Firm Heterogeneity in Skill Returns

Michael J. Böhm
Khalil Esmkhani
Giovanni Gallipoli*

THIS DRAFT: SEPTEMBER 6, 2020

Abstract

We combine matched employer-employee data and population records on cognitive and non-cognitive skills to show that identical skill endowments command very different returns across firms. This has implications for the assignment of workers with multidimensional skill sets to employers. Workers benefit from sorting into firms that reward their attributes but skill indivisibility prevents them from separately renting out their endowments to the highest return employer. We estimate firm-specific returns to each skill attribute and derive analytical results linking the Jacobian of the matching function to the covariance matrix of the within-firm distribution of skills. We use these results to examine the mechanism through which changes in multidimensional sorting patterns influence the distribution of wages.

Keywords: Firm Heterogeneity, Skills, Sorting, Bundling, Wages, Inequality
JEL Classification: D3, J23, J24, J31

*Böhm: University of Bonn (michael.j.boehm@uni-bonn.de); Esmkhani: UBC, Sauder School of Business (khalil.esmkhani@sauder.ubc.ca); Gallipoli: UBC, Vancouver School of Economics (gallipol@mail.ubc.ca). We are grateful to Chris Boehm, Matilde Bombardini, Matias Cortes, Gino Gancia, Georg Graetz, David Green, Magne Mogstad, Thomas Lemieux, and Dominic Rohner, as well as to many workshop participants, for helpful comments. We thank Shuahib Habib, Ronit Mukherji and Dongxiao Zhang for research assistance. Böhm acknowledges support from a research fellowship of the German Science Foundation (BO 4765/1-2). Gallipoli acknowledges support from Canada's SSHRC. An earlier draft was circulated with the title 'Firm Heterogeneity and the Return to Cognitive Skills'.

1 Introduction

The recognition that earnings distributions reflect both worker and firm heterogeneity dates back decades. In his overview of earnings functions, Robert Willis (1986) highlighted the presence of "an imbalance in the human capital literature which has emphasized the supply far more than the demand for human capital".

The growing availability of matched employer-employee records has led to a broader examination of firm-level differences.[1] A workhorse of the applied literature is the two-way fixed-effect model popularized by Abowd et al. (1999). This approach subsumes unobserved heterogeneity of workers and firms into additively separable measures whose contribution to the variance of earnings can be transparently quantified. In this context, the covariation between firm and worker fixed effects is often interpreted as evidence of non-random sorting of workers across employers, or lack thereof. An influential body of work on matching in the labor market,[2] however, cautions against drawing inference about match-specific productivity from fixed effect estimates. This follows from the observation that non-linear, and possibly non-monotonic, patterns of complementarity are hard to characterize within the boundaries of additively separable models of worker and firm heterogeneity.[3] These considerations inform richer empirical frameworks (e.g., Bonhomme et al., 2018; Lentz et al., 2018) that nest non-linear matching mechanisms within settings with two-sided unobserved heterogeneity. To preserve tractability these approaches resort to dimension reduction techniques based on grouping.

Across the range of methods and data sources, most studies lend support to the hypothesis that genuine firm-level earnings variation occurs above and beyond what is captured by industry and occupation heterogeneity. When subsumed within employer fixed or random effects, this variation reflects firm-specific premia for otherwise identical workers as well as unobserved differences in work force composition. Therefore the interpretation of evidence on firm-level va-

[1]E.g., Card et al. (2013), Card et al. (2016), Song et al. (2018), Lamadon et al. (2019).

[2]See Shimer and Smith (2000); Eeckhout and Kircher (2011).

[3]For this reason, several studies (e.g., Hagedorn et al., 2017; de Melo, 2018; Jarosch et al., 2019) pursue a different route and posit explicit structural assumptions to identify the productivity gains associated to specific matching patterns, while others highlight the role of amenities and non-pecuniary returns for matching (Sorkin, 2018).

riation depends on how one accounts for worker-level heterogeneity. A common approach is to rely on realized earnings to draw inference about worker ability; that is, ex-ante worker differences are teased out from ex-post realizations of labor market returns. By design, this implies that distinct dimensions of worker heterogeneity are lumped together, conflating returns to, and endowments of, different individual characteristics.[4]

In this paper we link two direct, ex-ante measures of worker skills to administrative earnings data from Sweden, matching employees to their employers. The skill measures summarize cognitive and non-cognitive abilities that are tightly linked to individual labor market outcomes, even over decade-long intervals. Our analysis reveals the presence of significant skill sorting across firms and, perhaps more importantly, suggests that worker sorting is associated to pronounced firm-level heterogeneity in returns to identical worker skills. That is, we document how similar skill bundles command very different returns across firms. This observation carries non-trivial implications since workers cannot rent out their different skill endowments to different employers; this basic indivisibility problem was originally examined in theoretical work by Mandelbrot (1962) and later revisited by Heckman and Scheinkman (1987). The latter paper also rejects the null hypothesis of homogeneous pricing of individual skills in PSID data.

We show that ex-ante measures of idiosyncratic ability are not only a helpful complement to account for heterogeneity in returns to identical bundles of skills but can be used to identify the firm-specific returns to each separate attribute. To this purpose, we develop a tractable two-step procedure that recovers estimates of the heterogeneous firm-level skill returns. Then, we use our estimates to characterize the sorting of workers to firms in a setting in which multi-dimensional heterogeneity implies varying degrees of worker-firm complementarity. This exercise is relevant in light of the growing interest in multidimensional sorting and mismatch in equilibrium assignment models (see Lindenlaub, 2017; Lindenlaub and Postel-Vinay, 2017; Guvenen et al., 2020).

Figure 1 illustrates some stark features of the distribution of skills in the cross-section of worker-firm matches. The left panel of Figure 1 shows a scatter of the averages of cognitive

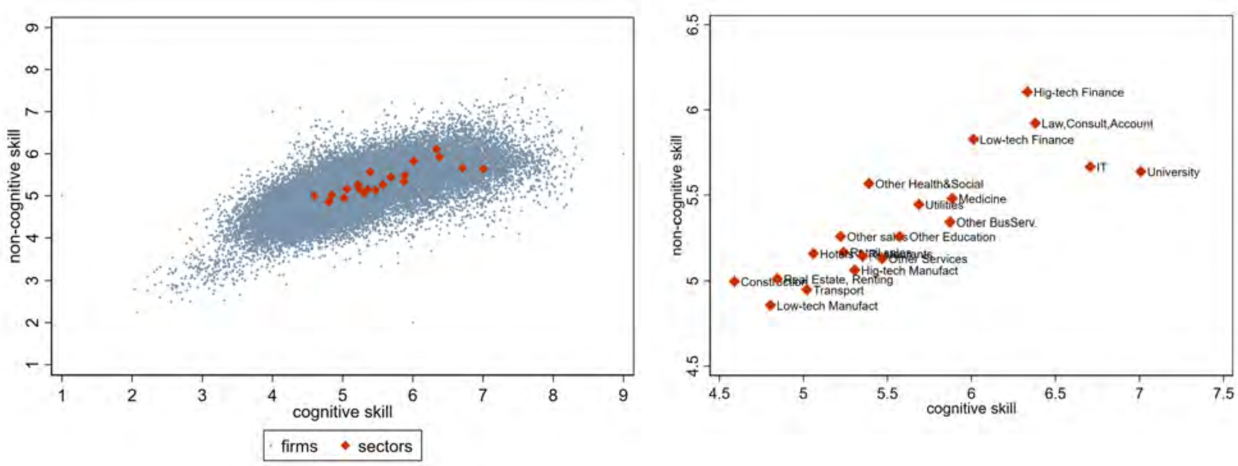
[4]Similar considerations affect the interpretation of estimates of the returns to schooling in the absence of direct proxies for ability. These have been the object of much debate for decades (see Griliches, 1977).

and non-cognitive worker skills across firms and industry sectors. It is apparent that: (i) the variation in skill intensity across firms is much wider than what is measured between two-digit industry sectors;[5] (ii) while cognitive or non-cognitive skill intensities are positively correlated, large differences occur in their relative composition.

Figure 1: Firms' Skill Intensities

(a) Firms

(b) Industry sectors



Notes: The figure plots average cognitive (x-axis) and non-cognitive (y-axis) skills of workers (males aged 20–60) in firms and 19 non-primary and non-public broad industry sectors. Both dimensions of skills are coded as Stanine, i.e., integers 1(lowest) to 9 (highest) and approximating the normal distribution. The pairwise correlation coefficient between cognitive and noncognitive skills in the underlying population of workers is 0.36. We condition on firms with at least on average 10 workers which exist for five years or more (31,613 unique firms during 1990–2017) to minimize idiosyncratic fluctuations in these statistics.

The right panel of Figure 1 zooms into the scatter plot, providing a coarser view of average skill intensity across industries. This highlights the presence of a few skill intensive sectors (i.e., where firms hire workers with generally higher skills), such as IT, law, consulting and accounting professions, as well as sectors with much lower average skills, such as construction or low-tech manufacturing. Sectors also exhibit variation in relative skill composition: for example, IT, university education, law and consulting are more intensive in cognitive ability.

Large differences in skill intensity and composition across firms suggest the presence of systematic selection and we examine this selection through the lens of a model with different layers

[5]To mitigate measurement error, we use firms with 10 or more employees which exist for at least five years. In the Appendix we show this variation also conditioning on detailed industry cells.

of heterogeneity in production arrangements. In Section 2 we report estimates of the distribution of firm-level wage premia for different skill bundles. These estimates show that the premia for skill bundles across firms are very heterogeneous and that workers sort into firms according to these premia.

These results, which rely on minimal assumptions, have two limitations: (i) they conflate in one skill premium both firm-specific return and worker skill endowment; (ii) they offer no information about marginal returns to a specific attribute (i.e. cognitive vs non-cognitive traits). In Section 4 we suggest a method-of-moments approach to unbundle the returns to different worker attributes. This is necessary to analyze the extent of empirical sorting on either trait. Using the notion of first order stochastic dominance we document how workers with different characteristics sort across firms with different returns. The latter exercise sheds light on the factors driving wage dispersion and the trade-offs faced by individual workers when matching with different employers.

Our estimates of firm-specific returns to individual worker attributes provide insights into the nature of firm heterogeneity, moving beyond a characterization where all heterogeneity is collapsed into an employer fixed effect. This allows us to draw attention to the growing polarization of skill returns across firms. Specifically, we highlight a rising covariation between cognitive and non-cognitive rewards in the cross-section of firms between 1990 and 2017. This, in conjunction with higher returns to non-cognitive skills, accounts for much of the convexification of the wage-skill distribution observed in the same period.

2 Stylized Facts about Wage Premia

2.1 Data

We use matched employer-employee data from Sweden between 1990 and 2017, including annual earnings from labor, employer, occupation and industry, and standard worker characteristics such as age, gender, and education. This information is directly linked to military enlistment tests, including measures of cognitive skills and assessments of non-cognitive traits based on interviews with trained psychologists. Prior research has shown that these test scores are highly significant at predicting earnings and other labor market outcomes, both on their own and conditionally on each other or any rich set of control variables (e.g., Lindqvist and Vestman, 2011; Fredriksson et al., 2018). Cognitive and non-cognitive measures are recorded on a Stanine (standard nine) scale which approximates the Normal distribution[6] and facilitates comparisons across birth cohorts.

We restrict the worker sample to males, for whom the skill measures are always available, aged 20–60. We also restrict attention to firms which exist for at least five years and employ at least ten such male workers on average. Our dataset reports both organization and workplace (i.e., organization by geographic location and industry) identifiers. We use workplaces as our "firms" as this is closest to the theoretical notion of a production unit and is consistent with existing approaches in the literature (e.g., Card et al., 2013). The resulting sample during 1990–2017 contains 415 thousand firm-year observations and more than 14 million individual worker-year pairs.

To reduce measurement error and ease interpretation, we granularize test scores for each attribute and divide them into three ranked groups (high, medium or low). Every worker has a bundle of attributes $s = (c,n)$, with the first letter denoting cognitive and the second non-cognitive traits. A skill type s is within the set $S = \{(c,n) | c \in \{L,M,H\}; n \in \{l,m,h\}\}$. The cognitive ranks are $\{L,M,H\}$ for high, medium and low, while $\{l,m,h\}$ is the rank set of non-cognitive attributes.[7] In summary, there are nine skill types, one for each combination of cognitive and non-cognitive ranks. Details about data and sample construction are in the appendix.

[6]Measures are standardized for each birth year in the population. A score of 5 is reserved for the middle 20 percentiles of the population taking the test. The 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.

[7]Low skill ranks correspond to stanine scores 1 to 3. Middle ranks subsume stanine 4 to 6. The high ranks refer to stanine scores 7 to 9.

2.2 Non-Parametric Wage Premia

Under the null hypothesis that firms have idiosyncratic skill yields, pecuniary labor market returns entail two firm-specific components: (i) a base wage common to all workers in a firm, irrespective of their attributes; (ii) a skill bundle premium. This generalizes the Abowd et al. (1999, AKM) specification, as it features worker-firm complementarities in addition to firm and worker fixed effects:

$$\ln(w_{ijt}) = \theta_j + \mu_i + \sum_{s_i \in S} \Delta_{js} 1[s_i = \{s\}] + X_{it} b_t + \varepsilon_{ijt}, \tag{1}$$

with skill types in $S = \{(L,m),(L,h),(M,l),(M,m),(M,h),(H,l),(H,m),(H,h)\}$. Wage premia are relative to the (omitted) type $s = (L,l)$, which corresponds to the lowest skill group and the firm's base wage. The controls $X_{it} b_t$ account for observable individual and economy-wide influences, whose effects may vary over time and by age as we fully stratify these variables and their interactions.

Interpreting parameters. For the subset of workers of baseline type $s = (L,l)$ (the omitted skill bundle), equation (1) collapses to a standard AKM specification with firm fixed effects θ_j, time-varying controls including year and age dummies, and a worker fixed effect μ_i. Intercepts θ_j are identified, up to a normalization, from wage changes of (L,l) workers upon switching firms.

For other skill bundles, (1) augments the AKM specification by allowing for firm-specific returns. These premia on the baseline wage are denoted by Δ_{js} with $s \in \{(L,m),(L,h),...\}$. Each premium Δ_{js} is identified from information about job switches by workers with skill bundle s, and by their wage changes *relative* to those of workers with base skill bundle (L,l).

The availability of worker-level measures of multidimensional skills is key for the estimation of heterogeneous returns. While additively separable worker fixed effects flexibly account for unobserved heterogeneity in wage levels, variation in ability measures helps identify the premium associated to observable skill types.[8] In the appendix we report results from event studies that illustrate the mechanics of how skill premia Δ_{js} are identified.

Estimation. Bringing equation (1) to data requires the estimation of a large number of employer-specific parameters from a sample of more than twenty thousand unique firms per period. The results of this massive estimation exercise are reported in the appendix. It is well-known, however, that estimation based on individual firm observations may suffer from measurement error and lead to biases due to limited mobility of workers across employers.[9] To address these issues, Bonhomme et al. (2018) suggest to group firms into a smaller number of bins; this alleviates biases and delivers consistent estimates of the within-bin average of the parameters of interest.

In light of these concerns, we develop a simple grouping estimator.[10] To reduce dimensionality, we rank firms into quintiles according to their relative hiring of each skill bundle. Each quintile corresponds to a categorical variable $g_s \in \{1,2,3,4,5\}$, with s indicating the skill bundle; the value of the categorical variable denotes the rank of the firm in that particular skill bundle intensity. Since employment of workers with the base skills (L,l) grows with the intercept term θ_j, we also group firms into quintiles according to employment relative to the base skills. These intercept groups are denoted by the categorical variable $g = g_\theta$.

The firm-specific parameters from the implementation of the grouping estimator in (1) reduce to $\theta_{g_\theta(j)}$ and $\Delta_{g_s(j)}$. For convenience, we refer to them by their shorthand θ_{js} and Δ_{js} wherever there is no ambiguity. Group assignments for different skill bundles are, by design, flexible and independent of one another. For example, two firms in the same quintile of the (H,h) bundle need not be in the same g_θ or g_s bins for the other skill bundles. This is the least restrictive assumption,

[8]Since s_i is time-invariant, the key restriction for the identification of Δ_{js}, up to a normalization, is that the μ_i do not vary across firms.

[9]Limited mobility may generate serious problems for smaller firms. An incidental parameters problem can introduce an upward bias for the variance of firm and worker fixed effects in AKM-style estimation, and a downward bias in their covariance (Andrews et al., 2008).

[10]This empirical approach is corroborated in the next section using a labor market model with two-sided multidimensional heterogeneity. The model implies that $\ln(\frac{q_{js}}{q_{j(L,l)}})$, the relative hiring of type s relative to the base skill, rises with the firm-specific skill premium Δ_{js}.

since each $\ln(\frac{q_{js}}{q_{j(L,l)}})$ is a sufficient statistic for grouping firms into bins with similar Δ_{js}, and it is consistent with the way we later cast heterogeneity in the model.

In the appendix we experiment with alternative grouping approaches such as clustering based on the k-means algorithm (Bonhomme et al., 2018), where each firm is allocated to one unique bin. These exercises deliver results similar to what we obtain for the benchmark grouping estimator; similar qualitative results hold even if we do not group at all and estimate (1) with fully flexible parameters for each individual firm. In Section 4, we verify that the estimates of Δ_{js} are robustly related to each firm's hiring of the corresponding skill type s.

2.3 Estimates

Table 1 summarizes the benchmark estimates of (1) between 1998 and 2006. The employment-weighted mean of θ_j is normalized to zero. The cross-sectional variation of θ_j around its average reflects differences in base wages (independent of skills) across firms, with a standard deviation of about 4 log points and a 5-95 interquantile range of 15 log points.

All wage premia parameters are relative to the base skill and rise with the skill rank within each bundle, e.g., $\bar{\Delta}_{(L,m)} = 0.09$, $\bar{\Delta}_{(M,m)} = 0.25$, and $\bar{\Delta}_{(H,m)} = 0.30$. This is reassuring and suggests that skill bundles based on cognitive and non-cognitive measures convey genuine information about abilities rewarded in the labor market.

The returns to the highest observable bundle (H,m) are on average 45 log points and they can reach 61 log points or more in firms that reward highly skilled workers the most. However, these estimates are not sufficient to establish whether cognitive or non-cognitive traits are more important for ranking skill returns. For example, while we find that $\bar{\Delta}_{(L,m)} = 0.09$ vs $\bar{\Delta}_{(M,l)} = 0.09$ or $\bar{\Delta}_{(L,h)} = 0.17$ vs $\bar{\Delta}_{(H,l)} = 0.20$, no individual skill return can be uniformly ranked using estimates for skill bundles. In Section 4 we explicitly tackle this question and show that some firms put more weight on cognitive traits while others reward non-cognitive attributes more heavily. By unbundling skills, we also distinguish between endowments and returns and characterize the distributions of firm-specific yields from individual attributes.

Table 1: Firm-specific skill bundle premia: estimates

	(sample period: 1999–2008)				
	mean	sd	p5	p50	p95
θ_j	0.00	0.04	-0.06	-0.00	0.07
$\Delta_{j(L,m)}$	0.10	0.04	0.03	0.09	0.16
$\Delta_{j(L,h)}$	0.17	0.06	0.08	0.17	0.24
$\Delta_{j(M,l)}$	0.09	0.04	0.02	0.07	0.15
$\Delta_{j(H,l)}$	0.19	0.07	0.08	0.19	0.32
$\Delta_{j(M,m)}$	0.26	0.04	0.22	0.23	0.37
$\Delta_{j(M,h)}$	0.28	0.05	0.21	0.28	0.40
$\Delta_{j(H,m)}$	0.30	0.05	0.23	0.30	0.41
$\Delta_{j(H,h)}$	0.46	0.06	0.37	0.45	0.59

Notes: The table shows moments of the distributions of estimated skill-bundle premia across firms. The sample statistics displayed are the mean, standard deviation, and percentiles 5, 50, and 95. The sample covers 25,604 unique firms between years 1999 and 2008.

Table 1 documents the significant dispersion of skill premia across firms, with standard deviations between 4 and 7 log points. The 5–95 interquantile ranges are large, between 13 and 24 log points. Means are close to medians and estimates are sufficiently precise so as to rule out implausible skill premia. For example, down to the 5th percentile of firms, there are no negative premia relative to (L,l) wages.

As we show in the appendix, when estimating (1) with no grouping restrictions, results are marginally different: in this case the 5th percentile skill premia are below the (L,l) baseline. This observation, together with much larger standard errors, indicate significant measurement error when estimating returns for individual (non-grouped) firms. In the appendix we also document that estimates are qualitatively similar for two alternative 9-year estimation periods (1990–1999 and 2008–2017). Skill premia generally rise from the earlier period and there is evidence of growing dispersion between 1998 and 2014. We also present evidence that (i) estimates of the between-firm heterogeneity in skill returns are robust and fairly conservative,[11] and (ii) significant worker sorting is associated to return heterogeneity.

While these findings lend to support to the view that skill premia are heterogeneous in the cross-section of firms, it is hard to draw inference about the returns to each attribute. In particular, one cannot assess if firm-specific marginal returns (or skill endowments) account for estimated premia, and whether there is a pattern in the evolution of cognitive versus non-cognitive yields (e.g., Deming, 2017). The unbundling of skill premia is the object of the next section.

3 A Labor Market with Two-Sided Heterogeneity

To examine the interaction of employer and employee heterogeneity we develop an empirically tractable model where workers have different cognitive and non-cognitive 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, although 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 within a labor market where employers choose how many workers to hire based on the demand for their output. Equilibrium in the labor market is assumed by imposing full employment.

3.1 Production and Market Structure

There is measure one of workers who differ in their cognitive (c) and non-cognitive (n) abilities and we let $G(c,n)$ denote the probability measure describing the distribution of worker types 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 $\boldsymbol{\lambda}$ produces according to $f_{\boldsymbol{\lambda}}(c,n)$, where the function $f_{\boldsymbol{\lambda}}$ describes

[11] In fact, the heterogeneity documented in Table 1 is likely a conservative estimate of the true cross-sectional variation in firm-specific skill bundle premia.

the output from the firm-worker match. Technology is CRS and a firm's output is the sum of all employees' products.[12] Firm $\boldsymbol{\lambda}$'s total output, as a function of its employment, is

$$y_{\boldsymbol{\lambda}}(M) = \int f_{\boldsymbol{\lambda}}(c,n)\, dM_{\boldsymbol{\lambda}}(c,n) \tag{2}$$

where $M_{\boldsymbol{\lambda}}(c,n)$ denotes the measure of different worker types employed by the firm. Worker types are observable and firms face upward sloping labor supply curves for each worker type. If firm $\boldsymbol{\lambda}$ decides to hire a share $q_{\boldsymbol{\lambda}}(c,n)$ of all workers of type (c,n) in the economy, it has to pay to each worker of this type a wage $w_{\boldsymbol{\lambda}}(c,n)$ such that:

$$\log(q_{\boldsymbol{\lambda}}(c,n)) = \log(h(c,n)) + \log(a_{\boldsymbol{\lambda}}) + \beta \log(w_{\boldsymbol{\lambda}}(c,n)) \tag{3}$$

where $a_{\boldsymbol{\lambda}}$ in (3) captures non-pecuniary benefits of working for firm $\boldsymbol{\lambda}$, which shifts the labor supply curve given a labor supply elasticity β.[13] The intercept $h(c,n)$ is an equilibrium outcome of market clearing, defined as

$$h(c,n) = \left[\int a_{\boldsymbol{\lambda}} e^{\beta w_{\boldsymbol{\lambda}}(c,n)}\, dF(\boldsymbol{\lambda})\right]^{-1} \tag{4}$$

where $dF(\boldsymbol{\lambda})$ is the density over the range of possible firm types. In the output market, firms face a downward sloping demand curve for their products. Firm $\boldsymbol{\lambda}$'s inverse demand is

$$\log(p_{\boldsymbol{\lambda}}) = \log(\phi_{\boldsymbol{\lambda}}) - \frac{1}{\sigma}\log(y_{\boldsymbol{\lambda}}) \tag{5}$$

where $p_{\boldsymbol{\lambda}}$ is product price, $y_{\boldsymbol{\lambda}}$ is output, $\phi_{\boldsymbol{\lambda}}$ is a firm-specific (inverse) demand intercept, and σ is the output demand elasticity w.r.t. price.

[12] 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 bundles. While convenient, the separability assumption is not crucial for our findings about sorting and returns heterogeneity.

[13] The labor supply equation can be micro-founded by aggregating workers' idiosyncratic preferences for firms.

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

$$\begin{aligned}
\max_{q_{\boldsymbol{\lambda}}(c,n)} \quad & p_{\boldsymbol{\lambda}} y_{\boldsymbol{\lambda}} - \textstyle\int w_{\boldsymbol{\lambda}}(c,n) q_{\boldsymbol{\lambda}}(c,n) \, dG(c,n) \\
s.t. \quad & y_{\boldsymbol{\lambda}} = \textstyle\int f_{\boldsymbol{\lambda}}(c,n) q_{\boldsymbol{\lambda}}(c,n) \, dG(c,n) \\
& \log(p_{\boldsymbol{\lambda}}) = \log(\phi_{\boldsymbol{\lambda}}) - \tfrac{1}{\sigma} \log(y_{\boldsymbol{\lambda}}) \\
& \log(q_{\boldsymbol{\lambda}}(c,n)) = \log(h(c,n)) + \log(a_{\boldsymbol{\lambda}}) + \beta \log(w_{\boldsymbol{\lambda}}(c,n))
\end{aligned} \tag{6}$$

This problem has a closed form solution, with equilibrium wages in firm $\boldsymbol{\lambda}$

$$w_{\boldsymbol{\lambda}}(c,n) = \frac{\left(\frac{\beta}{1+\beta}\right)^{\frac{\sigma}{\sigma+\beta}} f_{\boldsymbol{\lambda}}(c,n) \left(\frac{\sigma-1}{\sigma}\phi_{\boldsymbol{\lambda}}\right)^{\frac{\sigma}{\sigma+\beta}}}{\left[\int f_{\boldsymbol{\lambda}}(c,n)^{1+\beta} h(c,n) a_{\boldsymbol{\lambda}} \, dG(c,n)\right]^{\frac{1}{\sigma+\beta}}} \tag{7}$$

3.2 The Firm's Wage Intercept and Skill Premia

Firms' production choices can be characterized along the two input dimensions (cognitive and non-cognitive). In view of the ordinal nature of the empirical skill measures used in the descriptive data analysis, we categorize skill bundles again by assigning one of three levels (high, medium, or low) to each ability endowment. Every worker has a type within the set $S = \{(c,n) | c \in \{L,M,H\}; n \in \{l,m,h\}\}$, with the first letter denoting cognitive level and the second non-cognitive level. For example, a worker of type $s = (H,l)$ has high cognitive and low non-cognitive ability. The wage premium associated to skill bundle (c,n) in firm $\boldsymbol{\lambda}$ is

$$e^{\Delta_{\boldsymbol{\lambda}}(c,n)} = \frac{f_{\boldsymbol{\lambda}}(c,n)}{f_{\boldsymbol{\lambda}}(L,l)} \tag{8}$$

for all $(c,n) \in S$. The premium $e^{\Delta_{\boldsymbol{\lambda}}(c,n)}$ is proportional to the (measurable) productivity of a (c,n) worker in firm $\boldsymbol{\lambda}$ relative to a baseline worker of type (L,l). The parameter $\Delta_{\boldsymbol{\lambda}}(c,n)$ subsumes two sources of variation: (i) the skill endowment bundle (c,n), and (ii) the return to that bundle in firm $\boldsymbol{\lambda}$. By definition, $\Delta_{\boldsymbol{\lambda}}(L,l) = 0$ and one can redefine baseline match productivity in firm $\boldsymbol{\lambda}$ as $T_{\boldsymbol{\lambda}} = f_{\boldsymbol{\lambda}}(L,l)$, which is the output of workers of type (L,l). Using $T_{\boldsymbol{\lambda}}$ and $\Delta_{\boldsymbol{\lambda}}(c,n)$, we write the technology of firm $\boldsymbol{\lambda}$ as $y_{\boldsymbol{\lambda}} = T_{\boldsymbol{\lambda}} \sum_{s \in S} e^{\Delta_{\boldsymbol{\lambda}}(s)} q_{\boldsymbol{\lambda}}(s)$ and recast the profit maximization as a choice over a discrete set of skill bundles S.

Optimal hiring behavior in the discrete maximization problem implies:

$$w_{\boldsymbol{\lambda}}(s) = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{\frac{\sigma-1}{\sigma}\phi_{\boldsymbol{\lambda}} T_{\boldsymbol{\lambda}} \left(\frac{1}{y_{\boldsymbol{\lambda}}}\right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{e^{\Delta_{\boldsymbol{\lambda}}(s)}}_{\text{Skill Productivity}} \tag{9}$$

The latter expression captures different aspects of market structure. The marginal revenue is an increasing function of the firm's output demand $\phi_{\boldsymbol{\lambda}}$. 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. Crucially, an extra unit of skill s rescales marginal revenues proportionally to the firm's skill return $\Delta_{\boldsymbol{\lambda}}(s)$.

In log form, the equilibrium wage is the sum of a base intercept, a firm fixed effect, and a skill-specific return, lending theoretical support to empirical specifications like equation 1 in the previous section. That is:

$$\ln(w_{\boldsymbol{\lambda}}(s)) = \alpha + \theta_{\boldsymbol{\lambda}} + \Delta_{\boldsymbol{\lambda}}(s). \tag{10}$$

The intercept $\alpha \equiv \ln\left(\frac{\beta}{1+\beta}\frac{\sigma-1}{\sigma}\right)$ is common across firms and skills, while $\theta_{\boldsymbol{\lambda}} \equiv \ln\left(\phi_{\boldsymbol{\lambda}} T_{\boldsymbol{\lambda}} y_{\boldsymbol{\lambda}}^{-\frac{1}{\sigma}}\right)$ is the firm-specific baseline wage, which does not vary with worker skills, and $\Delta_{\boldsymbol{\lambda}}(s)$ is a *firm-specific return to skill bundle s*. Under the model's null hypothesis, the firm's demand intercept $\phi_{\boldsymbol{\lambda}}$ is subsumed in the fixed effect component $\theta_{\boldsymbol{\lambda}}$.

Optimal behavior implies that firms with higher returns to s-type skills tend to hire a larger share of s-type workers.[14] The latter observation suggests that firms with similar returns to a skill

[14]That is, for two firms denoted as $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, the following holds

$$\mathbb{E}\left[\log\left(\frac{q_{\boldsymbol{\lambda}_1}(s)}{q_{\boldsymbol{\lambda}_1}(L,l)}\right) - \log\left(\frac{q_{\boldsymbol{\lambda}_2}(s)}{q_{\boldsymbol{\lambda}_2}(L,l)}\right) |\Delta_{\boldsymbol{\lambda}_1}(s), \Delta_{\boldsymbol{\lambda}_2}(s)\right] = \beta(\Delta_{\boldsymbol{\lambda}_1}(s) - \Delta_{\boldsymbol{\lambda}_2}(s))$$

This condition is satisfied under the null hypothesis of our model, where the distributions of workers' preferences for firms and skill returns to specific skill bundles are independent. However, the result is robust to violations of the independence assumption. For example, it holds as long as the correlation between preferences for firms and skill returns is positive (that is, if skill-biased firms exhibit, on average, better non-pecuniary returns for higher type s workers). Weak negative correlations are also not sufficient to overturn the grouping result.

bundle can be grouped together based on their share of workers with that particular skill bundle, and that firm-specific returns to different skill bundles can be identified from a cross-section of grouped worker wages if direct skill measures are available, lending further support to our estimation approach for the skill type premia in the previous section.[15]

3.3 Separating Skill Endowments and Returns

The estimated wage premia presented in Section 2 are non-parametric and do not depend on functional assumption about the production technology. While this approach requires fewer assumptions, those estimates convey little or no information about the way the firm's premium to a particular skill bundle is shaped by heterogeneous returns to each skill attribute. In this section we impose the minimum amount of structure necessary to recover firms' marginal returns to different ability traits. To this purpose it helps to posit that skill endowments are comparable in a cardinal sense. We let c and n denote worker skill *levels*, defined over the cognitive and non-cognitive range. In the setting of our model this means that, for a given s-type worker, $s=(c,n)\in\mathbb{R}^2_+$ for all $s\in S$. A worker receives wages $w_{js}=wage_j(c,n)$ working at firm j, where $wage_j(\cdot,\cdot)$ denotes a (continuous) wage function of c and n. After a first-order approximation of $wage_j$, we obtain a bilinear log wage equation:

$$\ln(w_{js})\approx\lambda_j^0+\lambda_j^c c+\lambda_j^n n \tag{11}$$

where λ_j^c and λ_j^n are firm j's marginal returns to cognitive and non-cognitive endowments of a worker of type $s=(c,n)$. This approximation is exact for the widely used Cobb-Douglas production function.[16]

[15] An additional implication of firms' optimal hiring behaviour is that the total number of workers employed by a firm $\boldsymbol{\lambda}$ grows with the wage intercept θ_j. This follows from the observation that $\mathbb{E}\left[\log(q_{\boldsymbol{\lambda}_1}(L,l))-\log(q_{\boldsymbol{\lambda}_2}(L,l))|\theta_{\boldsymbol{\lambda}_1},\theta_{\boldsymbol{\lambda}_2}\right]=\beta(\theta_{\boldsymbol{\lambda}_1}-\theta_{\boldsymbol{\lambda}_2})$.

[16] The bilinear wage function in levels is $w_{js}=e^{\lambda_j^0}\cdot(e^c)^{\lambda_j^c}\cdot(e^n)^{\lambda_j^n}$. This is consistent with Cobb-Douglas output by worker type (c,n) at firm j in a model with a (constant) surplus sharing rule. This includes our model where the monopsony markdown is $\beta/(1-\beta)$).

An iterative method-of-moments estimator. We use a GMM approach to jointly estimate the linear returns in (11) and a set of skill endowments (up to an affine normalization) that span the full set of observed skill bundle premia. By matching variation in skill premia, rather than individual worker wages, we indirectly account for observed and unobserved covariates such as individual fixed effects, general life-cycle returns, flexible time effects, and more.

The wage equation expresses the non-parametric estimates of bundled premia as the product of marginal returns and skill endowments. As marginal skill returns are, by design, the loadings necessary to account for between firm variation in wage premia, the skill nodes serve the purpose of holding ability values fixed when estimating marginal skill returns. That is, returns can be identified by within-firm wage variation over a set of skill nodes that are common across firms. Given its simplicity, it is easy to explore departures from the baseline specification. For example, we verify that adding interaction effects $c \times n$ improves the empirical fit of the model very marginally.[17] Using skill bundle premia from the previous section as targets, the GMM optimization problem in the bilinear wage model is:

$$\min_{\substack{\{\lambda_j^0, \lambda_j^c, \lambda_j^n\}_{j=1}^J \\ s=(c,n)\in S}} \sum_j \sum_{s\in S} \left[\lambda_j^0 + \lambda_j^c c + \lambda_j^n n - \log(w_{js})\right]^2 \tag{12}$$

To estimate the (minimizing) parameter values we adopt an iterative procedure. First, we set a starting value for skill levels c and n, and solve for the firm-specific set of λ_j. Then, holding the λ_j fixed, we minimize the objective with respect to c and n to obtain updated values that can be entered into (12) to solve for a new set of λ_j. These steps are repeated until parameter estimates converge. The first-order conditions for the estimation parameters and the explicit solutions of the minimization problem are reported, in matrix notation, in the appendix. The objective function is convex and, in each step, we minimize it in the subspace of returns and skill levels. Since a global minimum exists, and given continuity of the objective function, the procedure converges to a minimum.[18]

[17]The interaction model is $\ln(w_{js}) = \lambda_j^0 + \lambda_j^c \cdot c + \lambda_j^n \cdot n + \lambda_j^{c\times n} \cdot c \cdot n$.

[18]In practice, the value of the objective function shrinks monotonically with each iteration. Estimation procedures for other functional forms (e.g., the $c \times n$ interaction model) work accordingly.

Empirical fit. The optimization in (12) has infinitely many solutions, as we can identify skill values and returns up to an affine transformation. For example, the cognitive skill premium $\lambda^c c$ might be high because the skill endowment is high or because returns are high (or some combination of the two). This requires a scale assumption and we normalize upper and lower bounds of each ability so that $l = L = 0$ and $h = H = 1$. Given the normalization, our approach delivers estimates of intermediate skill levels (nodes) within the bounded cognitive and non-cognitive ranges. These nodes, in conjunction with the firm-specific marginal returns, are sufficient to characterize the whole set of bundled skill premia.

The number of intermediate nodes estimated within each skill range depends on the number of skill bundles being targeted. Furthermore, the baseline specification posits three firm-specific parameters to target nine firm-specific moment conditions.[19] We also consider an extension that includes additional moments for $c \times n$ skill interactions. The baseline specification fits the moments in (12) remarkably well, accounting for over 96% of cross-sectional variation. The fit can hardly be improved by richer functional forms.

Estimates of skill endowments. Skill nodes help match the variation across the whole range of bundled skill premia. Pair-wise differences between nodes capture the skill endowment gaps between ordered ability categories. Given our nine-way grouping of skill bundles, and the normalization of the higher and lower bounds, this leaves us to estimate one intermediate node for each skill. These estimates lie in the upper half of the $[0,1]$-interval: the intermediate skill level in the cognitive dimension is 0.59, while the non-cognitive node is 0.66. The fact that the intermediate skill endowments are closer to their upper bounds implies that the (cardinal) skill increase when moving from low to middle is much larger than when moving from middle to high endowments. For reference, the intermediate values estimated for other periods are similar: in the 1990–1999 estimation period, they are 0.60 for cognitive and 0.59 for non-cognitive; in 2008–2017 they are 0.62 for cognitive and 0.61 for non-cognitive. While not restricted to be identical, the endowment node estimates are almost the same when we consider a model with higher-order

[19] The moment conditions are the nine skill bundle premia (the combination of three stanine groups for each skill).

interaction terms, such as $c \times n$. This indicates that estimated skill nodes are not the by-product of functional form alone.

Marginal returns in the cross-section of firms. Figure 2 plots the distribution of returns to cognitive and non-cognitive skill endowments in the population of firms, documenting that heterogeneity is stronger for the cognitive (λ_j^c) than the non-cognitive (λ_j^n) returns in all three estimation periods. The pecuniary incentive for skill sorting is, therefore, somewhat stronger in the cognitive dimension.

Comparing distributions over time, their locations—and thereby the levels of the returns to skills—have shifted notably across the estimation periods. The average λ_j^c in the population of firms increased from 18 log points in 1990–1999 to 25 log points during 2008–2017, whereas the average λ_j^n more than doubled from 12 to 25 log points. While cognitive skills remain a key driver of sorting (as we discuss in the following sections), non-cognitive skills have become substantially more important for pecuniary rewards. This evidence is consistent with the burgeoning literature on the rising labor market value of non-cognitive traits (Deming, 2017; Edin et al., 2018), and provides external corroboration of results from different data sources.

4 Implications for Matching

Breaking down wage premia into skill returns (prices) and endowments (quantities) is valuable to characterize the mechanics of matching in a setting with multiple skill dimensions. Table 2 shows that returns to cognitive and non-cognitive skills (c and n) are positively correlated in the cross-section of firms. This correlation grows significantly over the sample period, with returns λ_j^C and λ_j^N becoming more aligned within firms. Most of this increase in the cross-sectional correlation of skill returns occurred between 1990 and 2008, when it went from 2% to 28%.

How did the changing firm heterogeneity affect the allocation of workers to employers? To answer this question, we adapt the notion of assortative matching with multidimensional skills to a setting with many-to-one matching. In the process we derive several theoretical results linking the higher moments of within firm skill distributions to the intensity of worker-firm sorting and use them to gain insights about the extent of matching implied by our estimates of firm heterogeneity.

4.1 Matching and Firm-Specific Skill Distributions

To describe worker sorting across firms we adopt the characterization of multidimensional one-to-one matching proposed in Lindenlaub (2017) and describe worker-firm matching through a mapping that assigns heterogeneous $\boldsymbol{\lambda}$ firms to skill endowments (c,n). However, our framework allows for many workers matching to a single firm. For this reason, while we begin by defining the matching function over average skill levels rather than individual worker skills, we later expand the analysis to account for within firm heterogeneity. The latter step is instrumental to relating the theoretical restrictions to the rich employer-employee data that are used to examine sorting.

First, we introduce some notation. The model posits that firms can differ in several dimensions; namely, a labor supply intercept a and return-to-skill parameters λ^0, λ^c, and λ^n. Define $\boldsymbol{\lambda} = (a,\lambda^0,\lambda^c,\lambda^n)$ and let $F(\boldsymbol{\lambda})$ denote the probability measure describing the distribution of firms in their cross-section. Recall $G(c,n)$ denotes the measure of worker skill vectors in the working population.

From the labor supply equation, and imposing a bilinear wage function, firm $\boldsymbol{\lambda}$ hires a fraction $q_{\boldsymbol{\lambda}}(c,n)$ of the total workforce of type (c,n), where

$$\log(q_{\boldsymbol{\lambda}}(c,n)) = \log(h(c,n)) + \log(a) + \beta(\lambda^0 + \lambda^c c + \lambda^n n) \tag{13}$$

Define $Q_{\boldsymbol{\lambda}}$ to be the total number of workers in firm $\boldsymbol{\lambda}$. We have

$$Q_{\boldsymbol{\lambda}} = \int q_{\boldsymbol{\lambda}}(c,n)dG(c,n) \tag{14}$$

Table 2: Estimated Skill Returns

	$\overline{\lambda^c}$	$\overline{\lambda^n}$	sd(λ^c)	sd(λ^n)	corr(λ^c, λ^n)
1990–1999	0.18	0.11	0.06	0.05	0.02
1999–2008	0.23	0.20	0.06	0.04	0.28
2008–2017	0.25	0.25	0.05	0.04	0.26

Notes: The table reports descriptive statistics for the estimated skill returns and their changes over time. The standard deviations and correlations of skills in the population of workers are constant at $\mathrm{sd}(c) = 0.33$, $\mathrm{sd}(n) = 0.31$, and $\mathrm{corr}(c,n) = 0.31$ for all three periods.

We can also define average cognitive and non-cognitive ability of workers in firm $\boldsymbol{\lambda}$ as follows:

$$\begin{aligned} \bar{c}_{\boldsymbol{\lambda}} &= \int c \; \frac{h(c,n) a e^{\beta(\lambda^0 + \lambda^c c + \lambda^n n)}}{Q_{\boldsymbol{\lambda}}} dG(c,n) \\ &= \int c \; dM_{\boldsymbol{\lambda}}(c,n) \end{aligned} \tag{15}$$

$$\begin{aligned} \bar{n}_{\boldsymbol{\lambda}} &= \int n \; \frac{h(c,n) a e^{\beta(\lambda^0 + \lambda^c c + \lambda^n n)}}{Q_{\boldsymbol{\lambda}}} dG(c,n) \\ &= \int n \; dM_{\boldsymbol{\lambda}}(c,n) \end{aligned} \tag{16}$$

where $M_{\boldsymbol{\lambda}}$ is a probablity measure of the worker-skill distribution within firm $\boldsymbol{\lambda}$. It is worth noting that $M_{\boldsymbol{\lambda}}$ does not depend on a and λ^0 and only varies with cognitive and non-cognitive skill returns λ^c and λ^n.

Assortative Matching. We can define the matching function $\mu(\boldsymbol{\lambda}) = (\bar{c}_{\boldsymbol{\lambda}}, \bar{n}_{\boldsymbol{\lambda}})$, which maps firm returns into average worker skills. The notion of assortative matching, whether positive (PAM) or negative (NAM), is described as a set of properties of the matching function. In matching problems with one dimensional heterogeneity this boils down to the sign of one derivative only. With multidimensional skills, all elements of the Jacobian of the matching function play a role (see Lindenlaub, 2017).

Definition 1. *The sorting pattern is PAM (NAM) if for all* (λ^c, λ^n)

- $\frac{\partial \bar{c}}{\partial \lambda^c} > 0\ (< 0)$
- $\frac{\partial \bar{n}}{\partial \lambda^n} > 0\ (< 0)$
- $\frac{\partial \bar{c}}{\partial \lambda^c}\frac{\partial \bar{n}}{\partial \lambda^n} - \frac{\partial \bar{c}}{\partial \lambda^n}\frac{\partial \bar{n}}{\partial \lambda^c} > 0$

Proposition 1. *The Jacobian of the matching function evaluated at* $\boldsymbol{\lambda}$ *is equal to the covariance matrix of the worker-skill distribution at firm* $\boldsymbol{\lambda}$*.*

$$\frac{d\mu(\boldsymbol{\lambda})}{d(\lambda^c, \lambda^n)} = \beta\ COV_{M_{\boldsymbol{\lambda}}}[c,n] \tag{17}$$

where the covariance is taken under the $M_{\boldsymbol{\lambda}}$ *measure.*

Proof.

$$\begin{aligned}
\frac{\partial \bar{c}_{\boldsymbol{\lambda}}}{\partial \lambda^c} &= \frac{\int \beta\ c^2\ h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)}{\int h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)} - \beta\left[\frac{\int c\ h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)}{\int h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)}\right]^2 \\
&= \beta \int c^2\ dM_{\boldsymbol{\lambda}}(c,n) - \beta\left[\int c\ dM_{\boldsymbol{\lambda}}(c,n)\right]^2 \\
&= \beta \cdot \text{var}_{M_{\boldsymbol{\lambda}}}[c]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \bar{c}_{\boldsymbol{\lambda}}}{\partial \lambda^n} &= \frac{\int \beta\ cn\ h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n))}\ dG(c,n)}{\int h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n))}\ dG(c,n)} - \beta \frac{\int ch(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)}{\int h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)} \times \frac{\int nh(c,n)\ e^{\beta(\lambda^c n+\lambda^n n)}\ dG(c,n)}{\int h(c,n)\ e^{\beta(\lambda^c c+\lambda^n n)}\ dG(c,n)} \\
&= \beta \int cn\ dM_{\boldsymbol{\lambda}}(c,n) - \beta \int c\ dM_{\boldsymbol{\lambda}}(c,n) \times \int n\ dM_{\boldsymbol{\lambda}}(c,n) \\
&= \beta \cdot \text{cov}_{M_{\boldsymbol{\lambda}}}[c,n]
\end{aligned}$$

Similarly we can show $\frac{\partial \bar{n}_{\boldsymbol{\lambda}}}{\partial \lambda^n} = \beta \cdot \text{var}_{M_{\boldsymbol{\lambda}}}[n]$ and $\frac{\partial \bar{n}_{\boldsymbol{\lambda}}}{\partial \lambda^c} = \beta \cdot \text{cov}_{M_{\boldsymbol{\lambda}}}[c,n]$ □

Proposition (1) has implications for sorting. For example, given (arbitrary) upward sloping labor supply and a bilinear wage equation, PAM always holds. Perhaps more interestingly, the simple bilinear wage specification implies a natural analogy to a moment generating function: for example, one can show that the second derivative $\frac{\partial \bar{c}_{\boldsymbol{\lambda}}}{\partial \lambda^n \partial \lambda^n}$ is equal to the (uncentered) third moment of the within-firm distribution of cognitive abilities (a similar result holds for non-cognitive abilities).

These results suggest that the elements of the Jacobian function are tightly related to the prevailing distribution of skills within each firm. This is not surprising. Since the matching function is an equilibrium object, its exact functional form depends on the overall skills and returns distributions in the economy. However to the extent we can approximate this function with a linear one, we could, at least to a first order approximation, link the economy-wide skill dispersion to within firm dispersion.

$$\bar{c}_{\boldsymbol{\lambda}} \quad = \quad \delta_{1c} + \delta_{2c}\lambda^c + \delta_{3c}\lambda^n \tag{18}$$

This approximation together with proposition (1) imply all firms have the same within second moments. In particular $\delta_{2c} = \beta \cdot \mathrm{var}_{M_{\boldsymbol{\lambda}}}[c] \equiv \beta \cdot \overline{\mathrm{VAR}}_{\mathrm{w/n}}[c]$ and $\delta_{3c} = \beta \cdot \mathrm{cov}_{M_{\boldsymbol{\lambda}}}[c,n] \equiv \beta \cdot \overline{\mathrm{COV}}_{\mathrm{w/n}}[c,n]$. Taking expectations from (18) we have

$$\begin{aligned} \mathrm{var}[\bar{c}_{\boldsymbol{\lambda}}] = \beta^2 \Big\{ \quad & \overline{\mathrm{VAR}}_{\mathrm{w/n}}[c]^2 \cdot \mathrm{var}[\lambda^c] + \\ & \overline{\mathrm{COV}}_{\mathrm{w/n}}[c,n]^2 \cdot \mathrm{var}[\lambda^n] + \\ & 2\overline{\mathrm{VAR}}_{\mathrm{w/n}}[c] \cdot \overline{\mathrm{COV}}_{\mathrm{w/n}}[c,n] \cdot \mathrm{cov}[\lambda^c, \lambda^n] \Big\} \end{aligned} \tag{19}$$

We can decompose the total variance of cognitive skills across the economy (similar for non-cognitive skills) using the usual conditional variance formula as follows:

$$\mathrm{var}[c] = \mathrm{var}[\bar{c}_{\boldsymbol{\lambda}}] + \overline{\mathrm{VAR}}_{\mathrm{w/n}}[c], \tag{20}$$

where $\mathrm{var}[\bar{c}_{\boldsymbol{\lambda}}]$ is the variance between firms and $\overline{\mathrm{VAR}}_{\mathrm{w/n}}[c]$ the variance within firms (which is constant across firms in our approximation (18)).

By combining these two equations (19) and (20) we get

$$\begin{aligned}\text{var}[c]-\overline{\text{VAR}}_{\text{w/n}}[c]=\beta^2\Big\{\ & \overline{\text{VAR}}_{\text{w/n}}[c]^2\cdot\text{var}[\lambda^c]+\\ & \overline{\text{COV}}_{\text{w/n}}[c,n]^2\cdot\text{var}[\lambda^n]+\\ & 2\overline{\text{VAR}}_{\text{w/n}}[c]\cdot\overline{\text{COV}}_{\text{w/n}}[c,n]\cdot\text{cov}[\lambda^c,\lambda^n]\Big\},\end{aligned} \tag{21}$$

and an analogous result for the variance of n skills $\text{var}[n]-\overline{\text{VAR}}_{\text{w/n}}[n]$.

The Matching Jacobian in Data. Table 3 presents an empirical test of whether PAM holds, on average, in each of the sample subperiods. Specifically, we run regressions of firm-specific average skills onto their skill returns,

$$\begin{aligned}\bar{c}_{\boldsymbol{\lambda}} &= \delta_{1c}+\delta_{2c}\lambda^c+\delta_{3c}\lambda^n+\varepsilon_c\\ \bar{n}_{\boldsymbol{\lambda}} &= \delta_{1n}+\delta_{2n}\lambda^c+\delta_{3n}\lambda^n+\varepsilon_n\end{aligned} \tag{22}$$

These regressions estimate the elements of the Jacobian from cross-sectional variation in firm returns and average skills. To see this, note that the first regression provides the best linear approximation to the conditional expectations function $\text{E}(\bar{c}_\lambda|\lambda^c,\lambda^n)\approx\delta_{1c}+\delta_{2c}\lambda^c+\delta_{3c}\lambda^n$ and $\frac{\partial\text{E}(\bar{c}_\lambda|\lambda^c,\lambda^n)}{\partial\lambda^c}=\text{E}(\frac{\partial\bar{c}_\lambda}{\partial\lambda^c}|\lambda^c,\lambda^n)\approx\delta_{2c}$. Therefore, δ_{2c} is an average of the theoretical derivatives $\frac{\partial\bar{c}_\lambda}{\partial\lambda^c}=\beta\cdot\text{var}_{M_{\boldsymbol{\lambda}}}[c]$ from above across all λ combinations. Accordingly, $\text{E}(\frac{\partial\bar{c}_\lambda}{\partial\lambda^n}|\lambda^c,\lambda^n)\approx\delta_{3c}$, $\text{E}(\frac{\partial\bar{n}_\lambda}{\partial\lambda^c}|\lambda^c,\lambda^n)\approx\delta_{2n}$ and $\text{E}(\frac{\partial\bar{n}_\lambda}{\partial\lambda^n}|\lambda^c,\lambda^n)\approx\delta_{3n}$.[20]

The positive and significant coefficients on δ_{2c} and δ_{3n} throughout Table 3 show that the own-derivative conditions in Proposition 1 clearly hold in each estimation period. Moreover, the Jacobian is also positive definite, as the determinant $\delta_{2c}\delta_{3n}-\delta_{3c}\delta_{2n}$ is always and unambigously larger than zero. Therefore, PAM holds in all three estimation periods in the large set of Swedish

[20]Estimating the relationships in (22) using unconditional, bivariate regressions may give the wrong results. For example, running

$$\bar{c}_\lambda=\delta_{1c}^{short}+\delta_{2c}^{short}\lambda^c+\varepsilon_c^{short}$$

yields $\text{E}(\bar{c}_\lambda|\lambda^c)\approx\delta_{1c}^{short}+\delta_{2c}^{short}\lambda^c$ and $\frac{d\text{E}(\bar{c}_\lambda|\lambda^c)}{d\lambda^c}=\text{E}(\frac{d\bar{c}_\lambda}{d\lambda^c}|\lambda^c)\approx\delta_{2c}^{short}=\delta_{2c}+\delta_{3c}\frac{\text{cov}(\lambda^c,\lambda^n)}{\text{var}(\lambda^n)}$. The last summand is due to the omitted variables bias (OVB) formula and not in general zero. Therefore, the regression yields a comparison of changing average c skills across firms with higher λ^c *and* λ^n that tends to come with it. Instead, Proposition 1 makes a prediction about the partial derivative $\frac{\partial\bar{c}_\lambda}{\partial\lambda^c}$ only, holding λ^n constant.

Table 3: Estimated sorting parameters

		1990–1999		1999–2008		2008–2017	
		$\bar{c}_\lambda$	$\bar{n}_\lambda$	$\bar{c}_\lambda$	$\bar{n}_\lambda$	$\bar{c}_\lambda$	$\bar{n}_\lambda$
estimated δ_2 and δ_3:							
	λ^c	1.86	0.74	2.04	0.75	1.95	0.82
		(.01)	(.01)	(.01)	(.01)	(.01)	(.01)
	λ^n	0.70	1.36	0.87	1.55	1.15	1.46
		(.01)	(.01)	(.02)	(.01)	(.01)	(.01)
R^2		0.653	0.514	0.717	0.600	0.654	0.524
$\#firms$		19,634		25,249		21,755	

Notes: The table reports estimated sorting coefficients δ_2 and δ_3 from regression (22). Bootstrapped standard errors in parentheses.

firms. This result is far from trivial and it lends empirical support to our baseline model with bilinear returns. It also lends indirect support to our estimates of marginal skill returns, which were obtained solely from workers' wages while the left-hand side of (22) subsumes variation in the relative *quantities* of workers within firms.

Additional theoretical restrictions can also be verified in Table 3. In particular, $\delta_{3c} > 0$ and $\delta_{2n} > 0$ in period 1 indicate that there is positive cross-sorting such that higher returns to one skill tend to raise the average of the other, ceteris paribus.[21] Moreover, in none of the estimation periods we see that $\delta_{2c} < \delta_{3c}$ and $\delta_{3n} < \delta_{2n}$.[22] Results in Table 3 do not—and according to the theory should not—change when we include controls for firm employment size or bilinear intercepts λ^0. Finally, about 50–70 percent of the differences in average skills across firms can be explained by the estimated λ^c and λ^n returns alone.

Stochastic Dominance. An interesting corollary of PAM is that, for any given skill attribute, the distribution of higher-skilled workers over firm-specific returns should stochastically domi-

[21]Table 2 shows that in estimation period 1990–1999 the $\mathrm{cov}(\lambda^c, \lambda^n) \approx 0$. If this also implies independence of returns, one can prove that $\frac{\partial \bar{c}_{\boldsymbol{\lambda}}}{\partial \lambda^n} > 0$ and $\frac{\partial \bar{n}_{\boldsymbol{\lambda}}}{\partial \lambda^c} > 0$ when $\mathrm{cov}[c,n] > 0$.

[22]Because of the variance property $\mathrm{var}_{M_{\boldsymbol{\lambda}}}[c] \cdot \mathrm{var}_{M_{\boldsymbol{\lambda}}}[n] > 2 \cdot \mathrm{cov}_{M_{\boldsymbol{\lambda}}}[c,n]$ for any measure $M_{\boldsymbol{\lambda}}$, it can never be that both $\frac{\partial \bar{c}_\lambda}{\partial \lambda^c} < \frac{\partial \bar{c}_\lambda}{\partial \lambda^n}$ and $\frac{\partial \bar{n}_\lambda}{\partial \lambda^n} < \frac{\partial \bar{n}_\lambda}{\partial \lambda^c}$.

nate that of lower-skilled workers.[23] Figure 3 illustrates this result in our data, separately plotting the λ^c and λ^n cumulative distributions for the skill levels $c \in \{0, M_c, 1\}, n \in \{0, M_n, 1\}$. This is done for the pooled estimation period 1999–2008. The corollary states that, if we hold n , the proportional hiring of higher c workers should rise with λ^c. Equivalently, the hiring of higher n workers should increase with λ^n when holding c constant.[24]

The left panels of Figure 3 show that indeed, holding n constant, the CDF of λ^c shifts to the right for each higher level of c. A similar first-order stochastic dominance result exists when sorting workers by their n endowments and looking at their distributions over λ^n (right panels of the same figure). Therefore, all else equal, variation in each individual skill dimension is consistent with theoretical patterns of stochastic dominance in the distribution of the corresponding firm returns. This corroborates the finding of PAM and the sorting model with bilinear returns more generally.

Another corollary to the sorting theory is that its intensity should be stronger in the skill dimension exhibiting higher dispersion of returns. This is intuitive as workers become more selective when proportional differences in firm-specific returns increase. As shown in Table 2, cognitive returns are substantially more dispersed than non-cognitive ones. We also observe in Figure 3 that sorting on cognitive ability is more intense. This is consistent with the observation that wider differences in cognitive returns lead to more intense sorting on c attributes.

Before moving on to an analysis over time, it is worth examining our new theoretical result (17) in more detail. The $\frac{\partial \bar{c}_{\boldsymbol{\lambda}}}{\partial \lambda^c} = \beta \cdot \mathrm{var}_{M_{\boldsymbol{\lambda}}}[c]$ formula provides a concise description of how c skills in the firm will react to a change in returns λ^c. This has two components; the labor supply elasticity β which indicates that the more willing workers are to follow pecuniary incentives, the

[23]Lindenlaub and Postel-Vinay (2017) derive similar predictions In a competitive setting with search frictions and argue that the distribution for workers with higher c over the range of firm-specific returns λ^c should first-order stochastically dominate its counterpart for workers with lower c (similar results hold for n and λ^n).

[24]Formally, positive FOSD sorting in either cognitive or non-cognitive dimension implies:

$$\begin{aligned} &\text{if} \quad c_1 > c_2 \text{ then } CDF^c(c_1, n, \lambda^c) \leq CDF^c(c_2, n, \lambda^c) \text{ for all } n, \lambda^c \\ &\text{if} \quad n_1 > n_2 \text{ then } CDF^n(c, n_1, \lambda^n) \leq CDF^n(c, n_2, \lambda^n) \text{ for all } c, \lambda^n \end{aligned}$$

more skill sorting will differ for a given difference in returns. The second component is more subtle and related to the types of workers that the firm hires. The higher the dispersion of c skills inside the firm, the more diverse in terms of c skill is the set of workers that are at the margin of joining the firm. Thus, if in a high $\text{var}_{M_\lambda}[c]$ firm return λ^c rises, the newly attracted workers tend to come from high up in the c skill distribution (at least compared to $\bar{c}_\lambda$ of existing workers) and the more the average c skill in the firm is affected. In contrast, in a low $\text{var}_{M_\lambda}[c]$ firm, an increase in return λ^c draws in relatively similar marginal compared to existing workers and $\bar{c}_\lambda$ does not react much.

Table 4: Interacted sorting regression

	1990–1999		1999–2008		2008–2017	
	$\bar{c}_\lambda$	$\bar{n}_\lambda$	$\bar{c}_\lambda$	$\bar{n}_\lambda$	$\bar{c}_\lambda$	$\bar{n}_\lambda$
λ^c	1.53	0.55	1.84	0.50	1.58	0.55
	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)
λ^n	0.24	1.03	0.46	1.02	0.50	1.11
	(.02)	(.02)	(.02)	(.01)	(.02)	(.02)
$\lambda^c \cdot 1\{\text{var}_{M_\lambda}[c] > p50\}$	0.54		0.25		0.36	
	(.03)		(.02)		(.03)	
$\lambda^n \cdot 1\{\text{cov}_{M_\lambda}[c,n] > p50\}$	0.82		0.74		0.83	
	(.03)		(.03)		(.03)	
$\lambda^c \cdot 1\{\text{cov}_{M_\lambda}[c,n] > p50\}$		0.45		0.39		0.54
		(.02)		(.02)		(.02)
$\lambda^n \cdot 1\{\text{var}_{M_\lambda}[n] > p50\}$		0.71		0.67		0.54
		(.03)		(.02)		(.03)
R^2	0.696	0.563	0.744	0.680	0.727	0.618
$\#firms$	19,634		25,249		21,755	

Notes: The first column in each estimation period reports the $\delta_{2c}, \delta_{3c}, \delta_{6c}, \delta_{7c}$ coefficients from regression $\bar{c}_\lambda = \delta_{1c} + \delta_{2c}\lambda^c + \delta_{3c}\lambda^n + \delta_{4c}1\{\text{var}_{M_\lambda}[c] > p50\} + \delta_{5c}1\{\text{cov}_{M_\lambda}[c,n] > p50\} + \delta_{6c}\lambda^c \cdot 1\{\text{var}_{M_\lambda}[c] > p50\} + \delta_{7c}\lambda^n \cdot 1\{\text{cov}_{M_\lambda}[c,n] > p50\} + \varepsilon_c$. Regression for $\bar{n}_\lambda$ accordingly in the second columns. Bootstrapped standard errors in parentheses.

To explain the case of the cross-derivative $\frac{\partial \bar{c}_\lambda}{\partial \lambda^n} = \beta \cdot \text{cov}_{M_\lambda}[c,n]$, consider a firm where $\text{cov}_{M_\lambda}[c,n] < 0$. When λ^n rises, additional high n workers are drawn in. However, in this firm high n tends to come with low c and therefore $\bar{c}_\lambda$ declines. Conversely, when $\text{cov}_{M_\lambda}[c,n] > 0$, which is the case in our data, a rise of λ^n increases $\overline{c}_{\boldsymbol{\lambda}}$ by drawing in high n workers at the margin which tend to also be high c in this firm. This economic interpretation of (17) is of course the same for $\frac{\partial \overline{n}_{\boldsymbol{\lambda}}}{\partial \lambda^n} = \beta \cdot \text{var}_{M_{\boldsymbol{\lambda}}}[n]$ and $\frac{\partial \overline{n}_{\boldsymbol{\lambda}}}{\partial \lambda^c} = \beta \cdot \text{cov}_{M_{\boldsymbol{\lambda}}}[c,n]$.

We also test this new theoretical prediction in the data. Table (4) shows regressions like (22) but interacting the returns regressors with indicators of the variances $\text{var}_{M_{\boldsymbol{\lambda}}}[c]$, $\text{var}_{M_{\boldsymbol{\lambda}}}[n]$ and covariances $\text{cov}_{M_{\boldsymbol{\lambda}}}[c,n]$ above their median. We see that the baseline sorting coefficients decline compared to Table 3 but that the interaction effects with high variances or covariances of skills within firms are very substantial and significant throughout. Thus, indeed the economic prediction that effects of λ^c and λ^n increases are substantially and significantly stronger for firms with high variances and covariances of skills within them is borne out unequivocally in the data.

The results of Table (4) underscore once again that sorting theory is very powerful in explaining the differences in average skills across firms. It also points at the importance of the distribution of skills inside the firm, which goes beyond the representative average worker interpretation of the matching function. The next section will hone in on the economic importance of within-firm skill distributions and their changes over time.

4.2 Deconstructing Wage Variation

The bundling, and correlation, of different skill attributes are key to understand worker assignments to employers. Of course, this assignment is not without consequences for earnings. Nonlinearities due to the matching of high-skill workers to high-return firms help shape earnings' growth across the whole distribution of wage percentiles. As sorting increases, skilled workers accrue more of the gains from production complementarities and this induces a convexification of the earnings curve across different percentiles. To help interpret the impact of skill returns on the economy-wide earnings structure, we write the bilinear wage model (11) as:

$$\ln w_{\boldsymbol{\lambda}}(c,n) \approx \lambda^0 + \bar{\lambda}^c c + \tilde{\lambda}^c c + \bar{\lambda}^n n + \tilde{\lambda}^n n. \tag{23}$$

where $\tilde{\lambda}^c \equiv \lambda^c - \bar{\lambda}^c$ is the demeaned firm-specific return to cognitive skills, and similarly for $\tilde{\lambda}^n$. First, we observe that the linear returns to skills $\bar{\lambda}^c, \bar{\lambda}^n$ are independent of sorting. In contrast, wage dispersion does depend on the assignment of high skills to high returns and reflects wage non-linearities due to complementarities such as $(\lambda^c - \bar{\lambda}^c)c$ and $(\lambda^n - \bar{\lambda}^n)n$. As high-skilled workers sort into high-return firms, the wage curve becomes convex over the skill range.[25]

Using partial projections,[26] one can write the economy-wide (unconditional) return to cognitive skills c as an additively separable function of first and second moments of skill returns:

$$\underbrace{\frac{cov(\ln w_{\boldsymbol{\lambda}}(c,n),c)}{var(c)}}_{\text{overall}} = \underbrace{\bar{\lambda}^c}_{\lambda^c\text{-level}} + \underbrace{\bar{\lambda}^n \frac{cov(n,c)}{var(c)}}_{\lambda^n\text{-level}} + \underbrace{\frac{cov(\lambda^0,c)}{var(c)}}_{\lambda^0\text{-sorting}} + \underbrace{\frac{cov(\tilde{\lambda}^c c,c)}{var(c)}}_{\lambda^c\text{-sorting}} + \underbrace{\frac{cov(\tilde{\lambda}^n n,c)}{var(c)}}_{\lambda^n\text{-sorting}}. \tag{24}$$

The first two summands in the right-hand side of (24) capture the level effects due to the average λ_j^c and λ_j^n across firms. If workers were randomly allocated to firms, the average return to c attributes would equal $\bar{\lambda}^c$. Moreover, there would be an additional return to the worker's n traits, since n covaries with c. The covariation yield is $\bar{\lambda}^n \frac{cov(n,c)}{var(c)}$.

The remaining summands in (24) subsume the systematic sorting of skills across firms. The λ^0-sorting component is the contribution of cognitive skill covariation and employer-specific intercepts. The last two terms on the RHS reflect own- and cross-sorting on cognitive skills. The λ^c-sorting is the own-sorting into high cognitive return firms, which is substantial given the large coefficients in Table 3. Accordingly, the λ^n-sorting measures the cross-sorting into high non-cognitive return firms.

An equivalent decomposition can be obtained for the economy-wide return to non-cognitive skills.[27] Table 5 shows results from the empirical implementation of (24) and (25). During the

[25] Lindenlaub (2017, Proposition 6) formally derives the same result in a frictionless model.

[26] The coefficient from a regression of $\ln w_{\boldsymbol{\lambda}}(c,n)$ onto c is $\frac{cov(\ln w_{\boldsymbol{\lambda}}(c,n),c)}{var(c)}$. Given additive separability, the covariance of (23) with c, normalized by $var(c)$, delivers the decomposition result in (24).

[27] Specifically, for non-cognitive skills we have

$$\underbrace{\frac{cov(\ln w_{\boldsymbol{\lambda}}(c,n),n)}{var(n)}}_{\text{overall}} = \underbrace{\bar{\lambda}^c \frac{cov(c,n)}{var(n)}}_{\lambda^c\text{-level}} + \underbrace{\bar{\lambda}^n}_{\lambda^n\text{-level}} + \underbrace{\frac{cov(\lambda^0,n)}{var(n)}}_{\lambda^0\text{-sorting}} + \underbrace{\frac{cov(\tilde{\lambda}^c c,n)}{var(n)}}_{\lambda^c\text{-sorting}} + \underbrace{\frac{cov(\tilde{\lambda}^n n,n)}{var(n)}}_{\lambda^n\text{-sorting}}. \tag{25}$$

intermediate period in our sample (middle panel, 1999–2008) the overall return to cognitive skills is 35 log points. The average $\bar{\lambda}^c$ accounts for 23 log points, or roughly $2/3$ of the total variation. Another 6 log points are explained by high c workers having, on average, better n attributes. This leaves 6 log points, or about 17 percent of cognitive returns, unexplained. Most of this, 4 log points, is cognitive own-sorting due to the high standard deviation $\mathrm{sd}(\lambda^c)$, reported in Table 2. Finally, cross-sorting effects due to the positive covariation between skill returns and wage intercepts each add another 1 log point to the overall return to cognitive traits.

Table 5: Deconstructing the economy-wide returns to c and n skills.

	(1) overall	(2) λ_j^C-level	(3) λ_j^N-level	(4) λ_j^0-sorting	(5) λ_j^C-sorting	(6) λ_j^N-sorting
1990–1999						
C-return	0.25	0.18	0.03	0.01	0.04	0.00
N-return	0.21	0.06	0.12	0.00	0.02	0.01
1999–2008						
C-return	0.35	0.23	0.06	0.02	0.04	0.01
N-return	0.33	0.08	0.22	0.02	0.02	0.01
2008–2017						
C-return	0.38	0.25	0.07	0.02	0.03	0.01
N-return	0.39	0.08	0.25	0.02	0.02	0.01

Notes: The table shows the wage decompositions (24) and (25) for our three estimation periods. Column (1) gives the economy-wide (unconditional) return to each respective skill, which is split into its individual components in the remainder of the table. Columns (2) and (3) are the contributions of the level return to the respective skill averaged across firms. Column (4) is the correlation of workers' skills with firm wage intercepts. Columns (5) and (6) are the effects due to own-sorting (λ^c-sorting for c skills and λ^n-sorting for n skills) and cross-sorting (λ^c-sorting for n and λ^n-sorting for c).

Similarly, the economy-wide return to non-cognitive attributes between 1999 and 2008 is 33 log points. The average return $\bar{\lambda}^n$ accounts for 22 log points; 8 log points are explained by high-n workers also having better cognitive skills and therefore benefiting from the large $\bar{\lambda}^c$ return. Interestingly, for non-cognitive skills the effect of cross-sorting (2 log points) is more important than own-sorting (1 log points). This means that indirect matching effects account for much of the return from sorting on n skills.[28] Adding the λ^0-sorting effect due to covariation of skill returns and intercepts, we find that five out of 33 log points, or 15% of the return to non-cognitive skills, is explained by worker sorting.

In the earliest estimation period, 1990–1999, total returns to skills were substantially lower than in the main period. As shown in Figure 2, the change is stark in the case of returns to n skills.[29] These differences are mostly due to lower levels of skill returns, rather than lower contributions of own- or cross-sorting. In fact, sorting accounted for 20 and 14 percent of, respectively, cognitive and non-cognitive returns to skills between 1990 and 1998. Similar patterns are detected between 1999 and 2017, with a slightly higher effect of sorting on λ^0.

Wages: random assignment versus assortative matching. Table 5 documents how the assignment of skills to firms changes the distribution of labor market returns in the economy. Over time, the distribution of these returns across wage percentiles has become more convex.

To verify the impact of sorting on the distribution of returns, Figure 4 plots the wage distribution in the bilinear model under random empirical allocation versus the empirical distribution with positive assortative matching. The distribution with random allocation of workers to firms is less dispersed than the empirical one. The much fatter right tail of the empirical distribution implies higher wages (and better productive efficiency) on average. Endogenous sorting seems to increase wages in a first-order stochastic dominance sense.

5 Discussion and Conclusions

We explore a new dimension of firm-level heterogeneity and document that identical skill bundles command different returns across employers. Our empirical analysis relies on matched employer-

[28] This is easily reconciled in Table 2: while the correlations of n with λ^c and λ^n are similar, the dispersion of cognitive returns—$\text{sd}(\lambda^c) = 0.05$—is larger than the its non-cognitive counterpart—$\text{sd}(\lambda^n) = 0.04$.

[29] Comparing all three sample periods, average firm return to non-cognitive skills $\bar{\lambda}^n$ increased by 14 log points from an initial base of 11 log point. This is substantially stronger than the increase of $\bar{\lambda}^c$ of 7 log points from a base of 18, and it is consistent with technical change that is particularly biased toward n (or social) skills as argued in Deming (2017) and Edin et al. (2018).

employee population records in conjunction with direct measures of workers' cognitive and non-cognitive skills.

The fact that workers are unable to sell different factors (cognitive, non-cognitive) to different firms implies a non-trivial sorting problem which crucially depends on the firm-specific marginal returns to each skill attribute. We propose a procedure to recover the distribution of these marginal returns by unbundling the wage premia associated to different skill combinations. This procedure delivers estimates of the distribution of firm-specific rewards to cognitive and non-cognitive skills for different time periods between 1990 and 2017.

Having established basic facts about the distribution of firm-specific skill complementarities, we study the properties of a worker-to-firm assignment problem with multi-dimensional skills and returns. We show that key theoretical restrictions are borne out in data. In particular, we document that workers with higher endowments of specific skills populate firms with higher marginal returns to those skills. Crucially, we also show that the intensity of sorting on each skill dimension depends on the dispersion of that skill's return across firms: as dispersion grows, so does the incentive for more skilled workers to seek a better match.

When examining the empirical properties of the worker-firm assignment problem, we find evidence of significant cross-sorting, whereby workers with high cognitive skills experience a higher than average return in the non-cognitive dimension. The latter phenomenon is due to (i) the positive correlation of different skills in workers' skill bundles; and (ii) the growing alignment of cognitive and non-cognitive skill returns within firms.

Our estimates suggest that, over time, worker-firm matching due to skill complementarities has become more intense, resulting in a convexification of the wage distribution across the range of worker skills.

We find that returns to skills have grown significantly between 1990 and 2017, especially for non-cognitive traits. Sorting patterns have also become more intense over that period; as mentioned before, the latter observation can be attributed to cognitive and non-cognitive returns becoming progressively more aligned within firms. The most productive firms exhibit, by the end of our sample period, disproportionately large returns in both dimensions. This phenomenon has alleviated the frictions due to the fact that workers cannot unbundle their skills and separately rent them out to different firms. Stronger multi-dimensional matching in later years has bolstered the gains from firm-worker complementarities resulting in more inequality of earnings.

References

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

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.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2018): "A Distributional Framework for Matched Employer Employee Data," Mimeo, University of Chicago.

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.

DE MELO, R. L. (2018): "Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence," *Journal of Political Economy*, 126, 313–346.

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

EDIN, P.-A., P. FREDRIKSSON, M. NYBOM, AND B. OCKERT (2018): "The Rising Return to Non-Cognitive Skill," *IZA Discussion Paper*.

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

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.

GRILICHES, Z. (1977): "Estimating the returns to schooling: Some econometric problems," *Econometrica*, 1–22.

GUVENEN, F., B. KURUSCU, S. TANAKA, AND D. WICZER (2020): "Multidimensional skill mismatch," *American Economic Journal: Macroeconomics*, 12, 210–44.

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.

JAROSCH, G., J. S. NIMCZIK, AND I. SORKIN (2019): "Granular search, market structure, and wages," Tech. rep., National Bureau of Economic Research.

LAMADON, T., M. MOGSTAD, AND B. SETZLER (2019): "Imperfect Competition, Compensating Differentials and Rent Sharing in the US Labor Market," Tech. rep., National Bureau of Economic Research.

LENTZ, R., S. PIYAPROMDEE, AND J.-M. ROBIN (2018): "On Worker and Firm Heterogeneity in Wages and Employment Mobility: Evidence from Danish Register Data," .

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

LINDENLAUB, I. AND F. POSTEL-VINAY (2017): "Multidimensional Sorting under Random Search," in *Meeting Papers*, Society for Economic Dynamics.

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.

SHIMER, R. AND L. SMITH (2000): "Assortative matching and search," *Econometrica*, 68, 343–369.

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*," *The Quarterly Journal of Economics*, qjy001.

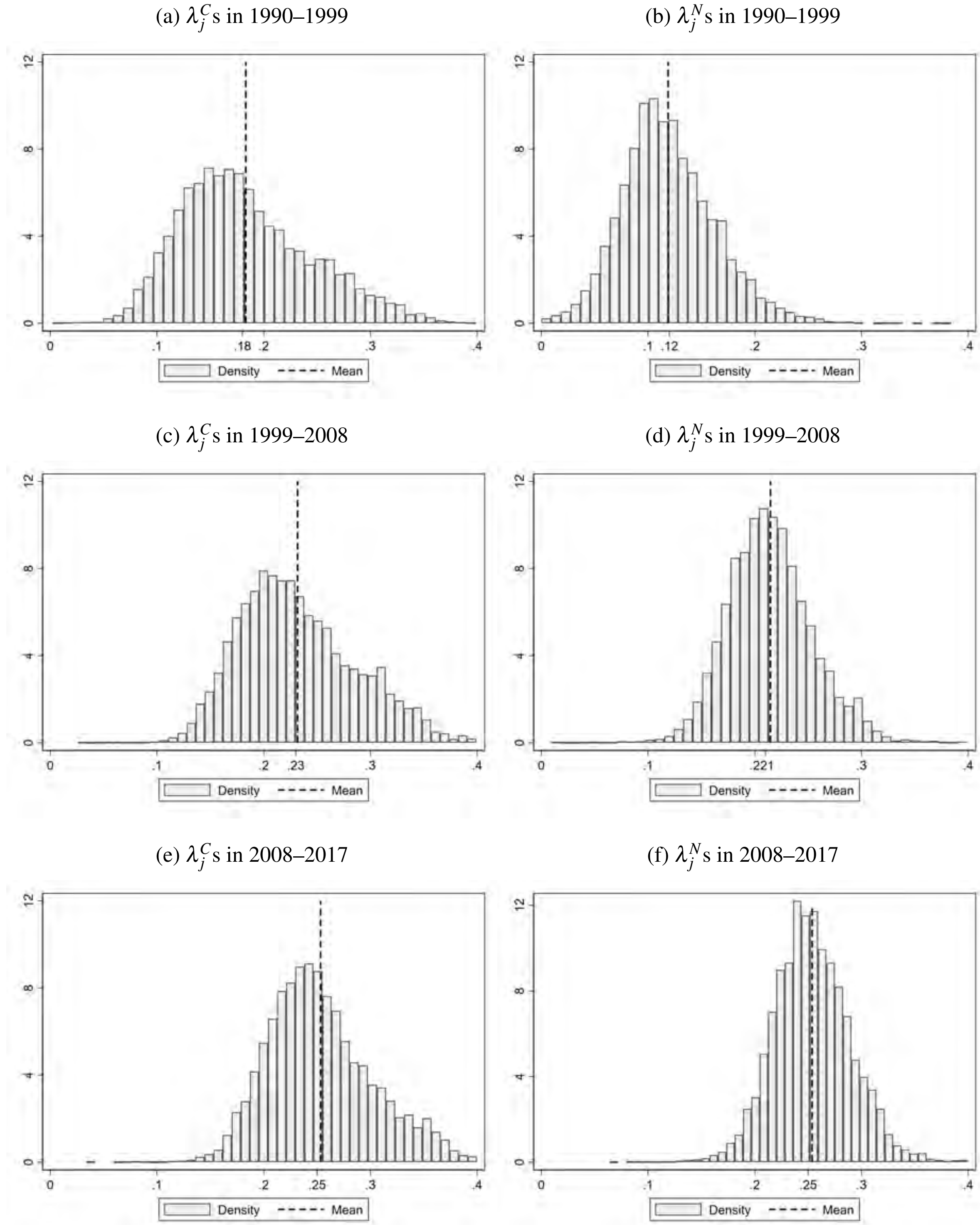
WILLIS, R. J. (1986): "Wage determinants: A survey and reinterpretation of human capital earnings functions," in *Handbook of labor economics*, Elsevier, vol. 1, 525–602.

Table 6: Average return to ability in different five-year periods (men, age 30 and 40). Standard errors in parenthesis.

	(1) **Age 30**	(2) **Age 40**	(3) **Age 30**	(4) **Age 40**	(5) **Age 30**	(6) **Age 40**
Rtrn to cognitive ability (1990-94)	**4.77**	**7.42**			**3.37**	**6.17**
	(0.00)	(0.00)			(0.00)	(0.00)
difference in 1995-99	0.05	0.50			0.23	-0.02
	(0.58)	(0.00)			(0.01)	(0.85)
difference in 2000-04	0.57	1.71			0.42	0.68
	(0.00)	(0.00)			(0.00)	(0.00)
difference in 2005-09	-0.78	0.83			-0.93	-0.09
	(0.00)	(0.00)			(0.00)	(0.32)
difference in 2010-14	-1.51	-0.01			-1.53	-0.90
	(0.00)	(0.91)			(0.00)	(0.00)
Rtrn to non-cogn. ablty (1990-94)			**5.69**	**6.27**	**4.15**	**4.15**
			(0.00)	(0.00)	(0.00)	(0.00)
difference in 1995-99			0.08	1.30	0.12	0.88
			(0.43)	(0.00)	(0.26)	(0.00)
difference in 2000-04			0.74	3.65	0.78	2.60
			(0.00)	(0.00)	(0.00)	(0.00)
difference in 2005-09			-0.14	3.79	0.45	3.34
			(0.15)	(0.00)	(0.00)	(0.00)
difference in 2010-14			-0.77	2.64	0.07	2.66
			(0.00)	(0.000	(0.53)	(0.00)
N	808,213	832,946	808,213	832,946	808,213	832,946
R-sq	0.190	0.273	0.196	0.264	0.196	0.264
Sample	Men	Men	Men	Men	Men	Men

Notes: p-values are under the coefficients.

Figure 2: Distribution of marginal returns to cognitive (c) and non-cognitive (n) skills in the population of firms



(a) λ_j^Cs in 1990–1999

(b) λ_j^Ns in 1990–1999

(c) λ_j^Cs in 1999–2008

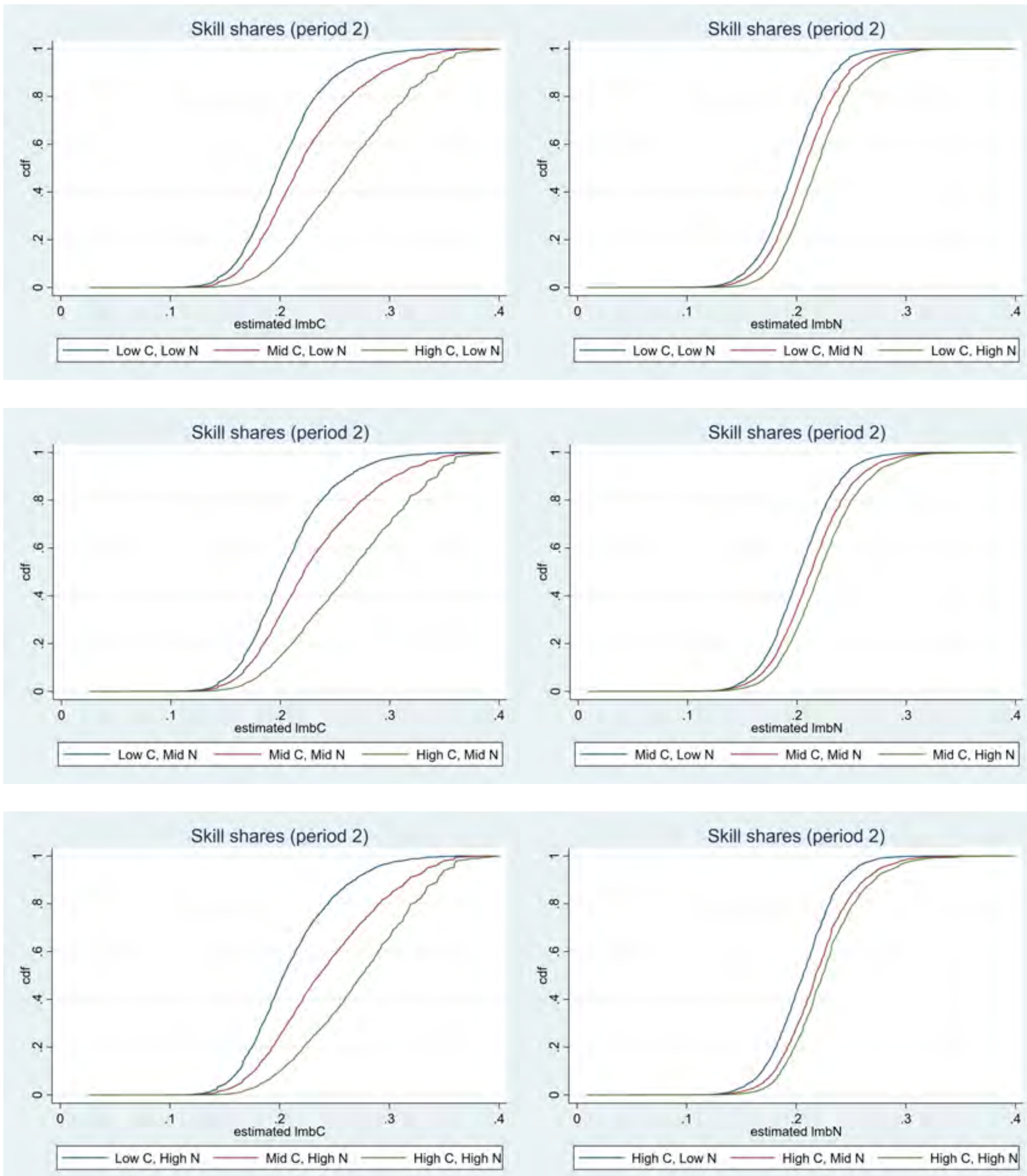
(d) λ_j^Ns in 1999–2008

(e) λ_j^Cs in 2008–2017

(f) λ_j^Ns in 2008–2017

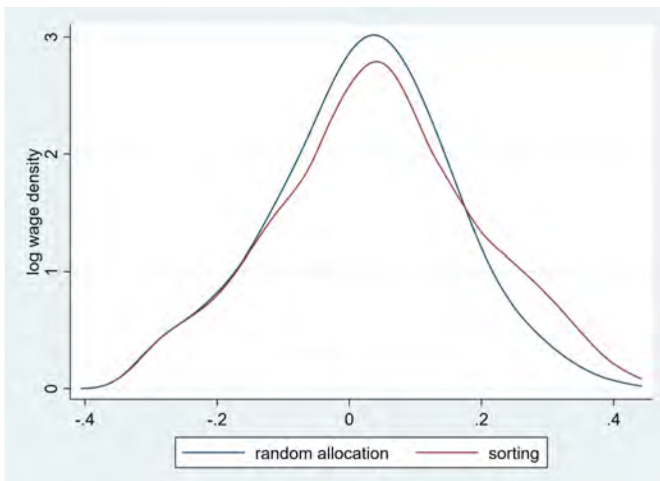
Notes: The figures plot histograms of λ_j^cs and λ_j^ns estimates in the cross-section of firms. The dashed line corresponds to the average in the respective period.

Figure 3: Own-sorting and first-order stochastic dominance. Cumulative shares of: c skills over λ^c (left panels); and n skills over λ^n (right panels). Years: 1999–2008.



Notes: The top left panel in this figure separately plots the cumulative distribution function (CDF) for three groups of workers with progressively higher c skills (i.e., low=0, mid=M_c, and high=1) over the estimated λ^c of the firms where they are employed. The CDF of high c workers first order stochastically dominates that of middle c workers; the latter dominates that of low c workers. The top left panel holds the n skill at its low value. Moving down the left panels of the figure, the same graph is plotted holding n skill at middle and high, respectively. The right panels show similar results for the CDF of low, middle and high n over the estimated λ^n.

Figure 4: Wage distribution with and without sorting (1999–2008)



Notes: The blue line in the figure depicts the wage distribution kernel when employment of c, n skills is allocated equally across firms in estimation period 2. That is, what one would observe, on average, under random allocation of workers to firms. The red line plots the observed wage distribution, i.e., with workers sorting into firms as in the actual data.