HCEO WORKING PAPER SERIES

Firm Heterogeneity in Skill Returns

Michael J. Böhm
Khalil Esmkhani
Giovanni Gallipoli

Working Paper

2020-082

11/2020; Revision 05/2022

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

Firm Heterogeneity in Skill Returns

Michael J. Böhm
Khalil Esmkhani
Giovanni Gallipoli *

MAY 2022

Abstract

We quantify firm heterogeneity in skill returns and present direct evidence of worker–firm complementarities. Using population data linked with cognitive and noncognitive skill measures, we estimate a model of firm-specific returns to these attributes. We find evidence of significant return heterogeneity, sorting, and earnings convexification: (1) Skills command different returns across employers; returns to the two skills correlate weakly within-firm. (2) Workers with large endowments of a skill populate firms with higher returns to it. Sorting intensity grows with cross-sectional dispersion of that skill return. (3) Complementarities and sorting have nonmonotonic effects, raising both level and skewness of earnings.

Keywords: Firm Heterogeneity, Skill Returns, Sorting, Earnings Distribution
JEL Classification: E24, J23, J24, J31

*Böhm: University of Bonn, Department of Economics (michael.j.boehm@uni-bonn.de); Esmkhani: University of British Columbia, Sauder School of Business (khalil.esmkhani@sauder.ubc.ca); Gallipoli: University of British Columbia, Vancouver School of Economics (gallipol@mail.ubc.ca). We thank many seminar and conference participants for comments and suggestions. We are grateful to Raffaele Saggio for discussions about the implementation of bias-correction methods and of related tests. Shuahib Habib, Ronit Mukherji and Dongxiao Zhang provided valuable research assistance. Böhm acknowledges funding from the German Science Foundation (DFG) under a Research Fellowship – BO 4765/1-2 – and Germany's Excellence Strategy – EXC 2126/1 – 390838866. Gallipoli acknowledges support from Canada's SSHRC. Any errors are the responsibility of the authors alone.

1 Introduction

The recognition that earnings distributions reflect both worker and firm heterogeneity dates back decades. Robert Willis notably warned about "an imbalance in the human capital literature which has emphasized the supply far more than the demand for human capital" (Willis, 1986). The availability of matched employer–employee records has brought about a renewed interest in firm-level differences (e.g., Card et al., 2013; Song et al., 2018; Sorkin, 2018; Lamadon et al., 2022). A workhorse of this literature is the Abowd et al. (1999, AKM) two-way fixed effect model, which subsumes unobserved heterogeneity of workers and firms into additively separable measures whose contributions to the dispersion of earnings can be transparently quantified. The correlation between firm and worker fixed effects is often interpreted as evidence of nonrandom sorting of workers across employers, or lack thereof. However, several studies (Eeckhout and Kircher, 2011; Hagedorn et al., 2017; Borovickova and Shimer, 2020) caution against drawing inference about match-specific productivity from fixed effect estimates, emphasizing that complementarity is hard to characterize within the boundaries of additively separable models of worker and firm heterogeneity. Such additively separable effects are also unsuitable to examine the endogenous skewness of wages emphasized in matching and assignment models. These considerations inform empirical frameworks that nest flexible matching mechanisms within two-sided unobserved heterogeneity (e.g., Bonhomme et al., 2019; Lentz et al., 2018).

This paper presents novel and direct evidence on worker–firm complementarities, matching, and their effects on earnings. To this end, we link cognitive and noncognitive test scores with population data on Swedish workers and firms, and employ distinct empirical approaches to robustly estimate firm-level returns to skill attributes. Our estimates reveal significant heterogeneity in skill returns, with some firms paying up to 35 log points more than others for similar cognitive and noncognitive attributes.[1] This leads to strong incentives for sorting of workers with different skill endowments across firms. We show that heterogeneous returns, and the induced sorting, materially impact both the level and the distribution of earnings.

[1] As we show, skill premia can be derived in labor market models with two-sided heterogeneity. Establishing their empirical prevalence and implications is, however, demanding in terms of data requirements and estimation.

The cognitive and noncognitive measures, elicited for almost all Swedish males, have been used in several studies that document their information content[2] and establish their relation to distinct productive attributes (Edin et al., 2022). In our high-dimensional estimation of firm returns, we employ these measures in conjunction with alternative approaches to address limited mobility biases in the estimation of returns and of their quadratic forms. One approach builds on clustering methods (Bonhomme et al., 2019) whereby we group firms into 100 classes based on the earnings and skills of their employees. The other delivers estimates of quadratic forms of the parameters of interest at the individual (non-grouped) firm level after explicitly correcting for biases (Kline et al., 2020). Each approach imposes different sample restrictions and assumptions. Results, however, are remarkably robust in the sense that the relative importance of skill returns, as opposed to conventional measures of firm heterogeneity based on fixed effects, is stable and does not depend on the approach or specific implementation choices. For either approach, estimation of different layers of firm heterogeneity requires significant computational work, which we discuss below.

To motivate our focus on the heterogeneity of skill returns, we begin by estimating standard fixed-effect models separately for high versus low skill workers. The hypothesis that earnings premia at a given firm are the same across skill levels is clearly rejected for both cognitive and noncognitive traits. Informed by this finding, we develop an empirical specification that flexibly allows for granular returns to each skill attribute. To facilitate comparisons to existing work, the specification is derived from a monopsonostic model of labour demand (Card et al., 2018; Lamadon et al., 2022). The model delivers a first-order approximation for a general wage function in which skill $\times$ firm interaction terms reflect heterogeneous returns, while firm intercepts capture skill-independent premia. As we show, standard Mincer returns are a key part of the empirical representation despite being subsumed into worker fixed effects.

Our estimates reveal considerable dispersion in returns across firms in either skill dimension, and relatively more in the cognitive one. The correlation between returns to different skills is positive but weak; this suggests that collapsing cognitives and noncognitives into a single composite

[2]For example, Lindqvist and Vestman (2011) show that the military test scores are highly significant in predicting earnings and unemployment conditional on any rich set of control variables. Fredriksson et al. (2018) use them to identify the effects of job–skill mismatch on labor mobility and life-cycle wage growth. See also our Appendix A.1.

index might be restrictive when examining complementarities and sorting. Estimates show that returns heterogeneity induces material gains from the assignment of workers to firms, generating earnings gaps of the same order of magnitude as those induced by firm intercepts.

To gauge the intensity of sorting we employ analytical notions developed in multidimensional assignment problems (Lindenlaub, 2017). Several testable restrictions implied by positive assortative matching are supported in the data. 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, 2020). Further corroboration of assortative matching is obtained by projecting firm-level returns onto skill measures (for a discussion of such projection methods, see Kline et al., 2020) as well as by regressing skills onto estimates of returns for different firm clusters. Sorting occurs along both skill dimensions but is stronger in the cognitive one where firm heterogeneity is larger.

The heterogeneity in returns, and the induced sorting, have significant but uneven effects on the moments of the earnings distribution. First, we show that matching increases aggregate efficiency and raises average earnings compared to a counterfactual random allocation of workers to firms. Moreover, earnings differences between different skill levels are strongly convexified by sorting. That is, earnings at the top are magnified by the interaction of skills and returns, while earnings of middle-to-low skill workers suffer a relative decline because they are frequently matched with low-return firms. Compared to random assignment, intermediate skill workers suffer more than the lowest-ability ones since the latter would hardly benefit from higher returns due to their meager skill endowments. Consistent with other studies (e.g., Bonhomme et al., 2019; Hagedorn et al., 2017; Lentz et al., 2018; Borovickova and Shimer, 2020; Lamadon et al., 2022), we find that match effects raise earnings levels and dispersion. A notable impact of worker–firm complementarities is on the skewness of earnings, which become more convex in skill levels. Such effects have long been discussed in the theoretical literature (Becker and Chiswick, 1966; Sattinger, 1993; Lindenlaub, 2017; Becker et al., 2018).

To validate the baseline findings we consider a few extensions and sensitivity checks. Notably, we find that estimates of the relative magnitude of skill returns do not visibly change with the number of firm classes when using the clustering approach. The same is true when we implement alternative sampling restrictions in the firm-level estimation with bias correction of quadratic-forms. Moreover, controlling for industry- or occupation-specific effects illustrates that returns heterogeneity across firms is quantitatively large even within narrow sectors and occupations.

To probe the nature of firm differences, and in keeping with our emphasis on direct measures, we link information from their balance sheets to the main data and show that employers exhibiting high cognitive returns have significantly different capital composition, with more intangible and intellectual assets (as opposed to physical capital) per worker. Moreover, after merging additional firm survey responses, we find that these firms invest more heavily in R&D and introduce product and process innovations more frequently. This lends support to the view that production and organizational arrangements play a key role in shaping the distribution of skill returns.

Our findings support the hypothesis that substantial worker–firm complementarities exist, that they bring about assortative matching, and that they influence earnings. In doing so, we draw attention to a less explored but important dimension of firm heterogeneity. More generally, we find direct evidence of efficient, albeit imperfect, skill assignment across employers. The use of skill measures complements existing studies of worker–firm interactions and presents a transparent counterpart as it does not require tight model restrictions for the identification of unobserved attributes. Resorting to informative skill proxies facilitates the measurement of gains from matching because pecuniary returns are not themselves used to recover skills ranks. This is especially advantageous when establishing which workers benefit or lose from returns' heterogeneity and sorting, as well as to identify the impact of complementarities on skewness.

One aspect that can play an important role in the imperfect assignment of skills to jobs is their multidimensional and bundled nature. A longstanding literature has examined selection and wages in settings where workers are endowed with multiple skills (see early work in Mandelbrot, 1962; Rosen, 1983; Heckman and Scheinkman, 1987). Our estimates suggest that cognitive and noncognitive returns heterogeneity have independent impacts on earnings, thus providing further motivation for research on the implications of 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).

Finally, our findings provide new evidence on the nature and evolution of returns to skills in the labor market. Previous work has shown that both cognitive and noncognitive attributes shape individual outcomes (e.g., Heckman et al., 2006; Lindqvist, 2012). Moreover, the *average* gains from these skills have changed over time (Beaudry et al., 2016; Deming, 2017; Edin et al., 2022). Our analysis implies that each worker's return will depend on both their skill and their match with an employer. For this reason, conventional measures of Mincerian returns are not equivalent to the averages of individual returns across employers. Rather, assortative matching can tangibly change overall skill premia in the economy while inducing uneven and nonmonotonic effects over the skill range. This points to the need to explore the determinants of firm heterogeneity in skill returns as well as its implications for matching over time.[3]

The paper is organized as follows. Section 2 describes the data and the methods used to account for incidental parameter biases in the estimation of firm-level variables. Section 3 derives the empirical specification within a labour market model with two-sided heterogeneity and presents new estimates on firm heterogeneity in skill returns. Section 4 illustrates how workers match with firms based on skills and returns, and provides different tests of the assortative matching hypothesis. Section 5 examines finer implications of heterogeneity and worker sorting for the distribution of earnings, and of gains and losses, relative to random assignment of skills. Section 6 explores the sensitivity of key results to alternative estimation approaches; in the same section we use ancillary information about firm heterogeneity (from surveys and balance sheets) to examine the correlation of firm returns with capital composition, production arrangements, and innovation activities. The last section concludes.

2 Data and Preliminary Evidence

2.1 Matched Earning Records and Skill Measures

Our data source consists of annual employer–employee matched records for the whole population of Swedish workers and firms during 1990–2017, including earnings, industry, occupation, and worker characteristics such as age, gender, and education. A key strength of these data are cognitive and noncognitive military enlistment tests that can be linked to individual workers. The

[3] In the context of unconditional firm premia, captured in wage intercepts, influential work has related firm heterogeneity to wage growth and inequality (e.g., Card et al., 2018; Lamadon et al., 2022), or linked it back to primitives such as market structure, institutions, and policy (De Loecker and Eeckhout, 2021; Dustmann et al., 2022).

tests were mandatory before 2007 and are available for almost 90 percent of males, across birth cohorts, in our sample.

The cognitive score is similar to an IQ measure and 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 highly significant at predicting workers' earnings and other labor market outcomes (e.g., Lindqvist and Vestman, 2011; Fredriksson et al., 2018; Edin et al., 2022), on their own as well as conditionally on each other and any rich set of control variables. Cognitive and noncognitive measures are recorded on a standard-nine (Stanine) scale, which approximates the Normal distribution and facilitates comparisons across birth cohorts.[4] In online Appendix A.1 we discuss these tests in detail and show that, while assessed at age 18–19, their scores are strongly associated with earnings over the entire life-cycle.

Due to the availability of test scores, we restrict the sample to males aged 20–60 with nonmissing scores. We also restrict attention to firms that employ an average of at least ten male workers over five years or more. We focus on estimates from 1999–2008 but 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.

Our dataset reports both organization and workplace identifiers. To identify "firms" we use the workplace with the highest income that year, since workplace is closest to the notion of a production unit and is consistent with existing work (e.g., Card et al., 2013). We also use the annual labor income at the firm, which is available for all workers and includes bonuses and performance pay, as our measure of earnings throughout. Details and descriptive statistics are in Section A of the online Appendix.

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 Sur-

[4] 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.

vey). 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 Effects Models

Our analysis requires estimation of models with many fixed effects for firms and workers as well as firm-specific returns to skill measures in matched employer–employee records. We then compute variance components of these parameters or project them onto firm observable characteristics. These steps require restrictions on the data samples and empirical methods we adopt.

Connected sets. To identify model parameters, firms need to be connected to each other through worker mobility in the final sample. This entails working with a connected component of the firm–worker graph (Abowd et al., 2002; Bonhomme et al., 2020). Distinct connected sets may exist within a large sample of employment matches and empirical analyses often focus on the largest set (or 'maximally connected subgraph'). When considering different skill levels (say high and low cognitive skills) the requirement is that we use a set which is connected for each skill level ("dual" or "double" connected in Card et al., 2016; Kline et al., 2020, respectively). As we show below, the connectedness restrictions become less stringent when observations are defined at the level of firm clusters rather than individual firms.

Limited mobility bias. While connectedness leads to unbiased identification of model parameters, researchers are usually interested in variance components. These are in general biased because of sampling error in individual parameter estimates that enter the variance components in a quadratic form. The squared sampling error is thus not mean zero and may not converge to zero as the number of firms increases. Intuitively, the bias arises from an insufficient number of movers into and out of the firm, hence "limited mobility bias", and it tends to overstate variances and understate covariances (Andrews et al., 2008). The magnitude of the bias is inversely related to the degree of connectivity of the firm-worker graph, with the graph being disconnected as limiting case (Jochmans and Weidner, 2019). For further details, see Bonhomme et al. (2020, Section 3).

Since we are interested in the dispersion and correlation of skill return parameters, our analysis is potentially subject to the limited mobility bias. Fortunately, the literature on panel data has made good progress in addressing this problem. One approach, suggested in Bonhomme, Lamadon, and Manresa (2019), defines the relevant level of firm unobserved heterogeneity as the "class" of a firm, corresponding to a cluster of similar employers. While the class can be made arbitrarily close to an individual firm, this may not be desirable because the number of job movers per firm will become smaller and result in an incidental parameters bias (i.e., reinstate the limited mobility problem). Under the assumptions of this approach, unobservable firm heterogeneity operates at the level of firm classes. The latter can be estimated in a first step through k-means clustering based on earnings and skills within each firm. This achieves two objectives: first, it enhances tractability; second, it delivers well-centered and accurate estimates of the contributions of worker and firm heterogeneity to earnings dispersion. Clustering trades off restrictions on the dimensionality of the underlying groups for increased connectedness between firm classes. Notably, this method does not require shedding observations to generate a connected set.

A different approach builds on variance component estimators designed for unrestricted linear models with heteroscedasticity of unknown form. This removes the bias by resorting to leave-out estimators of the variances of errors from the linear model. For each observation, an estimate of the error variance is obtained from a sample where that worker–firm match observation is left out. The leave-out procedure delivers unbiased estimates in finite samples (see Kline, Saggio, and Sølvsten, 2020) and facilitates tests of linear restrictions. It can be implemented as a simple variance component estimator plus a bias correction consisting of observation-specific error variances. The leave-out strategy to estimate the linear model parameters requires that firms remain connected by worker mobility when any single mover is dropped. This involves pruning the original sample to ensure that the connectedness condition is met by all leave-out subsamples.[5] Given the large scale computations involved in the estimation of leave-out quadratic forms, which are executed at the individual firm level, we resort to the random projection method of Achlioptas (2003). This reduces dimensionality and computation time (see Appendix B).

[5]We use Python NetworkX to identify the articulation points of the worker–firm graph and trim it to construct the double leave-one-out connected set. See Appendix A.2 for details on the construction of estimation samples.

Implementation of the clustering and bias correction approaches. Implementation of the grouping approach requires clustering firms into classes. We define 100 classes using a k-means algorithm based on average earnings as well as average cognitive and noncognitive skills of workers. Having a sufficiently large set of classes accommodates rich heterogeneity and ensures stability while still delivering a major dimension reduction. Using information beyond wages has been proposed in the structural literature (Eeckhout and Kircher, 2011; Hagedorn et al., 2017; Bartolucci et al., 2018; Bagger and Lentz, 2019). In our implementation this is further motivated by the theoretical restriction that firm-specific production arrangements affect both the skill composition of the workforce and their wages. Alternative clustering criteria (e.g., adding within-firm dispersion of skills and wages, or employment levels) as well as alternative numbers of classes deliver similar results (Section 6.3). The availability of skill measures makes it feasible to estimate specifications that feature firm effects in both levels and returns. Previous work has shown that class membership and fixed effects can be accurately estimated with sufficiently many workers. Using skill proxies also avoids incidental parameter biases in estimated returns due to few panel observations per worker.

Implementation of the bias correction approach relies on the leave-one-out double-connected set of firms. We prune the sample to contain firms that remain connected along both skill dimensions (cognitive and noncognitive) for different levels (high and low) when each single observation is dropped. The implementation accounts for correlation of error terms within an individual's spell at a given employer (Kline et al., 2020). This is done by averaging the data to the worker–firm match level. The resulting leave-match-out set is double-connected (in both skill dimensions) and smaller than the original sample but allows for estimation at the individual firm level. The extensive size of the Swedish population data assuages concerns about sample sizes. Appendix B discusses theoretical details of each approach and their implementation. In Table A.1 we report statistics for the underlying samples.

2.3 A First Glance at Skill Returns

We begin by examining whether labor market returns entail two firm-specific components: (i) a base wage common to all workers within a firm, irrespective of their attributes; (ii) a skill return.

The hypothesis of returns heterogeneity, in its most basic form, can be tested with binary skill levels (high or low test scores). Therefore we consider a well-known specification (Abowd et al., 1999, AKM) where firm fixed effects for high and low skill workers are allowed to differ.

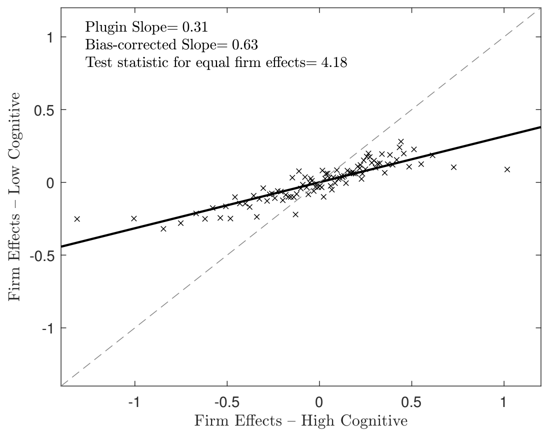
For the purposes of this section we construct subsamples corresponding to the largest connected sets of, respectively, high and low ability workers and we condition on firms that are in both of these sets (double-connectedness in skill levels). Since the analysis is carried out separately for cognitive and noncognitive attributes, this is relatively straightforward and does not require that firms be linked through mobility of both skill dimensions. However, in the following sections we examine set connectedness for the case where multiple skills are considered in the same specification.

We classify workers into high cognitive (Stanine $C = \mathbb{1}[c > 5]$) and high noncognitive ($N = \mathbb{1}[n > 5]$). Then, to exclude potential serial correlation within employment spells from estimated standard errors, we select observations within a two-year set and separately estimate linear binary models of worker and firm effects of the form:

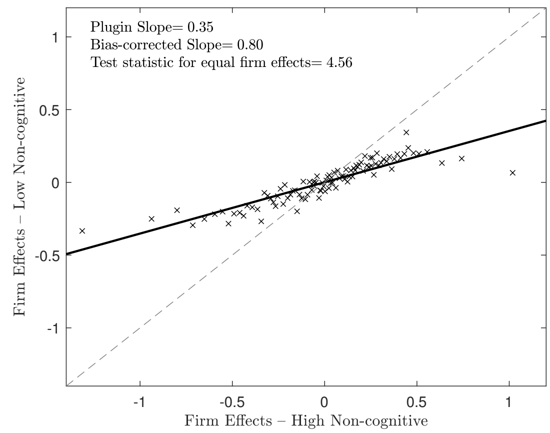
$$\log(w_{ijt}) = \mu_i^S + \theta_j^S + \varepsilon_{ijt}, \tag{1}$$

where $S \in \{C = 0, C = 1\}$ or $S \in \{N = 0, N = 1\}$ indicates the skill subsample while t takes on the values of the two years selected out of the 1999 to 2008 period. Figure 1 plots results for $t \in \{2004, 2007\}$. We use non-adjacent years (in fact, we employ pairs two years apart) to mitigate the impact of partial employment spells during contiguous years when workers switch firms. Various other year pairs are reported in Appendix B.3.

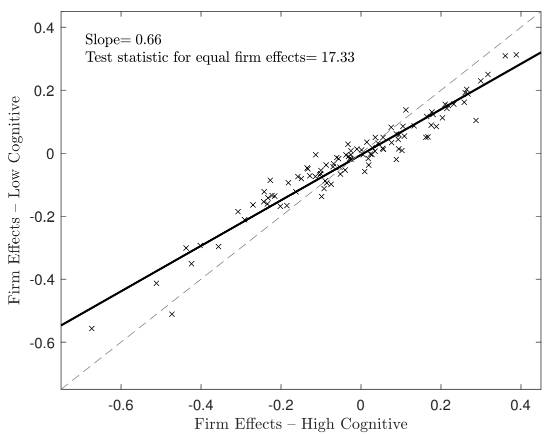
As noted before, even in connected samples one should be wary of incidental parameter bias due to limited worker mobility. A simple comparison of firm effects (θ_j^S) for high and low skill workers illustrates this point and shows that the statistics of firm effects (like their variances and correlations) can be biased if identified from few moves of workers into and out of each firm. Panels (a)–(b) in Figure 1 plot a scatter of estimated firm fixed effects for high-skill (x-axis) and low-skill (y-axis) workers. The samples consist of firms that are in the leave-one-out connected sets of both high and low ability workers. Each panel refers to a given skill attribute, covering the years 2004 and 2007. Panel (a) shows results for cognitive skills (9,268 firms) while Panel



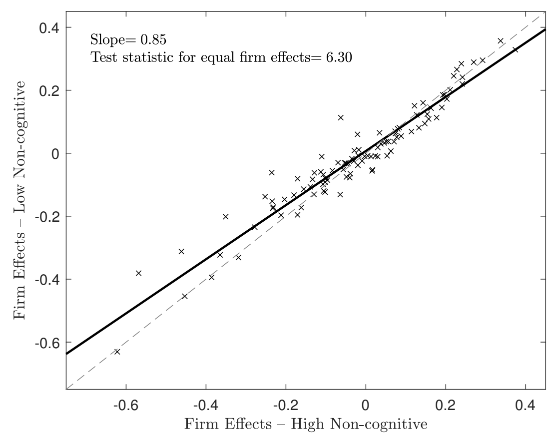
(a) High vs low cog skills – leave-one-out correction



(b) High vs low noncog skills – leave-one-out correction



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



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

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

Top panels: 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 $S \in \{C,N\}$. All sets of firm effects are demeaned. The sample in panel (a) consists of 9,268 firms that are leave-one-out connected in both high and low cognitive skills; in panel (b) we use 10,208 firms connected in both high and low noncognitive skills. The "plugin slope" is the coefficient from a person-year weighted projection of $\theta_j^{S=0}$ onto $\theta_j^{S=1}$. The "bias-corrected slope" adjusts the plug-in slope for attenuation bias by multiplying its value by the ratio of the plug-in estimate of the person-year weighted variance of $\theta_j^{S=1}$ to the bias-adjusted estimate of the same quantity. "Test Statistic" refers to the realization of $\hat{z}_{H_0}/\sqrt{v\hat{a}r(\hat{z}_{H_0})}$ where $\hat{z}_{H_0}$ is the quadratic form associated with the null hypothesis that the firm effects are equal across skill groups. From Theorem 2 in Kline et al. (2020), $\hat{z}_{H_0}/\sqrt{v\hat{a}r(\hat{z}_{H_0})}$ converges to a $N \sim (0,1)$ under the null hypothesis that $\theta_j^{S=0} = \theta_j^{S=1}$ for, respectively, all 9,268 and 10,208 firms.
Bottom panels: These 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.
Sample restriction: years 2004 and 2007 only. Tests for other year pairs are in Appendix Table B.1.

(b) plots those for noncognitives (10,208 firms). The figures show that ignoring estimation biases results in firm effects for high and low skill workers that are positively but weakly correlated within firms. The regression slope from mechanically projecting $\theta_j^{S=0}$ onto $\theta_j^{S=1}$ is 0.31 for cognitive traits and 0.35 for noncognitive. We refer to these slopes as the "plug-in" estimates and employ our two approaches to account for the attenuation biases.

The bias correction raises estimated slopes to 0.63 and 0.81, respectively. Under the null hypothesis of no heterogeneity in skill returns, however, the slopes should be statistically indistinguishable from one and the scatters should align along the dashed 45° lines. This is not the case, as the bias-corrected test statistics of equal firm effects for high and low skill workers have z-values above 4 for both cognitive and noncognitive returns. We therefore reject the hypothesis that firm effects are independent of worker skills. In fact, all our estimates indicate slopes that are well below one. Table B.1 in the appendix reports additional tests, which similarly reject the null hypothesis of homogeneous returns in several alternative samples.[6]

Panels (c)–(d) of Figure 1 show results when grouping firms into 100 clusters – as explained in Section 2.2 (see Bonhomme et al., 2019). Estimates of the slopes (0.66 for cognitive and 0.85 for noncognitive) are remarkably similar to those obtained using the quadratic form correction. Also in this case, the null hypotheses that firm effects are the same for high- and low-skill workers are strongly rejected, further discarding the notion of a homogeneous return to skills across firms. Tests for additional year pairs lead to similar conclusions and are presented in Table B.1.

3 Quantifying Variation in Skill Returns

The previous section emphasizes the significant differences in firm intercepts by skill level. However, to accurately examine the extent of variation in skill returns an empirical framework is needed that allows for granular differences in skill bundles while controlling for other sources of heterogeneity. To this end, we derive a richer empirical baseline from a simple model of demand

[6] All tests of parameters of equation (1) are based on an upper bound for the estimated error variance var(ε_{ijt}). This leads to conservative test statistics compared to the split-sample estimate in Figure 1 of Kline et al. (2020). Joint tests of the equal effects hypothesis across more than two periods are unfeasible as they introduce issues with clustering of errors at the firm level and no robust procedure is currently available to handle such issues. We thank Raffaele Saggio for feedback and discussions about implementing these tests.

for productive skills. Our theoretical restrictions aid in the interpretation of estimated parameters and facilitate comparisons with existing work.

3.1 Skill Demand by Heterogeneous Firms

We embed return heterogeneity in a model in which firms choose how many workers to hire based on demand for their output. We let firms differ in four dimensions: (i) cognitive returns; (ii) noncognitive returns; (iii) demand in their output market, where they have varying degrees of monopoly power; and (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, which underpins the cross-sectional variation in firm base-wages reflected in 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). These assumptions are sufficient to characterize the components of firm wage premia. To this structure we overlay a production technology with heterogeneous skill returns. Derivations are in Appendix C.

3.1.1 Production Complementarities

Consider an environment with two heterogeneous sides (workers, firms), with a measure one of workers who differ in their observable cognitive (c) and noncognitive (n) abilities. Firms are indexed by j and workers by the vector (c,n) of their skills. Firm j matched with a (c,n) worker produces output $y = f_j(c,n)$. Assuming technology is constant returns to scale (CRS) in worker headcounts, 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 $f_j(c_1,n_1) + f_j(c_2,n_2)$. [7] Hence the total output of firm j hiring fraction $q_j(c,n)$ of total (c,n) type workforce is

$$y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n). \tag{2}$$

where G is the population measure of different worker types in the economy.

[7] This is a variation of Eeckhout and Kircher (2018)'s assortative matching production setup for large firms and multiple skill inputs. The production function is defined at the level of the match (see Lise and Robin, 2017).

3.1.2 Labor Supply

A worker's utility from being matched with a specific firm depends on his wage plus a preference shock. For worker i of type (c,n), the utility of working at firm j with wage $w_j(c,n)$ is

$$u_{ij}(c,n) = \beta \log(w_j(c,n)) + \nu_{ij} \tag{3}$$

where ν_{ij} is the idiosyncratic preference for working at firm j. Shocks ν_{ij} are independent draws from a Type I Extreme Value distribution. This specification could be easily expanded to add firm-level variation in average amenities (Sorkin, 2018).

Workers choose the firms that give them the highest utility and, using standard arguments (McFadden, 1974), the share $q_j(c,n)$ of type (c,n) workers who choose firm j has logit form

$$\log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n)). \tag{4}$$

Equation (4) describes the upward sloping labor supply faced by firm j. The intercept $h(c,n)$ is determined in equilibrium and guarantees market clearing so that each worker gets a job,

$$h(c,n) = \left[\int w_j(c,n)^{\beta} \, \mathrm{d}F(j) \right]^{-1} \tag{5}$$

where $F(\cdot)$ is the measure describing the distribution of firms in the economy.

3.1.3 Technology and Wages

Given the simple structure outlined above, firm j's output is given by (2).

Final good. Each firm's output is an intermediate input for a final good Y produced by a representative firm through a CES technology, $Y = \left[\sum_{j=1}^{J} \phi_j y_j^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}$ where $\sigma > 1$ is the elasticity of substitution of intermediates. Each intermediate's share parameter ϕ_j is the marginal contribution of y_j to output Y and can be interpreted as the output market power of a firm.

Wages. In Appendix C we provide an analytical characterization of firm-specific wages offered to each skill set and define a stationary equilibrium in the labour market. We also show that a firm's optimal behavior implies:

$$w_j(c,n) = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{T_j \phi_j \frac{\sigma-1}{\sigma} \left(\frac{1}{y_j}\right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{e^{\Delta_j(c,n)}}_{\text{Skill Productivity}}. \tag{6}$$

The wage paid by firm j reflects different aspects of market structure and technology. The monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker. In turn, the marginal revenue is an increasing function of a firm's output market share ϕ_j and of its total factor productivity T_j. The latter parameter is normalized to $T_j = f_j(L,l)$, which is the output in firm j of a worker with the lowest cognitive and noncognitive ability. The premium $\Delta_j(c,n) = \log(f_j(c,n)/f_j(L,l))$ is the log output in firm j of a (c,n) type worker relative to a worker with the lowest skill endowment (L,l). The premium associated to the skill vector (c,n) depends on the firm's production technology and on (c,n)'s marginal contribution to output. Equation (6) is explicitly derived in Appendix C.

In logs, wages are the sum of a common level effect, a firm intercept and a skill return,

$$\log(w_j(c,n)) = \alpha + \Lambda_j + \Delta_j(c,n). \tag{7}$$

To obtain an empirical counterpart, we do not restrict the functional form of $f(\cdot)$, and hence of $\Delta_j(c,n)$, but rather use a first-order approximation that delivers a simple bilinear relationship for worker i in firm j.[8] Making the worker index i explicit, the empirical wage representation is:

$$\log(w_{i,j}(c,n)) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i, \tag{8}$$

where λ_j^0 is the baseline wage that a worker with the lowest endowment in both the cognitive and noncognitive dimension earns in firm j. Gradients λ_j^c and λ_j^n are firm-specific marginal returns,

[8]For instructive discussions of log-additive firm effects in wage specifications with bundled skills, see Choné and Kramarz, 2021. We explore higher order approximations featuring non-linear returns but the extra flexibility makes little difference. Notably, this type of 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).

above and beyond the baseline return λ_j^0. Finally, as we show below, the individual intercepts μ_i partly reflect the average (Mincerian) returns to a worker's skill endowments.

Normalizations. It is well known that specifications like (8) require linear restrictions on firm effects, since these are only identified relative to a reference firm (or set of firms). It follows that we can identify parameters up to a set of unknown constants $(\kappa_0, \kappa_\text{c}, \kappa_\text{n})$, such that:

$$\begin{aligned}
\lambda_j^0 &= \Lambda_j - \kappa_0 \\
\lambda_j^c &= \frac{\partial \Delta_j(c,n)}{\partial c} - \kappa_\text{c} \\
\lambda_j^n &= \frac{\partial \Delta_j(c,n)}{\partial n} - \kappa_\text{n} \\
\mu_i &= \alpha + \kappa_0 + \kappa_\text{c} \cdot c_i + \kappa_\text{n} \cdot n_i
\end{aligned} \tag{9}$$

We set $\kappa_0 = \bar{\Lambda}$, $\kappa_\text{c} = \frac{\partial \bar{\Delta}(c,n)}{\partial c}$, and $\kappa_\text{n} = \frac{\partial \bar{\Delta}(c,n)}{\partial n}$, 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. This normalization is quite conservative, since central moments yield the lowest variance of firm heterogeneity (intuitively, they minimize squared deviations).[9]

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 variation is a function of worker skill differences as firms with higher λ_j^c or λ_j^n exhibit higher skill premia.

3.2 Estimating Skill Returns

The empirical analysis relies on a sample of firms connected through worker mobility along both skill dimensions over the 1999–2008 period. The baseline representation becomes

$$\log(w_{ijt}) = \mu_i + \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i + \boldsymbol{X}_{it} \boldsymbol{b}_t + \varepsilon_{ijt}, \tag{10}$$

[9] More generally, estimates of central moments tend to be more robust relative to those of the extrema of the firm effects' distribution, which may suffer from non-trivial estimation error.

where λ_j^0 are skill-independent base earnings, λ_j^c and λ_j^n are skill gradients, and μ_i are worker fixed effects. To flexibly account for life-cycle and time variation by skill, we control for interactions of skill type, age, and years, denoted as $\boldsymbol{X}_{it}\boldsymbol{b}_t$ in (10).

Identification of firm effects. It is important to emphasize what we can, and cannot, identify from (10). While the availability of worker-level skill proxies provides a transparent way to estimate the distribution of firm returns, the distribution of workers' efficiency and firm returns could be jointly identified even in the absence of direct skill measures (Bonhomme et al., 2019; Lamadon et al., 2022). The latter approach would require the assumption that workers moving to a firm are not of similar quality as workers moving out of that firm. While such an assumption is easier to maintain when workers are ranked on a single index, it becomes less suitable in the presence of multiple skill dimensions where no unique ranking of workers is available. The general identification problem in these settings is that workers with different skill mixes may exhibit similar overall productivity. With direct proxies for different skills, returns λ_j^s for $s \in \{c,n\}$ can be identified upon a firm switch by the differential earning changes of workers with different skill levels. Thus the key requirement is that a sufficient number of such switches is observed. Identification of firm intercepts λ_j^0 is also obtained from earnings changes following firm switches.

Interpreting parameters. The level and dispersion of worker fixed effects μ_i partly reflect skill endowments. That is, μ_i includes the average skill return that a worker would get in any firm. Moreover, the empirical implementation of μ_i as a fixed effect flexibly accounts for residual dimensions of workers' skills that are priced homogeneously in the market.

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.[10]

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

The linear restrictions on firm effects imply that the lowest skill workers gain no employer premium above and beyond firm intercepts. Therefore, for the subset of workers with the lowest skill endowments $(c = 0, n = 0)$, equation (10) 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, (10) augments the double fixed-effect specification by explicitly allowing for heterogeneous returns to skills. By design, if we were to restrict attention to a single skill dummy S over a two year interval, with no other control variables, estimation of (10) would collapse back to the binary model in (1) 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 potential variation in average skill returns and significantly reduces computation times.[11] Conditional on the latter, worker fixed effects absorb time-invariant residual skill components, as discussed above.

Estimation. As discussed in Section 2.2, we report estimates for both the non-clustered leave-out samples and the clustered firm samples. When using the quadratic-form correction, the leave-out samples are defined so that each observation is a unique worker–firm match. That is, we collapse the data to the worker–firm level by averaging variables (earnings, age, time) within each spell a worker has at a given firm. This makes the estimator robust to serial correlation within clusters of observations and yields conservative variance estimates. To employ the group-level estimator, we use the k-means algorithm and partition firms into 100 clusters. The clustering is based on average cognitive and noncognitive scores of employees and on their earnings, consistent with the observation that different production arrangements lead to systematic variation in skill composition within firms. Results are robust to alternative clustering approaches.

[11] For example, estimation on 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 additional parameters. Allowing for time-varying returns to education 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.

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

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

Notes: The first two columns show standard deviations of parameters λ_j^0, λ_j^c, and λ_j^n estimated in (10). Column (1) quadratic-form corrects variances of the parameters estimated at the individual firm level and takes the square root. Column (2) assigns firms into 100 groups according to their average earnings and average c and n scores using the k-means algorithm. It then estimates (10) on this grouped data. Columns (3) and (4) multiply the estimated standard deviations with differences of skills between the 90th (c_i and n_i of 0.875) and 10th (c_i and n_i of 0.125) percentile. Estimation period: 1999–2008.

3.3 Estimates of Firm Parameters

Table 1 reports estimates of firm returns from specification (10) when skills are free to vary over their granular range (i.e., $c_i \in [0, 0.125, .., 1]$). While we initially focus on the dispersion of firm parameters, heterogeneity in skill returns has meaningful implications also for other moments of the earnings distribution through behavioural responses that result in assortative matching patterns. The latter effects are examined in the following sections.

Column (1) shows estimates for the leave-out (non-grouped) sample with bias correction. The first line, sd(λ_j^0) $= 0.22$, highlights that skill-independent premia vary significantly across employers, confirming the well-established relevance of such fixed effects. The estimates in the lines below document a less known layer of firm heterogeneity. In particular, they show that the standard deviations of skill returns are substantial, with sd(λ_j^0) $= 0.10$ for noncognitive skills and sd(λ_j^c) $= 0.15$ for cognitive ones.

Column (2) reports estimates based on the grouped-firms approach. As expected, standard deviations are lower since dispersion within each cluster is restricted to zero by treating all elements within it as a single representative employer. Nonetheless, while delivering a more conservative estimate of the absolute impact of skill returns heterogeneity, variation remains substantial. And, perhaps more interestingly, the relative magnitudes of returns are unchanged as the values of $\mathrm{sd}(\lambda_j^0)$, $\mathrm{sd}(\lambda_j^c)$ and $\mathrm{sd}(\lambda_j^n)$ all approximately halve. The finding of constant *relative* magnitudes is robust throughout the analysis, indicating 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. The hypothetical gains of such transitions are reported for high skill workers relative to low skill workers (90^{th} versus 10^{th} percentiles of skills). Based on bias-corrected firm-level sample estimates, moving to a firm that sits just a standard deviation higher in cognitive returns would result in a gain of 11 log points for a worker at the 90th cognitive percentile ($c_i = 0.875$) compared to a worker in the 10th percentile ($c_i = 0.125$). These are considerable differences in the gains from job mobility and, as discussed in Section 5, they are elicited through positive assortative matching.

Heterogeneity in noncognitive returns is somewhat lower but still 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 of n_i by seven log points. Jointly, a one-standard deviation change in both cognitive and noncognitive returns for workers at the 90th, rather than the 10th, percentile of each skill bring about an impact that is roughly as large as that of firm intercepts (see the cumulative effect in the last line of Table 1). The relative magnitude of the joint impact is similar for either of the estimation approaches.[12]

The finding of significant dispersion in skill returns is also robust in several respects. For example, Appendix D.1 shows that the bias correction approach in the leave-observation-out

[12] The joint estimated gains are 19 and 10 log points, respectively. By comparison, moving to a firm where λ_j^0 is one standard deviation larger raises a worker's earnings by 22 log points in the firm-level and 10 log points in the grouped estimates.

sample (rather than the leave-match-out sample) delivers even higher dispersion of skill returns. In sensitivity checks we also show that, when varying the number of firm clusters in the grouping estimator, the relative magnitude of skill returns and firm intercepts is unchanged.

The cross-section of skill returns. To characterize the cross-sectional distribution of firm returns we adopt as a baseline the estimates in column (2) of Table 1. Estimates based on the leave-out bias correction indicate even larger returns heterogeneity.

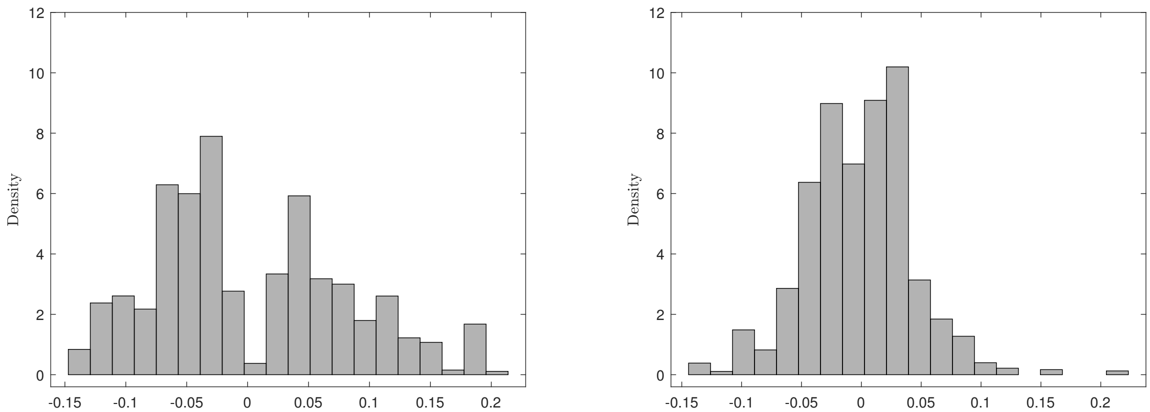
Figure 2 shows histograms of cognitive and noncognitive returns in the cross-section of firm clusters. As described earlier, the average λ_j^c and λ_j^n are normalized to zero. Return heterogeneity is significant in both dimensions although larger for cognitive traits, since $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.080$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.048$.

Dispersion is stable across time periods, with $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.095$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.052$ in 1990–1999 and $\mathrm{sd}(\lambda_j^{\mathrm{c}}) = 0.074$ and $\mathrm{sd}(\lambda_j^{\mathrm{n}}) = 0.048$ in 2008–2017 (see Appendix D.1). Edin et al. (2022) show that the *average* return to noncognitive skills increased while that of cognitive skills declined (see also Beaudry et al., 2016; Deming, 2017, for the U.S.). Our analysis suggests that, at the same time, the *heterogeneity* of skill returns across firms did not change differentially for cognitive and noncognitive skills.

The employment-weighted correlation of returns among firm clusters, $\mathrm{corr}(\lambda^{\mathrm{c}}, \lambda^{\mathrm{n}})$, is positive at 0.083.[13] Imperfect correlation lends support to the hypothesis that firm heterogeneity is genuinely multidimensional and that parameters can be independently identified through observable proxies that account for the skill-dependent ranking of workers.

Earnings gaps and skill premia. The plots in Figure 2 show that cognitive returns are concentrated between -15 and $+20$ log points. Relative to a worker from the 10th percentile of skills, a worker from the 90th percentile who moves from the bottom to the top of the returns distribution would gain 25 extra log points in earnings. That is, the difference in the cognitive premium between these workers is the skill difference $(0.875 - 0.125 = 0.75)$ multiplied by 35 log points. Complementarity of skills and returns implies that the earning function should be convex over skills because large earning effects accrue from matching high c workers to high

[13] Using the firm-level estimates of column (1) in Table 1, the bias-corrected correlation is 0.27.



(a) Cognitive, $\mathrm{sd}(\lambda^{\mathrm{c}}) = 0.080$

(b) Noncognitive, $\mathrm{sd}(\lambda^{\mathrm{n}}) = 0.048$

Figure 2: Histograms of Firm Returns (20 Bins)

Notes: Estimates of λ^{c} (left panel) and λ^{n} (right panel), based on 100 firm clusters weighted by employment. $\mathrm{corr}(\lambda^{\mathrm{c}}, \lambda^{\mathrm{n}}) = 0.083$. Grouped estimator for period: 1999–2008.

λ^{c} firms. Noncognitive returns can also add significantly to these earning differences. It follows that the impact of returns heterogeneity on the distribution of earnings hinges on the intensity of assortative matching and, in Section 4, we derive testable restrictions to gauge the prevalence of assortative matching in data. Then, in Section 5, we examine how firm heterogeneity, and the responses it elicits, shape the earnings distribution, and contrast our estimates to a counterfactual with random assignment of workers.

4 Matching

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

First, we introduce notation. Firms differ in three dimensions: their earnings intercept (λ_j^0) as well as cognitive (λ_j^{c}) and noncognitive (λ_j^{n}) returns. We define a matching function $\mu(\bar{c}_j, \bar{n}_j) = (\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})$, which maps a firm's average worker skills into its returns. One can show that, under the assumption of upward sloping firm-specific labor supplies (equation 4), the matching function μ

should be increasing in $\overline{c}_j$ and $\overline{n}_j$, and multidimensional PAM should hold as defined below. 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 properties of the matching function's derivatives. In matching problems 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.

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

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

Hence, to examine assortative matching we focus on the Jacobian of the matching function:

$$\frac{\mathrm{d}\mu(\overline{c}_j,\overline{n}_j)}{\mathrm{d}(\overline{c}_j,\overline{n}_j)} = \begin{bmatrix} \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{c}_j} & \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{c}_j} \\ \frac{\partial \lambda_j^{\mathrm{c}}}{\partial \overline{n}_j} & \frac{\partial \lambda_j^{\mathrm{n}}}{\partial \overline{n}_j} \end{bmatrix} \tag{11}$$

The Matching Jacobian in data. Intuitively, it does not matter for the empirical test of PAM whether firms choose workers or vice versa. This means that there are different ways to test the sorting hypothesis. We pursue two alternative routes, consistent with the previous analysis. First, we consider sorting regressions based on the Jacobian matrix defined above:

$$\begin{aligned} \lambda_j^{\mathrm{c}} &= d_{1\mathrm{c}} + d_{2\mathrm{c}}\overline{c}_j + d_{3\mathrm{c}}\overline{n}_j + e_j^{\mathrm{c}} \\ \lambda_j^{\mathrm{n}} &= d_{1\mathrm{n}} + d_{2\mathrm{n}}\overline{c}_j + d_{3\mathrm{n}}\overline{n}_j + e_j^{\mathrm{n}}. \end{aligned} \tag{12}$$

The linear forms in (12) are similar to the projections of fixed effect onto firm characteristics used in the applied literature (Kline et al., 2020). A strength of this specification is that, under general assumptions, the regression 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 d-parameters of (12). One can then use these linear projections to test for PAM in the cross-section of individual firms; this is true even if other statistics, such as the R^2, are potentially biased. One caveat is that, while point estimates from these regressions

Table 2: Projection of Individual Firms' Returns onto their Average Skills.

	Dependent Variables:					
	(1)		(2)		(3)	
	λ_j^{c}	λ_j^{n}	λ_j^{c}	λ_j^{n}	λ_j^{c}	λ_j^{n}
$\overline{c}_j$	0.29	-0.41	0.29	–0.41	0.16	–0.44
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.04)
$\overline{n}_j$	0.15	0.61	0.15	0.61	0.40	0.56
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
# firms	19,085		19,085		19,085	
Controls	No		# employees		No	
Weights	No		No		# employees	

Notes: The table reports sorting coefficients d_2 and d_3 from estimating (12) with individual firm λ_j^{c} and λ_j^{n}. Projections of individual coefficients in estimation period 1999–2008. Standard errors are corrected to account for the first-stage estimates of the outcome variable as in Kline et al. (2020, Section 4).

are generally unbiased, standard errors must be corrected for the correlation across the first-stage estimates of the outcome variable (firm parameters).[14]

Table 2 reports estimates from projections in (12), obtained from non-grouped firm-level data (employees' cognitive and noncognitive skills are averaged into firm-specific $\overline{c}_j$ and $\overline{n}_j$). It is apparent that PAM cannot be rejected since own-partial derivatives and the determinant of the Jacobian are positive throughout. The coefficients on $\overline{c}_j$ for λ_j^{c} are only about half as large as on $\overline{n}_j$ for λ_j^{n}. Flipping this around, $\overline{c}_j$ responds more to a given difference in returns, which implies stronger sorting on cognitive traits. Below we present additional evidence of uneven sorting patterns.

An alternate test of skill sorting. A different route to test PAM is based on a matching Jacobian where average skills $\overline{c}_j$ and $\overline{n}_j$ are projected onto firm returns. This builds on a definition of the matching function that maps firm gradients into worker characteristics, and provides a way to examine matching patterns where skill sorting in each dimension depends on both of the

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

employer's returns. In practice, we posit $\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = (\overline{c}_j, \overline{n}_j)$ and test Jacobian conditions[15] using the projections in (13). It is important to recognize that, if the λ parameters are measured with error due to limited mobility, estimation of (13) may deliver biased point estimates. To mitigate such concerns, we adopt a grouped-firm approach and project average skills (cognitive or noncognitive) onto the 100 cluster-specific returns. The grouping does not hardwire the relationships in (13) since cluster-level returns are free to vary. Table 3 reports estimates of the Jacobian for the following specifications:

$$\begin{aligned} \overline{c}_j &= \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{c}} \\ \overline{n}_j &= \delta_{1\mathrm{n}} + \delta_{2\mathrm{n}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{n}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{n}}. \end{aligned} \tag{13}$$

The regressions in (13) deliver the best linear approximation to the conditional expectations of $\overline{c}_j$ and $\overline{n}_j$. For instance, $\mathrm{E}(\overline{c}_j|\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}}$, so that the parameter $\delta_{2\mathrm{c}}$ is the expected value of the top-left element $\left(\frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{c}}}\right)$ 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 (13).

The positive and highly significant $\delta_{2\mathrm{c}}$ and $\delta_{3\mathrm{n}}$ in Table 3 imply that the own-derivative conditions for PAM are satisfied for both c and n. The Jacobian is also positive definite, with determinant $\delta_{2\mathrm{c}}\delta_{3\mathrm{n}} - \delta_{3\mathrm{c}}\delta_{2\mathrm{n}}$ larger than zero. This lends additional support to the hypothesis that PAM holds over the 1999–2008 period in our large sample of firms.

The positive $\delta_{2\mathrm{n}}$ in equation (13) indicates substantial cross-sorting of high n workers to high λ_j^{c} firms, which occurs when skill endowments are correlated. High c workers who sort into high cognitive return firms also have a higher endowment of n skills. Consistent with this observation, own-sorting in the c dimension is strong, as shown by the large $\delta_{2\mathrm{c}}$ estimates and in Figure 3 below. Cross-sorting of c to high λ_j^{n} firms is not present and sorting in the n dimension is substantially weaker (see also Figure 3 below).[16]

Results do not change after controlling for firm-specific employment size and firm intercepts λ^0, as shown in column (2). Neither do they change when weighting by employment size as

[15]The Jacobian becomes $\frac{\mathrm{d}\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})}{\mathrm{d}(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})} = \begin{bmatrix} \frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{n}}} \\ \frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{c}}} & \frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{n}}} \end{bmatrix}$.

[16]Correlation of λ_j^{c} and λ_j^{n} would affect cross-sorting estimates if we had not controlled for each respective other indirect return in (13).

Table 3: Projection of Average Skills onto Grouped Returns.

	Dependent Variables:					
	(1)		(2)		(3)	
	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$
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	

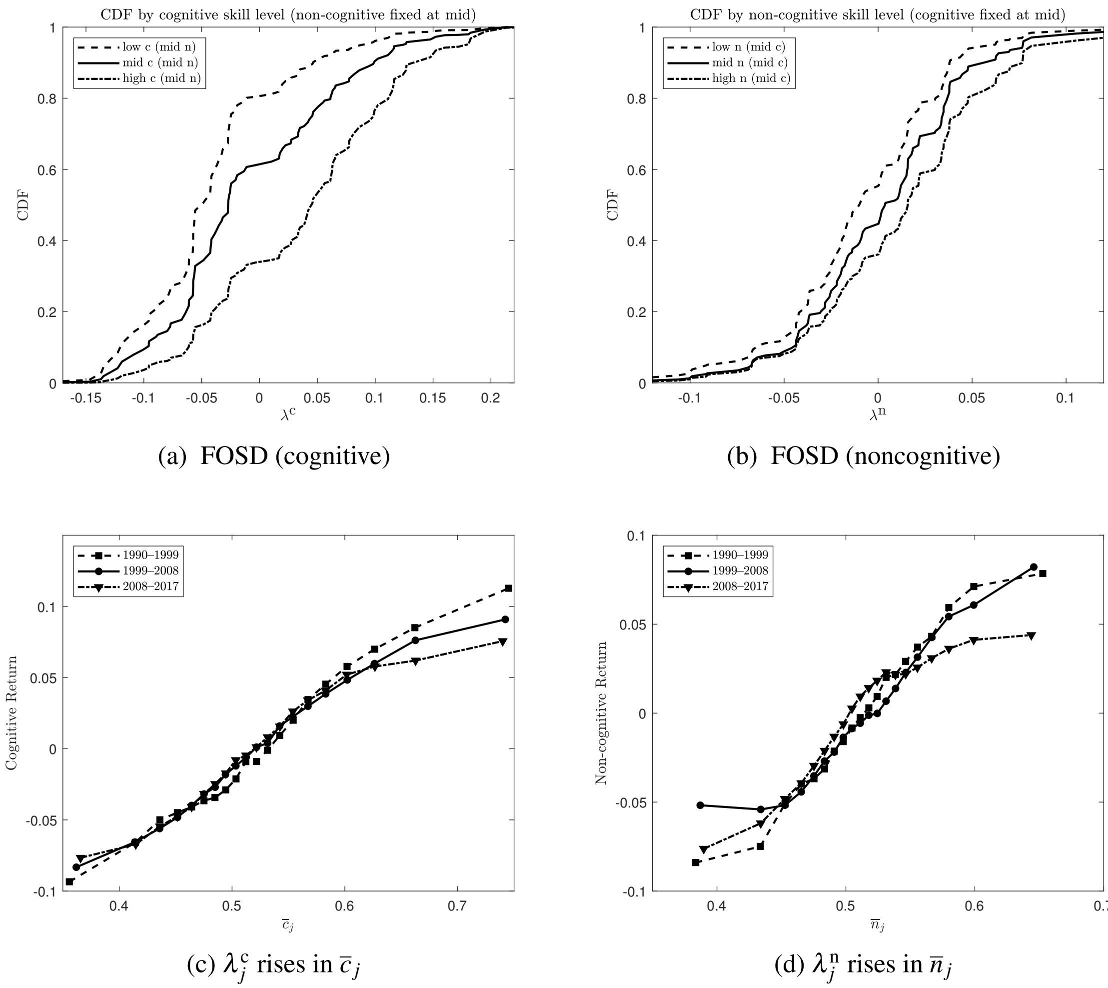
Notes: Column (1) reports sorting coefficients δ_2 and δ_3 from estimating (13). The specification in column (2) additionally controls for intercepts λ^0 and for the firms' total employment headcounts. Column (3) weights the observations by the firm's number of employees. Each firm is one observation. Robust standard errors clustered at the level of the 100 firm groups are in parentheses. Grouped estimator for period 1999–2008.

shown in column (3). Between 54 and 68 percent of the skill variation between firm clusters is accounted for by differences in estimated λ^{c} and λ^{n} returns alone.[17] When we weigh firms by their employment size and control for λ^0, the explained variation rises to 65–75 percent.

4.2 The Distribution of Workers over Returns

If workers with a high endowment of a particular skill match more frequently with firms with high returns to that skill (in the sense of first-order stochastic dominance or FOSD), then sorting is positive along that dimension (Lindenlaub and Postel-Vinay, 2020). A way to visualize such patterns is to compare the cumulative distribution functions (CDF) of returns for separate sets of workers (say, high versus low cognitive skills).

[17] A feature of the grouping approach is that the R^2 is essentially between firm classes, since average skills vary little within k-means clusters. Indeed, averaging $\overline{c}_j$ and $\overline{n}_j$ within groups and running the regression for group averages (i.e., 100 observations) gives only a tiny increase in R^2. Nonetheless, it is remarkable that returns can explain so much of the cross-clusters skill variation.



(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 3: Distributions of firm returns for different sets of worker skills.

Notes: Results from the grouped estimator. 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 3 illustrates 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.[18]

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, and 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.

In the bottom panels of Figure 3 we use an alternative way to visualize the distribution of skills over returns by plotting skill returns over within-firm average skills. These measures of central tendency confirm that returns increase monotonically with skill endowments, consistent with PAM. Between-firm differences in average skills are larger in the cognitive dimension, as expected given the higher dispersion of λ^{c} relative to λ^{n} and the stronger sorting incentives. Similar patterns hold for different estimation periods, suggesting that workers consistently sort across firms based on their attributes.

5 Complementarities and Earnings

Thought experiments where workers are parachuted into firms with higher returns, like the ones in Section 2.3, are not wholly informative about the actual impact of complementarities on the earnings distribution due to the non-random nature of firm assignment. In what follows, we cast earnings differences due to firm heterogeneity in terms of deviations from cross-sectional means that explicitly account for assortative matching.

[18] To ease exposition, 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 through the clustering approach. This is graphically convenient as it restricts variation on the x-axis. Estimates based on leave-out-samples without clustering deliver similar insights. Additional FOSD plots are in Appendix D.3.

5.1 Effects on the Distribution of Earnings

Equation (10) emphasizes that, after controlling for confounding effects, the return for worker i with skill bundle $s_i = (c_i, n_i)$ working in firm j can be represented as:

$$\log(w_j(s_i)) = \underbrace{\mu_i}_{\substack{\text{(a)} \\ \text{Person effect} \\ \text{(incl. Mincer returns)}}} + \underbrace{\lambda_j^0}_{\substack{\text{(b)} \\ \text{Firm intercept}}} + \underbrace{\lambda_j^{\mathrm{c}}\bar{c} + \lambda_j^{\mathrm{n}}\bar{n}}_{\substack{\text{(c)} \\ \text{Firm returns effect}}} + \underbrace{\lambda_j^{\mathrm{c}}\widetilde{c}_i + \lambda_j^{\mathrm{n}}\widetilde{n}_i}_{\substack{\text{(d)} \\ \text{Match effect}}} \;, \qquad (14)$$

where $\widetilde{x}_i$ denotes the deviation of skill x_i from its cross-sectional average $\bar{x}$.

Equation (14) has an intuitive interpretation: the term (a) contains the homogeneous Mincerian return $\kappa_{\mathrm{c}} c_i + \kappa_{\mathrm{n}} n_i$, which is often estimated in survey data when skill measures are available (the κ loadings are defined in Section 3); component (b) is a firm fixed effect that captures constant differences above and beyond the Mincerian return in (a). The elements (c) and (d) reflect, respectively, the direct impact of firm return hetorogeneity 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 $\lambda_j^{\mathrm{c}} c_i + \lambda_j^{\mathrm{n}} n_i$ and jointly subsume firm returns that vary with worker skills. The expected value of the (c) term in (14) is nil because $\mathrm{E}(\lambda_j^{\mathrm{c}}) = \mathrm{E}(\lambda_j^{\mathrm{n}}) = 0$. In contrast, the expected value of component (d) can be different from zero as it reflects the per capita wage gains due to assortative matching of workers to firms.

We note that, if skill measures were not available, components (b) and (c) would get conflated into the firm fixed effect, while the skill dependent variation would be absorbed within the person fixed effect μ_i. Separate identification of the impacts of heterogeneous returns and match-quality in summands (c) and (d) can be obtained only when proxies of skill endowments are available.

Lastly, through variance decompositions (Appendix D.2) it is possible to show that, if firm heterogeneity is restricted to fixed effects, a share of the earnings variance due to heterogeneous skill returns is improperly attributed to employer intercepts as if they were independent of skills.

Components of permanent heterogeneity. The impact of the components of equation (14) on earnings dispersion is summarized in Panel A of Table 4 where we present estimates for both the clustering approach and the firm-level estimation with bias correction.

Table 4: Contributions of Firm Heterogeneity to Dispersion and Levels of Earnings

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

Notes: Panel (A) shows the dispersion of each summand in equation (14), i.e., the standard deviations of person and firm intercepts, and the standard deviations of the products of returns and skills. Panel (B) shows the averages of the last two summands in equation (14), i.e., the contribution of matching to average earnings in the economy (through complementarity gains). Firm-level estimates in column (3) 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). Estimation period: 1999–2008.

In line with other studies, unobserved worker heterogeneity has a strong impact on earnings through fixed effects μ_i. The latter include the average returns to skills. The contribution of the heterogeneous components of skill returns to earnings dispersion is between 55% and 60% of that of firm fixed effects.

Returns heterogeneity and sorting lead, on average, to higher earnings. The latter gains can be measured through the covariance of skills and firm returns. For example, if we consider cognitive skills, we have that $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i) = \mathrm{cov}(\lambda_j^c, \bar{c}_j)$. This equivalence confirms that sorting determines the intensity of the average gain accruing from returns' heterogeneity.[19] Panel (B) of Table 4 shows estimates of the average gain from match effects, which are between 0.8 and 0.9 log points. The larger gains from heterogeneity in cognitive returns reflect the stronger sorting in that dimension, as documented in Section 4.

[19]The sorting parameters estimated in equation (12) are, in essence, just this gain standardized by the underlying variance of average skills across firms, $\frac{\mathrm{cov}(\lambda_j^c, \bar{c}_j)}{\mathrm{var}(\bar{c}_j)}$. The equality $\mathrm{E}(\lambda_j^c c_i) = \mathrm{cov}(\lambda_j^c, c_i)$ follows from the fact that excess skill returns have zero mean, that is $\mathrm{E}(\lambda_j^c) = 0$.

5.2 The Uneven Gains from Sorting

The gains from sorting are unevenly distributed and non-monotonic. They are positive and large for high skill workers, absent for the least skilled workers and negative for a wide range of intermediate skills. These patterns can be illustrated by taking expectations of equation (14) *after conditioning* on a given skill level.

For brevity, we discuss gains from cognitive skills but similar arguments hold for noncognitives. Given cognitive value c_i, the full earnings gain from sorting is

$$\underbrace{c_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Full sorting gain}} = \underbrace{\overline{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Firm returns effect}} + \underbrace{\widetilde{c}_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)}_{\text{Match effect}}, \tag{15}$$

where c_i is split into average $\overline{c}$ and deviation $\widetilde{c}_i$. Since the distribution of returns faced by each individual depends on their skill level, the expected return from firm heterogeneity changes non-linearly with skills. Estimates based on the clustering approach (column 1 of Table 5) illustrate that the marginal expected return $\mathrm{E}(\lambda_j^c \mid c_i)$ is increasing in c_i and thus deviates from the unconditional average, which is normalized to zero. The difference in expected marginal returns between top and bottom cognitive skill workers is almost 12 log points $(6.74 - (-5.13) = 11.87)$.

Marginal returns conditional on skills. The empirical distribution of gains is summarized in column (2) of Table 5. Top cognitive workers vastly benefit from higher conditional returns, which lead to earnings 7 log points higher than if they were matched with the average firm. To illustrate how much this return matters for skill premia, it is useful to consider a simple example where we compare the sorting gains gap between a top worker $(c_i = 1)$ and a low-middle (level 4 in Table 5, $c_i = 0.375$), which is 8 log points. The raw earnings difference between these two workers is on average 30 log points in our sample; this gap is reduced to $(30 - 8) = 22$ log points when sorting effects are taken out. Thus, sorting adds more than 1/3 ($\frac{8}{22}$) to the baseline gap and significantly amplifies between-skill earning differences.

Non-monotonicity of gains. Column (2) of Table 5 shows that gains are not monotonic in c_i. In particular, workers with low-to-middle skills lose out compared to a hypothetical situation where

Table 5: 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.

everyone is matched with the average return. To understand these losses, and why they wane as c_i approaches zero, equation (15) breaks down skill returns into a "return effect" $\bar{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ and a "match effect" $\tilde{c}_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)$.

Estimates of the return effect $\bar{c} \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ are shown in column (3) of Table 5 and 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. Hence the return effects measure the impact of firm heterogeneity net of any complementarity gains. Since high skill workers sort into high return firms, and low skill workers into low 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) in Table 5.

In contrast, the match effects $\widetilde{c}_i \cdot \mathrm{E}(\lambda_j^c \mid c_i)$ in column (4) raise earnings in the aggregate by eliciting incremental gains from worker-firm complementarity.[20] Unsurprisingly, match effects are large at the higher end of the skill distribution, where earnings are magnified compared to a random allocation (3.1 log points match effect for $c_i = 1$; 1.65 for $c_i = 0.875$).

Large match effects are also detected among low skill workers (e.g., 2.7 for $c_i = 0$; 1.9 for $c_i = 0.125$) since, as shown in (14), match effects are defined as deviations from the average-worker gains. That is, match quality effects measure returns in excess of those experienced by an average skill worker with the same employer. The result then follows from the observation that average skill workers experience a steeper loss from being matched with a low quality firm due to the higher opportunity cost of the mismatch.

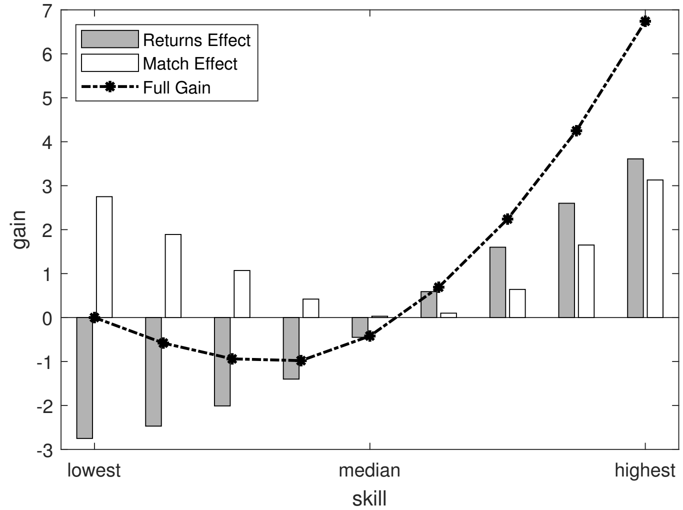
Firm-specific intercepts and gains from sorting. The last column of Table 5 shows the wage gains from matching with alternative intercepts λ_j^0, conditional on skill types c_i. These gains have a zero sum due to the lack of complementarity between skills and firm intercepts. Nonetheless, the differential assignment of workers across firms (and, hence, across λ_j^0) raises earning differences by an extent comparable to that due to sorting on returns in column (2). This reinforces overall inequality between skill levels as more able workers also tend to populate higher intercept firms. The purely redistributive nature of this effect induces, however, little or no additional convexification in the ability-wage space.[21]

A graphical representation. Figure 4 offers a concise summary of the distribution of sorting gains and their components. Workers with the lowest skills exhibit positive match quality effects because they do not lose like the average worker from being matched to a low return employer. By the same token, the gains turn negative for low to intermediate skill workers, who would benefit from matching with high return firms but are not assigned to such firms. These individuals would

[20]Both components are defined as surplus relative to the baseline in which firm heterogeneity is absent and all returns are equal to the population average. Hence, both positive and negative gains must be interpreted relative to a scenario where each worker is given the average return or, equivalently, where workers are randomly allocated across firms.

[21]Convexity of earnings is only due to skill complementarities. Further evidence of this point comes from the observation that, after conditioning on noncognitive skills n_i, the sorting across λ_j^0 results in modest effects that offset rather than reinforce the skewness of earnings across noncognitive endowments (see Appendix Figure D.3).

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



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 with detailed numbers in Table 5. Sample period: 1999–2008.

be better off in a world with no firm heterogeneity in skill returns. For workers with above average skills, both the return effects and the match effects are positive, which results in large gains at the top. Complementarities are key to deliver an earnings schedule that is convex in skills.

Estimated gains outweigh losses and matching raises aggregate earnings. A simple way to assess the intensity of matching in the data is to benchmark it against the maximum gain it could generate, given the estimated return and skill dispersion. Adopting this metric, assortative matching in the cognitive dimension generates 0.75 log points (Panel B, Table 4) as compared to a hypothetical maximum of 1.9 log points.[22] This simple calculation lends support to the view that the observed allocation of skills across employers, while imperfect, does deliver some of the gains associated to efficient matching. Gains from matching along noncognitive returns are smaller, yet they boost the aggregate match quality gain to 0.88 log points. All the estimates of the gains from sorting are robust to alternative normalizations of skills and returns (see Appendix D.4).

[22]Match effects are maximized when the correlation $\mathrm{corr}(\lambda_j^c, c_i) = \frac{\mathrm{cov}(\lambda_j^c, c_i)}{\mathrm{sd}(\lambda_j^c)\mathrm{sd}(c_i)} = 1$. Our 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 6: Moments due to skill returns under random versus actual sorting.

	Mean $\times 100$		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 distribution of skill returns assuming either the actual allocation or a counterfactual where workers are randomly assigned to firms. Mean earnings $\mu \equiv \mathrm{E}(\lambda_j^c c_i + \lambda_j^n n_i)$ and subcomponents rise due to matching in column (2) compared to (1). Dispersion $\sigma \equiv \mathrm{sd}(\lambda_j^c c_i + \lambda_j^n n_i)$ rises modestly in column (4) compared to random assignment (3). Skewness $\tilde{\mu}_3 \equiv \mathrm{E}[(\lambda_j^c c_i + \lambda_j^n n_i - \mu)/\sigma]^3$ in the actual is almost twice as large relative to random assignment (last two columns). Estimates based on the grouping approach. Sample period: 1999–2008.

Random assignment of returns. It is informative to compare the distribution of estimated skill returns to the one obtained under random assignment of workers to firms. We construct the random assignment counterfactual by sample-weighting all skill types within a firm according to their population share and we are careful to preserve the empirical firm size distribution. Table 6 illustrates the findings from this exercise by juxtaposing the first three moments of the empirical distribution to those obtained under random assignment.

Columns (1) and (2) report first moments. This reproduces the aggregate gains reported before, i.e., average log earnings effects are the same when randomly allocating workers or assigning them to the average firm. Columns (3) and (4) show that the standard deviations of skills premia are only marginally different: this is not surprising if one considers that higher between-skill inequality in the non-random allocation, seen in Figure 4, 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 an important and subtle distinction highlighted in the theoretical literature (Becker and Chiswick, 1966; Sattinger, 1993; Lindenlaub, 2017), which emphasizes how the most conspicuous changes induced by production complementarities may occur in the third moment of the earnings distribution. Columns (5) and (6) in Table 6 suggest this is indeed the case in our worker-firm sample, where the skewness of log earnings is twice as large under the non-random assignment of workers to firms.

More generally, high skill workers are not often observed in low return firms while differences in skill returns have little effect on low endowment workers. The latter observation translates into a fairly concentrated left tail of the earnings distribution when compared to random assignment. In contrast, the nonlinear effects from matching high skill individuals to high return firms result in a substantial thickening of the right tail of the earnings distribution, as shown in Figure 4.

To sum up, heterogeneity in skill returns 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.[23] Since the distribution of firm sizes is unchanged in our counterfactuals, sorting has no effect on the moments of firm intercepts λ_j^0, which are the same under the actual and random allocations.

6 Extensions and Robustness

Firm heterogeneity in skill returns encourages sorting and affects the earnings distribution. One may, however, question to what extent the assignment of workers to jobs occurs along the industry and occupation dimensions. This motivates a robustness exercise where we explicitly test for return heterogeneity within narrowly defined industry and occupation groups.

In addition, and to aid interpretation of our baseline findings, we examine the correlation of skill returns with a subset of firm-level measurements. This is facilitated by external data about firms' balance sheets, capital composition and innovation activities that can be linked to our sample of employers. The latter measures convey information about the nature of production arrangements that may underpin firm differences in skill returns.[24]

[23] Bonhomme et al. (2019) present an analogous counterfactual where workers are randomly allocated to firms. Our estimated gains from skill complementarities are of similar magnitude when compared to their match effects between unobserved worker and firm types. We find larger effects in the aggregate (almost 1% of earnings vs 0.5%). Part of the difference is accounted for by the more pronounced earnings convexification in our estimates, which disproportionately benefits workers with higher skill endowments.

[24] We focus on clustered firm returns (100 k-means groups) for brevity. Results for the leave-out firm-level samples are consistent with what we emphasize in this section (Appendix E).

Finally, we examine the robustness of estimates under the clustering approach to alternative choices about the number of firm classes and of variables used for grouping firms.

6.1 Industry and Occupation Specific Skill Returns

We begin by assessing whether skill returns simply reflect sector and job characteristics. We do so by adding to the baseline specification (10) a full set of industry $\times$ occupation interactions with cognitive and noncognitive skills. Detailed estimates are reported in Appendix E.1. We find that fine industry and occupation-specific skill returns, or returns that vary by industry $\times$ occupation group, account for a minor share of firm-specific heterogeneity. Sorting of both cognitive and noncognitive skills across returns remains strong. Results confirm that significant skill returns heterogeneity occurs at the firm level (as opposed to the more aggregate industry or occupation level). This remains true after conditioning on rather fine occupation measures.

Aggregating to industry or occupation. While most of the heterogeneity occurs at the firm level, some industries or occupations may still exhibit higher skill returns on average. In Appendix E.1 we explore this possibility by projecting the baseline λ_j^c and λ_j^n estimates on a broad set of industry-sector indicators and employment shares by occupation group. These projections are similar to those used to test the PAM hypothesis through the equations in (12) and one can show that they deliver generally unbiased point estimates. We find that high cognitive returns are frequent in the business services and IT sector 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, technical and services/sales jobs. These results hold in the firm-level leave-out samples and in the clustered samples.

6.2 Capital Composition, Innovation, and Skill Returns

Next, we link balance sheet and innovation data to the sample of employers. This lends some insight into potential sources of skill return heterogeneity.

Balance sheets and capital composition. Since differences in capital composition reflect systematic aspects of productive and organizational structure, we use balance sheet data to measure tangible and intangible capital per employee (as well as finer components). An advantage of the Swedish institutional setting is that a majority of private sector firms are limited liability corporations with publicly available financial statements. We thus aggregate the workplaces at the organization level where this information is reported. We refer to these aggregates as "firms" from now.

Table 7 reports estimates for projections of skill returns onto firms' tangible and intangible capital components. We focus on cognitive returns from the group-level estimates. Results for noncognitives are in Appendix E.2. To account for zero-value observations for finer capital items in the balance sheets, we use the inverse hyperbolic sine (arcsinh) transform.[25]

Capital composition is strongly associated to cognitive returns. Column (1) of Table 7 shows that tangible assets vary negatively with skill returns but intangible assets exhibit a strong positive correlation. Column (4) illustrates that the negative relationship holds strong for physical capital (buildings, land, and machinery) and the positive relationship is especially intense for intellectual capital (patents, licences, and capitalized *R&D* expenses). 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. Relatively high physical assets and machinery, on the other hand, are more frequent in firms that exhibit lower returns to cognitive skills.

These relationships are robust in several respects: they hold within industry sectors of the economy (columns (2) and (4) of Table 7) and if we weight with firm employment size (columns (3) and (6)). Appendix E.2 shows that they hold in the leave-out firm-level samples as well as when using dummy indicators (or logs) instead of the arcsinh transformation. Perhaps unsurprisingly if one considers production arrangements, firms that employ intangible and intellectual assets

[25] The arcsinh approximates $\log(2x_j) = \log(2) + \log(x_j)$. Estimates are interpreted as semi-elasticities (unit changes) for very small values of the transformed variable x_j, and as elasticities for larger values. See Bellemare and Wichman (2020) and note to Table 7. Findings are robust to alternative approaches; Appendix E.2 shows that similar results hold at the intensive margin (log transform of capital items) and at the extensive margin (firms with high cognitive returns are more likely to report nonzero intangible assets).

Table 7: Projection of Group Returns onto Firm Capital Composition.

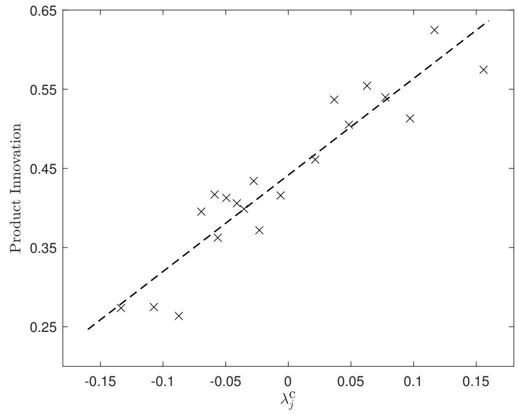
	Dependent variable: $\lambda_j^c \times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-0.83	-0.30	-0.53			
	(0.12)	(0.06)	(0.15)			
Buildings, Land, Machinery				-0.92	-0.35	-0.59
				(0.13)	(0.07)	(0.15)
Other tangible assets				0.13	0.05	0.04
				(0.07)	(0.05)	(0.14)
Intangible assets	1.02	0.63	0.85			
	(0.11)	(0.06)	(0.11)			
Patents, licences, capt. R&D				1.17	0.70	0.72
				(0.12)	(0.07)	(0.14)
Goodwill and other intangibles				0.52	0.36	0.52
				(0.09)	(0.06)	(0.09)
R-squared	0.10	0.33	0.08	0.11	0.33	0.09
Number of firms	14,339	14,339	14,339	14,339	14,339	14,339
Sector fixed effects	No	Yes	No	No	Yes	No
Employment weighted	No	No	Yes	No	No	Yes

Notes: Projections of cognitive skill returns onto capital components per employee, using firms' balance sheets. Tangible fixed assets comprise of buildings and land; machinery and equipment; and other. Intangible fixed assets include capitalized expenditure on research and development; patents, licenses, and concessions; goodwill; and other. All variables are transformed using inverse hyperbolic sine, i.e., $\text{arcsinh}(x_j) = \log\left(x_j + \sqrt{x_j^2 + 1}\right)$. The dependent variable λ_j^c is multiplied by 100. Estimates are based on the sample of clustered firms; period 1999–2008. Robust standard errors clustered at the level of each of the 100 firm groups.

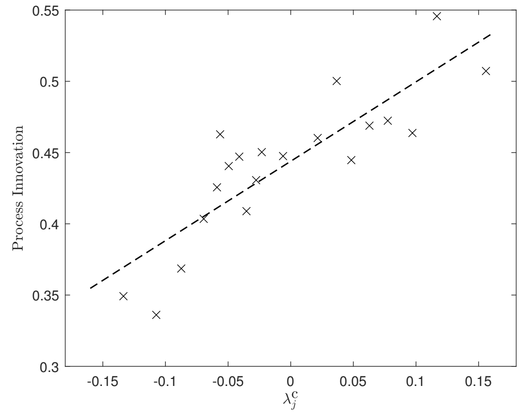
have substantially higher cognitive skill returns.[26] As we show in Appendix Table E.2, results for noncognitive skills are less pronounced and returns are modestly higher in firms with more physical capital. This lends support to the notion that skills should be modeled separately rather than collapsed into a single index.

Measures of innovation activities. Next, we use responses in the Swedish version of the European Community Innovation Survey (CIS) to study the relationship between skill returns and

[26]Even controlling for capital composition in equation (10), or allowing for interactions between capital and skill in parallel to occupation-specific skill returns in Section 6.1, has little impact on the heterogeneity of firm-specific skill returns that we uncover.



(a) λ_j^c and product innovation



(b) λ_j^c and process innovation

Figure 5: Cognitive skill returns and firm innovation.

Notes: The figure plots a binscatter of firms' innovation activities against cognitive skill returns (group-level estimates during 1999–2008). Innovation activities are measured as indicators whether a firm has conducted any product (including service, Panel a) or process (including organizational, Panel b) innovations. This information is from various waves of a representative firm survey (European Community Innovation Survey, CIS). We average the responses (i.e., indicators) for the waves 1998–2000, 2002–2004, 2004–2006, 2006–2008, 2008–2010 relevant to our estimation period. Underlying the plots are 4,138 unique firms. Regression slopes, controlling for a quadratic in firm employment, are $\beta = 1.21$ (clustered S.E. $= 0.13$) and $\beta = 0.55$ (clustered S.E. $= 0.10$) for product and process innovations, respectively.

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. (2021) show that the CIS provides direct, reliable, and broad measures for different types of firm-level technological change.

After linking the CIS survey responses to the administrative sample of employers, in Figure 5 we plot bin scatters of dummies (taking value one in the presence of product/process innovations in the firm) versus cognitive skill returns.[27] Firm innovation activities are positively, and almost linearly, associated with estimates of 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 about 25 to 65 percent. For process innovations

[27]We plot the raw relationship after controlling for (a quadratic in) employment, since the probability of engaging in any innovation rises with a firm's size. The controls do not substantively affect results. The corresponding relationships for noncognitives are weaker and reported in Appendix E.2.

the relationship is fainter and only borderline significant when we also condition on product innovation (Table E.4). However, innovation activities still differ by twenty percentage points between firms with the lowest and the highest skill returns.

In Appendix E.2 we show how results are qualitatively robust to alternative firm-level estimation approaches or when controlling for industry fixed effects. Moreover, innovation expenditures in the CIS survey are also larger for higher λ_j^c firms (especially in-house research and development), suggesting that high cognitive return firms differ in their ability to bring forward innovations. This evidence lends support to existing studies linking cognitive skills to worker level innovation activities (Aghion et al., 2017; Bell et al., 2018).

6.3 Changing the Cluster Design

When using estimators based on firm clusters, one question is whether results are sensitive to the grouping strategy. Next, we examine differences in the estimated contribution of firm heterogeneity to earnings dispersion under alternative assumptions about the number of firm classes and about the observables used to classify them.

Using only ten firm classes (Bonhomme et al., 2019; Lamadon et al., 2022) marginally lowers the absolute contribution of firm heterogeneity while raising the importance of skill returns relative to the intercepts. However, results remain similar to the baseline. After adding additional observables to the clustering criterion (namely, firm employment and the standard deviations of both earnings and skills within the firm), estimates of firm effects are comparable and in line with the benchmark.[28] If we discard information about worker skills and only use data about within firm earnings to define firm classes (see Bonhomme et al., 2019; Lamadon et al., 2022), we also find that estimates of skill returns' contribution do not change significantly relative to the case where many observables are used. Estimates from these exercises are in Appendix Table E.5.

Changing the number of clusters. The baseline grouping approach, with one hundred clusters, delivers conservative estimates of the contribution of firm heterogeneity to earnings dispersion. To illustrate how restrictions on the number of clusters affect estimates of firm effects,

[28]This check is performed using 100 firm classes. We also experiment with including value added, which is however reported at the organization level, and find similar results.

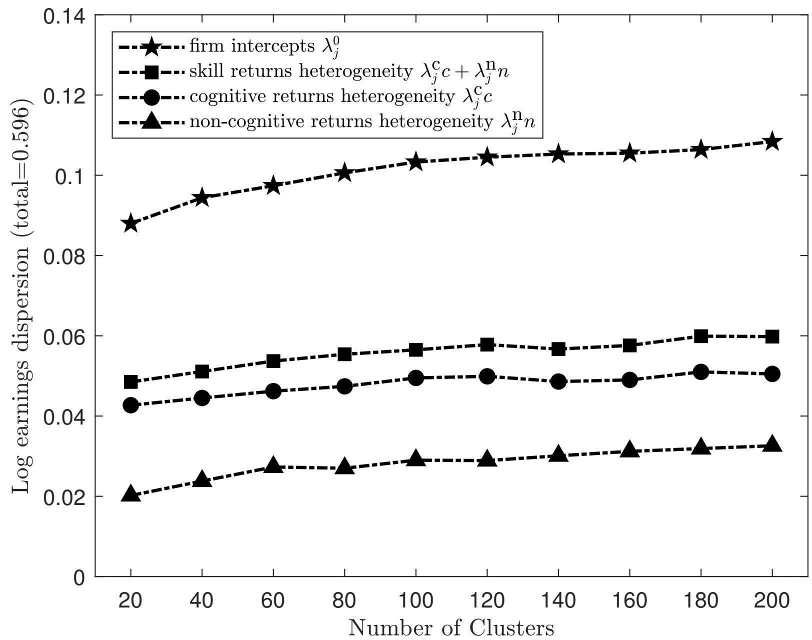


Figure 6: Dispersion due to firm heterogeneity (log earnings), by number of k-means groups.

Notes: The figure shows the earnings variation due to firm intercepts sd(λ^0), cognitive skill returns sd($\lambda_j^c c_i$), noncognitive skill returns sd($\lambda_j^n n_i$), and overall skill returns sd($\lambda_j^c c_i + \lambda_j^n n_i$) when we re-estimate the model with different numbers of k-means clusters. Estimation period: 1999–2008.

Figure 6 shows the standard deviation of log earnings that is 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 higher estimates of the impact of firm heterogeneity on overall inequality. As one might expect, the absolute values of firm effects estimated from finer clusters become similar to those obtained using the quadratic-form correction with no firm grouping. More notable, and perhaps less expected, is that 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 effectively unchanged (see Table 4). This is consistent with findings in Table 1 and indicates that the two empirical approaches deliver comparable estimates of the relative contribution of skill return heterogeneity to firm-level variation.

7 Conclusion

By examining distinct dimensions of firm heterogeneity, we estimate the extent of skill return variation across employers and present direct evidence of worker–firm complementarities. The analysis relies on alternative empirical approaches in administrative employer–employee population records that we link to high-quality information about the cognitive and noncognitive attributes of workers. Each approach imposes different restrictions to mitigate biases and carry out the large computational exercises. Nonetheless, estimates of the relative impact of skill returns, as opposed to conventional measures of firm heterogeneity based on fixed effects, are stable irrespective of the implementation.

Our key findings can be summarized as follows: (1) Similar skills command substantially different returns across firms. These differences occur along both the cognitive and non-cognitive skill dimension, but firm-level returns to each attribute only weakly correlate with one another. (2) Returns heterogeneity generates incentives for sorting and we document that, 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 unevenly distributed and non-monotonic in worker skills. High skill workers benefit from heterogeneity in returns while the least productive workers experience little loss from low-return employers. Considerable costs are borne by workers with intermediate skills who have a nontrivial opportunity cost of being matched with low quality firms.

More generally, we find evidence of economically meaningful complementarities between workers and firms, and of positive assortative matching in multiple skill dimensions. Sorting has material implications for earnings. In particular, the earnings distribution becomes more skewed due to the matching of high skills to high returns. By the same token, allocative efficiency improves and the average economy-wide skill premium rises.

A central lesson from studies of employer effects over the past two decades is that firm-level productivity and rent-sharing are key determinants of wage growth and inequality (see Card et al., 2018; Lamadon et al., 2022, for a summary). This has spawned interest in the role of market factors, institutions, and policies that may underpin such dependence (e.g., De Loecker and Eeckhout, 2021; Jäger et al., 2021; Dustmann et al., 2022). One implication of our findings is that labor market returns to skill attributes are nuanced and not solely driven by the supply of skills or by aggregate changes in demand (e.g., general skill biases in technical progress). Rather, the composition and evolving demands of firms play an important mediating role for the economy-wide returns to skills.

Using information from firms' balance sheets, we present ancillary evidence that firms with high cognitive returns engage in more innovation and hold more intellectual capital. This points to the presence of underlying complementarities in production and raises the possibility that mismatch between skills and firms may hamper the gains from innovation and productivity advances (Aghion et al., 2017; Bell et al., 2018). Evidence about the covariation of skill returns with intellectual capital and innovation suggests that a fruitful direction for further research may be to explicitly examine the nature and determinants of the extensive heterogeneity in skill returns.

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 (2017): "Living the American Dream in Finland: The Social Mobility of Inventors," Mimeo.

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.

Bagger, J. and R. Lentz (2019): "An empirical model of wage dispersion with sorting," *The Review of Economic Studies*, 86, 153–190.

Bartolucci, C., F. Devicienti, and I. Monzón (2018): "Identifying sorting in practice," *American Economic Journal: Applied Economics*, 10, 408–38.

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.

Becker, G. S., S. D. Kominers, K. M. Murphy, and J. L. Spenkuch (2018): "A theory of intergenerational mobility," *Journal of Political Economy*, 126, S7–S25.

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.

Bellemare, M. F. and C. J. Wichman (2020): "Elasticities and the inverse hyperbolic sine transformation," *Oxford Bulletin of Economics and Statistics*, 82, 50–61.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020): "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?" Tech. rep., National Bureau of Economic Research.

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., A. R. Cardoso, and P. Kline (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, 633–686.

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.

De Loecker, J. and J. Eeckhout (2021): "Global Market Power," Working Paper.

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

DUSTMANN, C., A. LINDNER, U. SCHÖNBERG, M. UMKEHRER, AND P. VOM BERGE (2022): "Reallocation effects of the minimum wage," *The Quarterly Journal of Economics*, 137, 267–328.

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

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

EECKHOUT, J. AND P. KIRCHER (2011): "Identifying Sorting In Theory," *The Review of Economic Studies*, 78, 872–906.

——— (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.

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.

JOCHMANS, K. AND M. WEIDNER (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

JÄGER, S., B. SCHOEFER, AND J. HEINING (2021): "Labor in the Boardroom," *Quarterly Journal of Economics*, 136, 669–725.

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

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 (2018): "On Worker and Firm Heterogeneity in Wages and Employment Mobility: Evidence from Danish Register Data," Working Paper.

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

LINDENLAUB, I. AND F. POSTEL-VINAY (2020): "Multidimensional Sorting under Random Search," Working Paper.

LINDNER, A., B. MURAKÖZY, B. REIZER, AND R. SCHREINER (2021): "Firm-level Technological Change and Skill Demand," *Working Paper*.

LINDQVIST, E. (2012): "Height and leadership," *Review of Economics and Statistics*, 94, 1191–1196.

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.

MANDELBROT, B. (1962): "Paretian distributions and income maximization," *The Quarterly Journal of Economics*, 76, 57–85.

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.

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.

Online Appendix Accompanying
"Firm Heterogeneity in Skill Returns"

Michael J. Böhm, Khalil Esmkhani, and Giovanni Gallipoli

MAY 2022

Contents

A Data and Samples Construction **2**
A.1 Data . . . 2
A.2 Estimation Samples . . . 6

B Overview of Econometric Methods **9**
B.1 Estimating Bias-Corrected Quadratic Forms . . . 9
B.2 Cluster-Based Estimation . . . 12
B.3 Testing Equality of Firm Effects across Worker Skills . . . 13

C A Labor Market with Two-Sided Heterogeneity and Heterogeneous Skills **15**
C.1 Production and Market Structure . . . 15
C.2 Base Pay and Skill Premia: Mapping Model to Firm Wages . . . 17

D Additional Estimation Results **19**
D.1 Results from Alternative Samples and Estimation Approaches . . . 19
D.2 Variance Accounting . . . 20
D.3 Implications for Matching of Workers with Firms . . . 23
D.4 The Uneven Gains from Sorting . . . 24

E Extensions and Robustness **28**
E.1 Industries and Occupations . . . 28
E.2 Capital Composition, Innovation, and Skill Returns . . . 35
E.3 Clustering Strategies and Number of Firm Clusters . . . 38

A Data and Samples Construction

A.1 Data

Base sample. The main data source for our analysis is the *Longitudinal Integrated Database for Health Insurance and Labor Market Studies* (LISA) by Statistics Sweden (SCB). LISA contains employment information (such as employment status, organization and workplace identifiers, industry and, from 2001, occupation), tax records (including labor and capital income) and demographic information (such as age, education) for all individuals 16 years of age and older domiciled in Sweden. LISA starts in 1990, with the most recent data including 2017.

Our measure of earning returns is annual labor income from the employer with highest recorded earnings. This is available for all workers, not top-coded, and includes end-of-year bonuses and performance pay. LISA reports a unique identifier for each individual's "company of employment", a so-called organization number, as well as a workplace identifier, which is the combination of organization number, employment location, and industry. To be consistent with the earning measure, and with the firm literature (see, among others, Card et al., 2013), we use the workplace with the highest earnings in a given year as the worker's "firm".

We keep workers dependently employed in the private nonprimary sector who earn above the *Prisbasbelopp* (the minimum amount of earnings that qualifies for the earnings-related part of the public pension system; see also Edin and Fredriksson, 2000). In 2008, the *Prisbasbelopp* was 41,000 kr or approximately 6,200 USD. We drop all observations with incomplete data (missing test scores, age, or workplace) and restrict the sample to 20–60 year old males. This process results in a sample of approximately 1 million unique workers, 26 thousand firms, and 6.6 million worker $\times$ year observations for the main sample period of 1999–2008.[1] Colum (1) of Table A.1 reports summary statistics for the resulting sample.

Measures of cognitive and noncognitive traits. A strength of our data source is that we have access to extensive and consistent measures of workers' cognitive and noncognitive attributes.

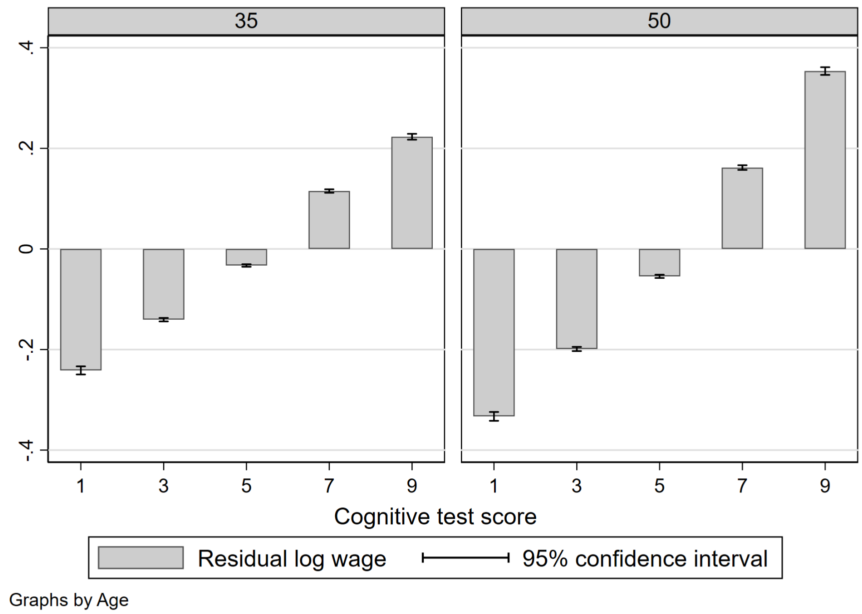
[1]We also document results for two alternative periods, 1990–1999 and 2008–2017.

This information comes from military enlistment tests, which were mandatory for Swedish males before 2007 and typically taken between age 18 and 19. In the early 2000s, Sweden started requiring progressively fewer males to do military service. The service was abolished in 2010. Before 2007, however, all males were required to take the military enlistment tests and test scores are available for almost 90 percent of males born up to the 1980s (e.g., see Figure A.1 in Böhm et al., 2022).

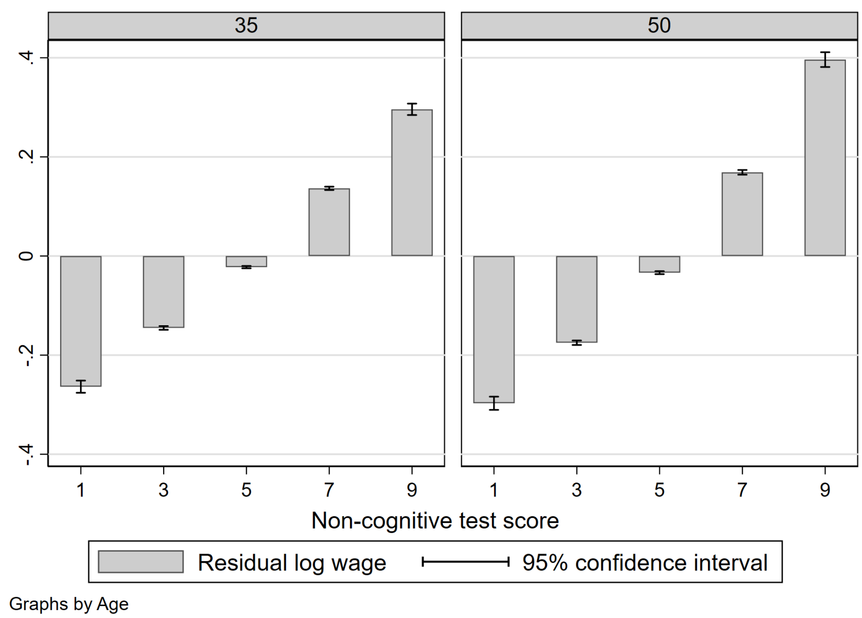
One might worry that certain individuals could deliberately perform badly on these tests to avoid military service. There are, however, several pieces of evidence suggesting this was not a major problem. In particular, we emphasize that employers routinely put considerable weight on military service performance and anecdotal evidence suggests that some positions – like being an officer in the navy – were important for the networks individuals would obtain; a substantial fraction of individuals working in influential positions within Swedish society went through these military service assignments. Consistent with these observations, and perhaps more importantly, military test scores have been shown to significantly predict future earnings at long intervals after the tests, as well as other labor market outcomes such as managerial positions and incidence of unemployment (see, e.g., Lindqvist and Vestman, 2011).

The enlistment process for military service spans two days and evaluates a person's medical and physical status as well as cognitive and mental abilities. We use the tests of cognitive and noncognitive ability, which are well established in the labor economics literature, for our analysis. The test of cognitive ability consists of four different parts (logic, verbal, spatial, and technical comprehension), each of which is constructed from 40 questions. These are aggregated into an overall score. The test is a rich measure of general competence and intelligence and it has a stronger fluid IQ component than the American AFQT, which focuses more on crystallized IQ. The aggregate cognitive score ranges from the integer value 1 (lowest) to 9 (highest), according to a STANINE (standard nine) scale that approximates a Normal distribution with a mean of 5 and standard deviation of 2 (meaning that a gap of two scores covers a standard deviation).

Noncognitive ability is assessed through a 25-minute semi-structured interview by a certified psychologist. Individuals are graded on, among others, their willingness to assume responsi-



(a) Cognitives



(b) Noncognitives

Figure A.1: Average Earnings of Males at Age 35 and 50, by Test Score Group.

Notes: Earnings for different test score ranks {1,3,5,7,9}; values are residualized using full age $\times$ year dummy interactions. Sample period: 1990–2017. 95% confidence intervals indicated by brackets.

bility, independence, outgoing character, persistence, emotional stability, and power of initiative (Swedish National Service Administration, referenced by, among others, Lindqvist and Vestman, 2011). The psychologist weighs these components together and assigns an overall noncognitive score on a STANINE scale. Lindqvist and Vestman (2011), on p. 108f, discuss in detail how the noncognitive score is related to, and different from, other measures often used in the literature on personality and labor market outcomes. Rather than assessing a unique trait, the noncognitive score assesses the ability to function in a demanding environment (military combat). Previous work provides robust evidence that these traits are also rewarded in the labor market.[2]

Test scores and later life outcomes. An important advantage of the military test scores is that they allow for a professional standardized measurement of different ability dimensions over a large population. Military enlistment scores are by design exogenous and predetermined with respect to individuals' career choices. Although cognitive and noncognitive ability are not fixed, they are hard for individuals to manipulate after late childhood or early adulthood (Hansen et al., 2004; Heckman et al., 2006). Crucially, as we show in Figure A.1, the tests are strongly associated to labor market outcomes and accurately predict earnings several decades later. Figure A.1 compares the earnings of workers with different STANINE scores in our sample (residualized using full age $\times$ year dummy interactions) and documents highly significant differences at ages 35 and 50, across both cognitive and noncognitive competencies. These plots emphasize the lasting informational content of the tests and their relevance for long term labor market outcomes. Strong significance at long lags is not always the case with ability tests in survey data and is partly due to the fine-grained and homogeneous nature of the procedures used to elicit different

[2]Individuals who score sufficiently high on the cognitive test are also evaluated for leadership ability, again on a STANINE scale. The leadership score is meant to capture the suitability to become an officer. Since leadership is only assessed for a subset of individuals, we focus on cognitive and noncognitive ability in our analysis. Noncognitive ability and leadership ability are also highly correlated; in our data the correlation is above 0.8, while the correlation of cognitive and noncognitive is 0.3.

attributes, resulting in measures that can be mapped into earnings for the whole population of interest over its working life cycle.[3]

A.2 Estimation Samples

Clustered estimation: sample and firm grouping. We concentrate on the largest set of firms that are connected via worker mobility. This corresponds to moving from column (1) to (2) in Table A.1, and is in fact not strictly necessary: for estimating clusters only mobility between firm classes is required, a condition almost trivially satisfied here. Nonetheless, we keep with existing literature and require connected firms; this is not a consequential sample restriction, as shown in Table A.1. The latter finding indicates that even our initial restrictions are enough to lead to a sample of relatively large and well-connected firms. Overall, there are 25,783 unique firms and 510,077 workers who move between firms at least once during 1999–2008 in column (2) of the table.

Next, we employ the k-means algorithm (see also Bonhomme et al., 2019, Section 4) to group firms into 100 clusters. We do this by using variation in mean earnings, mean cognitive, and mean noncognitive skills, which reflect the dimensions of firm heterogeneity that we are interested in. In particular, differing technologies should lead to variation in both firms' skill composition and earnings. We estimate model (10) using this sample and the definition of firm clusters (i.e., the j subscripts refer to the 100 clusters). Results are reported in Column (1) of Table 1. Section 6.3 in the paper and associated Appendix E.3 examine robustness with respect to alternative clustering criteria as well as to the number of firm classes.

Bias-correction estimation: leave-one-out match-level samples. The estimation of variance components with the bias correction requires a set of firms that are leave-one-out connected by mobility of high and low skill workers in both the cognitive and noncognitive dimension. We meet this condition by only sampling firms that are leave-one-out connected through: (i) low skill

[3] Aghion et al. (2017) further show that cognitive military test scores similar to ours strongly predict whether an individual becomes an inventor in Finland, another important later in life outcome and closely related to our analyses in Section 6.2.

Table A.1: Summary statistics for the estimation samples

	Full sample (1)	Largest connected (2)	Leave-one-out (3)	Match-level (4)
Number of observations	6,610,567	6,609,865	3,267,381	1,188,618
Number of stayers	578,146	578,146	-	-
Number of movers	510,077	510,077	477,424	477,424
Number of firms	25,839	25,783	19,085	19,085
Average log annual earnings	7.84	7.84	7.83	7.83
StDev log annual earnings	0.60	0.60	0.64	0.71
Average cognitive skill	5.28	5.28	5.44	5.44
Average noncognitive skill	5.13	5.13	5.23	5.23
Average age	37.32	37.32	36.35	36.35

Notes: Summary statistics for successively more restricted samples. Column (1) are all males aged 20–60 with non-missing employer, earnings, and test scores 1999–2008 at firms that exist at least five years with at least ten sample workers on average. Column (2) extracts the largest connected set of firms and their employees. Column (3) extracts the leave-one-out connected set of firms and removes workers who stay in the same firm in all years they are observed. Column (4) collapses the column (3) sample to worker–firm matches (summary statistics weighted by underlying frequencies). Earnings are real annual labor income in 2008 Swedish kronor. Cognitive and noncognitive scores are in Stanine scale. Our estimation samples are in bold font, (2) for clustering and (4) for the bias-correction approach.

workers ($c \leq 5, n \leq 5$), (ii) low in one and high in the other dimension workers ($c \geq 6, n \leq 5$ or $c \leq 5, n \geq 6$), and (iii) generally high-skill workers ($c \geq 6, n \geq 6$). A leave-one-out connected set of firms remains connected when any one worker is removed. This requires finding the workers that constitute cut vertices or articulation points in the corresponding bipartite network (Kline et al., 2020, Computational Appendix 2.1).

The algorithm to construct our estimation sample works as follows:

Step 1: We use Python's NetworkX package to identify the articulation points of the worker–firm graph, remove them and find the largest connected set that remains, then add back those articulation points that are connected to this largest leave-one-out connected set.

Step 2: We identify the largest leave-one-out connected set separately for the three skill groups (i)–(iii) and only keep those firms that are in the intersection of these sets.

We repeat Steps 1 and 2 until there is no reduction in the size of the graph (i.e., the three largest leave-one-out connected sets coincide). This final set is leave-one-out connected for the three skill groups.

We estimate the model at the worker–firm match level to account for potential serial correlation within worker–firm employment spells. That is, we collapse the data to means and drop workers who stay in the same firm throughout the period, since in the match-level estimation these do not contribute to identifying the firm effects. We thereby follow exactly Kline et al. (2020, Appendix A)'s recommendations for estimating variance components in panels of $T > 2$.

The final firm-level sample to estimate (10) is summarized in Column (4) of Table A.1. This consists of 19,084 unique firms and 477,423 mover workers within the firm-level sample. The leave-one-out connectedness requirement increases employer size as it reduces the number of firms (26%) relatively more than the number of workers (7%). However, these reductions seem to have moderate effects. The average and dispersion of earnings do not change much but workers in the firm-level sample with larger firms are slightly younger and more skilled. The smaller number of observations in the match-level sample, without stayers and collapsed to the worker–firm match level, also reduces the computational burden (see footnote 6 below). For comparison, we also show results for the leave-one-observation-out sample in Table D.1 and, as expected, estimated dispersions of firm returns are substantially larger. In that sense, the match-level results in the main text are conservative.

B Overview of Econometric Methods

Throughout the paper we use high-dimensional firm effects specifications featuring firm-specific returns to cognitive and noncognitive skills. Estimates from these models are employed to study quadratic forms of model parameters. The baseline linear model is[4]

$$\log(w_{ijt}) = \mu_i + \lambda_j^0 + c_i \cdot \lambda_j^c + n_i \cdot \lambda_j^n + \varepsilon_{ijt}.$$

Of particular economic interest is the set of second moments of firm and worker specific parameters. For instance, in the standard double fixed effect model, one might interpret $\text{cov}(\mu, \lambda^0)$ as a measure of sorting of high-type workers into high-type firms. However, the naive plug-in estimates of these moments are prone to biases. In fact, developing unbiased estimators of such quadratic forms is the object of several papers in the firm heterogeneity literature (Andrews et al., 2008; Bonhomme et al., 2019; Kline et al., 2020; Bonhomme et al., 2020). Since our interest i in studying similar second moments, in what follows we briefly overview some details about the methods we employ to estimate firm effects.[5]

B.1 Estimating Bias-Corrected Quadratic Forms

We begin by rewriting our baseline specification as:

$$\begin{aligned} \log(w_{ijt}) &= \mu_i + \lambda_j^0 + c_i \cdot \lambda_j^c + n_i \cdot \lambda_j^n + \varepsilon_{ijt}, \\ &\equiv \mathbb{X}_{ij}\beta + \varepsilon_{ijt} \end{aligned} \tag{B.1}$$

where $\beta = [\mu; \lambda^0; \lambda^c; \lambda^n]' \equiv [\mu_1, ..., \mu_I;\ \lambda_1^0, ..., \lambda_J^0;\ \lambda_1^c, ..., \lambda_J^c;\ \lambda_1^n, ..., \lambda_J^n]'$ is the parameter vector and $\mathbb{X}_{ij} = [\mathbf{1}_i,\ \mathbf{1}_j,\ c_i\mathbf{1}_j,\ n_i\mathbf{1}_j]$ is the data matrix.

[4] In the specifications studied in the main body we also include a broad set of control variables which are ignored here for notational simplicity.

[5] For in-depth discussions of these estimators see Kline et al. (2020) and Bonhomme et al. (2017, 2019).

The symbol $\mathbf{1}_i$ denotes a $I \times 1$ indicator vector whose elements are all zero except the i^{th} coordinate (corresponding to worker i) which is set to 1. Similarly $\mathbf{1}_j$ is a $J \times 1$ indicator vector for firm j.

Kline et al. (2020) suggest an unbiased estimator for arbitrary quadratic forms involving the coefficients of (B.1) in the form of $\beta' A \beta$, for given matrix A. By appropriately choosing the A matrix, one can recast all the second moments of firm parameters λ_j^0, λ_j^c, and λ_j^n into quadratic expressions of the form $\beta' A \beta$.

Constructing quadratic forms. We begin by definining three row vectors associated to different firm parameters: $\mathbb{X}_{ij}^0 = [0_{1\times I},\ \mathbf{1}_j,\ 0_{1\times J},\ 0_{1\times J}]$, $\mathbb{X}_{ij}^c = [0_{1\times I},\ 0_{1\times J},\ \mathbf{1}_j,\ 0_{1\times J}]$, and $\mathbb{X}_{ij}^n = [0_{1\times I},\ 0_{1\times J},\ 0_{1\times J},\ \mathbf{1}_j]$, where i identifies worker and j is firm. Also, we let $\mathbb{X}$ denote the matrix that results from vertically stacking all the observations in row vector $\mathbb{X}_{ij}$. Then, $\mathbb{X}^0$, $\mathbb{X}^c$, and $\mathbb{X}^n$ denote the matrices that result from vertically stacking $\mathbb{X}_{ij}^0$, $\mathbb{X}_{ij}^c$ and $\mathbb{X}_{ij}^n$. Finally, we define

$$
\begin{aligned}
A^0 &= \frac{1}{\sqrt{N}}\left(\mathbb{X}^0 - \overline{\mathbb{X}}^0\right)\\
A^c &= \frac{1}{\sqrt{N}}\left(\mathbb{X}^c - \overline{\mathbb{X}}^c\right)\\
A^n &= \frac{1}{\sqrt{N}}\left(\mathbb{X}^n - \overline{\mathbb{X}}^n\right)
\end{aligned}
$$

where $\overline{\mathbb{X}}^0 = \frac{1}{N}[0_{N\times I},\ 1_{N\times J},\ 0_{N\times J},\ 0_{N\times J}]$, $\overline{\mathbb{X}}^c = \frac{1}{N}[0_{N\times I},\ 0_{N\times J},\ 1_{N\times J},\ 0_{N\times J}]$, and $\overline{\mathbb{X}}^n = \frac{1}{N}[0_{N\times I},\ 0_{N\times J},\ 0_{N\times J},\ 1_{N\times J}]$. One can use the A matrices above to estimate second moments of interest, e.g. $\text{VAR}(\lambda^0) = \beta'(A^{0\prime}A^0)\beta$ or $\text{COV}(\lambda^c, \lambda^n) = \beta'(A^{c\prime}A^n)\beta$. In what follows we set $A = A_1' A_2$ to estimate $\theta = \beta' A \beta$, where A_1 and A_2 could be any of A^0, A^c, and A^n (depending on which moments we are interested in).

Plug-in estimator. The plug-in estimator $\hat{\theta}_{\text{PI}} = \hat{\beta}' A \hat{\beta}$ can be obtained by simply using the OLS estimates of $\hat{\beta}$ in the quadratic form defining θ. However, the plug-in estimator is biased and its expected value is

$$\mathbb{E}[\hat{\theta}_{\text{PI}}] = \theta + \text{trace}(A \times \text{VAR}[\hat{\beta}]) = \theta + \sum_{k=1}^{N} B_{kk}\sigma_k^2 \tag{B.2}$$

where $S = \mathbb{X}'\mathbb{X}$, B_{kk} is the k-th diagonal element of $B = \mathbb{X}S^{-1}AS^{-1}\mathbb{X}'$ corresponding to observation k, and σ_k^2 is the variance of error term of observation k. Therefore, the bias in the plug-in estimator can be corrected by using unbiased estimates of σ_k^2, which is the route we take when estimating the model at the level of individual firms.

Bias-corrected quadratic forms. We use leave-k-out OLS estimators of β, denoted by $\hat{\beta}_{-k}$, that are obtained from a sample where the observation k is excluded. This delivers an unbiased estimator of σ_k^2 such that

$$\hat{\sigma}_k^2 = y_k(y_k - x_k\hat{\beta}_{-k}), \tag{B.3}$$

where y_k is the dependent variable (i.e. log earnings) of observation k and x_k is the corresponding independent variables vector (i.e. row k of $\mathbb{X}$). Using the $\hat{\sigma}_k^2$ above, we compute the bias corrected estimator of θ as

$$\hat{\theta}_{KSS} = \hat{\beta}'A\hat{\beta} - \sum_{k=1}^{N} B_{kk}\hat{\sigma}_k^2. \tag{B.4}$$

Large Scale Computations. Estimating $\hat{\theta}_{KSS}$ is computationally expensive for large data-sets with many estimated parameters such as ours. Like Kline et al. (2020), we use a variant of the random projection method of Achlioptas (2003, known as Johnson-Lindenstrauss Approximation, or JLA) to estimate the $\hat{\sigma}_k^2$ and B_{kk} required in the estimation of $\hat{\theta}_{KSS}$. JLA suggests the following approximation:

$$\begin{aligned}
\hat{P}_{kk} &= \frac{1}{p}||R_P\mathbb{X}S^{-1}x_k||^2 \\
\hat{B}_{kk} &= \frac{1}{p}(R_BA_1S^{-1}x_k)'(R_BA_2S^{-1}x_k) \\
\hat{\sigma}_{k,JLA}^2 &= \frac{y_k(y_k - x_k\hat{\beta})}{1-\hat{P}_{kk}}\left(1 - \frac{1}{p}\frac{3\hat{P}_{kk}^3 + \hat{P}_{kk}^2}{1-\hat{P}_{kk}}\right),
\end{aligned}$$

where $p \in \mathbb{N}$ is a number much smaller than the total number of estimated parameters. That is, we can achieve a material reduction in the dimensionality of the prolem. The $R_P, R_B \in \{-1,1\}^{p \times N}$ are random matrices of order $p \times N$ featuring elements equal to +1 and -1 with equal probabilities. This makes computations significantly faster when parameters are estimated at the level of individual firms.[6]

B.2 Cluster-Based Estimation

Models with two sided heterogeneity rely on job movers to identify the unobserved firm and worker parameters. In typical employer–employee linked data sets the number of job movers per firm tends to be small, which leads to the well known limited mobility bias in quadratic forms of these estimates. To alleviate this problem, group based estimates have been suggested in the literature. In this approach, firm parameters are assumed to only vary across groups or clusters of firms, rather than individual firms. Under this assumption about the underlying data generating process and further assuming that the number of groups is limited, the number of job moves per group of firms is sufficiently large, which alleviates the small sample bias concern.

Partitioning returns across clusters. To adapt this framework to our setting, we begin by rewriting the baseline specification as

$$\log(w_{ijt}) = \mu_i + \lambda^0_{g(j)} + c_i \cdot \lambda^c_{g(j)} + n_i \cdot \lambda^n_{g(j)} + \varepsilon_{ijt},$$

where $g : \{1,...,J\} \to \{1,...,K\}$ is a partitioning function that maps firm j into cluster $g(j)$ that the firm j belongs to, and K is the total number of groups. These groups could in principle be the individual firms, i.e., $g(j) = j$, but only in models with a reduced number of groups is the limited mobility bias less of a concern.

[6] Estimating the bias-corrected second moments of parameters in model (10) on the data in Column (4) of Table A.1 takes about 20–30 hours using Python and the JLA approximation with $p = 50$ depending on the Swedish server's workload. Setting $p = 50$ is in line with Kline et al. (2020) and we have tested that further increasing p does not change our results.

Two-step estimation. We estimate the model in two steps i.e. (i) partition firms into K disjoint groups (ii) estimate the model using the firm groups. Bonhomme and Manresa (2015) show that a k-means estimator can consistently identify the firm classes up to a relabeling of groups. In the first step, as discussed in Section 2.2, we use the average earnings as well as average cognitive and noncognitive traits of their workers to group firms. Intuitively, the earnings and average skills in firms with identical intercepts and returns should be the same and one could then use these observed firm variables to define separate firm classes. The structural literature advocates going beyond earnings when clustering firms (Eeckhout and Kircher, 2011; Hagedorn et al., 2017; Bartolucci et al., 2018; Bagger and Lentz, 2019), since a classification may fail to be identified when two firm classes have identical earnings distributions in the cross section.[7]

Once firm groups are defined, firm and worker parameters are identified (up to the normalization discussed in Section 3.1 of the main text) under the assumptions of serial conditional independence of earnings and random job mobility (Bonhomme et al., 2019), and estimated using panel regressions in conjunction with skill proxies.

B.3 Testing Equality of Firm Effects across Worker Skills

In Section 2.3, and Figure 1, we choose the years 2004 and 2007 to test the equality of firm effects for high versus low skilled workers. Two years are selected to exclude potential serial correlation within employment spells due to estimated standard errors (see Kline et al., 2020, Computational Appendix 2.5).

Years are non-adjacent, in order to remove partial employment years when workers switch firms, while not too far apart to minimize any potential changes in firm effects over time. The sample is selected on firms that are leave-one-out connected in both high and low levels of the respective skill dimension ("double-connected").

To gauge robustness, we also replicate the analysis for alternative duplets of years. Like in Section 2.3, we focus on testing the null hypothesis that firm effects are equal for high and

[7] For example, a firm class may have higher intercepts and the other higher returns but worker sorting is such that observed earnings are the same. See also discussion in Bonhomme et al. (2019, page 14).

low skills groups; the hypotheses are separately tested for cognitive and noncognitive skills. Table B.1 shows the resulting test statistics and sample sizes of the respective double-connected individual firms for several year pairs using the bias-correction approach. The last column reports the corresponding test statistics among 100 firm classes using clustering as in Figure 1(c)–(d).

Table B.1: Tests for equality of firm effects by high- versus low-skill workers (by year combination and cognitive / noncognitive)

Year origin	Year destination	Skill	Test Statistic Firm-level	# Firms	Test Statistic Grouped
(1)	(2)	(3)	(4)	(5)	(6)
1999	2002	C	3.66	8,757	9.22
1999	2002	N	2.49	9,766	12.45
2000	2003	C	2.60	8,653	8.66
2000	2003	N	0.39	9,648	8.14
2001	2004	C	2.76	7,922	9.93
2001	2004	N	1.78	8,941	7.65
2002	2005	C	0.60	7,904	10.83
2002	2005	N	3.50	8,772	6.96
2003	2006	C	4.04	8,335	13.88
2003	2006	N	0.85	9,258	7.00
2004	2007	C	4.18	9,269	17.33
2004	2007	N	4.56	10,209	6.30
2005	2008	C	3.26	9,846	10.74
2005	2008	N	2.54	10,825	5.38

Notes: Table B.1 expands on Figure 1 to show test statistics associated with the null hypothesis that firm effects ($\theta_j^{S=0}$) and ($\theta_j^{S=1}$) are equal across skill level, where skill $S \in \{C,N\}$. Test statistic for firm-level bias-adjusted estimates as in Figure 1(a)–(b) are shown in column (4). The associated number of double-connected firms in each of the skill types and year combinations are reported in column (5). The last column reports the corresponding test statistic among 100 firm classes using the clustering approach as in Figure 1(c)–(d).

C A Labor Market with Two-Sided Heterogeneity and Heterogeneous Skills

To examine the interaction of employer and employee heterogeneity we develop an empirically tractable model featuring workers with different cognitive and noncognitive abilities. We consider a static setting with a continuum of firms, each producing its own distinct product using labor. All firms benefit from more able workers, but each firm exhibits an idiosyncratic return to skills. Firm-specific skill returns act as a force for sorting of high-skill workers into high-return firms, something that the matching literature has long emphasized. These layers of heterogeneity are embedded in a labor market where employers choose how many workers to hire based on the demand for their output. Equilibrium implies that the labor market clears.

C.1 Production and Market Structure

There is a measure one of workers who differ in their observable cognitive (c) and noncognitive (n) abilities and we let $G(c,n)$ denote the measure describing the distribution of worker types in the economy. A worker's utility from being matched with a specific firm depends on the wage they receive from that firm plus an idiosyncratic preference shock. For worker i of type (c,n), the utility of working at firm j with wage $w_j(c,n)$ is

$$u_{ij}(c,n) = \beta \log(w_j(c,n)) + \nu_{ij} \tag{C.1}$$

where ν_{ij} captures an idiosyncratic preference for working at firm j. We assume that shocks ν_{ij} are independent draws from a Type I Extreme Value distribution. This specification could be expanded adding firm-level variation of average amenities as in Sorkin (2018).

Given wages, workers choose the 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

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

Equation (C.2) delivers the upward sloping labor supply equation faced by firm j, with elasticity of supply β. The intercept $h(c,n)$ is determined in equilibrium and guarantees market clearing (every worker gets a job), that is

$$h(c,n) = \left[\int w_k(c,n)^{\beta} \, \mathrm{d}F(k)\right]^{-1} \tag{C.3}$$

where $F(\cdot)$ is the probability measure describing the distribution of firms in the economy.

As in Lise and Robin (2017), the production function is defined at the level of the match and we do not model complementarity between workers within a firm. A worker of type (c,n) employed at firm j produces according to $f_j(c,n)$, where the function f_j describes the output from the firm-worker match. Technology is CRS and a firm's output is the sum of all employees' products.[8] Firm j's total output is

$$y_j = \int f_j(c,n) q_j(c,n) \, \mathrm{d}G(c,n). \tag{C.4}$$

In the output market, firms face a downward sloping demand curve for their products. Firm j's inverse demand is

$$\log(p_j) = \log(\phi_j) - \frac{1}{\sigma}\log(y_j) \tag{C.5}$$

where p_j is product price, y_j is output, ϕ_j is a firm-specific (inverse) demand intercept, and σ is the output demand elasticity with respect to price.

The firm's problem. Given output demand and labor supply curves, a firm decides how many workers to hire for each skill type. Firm j's profit maximization problem is:

[8] Additive separability is often assumed in matching models with one-to-many sorting. In the empirical section we show how this technology specification delivers an accurate approximation of returns to different skill types. While convenient, the separability assumption is not crucial for our findings about sorting and returns heterogeneity.

$$
\begin{aligned}
\max_{q_j(c,n)} \quad & p_j y_j - \int w_j(c,n) q_j(c,n)\ \mathrm{d}G(c,n) \\
s.t. \quad & y_j = \int f_j(c,n) q_j(c,n)\ \mathrm{d}G(c,n) \\
& \log(p_j) = \log(\phi_j) - \tfrac{1}{\sigma} \log(y_j) \\
& \log(q_j(c,n)) = \log(h(c,n)) + \beta \log(w_j(c,n))
\end{aligned}
\tag{C.6}
$$

This problem has a closed form solution, with equilibrium wages in firm j

$$
w_j(c,n) = \frac{\left(\frac{\beta}{1+\beta}\right)^{\frac{\sigma}{\sigma+\beta}} f_j(c,n) \left(\frac{\sigma-1}{\sigma}\phi_j\right)^{\frac{\sigma}{\sigma+\beta}}}{\left[\int f_j(c,n)^{1+\beta} h(c,n)\ \mathrm{d}G(c,n)\right]^{\frac{1}{\sigma+\beta}}} \tag{C.7}
$$

C.2 Base Pay and Skill Premia: Mapping Model to Firm Wages

Firms' production choices can be characterized along the two input dimensions (cognitive and noncognitive). Every worker has a type within the set (c,n), with the first letter denoting cognitive level and the second noncognitive level. The wage premium associated to skill bundle (c,n) in firm j is

$$
e^{\Delta_j(c,n)} = \frac{f_j(c,n)}{f_j(L,l)} \tag{C.8}
$$

for all (c,n). This corresponds to the wage relative to the low-type worker (L,l), $\left(\frac{w_j(c,n)}{w_j(L,l)}\right)$, since everything else in the wage equation (C.7) cancels. The premium $e^{\Delta_j(c,n)}$ is proportional to the (measurable) productivity of a (c,n) worker in firm j relative to a baseline worker of type (L,l). The parameter $\Delta_j(c,n)$ subsumes two sources of variation: (1) the skill endowment bundle (c,n), and (2) the return to that bundle in firm j. By definition, $\Delta_j(L,l) = 0$ and one can redefine baseline match productivity in firm j as $T_j = f_j(L,l)$, which is the output of workers of type (L,l). Using T_j and $\Delta_j(c,n)$, we write the output of firm j as $y_j = T_j \sum_{(c,n)} e^{\Delta_j(c,n)} q_j(c,n) \mathrm{d}G(c,n)$, where $\mathrm{d}G(c,n)$ with some abuse of notation denotes the total number of (c,n) type workers, and recast the profit maximization as a choice over a discrete set of skill bundles (c,n).

Optimal hiring behavior in the discrete maximization problem implies:

$$w_j(c,n) = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{\frac{\sigma-1}{\sigma}\phi_j T_j \left(\frac{1}{y_j}\right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{e^{\Delta_j(c,n)}}_{\text{Skill Productivity}} \tag{C.9}$$

This expression captures different aspects of market structure. The marginal revenue is an increasing function of the firm's output demand ϕ_j. However, the monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker, with the fraction approaching one in more competitive markets where the labor supply elasticity β is larger. An extra unit of skill rescales marginal revenues proportionally to the firm's skill return $\Delta_j(c,n)$.

In log form, the equilibrium wage lends theoretical underpinning the empirical specifications in the paper. That is:

$$\log(w_j(c,n)) = \alpha + \Lambda_j + \Delta_j(c,n). \tag{C.10}$$

The intercept $\alpha \equiv \log\left(\frac{\beta}{1+\beta}\frac{\sigma-1}{\sigma}\right)$ is common across firms and skills, while $\Lambda_j \equiv \log\left(\phi_j T_j y_j^{-\frac{1}{\sigma}}\right)$ is the firm-specific baseline wage, which does not vary with worker skills; $\Delta_j(c,n)$ is a *firm-specific return to skill bundle* (c,n). Under the model's null hypothesis, the firm's demand intercept ϕ_j is subsumed in the fixed effect component Λ_j.

Optimal behavior implies that firms with higher returns to (c,n)-type skills tend to hire a larger share of (c,n)-type workers. This observation suggests that firms with similar returns to a skill type can be grouped together based on their share of workers with that particular type.

D Additional Estimation Results

D.1 Results from Alternative Samples and Estimation Approaches

Columns (1) and (2) in Table D.1 show the standard deviations of individual firm effects when we do not apply the quadratic-form correction (plug-in values) or when the sampling entails leaving single observations (worker–year) out, rather than the whole worker–firm spell as we do in Table 1 of the paper. As expected, when comparing to the baseline results, both these alternative specifications result in more pronounced firm heterogeneity. In this sense, our baseline estimates provide a conservative view of firm return variation. Details about the different sampling approaches (e.g. leaving out one worker-firm observation rather than the whole match) are discussed in the Appendix Section A.2. Colums (3) and (4) in Table D.1 show results when we

Table D.1: Standard deviations of firm parameters in alternative estimations.

	Firm-level (1999–2008):		Grouped (alt. periods):	
	Plug-in	Leave-obs-out	1990–1999	2008–2017
	(1)	(2)	(3)	(4)
sd(λ_j^0)	0.32	0.22	0.10	0.09
sd(λ_j^c)	0.40	0.21	0.10	0.07
$\times$ $90^{th} - 10^{th}$ *pct, cog score (c)*	*0.30*	*0.15*	*0.07*	*0.06*
sd(λ_j^n)	0.39	0.17	0.05	0.05
$\times$ $90^{th} - 10^{th}$ *pct, noncog score (n)*	*0.30*	*0.13*	*0.04*	*0.04*
$\times$ $90^{th} - 10^{th}$ *pct, cumulative (c+n)*	*0.59*	*0.28*	*0.11*	*0.09*
# unique firms	19,085	19,085	20,484	22,079

Notes: The table shows standard deviations of parameters λ_j^0, λ_j^c, and λ_j^n estimating (10) in alternative specifications and periods. Column (1) are plug-in estimates at the firm-level without quadratic-form correction. Column (2) quadratic-form corrects the firm-level variances leaving one observation (i.e., worker in a given year) rather than match (i.e., worker–firm spell) out at a time. Estimation period: 1999–2008. Columns (3) and (4) show the firm-clustered estimates in alternative periods 1990–1999 and 2008–2017. Otherwise notes to Table 1 apply.

re-estimate the model using the clustering approach for alternative sample periods. Dispersion of firm returns is slightly higher in 1990–1999 than in the baseline estimation period in Table 1. It is slightly lower in 2008–2017, alongside a lower standard deviation of firm intercepts. Overall, the dispersion of firm parameters appears remarkably stable over time.

D.2 Variance Accounting

To facilitate comparisons of our findings to existing work, it is useful to characterize the contribution of different layers of firm heterogeneity to the overall variance of earnings. We perform this exercise for either of the two estimation approaches (bias-correction and clustering); we also report similar decompositions for standard AKM estimators that do not explicitly account for firm-level heterogeneity in skill returns. Finally, to illustrate robustness of results under each approach, we carry out the analysis for the full sample (where each worker–firm match is observed for possibly multiple periods) and for the collapsed match-level samples (where each worker–firm match represents the average value over possibly several periods in which worker and employer are jointly observed).

To control for variation due to worker-only components, we define $\alpha_{it} \equiv \mu_i + \boldsymbol{X}_{it}\boldsymbol{b}_t$. This is consistent with the normalization adopted in Section 3.1 of the paper, where μ_i contains average returns $\kappa_c \cdot c_i + \kappa_n \cdot n_i$ across firms, and α_{it} accounts for both observed and unobserved worker-level variation. The firm-related component (possibly capturing interactions with worker skills) is defined as $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. In a standard AKM specification this latter component reduces to firm fixed effects.

Table D.2 shows the variance accounting exercise when we estimate the baseline equation (10) using 100 firm clusters for the 1999–2008 period. Results are similar for the 1990–1999 and 2008–2017 estimation periods, and comparable to grouping-based implementations for Sweden (Bonhomme et al., 2019) and the U.S. (Lamadon et al., 2022). As often found, worker-level heterogeneity accounts for much of the total earnings variation while the covariance between α and ψ is the second largest contributor to total variation (consistent also with Bonhomme

Table D.2: Variance decomposition of log earnings (shares $\times$ 100). Clustered firms approach with one hundred classes.

	$\frac{\text{var}(\alpha_{it})}{\text{var}(log(w_{ij}))}$ (1)	$\frac{\text{var}(\psi_{ij})}{\text{var}(log(w_{ij}))}$ (2)	$\frac{2\text{cov}(\alpha_{it},\psi_{ij})}{\text{var}(log(w_{ij}))}$ (3)		
Full sample				Obs. (million)	6.48
Full model	60.8	3.8	11.2	total	75.8
AKM	61.0	3.7	11.0	total	75.7
Match-level collapsed sample				Obs. (million)	1.19
Full model	62.0	4.5	14.0	total	80.5
AKM	62.1	4.4	13.9	total	80.4

Notes: Decomposition of the percentage in log earnings variance explained based on estimates from specification (10). We subsume worker-only contributions in $\alpha_{it} \equiv \mu_i + \boldsymbol{X}_{it}\boldsymbol{b}_t$ and firm/worker contributions in $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. We group firms into 100 clusters following the clustering approach as described in the text. AKM is alternative without heterogeneous firm returns; here it means fixed effects for the firm groups, not the individual firms. Estimation period: 1999–2008.

et al., 2020, who study several countries). Perhaps most interesting is the observation that one would obtain similar results when restricting the specification to a standard AKM with no skill interactions. This suggests that the economically significant heterogeneity in returns would be mistakenly attributed to employer and worker fixed effects. This is especially concerning when interpreting employer fixed effects as skill independent earnings shifts.

Table D.3 shows the variance accounting exercise when the coefficients in (10) are estimated using the bias-correction approach. Since this approach adjusts the quadratic forms for worker-only variation downward, the contribution from worker fixed effects is somewhat lower (relative to the clustering approach) although it remains the largest by far. Consistent with the estimates reported in Table 1 of the paper, the direct impact of firm heterogeneity on total variation is also larger but remains a smaller share of the total. Due to the downward rescaling of the quadratic forms, total explained variation is lower than for the clustered estimation. The comparison to

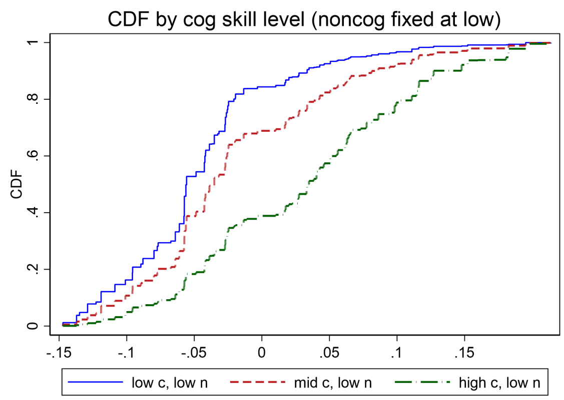
Table D.3: Variance decomposition of log earnings (shares $\times$ 100). Variance correction approach individual firm estimates (bias-corrected).

	$\frac{\text{var}(\alpha_i)}{\text{var}(log(w_{ij}))}$ (1)	$\frac{\text{var}(\psi_{ij})}{\text{var}(log(w_{ij}))}$ (2)	$\frac{2\text{cov}(\alpha_i,\psi_{ij})}{\text{var}(log(w_{ij}))}$ (3)		
Leave-one-out sample				Obs. (million)	3.27
Full model	49.4	8.4	5.8	total	63.6
AKM	42.7	8.0	5.3	total	55.9
Match-level collapsed sample				Obs. (million)	1.19
Full model	47.8	7.1	11.0	total	65.9
AKM	42.9	7.6	8.2	total	58.8

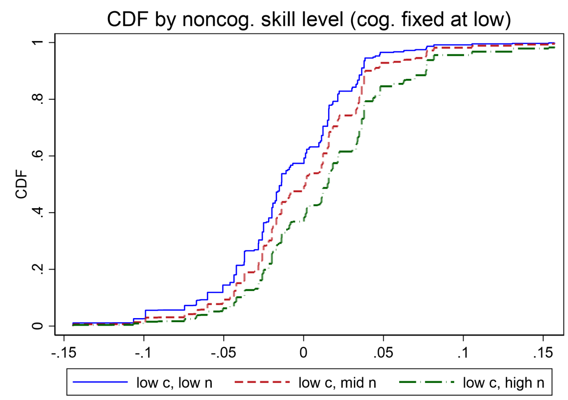
Notes: Decomposition of the percentage in log earnings variance explained based on estimates from specification (10). We capture worker-only contributions in $\alpha_i \equiv \mu_i$ and firm/worker contributions in $\psi_{ij} \equiv \lambda_j^0 + \lambda_j^c \cdot c_i + \lambda_j^n \cdot n_i$. Estimation period: 1999–2008.

the restricted AKM specification confirms that return heterogeneity is mistakenly conflated into separate employer and worker fixed effects. This concern becomes even more relevant when we observe that, under the full unrestricted model, the covariation of worker and firm effects also become larger.

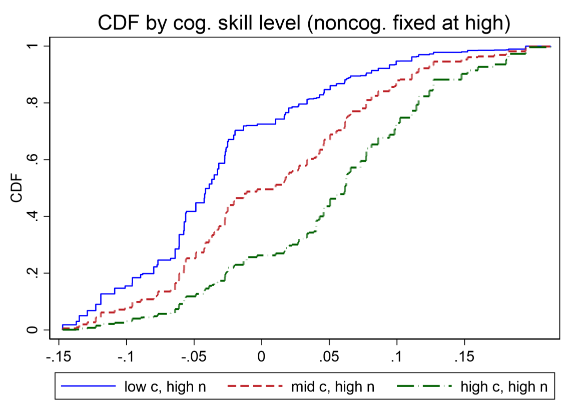
D.3 Implications for Matching of Workers with Firms



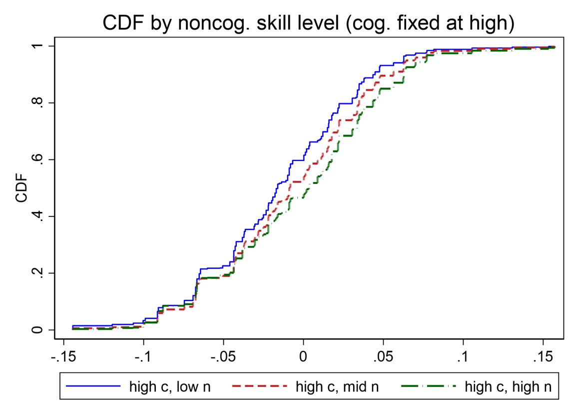
(a) λ_j^c, FOSD (Cognitive)



(b) λ_j^n, FOSD (Noncognitive)



(c) λ_j^c, FOSD (Cognitive)



(d) λ_j^n, FOSD (Noncognitive)

Figure D.1: Distribution of firm returns for different sets of worker skills.

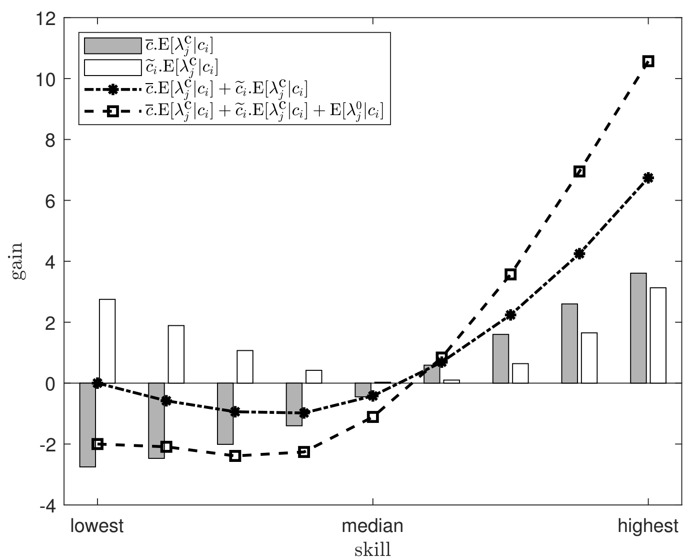
Notes: The figure shows cumulative distribution functions for workers with low ($c,n \leq 0.25$), mid ($0.25 < c,n < 0.75$), or high ($c,n \geq 0.75$) skill ranks over the range of firm returns. Period: 1999–2008. Results from the grouped estimator. FOSD: first-order stochastic dominance.

D.4 The Uneven Gains from Sorting

Robustness to rescaling of skills and returns. The sorting gains discussed in the main text are robust to rescaling of skills and returns since (i) multiplication of c_i by a non-zero factor would lead to a proportional change in the λ_j^c estimates as these would be scaled down by the same factor, leaving the product $\lambda_j^c c_i$ unchanged. (ii) Shifting the level of skills, by adding a constant x to c_i, leaves λ_j^c unchanged and shifts firm intercepts to $\lambda_j^0 - \lambda_j^c x$. Returns from working in firm j become $\lambda_j^c(c_i + x)$ but this is offset by $\lambda_j^0 - \lambda_j^c x$.

The total sorting gain, corresponding to the sum of both intercepts and returns ($\lambda_j^0 + \lambda_j^c c_i$), is hence fully invariant. This cumulative effect, calculated as the sum of columns (2) and (6) in Table 5, induces even larger inequality and skewness across the range of skill levels. Match effects are completely unaffected by rescaling, since they are defined relative to the demeaned $\widetilde{c_i}$.

Figure D.2: Gains from sorting across returns λ_j^c for different cognitive skill levels.



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 with detailed numbers in Table 5. Sample period: 1999–2008.

The dashed line in Figure D.2 shows the total sorting gain in the cognitive dimension, that is $\mathrm{E}(\lambda_j^0 \mid c_i) + c_i \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$. This induces even wider earning differences between skill levels and retains the strong convexity. The average effect, i.e., the aggregate gain from matching, is exactly the same as for the thick dotted line $c_i \cdot \mathrm{E}(\lambda_j^{\mathrm{c}} \mid c_i)$ already seen in the main text.

Gains from sorting on noncognitive returns. Table D.4 reports the effects from the sorting of noncognitive attributes n_i across noncognitive returns λ_j^{n}. These effects are comparatively smaller than in the cognitive dimension, which reflects the lower dispersion of noncognitive returns across firms (see Section 3) and the weaker sorting in that dimension (see Section 4). Nonetheless, there is clear evidence of sorting also in the noncognitive dimension.

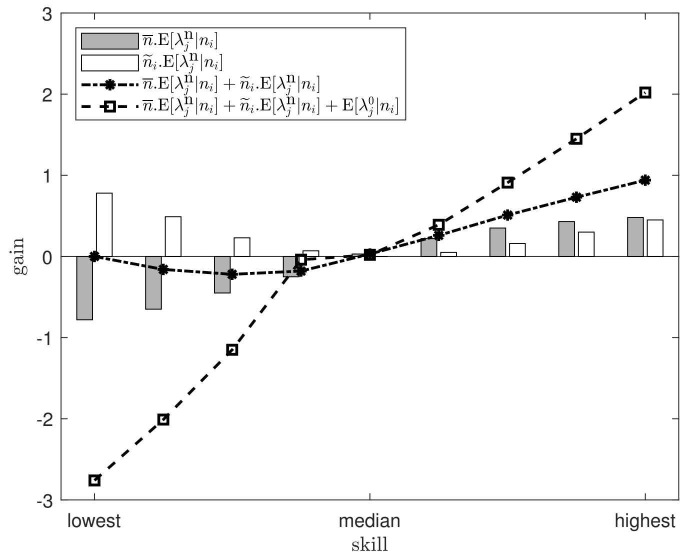
Table D.4: Gains from sorting across returns λ_j^{n} for different noncognitive skill levels.

	$\mathrm{E}(\lambda_j^{\mathrm{n}} \mid n_i)$ (1)	Full gain (2)	Returns effect (3)	Match effect (4)	$\mathrm{E}(\lambda_j^0 \mid n_i)$ (5)
skill level (n_i):					
1 (lowest, $n_i = 0$)	*-1.52*	0.00	-0.78	0.78	-2.76
2	*-1.26*	-0.16	-0.65	0.49	-2.01
3	*-0.87*	-0.22	-0.45	0.23	-1.15
4	*-0.48*	-0.18	-0.25	0.07	-0.04
5 (median, $n_i = 0.5$)	*0.06*	0.03	0.03	0.00	0.02
6	*0.42*	0.26	0.22	0.05	0.39
7	*0.68*	0.51	0.35	0.16	0.91
8	*0.84*	0.73	0.43	0.30	1.45
9 (highest, $n_i = 1$)	*0.94*	0.94	0.48	0.45	2.02
Aggregate	*0.00*	0.13	0.00	0.13	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.

Column (1) in Table D.4 shows that workers with higher noncognitive endowments sample from a distribution of employers with higher returns. Moving from $n_i = 0$ to $n_i = 1$ there is a 2.5 log points difference in $\mathrm{E}(\lambda_j^{\mathrm{n}} \mid n_i)$. This again leads to non-monotonic gains, since high-skill

Figure D.3: Gains from sorting across returns λ_j^{n} for different noncognitive skill levels.



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 with detailed numbers in Table D.4. Sample period: 1999–2008.

workers benefit the most from the sorting whereas the lowest-skill workers would benefit (or lose) little from any skill returns. The workers who experience steep losses are those with intermediate skills since they would gain from matching with high return firms but are not assigned to such firms. Match effects in column (4) reflect the complementarity of high-skill workers with high-return firms, and of low-skill workers with low-return firms, as well as the induced sorting. These are again positive and raise aggregate earnings by 0.13 log points. In the last column of Table D.4, inequality is further increased by the sorting of noncognitive attributes n_i over λ_j^0 intercepts.

Figure D.3 represents these effects visually. The earnings differences between skill levels are clearly convexified by the sorting (thick dotted line), albeit the convexification is not as pronounced as for cognitive traits. Interestingly, sorting over intercepts reverses this convexification (dashed line), since the least skilled workers face particularly low λ_j^0 (see column (5) of Table D.4). As we emphasized in the main body, and as we see here, the purely redistributive fixed effects (due to sorting into firm intercepts with no complementarity) do not in general induce skewness of the earnings distribution.

E Extensions and Robustness

E.1 Industries and Occupations

Using occupation and industry identifiers we can assess whether return heterogeneity is genuinely firm-specific. To this purpose we add industry and occupation interactions with cognitive and noncognitive skills to the specification (10). That is, $\boldsymbol{X}_{it}\boldsymbol{b}_t$ now contains $\lambda_o^c \cdot c + \lambda_o^n \cdot n$ as additional controls where each o indexes one industry or occupation cell.

Table E.1 reports the results, with the first column referring to the baseline specification from the main text for comparison. In column (2) we add industry-specific cognitive and noncognitive skill returns (for 19 different sectors). The contributions of firm intercepts and of returns heterogeneity to earnings dispersion decline very slightly – from 0.10 to 0.09 for sd(λ_j^0) and from 0.06 to 0.05 for sd($\lambda_j^c c_i + \lambda_j^n n_i$). The overall effects remain similar. Column (3) adds detailed five-digit industries, with up to 586 separate returns for each skill dimension; also in this case, the contributions of firm-level parameters to overall dispersion remain stable.

Occupation information is only available in the LISA data from 2001 onward (and only partially before then) so that the estimation sample shrinks. This can be seen, e.g., in the lower number of unique firms in the bottom row of Table E.1.

Introducing occupation-specific returns has more influence on the firm-level parameters. In column (4) of Table E.1 we allow for heterogeneous returns for eight major occupation groups (similar to those used in Acemoglu and Autor, 2011). In this specification the standard deviations of baseline cognitive and noncognitive returns, as well as their contributions to earnings dispersion, decline by about one third compared to the benchmark in column (1). This partly reflects variation in production arrangements within firms; to the extent this variation underpins firm-specific skill returns, it is natural to expect it to be captured by occupation-specific returns. Put differently, the firm-level occupation make-up is one of the primitives accounting for firm heterogeneity in skill returns and, therefore, is a legitimate component of the total firm return. Finer occupations in column (5) and even industry-sector $\times$ occupation-group interactions in col-

Table E.1: Dispersion of estimated effects under industry / occupation controls.

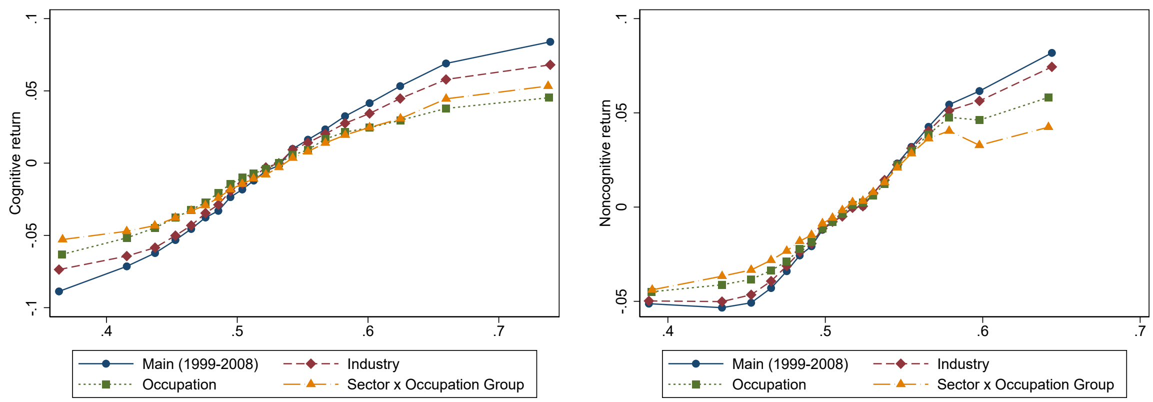
	Main (1)	Sector (2)	Industry (3)	Occup-Group (4)	Occupation (5)	Sec×OccGr (6)
sd(μ_i)	0.43	0.43	0.43	0.41	0.40	0.40
sd(λ_j^0)	0.10	0.09	0.09	0.10	0.09	0.09
sd(λ_j^c)	0.08	0.08	0.07	0.05	0.05	0.05
sd(λ_j^n)	0.05	0.05	0.04	0.04	0.04	0.04
sd($\lambda_j^c c_i$)	0.05	0.05	0.04	0.03	0.03	0.03
sd($\lambda_j^n n_i$)	0.03	0.03	0.03	0.03	0.02	0.02
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.06	0.05	0.05	0.04	0.04	0.04
# unique firms	25,783	25,783	25,783	23,999	24,168	23,973

Notes: Parallel to Tables 1 and 4, this table shows standard deviations of worker and firm effects but controlling for industry- or occupation-specific skill returns in equation (10). Column (1) repeats our specification from the main text without such controls. Column (2) adds broad industry sector specific skill returns (19 unique values per skill dimension). Column (3) adds detailed industry specific skill returns (up to 586 unique values per skill dimension). Column (4) adds broad occupation group specific skill returns (8 values, these groups can be seen in Figure E.3). Column (5) adds detailed occupation specific skill returns (113 values). Column (6) adds industry-sector × occupation-group specific skill returns (152 values). Group-level estimates in period: 1999–2008.

umn (6), which proxy for specific jobs in a firm, have little additional effect on the contribution of firm heterogeneity to earnings dispersion.[9]

Sorting patterns. Figures E.1 and E.2 show that the patterns of skill sorting across returns are effectively unchanged when we control for industry and occupation-specific interactions. The

[9]While results would not be much different than the industry-sector × occupation-group specification, we refrain from explicitly reporting estimates of detailed industry × detailed occupation-specific returns estimates. The reason is that this has additionally more than 21 thousand nonmissing cell-specific returns (almost as many as there are firms) for each skill dimension and thus reinstates an incidental parameter bias problem that the group-level estimation shown here circumvents.



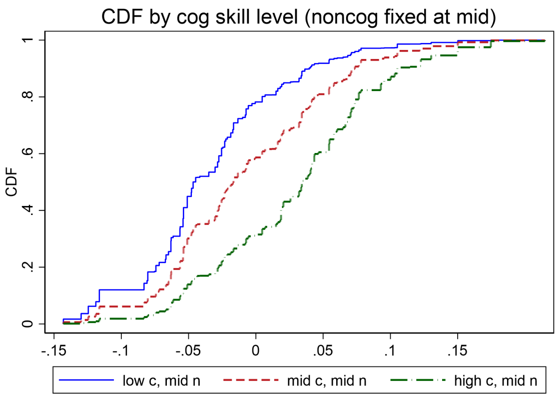
(a) λ_j^c rises in $\bar{c}_j$ (b) λ_j^n rises in $\bar{n}_j$

Figure E.1: Average skill by estimated return under different industry / occupation controls.

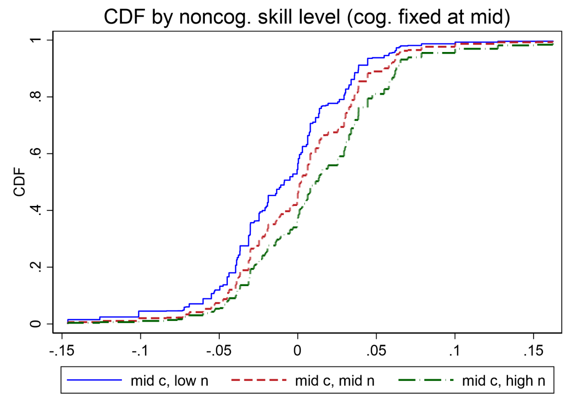
Notes: Parallel to Figure 3, the figure plots binned scatterplots of firm-specific skill returns (vertical axis) with average skills (horizontal axis) for the baseline specification shown in the main text; additionally controlling for detailed industry specific skill returns (up to 586 unique values per skill dimension) in equation (10); controlling for detailed occupation specific skill returns (113 values); and for industry-sector × occupation-group specific skill returns (152 values). Group-level estimates in period: 1999–2008.

range of variation of firm-level returns is only slightly smaller, in line with the reduction of dispersion in Table E.1. Sorting across firms remains strong and remarkably robust over the skill range.

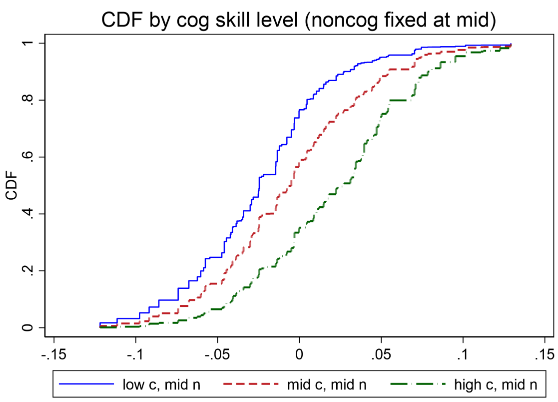
We conclude that firm-level differences are an important source of skill return heterogeneity. Accounting for industry and occupation heterogeneity provides further evidence of the large differences that persist at the firm level; these differences do not reflect purely sectoral or occupational variation. Rather, we find that even within the same narrow industries and occupations, skills command significantly different returns across employers.



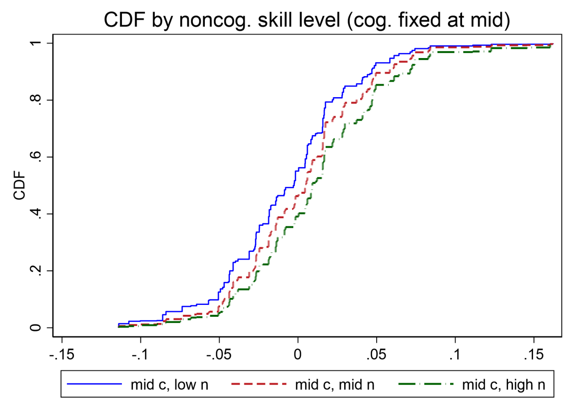
(a) λ_j^c, detailed industry



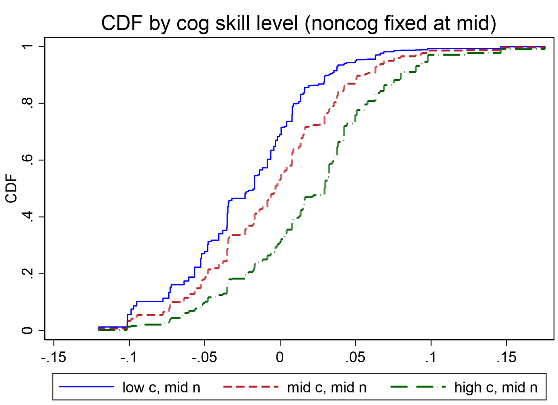
(b) λ_j^n, detailed industry



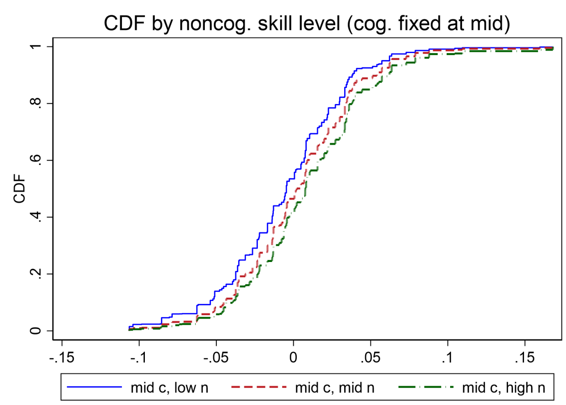
(c) λ_j^c, detailed occupation



(d) λ_j^n, detailed occupation



(e) λ_j^c, industry-sector $\times$ occupation-group



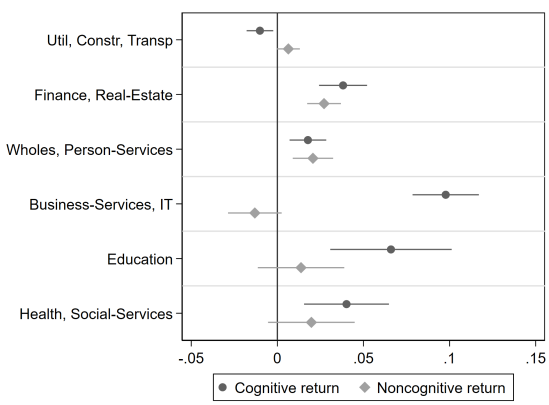
(f) λ_j^n, industry-sector $\times$ occupation-group

Figure E.2: FOSD sorting under different industry / occupation controls.

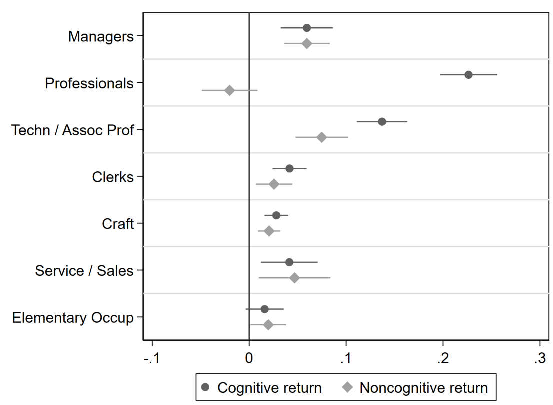
Notes: Parallel to main text Figure 3, this figure shows the cumulative distribution functions for workers with different skill ranks over the range of firm returns. The returns are now estimated controlling for industry- or occupation-specific skill returns in equation (10). The top row controls for detailed industry specific skill returns (up to 586 unique values per skill dimension); the middle row for detailed occupation specific skill returns (113 values); and the bottom row for industry-sector $\times$ occupation-group specific skill returns (152 values). FOSD: first-order stochastic dominance. Group-level estimates in period: 1999–2008.

Aggregating returns to the industry and occupation level. Whereas most of the heterogeneity occurs at the firm level, one may ask which industries or occupations exhibit higher skill returns on average. To answer this question, we first consider linear projections of baseline estimates of λ_j^{c} and λ_j^{n} on a full set of seven industry sector dummies. The projections are similar to those described in (12), where $\overline{c}_j$ and $\overline{n}_j$ are replaced by sector dummies, and yield the average cognitive and noncognitive return in the respective industry compared to the omitted "Manufacturing" sector.

Figure E.3a summarizes the results for the group-level estimates in the form of a coefficients plot. Cognitive returns are especially high in the business services and IT sector, noncognitive returns tend to be higher in wholesale and personal service related activities. By contrast, cognitive returns are rather low in the omitted manufacturing sector itself (represented by the zero line) and in utilities, transport, and services. Noncognitive returns in addition are remarkably low in business services and IT.



(a) Industry sectors



(b) Occupation shares

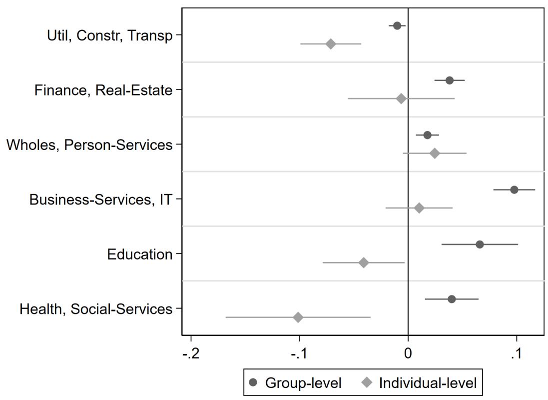
Figure E.3: Skill returns by industry and occupation composition.

Notes. Panel (a): coefficients from the projection of skill returns λ_j^{c} and λ_j^{n} onto seven broad industry-sector dummies. Sector dummies add up to one and the omitted sector is "Manufacturing", i.e., coefficients indicate difference in average skill return compared to average in manufacturing ($\lambda_j^{\mathrm{c}} = -0.029$ and $\lambda_j^{\mathrm{n}} = -0.004$ in that sector). Panel (b): coefficients from the projection of λ_j^{c} and λ_j^{n} onto a full set of eight broad occupation employment shares in each firm. Occupation group shares sum to one and the omitted group is "Operators / Assemblers". Returns are estimated for 100 firm classes. 95% confidence intervals based on robust standard errors clustered at the level of firm classes.

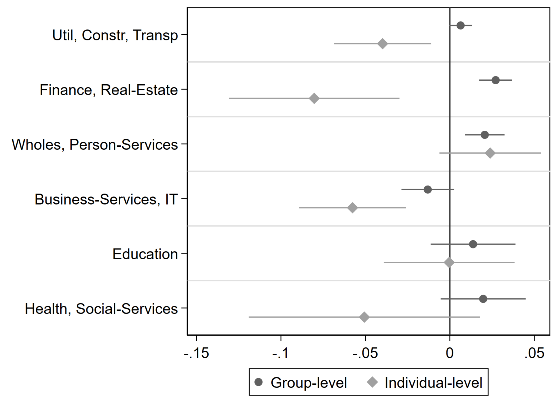
Figure E.3b shows corresponding results from a linear projection of estimates of λ_j^c and λ_j^n onto an exhaustive set of employment shares for eight broad occupation groups. The baseline omitted occupation are "Operators / Assemblers", a large manufacturing-type occupation group. As in Table E.1, and likely because they can vary within firms, occupations are somewhat more related to cognitive and noncognitive returns. That is, firms with large shares of professional, technical, and clerical workers have significantly higher cognitive returns compared to operator/assembler workers. Firms with larger shares of managers, technical workers, and services/sales workers have both high cognitive and noncognitive returns. As for business services and the IT sector, noncognitive skill returns are low among firms with a high share of professional workers.[10]

Finally, results are robust to alternatively considering the firm-level estimates of λ_j^c and λ_j^n from the (smaller) leave-match-out sample. This is shown in Figure E.4, next to the group-level estimates. This approach is less precise and has wider confidence intervals but remains broadly consistent with the group level projections. These exercises suggest that firms in certain industries, and with certain occupations, differentially reward particular skills. Yet, while such variation exists, skill returns (even conditional on, say, a given occupation) vary substantially across firms. In fact, occupation composition can itself be an outcome partly driven by return differences across firms.

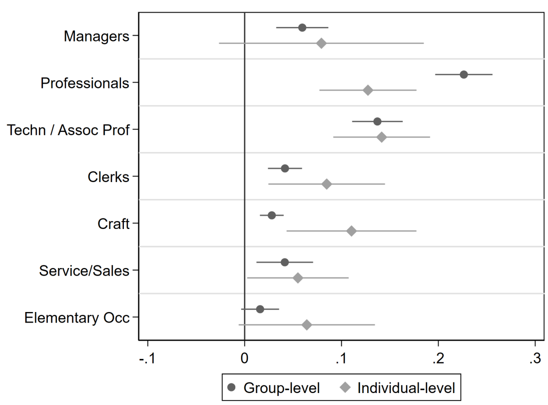
[10]These low noncognitive returns are consistent with the cross-sorting we found in Section 4 if the very high cognitive returns attract very cognitively able professionals to those firms. The professionals also have high noncognitive skills but they accept the low noncognitive returns in the "business / professional services" firms in exchange for the exceptional returns on their cognitives.



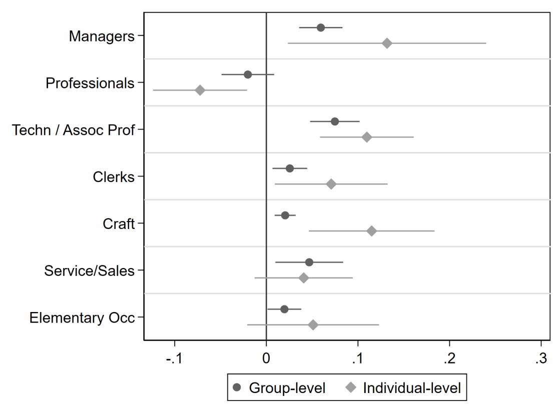
(a) Cognitive λ_j^c, industry sectors



(b) Noncognitive λ_j^n, industry sectors



(c) Cognitive λ_j^c, occupation shares



(d) Noncognitive λ_j^n, occupation shares

Figure E.4: Skill returns by industry and occupation, firm- versus group-level estimates.

Notes. See note to Figure E.3. Here we additionally plot the projections of firm-level λ_j^c and λ_j^n estimates onto broad industry sector dummies and occupational employment shares, and then compare them to the projections of group-level estimates from Figure E.3 separately by skill dimension.

E.2 Capital Composition, Innovation, and Skill Returns

Balance sheets and capital components. We use a commercial data product, the "Serrano" database provided by Bisnode AB, that collects and cleans information about each firm's financials. Up to now, and consistent with prior work, we have referred to workplaces as "firms". However, for the balance sheet analysis we aggregate workplaces up to the organization level (a broader notion of "firms") for which both financial accounts and innovation activity are reported. Since there are multi-workplace corporations, this reduces the number of observations by about one third (see, e.g., Table 7).

Table E.2: Projection of Group Returns onto Firm Capital Composition.

	Dependent variable: λ_j^c or λ_j^n $\times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-3.69		-1.06		0.14	
	(0.87)		(0.15)		(0.09)	
Buildings, Land, Machinery		-2.86		-0.80		0.07
		(0.81)		(0.21)		(0.10)
Other tangible assets		-2.40		0.25		0.12
		(0.35)		(0.10)		(0.05)
Intangible assets	2.95		0.97		0.03	
	(0.36)		(0.10)		(0.09)	
Patents, licences, capt. R&D		3.49		0.64		-0.24
		(0.44)		(0.12)		(0.10)
Goodwill and other intangibles		1.57		0.53		0.19
		(0.27)		(0.13)		(0.06)
Number of firms	14,339	14,339	5,496	862	14,339	14,339
Dependent variable	λ_j^c	λ_j^c	λ_j^c	λ_j^c	$\boldsymbol{\lambda_j^n}$	$\boldsymbol{\lambda_j^n}$
Independent variables as	**dummy**	**dummy**	**logs**	**logs**	arcsinh	arcsinh

Notes: Results from regressions of skill returns onto capital components per employee from firms' balance sheets. Observations (firms) are unweighted with no further control variables. Columns (1) and (2) use dummies for whether the firm reports a positive value of the respective capital item as opposed to zero. Columns (3) and (4) take logs of the items' values. Columns (5) and (6) use noncognitive instead of the cognitive return as dependent variable with independent variables in arcsinh as in the main text Table 7. Dependent variables λ_j^c, λ_j^n multiplied by 100. Grouped estimates in period: 1999–2008. Robust standard errors clustered at the level of the 100 firm groups.

Table E.2 shows the projections of skill returns estimated at the group level onto firm capital components per employee in various robustness specifications. First we employ alternatives to the arcsinh transformation of balance sheet items in the main text. Columns (1) and (2) use dummies, which take the value of one when a firm reports a positive value of the respective capital item as opposed to zero (missing values are still removed). We observe that tangible assets, and in particular physical capital, is significantly negatively associated with cognitive returns whereas intangible assets, and especially intellectual capital, is significantly positively related. As a flip-side of this "extensive margin", we also study the "intensive margin" where we use log transformations of the balance sheet items. As discussed, the number of non-missing observations now drops and especially so for the detailed distinctions within tangible and intangible assets in column (4). Nonetheless, qualitatively and statistically (as well as in terms of coefficient sizes) the results are comparable to those based on the arcsinh transformation in the main text.

Table E.3: Projection of Firm-Level Returns onto Firm Capital Composition.

	Dependent variable: $\lambda_j^c \times 100$ from firm-level estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Tangible assets	-0.45	-0.56	-0.88			
	(0.36)	(0.41)	(0.19)			
Buildings, Land, Machinery				-0.32	-0.45	-0.59
				(0.44)	(0.46)	(0.25)
Other tangible assets				-0.34	-0.38	-0.55
				(0.40)	(0.41)	(0.22)
Intangible assets	1.03	0.76	0.57			
	(0.32)	(0.33)	(0.16)			
Patents, licences, capt. R&D				1.33	1.03	0.75
				(0.46)	(0.47)	(0.21)
Goodwill and other intangibles				0.46	0.33	0.37
				(0.38)	(0.38)	(0.19)
Number of firms	10,258	10,258	10,258	10,258	10,258	10,258
Sector fixed effects	No	Yes	No	No	Yes	No
Employment weighted	No	No	Yes	No	No	Yes

Notes: Firm-level estimates of λ_j^c in period 1999–2008. Robust standard errors in parentheses. Other than that, see note to Table 7.

Table E.3 shows that the projection results onto capital components are remarkably robust even if we instead use the firm-level estimates of cognitive returns. Finally, we note that the relationships with noncognitive returns are weaker, as shown in columns (5) and (6) of Table E.2. If anything, patents, licences, and capitalized R&D appear slightly negatively related to noncognitive returns (goodwill and other intangibles positively). Overall, these results are consistent with firms exhibiting heterogeneous production arrangements, whereby capital and employment structure vary substantially and lead to different returns to skill attributes, with the stronger impacts holding in the cognitive skill dimension.

Innovation output. In what follows we conduct further tests on the data from the Community Innovation Survey (CIS). Similar to the preceding analyses, the first part of Table E.4 shows estimates from projecting cognitive skill returns onto both product and process innovations. As in the main text, we control for a quadratic in employment, since the probability of engaging in innovation rises with the firm's size. The results on product innovations remain strong, whether or not we use group-level (column 1) or firm-level (column 3) return estimates or we control for industry sector fixed effects (i.e., the 19 unique ones from Section E.1).

The relationship between skill returns and process innovations gets weaker when we condition on product innovations, and it is only borderline significant.

Specific innovation activities. Next, we examine firms' CIS-reported expenditures on specific types of innovation activities. This is, again, done by using the arcsinh transformation. Column (4) of Table E.4 shows that, consistent with the preceding findings, high cognitive returns firms spend significantly more on intramural (or in-house) research and development. They also spend somewhat more on purchasing external knowledge, and somewhat less on specific machinery. These findings are robust to adding industry sector fixed effects or using estimates of firm-level returns in columns (5) and (6).

Lastly, Figure E.5 shows the baseline binned-scatter plot of skill returns vis-a-vis product and process innovations for the noncognitives. Broadly in line with our prior findings, there is no

Table E.4: Projection of Skill Returns onto Firm Innovation Activities.

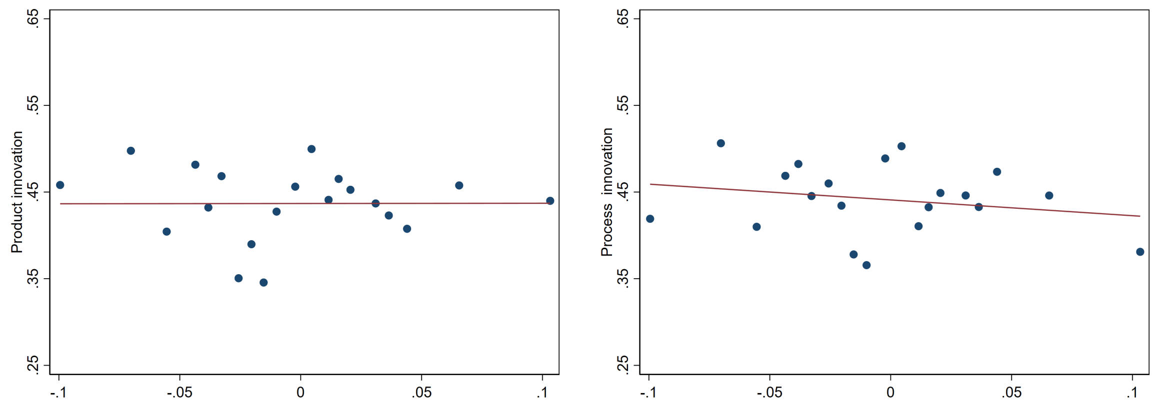
	Dependent variable: $\lambda_j^c \times 100$					
	(1)	(2)	(3)	(4)	(5)	(6)
Innovation output:						
Product innovation	3.73	2.89	7.77			
	(0.53)	(0.37)	(2.61)			
Process innovation	0.29	0.48	4.09			
	(0.30)	(0.25)	(2.73)			
Innovation spending:						
Intramural R&D				0.39	0.30	0.46
				(0.06)	(0.04)	(0.27)
Extramural R&D				-0.01	0.06	0.03
				(0.04)	(0.03)	(0.32)
Acquisition of machinery				-0.14	-0.02	0.05
				(0.04)	(0.03)	(0.27)
Other external knowledge				0.14	0.06	0.37
				(0.04)	(0.03)	(0.31)
Number of firms	4,138	4,138	3,344	3,857	3,857	3,123
Sector fixed effects	No	Yes	No	No	Yes	No
Estimates (level)	Group	Group	Firm	Group	Group	Firm

Notes: The first three columns report estimates from regressions of cognitive skill returns onto indicators for product and process innovations (as defined in the text and note to Figure 5), controlling for a quadratic in firm employment size. Column (1) uses group-level returns estimates, column (2) adds industry sector fixed effects, and column (3) uses firm-level returns estimates. The last three columns regress returns onto firms' innovation expenditure items, which are $\text{arcsinh}(x_j) = \log\left(x_j + \sqrt{x_j^2 + 1}\right)$ transformed. Otherwise specifications (4)–(6) are parallel to (1)–(3). Returns estimated in period 1999–2008. Robust standard errors in parentheses and clustered at the level of the 100 firm classes for the grouped estimates.

detectable relationship and, in contrast to λ_j^cs, the λ_j^ns do not actually predict higher innovation activity.

E.3 Clustering Strategies and Number of Firm Clusters

In what follows, we document how the dispersion (standard deviations) of firm-level parameters, and their contributions to earnings dispersion, vary under alternative restrictions on the number of clusters as well as on the observables used for the clustering.



(a) λ_j^{n} and product innovation

(b) λ_j^{n} and process innovation

Figure E.5: Noncognitive skill returns and firm innovation.

Notes: The figure plots a binscatter of firms' innovation activities against noncognitive skill returns (group-level estimates during 1999–2008). Innovation activities are measured as indicators whether a firm has conducted any product (including service, Panel a) or process (including organizational, Panel b) innovations. This information is from various waves of a representative firm survey (European Community Innovation Survey, CIS). We average the responses (i.e., indicators) for the waves 1998–2000, 2002–2004, 2004–2006, 2006–2008, 2008–2010 relevant to our sample period. Underlying the plots are 4,138 unique firms. Regression slopes, controlling for a quadratic in firm employment, are $\beta = 0.00$ (clustered S.E. $= 0.32$) and $\beta = -0.18$ (clustered S.E. $= 0.19$) for product and process innovation, respectively.

Table E.5 illustrates the key results. For comparison column (1) replicates the baseline specification in the main text. In column (2) we use only 10 (rather than the baseline 100) firm clusters; this number is the same as in the main analyses of Bonhomme et al. (2019); Lamadon et al. (2022). The contributions of firm intercepts and skill return heterogeneity to earnings dispersion marginally declines, while the *relative* contribution of returns rises. When we use a richer set of clustering variables – including firm employment size as well as the standard deviations of earnings and cognitive and noncognitive skills, in addition to the means of these variables – returns heterogeneity does not change significantly either in relative or absolute terms, as shown in column (3). Finally, column (4) shows that restricting the clustering strategy to earnings alone, through the use of quantiles of the earnings distribution within each firm (see Bonhomme et al., 2019; Lamadon et al., 2022), does not materially change the estimated impact of returns heterogeneity when compared to the other robustness checks. In all alternative specifications we also confirm the presence of positive assortative matching patterns.

Table E.5: Alternative clustering specifications.

	Main (1)	10 clusters only (2)	Adding variables (3)	Earning dist. only (4)
sd(μ_i)	0.43	0.44	0.43	0.43
sd(λ_j^0)	0.10	0.07	0.10	0.11
sd(λ_j^c)	0.08	0.06	0.07	0.06
sd(λ_j^n)	0.05	0.03	0.04	0.03
sd($\lambda_j^c c_i$)	0.05	0.04	0.04	0.04
sd($\lambda_j^n n_i$)	0.03	0.02	0.02	0.02
sd($\lambda_j^c c_i + \lambda_j^n n_i$)	0.06	0.04	0.05	0.05
# unique firms	25,783	25,783	25,711	25,783

Notes: Adding to the evidence in Tables 1 and 4, this table shows standard deviations of worker and firm effects under alternative clustering specifications. Column (1) repeats the baseline specification from the main text, for comparison. Column (2) shows estimates when using the means of earnings, cognitive and noncognitive skills within each firm but just ten clusters. Column (3) shows results for 100 clusters after adding standard deviations of earnings, cognitive and noncognitive skills and firm employment size as additional clustering variables. Column (4) shows results when we only use quantiles of the earnings distribution (10th, 30th, 50th, 70th, and 90th) within each firm for clustering and we impose 100 groups. All group-level estimates are based on the sample period: 1999–2008.

In the main text, Figure 6 shows estimates of the impact of firm heterogeneity for a range alternative restrictions on the number of clusters (which are let to vary between 20 and 200). The relative contributions of intercepts and skill returns change only marginally, lending further support to the results obtained under the baseline cluster design.

References

ACEMOGLU, D. AND D. AUTOR (2011): "Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 4, Part B, 1043 – 1171.

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 (2017): "Living the American Dream in Finland: The Social Mobility of Inventors," Mimeo.

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.

BAGGER, J. AND R. LENTZ (2019): "An empirical model of wage dispersion with sorting," *The Review of Economic Studies*, 86, 153–190.

BARTOLUCCI, C., F. DEVICIENTI, AND I. MONZÓN (2018): "Identifying sorting in practice," *American Economic Journal: Applied Economics*, 10, 408–38.

BÖHM, M., D. METZGER, AND P. STRÖMBERG (2022): "Since you're so rich, you must be really smart: Talent, rent sharing, and the finance wage premium," Tech. rep., ECONtribute Discussion Paper.

BONHOMME, S., K. HOLZHEU, T. LAMADON, E. MANRESA, M. MOGSTAD, AND B. SETZLER (2020): "How Much Should we Trust Estimates of Firm Effects and Worker Sorting?" Tech. rep., National Bureau of Economic Research.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2017): "Discretizing unobserved heterogeneity," *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

——— (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

BONHOMME, S. AND E. MANRESA (2015): "Grouped patterns of heterogeneity in panel data," *Econometrica*, 83, 1147–1184.

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.

EDIN, P.-A. AND P. FREDRIKSSON (2000): "Linda-longitudinal individual data for sweden," Tech. rep., Working Paper, Department of Economics, Uppsala University.

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

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying equilibrium models of labor market sorting," *Econometrica*, 85, 29–65.

Hansen, K. Y., J. J. Heckman, and K. J. Mullen (2004): "The effects of schooling and ability on achievement test scores," *Journal of Econometrics*, 121, 39–98.

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.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

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.

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.

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