Firm Heterogeneity in Skill Returns

Michael J. Böhm, *Technische Universität Dortmund, IZA, and SHoF*
Khalil Esmkhani, *Simon Fraser University*
Giovanni Gallipoli, *University of British Columbia, CEPR, HCEO, and RCEA*

Abstract

We quantify firm heterogeneity in skill returns and present direct evidence of worker–firm complementarities. Within a model of firms' demand for cognitive and noncognitive attributes we show that identification depends on the availability of skill measures. Linking administrative data to test scores we document worker sorting and convex earnings–skill relationships. We find that: (1) Both skills' returns vary substantially across employers and correlate weakly within-firm. (2) Workers with large endowments of a skill populate firms with higher returns to it. Sorting intensifies with the cross-sectional dispersion of returns. (3) Complementarities and sorting significantly influence the earnings distribution.

We are grateful to Raffaele Saggio for a helpful discussion about bias-correction methods. Shuaib Habib, Ronit Mukherji and Dongxiao Zhang have provided valuable research assistance. Böhm acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, BO 4765/1-1 and BO 4765/1-2). Gallipoli acknowledges support from Canada's SSHRC.

1 Introduction

The recognition that earnings distributions reflect the interaction of worker and firm heterogeneity dates back decades (Willis, 1986). Access to matched employer–employee data has rekindled interest in such interaction (e.g., Card et al., 2013; Song et al., 2018; Sorkin, 2018; Lamadon et al., 2022). In this paper we present evidence on worker–firm complementarities, matching, and their effects on earnings. We do so by linking cognitive and noncognitive test scores with population data on Swedish workers and firms,[1] from which we recover estimates of firm-level returns to each skill attribute.

Various studies examine match effects within the boundaries of a single skill index (among others, Sørensen and Vejlin, 2013; Woodcock, 2015; Bonhomme et al., 2019; Lachowska et al., 2020; Lentz et al., 2023). Unlike previous work, our analysis emphasizes what can, and cannot, be identified when considering more than one skill dimension. If unobserved skills are collapsed into a single index, identification requires a connected worker–firm graph (sufficient mobility between employers) and that the average skill of the workers moving to a firm is not the same as that of workers moving out of the firm (a rank condition; see Bonhomme et al., 2019; Lamadon et al., 2022). These requirements are no longer sufficient with multiple skill dimensions because the ranking of workers is not unique. In such settings returns are identified under general conditions if skill proxies are available.

We build on these insights to study how cognitive and noncognitive skill returns vary across firms and document their impacts on sorting and wages. Estimates reveal significant heterogeneity in returns both within and across skills. Some firms pay up to 35 log points more than others for similar cognitive or noncognitive attributes. The multidimensional and bundled nature of skills plays a role in the imperfect assignment of skills to jobs (Rosen, 1983; Heckman and Scheinkman, 1987). That cognitive and noncognitive returns have independent impacts high-

[1] Several studies document the information content of our cognitive and noncognitive measures (for a well-known example, see Edin et al., 2022) and their relation to productive attributes. Lindqvist and Vestman (2011) show that they are highly significant in predicting earnings and unemployment conditional on any rich set of controls. Fredriksson et al. (2018) use them to identify the effects of job–skill mismatch on mobility and wage growth. See also Appendix A.1.

lights workers' inability to separately rent out their skills to the highest bidder (Lindenlaub, 2017; Edmond and Mongey, 2021; Choné and Kramarz, 2021; Skans et al., 2022).

Our findings provide novel evidence on the nature of returns to skills in the labor market. Previous work has shown that cognitive and noncognitive attributes shape individual outcomes (Heckman et al., 2006; Lindqvist and Vestman, 2011; Deming, 2017) and that the *average* gains from these skills have changed over time (Beaudry et al., 2016; Deming, 2017; Edin et al., 2022). We show that each worker's return depends on both skills and on their match with an employer. For this reason, conventional Mincerian measures of returns are not equivalent to averages of individual returns across employers. Rather, assortative matching tangibly alters skill premia, inducing nonmonotonic effects that are convex in skills.

In the baseline analysis we address limited mobility biases through clustering methods (Bonhomme et al., 2019) whereby we group firms into 100 classes based on the earnings and skills of employees. To assess robustness we estimate quadratic forms of the parameters of interest at the individual firm level (non-grouped) after correcting for biases (Kline et al., 2020). Each approach imposes different sample restrictions and assumptions. Results, however, are remarkably robust insofar as the importance of skill returns, *relative* to conventional measures of firm fixed effects, is stable and does not depend on implementation choices. Under either approach, estimation requires significant computational work, which we discuss below.

To motivate the focus on skill returns, we begin by estimating fixed-effect models separately for high versus low skill workers and we reject the hypothesis that earnings premia at a given firm are the same across skill levels (either cognitive or noncognitive). We confirm the presence of firm heterogeneity in skill returns through event studies where we track wage changes for different worker types as they move across employers. Informed by these findings, we use a monopsonistic model of labor demand (Card et al., 2018; Lamadon et al., 2022) to derive an empirical specification that allows for granular returns to each skill attribute. The specification is the first-order approximation of a general wage function: skill $\times$ firm interaction terms reflect heterogeneous returns, firm intercepts capture skill-independent premia, and standard Mincer returns are subsumed into worker fixed effects.

Our estimates reveal considerable dispersion of returns across firms in either skill dimension. We find even larger heterogeneity in robustness checks where we implement firm-level estimation with bias-correction or where we account for the impact of measurement error in skill measures. The correlation between returns to different skills is positive but weak; this suggests that collapsing cognitives and noncognitives into a single index would be restrictive. The assignment of workers to firms with heterogeneous returns generates earnings gaps of the same order of magnitude as those induced by heterogeneity in firm intercepts. Through variance decompositions (see appendix) we show that allowing for skill returns' heterogeneity boosts the overall firm-specific contribution to wage inequality.

To gauge the intensity of sorting we adapt methods developed for multidimensional assignment problems (Lindenlaub, 2017). We find that the assignment of more able workers to high return employers stochastically dominates, in first-order, the assignment of lower skilled workers (Lindenlaub and Postel-Vinay, 2023). Several testable restrictions are consistent with positive assortative matching, which occurs along both skill dimensions. Sorting is stronger in the cognitive dimension where returns heterogeneity is larger. Worker–firm complementarities and sorting lead to a skewed wage distribution: more skilled workers match to firms with higher returns to skills. This raises average earnings and magnifies the earnings of high-skill workers. In contrast, middle-to-low skill workers earn less because they are frequently matched with low-return firms. These results are of interest to a long-standing debate on worker–firm interactions (Becker and Chiswick, 1966; Sattinger, 1993; Lindenlaub, 2017; Hagedorn et al., 2017; Bonhomme et al., 2019; Borovickova and Shimer, 2020; Lamadon et al., 2022; Lentz et al., 2023). Skill proxies facilitate the measurement of gains from matching because pecuniary returns are not, themselves, used to determine skills ranks. This helps establish which workers (and skill bundles) benefit or lose from returns heterogeneity.

In keeping with our emphasis on direct measures, we probe the nature of firm differences by linking balance sheet data to the main sample and show that employers who exhibit high cognitive returns have different capital composition, with more intangible and intellectual assets (as opposed to physical capital) per worker. Survey responses indicate that these firms invest more in

R&D and innovate frequently. This lends support to the view that production and organizational arrangements shape the distribution of skill returns.

2 Data and Preliminary Evidence

2.1 Matched Earning Records and Skill Measures

Our data consist of annual employer–employee matched records for the whole population of Swedish workers and firms during 1990–2017, including earnings, industry, occupation, worker characteristics such as age, gender, and education. A strength of these data are cognitive and noncognitive military enlistment tests that we link to individual workers. The tests were mandatory before 2007 and are available for almost 90 percent of males, across birth cohorts, in our sample. The cognitive score is assessed through tests covering logic, verbal, spatial, and technical comprehension. The noncognitive score is from a semi-structured interview with a certified psychologist who assesses willingness to assume responsibility, independence, outgoing character, persistence, emotional stability, and initiative.

Prior research shows that these scores are significant predictors of earnings and other labor market outcomes (Lindqvist and Vestman, 2011; Fredriksson et al., 2018; Edin et al., 2022). Cognitive and noncognitive measures are recorded on a standard-nine (Stanine) scale, which approximates the Normal distribution and facilitates comparisons across birth cohorts.[2] These tests are strongly associated with life-cycle earnings in our sample (see Appendix A.1).

We restrict the sample to males aged 20–60 with nonmissing scores and to firms that employ an average of at least ten male workers over five years or more. We focus on estimates from 1999–2008. Results are similar in alternative samples (1990–1999 and 2008–2017). The 1999–2008 sample consists of approximately 26,000 firms and 1,100,000 workers in all private nonprimary industries. The dataset reports both organization and workplace identifiers. To identify "firms"

[2] Measures are standardized for each birth year. A score of 5 denotes the middle 20 percentiles of the population taking the test. Scores of 6, 7, and 8, are given to the next 17, 12, and 7 percentiles, and the score of 9 to the top 4 percent of individuals. Scoring below 5 is symmetric.

we use the workplace with the highest income in the calendar year. This choice is closest to the notion of a production unit and is consistent with existing work (e.g., Card et al., 2013). We use the annual labor income at the firm as our measure of earnings: this is available for all workers and includes bonuses and performance pay throughout. Descriptive statistics are in Appendix A. In Section 6 we link information on firms' financial accounts (from a commercial data provider) and innovation activities (from the Swedish version of the European Community Innovation Survey). These data are reported at the organization level and, in the case of multi-workplace firms, we coarsen estimates to that level of aggregation.

2.2 Estimation of High-Dimensional Models of Skill Returns

We estimate models with many firm and worker fixed effects alongside firm-specific returns to skills. Depending on the estimation approach, different sampling restrictions are required (see Appendix A.2). We work with the largest connected component of the firm–worker graph (Abowd et al., 2002; Bonhomme et al., 2023). In order to account for different skill levels (e.g. high vs low cognitive skills), the set must be connected along each skill level (Kline et al., 2020). Connectedness delivers unbiased parameter estimates but the variance components may be biased if sampling errors in parameter estimates enter the quadratic forms of interest. This bias in quadratic forms may result in overstated variances and understated covariances (Andrews et al., 2008).

We use two different approaches to deal with this problem. First, in our baseline analysis, we characterize unobserved heterogeneity as the "class" of a firm, where each class corresponds to a cluster of similar employers (Bonhomme et al., 2019). We define 100 classes using a k-means algorithm based on the average earnings and average skills of workers (stayers and movers). This characterization reduces dimensionality, enhances tractability and delivers well-centered estimates of the contributions of worker and firm heterogeneity to earnings dispersion (Lamadon et al., 2022). Second, in robustness analysis, we use variance component estimators for linear models with heteroscedasticity of unknown form. This approach relies on leave-out estimators of the variances of the errors in the linear model. For each worker–firm match, we estimate the error variance from a sample where that match's observations are left out. The leave-out procedure delivers unbiased estimates in finite samples (Kline et al., 2020) and facilitates tests of linear restrictions.[3] Appendix B overviews the nuances of each approach; Table A.1 reports statistics for the underlying samples.

A concern is that skill proxies may be measured with error. To account for this possibility, we conduct robustness analysis to quantify the impact of incrementally stronger measurement error on estimates of skill returns (Appendix C). This exercise shows that our estimates of returns' heterogeneity would be even larger under conventional assumptions about noise in skill measures.

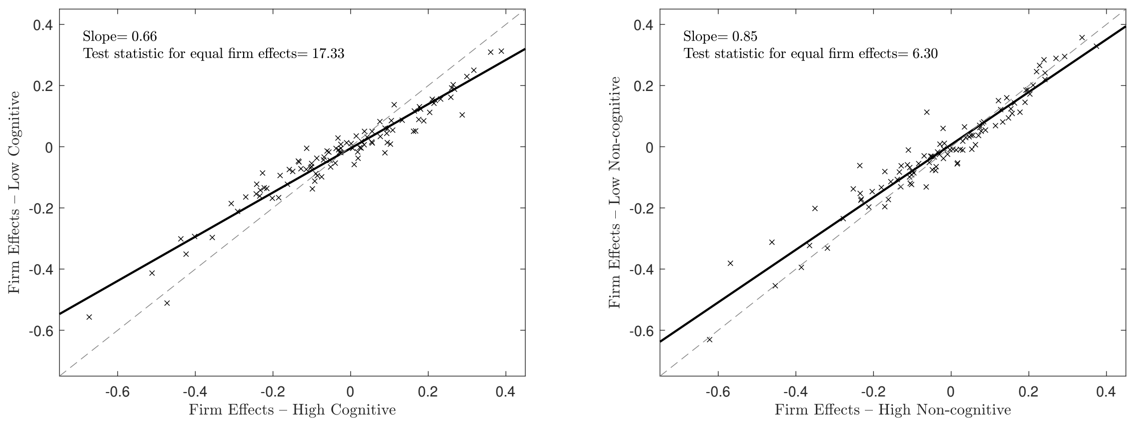
2.3 A First Glance at Skill Returns

Returns heterogeneity, in its most basic form, can be tested with binary skill levels (high vs low test scores) in each attribute. We consider a widely studied additive specification (Abowd et al., 1999, AKM) and allow firm fixed effects for high and low skill workers to differ.

We classify workers into high cognitive ($C = \mathbb{1}[c > 5]$) and high noncognitive ($N = \mathbb{1}[n > 5]$), where scores of 5 reflect the central moments of the Stanine scores (see Footnote 2). To account for serial correlation within employment spells, we select observations within a two-year set and separately estimate linear binary models of worker and firm effects of the form $\log(w_{ijt}) = \mu_i^S + \theta_j^S + \varepsilon_{ijt}$ for cognitive skills, $S \in \{C = 0, C = 1\}$ or noncognitive skills, $S \in \{N = 0, N = 1\}$. Subscript t takes on the values of the two years selected. We use non-adjacent years (in fact, two years apart) to mitigate the impact of partial employment spells during contiguous years when workers switch firms.

Figure 1 plots estimated firm fixed effects of high-skill (x-axis) and low-skill (y-axis) workers for $t \in \{2004, 2007\}$. Panel (A) shows results for cognitive skills while Panel (B) plots those for

[3] This analysis requires that firms remain connected by worker mobility when a single mover is dropped. Therefore, the original sample must be pruned to ensure that the connectedness condition is met by all leave-out subsamples. We use Python NetworkX to identify the articulation points of the worker–firm graph, then trim it to construct the double leave-one-out connected set (see Appendix A.3). Given the large computations necessary to estimate leave-out quadratic forms, we adapt the random projection method of Achlioptas (2003).



(A) High vs low cog skills – 100 firm clusters

(B) High vs low noncog skills – 100 firm clusters

Figure 1: Firm effects heterogeneity: cognitive and noncognitive skills.

The figures plot the averages of firm effects for low-skill workers ($\theta_j^{S=0}$) against the averages of firm effects for high skill workers ($\theta_j^{S=1}$), where $S \in \{C, N\}$. Firm effects are estimated for 25,783 firms grouped into 100 clusters based on workers' average cognitive skill, noncognitive skill, and earnings. Firm effects are demeaned. The "Test Statistic" is for the null hypothesis that estimated firm effects are equal across skill groups and should be evaluated against the standard normal distribution (see Appendix B.3).
Sample restriction: years 2004 and 2007 only. Tests for other year pairs are in Appendix Table B.1.

noncognitives. The sample consists of 25,783 firms grouped into 100 clusters based on workers' average cognitive skill, noncognitive skill, and earnings. The "Test Statistic" is for the null hypothesis that estimated firm effects are equal across skill groups. Estimated slopes are 0.66 for cognitive and 0.85 for noncognitive. Under the null hypothesis of no heterogeneity in skill returns, the slopes should be statistically indistinguishable from 1 and the scatters should align along the dashed 45° lines. This is not the case and the null hypotheses that firm effects are the same for high- and low-skill workers are strongly rejected: the test statistics of equal firm effects for high and low skill workers have values above 6 for both cognitive and noncognitive returns.

Since firm effects do not appear to be independent of worker skills, one must discard the notion of a homogeneous skill return across firms. Tests for alternative periods lead to similar conclusions and deliver slopes that are well below 1 (Table B.1). When we run similar tests on firm-specific estimates of slopes that are corrected for bias using the methods of Kline et al.

(2020), we again estimate slopes that are well below 1 and reject the null hypothesis of homogeneous returns across skills groups (Appendix B.3).

We find further evidence of heterogeneity in skill returns when performing event studies. In these exercises we track wage changes for different workers as they move across firms (Bonhomme et al., 2019; Lamadon et al., 2022). For example, moving to an employer with larger cognitive returns leads to increases in the relative wage of high-skill workers compared to low skill ones (Appendix B.4). Results are similar for noncognitive returns. We find no evidence of cross-skill effects (that is, no changes in the relative wages of high *cognitive* individuals who move between firms with different *noncognitive* skill returns).

3 Quantifying Variation in Skill Returns

To quantify heterogeneity in firm returns we develop an empirical framework that allows for granular differences in skill bundles while controlling for other sources of variation. The framework is derived from a model of demand for productive skills (see below and Appendix D).

3.1 Skill Demand by Heterogeneous Firms

We embed return heterogeneity in a model where firms differ in four dimensions: (i) cognitive returns; (ii) noncognitive returns; (iii) output market demand, where they have varying degrees of monopoly power; (iv) cost of labor in the input market, driven by differences in non-pecuniary firm characteristics valued by employees. Monopoly power in the output market implies a *skill-independent* firm surplus. This underpins the cross-sectional variation in base-wages reflected in firms' fixed effects. On the other hand, firm-specific labor supply curves (input market heterogeneity) imply rents for both workers and firms (Card et al., 2018; Lamadon et al., 2022). The model allows for a production technology with heterogeneous skill returns.

Production complementarities and labor supply. Consider an environment with two heterogeneous sides (workers, firms). Workers have measure one, differ in their cognitive (c) and noncognitive (n) abilities, and are indexed by their skill vector (c,n). Firms are indexed by j. A firm j matched with a (c,n) worker produces output $y = f_j(c,n)$. Given constant returns to scale (CRS) in worker headcounts (as in the production problem with multiple skill inputs of Eeckhout and Kircher, 2018), a j firm matched with k workers of type (c,n) produces $k \times f_j(c,n)$, while a j firm matched with one (c_1,n_1) and one (c_2,n_2) worker produces $f_j(c_1,n_1) + f_j(c_2,n_2)$.[4] Firm j's output, if it hires shares $q_j(c,n)$ of the workforce of type (c,n), is:

$$y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n). \tag{1}$$

where G is the population measure of different worker types in the economy. A worker's utility from matching with a firm depends on their wage plus a preference shock. Workers choose firms that give them the highest utility. Using standard arguments (McFadden, 1974), the share $q_j(c,n)$ of type (c,n) workers who choose firm j has a logit form:

$$\log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n)). \tag{2}$$

Equation (2) describes the upward sloping labor supply faced by firm j. The intercept $h(c,n)$, determined in equilibrium, guarantees market clearing. β is the sensitivity of labor supply to wages. Firm-specific wages for each skill set and the labor market equilibrium are defined in Appendix D.

Technology and wages. The wage paid by firm j reflects market structure and technology. The monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker. The marginal revenue is an increasing function of the firm's output market share and of its productivity.

In the appendix we show that logged wages are the sum of a common level effect plus a firm-specific intercept and skill premium,

[4] Production is defined at the level of the match since workers may not agree on the ranking of firms. That is, the production technology combines the skills of a worker and the technology of a firm (Lise and Robin, 2017). As in one-to-one matching problems, firms produce with every match separately.

$$\log(w_j(c,n)) = \alpha + \Lambda_j + \Delta_j(c,n). \tag{3}$$

The skill premium $\Delta_j(c,n) = \log(f_j(c,n)/f_j(L,l))$ corresponds to the output of a (c,n) skill bundle relative to the lowest skill bundle (L,l) within firm j. To obtain an empirical counterpart of (3), we do not restrict the functional form of the production function $f(\cdot)$, and hence of $\Delta_j(c,n)$, but rather use a first-order approximation that delivers a bilinear relationship for worker i in firm j. We also explore higher-order approximations with skill interactions but this makes a negligible difference.[5] Making the worker index i explicit, the empirical wage representation is:

$$\log(w_{i,j}(c,n)) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i, \tag{4}$$

where λ_j^0 is the baseline wage that a worker with the lowest bundle of skills earns in firm j. Gradients λ_j^c and λ_j^n are firm-specific marginal returns, above and beyond the baseline wage λ_j^0. Finally, as we show below, the individual intercepts μ_i capture the average (Mincerian) returns to a worker's skill endowments.

Normalizations. Identification of (4) requires linear restrictions on firm effects, which are defined relative to a reference firm (or set of firms). We identify parameters up to unknown constants $\{\kappa_0, \kappa_\mathrm{c}, \kappa_\mathrm{n}\}$, such that:

$$\begin{aligned}
\lambda_j^0 &= \Lambda_j - \kappa_0 \\
\lambda_j^c &= \frac{\partial \Delta_j(c,n)}{\partial c} - \kappa_\mathrm{c} \\
\lambda_j^n &= \frac{\partial \Delta_j(c,n)}{\partial n} - \kappa_\mathrm{n} \\
\mu_i &= \alpha + \kappa_0 + \kappa_\mathrm{c} \cdot c_i + \kappa_\mathrm{n} \cdot n_i
\end{aligned} \tag{5}$$

We set $\kappa_0 = \bar{\Lambda}$, $\kappa_\mathrm{c} = \frac{\partial \bar{\Delta}(c,n)}{\partial c}$, and $\kappa_\mathrm{n} = \frac{\partial \bar{\Delta}(c,n)}{\partial n}$, where the reference values $(\bar{\Lambda}, \frac{\partial \bar{\Delta}(c,n)}{\partial c}, \frac{\partial \bar{\Delta}(c,n)}{\partial n})$ correspond to the average employment-weighted firm effects. For example, κ_0 is normalized to

[5] For a discussion of log-additive firm effects in wage specifications with bundled skills, see Choné and Kramarz, 2021. Worker–firm complementarities can be micro-founded by restricting attention to the labor composition alone (e.g., learning and cooperation of workers as in Jarosch et al., 2021).

the average intercept ($\bar{\Lambda}$) in the population of firms. This is conservative, since central moments yield the lowest variance of firm heterogeneity (intuitively, they minimize squared deviations) and are less sensitive to estimation error than, e.g., extrema of the firm effects distribution. Unlike models with degenerate skill returns, firm premia are not restricted to be equal across skill groups. Under the model's null hypothesis, within-firm wage dispersion depends on workers' skills and firm's returns.

3.2 Identification and Estimation of Skill Returns

We study a sample of firms connected by worker mobility along both skill dimensions over the 1999–2008 period. The baseline wage specification is:

$$\log(w_{ijt}) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i + \boldsymbol{X}_{it}\boldsymbol{b}_t + \varepsilon_{ijt}, \tag{6}$$

where λ_j^0 are skill-independent earnings, λ_j^c and λ_j^n are skill gradients, and μ_i are worker fixed effects. We account for life-cycle and time variation through interactions of skill, age, and year, denoted as $\boldsymbol{X}_{it}\boldsymbol{b}_t$ in (6).

Identification of firm effects. It is useful to draw attention to what can, and cannot, be identified in specifications like (6). Lacking observable skill proxies, the assumption of a single skill index is necessary for identification alongside a connected worker–firm graph (see discussion in Appendix B.5). If unobserved skills are collapsed into a single index, identification amounts to a rank condition requiring that the average quality of the workers moving to a firm is not the same as that of workers moving out of the firm (see Remark 1 in the appendix and Bonhomme et al., 2019; Lamadon et al., 2022). This result does not hold with multiple skill dimensions because the ranking of workers is not unique and workers with different attributes may exhibit similar overall productivity. Identification of heterogeneous returns in such settings is not possible unless skill proxies are available. The intuition is that wage changes following moves across firms cannot be traced back to one single skill dimension unless the other dimension is controlled for (Lemma 1 and Remark 2, Appendix B.5). Returns in equation (6) are identified under general conditions if skill measures are available (Remark 3). To illustrate this point, we explicitly solve for cognitive returns as a function of the average skills and wages of workers who move between employers in Appendix B.5. Identification works similarly for noncognitive skills and firm intercepts.

Interpreting parameters. The level and dispersion of worker fixed effects μ_i reflect skill endowments. That is, μ_i includes the average (Mincerian) return to a worker's skills. Generally, the μ_i fixed effect accounts for worker skills that are priced homogeneously across firms.

We normalize the Stanine scores to take values on the unit interval. Setting a unit upper bound for skills is convenient because each skill return λ_j^s can be interpreted as the earnings gap separating the highest and lowest worker types.[6]

The linear restrictions on firm effects imply that the lowest skill workers gain no employer premium above and beyond firm intercepts. Put differently, for the subset of workers with the lowest skill endowments $(c = 0, n = 0)$, equation (6) reduces to a standard specification with firm fixed effects λ_j^0, time-varying controls $\boldsymbol{X}_{it}\boldsymbol{b}_t$, and worker fixed effects μ_i. For other skill types, (6) augments the double fixed-effect specification by allowing for heterogeneous returns to skills. If we restrict attention to a single skill dummy S over a two year interval with no other control variables, estimation of (6) collapses back to the binary model from Section 2.3 where $\lambda_j^0 = \theta_j^{S=0}$ and $\lambda_j^s = \theta_j^{S=1} - \theta_j^{S=0}$.

Interactions of skill, year, and age dummies (in $\boldsymbol{X}_{it}\boldsymbol{b}_t$) flexibly account for variation in average skill returns and reduce computation times. [7] Conditional on the latter, worker fixed effects absorb time-invariant residual skill components.

[6]The transformation is $(Stanine - 1)/8$. The distribution of normalized skills is carried over from the Stanine distribution. Normalized scores for c and n are defined on the grid $[0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1]$. Sampling restrictions have little impact on the distribution's moments relative to the population of test takers: e.g., $\bar{c} = 0.54$, $\bar{n} = 0.52$, $\text{sd}(c) = 0.24$, $\text{sd}(n) = 0.21$, $\text{corr}(c, n) = 0.36$.

[7] For example, estimation for the leave-out sample takes about 20–30 hours using Python and the JLA approximation. Adding stratified controls raises computation time proportionally to the number of added parameters. Allowing for time-varying returns to education, instead, does not affect results. Life-cycle profiles, by skill and time, are accounted for by the $cognitive \times noncognitive \times age \times year$ group interaction in $\boldsymbol{X}_{it}\boldsymbol{b}_t$. Dummies for $s \leq 0.25$, $0.375 \leq s \leq 0.625$, $0.75 \leq s$ for $s \in \{c, n\}$ are interacted with each other and age groups 20–25, 26–32, 33–42, 42–60 as well as two-year period dummies 1999–2000, 2001–2002, 2003–2004, 2005–2006, 2007–2008.

Estimation. Baseline estimation relies on the clustering approach. We use the k-means algorithm to partition firms into 100 clusters, based on the average skills and earnings of employees. Results are not sensitive to the use of alternative clustering approaches. To gauge robustness, we replicate the analysis using a different approach that delivers unbiased estimates of arbitrary quadratic forms of firm returns (Kline et al., 2020).

3.3 Estimates of Firm Parameters

Table 1 shows results from the estimation of (6). Skills are free to vary over their granular range (for example, $c_i \in [0, 0.125, .., 1]$). We initially focus on the dispersion of firm parameters. As we show below, heterogeneity in skill returns has implications for other moments of the earnings distribution through behavioral responses and assortative matching.

Column (1) reports estimates from the grouped-firms approach. The first line, $\text{sd}(\lambda_j^0) = 0.10$, highlights that skill-independent premia vary significantly across employers, confirming well-established evidence on firm fixed effects. The lines below reveal a less known layer of firm heterogeneity and show that the standard deviations of skill returns are $\text{sd}(\lambda_j^0) = 0.08$ for noncognitive skills and $\text{sd}(\lambda_j^c) = 0.05$ for cognitive ones. Column (2) shows that heterogeneity in skill returns is even larger when using the leave-out (non-grouped) approach. Interestingly, the relative magnitudes of parameter dispersion are unchanged as the values of $\text{sd}(\lambda_j^0)$, $\text{sd}(\lambda_j^c)$ and $\text{sd}(\lambda_j^n)$ all approximately double. The finding of stable *relative* magnitudes is robust throughout the analysis and indicates that estimates of the proportional contribution of each layer of firm heterogeneity do not depend on the estimation method.

A double differencing thought experiment. To convey the magnitude of skill premia, in columns (3) and (4) of Table 1 we consider thought experiments whereby workers with different skills are parachuted from their original firm to a different one in which returns are one standard deviation larger. We report the hypothetical change that such transitions imply for the wage gap between high skill workers (the 90^{th} percentile of skills) and low skill workers (the 10^{th} percentile of skills).

Table 1: Standard deviations of firm parameters: estimates from clustered sample and from firm-level sample with quadratic-form correction.

	Standard deviations		Standard deviations $\times$ ($90^{th} - 10^{th}$ skill percentile)	
	grouped (1)	firm-level (2)	grouped (3)	firm-level (4)
sd(λ_j^0)	0.10	0.22		
sd(λ_j^c)	0.08	0.15	0.06	0.11
sd(λ_j^n)	0.05	0.10	0.04	0.07
cumulative (cog+noncog score)			0.10	0.19
# unique firms	25,783	19,085		

Notes: The first two columns show standard deviations of parameters λ_j^0, λ_j^c, and λ_j^n in equation (6). Column (1), clustered firms estimation: group firms into 100 classes according to average earnings and average c and n scores (k-means algorithm). Then, estimate (6) on the grouped data. Column (2), quadratic-form correction: compute corrected variances of the parameters estimated at the individual firm level and take the square root. In Columns (3) and (4) we multiply the estimated standard deviations by the skill gap between the 90^{th} (c_i and n_i of 0.875) and 10^{th} (c_i and n_i of 0.125) percentiles. Sample period: 1999–2008.

Based on grouped estimates, moving to a firm that sits just a standard deviation higher in cognitive returns results in an extra gain of six log points for a worker at the 90th cognitive percentile ($c_i = 0.875$) compared to a worker at the 10th percentile ($c_i = 0.125$). These differences in the gains from job mobility are elicited through positive assortative matching (see Section 5). Heterogeneity in noncognitive returns is lower but economically significant. Parachuting a worker at the 90th percentile of n_i into a firm that is a standard deviation higher in noncognitive returns raises their earnings gap relative to someone at the 10th percentile by four log points. A one-standard deviation change in both cognitive and noncognitive returns for workers at the 90th, rather than the 10th, percentile of each skill brings about an impact that is as large as that of firm intercepts (see cumulative effect in the last line of Table 1). Absolute magnitudes are larger when we estimate at the firm (non-grouped) level.

Estimates of dispersion in skill returns are robust in several respects. For example, Appendix E.1 shows that bias correction in a leave–observation-out sample (rather than leave–match-out sample) delivers even higher dispersion of skill returns. In other sensitivity checks we find that, when varying the number of firm clusters in the grouping estimator, the relative magnitudes of skill returns and firm intercepts are unchanged. In Appendix E.2 we plot the dispersion of skill returns and show that it is stable across time periods. Moreover, we show that the employment-weighted correlation of returns among firms, $\text{corr}(\lambda^{\text{c}}, \lambda^{\text{n}})$, is positive but imperfect (less than 0.1 when we use the clustered-firms approach; less than 0.3 when we use the bias-correction approach). Such weak correlation lends support to the hypothesis that firm heterogeneity is multidimensional.[8]

4 Matching

Do cognitive and noncognitive traits matter for the assignment of workers to employers? And how do they affect the distribution of earnings? To examine these questions we characterize worker–firm matching in a setting with multiple skill attributes (Lindenlaub, 2017).

First, we introduce notation. Firms differ in three dimensions: earnings intercept (λ_j^0), cognitive return (λ_j^{c}) and noncognitive return (λ_j^{n}). We define a matching function $\varphi(\lambda_j^{\text{c}}, \lambda_j^{\text{n}}) = (\bar{c}_j, \bar{n}_j)$, which maps firm's returns into their average worker skills. Under the assumption of upward sloping firm-specific labor supplies (equation (2)), the matching function φ is increasing in λ_j^{c} and λ_j^{n}, and multidimensional positive assortative matching holds (see Appendix E.3). In what follows, we examine the empirical content of these restrictions.

4.1 Sorting Patterns

Assortative matching, whether positive (PAM) or negative (NAM), is characterized by the matching function's derivatives. With one-dimensional heterogeneity, this boils down to the sign of a single derivative. With multiple attributes, all elements of the Jacobian play a role.

[8]Plots in Appendix E.2 confirm the substantial dispersion of skill returns. In the grouped estimation, cognitive returns concentrate between -15 and $+20$ log points. Relative to the 10th percentile of skills, a worker from the 90th percentile who moves from bottom to top of the returns distribution gains 25 extra log points in earnings. Noncognitive returns have a similar range of variation.

Definition 1. *The sorting pattern is locally PAM if, for given* $(\lambda^{\mathrm{c}}, \lambda^{\mathrm{n}})$*, the following holds:*

$$(a)\ \frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{c}}} > 0; \qquad (b)\ \frac{\partial \bar{n}_j}{\partial \lambda_j^{\mathrm{n}}} > 0; \qquad (c)\ \frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{c}}}\frac{\partial \bar{n}_j}{\partial \lambda_j^{\mathrm{n}}} - \frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{n}}}\frac{\partial \bar{n}_j}{\partial \lambda_j^{\mathrm{c}}} > 0.$$

Definition 1 requires sorting patterns along each skill dimension; moreover, these patterns must not be offset by potential cross-sorting. To examine assortative matching, we focus on the Jacobian of the matching function:

$$\frac{\mathrm{d}\varphi(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})}{\mathrm{d}(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})} = \begin{bmatrix} \frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{n}}} \\ \frac{\partial \bar{n}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \bar{n}_j}{\partial \lambda_j^{\mathrm{n}}} \end{bmatrix} \tag{7}$$

The Matching Jacobian in data. We pursue two routes to test the sorting hypothesis. First, we evaluate the Jacobian by projecting average skills $\bar{c}_j$ and $\bar{n}_j$ onto firm returns. The advantage of this approach is that one can examine patterns where skill sorting in each dimension depends on both of the employer's returns. In practice, we estimate the following projections:

$$\begin{aligned} \bar{c}_j &= \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{c}} \\ \bar{n}_j &= \delta_{1\mathrm{n}} + \delta_{2\mathrm{n}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{n}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{n}}. \end{aligned} \tag{8}$$

Estimates of the δs provide a test of Jacobian conditions. The regressions in (8) deliver the best linear approximation to the conditional expectations of $\bar{c}_j$ and $\bar{n}_j$. For instance, $\mathrm{E}(\bar{c}_j|\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}}$, so that the parameter $\delta_{2\mathrm{c}}$ is the expected value of the top-left element $\left(\frac{\partial \bar{c}_j}{\partial \lambda_j^{\mathrm{c}}}\right)$ of the Jacobian taken over the sample of all firms. Similar arguments hold for $\delta_{3\mathrm{c}}$ and gradients in the second line of (8). If returns λ are measured with error (due to limited mobility), estimation of (8) may deliver biased point estimates. Our clustering approach mitigates such concerns as we project average skills (cognitive or noncognitive) onto the 100 cluster-specific returns. Table 2 reports estimates of the Jacobian's parameters.

Positive and significant estimates of $\delta_{2\mathrm{c}}$ and $\delta_{3\mathrm{n}}$ in Table 2 imply that the own-derivative conditions for PAM are satisfied for both c and n. High c workers sort with high λ^{c} returns firms, conditional on λ^{n} in (8); high n workers sort with high λ^{n} firms, conditionally on λ^{c}. This underscores the multidimensional nature of skill returns.

Table 2: Projection of average skills onto grouped returns.

	Dependent Variables:					
	(1)		(2)		(3)	
	$\bar{c}_j$	$\bar{n}_j$	$\bar{c}_j$	$\bar{n}_j$	$\bar{c}_j$	$\bar{n}_j$
λ_j^{c}	1.21	0.58	1.18	0.55	1.15	0.53
	(0.08)	(0.07)	(0.07)	(0.06)	(0.07)	(0.05)
λ_j^{n}	–0.15	0.61	–0.05	0.71	–0.14	0.61
	(0.11)	(0.08)	(0.10)	(0.07)	(0.11)	(0.07)
R^2	0.676	0.542	0.712	0.612	0.752	0.648
# firms	25,783		25,783		25,783	
Controls	No		λ_j^0, # employees		λ_j^0	
Weights	No		No		# employees	

Notes: Column (1) reports sorting coefficients δ_2 and δ_3 from estimating (8). The specification in column (2) additionally controls for intercepts λ^0 and for total employment headcounts within firms. Column (3) weights the observations by each firm's number of employees. One firm is one observation. Robust standard errors clustered at the level of the 100 firm groups (in parentheses). Grouped estimation. Sample period: 1999–2008.

The Jacobian is positive definite (the determinant $\delta_{2\mathrm{c}}\delta_{3\mathrm{n}} - \delta_{3\mathrm{c}}\delta_{2\mathrm{n}}$ is larger than zero), lending additional support to the hypothesis that PAM holds in our large sample of firms and workers between 1999 and 2008.

The positive $\delta_{2\mathrm{n}}$ in equation (8) indicates cross-sorting of high n workers to high λ_j^{c} firms. This occurs because skill endowments are correlated and high c workers, who sort into high cognitive returns, have on average a higher endowment of n skills. This observation suggests that the own-sorting in the c dimension is strong enough to trigger indirect effects, as confirmed by Figure 2 below. There is less evidence of cross-sorting of high c into high λ_j^{n} firms: this suggests that own-sorting in the n dimension is weaker and not sufficient to induce such indirect effects (confirmed, again, by Figure 2 below).[9] Results do not change when we control for firm-specific employment and intercepts λ^0, as shown in column (2) of Table 2. Neither do they change when we weigh by employment, as in column (3). Differences in estimated λ^{c} and λ^{n} returns account

[9]Correlation of λ_j^{c} and λ_j^{n} would affect cross-sorting estimates if we did not control for their respective indirect return in (8).

for roughly 60% percent of the skill variation between firm clusters.[10] When we weigh firms by their employment and control for λ^0, the explained variation rises to 65–75 percent.

A different test of sorting. The second route to test for sorting hinges on Jacobian parameters estimated from the following specification (see Appendix E.3):

$$\begin{aligned} \lambda_j^{\mathrm{c}} &= d_{1\mathrm{c}} + d_{2\mathrm{c}}\bar{c}_j + d_{3\mathrm{c}}\bar{n}_j + e_j^{\mathrm{c}} \\ \lambda_j^{\mathrm{n}} &= d_{1\mathrm{n}} + d_{2\mathrm{n}}\bar{c}_j + d_{3\mathrm{n}}\bar{n}_j + e_j^{\mathrm{n}}. \end{aligned} \tag{9}$$

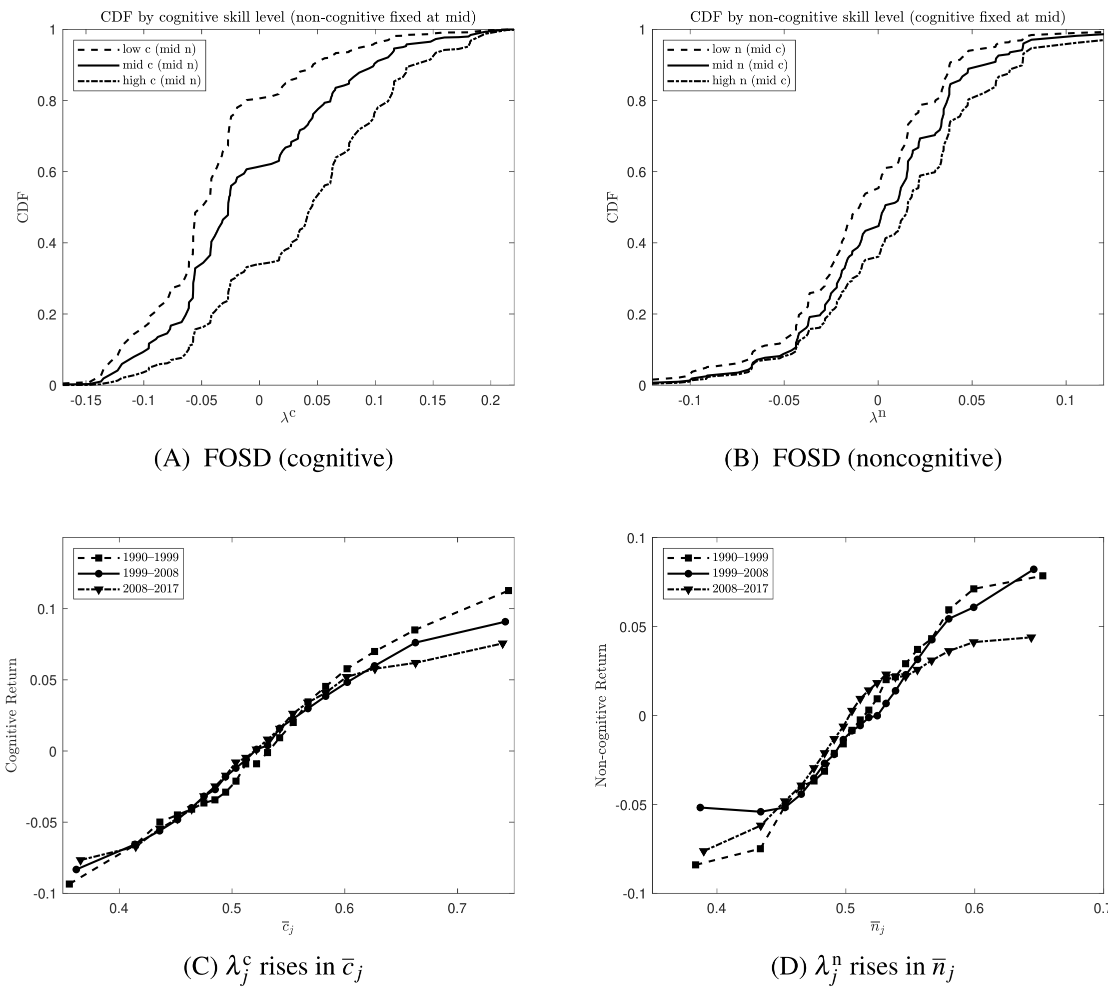
The linear forms in (9) are reminiscent of the projections of fixed effect on firm characteristics used in the applied literature (Kline et al., 2020). Under standard assumptions, the parameters can be correctly estimated from a cross-section of individual (non-grouped) firms. If returns are measured with error, having λ_j^{c} and λ_j^{n} on the left-hand-side avoids biases in the estimation of the d-parameters in (9). We use these projections to test for PAM. One caveat applies: while point estimates are generally unbiased, standard errors must be corrected for the correlation across the first-stage estimates of the outcome variable (firm parameters).[11] Appendix E.3 reports point estimates and standard errors for the projections in (9), based on firm-level data (employees' cognitive and noncognitive skills are averaged into firm-specific $\bar{c}_j$ and $\bar{n}_j$). PAM cannot be rejected since the own-partial derivatives and the determinant of the Jacobian are positive throughout. Sorting again appears to be stronger in the cognitive dimension.

4.2 The Distribution of Workers over Returns

If a high-skill workers matches more frequently with firms exhibiting high returns to that skill (in the sense of first-order stochastic dominance, FOSD), then sorting is positive along that dimension (Lindenlaub and Postel-Vinay, 2023). To visualize these patterns, we compare the cumulative distribution functions (CDF) of returns for separate sets of workers.

[10]The R^2 refers mostly to between-firm variation, since average skills vary little within k-means clusters. It is remarkable that returns can explain so much of the skill variation between the clusters.

[11]We use the correction proposed in equation (7) of Kline et al. (2020) to construct adjusted standard errors.



(A) FOSD (cognitive)

(B) FOSD (noncognitive)

(C) λ_j^c rises in $\overline{c}_j$

(D) λ_j^n rises in $\overline{n}_j$

Figure 2: Distributions of firm returns for different sets of worker skills.

Notes: Results from grouped estimation. Panels (A) and (B) show cumulative distribution functions for workers with low ($c, n \leq 0.25$), mid ($0.25 < c, n < 0.75$), or high ($c, n \geq 0.75$) skill ranks over the range of firm returns. Period: 1999–2008. FOSD: first-order stochastic dominance.

Panels (C) and (D) show binned scatterplots of firm-specific skill returns (vertical axis) with average skills (horizontal axis) for three ten-year estimation periods: 1 (1990–1999), 2 (1999–2008), 3 (2008–2017).

First-order stochastic dominance. Figure 2 shows sorting patterns along either cognitive or noncognitive attributes, using the grouped-firm estimates. The top panel plots the CDF for workers in three coarse skill-specific ranks (low, medium or high). The CD functions are defined over the ordered set of estimated firm returns.[12]

In Panel (A) we condition on medium noncognitive skills and show that workers with higher cognitive attributes match with higher cognitive returns λ_j^{c}. The CDF of high cognitive workers dominates all other types; the CDF of medium cognitive workers dominates the CDF of low cognitive workers. Panel (B) shows FOSD patterns across ranks of noncognitive attributes (n), holding cognitive attributes fixed at the medium rank. Sorting patterns on noncognitive traits are less striking but clearly discernible. The bottom panels of Figure 2 show the distributions of skill returns over the range of within-firm average skills. These plots confirm that returns increase monotonically with skill endowments, consistent with PAM. Between-firm differences in average skills are larger in the cognitive dimension, which is expected given the higher dispersion of λ^{c} relative to λ^{n} and the stronger sorting incentives. Similar patterns hold for other sample periods.

5 Complementarities and Earnings

Considering the non-random nature of firm assignments, it is useful to distinguish between the return that an average worker gets from a firm and the excess gains derived by matching different skills to different employers.

In what follows, we cast firm heterogeneity in terms of deviations from cross-sectional means and explicitly account for assortative matching.

[12]For clarity, we coarsen the skill levels to low ($c,n \leq 0.25$), mid ($0.25 < c,n < 0.75$), and high ($c,n \geq 0.75$). Returns are estimated by clustering firms. This is graphically convenient as it restricts variation on the x-axis. Estimates at the non-clustered firm level deliver similar insights. Additional FOSD plots in Appendix E.3.

5.1 Effects on the Distribution of Earnings

We rewrite Equation (6) so that the log wage for skill bundle $s_i = (c_i, n_i)$ of worker i in firm j is:

$$\log(w_j(s_i)) = \underbrace{\mu_i}_{\substack{\text{(a)}\\ \text{Person effect}\\ \text{(incl. Mincer returns)}}} + \underbrace{\lambda_j^0}_{\substack{\text{(b)}\\ \text{Firm intercept}}} + \underbrace{\lambda_j^{\mathrm{c}}\overline{c} + \lambda_j^{\mathrm{n}}\overline{n}}_{\substack{\text{(c)}\\ \text{Firm returns effect}}} + \underbrace{\lambda_j^{\mathrm{c}}\widetilde{c}_i + \lambda_j^{\mathrm{n}}\widetilde{n}_i}_{\substack{\text{(d)}\\ \text{Match effect}}}, \tag{10}$$

where $\widetilde{x}_i$ denotes the deviation of skill x_i from its cross-sectional average $\bar{x}$. Equation (10) has an intuitive interpretation: the term (a) summarizes the homogeneous Mincerian return (the $\kappa_{\mathrm{c}} c_i + \kappa_{\mathrm{n}} n_i$ in the normalization of Section 3), often estimated from survey data; component (b) is a firm fixed effect. The elements (c) and (d) reflect, respectively, the direct impact of returns' heterogeneity on the earnings of an average-skill person, and the more nuanced effect of assortative matching. Terms (c) and (d) add up to the premium $\lambda_j^{\mathrm{c}} c_i + \lambda_j^{\mathrm{n}} n_i$. They encapsulate firm returns that vary with worker skills. The expected value of (c) in (10) is nil because $\mathrm{E}(\lambda_j^{\mathrm{c}}) = \mathrm{E}(\lambda_j^{\mathrm{n}}) = 0$. In contrast, the expected value of (d) can be different from zero as it reflects the per capita wage gains due to assortative matching.

In the absence of skill measures, (b) and (c) would be conflated into the firm fixed effect and the skill dependent variation would be absorbed within the person fixed effect μ_i. Identification of heterogeneous returns and match-quality in summands (c) and (d) is obtained only when proxies of skill endowments are available.

Permanent heterogeneity decomposition. Panel (A) of Table 3 summarizes the influence of heterogeneous components (equation (10)) on earnings dispersion. Worker fixed effects μ_i have a strong influence on earnings. Estimates of worker fixed effects reflect the cross-sectional average of skill returns, as discussed in Section 3.1. Estimates of the dispersion of firm intercepts, as a share of total variation, are in line with existing evidence both at the firm level (Kline et al., 2020) and at the group level (Bonhomme et al., 2019; Lamadon et al., 2022). Table 3 highlights the magnitude of heterogeneous skill returns, which add up to a value close to the estimate of firm fixed effects.

Table 3: Contributions of firm heterogeneity to dispersion and levels of earnings.

Panel (A)	Dispersion Components: grouped (1)	firm-level (2)	Panel (B)	Levels ($\times 100$): grouped (3)	firm-level (4)
sd(μ_i)	0.43	0.49		—	—
sd(λ_j^0)	0.10	0.22		—	—
sd($\lambda_j^c c_i$)	0.05	0.09	E($\lambda_j^c c_i$)	0.75	0.66
sd($\lambda_j^n n_i$)	0.03	0.06	E($\lambda_j^n n_i$)	0.13	0.17
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.06	0.12	E($\lambda_j^c c_i + \lambda_j^n n_i$)	0.88	0.83
# unique firms	25,783	19,085		25,783	19,085

Notes: Panel (A) shows the dispersion of each summand in equation (10). Namely, the standard deviations of: (i) person and firm intercepts; (ii) interactions of returns and skills. Panel (B) shows the averages of the last two summands in equation (10). Namely, the contribution of matching to average earnings in the economy (complementarity gains). Firm-level estimates in column (4) are based on the observation-level, rather than the match-level, leave-out sample to capture the gains from matching in the population of workers. The averages of person and firm intercepts are uninformative due to the normalization of firm parameters and are omitted from Panel (B). Sample period: 1999–2008.

In Appendix E.4 we document, through variance decompositions, the importance of worker and firm effects, and of their covariation. This exercise shows that skill returns' heterogeneity accounts for a sizable share (about 1/4) of the firm-specific contributions to wage inequality. Restricting firm heterogeneity to fixed effects mechanically attributes part of that variation to employer intercepts.

Heterogeneous skill returns and sorting increase average earnings. These gains can be measured from the covariance of skills and firm returns. For example, for cognitive skills, we have that $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i) = \mathrm{cov}(\lambda_j^c, \overline{c}_j)$, which shows that sorting determines the intensity of the average gain accruing from returns' heterogeneity.[13] Panel (B) of Table 3 shows estimates of the average gain from match effects, which is 0.8–0.9 log points. The larger gains from cognitive returns reflect the stronger sorting in that dimension, also documented in Section 4.

5.2 The Uneven Gains from Sorting

The gains from sorting are nonmonotonic and convex. They are positive and large for high skill workers, absent for the least skilled workers and negative for a wide range of intermediate skills. To illustrate these patterns, we take an expectation of equation (10) *after conditioning* on skill level. For brevity, we overview gains from cognitive skills; similar arguments hold for noncognitives. Given skill level c_i, the full earnings gain from sorting is

$$\underbrace{c_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Full sorting gain}} = \underbrace{\overline{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Firm returns effect}} + \underbrace{\widetilde{c_i} \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Match effect}}, \tag{11}$$

where we split c_i into average $\overline{c}$ and deviation $\widetilde{c_i}$. The distribution of returns faced by each individual depends on the skill level and the expected return from firm heterogeneity changes nonlinearly with skills. Baseline estimates (Column 1, Table 4) illustrate that the marginal expected return $\mathrm{E}(\lambda_j^c \mid c_i)$ is increasing in c_i and deviates from the unconditional average, which is normal-

[13]The sorting parameters estimated in (9) are, in essence, this gain standardized by the variance of skills across firms, $\frac{\mathrm{cov}(\lambda_j^c, \overline{c}_j)}{\mathrm{var}(\overline{c}_j)}$. The equality $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i)$ follows from $\mathrm{E}(\lambda_j^c) = 0$ (excess returns have zero mean).

Table 4: Gains from sorting across returns λ_j^c for different cognitive skill levels.

	$\mathrm{E}(\lambda_j^c \mid c_i)$ (1)	Full gain (2)	Return effect (3)	Match effect (4)	$\mathrm{E}(\lambda_j^0 \mid c_i)$ (5)
skill level (c_i):					
1 (lowest, $c_i = 0$)	*-5.13*	0.00	-2.75	2.75	-2.00
2	*-4.61*	-0.58	-2.47	1.89	-1.51
3	*-3.75*	-0.94	-2.01	1.07	-1.45
4	*-2.61*	-0.98	-1.40	0.42	-1.28
5 (median, $c_i = 0.5$)	*-0.85*	-0.42	-0.45	0.03	-0.69
6	*1.10*	0.69	0.59	0.10	0.15
7	*2.98*	2.24	1.60	0.64	1.33
8	*4.86*	4.25	2.60	1.65	2.70
9 (highest, $c_i = 1$)	*6.74*	6.74	3.61	3.13	3.83
Aggregate	*0.00*	0.75	0.00	0.75	0.00

Notes: Gains are multiplied by 100 (i.e., in log points) for readability. All returns are differences relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach. Sample period: 1999–2008. Column (1): expected marginal return conditional on skill. Column (2): total gain from sorting. Column (3): gain from sorting for the average-skill worker. Column (4): gain from sorting in excess of an average-skill worker with the same employer. Column (5): gain from sorting into intercepts.

ized to zero. The difference in expected marginal returns between top and bottom cognitive skills is almost 12 log points ($6.74 - (-5.13) = 11.87$).

Marginal returns conditional on skills. Column (2) of Table 4 summarizes the distribution of gains. Top cognitive workers benefit from higher returns and earn 7 log points more than if they were matched with the average firm. This return matters for skill premia: we compare the sorting gains gap between a top worker ($c_i = 1$) and a low-middle (level 4 in Table 4, $c_i = 0.375$), which is 8 log points. The raw earnings difference between these two workers is on average 30 log points; the gap is reduced to $(30 - 8) = 22$ log points when sorting effects are taken out. Therefore, sorting adds more than 1/3 ($\frac{8}{22}$) to the baseline gap.

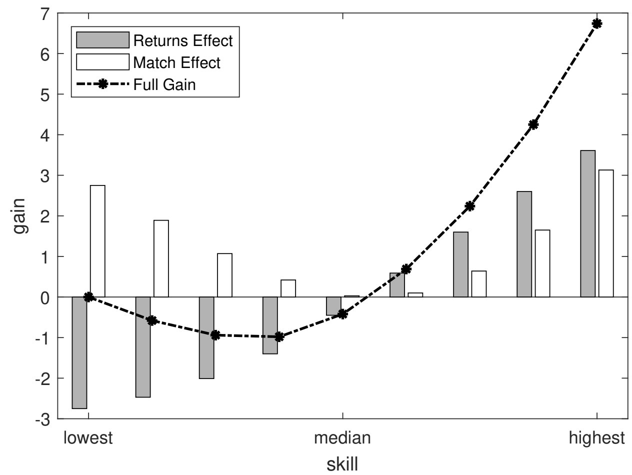
Nonmonotonicity of gains. Gains are not monotonic in skills (Column (2), Table 4). Workers with low-to-middle skills lose out compared to a hypothetical situation where everyone is matched with the average return. To understand why these losses wane as c_i approaches zero, we break down skill returns into a mechanical "return effect" $\bar{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ and a "match effect" $\widetilde{c_i} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ (see equation (11)). Estimates of the return effect $\bar{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ reflect the gain that a worker i, whose skill endowment is equal to the cross-sectional average, derives from being assigned to different expected returns (column (3), Table 4). Hence, the return effects measure the impact of firm heterogeneity net of complementarity. Since high skill workers sort into high return firms, estimates of the return effects grow monotonically with skills. This raises inequality compared to a random allocation and results in a zero-sum redistribution of returns, as evidenced by the aggregate nil effect reported in the bottom row of column (3) of Table 4. In contrast, match effects $\widetilde{c_i} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ in column (4) raise aggregate earnings by eliciting incremental gains from worker–firm complementarity.[14] Unsurprisingly, match effects are large at the high end of the skill distribution, where earnings are magnified compared to the random allocation (3.1 log points match effect for $c_i = 1$; 1.7 for $c_i = 0.875$). We detect large match effects also for low skill workers (2.7 for $c_i = 0$; 1.9 for $c_i = 0.125$) since match effects are defined as deviations from the average-worker gain (see (10)). That is, match effects reflect gains in excess of those experienced by an average-skill worker with the same employer (note that average-skill workers experience a steeper loss from matching with a low quality firm due to their higher opportunity cost).

Firm-specific intercepts and gains from sorting. The last column of Table 4 shows the wage gains due to matching with intercepts λ_j^0, conditional on skills. These gains are zero-sum due to the lack of complementarity between skills and firm intercepts. Nonetheless, the assignment of workers across firms (hence, across λ_j^0) raises earning gaps by an extent comparable to that due to skill sorting across returns (column (2)). This reinforces between-skills inequality as more able workers populate high λ_j^0 firms.

A graphical representation. Figure 3 summarizes the distribution of returns by skill. Low-skill workers experience positive match quality effects because they do not lose as much as

[14]Both components are defined as surplus relative to a baseline with no firm heterogeneity where all returns are equal to the population average. Hence, both positive and negative gains must be interpreted relative to a scenario where each worker gets the average return or, equivalently, where workers are randomly matched with firms.

Figure 3: Gains from sorting for workers with different cognitive skill ranks.



Gains are multiplied by 100 (i.e., in log points). Differences are relative to a scenario with no heterogeneity in firm returns. Estimates are based on the grouping approach (see Table 4). Sample period: 1999–2008.

the average worker from matching with low return employers. Gains turn negative for low-to-intermediate skill workers, who would benefit from matching with high return firms but do not. These individuals would be better off in a world with no firm heterogeneity in skill returns. Workers with above average skills experience large gains: both their return effects and their match effects are positive. Complementarities convexify the earnings-skills schedule. Overall, the gains outweigh the losses and matching raises aggregate earnings. To gauge the intensity of matching in the data, we benchmark it against the maximum gain achievable from the estimated returns and skill dispersion.[15] Assortative matching in the cognitive dimension generates 0.75 log points as compared to a hypothetical maximum of 1.9 log points. That is, the observed allocation of skills across employers delivers about 40% of the potential gains from cognitive sorting. Adding the returns from matching on noncognitives boosts aggregate match quality, up top 0.88 log points (Panel B, Table 3). Estimated gains from sorting are robust to alternative normalizations of skills and returns (Appendix E.5).

[15]Match effects are maximized when the correlation $\mathrm{corr}(\lambda_j^c, c_i) = \frac{\mathrm{cov}(\lambda_j^c, c_i)}{\mathrm{sd}(\lambda_j^c)\mathrm{sd}(c_i)} = 1$. Baseline (grouped) estimates imply an upper bound for match effects in the cognitive dimension of $\mathrm{sd}(\lambda_j^c) \times \mathrm{sd}(c_i) = 0.08 \times 0.24 = 0.019$.

Table 5: Moments due to skill returns under random versus actual sorting.

	Mean ×100		Standard deviation		Skewness	
	Random	Actual	Random	Actual	Random	Actual
	(1)	(2)	(3)	(4)	(5)	(6)
$\lambda_j^c c_i$	0.00	0.75	0.05	0.05	0.52	0.90
$\lambda_j^n n_i$	0.00	0.13	0.03	0.03	0.34	0.68
$\lambda_j^c c_i + \lambda_j^n n_i$	0.00	0.88	0.05	0.06	0.28	0.55

Notes: Central moments of the distribution of skill returns assuming either the actual allocation or a counterfactual where workers are randomly assigned to firms. Mean earnings $\mu \equiv \mathrm{E}(\lambda_j^c c_i + \lambda_j^n n_i)$ rise due to matching (compare Columns (2) and (1)). Dispersion $\sigma \equiv \mathrm{sd}(\lambda_j^c c_i + \lambda_j^n n_i)$ rises modestly (Column (4)) compared to random assignment (Column (3)). Skewness $\tilde{\mu}_3 \equiv \mathrm{E}[(\lambda_j^c c_i + \lambda_j^n n_i - \mu)/\sigma]^3$ is almost twice as large relative to random assignment (last two columns). Baseline grouped estimation. Sample period: 1999–2008.

Random assignment. To construct a random assignment counterfactual, we re-weight all skill types within a firm according to their population share while preserving the observed firm size distribution. Table 5 compares the first three moments of the distribution of earnings to moments obtained under the assumption of random assignment.

As expected, the average returns' gap (comparing columns (1) and (2) in Table 5) is in line with the efficiency gains reported in Tables 3 and 4. The standard deviations of skills premia are only marginally different (columns (3) and (4)): this is not surprising since higher between-skill inequality in the non-random allocation (Figure 3) is offset by declines in within-skill inequality due to the similarity of worker skills within firms. The muted changes in the second moment of the distribution point to a subtle distinction highlighted in the theoretical literature (Becker and Chiswick, 1966; Sattinger, 1993; Lindenlaub, 2017): the most conspicuous changes induced by complementarities occur in the third moment of the earnings distribution. Columns (5) and (6) suggest that the skewness of log earnings is twice as large under the non-random allocation of workers. The matching of high skill individuals to high return firms results in a thickening of the right tail of the earnings distribution (Figure 3). To sum up, returns' heterogeneity provides a natural way to interpret the asymmetries in the distribution of earnings and reconcile models of sorting with the well-established evidence on between-firm variation. Since the distribution of firm sizes is unchanged in our counterfactuals, sorting has no effect on the moments of firm intercepts λ_j^0.

6 Extensions and Robustness

Does the assignment of skill vary with industry and occupation? In robustness exercises we test for return heterogeneity within narrowly defined industry and occupation groups. To aid interpretation of our baseline findings, we also examine the correlation of skill returns with firm characteristics. Specifically, we link external data about firms' balance sheets and innovation to our sample of employers and examine the relationship between capital composition and skill returns. Finally, we assess the sensitivity of our estimates to alternative firm clustering approaches and to potential measurement error in skill proxies.

6.1 Industry and Occupation Specific Skill Returns

To explore whether skill returns reflect sector and job characteristics, we augment the baseline specification (6) with a full set of industry $\times$ occupation interactions with cognitive and noncognitive traits. Estimates in Appendix F.1 show that fine industry and occupation-specific skill returns (returns that vary by industry $\times$ occupation group) account for a minor share of firm-level heterogeneity. This confirms that most of the returns' heterogeneity occurs at the firm level, as opposed to the more aggregate industry or occupation level.

Aggregating to industry or occupation. While most return heterogeneity occurs at the firm level, some industries or occupations may exhibit higher skill returns on average. To explore this possibility, we project the baseline λ_j^c and λ_j^n estimates on industry-sector indicators and on employment shares for occupation groups (Appendix F.1). We find that high cognitive returns are frequent in Business-Services and IT as well as in firms with a large share of professional occupations. Noncognitive returns tend to be higher in the personal services sector and in firms that have large shares of managerial and services/sales jobs.

6.2 Capital Composition, Innovation, and Skill Returns

We explore some potential sources of returns' heterogeneity by linking balance sheet and innovation data to the sample of employers.

Capital composition. Differences in capital composition may reflect systematic aspects of productive and organizational structure. An advantage of the Swedish setting is that most private sector firms are limited liability corporations with publicly available financial statements. We use their balance sheet data to measure different capital components per employee. We aggregate workplaces to the organization level where this information is reported. In what follows we refer to these aggregates as firms. Results in Appendix F.2 show that intangible capital (especially patents, licenses, and capitalized R&D expenses) is strongly positively associated to cognitive returns. The notion that intangible capital and intellectual property are complementary to high skilled labor within a firm is consistent with production arrangements that leverage innovation. Physical assets and machinery, on the other hand, are larger items in firms that exhibit lower returns to cognitive skills. This is unlike noncognitive skills, whose returns are modestly higher in firms with more physical capital. The findings support the view that skills should be modeled separately rather than collapsed into a single index.

Measures of innovation activities. To further qualify these findings, we use the Swedish version of the European Community Innovation Survey (CIS) and examine the relationship between skill returns and innovation activities. In each wave of the CIS, a representative sample between 2,000 and 5,000 firms reports whether they conducted any product (including new services) or process (including organizational structure) innovations in the survey year or the preceding two years. Lindner et al. (2022) argue that the CIS provides direct and reliable measures for different types of firm-level technological change. Results in Appendix F.2 show that innovation activities are positively, and almost linearly, associated with cognitive returns. This is especially apparent in the case of product innovations where, moving from the lowest to the highest λ_j^c firms, the share of firms which introduce such innovations rises from 25 to 65 percent. Innovation expenditures (inputs) are larger for higher λ_j^c firms, suggesting that high return firms differ in their ability to bring forward innovations. This reinforces the evidence from studies linking cognitive skills to worker-level innovation (Aghion et al., 2023; Bell et al., 2018).

6.3 Changing the Cluster Design

Does the firm heterogeneity's contribution to earnings dispersion vary with the number of firm classes and the observables used to cluster firms? Using only ten classes (Bonhomme et al., 2019; Lamadon et al., 2022) marginally lowers the absolute contribution of firm heterogeneity and raises the importance of skill returns relative to the intercepts. Using more observables to cluster firms (e.g., firm employment and the standard deviations of earnings and skills within the firm) delivers estimates of firm effects in line with baseline estimates. If we only use data on within firm earnings to define classes (Bonhomme et al., 2019; Lamadon et al., 2022), skill returns' contribution does not change significantly relative to the case where many observables are used. All these estimates are in Appendix Table F.6.

Number of clusters. The baseline grouping with one hundred clusters delivers conservative estimates of the contribution of firm heterogeneity to earnings dispersion. Appendix Figure F.6 shows the standard deviation of log earnings attributed to different layers of firm heterogeneity when we increase the number of clusters from 20 to 200. Under the assumption of only twenty firm clusters, the impact of skill return heterogeneity is substantial, with a contribution of 5 log points to dispersion as opposed to 9 log points due to firm intercepts. Increasing the number of clusters results in a stronger influence of firm heterogeneity on overall inequality, and the absolute values of firm effects estimated from finer clusters become similar to those obtained using the bias correction approach with no clustering. Interestingly, the relative contribution of each layer of heterogeneity is stable throughout. When no clustering is imposed and estimates are adjusted using the bias correction method, the absolute impact of firm heterogeneity on earnings is larger but the relative impact of different components (intercept vs skill returns) is unchanged (Table 3). This confirms that alternative empirical approaches deliver comparable estimates of the relative contribution of skill return heterogeneity to firm-level variation.

6.4 Measurement Error in the Skill Measures

In Appendix C we examine the potential impact of measurement error in skill proxies. First, we consider the impact of different degrees of (classical) measurement error on estimates of returns' dispersion. In this way we show that measurement error has a discernible effect but does not alter the key insights of our analysis. For example, halving the reliability ratio of cognitive measurements (denoted by r^c, see definition in appendix) reduces the estimated dispersion of cognitive returns by about 30%. Similar impacts are present in noncognitive returns when we halve the reliability of the noncognitive measure (denoted by r^n). Introducing additional noise in the skill measures leads to roughly linear responses in the downward bias of the estimated dispersion of returns, $\text{sd}(\lambda^c)$ and $\text{sd}(\lambda^n)$. We find no evidence of spillover effects: that is, additional noise in one attribute (say, cognitive) does not materially affect the dispersion of the other attribute's returns (say, noncognitive).

Given these findings, in the second part of Appendix C we suggest a procedure to quantify the impact of measurement error on our baseline estimates. This analysis accommodates alternative assumptions about the intensity of measurement error. Assuming conventional reliability ratios for the skill proxies (Lindqvist and Vestman, 2011; Grönqvist et al., 2017), we show that our baseline estimates are rather conservative. For example, setting reliability ratios to $r^c \in [75\%, 85\%]$ and $r^n \in [50\%, 70\%]$, cognitive returns' dispersion would be around 0.09–0.095 and noncognitive returns' dispersion around 0.06–0.065. A comparison of these values to the estimates reported in Table 1 (namely, 0.08 and 0.05) indicates that the contribution of skill returns to wage dispersion may be larger than our baseline estimate. Further lowering the reliability of skill measures results in even higher heterogeneity in skill returns.

7 Conclusion

We present evidence of skill returns' variation across employers and of worker–firm complementarities. To identify the distinct layers of firm heterogeneity, we link administrative employer–employee population records to high-quality information about the cognitive and noncognitive attributes of workers. We adopt several approaches to estimate firm-level parameters. Each approach imposes different restrictions; however, estimates of the relative magnitude of skill returns, as opposed to skill-independent firm fixed effects, are stable throughout.

Our findings can be summarized as follows: (1) Similar skills command different returns across firms. Differences occur along both the cognitive and noncognitive dimension. Within-firm returns to each attribute are weakly correlated with one another. (2) Returns heterogeneity generates incentives for sorting; indeed, workers with larger endowments of cognitive and noncognitive skills populate firms with higher returns to those attributes. The intensity of sorting in each skill dimension depends on the dispersion of that skill's return across firms; as dispersion grows, so does the incentive for skilled workers to seek a better match. (3) The gains from sorting across employers are nonmonotonic in worker skills. High skill workers benefit from heterogeneity in returns. Considerable costs are borne by workers with intermediate skills because of the opportunity cost from matching with less productive firms. The least skilled workers experience little loss from firm heterogeneity as their skill returns are low regardless of the employer. (4) Positive assortative matching has implications for the distribution of earnings, which becomes more skewed due to the matching of high skills to high returns. Sorting boosts average earnings and the economy-wide skill premium. (5) Data from firms' balance sheets indicate that firms with high cognitive returns engage in more innovation and hold more intellectual capital. Noncognitive returns, however, do not vary with intellectual capital and are only modestly higher in firms with more physical capital. This discrepancy lends support to the view that skills should be modeled separately rather than collapsed into a single index.

References

ABOWD, J. M., R. H. CREECY, F. KRAMARZ, ET AL. (2002): "Computing person and firm effects using linked longitudinal employer-employee data," Tech. rep., Center for Economic Studies, US Census Bureau.

ABOWD, J. M., F. KRAMARZ, AND D. N. MARGOLIS (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

ACHLIOPTAS, D. (2003): "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of computer and System Sciences*, 66, 671–687.

AGHION, P., U. AKCIGIT, A. HYYTINEN, AND O. TOIVANEN (2023): "Parental education and invention: the Finnish enigma," *International Economic Review*, 64, 453–490.

ANDREWS, M. J., L. GILL, T. SCHANK, AND R. UPWARD (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, 673–697.

BEAUDRY, P., D. A. GREEN, AND B. M. SAND (2016): "The Great Reversal in the Demand for Skill and Cognitive Tasks," *Journal of Labor Economics*, 34, 199–247.

BECKER, G. S. AND B. R. CHISWICK (1966): "Education and the Distribution of Earnings," *The American Economic Review*, 56, 358–369.

BELL, A., R. CHETTY, X. JARAVEL, N. PETKOVA, AND J. VAN REENEN (2018): "Who Becomes an Inventor in America? The Importance of Exposure to Innovation*," *The Quarterly Journal of Economics*, 134, 647–713.

BONHOMME, S., K. HOLZHEU, T. LAMADON, E. MANRESA, M. MOGSTAD, AND B. SETZLER (2023): "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics*, 41, 291–322.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

BOROVICKOVA, K. AND R. SHIMER (2020): "High Wage Workers Work for High Wage Firms," Working Paper.

CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

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

CHONÉ, P. AND F. KRAMARZ (2021): "Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling," Tech. rep., Center for Research in Economics and Statistics.

DEMING, D. J. (2017): "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 132, 1593–1640.

EDIN, P.-A., P. FREDRIKSSON, M. NYBOM, AND B. ÖCKERT (2022): "The Rising Return to Noncognitive Skill," *American Economic Journal: Applied Economics*.

EDMOND, C. AND S. MONGEY (2021): "Unbundling Labor," Tech. rep., Working Paper.

EECKHOUT, J. AND P. KIRCHER (2018): "Assortative matching with large firms," *Econometrica*, 86, 85–132.

FREDRIKSSON, P., L. HENSVIK, AND O. N. SKANS (2018): "Mismatch of talent: Evidence on match quality, entry wages, and job mobility," *American Economic Review*, 108, 3303–38.

GRÖNQVIST, E., B. ÖCKERT, AND J. VLACHOS (2017): "The intergenerational transmission of cognitive and noncognitive abilities," *Journal of Human Resources*, 52, 887–918.

HAGEDORN, M., T. H. LAW, AND I. MANOVSKII (2017): "Identifying equilibrium models of labor market sorting," *Econometrica*, 85, 29–65.

HECKMAN, J. AND J. SCHEINKMAN (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies*, 54, pp. 243–255.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor Economics*, 24, 411–481.

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

KLINE, P., R. SAGGIO, AND M. SØLVSTEN (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

LACHOWSKA, M., A. MAS, AND S. A. WOODBURY (2020): "Sources of displaced workers' long-term earnings losses," *American Economic Review*, 110, 3231–66.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2022): "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market," *American Economic Review*, 112, 169–212.

LENTZ, R., S. PIYAPROMDEE, AND J.-M. ROBIN (2023): "The Anatomy of Sorting – Evidence from Danish Data," *Econometrica*, forthcoming.

LINDENLAUB, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84, 718–789.

LINDENLAUB, I. AND F. POSTEL-VINAY (2023): "Multi-dimensional sorting under random search," *Journal of Political Economy*, forthcoming.

LINDNER, A., B. MURAKÖZY, B. REIZER, AND R. SCHREINER (2022): "Firm-level Technological Change and Skill Demand," *Working Paper*.

LINDQVIST, E. AND R. VESTMAN (2011): "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment," *American Economic Journal: Applied Economics*, 3, 101–128.

LISE, J. AND J.-M. ROBIN (2017): "The macrodynamics of sorting between workers and firms," *American Economic Review*, 107, 1104–35.

MCFADDEN, D. (1974): "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, 105–142.

ROSEN, S. (1983): "A note on aggregation of skills and labor quality," *The Journal of Human Resources*, 18, 425–431.

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

SKANS, O. N., P. CHONÉ, AND F. KRAMARZ (2022): "When Workers? Skills Become Unbundled: Some Empirical Consequences for Sorting and Wages," *Working Paper*.

SONG, J., D. J. PRICE, F. GUVENEN, N. BLOOM, AND T. VON WACHTER (2018): "Firming up inequality," *The Quarterly Journal of Economics*, 134, 1–50.

SØRENSEN, T. AND R. VEJLIN (2013): "The importance of worker, firm and match effects in the formation of wages," *Empirical Economics*, 45, 435–464.

SORKIN, I. (2018): "Ranking firms using revealed preference," *Quarterly Journal of Economics*, 133, 1331–1393.

WILLIS, R. J. (1986): "Wage determinants: A survey and reinterpretation of human capital earnings functions," in *Handbook of labor economics*, Elsevier, vol. 1, 525–602.

WOODCOCK, S. D. (2015): "Match effects," *Research in Economics*, 69, 100 – 121.