/\chi}\Phi$, so that $p = \iota^{-1}\left(\frac{m}{\Phi}\right)^{\chi}$ and $dp = \iota^{-1}\chi\left(\frac{m}{\Phi}\right)^{1-\chi}\frac{1}{\Phi}dm$. Performing the integration by substitution,

$$
\begin{aligned}
\lambda^{1-\eta} &= \frac{1}{\chi} \iota^{\frac{1}{\chi}} \Phi \int_0^\infty \left(\iota^{-1} \left(\frac{m}{\Phi} \right)^{\chi} \right)^{\frac{1}{\chi} - \eta} \exp(-m) \iota^{-1} \chi \left(\frac{m}{\Phi} \right)^{1-\chi} \frac{1}{\Phi} dm \\
&= \iota^{\eta-1} \Phi^{-\chi(1-\eta)} \int_0^\infty m^{\chi(\eta-1)} \exp(-m) dm \\
&= \iota^{\eta-1} \Phi^{-\chi(1-\eta)} \Gamma(1 + \chi(1-\eta))
\end{aligned}
$$

where $\Gamma(\cdot)$ is the Gamma function, evaluated at the argument $1 + \chi(1-\eta)$. Convergence of this integral requires that $1 + \chi(1-\eta) > 0$. Intuitively, tasks may not be so substitutable that workers' time is concentrated on tasks that takes them close to zero time to perform, in which case final good production could be unbounded (cf. Alvarez and Lucas, 2007, Footnote 3). Finally, since $\iota = \Gamma(1 + \chi - \eta\chi)^{\frac{1}{1-\eta}}$, we obtain

$$
\lambda = (\Phi)^{-\chi} = \left(\sum_{i\in\mathcal{S}} \left(\frac{a_1 x_i}{\lambda_i^L} \right)^{1/\chi} \right)^{-\chi},
$$

as stated in equation (11) in the main text.

PART (II). What is the probability that worker i has the lowest shadow cost of providing a task $\tau \in \mathcal{T}$? Because productivity draws are iid, and since tasks are on a continuum, by the law of large numbers this probability will be equal to the fraction of tasks that i produces according to the optimal plan. Define this share as

$$
\begin{aligned}
\pi_i &:= \Pr\left\{\tilde{\lambda}_i(\tau) \leq \min_{k \in \mathcal{S} \setminus i} \tilde{\lambda}_k(\tau)\right\} \\
&= \Pr\left\{\min_{k \in \mathcal{S} \setminus i} \geq p\right\} dG_i(p) \\
&= \int_0^\infty \Pr\left\{\cap_{k \in \mathcal{S} \setminus i} \tilde{\lambda}_k(\tau) \geq p\right\} dG_i(p) \\
&= \int_0^\infty \Pi_{k \in \mathcal{S} \setminus i}(1 - G_k(p)) dG_i(p)
\end{aligned}
$$

Now substitute for the distribution of shadow costs:

$$
\begin{aligned}
\pi_i &= \int_0^\infty \left[1 - \exp\left(-\left(\frac{\lambda_k^L}{\iota a_1 p x_k}\right)^{\frac{1}{\chi}}\right)\right] dp \left[\frac{d}{dp}\left([1 - \exp\left(-\left(\frac{\lambda_i^L}{\iota a_1 p x_i}\right)^{-\frac{1}{\chi}}\right)\right]\right. \\
&= \left(\frac{\lambda_i^L}{\iota a_1 x_i}\right)^{-\frac{1}{\chi}} \int_0^\infty \left(\frac{1}{\chi} \iota^{\frac{1}{\chi}} p^{\frac{1}{\chi}-1}\right) \left(\exp(-(p\iota)^{\frac{1}{\chi}} \Phi_n)\right) dp \\
&= \frac{(a_1 x_i)^{\frac{1}{\chi}} \left(\lambda_i^L\right)^{-\frac{1}{\chi}}}{\Phi_n} \left[-\exp(-(\iota p)^{\frac{1}{\chi}} \Phi_n)\right]_0^\infty \\
&= \frac{(a_1 x_i)^{\frac{1}{\chi}} \left(\lambda_i^L\right)^{-\frac{1}{\chi}}}{\Phi_n}.
\end{aligned}
$$

Lastly, we need to show that π_i is not only the fraction of *tasks* that i produces; it is also the fraction of the *value* of tasks. This result obtains because under the Fréchet distribution, the distribution of shadow prices of tasks that worker n actually uses from any team member will be the same. Here is a proof, adapted from Allen and Arko (2019). The probability that the shadow cost of i providing a task τ is lower than $\tilde{p}$, conditional on that cost being the lowest for i among all team members, is

$$
\Pr\left\{\tilde{\lambda}_i(\tau) \leq \tilde{p} | \tilde{\lambda}_i(\tau) \leq \min_{k \in \mathcal{S} \setminus i} \tilde{\lambda}_k(\tau)\right\} = \frac{1}{\pi_i} \int_0^{\tilde{p}} \Pr\left\{\min_{k \in \mathcal{S} \setminus i} p_k(\tau) \geq p\right\} dG_i(p),
$$

were the first term on the right-hand side is the inverse probability that i has the lowest shadow cost of producing a given task, and the second term is the probability that the firm can be provided with a a task at a shadow cost lower than $\tilde{p}$ by a team member other than i.

Then using the same logic as in the derivation of π_i, we find that this is equal to

$$
\begin{aligned}
\Pr\left\{\tilde{\lambda}_i(\tau) \leq \tilde{p} | \tilde{\lambda}_i(\tau) \leq \min_{k \in \mathcal{S}\setminus i} \tilde{\lambda}_k(\tau)\right\} &= \int_0^{\tilde{p}} \Pi_{k\in\mathcal{S}\setminus i}\Big(1 - G_k(p)\Big) dG_i(p), \\
&= \frac{1}{\pi_i} \frac{(a_1 x_i)^{\frac{1}{\chi}} \left(\lambda_i^L\right)^{-\frac{1}{\chi}}}{\Phi} \left[-\exp(-(\iota p)^{\frac{1}{\chi}} \Phi)\right]_0^{\tilde{p}} \\
&= \frac{1}{\pi_i} \frac{(a_1 x_i)^{\frac{1}{\chi}} \left(\lambda_i^L\right)^{-\frac{1}{\chi}}}{\Phi} \Big[[-\exp(-(\iota \tilde{p})^{\frac{1}{\chi}} \Phi) - (-\exp(0))\Big] \\
&= \frac{1}{\pi_i} \pi_i \Big(1 - \exp(-(\iota\tilde{p})^{\frac{1}{\chi}} \Phi)\Big) \\
&= G(\tilde{p}),
\end{aligned}
$$

which is independent of i.

Intuitively, the planner makes team members with an absolute advantage provide a greater range of tasks exactly up to the point where the distribution of shadow costs associated with producing tasks is the same as the overall distribution of shadow costs.

PART (III). The final result immediately follows since

$$
Q_i = \pi_i \Big(\int_{\mathcal{T}} \tilde{\lambda}(\tau) q(\tau) d\tau \Big) = \pi_i \int_{\mathcal{T}} Q(\tau) d\tau = \pi_i Q.
$$

A.1.3 Proposition 1

When $\lambda = 1$, then part (i) of Lemma 1 implies that

$$
1 = \sum_{i \in \mathcal{S}} \left(\frac{a_1 x_i}{\lambda_i^L} \right)^{\frac{1}{\chi}}, \tag{A.11}
$$

Hence, from part (ii) of Lemma 1,

$$
\pi_i = ((a_1 x_i)/\lambda_i^L)^{\frac{1}{\chi}}.
$$

Next, if $y_i(\tau) > 0$, then after some straightforward manipulation the FOC w.r.t. $l_i(\tau)$ reads

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

Integrating and using the time constraint (4) implies that

$$\int_{\mathcal{T}} \lambda_i(\tau) q_i(\tau) = \lambda_i^L.$$

Since $\tilde{\lambda}(\tau) = \lambda_i(\tau)$ when $y_i(\tau) > 0$, this also means that

$$Q_i = \lambda_i^L,$$

which says that the shadow value of worker i's time is equal to the shadow value of all tasks produced by that worker. Hence, from part (iii) of Lemma 1, and given the normalization $\lambda = 1$,

$$\begin{aligned} \lambda_i^L &= \pi_i Q, \\ \Leftrightarrow \lambda_i^L &= \pi_i Y, \\ \Leftrightarrow \left(\frac{a_1 x_i}{\lambda_i^L}\right)^{\frac{1}{\chi}} &= \lambda_i^L / Y. \end{aligned}$$

Substituting this expression into equation (A.11) and rearranging for Y yields

$$Y = n^{1+\chi} \left(\frac{1}{n} \sum_{i=1}^{n} (a_1 x_i)^{\frac{1}{1+\chi}} \right)^{1+\chi}$$

A.2 Team hiring

A.2.1 Population dynamics

For any type x, the measure of unemployment satisfies

$$\delta\left(d_{m.1}(x) + \int d_{m.2}(x, \tilde{x}') d\tilde{x}'\right) = d_u(x) \lambda_u \left(\int \frac{d_{f.0}}{v} h(x, \tilde{y}) + \int \frac{d_{m.2}(\tilde{x}')}{v} h(x|\tilde{x}') d\tilde{x}' \right). \tag{A.12}$$

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

$$d_{m.1}(x)\left(\delta + \lambda_{v.u} \int \frac{d_u(\tilde{x}')}{u} h(\tilde{x}'|x) d\tilde{x}'\right) = d_u(x) \lambda_u \frac{d_{f.0}}{v} h(x) + \delta \int d_{m.2}(x, \tilde{x}') d\tilde{x}'. \tag{A.13}$$

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'),[A.1]

$$2\delta d_{m.2}(x, x') = d_u(x)\lambda_u \frac{d_{m.1}(x')}{v} h(x|x') + d_u(x')\lambda_u \frac{d_{m.1}(x)}{v} h(x'|x). \tag{A.14}$$

The economic intuition parallels the aforementioned reasoning.

A.2.2 Derivation of surplus equations

A.2.2.1 $S(x)$**.** We start with the definition of $S(x)$, repeated here for convenience:

$$S(x) = \Omega_1(x) - V_u(x) - V_{f.0} \tag{A.15}$$

Consider equation (30). The term in $[\cdot]$ obviously corresponds to $S(x)$. Rearranging the surplus sharing equations furthermore implies that

$$\big(-\Omega_1(x) + V_{e.2}(x|x') + V_{f.1}(x, x')\big) = (1-\omega)S(x'|x)$$

In words, the joint value of firm and x from being with another worker x' minus their joint outside option is equal to $(1-\omega)$ of the surplus of adding x'.

Hence, we can write

$$\rho\Omega_1(x) = f_1(x) - \delta S(x) + \lambda_{v.u}(1-\omega) \int \frac{d_u(\tilde{x}')}{u} S(\tilde{x}'|x)^+ d\tilde{x}'$$

Substituting this expression into equation (A.15) multiplied by ρ yields

$$\big(\rho + \delta\big)S(x) = f_1(x) - \rho\big(V_u(x) + V_{f.0}\big) + \lambda_{v.u}(1-\omega) \int \frac{d_u(\tilde{x}')}{u} S(\tilde{x}'|x)^+ d\tilde{x}'. \tag{A.16}$$

A.2.2.2 $S(x|x')$**.** The first equation we need is a relationship between the surplus and the total joint value. Start with the definition of $S(x|x')$, substitute using the sharing rule (21), and

[A.1] From an accounting perspective, one of the preceding three equations is redundant given because the adding-up constraint represented by equation (18) must hold and the distribution of worker types is exogenous. Similarly, $d_v(y)$ can be backed out from equation (19) given $d_f(y)$.

simplify:

$$\begin{aligned}
S(x|x') &= \Omega_2(x,x') - \Big(V_{e.1}(x') + V_{f.1}(x')\Big) - V_u(x) \\
&= \Omega_2(x,x') - \big(V_u(x') + \omega S(x') + V_{f.0} + (1-\omega)S(x')\big) - V_u(x) \\
&= \Omega_2(x,x') - S(x') - \big(V_u(x) + V_u(x') + V_{f.0}\big)
\end{aligned}$$

Hence,

$$\Omega_2(x,x') = S(x|x') + S(x') + V_u(x) + V_u(x') + V_{f.0}. \tag{A.17}$$

and analogously

$$S(x'|x) = \Omega_2(x,x') - \big(S(x) + V_u(x) + V_u(x') + V_{f.0}\big). \tag{A.18}$$

Second, consider the recursion for $\Omega_2(x,x')$ and substitute in the definitions of surplus to get a second expression for the latter:

$$\begin{aligned}
\rho\Omega_2(x,x') &= f_2(x,x') + \delta\Big[-S(x|x') - S(x'|x)\Big] \\
\Leftrightarrow \delta S(x|x') &= f_2(x,x') - \rho\Omega_2(x,x') - \delta S(x'|x)
\end{aligned}$$

Substituting for $S(x'|x)$ using equation (A.18) as well as for $\Omega_2(x,x')$ using equation (A.17), and collecting terms yields

$$\begin{aligned}
\delta S(x|x') &= f_2(x,x') - \rho\Omega_2(x,x') - \delta\big[\Omega_2(x,x') - \big(S(x) + V_u(x) + V_u(x') + V_{f.0}\big)\big] \\
&= f_2(x,x') - (\rho+\delta)\Omega_2(x,x') + \delta\big[S(x) + V_u(x) + V_u(x') + V_{f.0}\big] \\
&= f_2(x,x') - (\rho+\delta)\Big(S(x|x') + S(x') + V_u(x) + V_u(x') + V_{f.0}\Big) \\
&+ \delta\big[S(x) + V_u(x) + V_u(x') + V_{f.0}\big]
\end{aligned}$$

Simplifying yields

$$S(x|x')\big(\rho + 2\delta\big) = f_2(x,x') - \rho\big(V_u(x) + V_u(x') + V_{f.0}\big) + \delta S(x) - (\rho+\delta)S(x'). \tag{A.19}$$

A.2.3 Derivation of the wage function

The value of employment for worker x when their coworker is of type x' is

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

Combining with the surplus sharing rule (25) yields

$$w(x|x') = \rho V_u(x) + (\rho + 2\delta)\omega S(x|x') - \delta\omega S(x) \tag{A.20}$$

We can simplify this expression further. First, substitute for $(\rho + 2\delta)S(x|x')$ from equation (A.19) to get

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

Next, substitute for $S(x')$ from equation (A.16) to obtain

$$\begin{aligned} w(x|x') &= \omega f_2(x,x') + (1-\omega)\rho V_u(x) - \omega\rho\big(V_u(x') + V_{f.0}\big) \\ &\quad - \omega\left[f_1(x') - \rho\big(V_u(x') + V_{f.0}\big) + \lambda_{v.u}(1-\omega)\int \frac{d_u(\tilde{x}'')}{u} S(\tilde{x}''|x')^+ d\tilde{x}''\right] \\ &= \omega\big(f_2(x,x') - f_1(x')\big) + (1-\omega)\rho V_u(x) - \omega(1-\omega)\lambda_{v.u}\int \frac{d_u(\tilde{x}'')}{u} S(\tilde{x}''|x')^+ d\tilde{x}''. \end{aligned} \tag{A.21}$$

A.2.4 Monotonicity Lemma

The following lemma implies that Hagedorn *et al.*'s (2017) non-parametric ranking algorithm extends to the present environment.

Lemma 2. *Assume that* $\frac{f_1(x)}{\partial x} > 0$, $\frac{\partial f_2(x,x')}{\partial x} > 0$, *and* $\omega > 0$. *Then: (i) The value of unemployment* $V_u(x)$ *is monotonically increasing in* x*, and (ii) so is the wage function* $w(x|x')$.

The proof is lengthy but straightforward, involving substitution and differentiation of the surplus equations. It was included in a working paper version but is omitted here for sake of brevity.

A.3 Closed-form results for a stylized version

The full, quantitative model does not admit closed-form characterization of matching decisions and the distribution of wages. The issue lies with the value of either side's outside option (waiting) which at once informs the matching decision and at the same time is endogenous to the matching decisions taken by all agents in the economy. After all, it is those decisions that pin down who is available for meetings if you do wait. This general-equilibrium interplay between individual decisions and distributional dynamics is characteristic of the class of mean-field games to which the present model belongs. To solve for matching decisions and distributions by hand, then, we need to sever their mutual interdependence. To do so, I adapt the model of Eeckhout and Kircher (2011).[A.2] I focus on matching of pairs of workers into a team, each managed by a firm, rather than matching between one worker and one firm. I show how to derive closed-form expressions for the degree of coworker sorting and firm-level inequality.

A.3.1 Environment

As in the main model, there is unit mass of workers, with types uniformly distributed over $[0, 1]$. There is exactly half as many firms which, as in the main text, are ex-ante homogeneous with productivity normalized to unity.

Production with a single worker is normalized to be zero and $f_2(x, x') = x + x' - \gamma(x - x')^2$, where $\gamma > 0$ again indicates the strength of complementarities. This specification of team production is motivated by a second-order Taylor approximation to the micro-founded CES function in Proposition 1 around the mean worker type, i.e., around $\frac{x+x'}{2}$. Then $(x - x')^2$ is proportional to the variance of types and γ measures the production loss due to coworker mismatch.

Consider a finite-horizon setup in which there are three stages $s \in \{0, 1, 2\}$. All the focus will be on the outcomes in $s = 1$; the other two stages merely serve to frame the decision problem in $s = 1$. Everyone starts out unmatched. In $s = 0$, each firm meets and matches with one worker. Our focus is on the following matching decision, to be taken in $s = 1$. Each worker-firm pair coming into $s = 1$, of which there is a mass $\frac{1}{2}$, is randomly paired with one of the remaining unmatched workers. *Either* they match, produce, and share the output. The agents are then idle in $s = 2$. *Or* they decide to *wait*, in which case all agents separate, each worker pays a *fixed search cost* c, and they can be active in stage $s = 2$. Specifically, in return for paying the search cost,

[A.2] The spirit of this exercise is similar to Coles and Francesconi's (2019) observation about search-and-matching models with ex-ante heterogeneous agents: "[M]uch can be learned about the structure of steady state equilibria from considering the partial equilibrium conditions [for an exogenously given population of unmatched agents] in isolation." The same reasoning applies here, except that I do not adopt a "clones assumption" (Burdett and Coles, 1999) but, instead, use Eeckhout and Kircher's (2011) trick to gain tractability.

each worker in the final stage is paired with their optimal match. Finally, payoffs are determined according to the following, simplified protocol. Firms have no bargaining power and each worker receives their outside option plus half the surplus generated by the match.

Relative to the full model, we have complete destruction of matches after the production stage ($\delta = 1$); every firm is guaranteed to meet one worker in every period; and there is zero discounting. As in Atakan (2006), the search cost is *explicit* instead and, as such, *type-independent*. This contrasts with the full model, where search costs are *implicit*. In that case, more productive agents have greater opportunity costs of time due to discounting, hence search frictions disproportionately erode their value of search (Sandmann and Bonneton, 2022).

A.3.2 Solving the model by backward induction

FRICTIONLESS MATCHING IN $s = 2$. Matching in the second stage is frictionless. As described e.g. in Boerma *et al.* (2021), the problem of a firm is to choose workers x and x' to maximize profits, taking the wage schedule for workers, $w : [0, 1] \to \mathbb{R}_+$, as given:

$$v = \max_{x,x'} \left(f_2(x, x') - w(x) - w(x') \right).$$

Any worker x chooses an employer (among which they are indifferent) and coworker x' to maximize their wage income, taking as given the firm value, v, and the coworker's wage schedule:

$$w(x) = \max_{x'} \left(f_2(x, x') - w(x') - v \right).$$

An equilibrium is a tuple (π, w, v), with probability measure π over workers and coworkers indicating the assignment, such that firms solve their profit maximization problem, each worker solves their respective problem, and payoffs are feasible with respect to production. The firm's first-order condition for a worker is $\frac{\partial f_2(x,x')}{\partial x} - \frac{\partial w(x)}{\partial x} = 0$. From the second-order conditions, per Becker (1973), we know that when the production function is supermodular in worker types, i.e., when $\gamma > 0$, then the optimal assignment among workers features positive assortative matching between coworkers. That is, we have a deterministic coupling, or matching function, $\mu : [0, 1] \to [0, 1]$ and, specifically, $\mu^*(x) = x$. The wage schedule is then derived by integrating over the first-order condition under the equilibrium assignment. Under our parametric assumptions on the production function and the bargaining process, this yields $w^*(x) = x$ for any x and $v = 0$.

STAGE-1 MATCHING DECISION. Each firm with one worker, the latter being denoted x', is randomly matched with an unmatched worker whose type is x. If they match, the team produces and the output value is shares. Then they are idle in the final stage. Else, the partners are all unmatched, each worker pays a fixed search cost c, and all agents actively participate in the stage-2 matching process.

A firm with worker x' that is randomly matched a worker x decides to hire them if the joint value of production is (weakly) greater than the sum of the respective outside options, taking into account the cost of search.[A.3] Using the same notation as in the full model, this means that $h(x|x') = 1 \Leftrightarrow S(x|x') > 0$, where[A.4]

$$S(x|x') = f_2(x, x') - \left[w^*(x) + w^*(x') + v^* - 2c\right].$$

Substituting for the last-period payoffs, i.e., $w^*(x) = x$ and $v^*(y) = 0$, and simplifying shows that a match is formed whenever $|x' - x| > s^*$, where the equilibrium threshold s^* satisfies

$$s^* = \sqrt{2c/\gamma}. \tag{A.22}$$

Equivalently, the (symmetric) matching set is $M(x') = \{x \in \mathcal{X} : x' - s^* > x < x' + s^*\}$.[A.5]

It immediately follows that greater complementarities, corresponding to a higher value of γ, render the matching set narrower, whereas greater search costs, c, render the matching set wider.

STAGE-0 MATCHING DECISION. At the very beginning, each firm is randomly paired with one worker. It is optimal to match, since the opportunity cost of doing so is zero.[A.6]

A.3.3 Characterization

Next, we can characterize the sorting patterns and wage-distributional outcomes implied by this matching rule in closed-form. To ensure maximum comparability with the full model, I take the following approach: I focus on the outcomes for workers that are part of matched formed in stage 1; and I re-weight these workers according to the (uniform) population distribution. That is, the marginal distribution of worker types participating in stage-1 matches is taken to be uniform. In this way, the results are not biased by the fact that stage-2 outcomes exhibit no mismatch by construction; nor is the workforce composition mechanically biased towards workers with intermediate quality levels who accept a wider range of partners than those at the tails of the

[A.3] I assume that when the parties are indifferent between producing now and waiting they opt for the former. This assumption does not, of course, affect the substantive results.

[A.4] As discussed in Footnote 7 of Eeckhout and Kircher (2011), for low types the surplus in the next period may not exceed the total waiting cost. In order to avoid keeping track of endogenous entry, I assume that people will search even if that is the case.

[A.5] This simple formulation also emerges as a special case in Eeckhout and Kircher (2011). The result also resembles circular production models à la Marimon and Zilibotti (1999).

[A.6] To resolve the indifference case, we could of course assume an infinitesimally small positive production value.

quality distribution .[A.7]

MATCHING PATTERNS. The key step affording analytical tractability is to observe that for any threshold level s, the distribution of coworker types conditional on type x, denoted $\Phi(x'|x)$, takes the form of a piecewise uniform distribution:

Lemma 3 (Conditional type distribution)**.** *Given a threshold distance s, the conditional distribution of coworkers for $x \in \mathcal{X}$ is*

$$\Phi(x'|x) = \begin{cases} 0 & \textit{for } x' < \sup\{0, x-s\} \\ \frac{x-\sup\{0,x-s\}}{\inf\{x+s,1\}-\sup\{0,x+s\}} & \textit{for } x' \in [\sup\{0, x-s,\}, \inf\{x+s, 1\}] \\ 1 & \textit{for } x' > \inf\{x+s, 1\} \end{cases}$$

Proof. The distribution of coworkers for a worker of type x is

$$\Phi(x'|x) = \frac{\int_{\mathcal{X}} \mathbf{1}\{\tilde{x} \in M(x)\} d\Phi(\tilde{x})}{\Phi(\overline{m}(x)) - \Phi(\underline{m}(x))}$$

where $\overline{m}(x)) = \sup M(x)$ and $\underline{m}(x)) = \inf M(x)$; given the assumption of uniform weights on types among the matched $\Phi(x)$ is the uniform cdf.

Assuming that potential coworkers are uniformly distributed, the equilibrium matching rule then implies that

$$x' \sim \begin{cases} U[0, x+s] & \text{for } x \in [0, s) \\ U[x-s, x+s] & \text{for } x \in [s, 1-s] \\ U[x-s, 1] & \text{for } x \in [1-s, 1] \end{cases}$$

Writing this out more carefully for three different segments, for $x \in [0, s)$:

$$\Phi(x'|x) = \begin{cases} 0 & \text{for} & x' = 0 \\ \frac{x'}{x+s} & \text{for} & x' \in [0, x+s) \\ 1 & \text{for} & x' > x+s. \end{cases}$$

[A.7]Concretely, since the matching set is $M(x) = \{x' \in \mathcal{X} : x - s^* < x > x + s^*\}$, for any $s^* < 1$, workers outside the interval $[s^*, 1 - s^*]$ would be under-represented relative to those inside that interval.

Next, for $x \in [s, 1-s]$:

$$\Phi(x'|x) = \begin{cases} 0 \quad \text{for } x' < x - s \\ \frac{x'-(x-s)}{1-(x+s)} \quad \text{for } x' \in [x-s, x+s] \\ 1 \quad \text{for } x' > x + s. \end{cases}$$

Finally, for $x \in (1-s, 1]$:

$$\Phi(x'|x) = \begin{cases} 0 \quad \text{for } x' < x - s \\ \frac{x'-(x-s)}{1-(x-s)} \quad \text{for } x' \in (x-s, 1] \\ 1 \quad \text{for } x' = 1. \end{cases}$$

□

Using Lemma 3, we can describe coworker sorting patterns in closed-form as follows:

Proposition A.1 (Coworker sorting). *For a given threshold s,*

(i) *the coworker correlation is given by*

$$\rho_{xx} = (2s+1)(s^2-1)^2;$$

(ii) *the average coworker type is*

$$\hat{\mu}(x) = \hat{\mu}(x) = \begin{cases} \frac{x+s^*}{2} \quad \textit{for} \quad x \in [0, s^*) \\ x \quad \textit{for} \quad x \in [s^*, 1-s^*] \\ \frac{1+x-s^*}{2} \quad \textit{for} \quad x \in (1-s^*, 1]. \end{cases}$$

Proof. Part (ii) immediately follows from 3. For part (ii), notice first that under the uniform distribution, the mean type is $\bar{x} = \frac{1}{2}$ and the variance is $\frac{1}{12}$. We need to derive the covariance term, which is defined as

$$\bar{c}_{xx} = \int_X \int_X (x - \bar{x})(x' - \bar{x}) d\Phi(x'|x) d\Phi(x).$$

Step 1. Substituting for the piece-wise conditional match distribution and the uniform population weights, for a threshold s,

$$\begin{aligned}
\bar{c}_{xx} &= \int_0^s \int_0^{x+s} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{x+s} dx' dx \\
&+ \int_s^{1-s} \int_{x-s}^{x+s} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{2s} dx' dx \\
&+ \int_{1-s}^{1} \int_{x-s}^{1} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{1-(x-s)} dx' dx
\end{aligned}$$

Step 2. Integrate over x'.

The different components are:

$$\begin{aligned}
\int_0^{x+s} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{x+s} dx' &= \frac{1}{2}(x - \bar{x})(s + x - 2\bar{x}) \\
\int_{x-s}^{x+s} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{2s} dx' &= (x - \bar{x})^2 \\
\int_{x-s}^{1} (x - \frac{1}{2})(x' - \frac{1}{2}) \frac{1}{1-(x-s)} dx' &= \frac{1}{2}(x - \bar{x})(1 - s + x - \bar{x})
\end{aligned}$$

Step 3. Integrate over x.

$$\begin{aligned}
\bar{c}_{xx} &= \int_0^s \frac{1}{2}(x - \bar{x})(s + x - 2\bar{x}) dx + \int_s^{1-s} (x - \bar{x})^2 dx + \int_{1-s}^{1} \frac{1}{2}(x - \bar{x})(1 - s + x - \bar{x}) dx \\
&= \left[\frac{5}{12}s^3 - \frac{5}{8}s^2 + \frac{1}{4}s\right] + \frac{(2s-1)^3}{12} + \left[\frac{5}{12}s^3 - \frac{5}{8}s^2 + \frac{1}{4}s\right] \\
&= \frac{1}{12}\left[(2s+1)(s-1)^2\right]
\end{aligned}$$

Hence, the *correlation* is

$$\rho_{xx} = (2s+1)(s^2-1)^2.$$

□

Part (i) of Proposition A.1, together with equation (A.22) summarizing the equilibrium matching decision, demonstrates that the coworker correlation coefficient, ρ_{xx} is *increasing* in γ – stronger complementarities give rise to more pronounced coworker sorting. In particular, $\rho_{xx} \to 1$ as $\gamma \to \infty$. Conversely, greater search costs dilute the incentive to wait for the best match and accordingly lower the degree of sorting observed in the economy. This result therefore sharply summarizes the tradeoff between complementarities and search costs in determining coworker sorting patterns in the economy.[A.8]

Part (ii) of Proposition A.1 paints a more disaggregated picture of matching patterns that reveals *non-linearities* arising from *asymmetries* in the matching set. It defines for every worker type the average coworker type for a given threshold level s. Figure 4a in the main text graphically illustrates. The dashed-dotted and dotted lines describe matching patterns under, respectively, the deterministic coupling $\mu(x) = x$ that captures matching decisions under PAM in the frictionless economy (dashed 45-degree line); and the independent coupling that intuitively corresponds to a random matching process (dotted, horizontal line). The solid line describes $\hat{\mu}(x)$ for a low value of γ and the dashed line illustrates $\hat{\mu}(x)$ for a higher value of γ.

We can make three observations. First, for "middle types" such as $x = \frac{1}{2}$, the average coworker type is invariant to changes in γ. Intuitively, stronger complementarities mean that such an agent is less likely to be in a team with a much better agents but the likelihood of being teamed up with a much worse agent shrinks in symmetric fashion. Second, the presence of search costs means that low types are, on average, paired up with agents better than them, whereas high types are typically paired up with coworkers *worse* than them. Lastly, a strengthening of complementarities *increases* the average coworker type for the *best*; and it *lowers* it for the worst. This force has thus the potential to engender polarizing dynamics wherein firms with "superstar teams" pull away while "laggards" fall behind.

WAGE DISTRIBUTION.

The production function, matching patterns, and wage sharing rule jointly determine the distribution of wages in the economy. Here we are particularly interested in the fraction of wage dispersion that occurs *between* firms. A tedious but otherwise straightforward sequence of integration and algebra steps yields the following result.

Proposition A.2. *Given a threshold distance s and a value of γ, the between-firm share of the variance of wages is equal to*

$$\frac{-\frac{13\,\gamma^2\,s^5}{2400} + \frac{\gamma^2\,s^4}{80} + \frac{5\,s^3}{36} - \frac{s^2}{6} + \frac{1}{12}}{\frac{\gamma^2\,s^4}{45} - \frac{4897\,\gamma^2\,s^5}{10800} - \frac{\gamma^2\,s^6}{324} + \frac{19\,\gamma^2\,s^5\,\ln(2)}{30} + \frac{1}{12}}. \tag{A.23}$$

Proof. To find the between-team share of the variance of wages, first compute the total wage dispersion as the sum of the variance of wages between and within *types*. Then compute the variance of the average wage by team, and the between-share is the ratio of the latter over the

[A.8] In practice, the worker types x are not observable, of course. Defining an observable measure of types following Borovičková and Shimer (2020), and denoted $\lambda(x)$, yields to a coworker sorting correlation coefficient that is virtually indistinguishable from ρ_{xx} (in population). It is also characterizable in closed form but less intuitively, being equal to $\rho_{\lambda\lambda} = \frac{10\gamma^2 s^6 - 3\gamma^2 s^5 + 3240 s^3 - 4860 s^2 + 1620}{-60\gamma^2 s^6 + 36\gamma^2 s^5 + 1620}$.

former. Many of the steps are straightforward but lengthy, following similar steps as in the preceding section, hence I only sketch the procedures and intermediate results.

Total wage dispersion. The average wage can be computed as

$$\bar{\lambda} = \int_0^s \lambda^b(x)dx + \int_s^{1-s} \lambda^m(x)dx + \int_{1-s}^1 \lambda^h(x)dx$$

where

$$\lambda(x) = \int_0^1 w(x,x')d\Phi(x'|x).$$

Performing the relevant integration steps, utilizing Lemma 3, and simplifying yields

$$\lambda(x) = \begin{cases} \lambda^b(x) := x - \frac{\gamma}{6}\left[(x-s)^2 + xs\right] & \text{for } x \in [0,s) \\ \lambda^m(x) := x - \frac{\gamma}{6}s^2 & \text{for } x \in [s, 1-s] \\ \lambda^h(x) := x - \frac{\gamma}{6}\left[1 + s - 2x - s^2 - x^2 - sx\right] & \text{for } x \in (1-s, 1]. \end{cases}$$

After some further integration and simplification, we obtain

$$\bar{\lambda} = \frac{1}{2} - \frac{1}{6}\gamma s^2(1 - \frac{s}{3})$$

The between-type variance is equal to

$$\bar{\sigma}^2_{\text{between-type}} = \int_0^s \left(\lambda^b(x) - \bar{\lambda}\right)^2 dx + \int_s^{1-s} \left(\lambda^m(x) - \bar{\lambda}\right)^2 dx + \int_{1-s}^1 \left(\bar{\lambda}^h(x) - \bar{\lambda}\right)^2 dx,$$

which corresponds to the contribution of worker heterogeneity to total wage inequality.

After more integration and algebra, we find that this is equal to

$$\bar{\sigma}^2_{\bar{\lambda}} = \frac{1}{12} - \frac{\gamma^2 s^5}{12}(\frac{s}{27} - \frac{1}{45})$$

Conditional on a type x, the variance of wages is equal to

$$\sigma^2_{\lambda(x)} = \int_0^1 \left(w(x|x') - \lambda(x)\right)^2 d\Phi(x'|x).$$

Calculating this expression for the three segments of worker types – bottom, middle, and high –

we then obtain the average within-type wage variance as

$$\sigma^2_{\text{within }\lambda} = \int_0^s \sigma^2_{\lambda^b(x)} dx + \int_s^{1-s} \sigma^2_{\lambda^m(x)} dx + \int_{1-s}^1 \sigma^2_{\lambda^h(x)} dx.$$

Performing the usual piece-wise integration yields

$$\sigma^2_{\text{within-type}} = \frac{\gamma^2\, s^4\, \left(2280\, s\, \ln(2) - 1639\, s + 80\right)}{3600},$$

and, hence, the total wage variance is equal to

$$\sigma^2_w = \frac{1}{12} + \frac{\gamma^2\, s^4}{45} - \frac{4897\, \gamma^2\, s^5}{10800} - \frac{\gamma^2\, s^6}{324} + \frac{19\, \gamma^2\, s^5\, \ln(2)}{30}.$$

Between-share of wage inequality. Finally, we compute the variance of the average wage in a firm, which given our assumptions on the firm earning zero return just corresponds to output per worker (i.e., productivity). That is,

$$\begin{aligned}
\sigma^2_{w,\text{between-firm}} &= \int_0^s \left[\int_0^{x+s} \left(\frac{f_2(x,x')}{2} - \bar{\lambda}\right)^2 \frac{1}{x+s} dx'\right] dx \\
&+ \int_s^{1-s} \left[\int_{x-s}^{x+s} \left(\frac{f_2(x,x')}{2} - \bar{\lambda}\right)^2 \frac{1}{2s} dx'\right] dx \\
&+ \int_{1-s}^{1} \left[\int_{x-s}^{1} \left(\frac{f_2(x,x')}{2} - \bar{\lambda}\right)^2 \frac{1}{1+s-x} dx'\right] dx \\
&= \frac{1}{12} - \frac{13\, \gamma^2\, s^5}{2400} + \frac{\gamma^2\, s^4}{80} + \frac{5\, s^3}{36} - \frac{s^2}{6},
\end{aligned}$$

where the last equality follows after another sequence of integration and algebra. The result in Proposition A.2 then follows by taking the ratio $\frac{\sigma^2_{w,\text{between-firm}}}{\sigma^2_w}$. □

Proposition A.2 makes precise the following two points. First, the between-firm variance share unambiguously increases with the strength of coworker complementarities, captured by γ. Second, for $s = 0$, which in particular obtains in equilibrium when search costs are absent, between-firm inequality accounts for all of the dispersion in wages.

B Empirics

B.1 Data

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). A detailed description can be found in vom Berge *et al.* (2020). To ensure best practices, I extensively rely on publicly available code by Eberle and Schmucker (2017) and Dauth and Eppelsheimer (2020). Individual records originate in labour administration and social security data processing.[B.1]

The SIEED covers every worker at a random sample of 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 the panel establishments, but instead require a minimum number of persons in every establishment-year cell (see below). Variables available include a worker's establishment and average daily wage alongside a rich set of other characteristics, including employment status, age, gender, tenure, occupation, and education, among others. Throughout, I use the KldB-1988 2-digit occupational classification, which the IAB reports in harmonized form throughout the entire sample period.

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).

My sample selection criteria are similar to other studies using this dataset or studying similar topics (e.g., Card *et al.*, 2013). In a first step, I select employees aged 20-60 with workplaces in West German states who are liable to social security and are not in part-time or marginal employment (i.e., I limit the sample to full-time employeese). I also drop jobs with real daily earnings of less than 10 Euros. I drop observations in select industries and create a consistent industry classification at the 2-digit level of the OECD STAN-A38 nomenclature.[B.2]

A well-known and non-trivial limitation of the German matched employer-employee data is that the earnings variable is top-coded at the so-called "contribution assessment limit" of the social security system ("Beitragsbemessungsgrenze"). To impute right-censored wages, I implement best practices, specifically following Card *et al.* (2013), who build on Gartner (2005) and Dustmann *et al.* (2009). This approach involves 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

[B.1] Relative to the LIAB dataset, which is more commonly used, including in a previous version of this paper, the SIEED is a new product offered by the IAB. It does not include information from the IAB Establishment Panel survey but, crucially, comes with a much larger coverage.

[B.2] I drop Agriculture, forestry and fishing (1); Mining and quarrying (5); Utilities (35-36); as well as Activities of households as employers; undifferentiated activities of households for own use (97) and Activities of extraterritorial organizations and bodies (99). The selection aims to ensure consistency with analyses for the Portuguese case.

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, following Dauth and Eppelsheimer (2020), whose publicly available code I heavily rely on, I limit imputed wages at 10 times the 99th percentile. In a second sample restriction step, I then drop establishment-year cells with fewer than ten full-time employees or worker-year observations and restrict attention to the largest connected set. The final sample (1985-2017) includes 17,126,027 person-year observations for 1,982,239 unique persons, whose average age is 38.49. The median unweighted establishment size is 29. The average real daily wage in 2010 is 106.48 Euros.

B.1.2 Portuguese micro-data

The construction of the Portuguese dataset is described in detail in Criscuolo *et al.* (2023), from which the results relating to Portugal that are shown in this paper are excerpted.[B.3] Here I provide a brief summary. After an extensive data cleaning procedure to handle duplicate person and employer identifiers, we impose sample restrictions similar to those described in the preceding section. All observations retained relate to persons employed by third parties. Real hourly wages include base pay, regular benefits, and bonus pay. As wages are not top-coded, no imputation procedure is needed. I drop observations below the statutory minimum monthly pay, as in Cardoso *et al.* (2018). By merging in balance sheet and income statement information from the Informação Empresarial Simplificad (SCIE), we can construct a measure of value-added per worker to study dispersion in firm-level productivity. Given the sampling frame of the SCIE, I restrict myself to the non-financial corporate sector. While primarily focusing on the 2010-2017 sample, where longer periods are considered I use a combination of official and unofficial but widely used harmonization crosswalks for occupations and industries. The final sample (1986-2017) includes 21,200,256 person-year observations, of which 6,930,892 belong to the 2010-2017 sample. In 2017, there are 848,970 workers employed at 19,391 unique firms. Throughout, as a measure of a worker's type I use the AKM method. The same analysis but using the non-parametric ranking approach yields very similar results when comparing methods in a subsample containing only one gender (the full sample is too large for the non-parametric method to be feasible).

Regarding the industry-level proxy for coworker skill complementarity used in Section 4.3, I first construct the "CTIC" measures (Communication, Impact, Commmunication, Contact) for each O*NET SOC occupation based on the average score assigned by the O*NET experts on a

[B.3] In addition to Chiara Criscuolo and Peter Gal, I thank the following individuals for their generous answers to my questions: Ana Rute Cardoso, Priscilla Fialho, Paulo Guimarães, Pedro Portugal, Pedro Raposo and Marta Silva.

Likert scale from one to five, and taking the mean across the four different dimensions. Then, after cross-walking to the ISCO-2d classification available in the QP data using publicly available crosswalks, I construct an employment-weighted mean score for each 4-digit industry. The reader is referred to Bombardini *et al.* (2012) for details on the questions selected from O*NET.

B.2 Methodology

B.2.1 Measuring worker types

As stated in Section 4.1, I implement two alternative approaches to estimate each worker's time-invariant quality type in the data. One ranks workers by their fixed effect from Abowd *et al.* (1999, AKM hereafter) style log-linear wage regressions. An alternative that is less common but more consistent with the structure of the model uses the non-parametric ranking algorithm proposed in Hagedorn *et al.* (2017). This section provides details on the implementation of both methods.

Under either method, given a individual's unique rank in the (sample-period specific) type distribution, I compute an individual's decile rank in the year-specific distribution. Hence, types are uniformly distributed in each year, even as the number of workers varies across years (different from the model). In any case, the binning procedure also implies that a worker's decile rank generally does not change across years. I omit a time subscript to underline that $\hat{x}_i$ denotes an individual characteristic that is time-invariant modulo sample composition changes.

B.2.1.1 AKM models. Estimation of two-way fixed-effects regressions in the spirit of AM is a popular approach to account for unobservable worker and firm effects, is widely used in empirical studies of wage inequality and can, at a minimum, be viewed as a useful diagnostic tool. I estimate the AKM model separately for five overlapping sample periods. This estimation is implemented in Stata using the *reghdfe* package (Correia, 2017).

My implementation takes care to mitigate limited mobility bias, which is a form of incidental parameter bias arising from the fact that a large number of firm-specific parameters (i.e., fixed effects) that are solely identified from workers who move across employers. To this end I follow Bonhomme *et al.* (2019) and reduce the dimensionality of the estimation problem by clustering similar firms.[B.4] Clusters are found by solving a weighted k-means problem,

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

[B.4]For Portugal, I have experimented with a variety of bias-correction methods. In particular, the method discussed here as well as the approaches of Andrews *et al.* (2008) and Kline *et al.* (2020) yield very similar findings. Moreover, the uncorrected version appears significantly biased, overestimating the contributions of firm-specific pay premia and underestimating the degree of worker-firm and worker-worker segregation.

where $k(1), ..., k(J)$ constitutes a partition of firms into K known classes; $\hat{F}_j$ 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, as reported also by Bonhomme *et al.* (2019)). I use firms' wage distributions over the entire sample period on a grid of 20 percentiles for clustering.

Different from Bonhomme *et al.* (2019), and in similarity to Palladino *et al.* (2021), I then stick to the two-way fixed effect regression approach rather than estimating a correlated randon effects model. That is, after imputing a cluster to each worker-year observation, I estimate the following regression, which uses cluster effects instead of firm effects:

$$\ln(w_{it}) = \alpha_i + \sum_{k=1}^{K} \psi_k \mathbf{1}(j(i,t) = k) + \epsilon_{it} \tag{B.2}$$

where $\mathbf{1}(J(i,t) = k)$ are dummies indicating which cluster k firm the employer of i in period t, $j(i,t)$ has been assigned to. I associate with each firm j the fixed effect of the cluster to which j belongs and denote it ψ_j. For the sake of exposition, I abstract from observables, but in practice I control for an index of time-varying characteristics that includes a cubic in age and a quadratic in job tenure, exactly as in the construction of the residualized wages described in Section B.1.1.

B.2.1.2 Non-parametric ranking algorithm. As is well known (Eeckhout and Kircher, 2011; Lopes de Melo, 2018; Bonhomme *et al.*, 2019), the restrictions imposed by the AKM model are inconsistent with common structural models, including that laid out in Section 3. As an alternative way of estimating worker quality types, I therefore implement a version of the non-parametric ranking algorithm proposed in Hagedorn *et al.* (2017). For a lemma indicating the theoretical applicability of this method, see Appendix section A.2.4. Intuitively, since both production and the value of unemployment are increasing in worker productivity, wages within firms are also increasing in worker type. We can then infer a partial ranking of employees in a given firm from their wages (Anna and Bob both work at Zara in Cambridge; if Anna's wage is higher than Bob's, the algorithm interprets that as Anna being ranked above Bob). If no firm matches with all workers, we have to aggregate the partial within-firm rankings to a global one by exploiting the logic of transitivity and worker mobility across employers (Bob changes job to work for H&M and has a new coworker, Carl; if Bob is ranked above Carl, then we infer Anna's rank to be greater than Carl's.)

To handle potential inconsistencies across partial rankings, measurement error being one relevant source, Hagedorn *et al.* (2017, Online Appendix C) adapt the Kemeny-Young method, which minimizes the sum of Kendall-tau distances between two rankings, and propose a computationally feasible approximation to the NP-hard problem of uncovering the exact solution. In my implementation for convergence I require the correlation between the rankings obtained from consecutive iterations to be at least 0.995. The implementation in Octave, run on the FDZ-IAB servers, takes around three weeks when a ranking is created separately for each of five sample periods.

B.2.2 Approximating the cross-partial wage derivative

The cross-partial derivative of the wage function, denoted $\frac{\partial^2 w(x|x')}{\partial x \partial x'}$, is a high-dimensional object. The main text, specifically Section 4.2, discussed two methods to approximate the average cross-partial: a regression-based approach and a non-parametric (NP) method. Here I first briefly summarize the practical implementation of the NP method, and then show that in simulated data generated by the structural model, both methods approximate the truth quite well.

To non-parametrically approximate the average cross-partial derivative of the wage function, I proceed as follows:

(i) Construct the non-parametric wage function, which given an approximation of the worker type space on n_x grid points is a $n_x \times n_x$ matrix. Denote as w_{ij} the average wage of a worker in decile i of the type distribution whose average coworker is in decile j of the coworker type distribution, $i = 1, ..., n_x$ and $j = 1, \ldots, n_x$.

(ii) To prevent local nonlinearities in the conditional wage function from biasing the estimate, fit a polynomial separately for each worker type x. Denote the predicted values $\tilde{w}_{ij}$.(In that sense, the approach here is not completely non-parametric.)

(iii) Given a matrix of these predicted values, numerically compute the cross-partial derivative based on the matrix $[\tilde{w}_{ij}]_{i=1,\ldots,n_x}^{j=1,\ldots,n_x}$ using finite difference methods. Specifically, using forward differences for for $i, j = 1$; using backward differences for $i, j = n_x$; and using central differences everywhere else.

(iv) Finally, compute the (unweighted) average; $\widehat{\frac{\partial^2 w(x|x')}{\partial x \partial x'}} = \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \left[\widehat{\frac{\partial^2 w(x|x')}{\partial x \partial x'}} \right]_{ij}$.

For validation purposes, I evaluate both the NP and the regression-based method in data generated from the structural model. In the latter model, I can compute the exact average cross-partial derivative of the wage function given the parametrized production function and the mapping between production cross-partial and wage cross-partial in Corollary 3. Figure B.1 shows that both methods slightly under-estimate the exact value for any given degree of

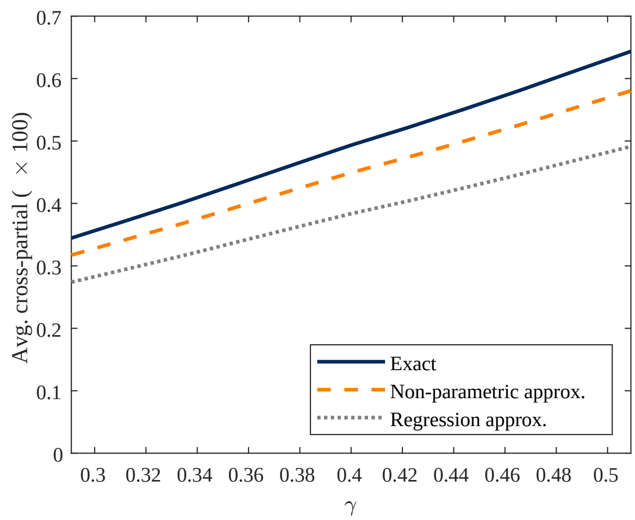


Figure B.1: Recovering the average cross-partial derivative

Notes. This figure shows the true average cross-partial (solid line) alongside approximations obtained using the non-parametric (dashed) and regression-based (dotted) methods, respectively. The model parameters are those indicated in Table 3, except that the elasticity of complementarity γ is varied, as indicated along the horizontal axis.

complementarity, but replicate the positive relationship very well. The slight under-estimation comes from the fact that neither method fully captures that the cross-partial is increasing in the distance between x and x'. For the approximation to work well, it turns out to be important to weight observations by the inverse probability that the worker-coworker combination associated with that observation occurs in the data. Else, the parts of the state space where matches are unlikely to occur is underweighted.

B.3 Additional descriptive results

B.3.1 Cross-country evidence on the "firming up" of between-firm inequality

Figure B.2 illustrates cross-country trends for the between-firm share of wage inequality, drawing on aggregated statistics kindly made available by Tomaskovic-Devey *et al.* (2020). While the levels cannot be straightforwardly compared across countries due to variation in the measure of earnings used (e.g., hourly vs. daily vs. monthly earnings), one can observe a consistent upward trend for almost all countries.

Interestingly, even in a country like France, where total wage inequality has broadly flatlined over the past few decades, the between-firm component tends to have increased due to rising sorting and segregation, whereas within-firm inequality has declined for a variety of reasons (Babet *et al.*, 2022).

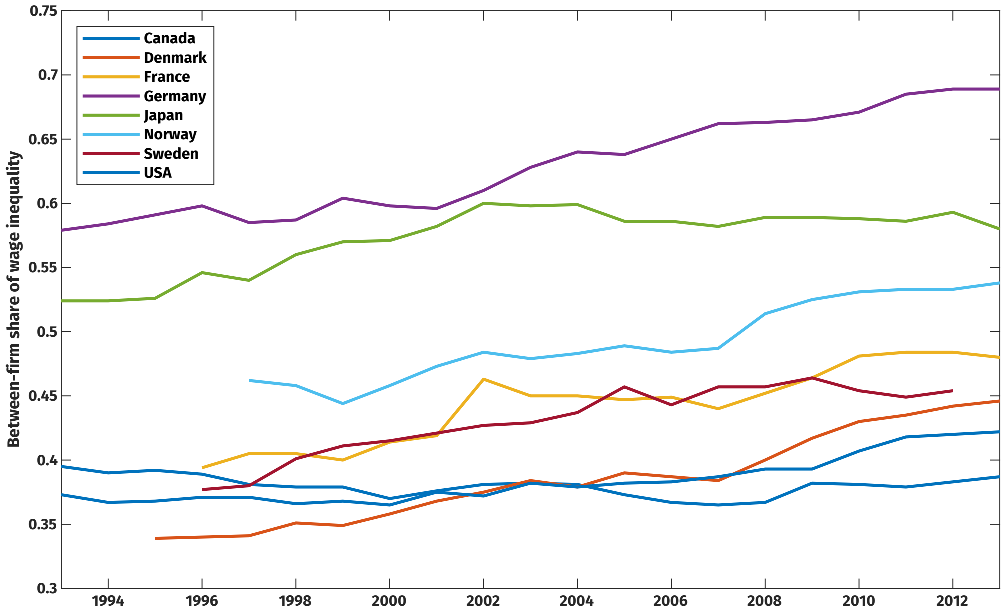


Figure B.2: Cross-country evidence on the between-firm share of wage inequality

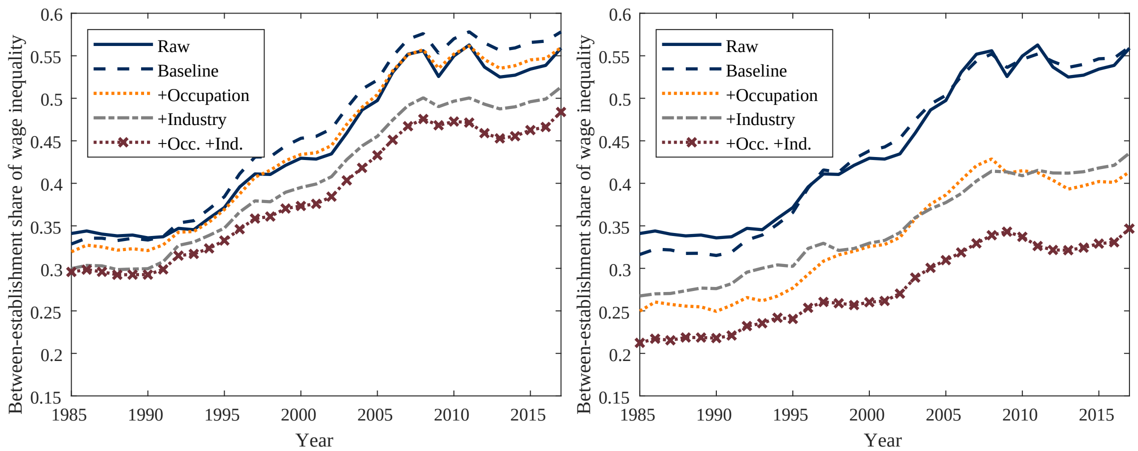
Notes. This figure reports the evolution of the between-firm share of wage inequality for a set of OECD economies. Data source: Tomaskovic-Devey *et al.* (2020).

B.3.2 Wage inequality: alternative wage variables

This section examines how robust the increase in the between-establishment share of the variance of log wages in Germany is to alternative measures of individual wages. Figure B.3 provides a graphical summary and indicates that controlling for different sets of covariates does not alter the main conclusion of a rising between-employer component.

The solid line in Figure B.3, in either panel, depicts the standard between-within establishment decomposition of the variance of log raw wages, $\ln(\tilde{w})_{it}$. In the left panel, the dashed line in the same colors swaps in the baseline measure of residualized wages used in the main text. Next, the dotted, dashed-dotted, and dotted-crossed lines also remove, respectively, (2-digit) occupation FEs, (2-digit) industry FEs, or both. Throughout, by including the worker FEs in the "residual" component, it is ensured that we do not omit variation in time-invariant individual earnings that is also present in – indeed, central to – the structural model. For robustness, the right panel in the figure repeats the same exercise but without including these person FEs when residualizing wages.

Several observations stand out. First, and reassuringly, the rise in the between-establishment share of wage inequality is highly robust across all specifications. While the *level* of this share varies, the increase over time is quite uniform, ranging from 13 to 25 percentage points (1985-2017). Second, and unsurprisingly, both the level and the percentage point increase in the



(a) With worker fixed effects

(b) Without worker fixed effects

Figure B.3: Rising between-establishment share of wage inequality: alternative wage variables

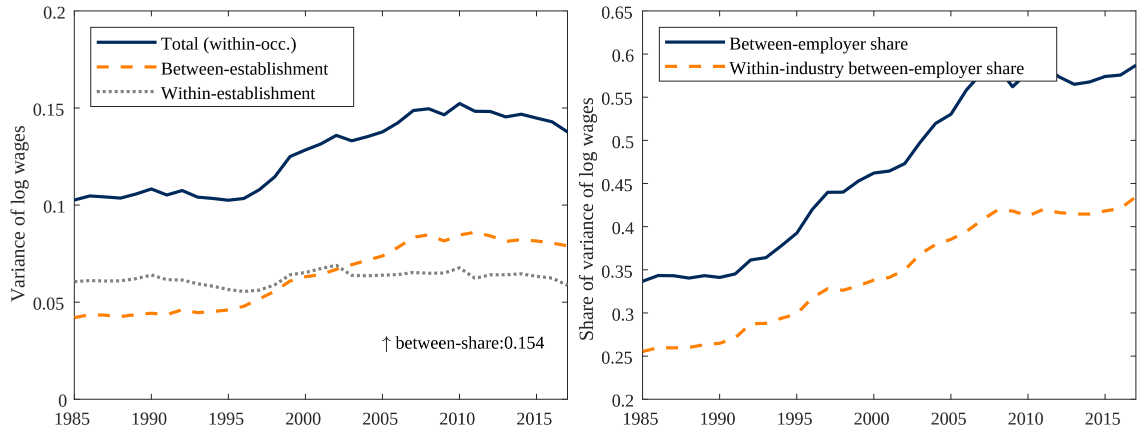
Notes. This figure depicts the share of the variance of log wages occurring between as opposed to within establishments for different measures of wages: raw wages, the baseline residuals used in the main text, and alternatives that also incorporate, respectively, occupation or industry fixed effects, or both. The relevant regressions are performed for the entire sample, then the decomposition is performed year-by-year.

between-establishment share are smaller in magnitude when worker FEs are not accounted for. Third, turning to the role of occupations and industrie, when controlling for both occupation and industry FEs, the between share rises from 0.21 to 0.35, with most of the increase occurring up until the Global Financial Crisis. That the increase in the between share is smaller, in percentage point terms, when taking out occupation and industry FEs is consistent with two points made in the literature. The first point is that between-firm inequality partly arises from between-industry differences in average pay (Haltiwanger and Spletzer, 2020). The second is that some of the rise in between-firm inequality is due to occupational outsourcing. Goldschmidt and Schmieder (2017), in particular, document that outsourcing of cleaning, security, and logistics services accounts for around 9% of the increase in German wage inequality since the 1980s.[B.5] Moreover, through the lens of an AKM model – considered in the preceding Appendix section – outsourcing manifests in a reduction in all three of the between components. In summary, additional controls can account for some but not the majority of the rise in the between-firm share of wage inequality.

B.3.3 Wage inequality: within-occupation and within-industry decomposition

Figure B.4a depicts a within-occupation decomposition of the variance of log wages into the between-employer and within-employer components. To this end, the between-within de-

[B.5] Bilal and Lhuillier (2021) argue, in the French context, that outsourcing leads to rising labor market sorting but relatively stable wage inequality, due to an offsetting general-equilibrium mechanism associated with pro-competitive effects of contractors at the bottom of the job ladder.



(a) Within occupation **(b)** Within industry

Figure B.4: Rising wage inequality: occupation and industry perspective

Notes. The left panel shows a decomposition of the within-occupation variance of log wages, as described in the main text. The right panel indicates the between-employer share of the variance of log wages when computed in the economy as a whole, including between-industry differences, and when computed within-industry. All results are based on the baseline measure of residualized log wages and are weighted by person-years.

composition is conducted separately for each two-digit occupation, then a person-weighted year-by-year average is taken across occupations. It can be seen that the between-employer share has risen substantially also when controlling for between-occupation effects in this way.

Figure B.4b considers the role of industry differences. I decompose the variance of log wages into three components: between-industry, within-firm within-industry, and between-firm within-industry. The analysis is done at the two-digit STAN-A38 level. The solid line shows the total between-firm share of the variance of log wages (which is the fraction of the total variance due to both between-industry and between-firm within-industry components), while the dashed line shows the share of the within-industry wage variance due to the between-firm within-industry component. Unsurprisingly, between-firm wage differences play a somewhat smaller role when controlling for industry (dashed line) compared to the economy-wide analysis (solid line). The rise captured by the dashed line clearly indicates, however, that the between-employer share has risen also within industry.

B.3.4 AKM-based variance decompositions

The log-linear structure of the AKM model facilitates widely used diagnostic decompositions. Given estimated worker and employer fixed effects, α_i and ψ_j, we can decompose the variance of log wages as in Song *et al.* (2019):

$$\mathrm{Var}(w_{it}) = \underbrace{\mathrm{Var}(\alpha_i - \bar{\alpha}_j) + \mathrm{Var}(\epsilon_{it})}_{\text{within-component}} + \underbrace{\mathrm{Var}(\psi_j)) + 2\mathrm{Cov}(\bar{\alpha}_j, \psi_j) + \mathrm{Var}(\bar{\alpha}_j)}_{\text{between-components}}, \tag{B.3}$$

where $\bar{\alpha}_j$ is the average worker FE in firm j. A common interpretation then relates $\mathrm{Var}(\psi_k)$ to firm/cluster-specific "wage premia", $\mathrm{Cov}(\bar{\alpha}_j, \psi_j)$ to "worker-firm sorting", and $\mathrm{Var}(\bar{\alpha}_j)$ to "worker segregation."

Before presenting the empirical decompositions, I stress that through the lens of a structural model with coworker complementarities the three components just mentioned are not neatly separable. In particular, the AKM model will attribute correlations in output between workers at the same employer, insofar as they are reflected in wages (and conditional on worker types) to a common employer component, even if these correlations are, in reality, due to coworker complementarities. This has at least two implications. First, the "worker-firm sorting" component may partly pick up coworker sorting. Second, the common interpretation of the employer fixed effect as reflecting firm wage premia or wage-setting practices (as opposed to worker characteristics) is not necessarily warranted.

A natural follow-up question is whether these conjectures are borne out when estimating AKM regressions on panel data simulated from the structural model. Unfortunately, a quantitatively credible answer to this question requires a version of the model that allows for a firm size n significantly above two. Otherwise, and imposing empirically consistent patterns of worker mobility, there is insufficient variation for an AKM-type model to stand a fair chance at identifying the employer fixed effects. Qualitatively, exercises confirm that estimating an AKM model on simulated panel data yields positive contributions of the variance of employer fixed effects and the covariance between worker and employer fixed effects. Solving a model with unrestricted within-firm coworker interactions in production and "large" firms remains an unresolved computational challenge (given the exploding state space), which these exercises suggests may be a promising avenue.

Turning to the empirical variance decompositions, the left panel in Figure B.5 depicts the results for the German economy on a period-by-period basis. We see that all three between-components have risen over time. Moreover, the Kremer-Maskin segregation index (Kremer and Maskin, 1996), which indicates the share of the variance of worker FEs that occurs between rather than within establishments – and, thus, represents an alternative measure of coworker sorting – has risen over time. The picture looks very similar if occupation and industry fixed effects are accounted for as well.[B.6] Overall, changes in workforce composition account for more

[B.6]Interestingly, if one performs the decomposition separately by major sector groups, the upward trend is most

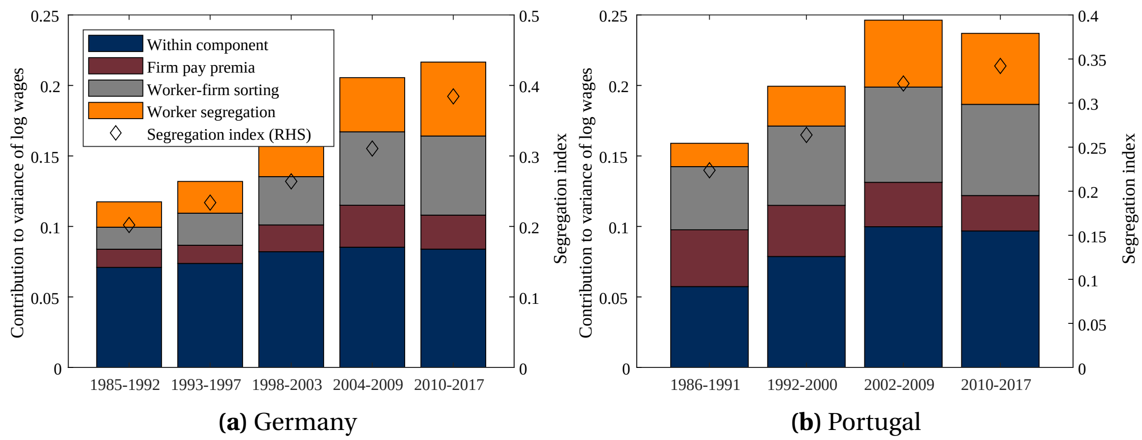


Figure B.5: AKM-based wage variance decomposition

Notes. These panels report the components of the variance of log wages, following equation (B.3), based on the estimation of the AKM model described in equation (1). The left panel is for Germany, the right panel for Portugal (where I only use four sample periods to better span missing years in 1990 and 2001).

than fourth fifths of the total increase in the between-employer component, with changes in what are commonly labelled "firm pay premia" accounting for the remainder.[B.7]

For comparison purposes, the right-hand panel reports the variance decomposition for the Portuguese economy. A key takeaway is that even though the between-firm share of the variance has remained roughly constant since the late 1990s, both $\mathrm{Cov}(\bar{\alpha}_j, \psi_j)$ and $\mathrm{Var}(\bar{\alpha}_j)$ have increased, as has the segregation index – similar to the German economy. Controlling for changes in the firm-specific pay premia, therefore, the sorting and segregation components would have pushed up between-firm inequality, in line with what we observe for Germany and many other advanced economies. Silva *et al.*'s (2022) analysis of the Portuguese economy come to a similar conclusion.

B.3.5 Schooling: Labor market sorting and Mundlak decomposition

This section performs a Mundlak decomposition of the year-by-year OLS return to schooling, replicating unpublished material from Card *et al.* (2013).[B.8] The analysis also reveals that labor market sorting has risen also when worker productivity is proxied by years of schooling.

---

pronounced for knowledge-intensive services.

[B.7] I thus find a slightly smaller contribution of firm pay premia than Card *et al.* (2013), my numbers being more similar to those by Song *et al.* (2019) for the U.S. One potential reason is that I use a dimension-reduction technique following Bonhomme *et al.* (2019) to mitigate limited-mobility bias. In addition, I do not include education as a control variable. My results remain similar when estimating a higher-dimensional model that also takes out occupation and industry fixed effects.

[B.8] Also see `https://doku.iab.de/veranstaltungen/2020/PhD_2020_Card_Keynote.pdf` (last accessed: 2023-10-05).

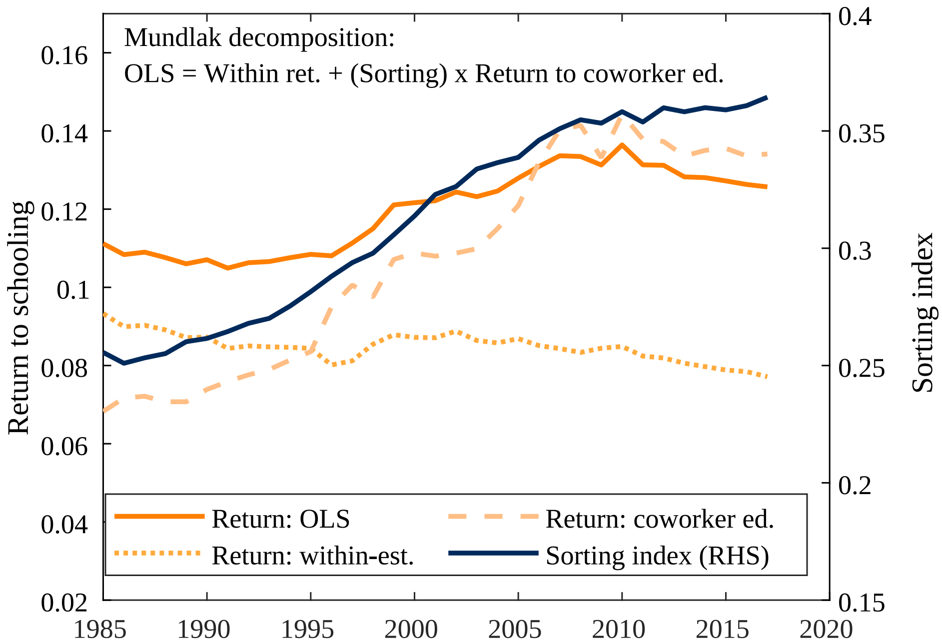


Figure B.6: Mundlak decomposition of the returns to schooling

Notes. This figure depicts the year-by-year Mundlak decomposition of the OLS return to years of schooling. "Return: OLS" corresponds to β_t^{ols} in the text; "Return: coworker ed." to $\beta_t^{\text{estab.}}$; "Return: within-est" to β_t^{within}; and "Sorting index" to ρ_t. The wage regressions also control for a quadratic of potential experience and a gender dummy.

The decomposition is performed in two steps. First, consider a regression of the log wage of individual i in year t on their years of schooling, S_i, and the average years of schooling among workers in their establishment, $\bar{S}_{j(i,t),t}$:

$$\ln w_{it} = \beta_0 + \beta_t^{\text{within}} S_i + \beta_t^{\text{estab.}} \bar{S}_{j(i,t),t} + e_{it}. \tag{B.4}$$

Second, recover the partial correlation coefficient ρ_t from a regression of $\bar{S}_{j(i,t),t}$ on S_i:

$$\bar{S}_{j(i,t),t)} = \beta_0^{\rho} + \rho_t S_i + \tilde{e}_{i,t}. \tag{B.5}$$

Then the sum of the coefficients on S_i and on $\bar{S}_{j(i,t),t}$, weighting the latter by ρ_t, is equal to the OLS return to schooling:

$$\beta_t^{\text{ols}} = \beta_t^{\text{within}} + \rho_t \times \beta_t^{\text{estab.}},$$

where β_t^{ols} is estimated from the regression

$$\ln w_{it} = \beta_0 + \beta_t^{\text{ols}} S_i + \check{e}_{it}. \tag{B.6}$$

Here, β_t^{within} measures the within-establishment return to schooling, while $\beta_t^{\text{estab.}}$ is an estimate of the individual wage return to average establishment schooling, and ρ_t measures coworker sorting by schooling. Notice that the last term captures a property of the joint *distribution* of schooling and coworker schooling, while the first two terms measure the *wage return* to schooling.

Figure B.6 plots this decomposition on a year-by-year basis. Two observations stand out. First, labor market sorting by schooling has consistently increased over time. Second, underpinning the modest rise in the OLS return to schooling is a more substantial rise in the return to coworker education and a mildly declining within-establishment return to schooling.

B.3.6 Additional evidence for a rise in specialization

This section reviews additional evidence regarding changes in specialization over time.

First, and briefly, other studies that impose stronger, structural assumptions to quantify specialization likewise point to a strengthening of specialization over time. Most importantly, Grigsby (2023) semi-parametrically estimates the economy-wide, multidimensional skill distribution using U.S. panel data on wages and occupation choices. Grigsby's (2023) findings are consistent with an increase in specialization: The average within-worker variance of productivities across tasks has grew by around 50% between the 1980s and 2000s, and transferability of skills across jobs has declined amongst occupations requiring high-skill training, manual and social skills.[B.9]

Second, a broader rise in the prevalence of teamwork is likewise consistent with specialization having become more important. Beyond the case of science discussed in the main text, such a rise is documented in surveys, such as Lazear and Shaw (2007) for the U.S., and Wood and Bryson (2009) for the UK, as well as Bresnahan *et al.* (2002) and Bloom and Reenen (2011).

Third, we can use the SIEED matched employer-employee and BIBB survey data jointly to study how individuals' employer-to-employer transitions reallocate them across jobs with more or less similar task requirements, asking in particular how these movements in task space have evolved over time.

I proceed in four steps, extending the analysis in Gathmann and Schönberg (2010) to a time series dimension. First, consider the set of occupation O. Considering fifteen harmonized tasks in the BIBB, the task content for each occupation $o \in O$ is defined as the vector $\bar{\mathbf{l}}o = (\bar{l}_{o1}, \ldots, \bar{l}_{o15})$, where $\bar{l}_{o\tau}$ denotes the fraction of workers in occupation o performing task $\tau \in \{1, ..., 15\}$. Second, for any two occupations o and o', I compute the distance in task space as

[B.9] An increase in specialization is also consistent with Braxton *et al.* (2021), who document an increase in persistent earnings risk since the 1980s (in U.S. data), concentrated in occupations involving more non-routine cognitive tasks.

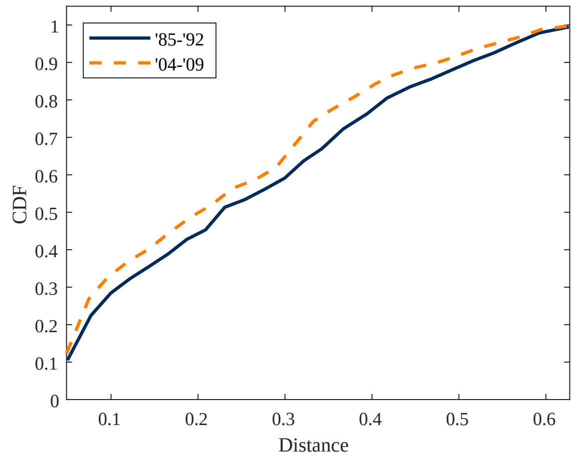


Figure B.7: Distribution of task-distances for occupational movers

Notes. This figure plots the cumulative distribution of the distances in task space associated with occupational switchers, as defined in the text. The CDF is approximated at 25 points for data protection reasons.

the (normalized) angular distance,

$$\varphi(\bar{\mathrm{I}}o, \bar{\mathrm{I}}o') = \frac{1}{\pi}\cos^{-1}\left(\frac{\bar{\mathrm{I}}'o\bar{\mathrm{I}}o'}{||\bar{\mathrm{I}}o|| \cdot ||\bar{\mathrm{I}}o'||}\right) \in [0,1].$$

Steps one and two are performed separately for each of the waves of the BIBB survey. To construct a matrix of pairwise distances that is invariant over time and thus focus on changes in movement patterns, I take the average across different survey waves.

Third, in the SIEED I identify the subsample of occupational movers. An individual i who in period t is employed at j in occupation o counts as an *occupational mover* if in $t+1$, i is employed at $j' \neq j$ and $o' \neq o$.[B.10] Fourth, the distances $\{\varphi(\bar{\mathrm{I}}_o, \bar{\mathrm{I}}_{o'})\}_{oo'}$ from the BIBB are merged into the mover subsample of the SIEED.

Following these steps, Figure B.7 plots the cumulative distribution of distances for the periods 1985-1992 and 2004-2009. We observe that the distribution of moves in the earlier period stochastically dominates that in the latter period. The unconditional average distance travelled has fallen by 10%, declining from 0.253 to 0.227.[B.11]

B.4 Additional results

[B.10]Considering the subsample of occupational movers instead of the occupational switching rate itself effectively controls for shocks that move the latter around and which the existing literature has argued are important (Kambourov and Manovskii, 2009).

[B.11]This result is robust to controlling for individual potential experience, origin occupation, gender as well as the state of the business cycle proxied by the overall unemployment rate. Moreover, quantile regressions confirm that the decline can be observed across different points of the distribution. Results are available upon request.

B.4.1 Non-linear coworker aggregation

The main text explains that for each person-year observation I construct an average coworker type by computing the unweighted arithmetic mean of all coworkers' types, and notes that this approach ignores a non-linearity in the aggregation that is implied by the structural model. Here I elaborate on this approach and discuss robustness results which suggest that the bias from ignoring this non-linearity is small.

To recap, my baseline approach is to average over coworkers without weighting, that is I compute $\hat{x}_{-it} = \left(\frac{1}{|S_{-it}|}\sum_{k\in S_{-it}}\hat{x}_k\right)$, where S_{-it} is the set of i's coworkers in year t. However, the structural model and specifically the micro-foundation in Section 3.1 suggest that we ought to aggregate using a power mean that assigns disproportionate weight to low-type coworkers insofar as coworker complementarities are important. Denote this alternative by $\tilde{x}_{-it} = \left(\frac{1}{|S_{-it}|}(\hat{x}_k)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$. Some intuition for what is lost by ignoring this non-linear aggregation can be gained by taking a second-order Taylor approximation around $\hat{x}_{-it}$, which shows that $\tilde{x}_{-it} - \hat{x}_{-it} \approx -\frac{1}{2}\gamma\frac{\sigma_x^2}{\hat{x}_{-it}}$, where σ_x^2 is the variance of coworker types. This shows that the unweighted average is upward biased in proportion to the product of complementarities and the dispersion among coworkers. From the perspective of the structural model, this offers reassurance, since dispersion – which is another manifestation of imperfect coworker sorting — should be low precisely when γ is high.

To evaluate the magnitude of bias empirically, I compare the estimates of coworker sorting and complementarities for the final sample period under different weighting schemes.[B.12] In addition to the baseline relying on linear aggregation, I recompute the coworker correlation and estimate regression (37) when computing a weighted average coworker, for each person-year, under three different values of γ: 0.25, 0.50 and 0.75. Table B.1 summarizes the results. It demonstrates that all four approaches yield very similar estimates. I have, in addition, performed Monte Carlo simulations using a statistical model of team formation, similar to that introduced in Section C.1.3, an exercise that likewise suggests a bias in the estimated complementarities that is positive but small.

B.4.2 Additional cross-sectional validation results

This section reports additional, empirical analysis using the Portuguese micro-data.

First, the richness of the Portuguese data facilitates an alternative operationalization of who an employees' coworkers are: those who work at the same hierarchical level of a firm. From

[B.12] Briefly, it is tempting to consider an iterative approach that alternates between, on the one hand, guessing a value of γ and constructing a weighted average coworker based on that value, and etimating wage complementarities and inferring a new value of γ, on the other hand. While this is possible in principle, note that this inference ultimately relies on the entire structure of the model, since γ is not uniquely pinned down by a given estimate of wage complementarities. Apart from the computational load, it is not possible to estimate the structural model directly on the servers of the IAB, where the micro data are stored.

	Unweighted	Weighted: (0.25)	Weighted (0.50)	Weighted (0.75)
Sorting	0.613	0.617	0.616	0.614
Complementarity	0.00905	0.00896	0.00882	0.00870

Table B.1: Robustness: coworker sorting and complementarities under non-linear aggregation

Notes. This table summarizes estimates of coworker sorting and complementarities when the average coworker type is computed, respective, as the unweighted arithmetic mean or the weighted power mean, using three different weighting parameters γ. The "Sorting" indicates the correlation coefficient between own type and average coworker type, while "Complementarity" refers to the point estimate of the interaction coefficient in regression (37). The sample period is 2010-2017.

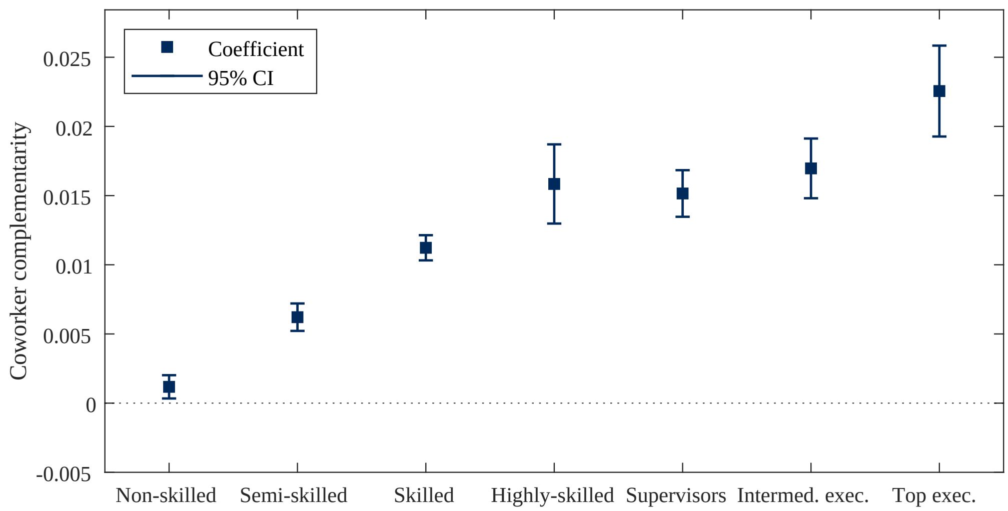


Figure B.8: Complementarities across different hierarchical layers

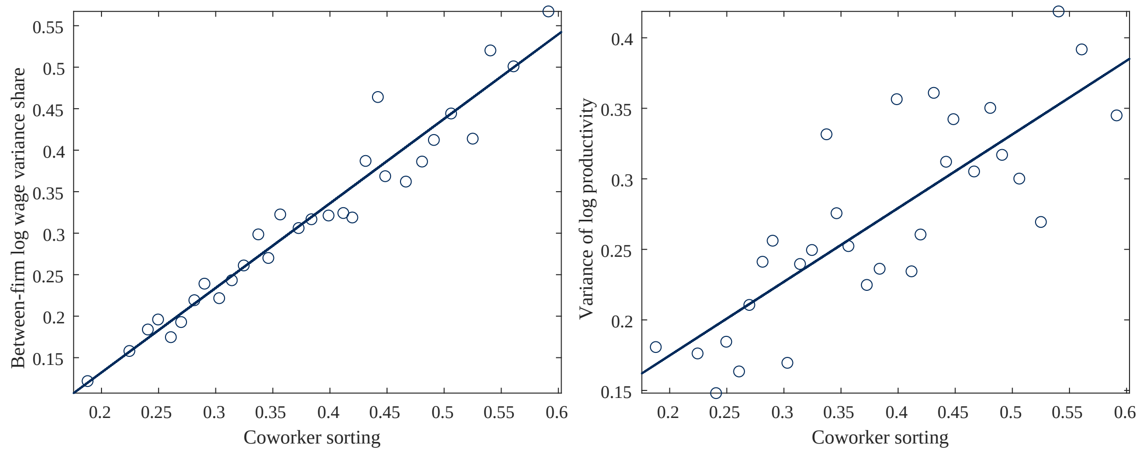
Notes. This figure reports the point estimate for the coefficient β_c, alongside confidence intervals, when regression (37) is estimated separately for workers belonging to different hierarchy levels. To avoid overlap across hierarchy levels, the set of employee *i*'s coworkers here is restricted to those employees in the same employer-year-layer cell.

2010 onward, we can use a variable that assigns workers according to a consistent definition into seven different, vertically differentiated layers, grouped by similarity in the complexity of tasks and skills required. These range from top executives to non-skilled.[B.13] If we interpret these hierarchical layers as an ordinal proxy for χ, in the spirit of Garicano (2000), then we expect coworker wage complementarity to be greater in higher layers.

To test this hypothesis, I run a variation of the regression specification in (37), for each $l = 1, \ldots, 7$.[B.14] Figure B.8 shows the estimated value of β_c alongside 95% confidence bands for

[B.13] Per Decree-Law 380/80, firms should indicate for each employee the qualification level indicated in the relevant Collective Agreement. If this is not available, firms should select the qualification level of the worker. Table B-1 in Mion and Opromolla (2014) provides details on tasks and skills. Also see Caliendo *et al.* (2020). I exclude apprentices/interns/trainees.

[B.14] I deliberately estimate the regression separately for each layer, as opposed to pooling samples and including a



(a) Between-firm share of wage dispersion

(b) Productivity dispersion

Figure B.9: Coworker sorting is associated with greater between-firm inequality

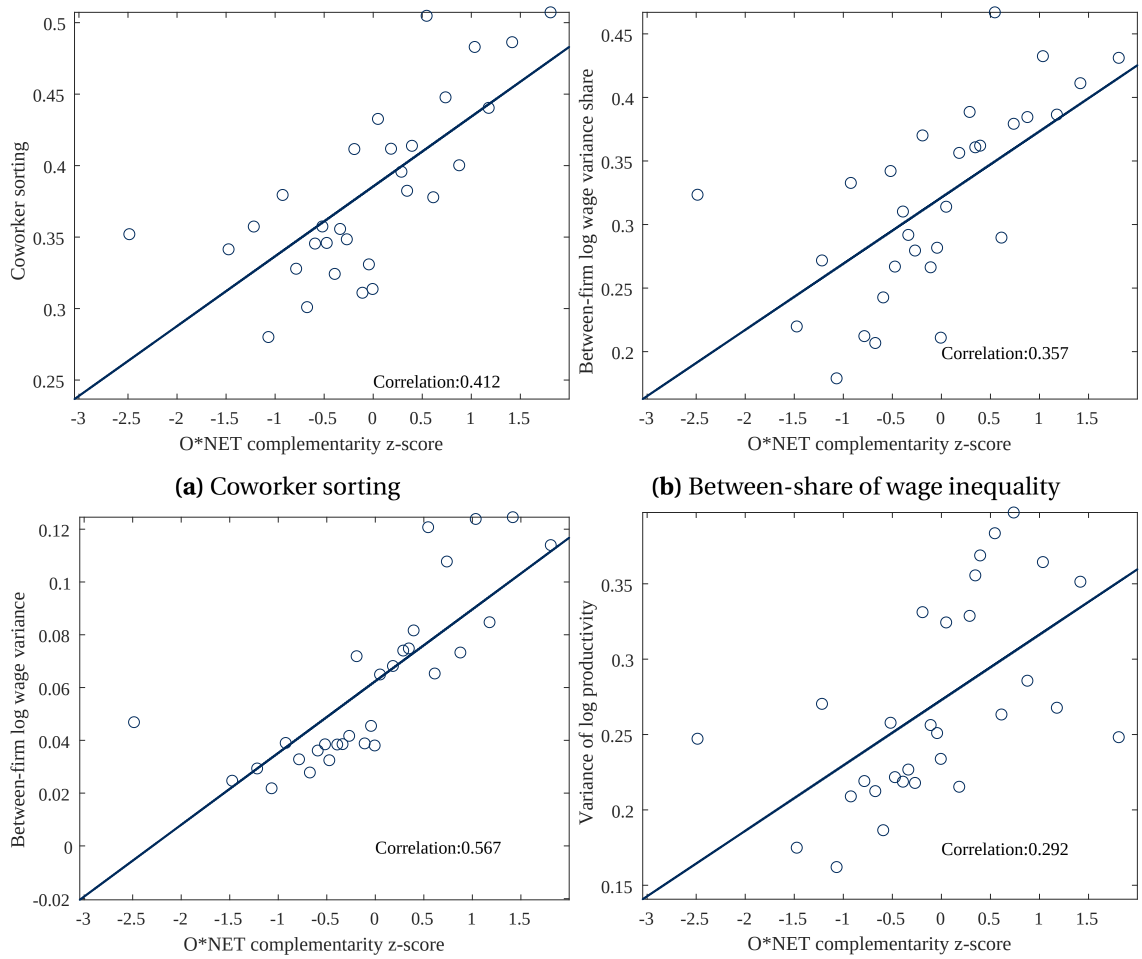
Notes. This figure plots the variance of log wages (left panel) and the between-firm share of the variance of log wages (right panel) against the coworker sorting correlation. The unit of observation is a 4-digit NACE industry, with at least 5,000 person-year observations and where the industry-level proxy for complementarity is within two standard deviations. Observations are grouped into 30 bins. The linear regression line is fitted based on unweighted, non-binned observations.

each layer. The point estimates are almost monotonically increasing in the hierarchical layer, ranging from effectively zero for those classified as non-skilled to above 0.02 for top executives. This is not due to the fact that the left-hand side of the regression is the wage in levels and higher-layer workers earn more, so that β_c is mechanically increasing in the layer, as the dependent variable is the wage level divided by the layer-specific average wage.

Next, Figure B.9 expands on the cross-industry findings from Section 4.3. It shows in industries which feature strong coworker sorting a greater share of the variance of log wages occurs between as opposed to within firms. This is unsurprising but reassuring. In addition, firm-level productivity is more dispersed in these industries.

These results echo a literature that has highlighted the important role of differences in workforce composition across firms in explaining between-firm wage inequality. In addition to the aforementioned Card *et al.* (2016) and Song *et al.* (2019), Håkanson *et al.* (2021) use direct measures of workers skills from military enlistment tests in Sweden and find that between-firm skill inequality is a key driver behind between-firm wage inequality. In addition, Sorkin and Wallskog (2021) find for the U.S., that productivity dispersion, between-firm earnings inequality, and coworker sorting are successively higher for more recent cohorts of firms in the U.S (also see Berlingieri *et al.* (2017)).

double interaction between the interaction term and l, to avoid restricting potential variation in β_1 and β_2 across l. The downside is that, as a result, we do not exploit variation from switchers across layers within firms.



(a) Coworker sorting

(b) Between-share of wage inequality

(c) Between-firm wage inequality

(d) Productivity dispersion

Figure B.10: Bombardini et al. (2012) proxy for complementarity and industry-level outcomes

Notes. Each panel plots the moment indicated in the subtitle of the respective plot against the industry-level proxy for complementarity constructed from O*NET data following Bombardini *et al.* (2012). The unit of observation is a 4-digit NACE industry, with at least 5,000 person-year observations and where the industry-level proxy for complementarity is within two standard deviations. Industry-level observations are grouped into 30 bins. Industries are binned into 30 cells. The linear regression line is fitted based on unweighted, non-binned observations.

	Sorting		Complementarities	
Period	Spec. 1	Spec. 2	Spec. 1	Spec. 2
1985-1992	0.47	0.38	0.001	0.000
1993-1997	0.56	0.46	0.002	0.001
1998-2003	0.60	0.48	0.004	0.002
2004-2009	0.65	0.50	0.005	0.002
2010-2017	0.68	0.51	0.005	0.004

Table B.2: Coworker sorting & complementarities: results based on non-parametric ranking method

Notes. This table indicates, under the column "Sorting" the correlation between a worker's estimated type and that of their average coworker, separately for five sample periods. The column "Complementarities" indicates the point estimate of the regression coefficient β_c in regression (37). Under "Specification 1" workers are ranked economy wide, while under "Specification 2" they are ranked within two-digit occupations. Worker rankings are based on the non-parametric method described in Appendix B.2.1.2.

Next, Figure B.10 shows that the industry-level measure of the importance of teamwork, impact on coworker output, communication, and contact following Bombardini *et al.* (2012), discussed in Section 4.3, is positively associated with coworker sorting, the between-share of wage inequality, overall between-firm wage inequality, and productivity dispersion.

B.4.3 Results using non-parametric ranking algorithm

Table B.2 shows the evolution of coworker sorting and coworker complementarity when worker types that are estimated using the non-parametric method described in Appendix B.2.1.2, instead of relying on AKM-based fixed effects. In comparison, the levels of sorting are slightly higher and those of complementarities somewhat lower. The trends are qualitatively the same, though.

B.4.4 Complementarity over time: ranking workers within occupations

Figure B.3 reports the point estimates for the coefficient on the interaction term, β_c, in regression (37) for each of the five sample periods. In addition to the baseline, where workers are ranked economy-wide, results for the alternative specification, whereby workers are ranked within two-digit occupations, are included also. The results are discussed in the main text.

B.4.5 Measuring complementarity using years of education as a quality measure

As noted in Section 4.2, one concern regarding the evidence for coworker complementarity is that the dependent variable in regression (37) is the period-t wage but, in addition, the independent variables of interest (own type and average 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,

	Complementarity	
Period	Spec. 1	Spec. 2
1985-1992	0.0036	0.0022
1993-1997	0.0042	0.0031
1998-2003	0.0070	0.0045
2004-2009	0.0082	0.0054
2010-2017	0.0091	0.0058

Table B.3: Complementarity over time

Notes. This table indicates the point estimate of the coefficient β_c in regression (37), estimated separately for five sample periods. Under "Spec. 1" the worker ranking is economy wide, under "Spec. 2" it is within-occupation.

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 when worker quality is proxied by education – similar to Nix (2020) – and thus a non-wage characteristic, instead, the main results remain robust.

	'85-'92	'93-'97	'98-'03	'04-'09	'10-'17
Interaction	0.0063***	0.0060***	0.0099***	0.0112***	0.0129***
	(0.0008)	(0.0007)	(0.0008)	(0.0007)	(0.0009)
Obs. (100,000s)	3,613	2,508	2,694	3,836	4,376
R^2	0.5033	0.5451	0.5746	0.6330	0.6425

Table B.4: Regression evidence on coworker complementarity using years of schooling

Notes. Dependent variable is the wage level over the year-specific average wage. Independent variables are a constant, years of schooling, coworker years of schooling, and the interaction between those two terms. All regressions include industry-year, occupation-year and employer fixed effects. Employer-clustered standard errors in parentheses. Observations are unweighted. The sample is unchanged from the main text, except that 96,517 observations with missing years of schooling are dropped. Observation count rounded to 100,000s.

Specifically, I adapt equation (37) by regression the wage w_{it} of worker i in period t on their years of education, denoted s_i, their coworkers' average years of education, denoted s_{-it}, and the interaction of these two terms:

$$\frac{w_{it}}{\bar{w}_t} = \beta_0 + \beta_s \hat{s}_i + \beta_{s'} s_{-it} + \beta_c \left(s_i \times s_{-it}\right) + \psi_{j(it)} + \nu_{o(i)t} + \xi_{s(i)t} + \epsilon_{it}, \tag{B.7}$$

where $s_{-it} = \frac{1}{|S_{-it}|} \sum_{k \in S_{-it}} \hat{s}_k$. To implement this regression, I follow Card *et al.* (2013) in estimating the years of education based on the level of training indicated in the SIEED (on which see Fitzenberger *et al.*, 2006). Notice that the interpretation of the regression coefficients changes accordingly compared to the baseline, in which worker quality is measured in decile ranks. As before, I estimate equation (B.7) separately for 5 sample periods.

Table B.4 summarizes the estimation results. It shows, firstly, that the regression coefficient of interest, β_c, is positive across all periods, and, secondly, points to an increase over time in line with the baseline estimates. The magnitude of the increase in β_c is similar in proportional terms to the estimates reported in the main text, rising by about 74% from the first to the final sample period.

B.4.6 Complementarity regression using lagged types

This section describes a robustness check to the estimation of wage complementarities based on equation (37). To rule out a mechanical relationship between wages, on the left-hand side, and (within-period time-invariant) types, estimated from wages themselves, we can take the following approach. We assign to each individual i in periods $p \in \{2,3,4,5\}$ the fixed effect estimated for i in period $p-1$, and then re-compute worker deciles and average coworker types, $\hat{x}_i^{p-1}$ and $\hat{x}_{-it}^{p-1} = (|S_{-it}|)^{-1} \sum_{k \in \mathcal{S}} \hat{x}_k^{p-1}$, on that basis. The correlation between $\hat{x}_i$ and $\hat{x}_i^{p-1}$ is 0.85, consistent with the idea that unobserved individual characteristics captured by the fixed effect are very persistent. Next, the original wage complementarity regression is re-estimated for each period $p \in \{2,3,4,5\}$, once using the baseline measure of types and once using the lagged types. To compare like with like, both specifications are estimated on the same subsample of individuals for whom a type could be estimated in the preceding period.

Table B.5 reports on the results. Throughout, the coefficient on the interaction terms is positive but estimated to be roughly 50% smaller, on average, when using lagged types, compared to the baseline. It is difficult to gauge to what extent this difference in magnitude reflects absence of endogeneity issues from simultaneity or greater attenuation bias due to measurement error when using lagged types. Reassuringly, though, the evolution over time is similar.

B.4.7 Results when worker types are constructed without residualizing for observables

This section addresses the concern – noted in Footnote 45 — that, in the baseline, worker types were constructed from wages residualized for individual-level observables like age and tenure that, from a production perspective, we may not want to control for.

As a robustness measure, I re-estimate the AKM models without controlling for polynomials in age and tenure, then construct worker and coworker types from the worker FEs thus obtained. The correlation between baseline type measure and this alternative version is 0.83; the same comparison for the coworker types is higher still, at 0.93.

Turning to labor market outcomes, Table B.6 compares the evolution of coworker sorting using the baseline type measures and the robustness check. As can be seen, the two approaches

	'93-'97		'98-'03		'04-'09		'10-'17	
	Baseline	Alt.	Baseline	Alt.	Baseline	Alt.	Baseline	Alt.
Interaction	0.0040***	0.0019***	0.0061***	0.0027***	0.0074***	0.0043***	0.0069***	0.0038***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
R^2	0.8027	0.7417	0.7972	0.7361	0.8254	0.7656	0.8173	0.7471
Obs. (100,000s)	1,986		1,959		2,807		3,050	

Table B.5: Evidence on coworker complementarity: robustness check with lagged types

Notes. The columns labelled "Alt." represent regressions using lagged types, compared to "Baseline" estimated on the same sample. The dependent variable is the wage level over the year-specific average wage. Independent variables are a constant, worker type, coworker type, and the interaction between those two terms. All regressions include industry-year, occupation-year and employer fixed effects. Employer-clustered standard errors are given in parentheses. Observation count rounded to 100,000s.

	Coworker sorting	
Period	Baseline	Alternative
1985-1992	0.427	0.433
1993-1997	0.458	0.479
1998-2003	0.495	0.518
2004-2009	0.547	0.572
2010-2017	0.617	0.609

Table B.6: Coworker sorting: robustness check of constructing types without residualizing for observables

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. In the "Baseline" column, worker types are estimated by controlling for time-varying individual-level observables, whereas in the "Alternative" those controls are omitted. In both cases, the AKM method is used and the worker ranking is performed within the economy as a whole.

yield extremely similar results. In addition, when re-estimating the complementarity regression (37) using raw log wages on the left-hand side, and including alternative worker- and coworker-type measures as well as age and tenure polynomials as separate covariates, the findings are similar to those reported in Table 2.

In summary, though the treatment of time-varying individual-level observables like age and tenure in the mapping between theoretical and empirical coworker types is ambiguous, in practice, sorting and complementarity patterns are very similar irrespective of whether those observables are held constant or not.

C Quantitative analysis

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. I proceed in four main steps. First, I convert the spell-level data into a monthly panel. Second, I restrict the sample to approximate the selection criteria used in the other empirical analysis but without being limited to full-time employees. Specifically, 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 validate the identification of the vector of jointly estimated parameters, ψ, I conduct two exercises, following Bilal *et al.* (2022). First, to support the argument laid out in the main text that each element of 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. For the second exercise, let a given parameter ψ_i vary around the estimated value ψ_i^* and plot the distance function $\mathcal{G}(\psi_i, \psi_{-i}^*)$, where the remaining parameters are allowed to adjust to minimize $\mathcal{G}$. Figure C.2 indicates that $\mathcal{G}(\psi_i, \psi_{-i}^*)$ has a steep U-shape, suggesting that ψ is indeed well-identified.

C.1.3 Between-share adjustment procedure

Since in the structural model the number of workers in each production unit is two, the between-unit share of the variance of wages – or, for that matter, that of types – will be greater than zero even under random matching. More generally, for any degree of coworker sorting less than unity, i.e. $\rho_{xx} < 1$, the *level* of the between-share in a model with team size $n = 2$ will be biased upward relative to the case of $n > 2$ and, in particular, $n \to \infty$. These observations are simply an outgrowth of the law of large numbers not applying within production units. As such, it is a statistical phenomenon rather than an economically interesting mechanism. Furthermore, the upward bias is *greater* when the coworker correlation is *lower*. One can immediately verify this result intuitively by noting that for $\rho_{xx} = 1$, *all* dispersion is *across* and *none* within units, regardless of the value of n. Figure C.3 illustrates these ideas graphically (its construction is described below).

This statistical bias has two implications. First, without any further adjustment, the *level* of the between-share predicted by the estimated model, which assumes $n = 2$, will be excessively

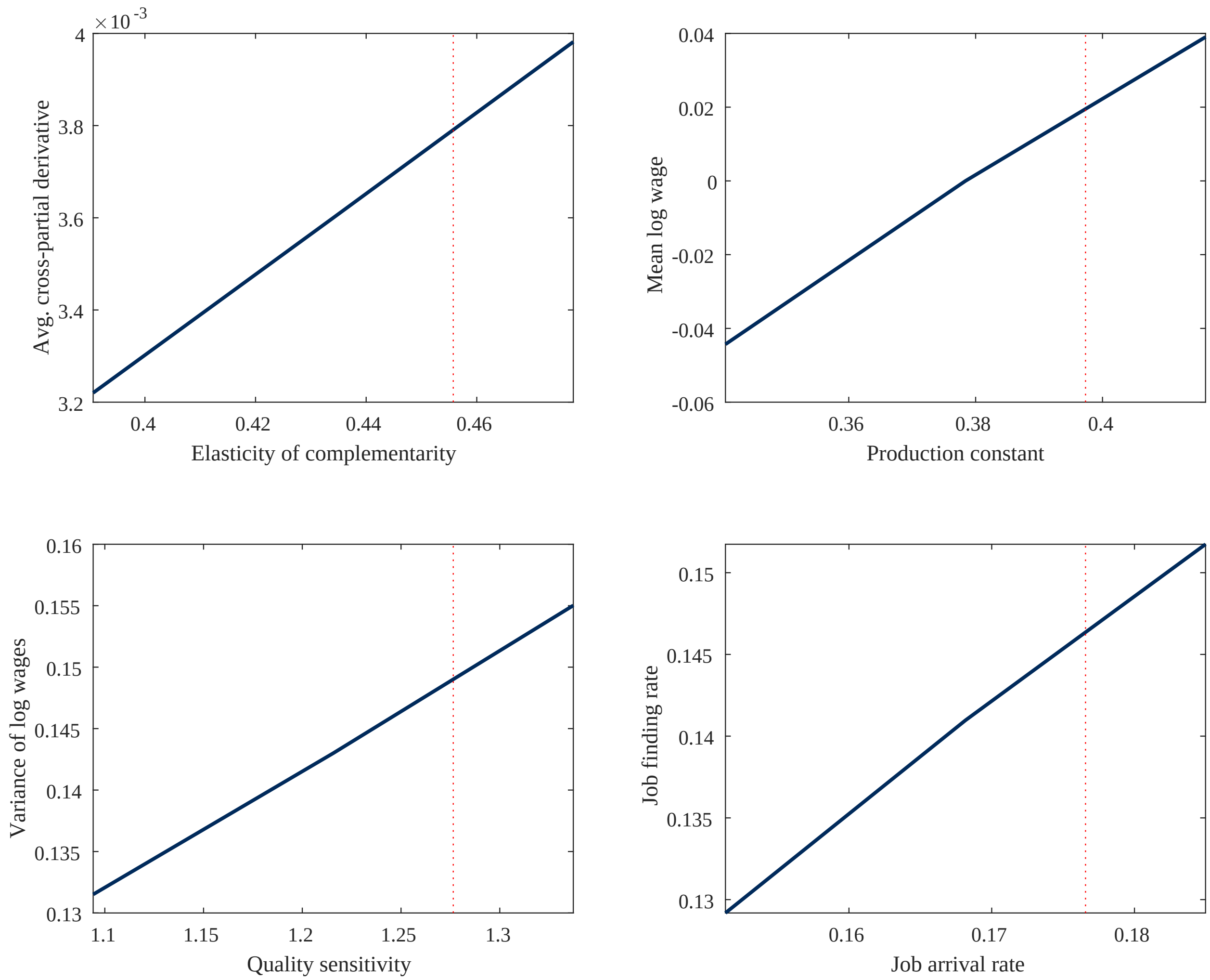


Figure C.1: Validation of identification method: moment against parameter

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

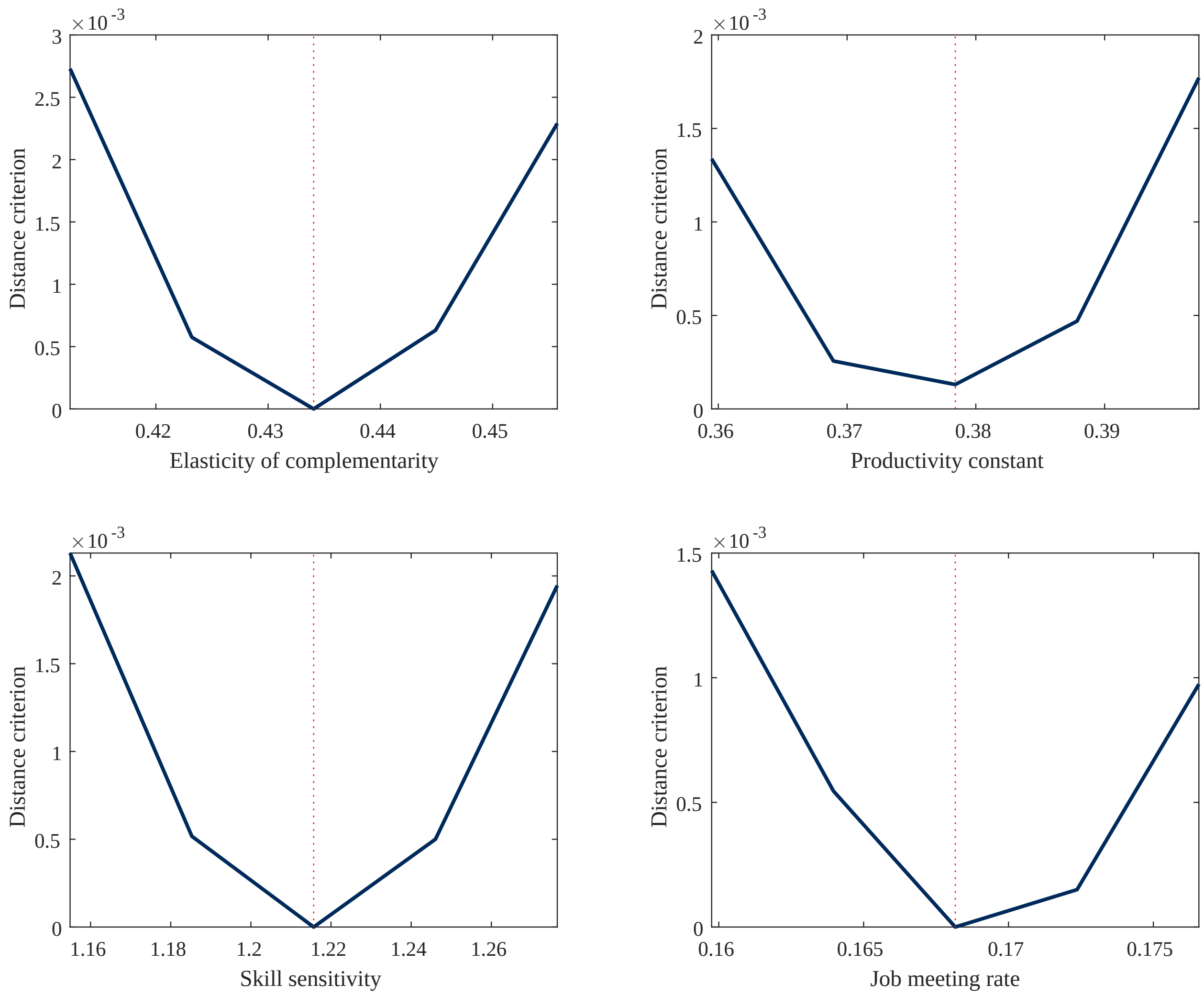
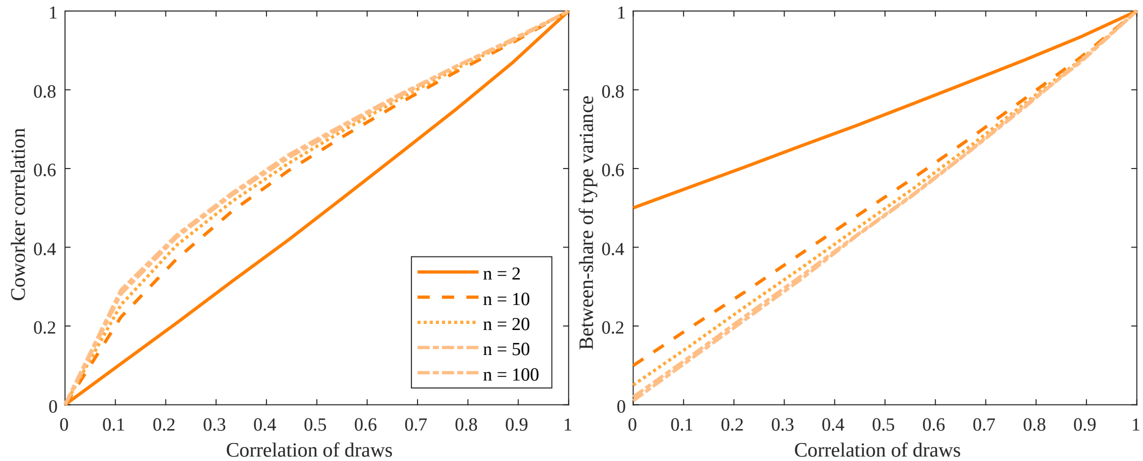


Figure C.2: Validation of identification method: distance criterion

Notes. This figure plots the distance function $\mathcal{G}(\psi_i, \psi^*_{-i})$ when varying a given parameter ψ_i around the estimated value ψ^*_i. The remaining parameters are allowed to adjust to minimize $\mathcal{G}$.



(a) Coworker sorting

(b) Between-share of type var.

Figure C.3: 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.

high relative to the real world, in which $m > 2$. Second, insofar as the estimated model predicts greater stronger coworker sorting than the earlier period, the predicted increase in the between-share of wage inequality is a *lower bound*, because the statistical upward bias in the later period will be smaller than in the later period. This section proposes an approach to correct the level of the between-share, but all results for *changes over time* that are reported in the main text are a conservative estimate, insofar as I apply the same correction factor to both periods.

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 $\Sigma(\rho_c)$. To map this onto our empirical 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.

This setup affords us with a closed-form description of the population between-team share of the variance of types as a function of n and ρ_c. Since the marginals of the Gaussian copula are simply continuous uniforms defined over the unit interval, 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,

$\bar{X}$. That is, 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 $\sigma^2_{x,\text{between-share}}(\rho_c, m) = \frac{1}{m}\left(1 + (m-1)\rho_c\right)$.

In the main text, and letting 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 $= \frac{1}{2}\left(1+\rho_c\right) - \frac{1}{\hat{n}}\left(1 + (\hat{n}-1)\rho_c\right)$,

The value of ρ_c I feed into this formula is the average coworker correlation value in the period-1 sample.[C.1] The implied value of the correction factor is approximately 0.25.

A concern with this approach is that ρ_c is not the same measure as the coworker correlation, ρ_{xx}, that we typically consider in both the empirical analysis and the structural model. 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 our 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.3 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 $\rho_{xx} > \rho_{cc}$.[C.2]

Another second concern is that the adjustment approach pertains, strictly speaking, to 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 seems to accomplish the desired goal.

[C.1] I choose to base the correction on the earlier sample, yielding a bigger downward adjustment, to avoid overstating the degree of between-firm inequality that the model can generate without assuming ex-ante firm heterogeneity. In addition, I use $\hat{n} = 15$. (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.)

[C.2] To match the structural model as closely as possible, I binned the draws in the same way as I did in the structural model and for empirical analyses. Of course, the statistical environment makes it possible to examine the implications of such binning and, reassuringly, it makes little difference to the results.

Parameter	Baseline	Within-occ ranking	OJS	OJS: constant EE rates
ρ		0.008		
ω		0.5		
a_2		1.1		
δ		0.0071 | 0.0085		
γ	0.434 | 0.836	0.277 | 0.537	0.233 | 0.315	0.215 | 0.728
a_0	0.378 | 0.238	0.397 | 0.239	0.400 | 0.274	0.397 | 0.268
a_1	1.216 | 1.559	1.194 | 1.566	1.216 | 1.559	1.204 | 1.550
$\bar{b}$	0.739 | 0.664	0.742 | 0.669	0.720 | 0.654	0.721 | 0.648
λ_u	0.168 | 0.230	0.155 | 0.214	0.152 | 0.179	0.152 | 0.189
λ_e	0	0	0.054 | 0.103	0.057 | 0.063

Table C.1: Model parameters: baseline and robustness

Notes. This table summarizes the parameterization of the model under alternative specifications. The value on the left side of | is for period 1, that on the right side for period 2. Discount rate, bargaining weight, team benefit, and separation rates are identical across specifications.

C.2 Additional results and robustness checks

C.2.1 Re-estimation of model under alternative specifications

Table C.1 summarizes the parameterization of the model under the different specifications considered in Section 5. Note that four parameters, namely ρ, ω, a_2 and δ are the same across specifications.

C.2.2 Version with on-the-job search

SETUP. This section briefly summarizes the version of the model with on-the-job search (OJS). In terms of setup, an employed worker may now meet vacancies at a Poisson rate λ_e. The unconditional rate at which a vacancy, in turn, meets an employed worker of any type is denote $\lambda_{v.e} = \frac{\lambda_e e}{v}$. The equations become a bit lengthy, so I restrict myself to giving two examples, with a complete description available upon request.

For example, the joint value of a firm and a worker of type x is now

$$\begin{aligned}
\rho\Omega_1(x) &= f_1(x) - \delta S(x) \\
&\quad + (1-\omega)\Bigg[\underbrace{\left(\lambda_{v.u}\int \frac{d_u(\tilde{x}')}{u} S(\tilde{x}'|x)^+ d\tilde{x}'\right)}_{\text{meet unemployed}} \\
&\quad + \underbrace{\left(\lambda_{v.e}\left(\int \frac{d_{m.1}(\tilde{x}')}{e}\{S(\tilde{x}'|x) - S(\tilde{x}')\}^+ d\tilde{x}' + \int\int \frac{d_{m.2}(\tilde{x}',\tilde{x}'')}{e}\{S(\tilde{x}'|x) - S(\tilde{x}'|\tilde{x}'')\}^+ d\tilde{x}' d\tilde{x}''\right)\right)}_{\text{meet employed}}\Bigg] \\
&\quad + \omega\lambda_e\left(\int \frac{d_{f.0}}{v} \times \underbrace{\{S(x) - S(x)\}^+}_{x \text{ joins different 0-worker firm}} + \int\int \frac{d_{m.1}(\tilde{x}')}{v} \times \underbrace{\{S(x|\tilde{x}') - S(x)\}^+}_{x \text{ joins different 1-worker firm}} d\tilde{x}'\right).
\end{aligned}$$

Turning to flows, in the stationary equilibrium the 'density' of matches consisting of a firm with two workers of types x and x' satisfies the following balance condition, where for instance $h_{2.1}(x, x''|x') = \mathbf{1}\{S(x|x') - S(x|x'') > 0\}$ indicates whether a worker x in a two-worker firm with coworker x'' would move to an employer that currently has one employee of type x'.

$$\begin{aligned}
& d_u(x)\lambda_u \frac{d_{m.1}(x')}{v} h_{u.1}(x|x') \\
& + d_{m.1}(x)\Bigg\{\lambda_e \frac{d_{m.1}(x')}{v} h_{1.1}(x|x') + \lambda_{v.u}\frac{d_u(x')}{u} h_{u.1}(x'|x) + \lambda_{v.e}\Bigg(\frac{d_{m.1}(x')}{e} h_{1.1}(x'|x) \\
& \quad + \int \frac{d_{m.2}(x',\tilde{x}'')}{e} h_{2.1}(x',\tilde{x}''|x) d\tilde{x}''\Bigg)\Bigg\} \\
& + \int d_{m.2}(x,\tilde{x}'')\lambda_e \frac{d_{m.1}(x')}{v} h_{2.1}(x,\tilde{x}''|x') d\tilde{x}'' \\
& = d_{m.2}(x,x')\Bigg\{2\delta + \lambda_e\left(\int \frac{d_{f.0}}{v} h_{2.0}(x,x') + \int\int \frac{d_{m.1}(\tilde{x}'')}{v} h_{2.1}(x,x'|\tilde{x}'') d\tilde{x}''\right) \\
& \quad + \lambda_e\left(\int \frac{d_{f.0}}{v} h_{2.0}(x',x) + \int\int \frac{d_{m.1}(\tilde{x}'')}{v} h_{2.1}(x',x|\tilde{x}'') d\tilde{x}''\right)\Bigg\}.
\end{aligned}$$

IMPLICATIONS FOR MATCHING PATTERNS. Figure 10 seems to indicate that the version with OJS does a better job at capturing empirical coworker sorting patterns. Here I offer a brief explanation. In essence, with OJS matching patterns better reflect the heterogeneity in surplus across matches. Absent OJS, match outcomes do not fully reflect preferences over these outcomes. Specifically, when the match outcome is a binary comparison of the values from being unmatched or matched, then these outcomes do not express information about who among the accepted matching partners I would rather match with; and who among the rejected ones I like even less than the other. By contrast, with OJS, preferences over match outcomes are also encoded in job-to-job transitions, which are determined by a comparison of match surpluses at alternative employers, respectively with different coworkers. These moves thus contain additional information about heterogeneity in surplus among the partners inside the acceptance set coming out of unemployment. In that sense, the stationary match distribution better reflects preferences over match outcomes.

TRADE-OFFS AND FUTURE RESEARCH. While incorporating OJS improves the model's capacity to fit empirical coworker sorting patterns, this comes at a twofold cost. First, unlike the baseline model, the version with OJS predicts a *level* of the between-firm share of wage inequality that exceeds what is observed in the data (even after correcting for a mechanical upward bias discussed in Appendix C.1.3). While this does not impair the model's ability to speak to *changes* over time, it does highlight a shortcoming of the model. I conjecture that this limitation partly reflects that the model does not incorporate a vertical dimension of firm organization as in Garicano and Rossi-Hansberg (2006), which would plausibly generate more within-firm wage dispersion. Future work that takes serious within-firm worker heterogeneity both across and within layers of the firm, perhaps by modelling the interaction between multiple teams, might improve on this dimension.

The second issue relates to the mapping between production and wage complementarities introduced in Corollary 3. While "origin" effects that differentiate workers' in terms of their bargaining position (e.g., unemployment vs. jobs yielding varying levels of surplus) leave this mapping unaffected – at least under the bargaining protocol assumed here – a different issues arises relating to anticipation effects. Specifically, the interaction between own type and coworker type influences the wage not only *directly* through the production value but also *indirectly*. The indirect effect arises because the coworker match quality influences the probability that either team member switches jobs conditional on receiving an outside offer. I next show this more carefully. Under the assumption that wages are continuously renegotiated under Nash bargaining, with unemployment as a worker's outside option, it is still possible to characterize the wage of a worker of type x employed at a firm with another worker of type x' in closed form:

$$\begin{aligned} w(x|x') = {} & \rho V_u(x) + (\rho + 2\delta)\omega S(x|x') \\ & - \omega S(x) \times \left(\delta + \lambda_e \int \frac{d_{f.0}}{v} h_{2.0}(x', x) + \lambda_e \int \int \frac{d_{m.1}(\tilde{x}'')}{v} h_{2.1}(x', x|\tilde{x}'') d\tilde{x}'' \right) \\ & - \omega \lambda_e \int \frac{d_{m.1}(\tilde{x}'')}{v} h_{2.1}(x, x'|\tilde{x}'') S(x|\tilde{x}'') d\tilde{x}'' \end{aligned}$$

Here, the term $h_{2.1}(x', x|\tilde{x}'')$ is directly shaped by $S(x|x')$ and similarly for $h_{2.1}(x, x'|\tilde{x}'')$. Hence, $\frac{\partial^2 w(x|x')}{\partial x \partial x'}$ moves more than proportionately with $\frac{\partial^2 f(x|x')}{\partial x \partial x'}$ insofar as $\lambda_e > 0$. What underlies this effect is the assumption that wage-setting is very forward-looking, with wages fully reflecting already today the expected change in future values of employer and worker(s) due to future events such as exogenous job separations or endogenous job moves. The reasoning in Hall and Milgrom (2008), among other things, leads me to doubt the empirical force of such forward-looking effects. But introducing a "partially myopic", alternative wage sharing rule à la Hall and Milgrom (2008) is difficult to reconcile with match formation being privately efficient and based on joint surplus maximization. I view this as a potentially important albeit challenging direction for future theoretical work, alongside an exploration of alternative bargaining protocols (e.g., Cahuc *et al.*, 2006).

Of course, in the structural estimation it is still possible to treat $\frac{\partial^2 w(x|x')}{\partial x \partial x'}$ as moment that is highly informative about the strength of production complementarities. That is indeed how I proceed in the re-estimation of an extended model that includes OJS. Since, however, the estimation targets an *increase* in EE rates from 0.0076 to 0.0106, matching which requires a rise in λ_e, this procedure yields a smaller increase in the estimated strength of *production* complementarities than the models without OJS do. The final column in Table C.1 shows that if, instead, the EE rate is kept constant, then the model with OJS predicts an increase in γ that is very similar to the baseline model. Under this calibration, the model predicts 12.5 percentage point increase in the between share from period 1 to period 2, of which only 2.1 points would have occurred had the estimated value of γ remained at the period-1 level. Thus, in this model complementarities rationalize around 45% of the empirically observed rise in the between-firm share of wage inequality, a value that is very similar to the baseline model.

C.2.3 Extension: $n_{\max} = 3$

In the baseline model, firms can have zero, one or, at most, two employees. The associated assumption of extreme decreasing returns to scale setting in at $n_{\max} = 2$ facilitates a transparent discussion of the key link between complementarities and matching decisions, but it is primarily imposed for reasons of tractability. In particular, as the number of potential employees increases, the potential combinations of worker types and, hence, the state space explodes.

This section shows that the key model properties are robust to allowing at least for $n_{\max} = 3$.[C.3] In particular, Figure C.4 plots the average coworker quality for different types of workers, under alternative model specifications. Depicted in solid grey is the baseline model; in dashed orange the model with $n_{\max} = 3$, considering teams of size 2 or 3; and in dashed-dotted light orange the model with $n_{\max} = 3$ but considering only teams of size 2. As can be seen, the equilibrium sorting implications are very similar when raising the maximum team size.

[C.3] The equations describing the stationary equilibrium become messier but are conceptually straightforward extensions of the $n_{\max} = 2$ case.

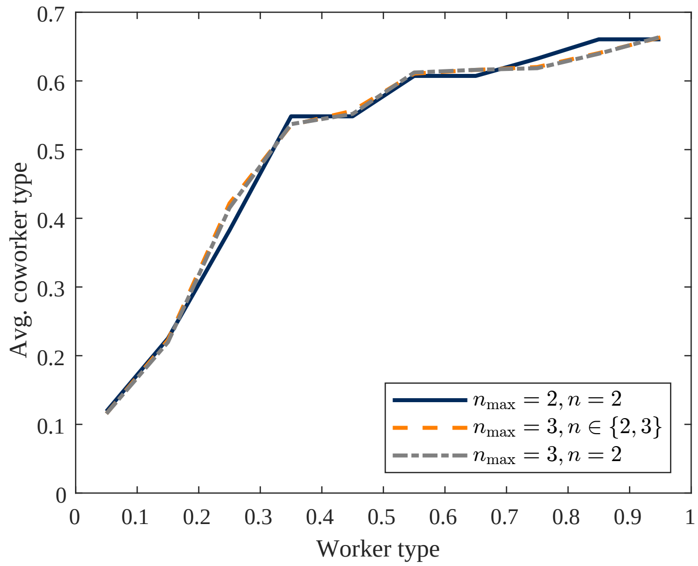


Figure C.4: Equilibrium coworker sorting patterns when $n_{\max} = 3$

Notes. This figure plots the average coworker type (vertical axis) for different worker types (horizontal axis), as implied by the equilibrium matching patterns emerging from the baseline model with maximum team size 2 ($n_{\max} = 2$) and the alternative model with maximum team size 3 ($n_{\max} = 3$).

C.2.4 Job-to-job transitions and coworker sorting

This section documents how, in the data, workers transition across jobs characterized by different coworker quality types. It also compares these patterns to the predictions of the calibrated version of the model version with OJS, considering also how these movements have changed over time.

According to the theoretical model, job-to-job (EE) transitions tend to move workers in a surplus-maximizing direction, which in the presence of coworker complementarities means that they tend to increase sorting. Hence, the *change* in coworker quality (defined as $\Delta\hat{x}_{-it} = \hat{x}_{-i,t} - \hat{x}_{-i,t-1}$), should be *positively* correlated with $\hat{x}_i$.[C.4]

To evaluate this prediction, I first transform the spell-level version of the SIEED into a *monthly* worker-originMonth-destinationMonth-originJob-destinationJob dataset, which also contains information on the characteristics of origin and destination job, including in particular the average coworker quality. For computer memory management purposes on the administrative data servers, I limit this analysis to the first and the last sample period, i.e., the periods 1985-1992 and 2010-2017. I then consider an individual as undergoing an EE transition if they are employed at establishment j in month t and $j' \neq j$ in month $t+1$.

Turning to the analysis, Figure C.5 describes the properties of EE transitions by worker type in a non-parametric manner. It plots, for each worker type and origin coworker type, the average

[C.4] In the theoretical model, $h_{2.1}(x, x''|x') = 1$, so that a worker x in a two-worker firm with coworker x'' would move to an employer that currently has one employee of type x', if and only if $S(x|x') - S(x|x'') > 0$.

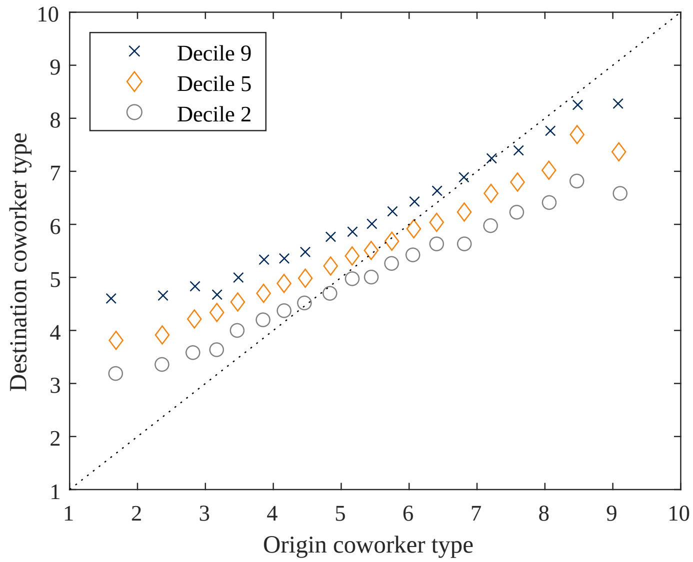


Figure C.5: Origin-destination plot by worker type

Notes. This figure plots, for each origin coworker type (decile; horizontal axis), the average coworker destination type (decile; vertical axis), separately for worker deciles 2, 5, and 9. Figures are averaged across 1985-1992 and 2010-2017.

coworker destination type, computed from the pooled sample. Consistent with the prediction of the theoretical model, EE transitions tend to lead greater coworker sorting: for any given origin, high-type move to workplaces with better coworkers than lower-type workers do.

Next, I use a regression approach to answer the following question: Have EE transitions become *more* like to reallocate workers in a direction consistent with positively assortative coworker matching? To this end, I regress $\Delta\hat{x}_{-it}$, scaled by the standard deviation of of coworker quality changes, σ_Δ, on the worker type, controlling for origin coworker type as well as year effects.

$$\frac{\Delta\hat{x}_{it}}{\sigma_\Delta} = \beta_0 + \beta_1\hat{x}_i + \beta_2\hat{x}_{-i,t-1} + \eta_t + \epsilon_{it}.$$

The second main column displays the *model-implied* coefficient of interest, β_1, for both sample periods. As complementarities have increased over time, β_1 be greater in 2010-2017 than in 1985-1992. In the first main column, this prediction is empirically tested, showing that empirically both the sign and change over time are qualitatively consistent with the theoretical prediction. Quantitatively, the magnitude of the estimate, $\hat{\beta}_1$, lies below the model prediction in both periods. This points to a limitation of the model, wherein *all* EE transitions reallocate toward more positive sorting, whereas in the data there are also countervailing flows. To improve the model's fit with the data, incorporating match-specific productivity shocks as well as idiosyncratic preference shocks would be useful.

	Data		Model	
Change in coworker type	’85-’92	’10-’17	Period-1	Period-2
Own type	0.0883***	0.118***	0.214	0.270
	(0.000799)	(0.000918)		
Controls	Year FEs, Origin	Year FEs, Origin	Origin	Origin
Obs.	196,100	282,700	∞	∞
Adj. R^2	0.284	0.204		

Table C.2: Change in coworker type due to EE moves is positively related to own type

Notes. For the data columns, individual-level clustered standard errors are given in parentheses; model counterparts are computed simulation-free in population. The dependent variable is scaled throughout by the standard deviation of the change in coworker type. Observation count rounded to 100s. The model parametrization is summarized in Table C.2.1.

D The task content of production in Germany

To document patterns in the task content of production I draw on the Employment Surveys (ES) carried out by the German Federal Institute for Vocational Training (Bundesinstitut fuer Berufsbildung, BIBB; Hall and et al. (2018)).[D.1] Following the influential methodology introduced in Autor *et al.* (2003) and first applied to the ES by Spitz-Oener (2006), I use the tasks that employees report to have performed to measure variations in the nature of work.

The BIBB surveys have several attractive features: they provide detailed information on tasks performed at work; the survey has been run, in repeated waves, since 1985 (1985/86, 1991/92, 1998/99, 2006, 2012, and 2018), facilitating time series analyses; each wave has a large sample size between 20,000 to over 30,000 respondents per wave, facilitating between-group comparisons; responses are at the *worker-level* and consistent occupation codes can be used across multiple waves, making it possible to capture changes in nature of work not only associated with employment shifts across occupations but also within-occupation (on the importance of which see, e.g., Spitz-Oener (2006); Atalay *et al.* (2020)); and a supplemental survey in 2012 allows enriching binary task indicators with information on the actual shares of time spent by employees in different occupations on various tasks.

The analysis uncovers several key findings, which I summarize next before discussing details:

(i) The aggregate usage share of complex tasks in workers’ activities has monotonically risen since 1985/86; the 1990s saw a particularly sharp increase.

(ii) This trend is prevalent across different levels of education. It is not driven by occupational employment effects either, instead the majority of the increase occurs within-occupation.

[D.1] I thank the Research Data Center of the BIBB for providing access to scientific use files as well as guidance. All remaining errors are my own.

(iii) In the cross-section, the task portfolio of more educated individuals tends to be disproportionately skewed toward complex tasks compared to less educated individuals. The ranking of different occupations is intuitive and likewise reveals large variation in task shares.

(iv) Cross-sectional differences are robust to using measures of time spent on different tasks.

Altogether, these results provide reassurance that the time trend is robust and not merely the result of composition effects, and that there is substantial cross-sectional variation that can be exploited as done in the main text.

D.1 Methodology

SAMPLE RESTRICTIONS. As detailed in Rohrbach-Schmidt and Tiemann (2013) and Hall and Rohrbach-Schmidt (2020), time comparisons with the BIBB/IAB surveys require a standardization of the sample basis. To that end, I follow the steps detailed in those reports and focus on employees from West Germany, aged 15 to 65, who belonged to the labor force (defined as having a paid employment situation) with a regular working time of at least ten hours per week.[D.2] The final sample comprises 91,152 worker-year observations.

In addition, I entirely omit the 1998/99 wave from my analysis, because the number of activities queried in that wave is substantially lower than in the other surveys. While doing so reduces the overall sample size, this choice avoids bias to the results due to the limited comparability in tasks. For example, none of the activities "accommodating", "caring", "storing", "protecting", "programming" and "cleaning" were queried in the 1998/99 survey.[D.3]

TASK CLASSIFICATION. As the careful discussion in Rohrbach-Schmidt and Tiemann (2013) makes clear, comparisons over task intensities using the BIBB ES over time need to be implemented carefully and account for variation over time in what tasks are queried and whether their content has changed in meaning. Writing in the context of typical studies that compare task items in the categories non-routine analytical, non-routine interactive, non-routine manual, routine-cognitive and routine-manual, the authors highlight in particular that routine-cognitive tasks are difficult to classify (e.g. "measuring" may be routine-cognitive or routine-manual; also see the findings of Antonczyk *et al.* (2009) in comparison to those by Spitz-Oener (2006)).[D.4]

[D.2] In addition, I drop observations who report having performed none of the activities queried in at least two waves. Given the extensive use of occupational codes, I also drop any occupations with fewer than thirty observations across all waves.

[D.3] I thank Daniela Rohrbach-Schmidt for her generous advice on how to handle the older waves and for sharing programs illustrating how the scientific use files can be rendered maximally consistent with the original data.

[D.4] Autor and Handel (2013) also treat the "physical" dimension of tasks as a combined measure of physical and routine tasks. Meanwhile, Acemoglu and Autor (2011) subsume non-routine analytical and non-routine interactive into "abstract", while routine-cognitive and routine-manual tasks are subsumed into "routine".

Given my focus on non-routine or complex tasks, these classification problems are less severe though, as “these items are regularly observable throughout the cross-sections, their content did not change significantly from year to year, and measurement validity is comparatively strong,” as Rohrbach-Schmidt and Tiemann (2013) note when suggesting to researchers to focus on the increase in these tasks.

Task classification	**Task name**	**Description**
Complex/abstract	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 D.1: Classification of tasks in the BIBB Employment Surveys

Notes. This table summarizes the classification of tasks into two groups: “Complex/abstract” and ”Other.”

As summarized in Table D.1, I therefore collect task items in the index of abstract tasks — guided by the classification of non-routine tasks in Spitz-Oener (2006), Rohrbach-Schmidt and Tiemann (2013) and Atalay *et al.* (2020) – and compare those with all other tasks, i.e., those that are broadly categorized as routine or manual.[D.5]

TASK INDEX. Given this classification, I then define an index capturing the usage of abstract/-complex tasks for worker i in period t, following Antonczyk *et al.* (2009):

$$T_{it}^{\text{abstract}} = \frac{\text{number of activities performed by } i \text{ in task category "abstract" in sample year } t}{\text{total number of activities performed by } i \text{ in sample year } t}$$

[D.5] I do not use the task items “managing”, “applying law” and “negotiating”, because they are only measured in the early waves. Moreover, I associate buying/selling with “other”, since even though they may be hard to automate (even that seems questionable in light of self-checkouts and e-commerce), they are arguably not among the most complex activities. This decision makes no practical difference to the results.

To illustrate, if worker i performs five distinct activities in sample period t and two of those belong to the category of abstract/complex tasks, then the complexity index for her work is 0.4.

OCCUPATIONAL CLASSIFICATION. To ensure a consistent classification of occupations when using information from multiple waves, I use the German Classification of Occupations 1988 (KldB88). As the oldest classification available in the two most recent waves (2012 and 2018) is the KldB92 classification (Hall and Rohrbach-Schmidt, 2020, cf. Table 9), in processing these two waves I rely on a conversion table KldB92→KldB92; the conversion quality is high as the two classifications are very similar.[D.6]

D.2 Results

PATTERNS OVER TIME. The first column in Table D.2 indicates that the aggregate usage share of complex tasks in workers' activities has monotonically increased from 1986 to 2018, with the increase being particularly pronounced in the first half of the time period.

The second to fourth columns decompose the period-by-period change in the importance of complex tasks into two components: a "between" component that captures shifts in occupational employment shares and a "within" component that measures changes in the task content within occupations. Formally, as in Atalay *et al.* (2020), I decompose changes in the usage of abstract tasks between periods t and $t-1$ according to the equation

$$\Delta \bar{T}_t^{\text{complex}} = \sum_o \omega_{o,t-1}(\bar{T}_{t,o}^{\text{abstract}} - \bar{T}_{t-1,o}^{\text{complex}}) + \sum_o (\omega_{o,t} - \omega_{o,t-1})\bar{T}_{t,o}^{\text{abstract}}$$

where $\bar{T}_{t,o}^{\text{complex}}$ measures the average usage of complex tasks by members of occupation o in period t and $\omega_{o,t}$ is the period- t employment share of occupation o.

Consistent with the findings of Atalay *et al.* (2020) for the US, this decomposition reveals that about three quarters of the increase in complex tasks over the sample period have occurred *within* occupations.

Education offers an alternative lens through which to view the changing task content. As shown already in Figure 2a in the main text, the share of complex tasks in the portfolio of university-educated individuals is substantially greater than that of persons with less formal education. The *increase* over time takes place across the board, however.

CROSS-SECTIONAL PATTERNS. The share of complex tasks also varies substantially by occupation. I compute the average complex-task shares at the ISCO-08 2-digit level in the waves 2012 and 2018. Table D.3 lists the bottom 5 and top 5 occupations. In addition, I also show the

[D.6]This crosswalk is based on the Klassifikationsserver der Statistischen Ämter des Bundes und der Länder, current occupations coded in the 2006 wave in which both KldB88 and KldB92 are available as well as personal judgements. I thank Anett Friedrich for creating and sharing the crosswalk.

	Total	Between	Within	Within-share
1986 level	0.252			
1986-1992	0.025	0.002	0.022	0.906
1992-2006	0.298	0.057	0.241	0.809
2006-2012	0.019	0.002	0.017	0.890
2012-2018	0.053	0.028	0.025	0.476
Total change	0.395	0.089	0.306	0.775

Table D.2: The evolving task content of production in Germany

Notes. This table reports the within-between occupation decomposition of the change in the share of complex tasks over time. The "Total" column aggregates across all individuals. The decomposition is performed at the level of KldB-1988 2-digit occupations.

non-routine abstract score from Mihaylov and Tijdens (2019) used in Section 4.3. The comparison reveals large and consistent variation in task shares across occupations according to either measure.

TIME USAGE. One concern is that the analysis only considered whether a given task represents an important activity in the respondent's job, as opposed to measuring how important that activity is relative to others. To address this concern, I draw on a supplemental survey from 2012 that precisely details the amount of time a subset of workers spent on the different tasks on a given day. Figure D.1 charts the shares of time spent on the seven abstract/complex activities for different occupational groups. Specifically, I rank occupations according to their task index and group them into 4 equally sized groups. Drawing on the supplemental survey, I then compute the average share of time members of these occupational groups spent on the various tasks. Figure D.1 shows that at least in more recent periods, each occupational group spends some time on such tasks as organizing or using the computer ("programming"). Crucially, the fraction of time spent on each of these tasks is multiples greater for the top quartile than for the bottom quartile. No one single task drives the overall increase in the complexity index.

ISCO-08 2-digit occupation	$\bar{T}_o^{\text{complex}}$	MT-NRA
Business and administration professionals	0.84	0.47
Legal, social and cultural professionals	0.83	0.67
Business and administration associate professionals	0.82	0.29
Teaching professionals	0.81	0.57
Administrative and commercial managers	0.81	0.58
...	...	...
Drivers and mobile plant operators	0.2	0
Agricultural, forestry and fishery labourers	0.14	0
Food preparation assistants	0.14	0
Market-oriented skilled forestry, fishery and hunting workers	0.12	0
Cleaners and helpers	0.12	0

Table D.3: Top- and bottom-5 occupations in terms of task complexity

Notes. This table reports the top-5 and bottom-5 ISCO-08 2-digit occupations when ranked by $\bar{T}_o^{\text{complex}}$ in pooled 2012 and 2018 waves. The column "MT-NRA" shows the non-routine abstract score taken from Mihaylov and Tijdens (2019) after collapsing to the ISCO-08-2d level using occupational employment shares (computed using the Portuguese data, in lne with the application in Section 4.3.

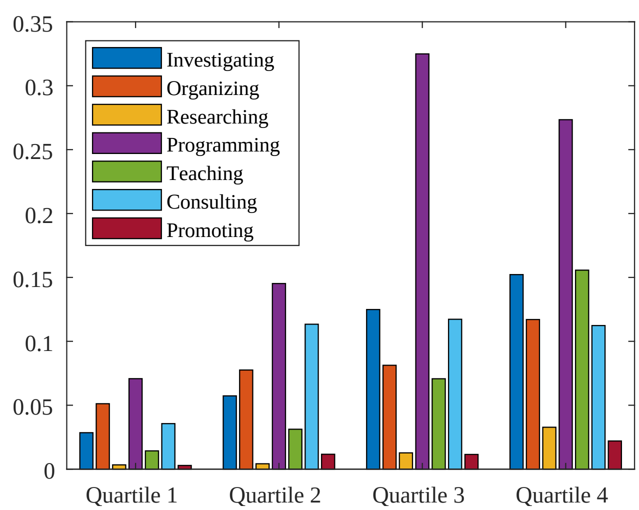


Figure D.1: Allocation of time to complex tasks by occupational groups

Notes. This table reports the share of time spent on complex tasks by different occupational groups. Occupations are first ranked according to their task complexity index and grouped into four groups of approximately equal size. Then the average share of time members of these occupational groups spend on the various tasks labelled as "complex" in Table D.1 is computed.