Firm Heterogeneity in Skill Returns

Michael J. Böhm
Khalil Esmkhani
Giovanni Gallipoli*

THIS DRAFT: NOVEMBER 15, 2020

Abstract

This paper presents new evidence on worker–firm complementarities. We combine matched employer–employee data with direct measures of workers' cognitive and noncognitive skills, and propose an empirical approach that separately identifies the firm-level return for each attribute. We find that similar skills command different returns across employers and that workers' sorting into firms depends on returns to both attributes. We derive theoretical restrictions that characterize many-to-one matching in employer–employee data, linking within-firm skill dispersion to between-firm differences in average skills. Estimates support these restrictions. Firm heterogeneity in skill returns raises both the average level and dispersion of earnings.

Keywords: Firm Heterogeneity, Skill Returns, Sorting, Wages, Inequality
JEL Classification: D3, 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 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.

1 Introduction

The recognition that earnings distributions reflect both worker and firm heterogeneity dates back decades. In 1986 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". 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., 2019). A workhorse of the applied 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 variance of log earnings can be transparently quantified. In this context, the covariation 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. These considerations inform rich empirical frameworks that nest flexible matching mechanisms within two-sided unobserved heterogeneity (e.g., Bonhomme et al., 2019; Lentz et al., 2018).[1]

This paper presents new results on worker–firm complementarities. We directly estimate firm-specific returns to different skill attributes by linking cognitive and noncognitive military test scores to the population of workers and firms in Sweden. Several studies have convincingly established the empirical content of these test scores.[2] We employ a two-step procedure to separately identify workers' skill endowments and firms' returns to those endowments. In the first step, we build on the AKM specification to recover nonparametric estimates of firm-specific wage premia for different skill bundles; these are the input to the second step, in which we separately estimate the firm-level returns and worker skill endowments that span the set of wage premia from

[1] To preserve tractability, these approaches resort to dimension reduction techniques based on grouping.

[2] For example, Lindqvist and Vestman (2011) document that the military test scores are highly significant at predicting labor market 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.

the previous step. With endowments and returns in hand, we are able to explicitly describe the interaction between heterogeneous attributes of workers and firms.

We document significant variation in skill returns. Firms at the top of the distribution of returns pay up to 20 log points more than firms at the bottom for similar cognitive and noncognitive attributes. This heterogeneity introduces an incentive for high-skill workers to match with high-return employers. However, as emphasized by Lindenlaub (2017) and Lindenlaub and Postel-Vinay (2020), the multidimensional nature of skill bundles and firm returns limits workers' ability to rent out each attribute to the employer with the highest return to it. To characterize matching in the many-to-one setting found in employer–employee data, we derive testable restrictions linking within-firm skill dispersion to differences in average skills between firms.[3] Our estimates lend empirical support to these restrictions and show that worker sorting depends on firm-specific skill returns, as implied by positive assortative matching.

The use of direct skill measures to study worker–firm complementarities and sorting is promising and, to our knowledge, little explored. Our findings draw attention to distinct dimensions of firm heterogeneity and support the view that worker–firm complementarities are quantitatively important. In this respect, we also show that variation in skill returns and sorting help account for the thick right tail in the distribution of earnings, driving up both dispersion and efficiency.

2 Firm Heterogeneity and Skill Types

2.1 Data

We use 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. To this we link scores from cognitive and noncognitive military enlistment tests, which were mandatory for males before 2007. These are the basis for our direct skill measures and a key strength of the Swedish setting.

[3] More generally, the matching function can be fully described by moments of the within-firm skill distributions.

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 test scores are highly significant at predicting workers' earnings and other labor market outcomes (e.g., Lindqvist and Vestman, 2011; Fredriksson et al., 2018), on their own and conditionally on each other or 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 the online Appendix we discuss additional details about these tests and show that, while assessed at age 18–19, test scores strongly predict wages over the whole life-cycle in our sample.

We restrict the sample to males aged 20–60 with nonmissing test scores. We also restrict attention to firms that employ an average of at least ten male workers for at least five years. 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, 1,100,000 workers, and 6,600,000 worker–firm years. Details about data and descriptive statistics are in the online Appendix.

Our dataset reports both organization and workplace identifiers. We use the latter to identify "firms", since workplace is closest to the notion of a production unit and is consistent with approaches in the literature (e.g., Card et al., 2013). We use annual labor earnings to measure employment returns and have verified that results are not sensitive to using full-time equivalent monthly wages, which are available for a subset of surveyed firms.

To reduce measurement error, and simplify interpretation, we granularize test scores for each attribute into three ranked types (high, medium, or low). Every worker has a bundle of attributes $s = (c, n)$, with the first letter denoting cognitive and the second noncognitive. A skill type s is within the set $S = \{(c, n) \mid c \in \{L, M, H\}; n \in \{l, m, h\}\}$. The cognitive ranks are $\{L, M, H\}$ for

[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.

high, medium and low, while $\{l,m,h\}$ is the set of noncognitive attributes.[5] There are nine skill types, one for each combination of cognitive and noncognitive ranks.

2.2 Firm-Level Wage Premia

We examine the hypothesis that heterogeneous 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 type premium. This generalizes the Abowd et al. (1999, AKM) specification as it features worker–firm complementarities in addition to firm and worker fixed effects:

$$\log(w_{ijt}) = \theta_j + \mu_i + \sum_{s\in S\setminus\{(L,l)\}} \Delta_j(s)\mathbb{1}[s_i = s] + \boldsymbol{X}_{it}\boldsymbol{b}_t + \varepsilon_{ijt}. \tag{1}$$

Wage premia $\Delta_j(s)$ are relative to the lowest skill type (L,l), which is the omitted group.

Interpreting parameters. For the subset of workers of base type (L,l), equation (1) collapses to a standard AKM specification with firm fixed effects θ_j, time-varying controls $\boldsymbol{X}_{it}\boldsymbol{b}_t$ (skill type, year, and age fully stratified), and worker fixed effect μ_i. Base wages θ_j are identified, up to a normalization, from wage changes for (L,l) workers switching firms. For other skill types, (1) augments the AKM specification by allowing for firm-specific premia on top of base wages. Each premium $\Delta_j(s)$ is identified from information on between-firm switches by workers with skill type s, and by their wage changes *relative* to those of workers with base type (L,l).

Fixed effects μ_i and the covariates $\boldsymbol{X}_{it}\boldsymbol{b}_t$ control for wage variation that is worker-specific. We interact all skill type, year, and age dummies in $\boldsymbol{X}_{it}\boldsymbol{b}_t$ to flexibly account for skill premia and their evolution over age and time. Conditional on this, the worker fixed effects absorb time-invariant residual skill components.[6] The *variation* in $\Delta_j(s)$ therefore represents firm-specific heterogeneity in skill premia around an economy-wide mean controlling for age and time effects. Without loss of generality, we normalize the μ_i and $\boldsymbol{X}_{it}\boldsymbol{b}_t$ in the estimation so that the *level* of

[5] Lower ranks are STANINE scores 1 to 3, middle ranks 4 to 6, and high ranks 7 to 9.

[6] Allowing for time-varying returns to years of education does not change the results. Life-cycle profiles (by skill and time) are accounted for by the age×year×skill type interaction in $\boldsymbol{X}_{it}\boldsymbol{b}_t$.

$\Delta_j(s)$ represents the average (conditional) wage difference between skill s and the base type (i.e., the "average skill premium" of type s).

The availability of worker-level skill measures is key for the estimation of heterogeneous skill 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.

Grouping firms. 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. It is well-known that estimation based on individual firm observations may suffer from measurement error and lead to biases due to limited mobility of workers across employers (e.g., Andrews et al., 2008). To address these issues, Bonhomme et al. (2019) 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.[7]

In light of these concerns, we develop a simple grouping estimator. To reduce dimensionality, we rank firms into quintiles according to their hiring of each skill type relative to the hiring of base type (L,l). The ranking for the base type is with respect to the absolute number of (L,l) workers. Group assignments for different skill types are, by design, flexible and independent of one another. For example, two firms in the same quintile of the (H,h) type need not be in the same quintile for the other skill types.

Grouping based on employment shares is consistent with the observation that, as we show in the next section, skill premia $\Delta_j(s)$ are tightly related to relative employment of the corresponding skill types s. This relationship can be derived from a simple labor market model with worker and firm heterogeneity (see online Appendix A). However, we also verified that results are qualitatively robust if we impose no grouping and estimate (1) with fully flexible parameters for each individual firm. As expected, the additional measurement error implies larger heterogeneity among these estimates.

[7]Hagedorn et al. (2017) employ a theory-driven grouping approach.

Table 1: Distribution of base wages and skill premia in the cross-section of firms

	Sample period: 1999–2008				
	mean	sd	P_5	P_{50}	P_{95}
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
$\Delta_j(M,m)$	0.26	0.04	0.22	0.23	0.37
$\Delta_j(M,h)$	0.28	0.05	0.21	0.28	0.40
$\Delta_j(H,l)$	0.19	0.07	0.08	0.19	0.32
$\Delta_j(H,m)$	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: Distributions of estimated base wages and skill premia across firms. The sample statistics displayed are the mean, standard deviation, and the 5th, 50th, and 95th percentiles. The sample covers 25,604 unique firms between the years 1999 and 2008.

2.3 Nonparametric Estimates of Wage Premia

Table 1 summarizes sample statistics from the distributions of base wages and skill premia across firms. The employment-weighted mean of θ_j is normalized to zero; the cross-sectional variation of θ_j around its mean reflects differences in base wages across firms, with a standard deviation of about 4 log points and a 5–95 interquantile range of 13 log points.[8]

Skill premia estimates are relative to the base wage and rise with the skill rank within each type, e.g., $\bar{\Delta}(L,m) = 0.10$, $\bar{\Delta}(M,m) = 0.26$, and $\bar{\Delta}(H,m) = 0.30$. The average premium for the top type (H,h) is 46 log points, and it can exceed 59 log points in the firms with the highest returns to high skills. The sizes and ranks of skill premia underscore that our measures of cognitive and noncognitive attributes convey genuine information about traits that are rewarded in the labor market.

Table 1 documents the significant dispersion of skill premia across firms: standard deviations are 4–7 log points and the 5–95 interquantile ranges are large, between 13 and 24 log points. The

[8] Online Appendix Table B.3 reports the contributions of firm and worker heterogeneity to the total variance of earnings, based on specification (1). Results are comparable to similar variance-accounting exercises for Sweden (Bonhomme et al., 2019) and the U.S. (Lamadon et al., 2019).

overall wage variation is substantially larger than what is captured by conventional firm fixed effects. In fact, as shown in Section 4, if we did not account for $\sum_{s \in S \setminus \{(L,l)\}} \Delta_j(s) \mathbb{1}[s_i = s]$ in the estimation of (1), only a fraction of the skill premia heterogeneity would be loaded onto fixed effects θ_j, with the remainder ending up in either worker fixed effects or residuals ε_{ijt}.

Estimates are sufficiently precise to rule out implausible skill premia. For example, down to the 5th percentile of firms, there are no negative premia relative to the (L,l) base wages. In the online Appendix we check robustness by adding fine occupation and industry interactions with skill types and confirm that heterogeneity at the firm level accounts for most of the estimated wage variation. We also verify that the rank correlation between firm-specific premia and skill intensity is positive and monotonic, which lends further support to the grouping approach and anticipates some of the sorting results below.

3 Returns versus Endowments

The firm-specific wage premia in Section 2.3 are nonparametric estimates. Firm j pays a base wage plus skill premia depending on worker's type s: $\log(w_j(s)) = \theta_j + \Delta_j(s)$.[9] These reduced-form wage premia do not impose functional restrictions on the production technology and their nonparametric estimation accounts for the ordinal nature of the test scores. However, estimates of the $\Delta_j(s)$ premia do not distinguish between workers' skill endowments and firms' returns to these endowments. Moreover, they do not distinguish between returns to the cognitive and noncognitive attributes. Similar drawbacks, which also apply to commonly used fixed effect specifications, limit our understanding of the role of returns versus endowments and of the importance of different skills. In this section we impose the minimum amount of structure necessary to recover firms' marginal returns to, and workers' endowments of, cognitive and noncognitive attributes.

[9] This specification is also derived from the model with two-sided heterogeneity discussed in the online Appendix.

3.1 Separating Skill Endowments and Returns

We posit that skill endowments, c for cognitive and n for noncognitive, can be compared in a cardinal sense. An s-type worker, with $s = (c,n) \in \mathbb{R}^2_+$, receives wages $w_j(s)$ at firm j, which is a continuous function of c and n. A first-order approximation of w_j results in a bilinear log wage equation:

$$\log(w_j(s)) = \theta_j + \Delta_j(s) \approx \lambda_j^0 + \lambda_j^{\mathrm{c}} c + \lambda_j^{\mathrm{n}} n \tag{2}$$

where λ_j^{c} and λ_j^{n} are firm j's marginal returns to cognitive and noncognitive endowments of a worker of type $s = (c,n)$, and the function approximates the nonparametric wage premia. The approximation is exact for the widely used Cobb–Douglas production function.[10]

An iterative method-of-moments estimator. We use a GMM approach to jointly estimate the linear returns in (2) and a set of skill endowments (up to an affine normalization) that span the full set of nonparametric skill premia. By matching variation in wage premia, rather than individual wages, we do not target wage variation that does not depend on firm characteristics, including worker fixed effects, life-cycle profiles, and time effects.

Equation (2) expresses the nonparametric estimates of bundled wage premia as the product of marginal returns and skill endowments. Since returns are, by design, the loadings necessary to account for between-firm variation in wage premia, the endowments serve the purpose of holding skills fixed when estimating marginal returns. That is, returns can be identified by within-firm wage variation over a set of skill endowments that are common across firms. Given its simplicity, it is easy to explore departures from this baseline wage specification. For example, we verify that adding interaction effects $c \times n$ improves the empirical fit of the model very marginally.[11] Targeting nonparametric skill premia estimates from the previous section, the GMM optimization problem reduces to choosing a set of firm-specific returns $\left\{\lambda_j^0, \lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}\right\}_{j=1}^{J}$ and a common grid

[10] For example, this is consistent with Cobb–Douglas output by worker type (c,n) at firm j in a model with constant surplus sharing. This includes the model in the online Appendix, where the monopsony markdown is fixed.

[11] The interaction model is $\log(w_j(s)) = \lambda_j^0 + \lambda_j^{\mathrm{c}} \times c + \lambda_j^{\mathrm{n}} \times n + \lambda_j^{\mathrm{cn}} \times c \times n$.

of cognitive (L,M,H) and noncognitive (l,m,h) nodes, to minimize

$$\begin{aligned} &\sum_j \sum_{(c,n)\in S} \left[\lambda_j^0 + \lambda_j^{\mathrm{c}} c + \lambda_j^{\mathrm{n}} n - \log(w_j(c,n))\right]^2 \\ &\text{s.t.} \qquad S = \{(c,n) \mid c \in \{L,M,H\}, n \in \{l,m,h\}\} \end{aligned} \tag{3}$$

To estimate the parameter values we adopt an iterative procedure. We guess a starting value for skill levels c and n, and solve for the firm-specific set of lambdas. Then, holding lambdas fixed, we minimize the objective with respect to c and n, and obtain updated guesses that can be entered into (3) to solve for a new set of lambdas. These steps are repeated until parameter estimates converge. A graphical illustration of the identification approach, with FOCs and explicit solutions of the minimization problem, are in online Appendix C.

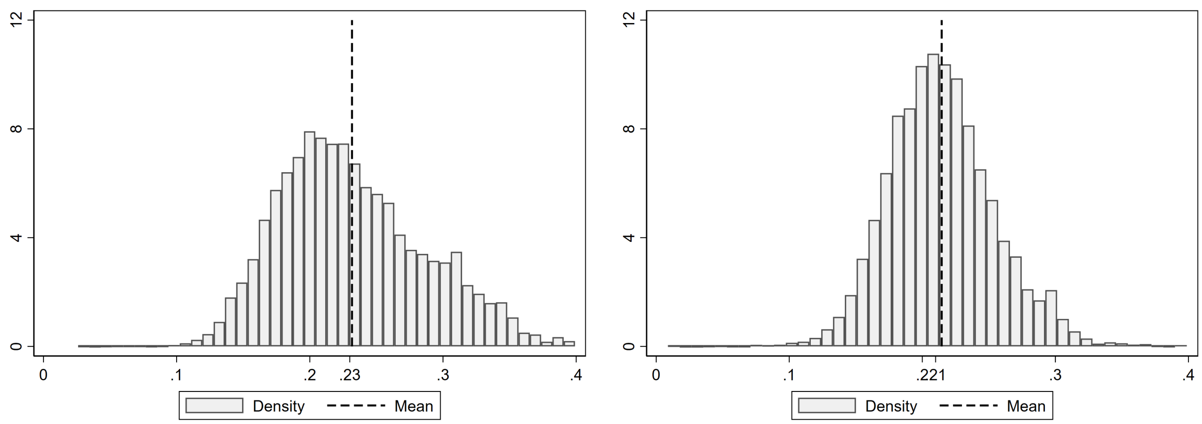
Normalization. The optimization in (3) has infinitely many solutions, as one can identify skill nodes and returns up to an affine transformation. Through a scale assumption we normalize the upper and lower bound of each skill attribute to $l = L = 0$ and $h = H = 1$. Given the normalization, our approach delivers estimates of intermediate skill levels within the bounded cognitive and noncognitive ranges. These levels, in conjunction with the firm-specific marginal returns, are sufficient to characterize the whole set of bundled skill premia.[12]

3.2 Estimation Results

The bilinear model fits the moments in (3) remarkably well, accounting for over 92% of variation. This fit can hardly be improved by richer functional forms.

Skill endowments. Estimates of intermediate skill nodes lie in the upper half of the $[0,1]$ interval: their levels are 0.59 in the cognitive dimension and 0.66 in the noncognitive. Intermediate endowments close to their upper bounds imply that the cardinal skill change when moving from low to middle is larger than when moving from middle to high endowments. Estimates for

[12]The estimation approach can be generalized to a finer skill grid given enough within-firm skill variation. Our sparse grid is, however, sufficient to obtain an accurate match of estimated wage premia.



(a) $\overline{\lambda}^{\mathrm{c}} = 0.232$, $\mathrm{sd}(\lambda^{\mathrm{c}}) = 0.055$

(b) $\overline{\lambda}^{\mathrm{n}} = 0.215$, $\mathrm{sd}(\lambda^{\mathrm{n}}) = 0.041$

Figure 1: Distribution of returns to cognitive (c) and noncognitive (n) skills (log scale).

Notes: Estimates of λ^{c} (left panel) and λ^{n} (right panel) in the population of firms. Dashed lines are averages; $\mathrm{corr}(\lambda^{\mathrm{c}}, \lambda^{\mathrm{n}}) = 0.248$. Estimation period: 1999–2008.

alternative estimation periods are similar: in 1990–1999 (2008–2017) they are 0.60 (0.62) for cognitive, and 0.59 (0.61) for noncognitive. Results are almost identical when including a $c \times n$ interaction term.

Estimates are not hardwired to be similar across periods and functional forms, suggesting that skill nodes capture stable endowments spanning the space of wage premia.

Marginal returns. Figure 1 plots the distribution of (log-scale) cognitive and noncognitive returns in the population of firms. Average returns to the two attributes are similar, with considerable heterogeneity in both dimensions. Most cognitive returns are between 0.1 and 0.4 log points, with a dispersion that is almost a quarter of the mean. This implies strong efficiency gains from matching high c workers to high λ^{c} firms. Noncognitive returns are somewhat less dispersed, but still very heterogeneous.

The correlation of returns in the cross-section of firms is positive, with $\mathrm{corr}(\lambda^{\mathrm{c}}, \lambda^{\mathrm{n}}) = 0.25$, but still far from perfect. Limited correlation suggests that firm heterogeneity is genuinely multidimensional and that worker skills cannot be collapsed into a single index without loss of information; as we discuss below, worker wage rankings are not constant across firms (see also

Lindenlaub, 2017; Lindenlaub and Postel-Vinay, 2020) and sorting occurs along multiple dimensions.

The correlation between intercepts λ^0 and returns is also weakly positive, with $\mathrm{corr}(\lambda^0, \lambda^{\mathrm{c}}) = 0.14$ and $\mathrm{corr}(\lambda^0, \lambda^{\mathrm{n}}) = 0.12$. This suggests that a rent sharing model (consistent with the AKM notion where high-wage firms tend to pay higher wages to everyone) describes firm heterogeneity marginally better than an assignment model (where high returns firms want to set low λ^0 to discourage low-skill and only hire high-skill workers).

The heterogeneity of skill returns is stable across periods (1990–1999 and 2008–2017 are shown in the online Appendix). Cross-firm dispersion, and associated incentives for sorting, is higher for cognitive traits throughout the sample. However, the average λ^{n} has steadily grown, relative to λ^{c}. The latter observation is consistent with the rising labor market value of noncognitive traits (Deming, 2017; Edin et al., 2018).

4 Implications for Matching and Inequality

Does firm heterogeneity matter for the allocation of workers across employers? And how does it affect the distribution of earnings? To examine these questions we adopt the analytical characterization of matching proposed in Lindenlaub (2017), which describes one-to-one worker–firm matching in a setting with multiple skill attributes. Moreover, to account for within-firm skill dispersion, we derive new results linking the higher moments of within-firm skill distributions to between-firm skill differences. These restrictions are testable in the many-to-one settings typically examined in matched employer–employee datasets like ours.

4.1 Skill Variation Between and Within Firms

First, we introduce some notation. Firms differ in three dimensions: each firm has a different wage intercept (λ^0) as well as cognitive (λ^{c}) and noncognitive (λ^{n}) returns. We call $G(c,n)$ the measure of skills in the working population and let $q_j(c,n)$ be the fraction of the total workforce of type (c,n) hired by firm j. Given wage equation (2), we posit that $q_j(c,n)$ is such that

$$\log(q_j(c,n)) = \log(h(c,n)) + \beta(\lambda_j^0 + \lambda_j^{\mathrm{c}} c + \lambda_j^{\mathrm{n}} n), \tag{4}$$

where $\beta > 0$ is an elasticity of skill supply to pecuniary returns and $h(c,n)$ captures the relative scarcity of each skill bundle (c,n). The employment relationship (4) concisely describes an (empirically-consistent) upward sloping labor supply of type (c,n) to firm j. One can derive this relationship from a random utility model of the labor market with two-sided heterogeneity and multiple productive skills (see online Appendix A).

Finally, we define Q_j, the total number of workers in firm j, as

$$Q_j = \int q_j(c,n)\mathrm{d}G(c,n).$$

Using this notation, the average cognitive and noncognitive ability of workers in firm j are:

$$\overline{c}_j = \int c \, \frac{h(c,n)e^{\beta(\lambda_j^0+\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}}{Q_j}\mathrm{d}G(c,n) = \int c \, \mathrm{d}M_j(c,n)$$

$$\overline{n}_j = \int n \, \frac{h(c,n)e^{\beta(\lambda_j^0+\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}}{Q_j}\mathrm{d}G(c,n) = \int n \, \mathrm{d}M_j(c,n)$$

where M_j is the probability measure of the skills' distribution in firm j. The within-firm measure M_j does not depend on λ_j^0 and only varies with cognitive and noncognitive returns λ_j^{c} and λ_j^{n}.

Assortative matching. Assortative matching, whether positive (PAM) or negative (NAM), can be characterized by the properties of the matching function's derivatives. We define matching function $\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = (\overline{c}_j, \overline{n}_j)$, which maps firms' returns into their average worker skills. 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 (see Lindenlaub, 2017).

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

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

Proposition 1. *The Jacobian of the matching function evaluated at $(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})$ is equal to the covariance matrix of the worker-skill distribution within firms with returns $(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})$. That is,*

$$\frac{\mathrm{d}\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})}{\mathrm{d}(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})} = \beta\, \mathbf{cov}_{M_j}[c,n] = \beta \begin{bmatrix} \mathrm{var}_{M_j}[c] & \mathrm{cov}_{M_j}[c,n] \\ \mathrm{cov}_{M_j}[c,n] & \mathrm{var}_{M_j}[n] \end{bmatrix} \tag{5}$$

where the covariance is taken under the within-firm measure M_j.[13]

We emphasize two implications of Proposition 1. First, given a positive elasticity β in (4), PAM must hold as covariance matrices are positive semi-definite. The derivatives are also larger for firms with higher variance and/or covariance of skills within. Notably, the matching function with bilinear returns offers a natural analogy to a moment generating function. For example, the second derivative of $\mu(\cdot)$ with respect to λ_j^{c} is equal to the third moment of the within-firm distribution of cognitive skills (a similar result holds for noncognitive skills). More generally, the matching function with bilinear returns can be fully described using moments of the within-firm distribution of skills, which are often available in matched employer-employee data sets (see general Proposition D.1 in the online Appendix).

The Matching Jacobian in data. Table 2 reports estimates of the Jacobian elements, based on the empirical derivatives of skill averages with respect to returns, in the cross-section of firms. Column (1) shows estimates for the following specifications:

$$\begin{aligned} \bar{c}_j &= \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{c}} \\ \bar{n}_j &= \delta_{1\mathrm{n}} + \delta_{2\mathrm{n}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{n}}\lambda_j^{\mathrm{n}} + \varepsilon_j^{\mathrm{n}}. \end{aligned} \tag{6}$$

These regressions deliver the best linear approximation to the conditional expectations of $\bar{c}_j$ and $\bar{n}_j$. For instance, $\mathrm{E}(\bar{c}_j|\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = \delta_{1\mathrm{c}} + \delta_{2\mathrm{c}}\lambda_j^{\mathrm{c}} + \delta_{3\mathrm{c}}\lambda_j^{\mathrm{n}}$, so that the parameter $\delta_{2\mathrm{c}}$ is the expected

[13] The proof is a simple differentiation of $(\bar{c}_j, \bar{n}_j)$ with respect to $(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}})$. Details in online Appendix D.

Table 2: Jacobian Estimates: Average Sample Derivatives.

	Dependent Variables:					
	(1)		(2)		(3)	
	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$
λ^{c}	2.04	0.75	1.95	0.68	1.84	0.50
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
λ^{n}	0.87	1.55	0.77	1.47	0.46	1.02
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
$\lambda^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[c] > P_{50}\right\}$					0.25	
					(0.03)	
$\lambda^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c,n] > P_{50}\right\}$					0.74	
					(0.03)	
$\lambda^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c,n] > P_{50}\right\}$						0.39
						(0.02)
$\lambda^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[n] > P_{50}\right\}$						0.67
						(0.03)
R^2	0.717	0.600	0.801	0.692	0.744	0.680
# firms	25,249		25,249		25,249	
Controls	No		λ_j^0, # employees		No	

Notes: Estimation period: 1999–2008. Column (1) reports sorting coefficients δ_2 and δ_3 from estimating (6). The specification in column (2) additionally controls for intercepts λ^0 and for the firms' total employment headcounts. Column (3) adds interactions of λ^{c} and λ^{n} with dummies indicating whether the jth within-firm skill variance (or covariance) is above the cross-sectional median. Each firm is one observation. Main effects of $\mathbb{1}\{\cdot\}$ dummies are not displayed for brevity. Robust standard errors in parentheses.

value of the top-left element of the Jacobian $\left(\frac{\partial \overline{c}_j}{\partial \lambda_j^{\mathrm{c}}}\right)$ taken over the sample of all firms. Similar arguments hold for $\delta_{3\mathrm{c}}$ and gradients in the second line of (6).

The positive $\delta_{2\mathrm{c}}$ and $\delta_{3\mathrm{n}}$ in Table 2's column (1) indicate that the own-derivative conditions for PAM in Proposition 1 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. We conclude that PAM holds over the 1999–2008 period in our large sample of Swedish firms.

Positive $\delta_{3\mathrm{c}}$ and $\delta_{2\mathrm{n}}$ in equation (6) indicate substantial cross-sorting, as higher returns to one attribute raise the average quality of the other. This happens when skill endowments are correlated in such a way that high c workers, who sort into high λ_j^c firms, also carry a higher endowment of n skills with them (and similarly for high n workers sorting into high λ_j^n).[14]

These results do not, and according to Proposition 1 should not, change after controlling for firm-specific employment sizes or intercepts λ^0, as we show in column (2). Between 60 and 72 percent of the between-firm skill variation is accounted for by differences in estimated λ^c and λ^n returns alone.

Within-firm skill dispersion versus between-firm average skill gaps. Column (3) of Table 2 probes the empirical content of Proposition 1. Above-median variances and covariances of within-firm skill distributions are associated with decidedly larger derivatives of the Jacobian.[15] This is consistent with Proposition 1, which states that $\overline{c}_j$ or $\overline{n}_j$ should respond more strongly to differences in λ^c or λ^n when firms exhibit higher skill dispersion within. Intuitively, such firms hire a more diverse pool of workers and differences in returns will induce them to sample skills further away from the firm average.

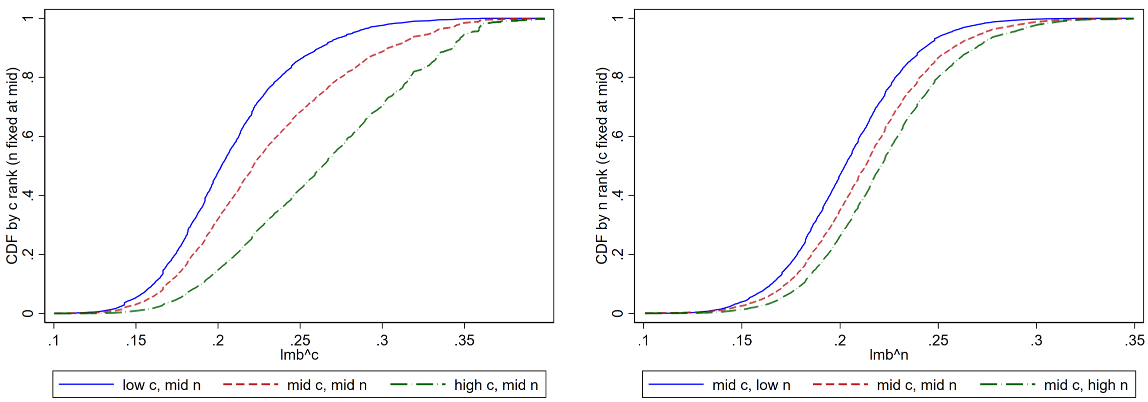
Results in Table 2 are based on variation in workers' skills (from *headcounts*) and on their response to skill returns (from *wages*). The fact that these results consistently line up with theoretical restrictions lends ancillary support to our estimates of skill returns and the employment relationship in (4).

The distributions of worker skills. A natural way to visualize sorting patterns is to compare the distribution of firms that each worker type is matched to. If workers with higher skills in one dimension are more frequently matched to firms with higher returns to those skills (in the sense of first-order stochastic dominance, or FOSD), then sorting is positive along that dimension (Lindenlaub and Postel-Vinay, 2020). The top panels of Figure 2 document FOSD along both

[14]Positive correlation of λ_j^c and λ_j^n would reinforce cross-sorting effects if each indirect return were not controlled for in (6).

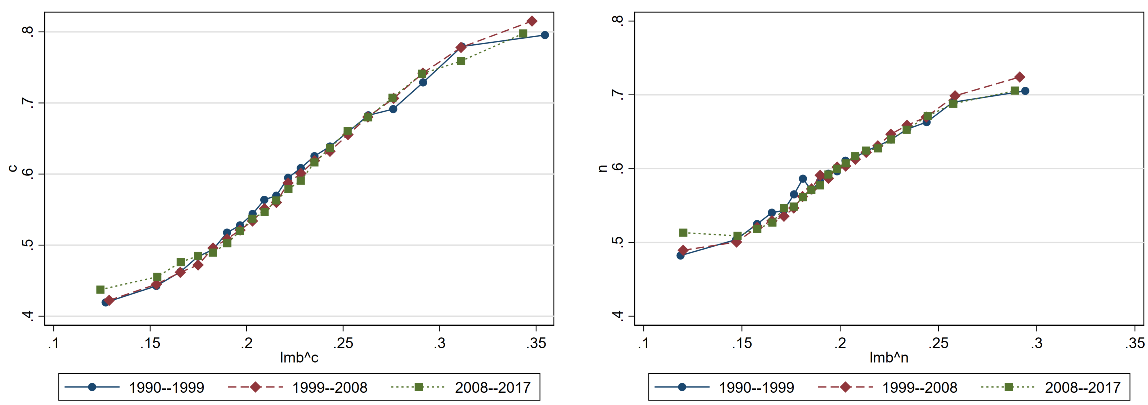
[15]Online Appendix D confirms this relationship, ranking firms into quintiles based on their within-firm skill distributions.

cognitive and noncognitive attributes/returns, plotting CDFs for workers with different skill ranks over the range of firm returns.[16]



(a) FOSD (cognitive)

(b) FOSD (noncognitive)



(c) $\bar{c}_j$ rises in λ^c

(d) $\bar{n}_j$ rises in λ^n

Figure 2: Worker skills versus returns in the cross-section of firms.

Notes: Panels (a) and (b) show cumulative distribution functions for workers with low, middle, or high skill ranks over range of firm returns. Period: 1999–2008. FOSD: first-order stochastic dominance.
Panels (c) and (d) show binned scatterplots of firm-specific skill returns (horizontal axis) with average skills (vertical axis) for three periods: 1990–1999, 1999–2008, 2009–2017.

The bottom panels of Figure 2 expand on these findings, plotting firm-level average skills over firm-level returns. Average skill endowments grow monotonically with returns, as suggested by PAM. Between-firm differences in average skills are larger in the cognitive than in the noncog-

[16]More plots documenting FOSD are in the online Appendix D.

nitive dimension, consistent with the higher dispersion of λ^{c} relative to λ^{n}. Similar patterns hold in different sample periods and skill differences are large.

4.2 Firm Heterogeneity and the Distribution of Earnings

To assess the influences of firm heterogeneity on the structure of earnings, we recast the latter in terms of cross-sectional deviations. Denoting as $\widetilde{x}_j$ the deviation of variable x in firm j from its cross-sectional mean $\bar{x}$, we express the returns to skill bundle (c,n) in firm j as

$$\log(w_j(s)) = \underbrace{\overline{\lambda}^{\mathrm{c}} c + \overline{\lambda}^{\mathrm{n}} n}_{\substack{(1)\\ \text{Mincer returns}}} + \underbrace{\lambda_j^0}_{\substack{(2)\\ \text{firm intercepts}}} + \underbrace{\widetilde{\lambda}_j^{\mathrm{c}} \overline{c} + \widetilde{\lambda}_j^{\mathrm{n}} \overline{n}}_{\substack{(3)\\ \text{heterog.returns}}} + \underbrace{\widetilde{\lambda}_j^{\mathrm{c}} \widetilde{c} + \widetilde{\lambda}_j^{\mathrm{n}} \widetilde{n}}_{\substack{(4)\\ \text{match effects}}} . \tag{7}$$

This representation emphasizes the distinct layers of firm heterogeneity that contribute to earnings variation. Element 1 captures the homogeneous Mincerian returns to skills, which are estimated from standard survey data when direct skill measures are available. Element 3 captures the novel dimension of firm heterogeneity that we estimate.

If no skill measure is available but the data consists of matched employer–employee records, components 2 and 3 are conflated into the firm fixed effect, while the Mincerian variation in 1 is absorbed into the worker fixed effect. Finally, element 4 reflects genuine match-specific returns and can be identified if one has access to both employer–employee linked data and skill proxies. Matching induces a convexification of the earnings curve over the skill range whenever high-skilled workers sort more frequently into high-return firms.[17]

Counterfactual earning distributions. How does firm heterogeneity affect earning distributions? The left panel of Figure 3 plots the counterfactual earnings distributions that prevail when we separately allow for firm-specific heterogeneity only through intercepts (element 2) or skill returns (elements 3 plus 4) in equation (7). Heterogeneous returns generate a thicker right tail than heterogeneous intercepts, and push the average up. This suggests that conventional fixed

[17]See Proposition 6 in Lindenlaub (2017) for a discussion of these effects in a frictionless model.

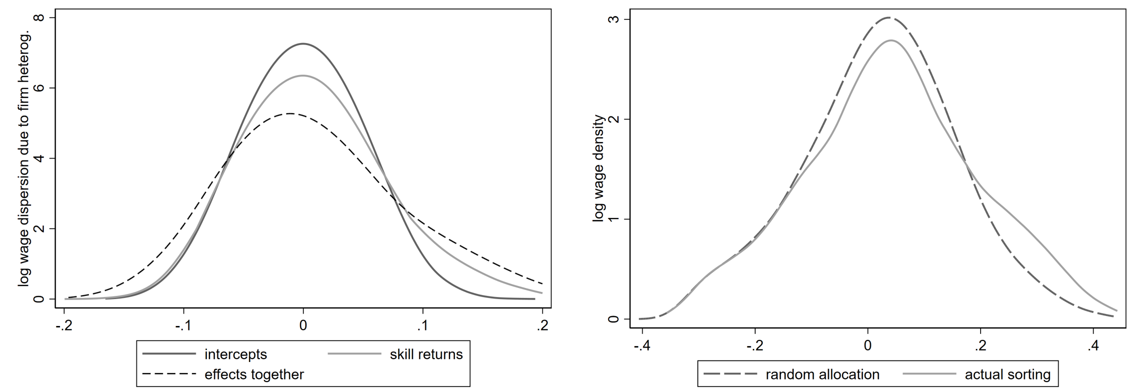


Figure 3: Firm heterogeneity and the distribution of earnings.

Notes: Left panel: Counterfactual distributions when only one layer of firm heterogeneity is active. $\mathrm{sd}(\lambda^0) = 0.033$; $\mathrm{sd}(\widetilde{\lambda}^{\mathrm{c}}\overline{c} + \widetilde{\lambda}^{\mathrm{n}}\overline{n}) = 0.042$; $\mathrm{sd}(\widetilde{\lambda}^{\mathrm{c}}\widetilde{c} + \widetilde{\lambda}^{\mathrm{n}}\widetilde{n}) = 0.020$; $\mathrm{sd}(\lambda^0 + \widetilde{\lambda}^{\mathrm{c}}\overline{c} + \widetilde{\lambda}^{\mathrm{n}}\overline{n} + \widetilde{\lambda}^{\mathrm{c}}\widetilde{c} + \widetilde{\lambda}^{\mathrm{n}}\widetilde{n}) = 0.068$.
Right panel: Earnings with random worker–firm matching versus actual distribution. The former is computed by sample-weighing every skill type in every firm according to their population shares. Estimation period: 1999–2008.

effects may underestimate the impact of firm heterogeneity. For comparison, the figure also plots the earnings distribution when all layers of firm heterogeneity (2, 3 and 4) are active.

The right panel in Figure 3 juxtaposes the counterfactual distribution under random assignment of workers to firms with the empirical earnings distribution. As expected, random allocation of workers implies less dispersion than what we observe in reality. The thicker right tail of the empirical distribution, due mostly to match effects, induces higher wages and productive efficiency. Mechanically, differences are less noticeable on the left tail where lower skills and returns combine into negligible interaction effects.

5 Discussion

We examine distinct, and little explored, dimensions of firm-level heterogeneity and present new evidence of worker–firm complementarities. The analysis relies on matched employer–employee population records in conjunction with direct measures of workers' cognitive and noncognitive skills.

We find that similar skills command very different returns across firms. This generates clear incentives for sorting. We document that, indeed, workers with higher endowments of cognitive and noncognitive skills populate firms with higher returns to those skills. 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 skilled workers to seek a better match. Moreover, consistent with theoretical restrictions derived in our many-to-one matching environment, higher skill dispersion within firms is associated to wider skill gaps between firms. Overall, there is strong evidence of complementarities between workers' heterogeneous skills and firms' heterogeneous returns, and of positive assortative matching.

These results have potentially important implications. First, we show that the wage distribution becomes more unequal because of matching. At the same time, allocative efficiency improves and average wages rise. In ongoing work, we also find that firms with high skill returns engage in more product and process innovation and have higher value added per worker. This lends further support to the presence of complementarities in production. It also implies that mismatch between skills and firms may hamper innovation and productivity.

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

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.

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

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," 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.

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.

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

Current draft: November 15, 2020

Contents

A A Labor Market Model with Two-Sided Heterogeneity and Multiple 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.

A.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

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

where ϵ_{ij} captures an idiosyncratic preference for working at firm j. We assume that shocks ϵ_{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

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

Equation (A.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

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

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.[1] Firm j's total output is

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

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

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

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:

[1] 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.

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

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

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

A.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). In view of the ordinal nature of the skill measures used in the descriptive data analysis, we categorize skill bundles 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 noncognitive level. For example, a worker of type $s = (H, l)$ has high cognitive and low noncognitive ability. The wage premium associated to skill bundle (c,n) in firm j is

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

for all $(c,n) \in S$. 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 technology of firm j as $y_j = T_j \sum_{s \in S} e^{\Delta_j(s)} q_j(s)$ 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:

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

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 s rescales marginal revenues proportionally to the firm's skill return $\Delta_j(s)$.

In log form, the equilibrium wage lends theoretical underpinning to empirical specifications like equation (1) in the paper. That is:

$$\log(w_j(s)) = \alpha + \theta_j + \Delta_j(s). \tag{A.10}$$

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

Optimal behavior implies that firms with higher returns to s-type skills tend to hire a larger share of s-type workers.[2] This observation suggests that firms with similar returns to a skill bundle can be grouped together based on their share of workers with that particular bundle, and that firm-specific returns to different skill bundles can be identified from a cross-section of

[2]That is, for two firms denoted as j and k, the following holds

$$\mathbb{E}\left[\log\left(\frac{q_j(s)}{q_j(L,l)}\right) - \log\left(\frac{q_k(s)}{q_k(L,l)}\right) |\Delta_j(s), \Delta_k(s)\right] = \beta(\Delta_j(s) - \Delta_k(s)).$$

grouped worker wages if direct skill measures are available, corroborating the approach to the estimation of skill type premia in Section B.2.[3]

[3]Another implication of firms' optimal hiring behaviour is that the total number of workers employed by a firm j grows with the base wage θ_j. This follows from observing that $\mathbb{E}\left[\log(q_j(L,l)) - \log(q_k(L,l))|\theta_j,\theta_k\right] = \beta(\theta_j - \theta_k)$.

B Firm-Level Wage Premia by Skill Type: Estimates

B.1 Data

Sample construction. 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.

The main measure of wage compensation is annual labor income from the employer with highest recorded earnings. Earnings are not top-coded and include end-of-year bonuses and performance pay. In robustness checks, we used workers' full-time equivalent monthly wages, available through a survey of a subset of private sector firms (*Lönestrukturstatistik*), and we found similar results.

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 wage 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×year observations for the main sample period of 1999–2008.[4] Table B.1 reports summary statistics for the final estimation sample.

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

Table B.1: Summary statistics for the main estimation sample

	mean	sd	p5	p25	p50	p75	p95
Real wages ('000 kr)	323.3	237.5	79.5	231.2	294.2	373.5	621.3
Log (wages)	5.6	0.6	4.4	5.4	5.7	5.9	6.4
Age	37.3	9.2	22	30	37	45	52
Cognitive	5.3	1.9	2	4	5	7	8
Noncognitive	5.1	1.7	2	4	5	6	8
Years of schooling	12.2	2.4	9	10.5	12	13.5	16
Observations	6610567						

Notes: Summary statistics for main estimation sample: 1999–2008; males aged 20–60 with nonmissing employer, earnings, and test scores; employers exist at least five years, with at least ten sample workers on average. Wages are real annual labor earnings in 2008 Swedish kronor. Cognitive and noncognitive scores are in STANINE scale. Sample: 1,088,415 unique workers and 25,839 firms.

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. 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.

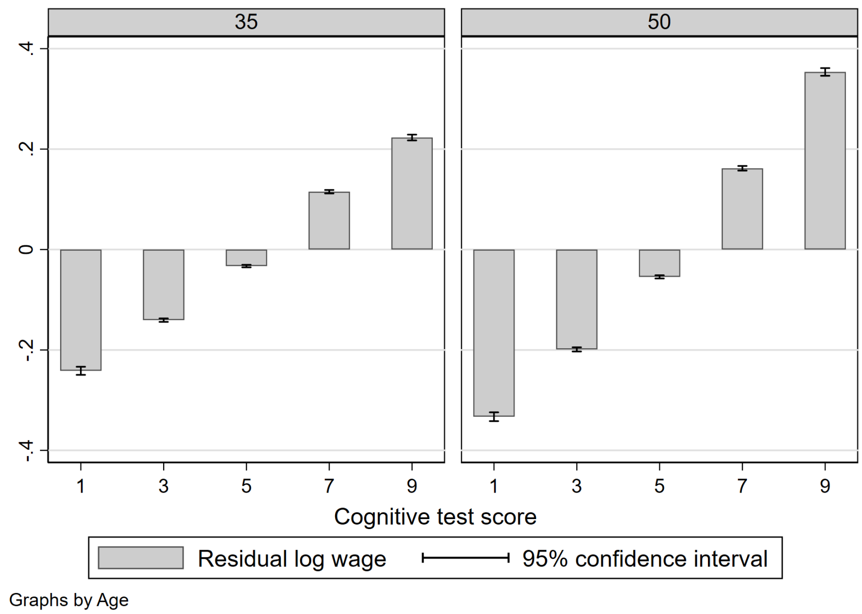
One might worry that certain individuals could deliberately perform badly on these tests to avoid military service. There are, however, several observations suggesting this was not a major problem. In particular, we emphasize that employers routinely put considerable weight on military service performance 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 wages 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).

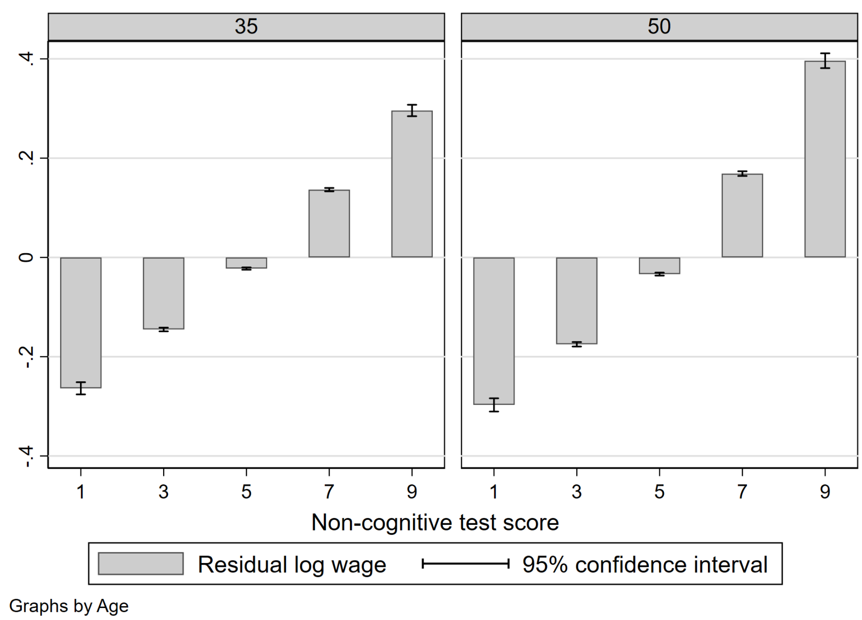
Noncognitive ability is assessed through a 25-minute semi-structured interview by a certified psychologist. Individuals are graded on, among others, their willingness to assume responsibility, independence, outgoing character traits, persistence, emotional stability, and power of initiative (Swedish 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.[5]

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 B.1, the tests are strongly associated

[5]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.



(a) Cognitives



(b) Noncognitives

Figure B.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×year dummy interactions. Sample period: 1990–2017. 95% confidence intervals indicated by brackets.

to labor market outcomes and accurately predict earnings several decades later. Figure B.1 compares the earnings of workers with different STANINE scores in our sample (residualized using full age×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 attributes, resulting in rank measures that can be mapped into earnings for the whole population of interest over its working life cycle.

B.2 Estimation of Skill-Bundle Premia across Firms

Estimates using employment of (M, m) as base type. The interpretation of skill premia is most natural when (L, l) is set as the base skill type, like we do in the main text. However, (L, l) skill bundles are less frequent across firms than (M, m).[6] Grouping based on employment shares relative to the (L, l) base type can, therefore, lead to lower estimation precision for firms that hire few or no (L, l)-type workers during the estimation period. Since almost all firms in our sample employ (M, m) workers during 1999–2008, we estimate (1), without loss of generality, using the specification

$$\log(w_{ijt}) = \theta_j + \mu_i + \sum_{s_i \in S \setminus \{(M,m)\}} \Delta_j(s) \mathbb{1}[s_i = s] + \boldsymbol{X}_{it} \boldsymbol{b}_t + \varepsilon_{ijt}. \tag{B.1}$$

where (M, m) is the base type. These estimates are displayed in Table B.2. The observation counts indicate the number of firms that hire at least one worker of the respective type.[7] In our sample of firms (existing for at least five years with on average at least ten workers), all have at least one

[6]Given the STANINE scale and correlation of 0.3 between the cognitive and noncognitive test scores, the frequencies of skill types are: $(L, l) = 6.7\%$, $(L, m) = 10.3\%$, $(L, h) = 1.2\%$, $(M, l) = 8.0\%$, $(M, m) = 35.9\%$, $(M, h) = 10.8\%$, $(H, l) = 2.1\%$, $(H, m) = 15.2\%$, $(H, h) = 9.7\%$.

[7]Firms that hire none of a given type are in the bottom grouping quintile, with the exception of (H, l) and (L, h) where firms that hire none of those types reach up into the second-lowest quintile.

(M, m) worker during the 1999-2008 period, with employment of skill types in each individual firm varying significantly. This is not a sample restriction and corroborates our practical choice of base skill type.

The bundled skill premia reported in Table 1 of the main paper are a re-normalization based on the premia $\Delta_j(L, l)$. In particular, for each firm we set $\Delta_j(M, m) = -\widehat{\Delta}_j(L, l)$ and, for $s \notin \{(L, l), (M, m)\}$, $\Delta_j(s) = \widehat{\Delta}_j(s) - \widehat{\Delta}_j(L, l)$. The base wage is $\theta_j = \widehat{\theta}_j + \widehat{\Delta}_j(L, l)$. The cross-sectional average of the base wage is de-meaned to zero accounting for firm-specific employment weights.

For ease of interpretation, we display the premia normalized by the $\Delta_j(L, l)$ as base type in the main text. However, the empirical results of the cross-sectional analysis are not sensitive to the change of base type. All the estimates in Table B.2 are robust: wage premia rise with the skill rank within each type, e.g., $\bar{\Delta}(H, l) = -0.06$, $\bar{\Delta}(H, m) = 0.05$, and $\bar{\Delta}(H, h) = 0.20$. The premia are all substantial: for example, the difference between the premium for the highest (H, h) and lowest (L, l) type is on average 46 log points and it can become substantially higher in certain firms. The heterogeneity of skill premia across firms is also substantial, with the dispersion for each single skill bundle return across firms anywhere between 1/3 and 5/6 of the dispersion of base wages.

Details about grouping. To reduce dimensionality, we rank firms into quintiles according to their relative hiring of each skill type. The quintiles correspond to categorical variables $g_s \in \{1, 2, 3, 4, 5\}$ where s is a skill type. The value of each categorical variable denotes the rank of the firm in that particular skill type intensity. This empirical approach is consistent with the theoretical result that the employment of type s relative to the base skill rises with the firm-specific skill premium $\Delta_j(s)$. That is, firms with higher returns to s-type skills tend to employ a larger share of s-type workers. Since employment of workers with the base skill (M, m) grows with the intercept term θ_j, as shown in the model Section A above, we also group firms into quintiles according to their total employment of the base skill (M, m). These base-type (firm-specific intercept) groups are denoted by the categorical variable $g = g_\theta$.

Table B.2: Firms-specific type premia: raw estimates with (M, m) base skill

	(sample period:1999–2008)							
	count	mean	sd	p5	p25	p50	p75	p95
$\widehat{\theta_j}$	25,604	-0.00	0.06	-0.09	-0.05	-0.01	0.04	0.12
$\widehat{\Delta}_j(L,l)$	19,416	-0.26	0.04	-0.37	-0.29	-0.23	-0.22	-0.22
$\widehat{\Delta}_j(L,m)$	22,028	-0.16	0.03	-0.23	-0.18	-0.15	-0.13	-0.13
$\widehat{\Delta}_j(L,h)$	10,773	-0.08	0.04	-0.18	-0.14	-0.05	-0.05	-0.05
$\widehat{\Delta}_j(M,l)$	22,673	-0.17	0.03	-0.22	-0.18	-0.17	-0.15	-0.15
$\widehat{\Delta}_j(M,h)$	24,015	0.03	0.02	-0.01	0.02	0.03	0.06	0.06
$\widehat{\Delta}_j(H,l)$	13,655	-0.06	0.05	-0.14	-0.09	-0.05	-0.01	-0.01
$\widehat{\Delta}_j(H,m)$	23,794	0.05	0.03	0.01	0.04	0.05	0.07	0.08
$\widehat{\Delta}_j(H,h)$	21,115	0.20	0.03	0.15	0.19	0.20	0.23	0.23

Notes: The table shows the distributions of estimated base wages and skill premia across firms. The sample covers 25,604 unique firms between years 1999 and 2008.

The firm-specific estimates from the implementation of the grouping estimator in (B.1) result in five (one for each quintile) values for $\theta_{g_\theta(j)}$ and for the different $\Delta_{g_s(j)}$.[8] We refer to them by their shorthand θ_j and $\Delta_j(s)$ throughout the paper. Group assignments for different skill types are, by design, flexible and independent of one another. For example, two firms in the same quintile of the (H, h) type need not be in the same g_θ or g_s bins for the other skill types. This is the least restrictive assumption, since each skill type's relative employment is a sufficient statistic for grouping firms into bins with similar $\Delta_j(s)$. In the next section we verify that estimates of $\Delta_j(s)$ do, in fact, rise with the relative hiring of the corresponding skill type s.

AKM-based variance decomposition. Table B.3 reports results from a wage-variance accounting exercise based on estimation of (B.1) for the 1999–2008 period. We denote worker-only components as $\alpha_{it} \equiv \mu_i + \boldsymbol{X}_{it}\boldsymbol{b}_t$ and worker/firm components as $\psi_{ij} \equiv \theta_j + \sum_{s_i} \Delta_j(s)\mathbb{1}[s_i = s]$. As often found, worker-specific heterogeneity accounts for most of total wage variation. The second largest contribution comes from firm-worker terms, which account for six percent of wage variation. The covariance between α and ψ components makes up only five percent of total variation,

[8] Estimates of firm-specific wage intercepts are the sum of $\theta_{g_\theta(j)}$ and of the base (noninteracted) return for each g_s group (the $\Delta_{g_s(j)}$ are the groups' interaction effects with s). Therefore, the base wage captures more than just base-type rank and is indeed more accurately denoted by θ_j than $\theta_{g_\theta(j)}$.

partly because our interacted specification directly accounts for much of worker/firm complementarity (accordingly, their joint contribution obtained by adding up columns 2 and 3 is relatively large, around 10%). Results are similar for the 1990–1999 and 2008–2017 estimation periods, and they are comparable to grouping-based AKM implementations for Sweden (Bonhomme et al., 2019) and the U.S. (Lamadon et al., 2019).

Table B.3: Decomposition of log earnings variance

Variance decomposition (in %)			
(1)	(2)	(3)	(4)
$\frac{\text{Var}(\alpha)}{\text{Var}(\log(w))}$	$\frac{\text{Var}(\psi)}{\text{Var}(\log(w))}$	$\frac{2\text{cov}(\alpha,\psi)}{\text{Var}(\log(w))}$	$\frac{\text{Var}(\varepsilon)}{\text{Var}(\log(w))}$
63.7	6.0	4.9	25.4

Notes: Decomposition of the log earnings variance based on estimates from specification (B.1). We subsume worker-only contributions in $\alpha_{it} \equiv \mu_i + \boldsymbol{X}_{it}\boldsymbol{b}_t$ and worker/firm contributions in $\psi_{ij} \equiv \theta_j + \sum_{s_i} \Delta_j(s)\mathbb{1}[s_i = \{s\}]$. Estimation period: 1999–2008.

B.3 Nonparametric Wage Premia

Figure B.2 plots estimates of skill-bundle premia taken from Table B.2 and ordered by their quintile rank. As discussed above, there are five parameter estimates for each skill bundle (four in the case of (H, l) and (L, h) where the bottom quintile of firms does not hire any of those types). It is reassuring that skill premia are consistently growing along the rank order within each type, e.g., $\widehat{\Delta}(L, l) < \widehat{\Delta}(M, l) < \widehat{\Delta}(H, l)$. There is no way of ranking bundles with inverted order attributes and, indeed, we find that $\widehat{\Delta}(M, l)$ and $\widehat{\Delta}(L, m)$ do cross over the quintile range.

Sorting and skill-bundle premia. Higher-ranked skill bundles command bigger premia in the vertical dimension of Figure B.2. In addition, larger wage premia are positively associated to the relative employment along quintile groups, and often quite substantially. Figure B.3 shows this by using separate box plots of worker shares together with wage premia, over the range of

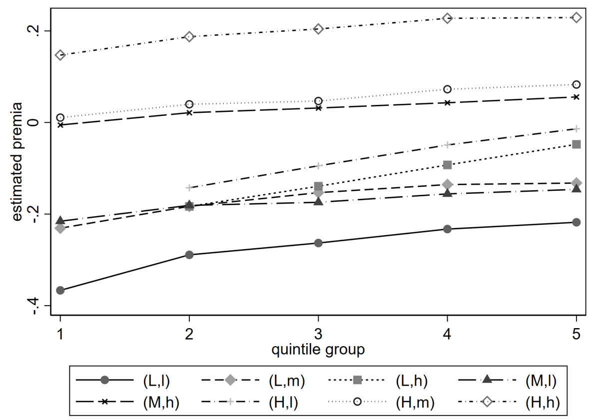


Figure B.2: Estimated skill premia $\Delta_j(s)$ by quintile group g_s.

Notes: Skill premia from the estimation of equation (B.1), relative to the omitted (M,m) base wage, by skill type's quintile group. Estimation period: 1999–2008.

relative employment quintiles. Box plots show the share of each skill type relative to firm's total employment, lending empirical support to the theoretical restriction of the model in Section A and suggesting positive rank-correlation between relative employment of s-type workers and the premium $\widehat{\Delta}(s)$ paid to them. Such correlation appears to be robust, with most of the differences highly significant. This finding corroborates the grouping approach adopted in the estimation of skill bundle premia, and also anticipates the sorting results of Section 4. That is, the nonparametric estimates of skill-bundle premia confirm that workers consistently sort into firms where the wage premium for their type is higher.

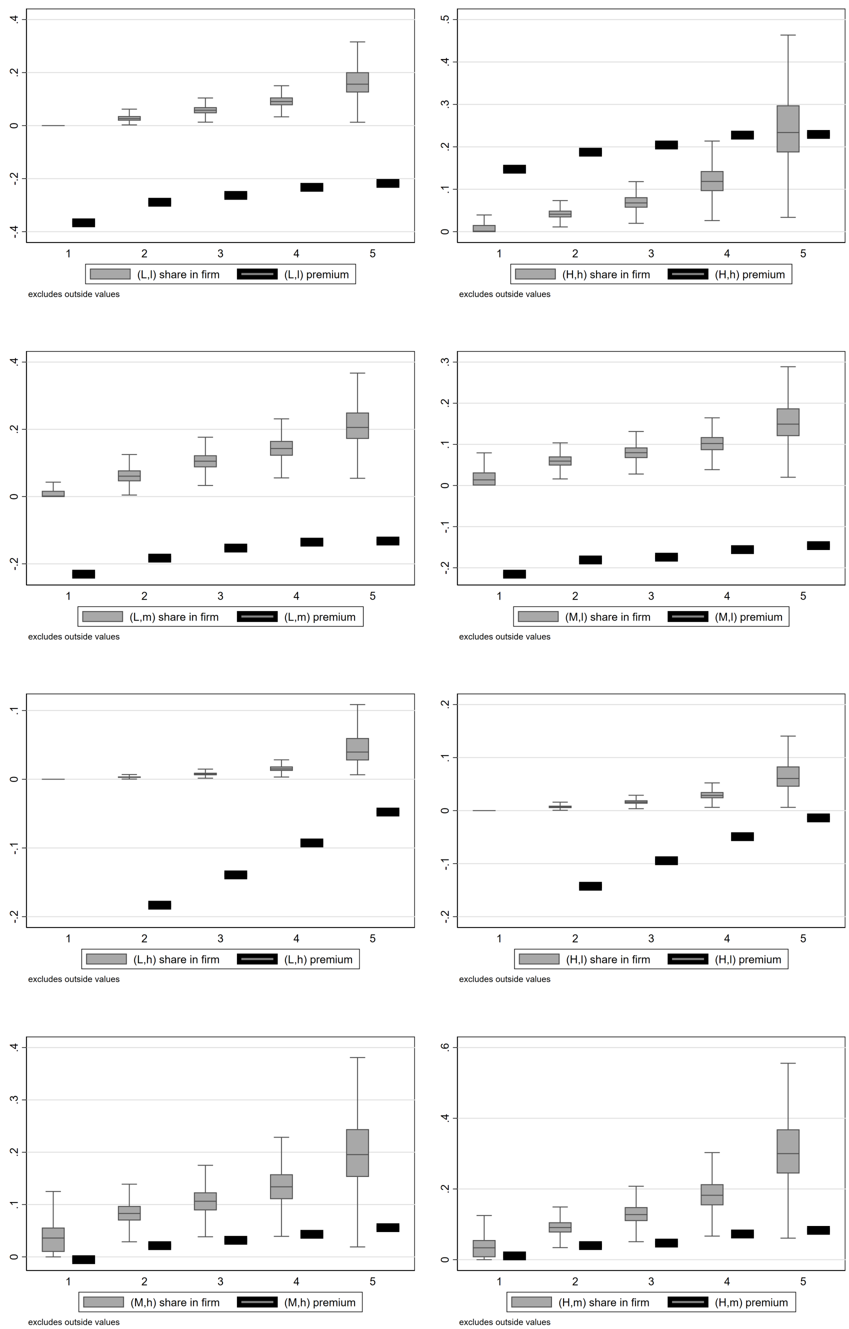


Figure B.3: Estimated skill premia ($\Delta_j(s)$) and firms' employment shares of different skill types (s), by quintile group

Notes: Wage premia of different skill types (relative to (M,m)) by quintile group. Wage premia are unique values but bars plotted thicker for visibility. Premia generally increase with quintile rank (i.e. they grow with each type's relative employment). Premia tend to be negative for the (L,l), (L,m), (M,l), (L,h), and (H,l) bundles; and tend to be positive for (H,h), (H,m), and (M,h). Box plots show the share of each skill type relative to firm's total employment. Therefore, box plots can overlap across quintiles.

C Skill Returns vs Endowments: Identification and Estimation

C.1 Moment Minimization Procedure

We restate the GMM optimization problem in the paper's eqt. (3) as:

$$\min_{\substack{\{\lambda^0_j,\lambda^c_j,\lambda^n_j\}_{j=1}^J \\ L,M,H \\ l,m,h}} \quad \sum_{j=1}^{J} \sum_{(c,n)\in S} \left[\lambda^0_j + \lambda^c_j c + \lambda^n_j n - \log(w_j(c,n))\right]^2$$

$$\text{s.t.} \qquad S = \{(c,n) |\ c \in \{L,M,H\}, n \in \{l,m,h\}\}$$

Taking first order conditions we get:

$$\text{FOC}[\lambda^0_k]: \qquad \sum_{(c,n)\in S} \left[\lambda^0_k + \lambda^c_k c + \lambda^n_k n - \log(w_k(c,n))\right] = 0$$

$$\text{FOC}[\lambda^c_k]: \qquad \sum_{(c,n)\in S} c\left[\lambda^0_k + \lambda^c_k c + \lambda^n_k n - \log(w_k(c,n))\right] = 0$$

$$\text{FOC}[\lambda^n_k]: \qquad \sum_{(c,n)\in S} n\left[\lambda^0_k + \lambda^c_k c + \lambda^n_k n - \log(w_k(c,n))\right] = 0$$

$$\text{FOC}[c]: \qquad \sum_{j=1}^{J} \sum_{n\in\{l,m,h\}} \lambda^c_j \left[\lambda^0_j + \lambda^c_j c + \lambda^n_j n - \log(w_j(c,n))\right] = 0$$

$$\text{FOC}[n]: \qquad \sum_{j=1}^{J} \sum_{c\in\{L,M,H\}} \lambda^n_j \left[\lambda^0_j + \lambda^c_j c + \lambda^n_j n - \log(w_j(c,n))\right] = 0$$

Using the first three FOCs and conditional on values of skill levels, we can solve for returns as follows:

$$\begin{bmatrix} \lambda^0_k \\ \lambda^c_k \\ \lambda^n_k \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{3}\sum\limits_{C\in\{L,M,H\}} C & \frac{1}{3}\sum\limits_{N\in\{l,m,h\}} N \\ \frac{1}{3}\sum\limits_{C\in\{L,M,H\}} C & \frac{1}{3}\sum\limits_{C\in\{L,M,H\}} C^2 & \frac{1}{9}\sum\limits_{(C,N)\in S} C.N \\ \frac{1}{3}\sum\limits_{N\in\{l,m,h\}} N & \frac{1}{9}\sum\limits_{(C,N)\in S} C.N & \frac{1}{3}\sum\limits_{N\in\{l,m,h\}} N^2 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{9}\sum\limits_{(C,N)\in S} \log(w_k(C,N)) \\ \frac{1}{9}\sum\limits_{(C,N)\in S} C.\log(w_k(C,N)) \\ \frac{1}{9}\sum\limits_{(C,N)\in S} N.\log(w_k(C,N)) \end{bmatrix} \tag{C.1}$$

And conditional on marginal returns we can solve for skill levels as follows:

$$\begin{bmatrix} L \\ M \\ H \\ l \\ m \\ h \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{J}\sum_{j=1}^{J}(\lambda_j^{\mathrm{c}})^2\right)\mathbf{I}_{3\times 3} & \left(\frac{1}{3J}\sum_{j=1}^{J}\lambda_j^{\mathrm{c}}\lambda_j^{\mathrm{n}}\right)\mathbf{1}_{3\times 3} \\ \left(\frac{1}{3J}\sum_{j=1}^{J}\lambda_j^{\mathrm{c}}\lambda_j^{\mathrm{n}}\right)\mathbf{1}_{3\times 3} & \left(\frac{1}{J}\sum_{j=1}^{J}(\lambda_j^{\mathrm{n}})^2\right)\mathbf{I}_{3\times 3} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{3J}\sum_{j=1}^{J}\sum_{N\in\{l,m,h\}}\lambda_j^{\mathrm{c}}\log(w_j(L,N)) \\ \frac{1}{3J}\sum_{j=1}^{J}\sum_{N\in\{l,m,h\}}\lambda_j^{\mathrm{c}}\log(w_j(M,N)) \\ \frac{1}{3J}\sum_{j=1}^{J}\sum_{N\in\{l,m,h\}}\lambda_j^{\mathrm{c}}\log(w_j(H,N)) \\ \frac{1}{3J}\sum_{j=1}^{J}\sum_{C\in\{L,M,H\}}\lambda_j^{\mathrm{n}}\log(w_j(C,l)) \\ \frac{1}{3J}\sum_{j=1}^{J}\sum_{C\in\{L,M,H\}}\lambda_j^{\mathrm{n}}\log(w_j(C,m)) \\ \frac{1}{3J}\sum_{j=1}^{J}\sum_{C\in\{L,M,H\}}\lambda_j^{\mathrm{n}}\log(w_j(C,h)) \end{bmatrix}, \tag{C.2}$$

where $\mathbf{I}_{n\times n}$ denotes the n by n identity matrix and $\mathbf{1}_{m\times n}$ denotes the m by n matrix whose elements are all equal to one. The optimization in (3) has infinitely many solutions, and we can only identify skill values and their marginal returns up to an affine transformation. This requires a scale assumption; therefore, we impose that skill units are measured between zero and 1 (we normalize lower and upper bounds of each ability to zero and one, respectively; that is $L = l = 0$ and $H = h = 1$). Numerical minimization of the objective function is achieved iteratively: starting with a guess for M and m (since the other skill levels are already normalized to either zero or one) we use (C.1) to estimate firm specific returns i.e. the lambdas. Next, we use (C.2) and estimated returns to update our estimates of M and m. We iterate this procedure until it converges to the optimal solution. 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, the procedure converges to this optimum.

Our approach delivers estimates of intermediate skill levels within bounded cognitive and noncognitive ranges. Estimates of skill nodes play a key role in matching variation across the whole range of skill-bundle premia, conditional on the marginal returns. It is worth emphasizing two things: (1) the number of intermediate nodes estimated depends on the number of skill bundles that one targets, as pair-wise differences between nodes capture the skill endowment gaps between stanine group categories; (2) while marginal skill returns are, by design, the loadings necessary to account for between firm variation in skill bundle premia, the skill nodes serve the purpose of holding skill endowments constant across firms when estimating marginal skill returns. This identification argument is reflected in the recursive estimation approach, where we compute estimates of skill returns and node locations conditional on each other, until convergence is achieved. Estimation delivers a set of skill nodes that are common to all firms. These nodes, in conjunction with the firms specific marginal returns, are sufficient to span the whole distribution of skill-bundle premia.

C.2 A Graphical Illustration

To separately estimate skill endowments and firm-specific returns we adopt an iterative procedure whereby we target the nonparametric skill-bundle premia $\Delta_j(s)$. Figure C.1 illustrates this approach, and the underlying identification argument, for the simple case of one-dimensional heterogeneity in cognitive skills and returns.[9]

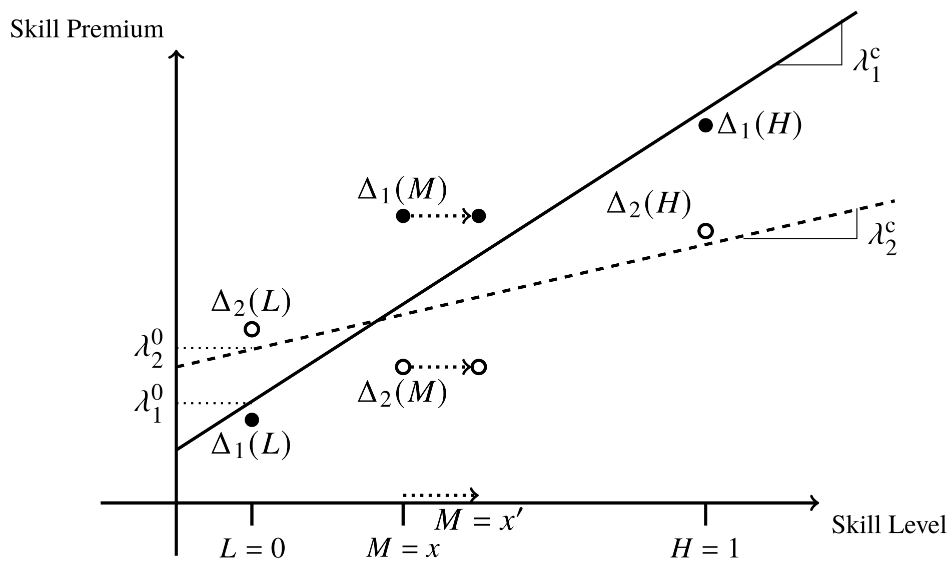


Figure C.1: Identification of returns vs endowments: a graphical example (c only)

In each iteration of this procedure, we first condition on the skill nodes (L, M, and H) and estimate firm-specific intercept λ_j^0 and gradient λ_j^c by least squares projections of empirical skill premia $\Delta_j(s)$ on skill endowments. This is done for all firms separately. In figure C.1 we illustrate

[9]The procedure and intuition with two sets of skills, and returns, is the same. We present the one-dimensional example because it is easily explained through a simple graph.

this for two firms, where we show one OLS line for each firm. In the next step, conditional on estimates of firm-specific returns, we search for the endowment values that minimize the sum of squared errors across *all* firms, simultaneously. Given the normalization $L = 0$ and ($H = 1$), in our illustrative example we choose the value M to minimize the sum of squared vertical distances of all firm premia from the firm-specific regression line, estimated in the previous step. These two steps are repeated until convergence to the global minimum. Given the convexity of the objective function, a partial gradient descent algorithm always converges to the global minimum.

D Matching Patterns and Inequality

D.1 Proof of Proposition 1

$$
\begin{aligned}
\frac{\mathrm{d}\overline{c}_j}{\mathrm{d}\lambda_j^{\mathrm{c}}} &= \frac{\displaystyle\int \beta\, c^2\; h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}{\displaystyle\int h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)} - \beta\left[\frac{\displaystyle\int c\; h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}{\displaystyle\int h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}\right]^2 \\
&= \beta\int c^2\; \mathrm{d}M_j(c,n) - \beta\left[\int c\; \mathrm{d}M_j(c,n)\right]^2 \\
&= \beta \cdot \mathrm{var}_{M_j}[c]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}\overline{c}_j}{\mathrm{d}\lambda_j^{\mathrm{n}}} &= \frac{\displaystyle\int \beta\, cn\; h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}{\displaystyle\int h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)} \\
&\quad - \beta\frac{\displaystyle\int ch(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}{\displaystyle\int h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)} \times \frac{\displaystyle\int nh(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} n+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)}{\displaystyle\int h(c,n)\; e^{\beta(\lambda_j^{\mathrm{c}} c+\lambda_j^{\mathrm{n}} n)}\; \mathrm{d}G(c,n)} \\
&= \beta\int cn\; \mathrm{d}M_j(c,n) - \beta\int c\; \mathrm{d}M_j(c,n) \times \int n\; \mathrm{d}M_j(c,n) \\
&= \beta \cdot \mathrm{cov}_{M_j}[c,n]
\end{aligned}
$$

Similarly we can show $\frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{n}}} = \beta \cdot \mathrm{var}_{M_j}[n]$ and $\frac{\partial \overline{n}_j}{\partial \lambda_j^{\mathrm{c}}} = \beta \cdot \mathrm{cov}_{M_j}[c,n]$.

D.2 Generalization of Proposition 1

Proposition 1 links the first derivatives of the matching function $\mu(\lambda_j^{\mathrm{c}}, \lambda_j^{\mathrm{n}}) = (\overline{c}_j, \overline{n}_j)$, defined in the paper, to the second moments of the skill distribution within firms. That proposition is, in fact, a special case of a more general statement. The following Proposition D.1 describes the general relationship between the higher order derivatives of the matching function and the within-firm skill distribution of cognitive and non cognitive skills M_j, defined in the paper.

Proposition D.1. *For any integer pair $(u, v) \geq 0$, where $u + v \geq 1$, the following relationships describe the moments of the within-firm distributions of cognitive skills c and noncognitive skills* *n:*

$$\frac{\mathrm{d}^{u+v}\ \overline{c}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^u\ \mathrm{d}(\lambda_j^{\mathrm{n}})^v} = \beta^{u+v}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v\right] \tag{D.1}$$

$$\frac{\mathrm{d}^{u+v}\ \overline{n}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^u\ \mathrm{d}(\lambda_j^{\mathrm{n}})^v} = \beta^{u+v}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u}(n-\overline{n}_j)^{v+1}\right] \tag{D.2}$$

where the expectations are taken using the within-firm skill distribution M_j of firm j, as defined in the paper.

Proof. Here we only show D.1 by induction and the proof for D.2 is similar. Proposition 1 already establishes that D.1 holds for $(u,v)=(1,0)$ and $(u,v)=(0,1)$. It only remains to show that if D.1 holds for (u,v) then it has to hold for $(u+1,v)$ and $(u,v+1)$. First for $(u+1,v)$ we have:

$$\begin{aligned}
\frac{\mathrm{d}^{(u+1)+v}\ \overline{c}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^{u+1}\ \mathrm{d}(\lambda_j^{\mathrm{n}})^v} &= \frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{c}}}\ \frac{\mathrm{d}^{u+v}\ \overline{c}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^u\ \mathrm{d}(\lambda_j^{\mathrm{n}})^v} \\
&= \frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{c}}}\ \beta^{u+v}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v\right] \\
&= \beta^{u+v}\frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{c}}}\ \frac{\displaystyle\int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)} \\
&= \beta^{u+v}\left\{\frac{\displaystyle\int \beta c(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right. \\
&\quad\left. - \frac{\displaystyle\int \beta c h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)\ \int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\left[\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)\right]^2}\right\} \\
&= \beta^{u+1+v}\left\{\frac{\displaystyle\int c(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right. \\
&\quad\left. - \frac{\displaystyle\int c h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\frac{\displaystyle\int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\displaystyle\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right\} \\
&= \beta^{u+1+v}\left\{\mathbf{E}_{M_j}\left[c(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v\right] - \overline{c}_j\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v\right]\right\} \\
&= \beta^{u+1+v}\mathbf{E}_{M_j}\left[c(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v - \overline{c}_j(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^v\right]
\end{aligned}$$

$$= \beta^{(u+1)+v}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{(u+1)+1}(n-\overline{n}_j)^{v}\right]$$

Now for $(u, v+1)$ we have:

$$\begin{aligned}
\frac{\mathrm{d}^{u+(v+1)}\ \overline{c}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^u\ \mathrm{d}(\lambda_j^{\mathrm{n}})^{v+1}} &= \frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{n}}}\ \frac{\mathrm{d}^{u+v}\ \overline{c}_j}{\mathrm{d}(\lambda_j^{\mathrm{c}})^u\ \mathrm{d}(\lambda_j^{\mathrm{n}})^{v}} \\
&= \frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{n}}}\ \beta^{u+v}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}\right] \\
&= \beta^{u+v}\frac{\mathrm{d}}{\mathrm{d}\lambda_j^{\mathrm{n}}}\ \frac{\int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)} \\
&= \beta^{u+v}\left\{\frac{\int \beta n(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right. \\
&\quad\left. - \frac{\int \beta n h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)\ \int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\left[\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)\right]^2}\right\} \\
&= \beta^{u+v+1}\left\{\frac{\int n(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right. \\
&\quad\left. - \frac{\int n h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\frac{\int (c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}{\int h(c,n)e^{\beta(\lambda_j^{\mathrm{c}}c+\lambda_j^{\mathrm{n}}n)}\mathrm{d}G(c,n)}\right\} \\
&= \beta^{u+v+1}\left\{\mathbf{E}_{M_j}\left[n(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}\right] - \overline{n}_j\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}\right]\right\} \\
&= \beta^{u+v+1}\mathbf{E}_{M_j}\left[n(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v} - \overline{n}_j(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{v}\right] \\
&= \beta^{u+(v+1)}\mathbf{E}_{M_j}\left[(c-\overline{c}_j)^{u+1}(n-\overline{n}_j)^{(v+1)}\right]
\end{aligned}$$

□

D.3 First-Order Stochastic Dominance

A corollary of PAM is that, for each skill attribute, the distribution of higher-skilled workers over firm returns should (first-order) stochastically dominate that of lower-skilled workers. For example, holding constant non cognitive traits n, the share of employees with higher cognitive traits c should rise with a firm's λ^{c}. Equivalently, the frequency of higher n workers should increase with λ^{n}, holding c constant. Formally, positive FOSD implies:

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

In Figure D.1 we examine whether such patterns occur in our sample. Using estimates of skill nodes and returns, we separately plot the λ^{c} and λ^{n} cumulative distribution functions of workers in different skill groups. The left panels show that, holding n constant, the CDF of λ^{c} shifts to the right when we consider higher endowments of c. A similar finding emerges when looking at the CDF of noncognitive endowments (n) over λ^{n} (right panels). All else equal, variation in each skill dimension is consistent with stochastic dominance over the distribution of the corresponding firm returns. This supports the PAM hypothesis and further validates the bilinear matching-return function.

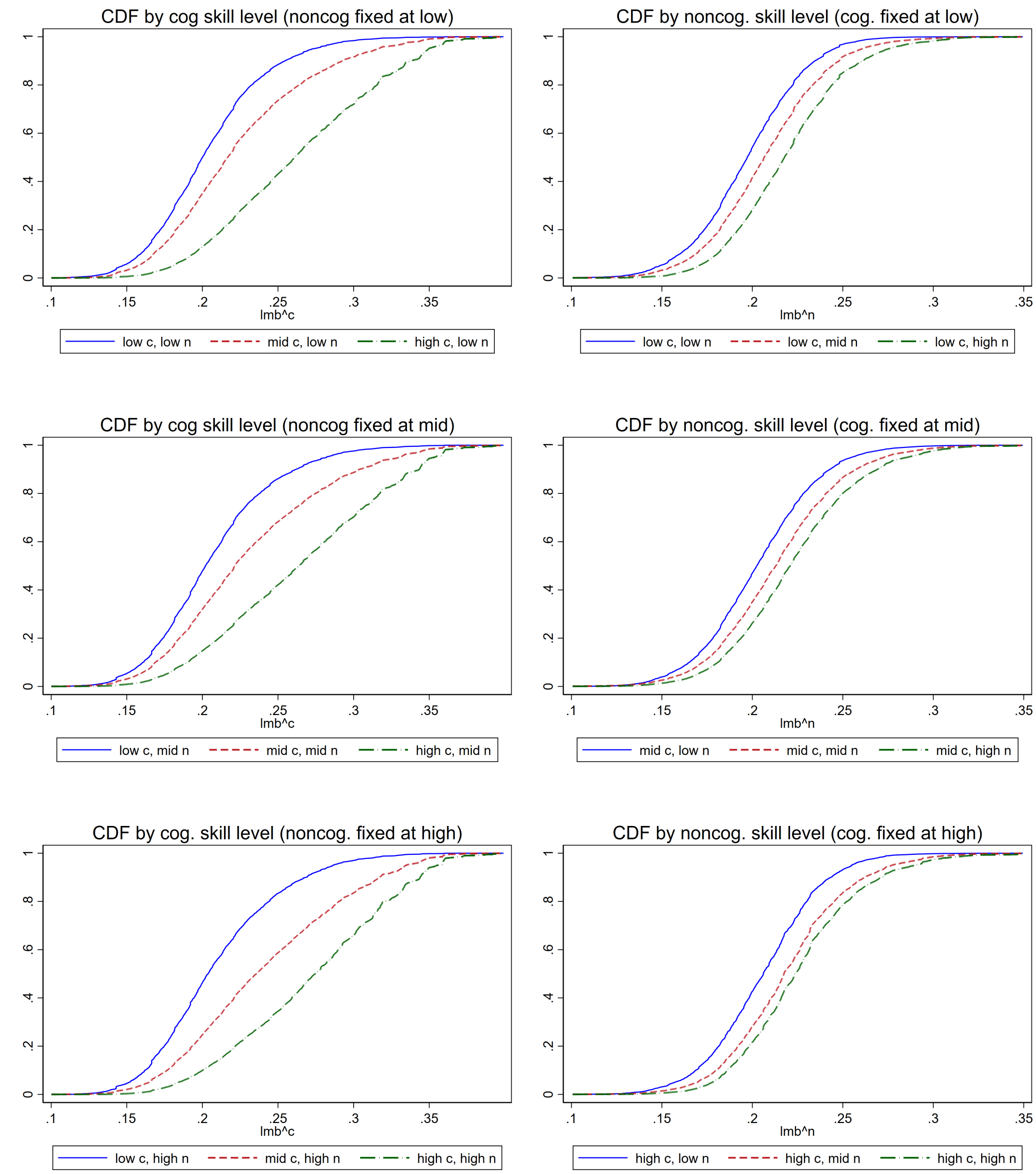


Figure D.1: First-order stochastic dominance. Left panels: CDF of cognitive (c) traits over returns λ^{c}. Right panels: CDF of noncognitive (n) traits over returns λ^{n}. Sample period: 1999–2008.

Notes: Top left panel plots cumulative distribution functions (CDF) of three sets of workers ranked by their cognitive rank (low C, mid C, high C) over the range of estimated firm returns λ^{c}. The CDF of high C workers stochastically dominates that of middle C workers; the latter dominates that of low C workers. In the top left panel we condition on the lowest non cognitive traits N. Moving down the panels on the left, the CDFs hold N attributes at middle and high ranks, respectively. The right panels show similar results for the CDFs of low, middle and high N over returns λ^{n}, holding cognitive traits fixed.

D.4 Skill Variation Between and Within Firms

Table D.1: Jacobian Estimates: Quintile interactions.

	(1)		(2)		(3)	
	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$	$\overline{c}_j$	$\overline{n}_j$
λ_j^{c}	1.36	0.42	2.12	0.80	1.59	0.47
	(.02)	(.01)	(.01)	(.01)	(.02)	(.01)
λ_j^{n}	0.31	0.71	0.83	1.61	0.25	0.91
	(.03)	(.02)	(.01)	(.01)	(.02)	(.02)
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[c] \in (p20, p40]\right\}$	0.48				0.40	
	(.03)				(.03)	
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[c] \in (p40, p60]\right\}$	0.89				0.61	
	(.03)				(.03)	
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[c] \in (p60, p80]\right\}$	0.74				0.54	
	(.03)				(.03)	
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[c] \in (p80, p100]\right\}$	0.72				0.59	
	(.04)				(.04)	
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p20, p40]\right\}$	0.07				0.14	
	(.04)				(.03)	
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p40, p60]\right\}$	0.35				0.63	
	(.04)				(.03)	
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p60, p80]\right\}$	0.70				0.87	
	(.04)				(.04)	
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p80, p100]\right\}$	1.26				1.27	
	(.04)				(.04)	
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p20, p40]\right\}$		0.35				0.20
		(.04)				(.03)
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p40, p60]\right\}$		0.65				0.43
		(.04)				(.03)
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p60, p80]\right\}$		0.94				0.58
		(.04)				(.04)
$\lambda_j^{\mathrm{c}} \times \mathbb{1}\left\{\mathrm{cov}_{M_j}[c, n] \in (p80, p100]\right\}$		0.94				0.88
		(.04)				(.04)
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[n] \in (p20, p40]\right\}$		0.08				0.15
		(.02)				(.01)
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[n] \in (p40, p60]\right\}$		0.20				0.31
		(.02)				(.02)
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[n] \in (p60, p80]\right\}$		0.39				0.44
		(.02)				(.02)
$\lambda_j^{\mathrm{n}} \times \mathbb{1}\left\{\mathrm{var}_{M_j}[n] \in (p80, p100]\right\}$		0.71				0.70
		(.03)				(.02)
R^2	0.763	0.725	0.811	0.710	0.840	0.795
$\# firms$	25,249		25,249		25,249	
Obs. Weights	No		# of empl.		# of empl.	

Notes: Estimation period: 1999–2008. Table reports coefficients on: (i) λ^{c} and λ^{n} from estimation of eqt. (6) in the paper, and (ii) lambdas interacted with indicators of whether the firm is in the respective quintile of the distribution of within-firm skill variances (or covariances). Each firm is one observation in column 1. In column 2 we include no interaction effects, weighting observations by firm employment; in column 3 we estimate model with interactions, weighting by firm employment. Robust standard errors in parentheses.

E Robustness: Alternative Samples and Specifications

Alternative sample periods: We estimate nonparametric skill-bundle premia in equation (1) for sample periods 1990–1998 and 2008–2017. Tables E.1 and E.2 report base wages and skill premia. As in baseline results, skill premia monotonically rise with the rank of each skill bundle and are substantial.[10] Dispersion of skill premia across firms is similarly very significant and (almost) all estimated premia are positive and plausible. Overall, firm heterogeneity in skill returns is rather stable across alternative estimation periods.

Next, we collect all estimates θ_j and $\Delta_j(s)$ and proceed with the iterative estimation of skill endowments and returns λ_j^{c} and λ_j^{n}, as in the main text. Table E.4 shows the resulting estimates of the Jacobian matrix elements for different sample periods. The own-derivatives on the main diagonal are clearly positive and similar in size to the benchmark specification in Table 2. PAM holds throughout.

As shown in Figure 2c and 2d of the paper, binscatter plots confirm that sorting patterns are similar across alternative estimation periods (1990–1999, 1999–2008, and 2008–2017).

Adding detailed (industry × occupation) interactions. We verify that skill returns are indeed firm-specific and do not simply reflect industry- or occupation-level skill returns. To this purpose we add detailed industry × occupation interactions with skill types to the specification (1). That is, $\boldsymbol{X}_{it}\boldsymbol{b}_t$ now contains $\sum_{o \in O} \Delta_o(s) \mathbb{1}[s_i = s]$ as additional controls where each o indexes one five-digit industry × three-digit occupation cell. This is a demanding specification as we have to control for a total of 30,738 nonmissing cell-specific wage premia for each skill type. The results reported in Table E.3 are comparable to baseline estimates.[11] As in the main text, wage premia generally rise with the skill rank within each type. As expected, estimated premia are somewhat

[10] Average skill premia differences are somewhat smaller in 1990–1998 than in the baseline, and somewhat larger in 2008–2017. This is partly because cohorts who took the military test become older over time and skill returns rise over the life cycle. In unreported analyses we use a common normalization of $\boldsymbol{X}_{it}\boldsymbol{b}_t$ (see also discussion in Section 2.2) to make the levels of $\Delta_j(s)$ comparable across sample periods. Findings indicate that average skill returns are generally stable over time although shifting toward higher relative returns for noncognitives.

[11] Results are also similar if we only control for broad industry sectors × occupations groups, or separately for specific premia of (592 unique) industries and (113 unique) occupations.

smaller, since firm-level skill returns are partly absorbed by the $\Delta_o(s)$. However, substantial level differences remain: e.g., the premium for the highest type (H, h) is still on average 28 log points, and it reaches 37 log points or more in firms that reward highly skilled workers the most. Table E.3 also shows that significant dispersion of skill premia across firms remains despite industry × occupation controls. Standard deviations are 3–5 log points. The 5–95 interquantile ranges are large, between 11 and 17 log points, and the estimated skill premia remain plausible.

Column 3 in Table E.4 reports the Jacobian matrix estimates when we include the industry × occupations controls to the estimation of equation (1). Again, the own-derivatives on the diagonal are clearly positive and similar in size to our benchmark specification in Table 2. PAM also holds. The results in Tables E.3 and E.4 show that even adding (extremely) detailed industry × occupations controls cannot turn around our main results that (1) pecuniary skill returns are very heterogeneous across firms and (2) multidimensional sorting strongly responds to these returns.

Finally, Figure E.1 plots a version of the binscatter (sorting) plot based on these estimates.[12] Average levels, especially for cognitive returns, are lower than in the baseline specification (compare Figure 2c and 2d), indicating that industry × occupation effects account for part of the baseline returns.[13] Nonetheless, average skills in firms' grow strongly, like in the baseline results, and they are generally monotonic in skill returns.

[12] This is done only for the 1999–2008 period, as occupation information is not available in 1990–1999 and industry codes change in 2010.

[13] Heterogeneity remains substantial, as seen in Table E.3 and, interestingly, industry × occupation controls in estimation (1) do not account for much of the level of returns to noncognitives.

Table E.1: Distribution of base wages and skill premia across firms

	Sample period: 1990–1999				
	mean	sd	p5	p50	p95
θ_j	0.00	0.05	-0.10	0.00	0.07
$\Delta_j(L,m)$	0.05	0.06	-0.06	0.04	0.19
$\Delta_j(L,h)$	0.10	0.07	0.00	0.11	0.25
$\Delta_j(M,l)$	0.07	0.05	0.00	0.07	0.22
$\Delta_j(M,m)$	0.17	0.06	0.13	0.14	0.33
$\Delta_j(M,h)$	0.18	0.06	0.11	0.17	0.34
$\Delta_j(H,l)$	0.18	0.08	0.04	0.19	0.35
$\Delta_j(H,m)$	0.19	0.07	0.12	0.19	0.36
$\Delta_j(H,h)$	0.32	0.06	0.23	0.30	0.48

Notes: Distributions of estimated base wages and skill premia across firms. The sample statistics displayed are the mean, standard deviation, and the 5th, 50th, and 95th percentiles. The sample covers 20,269 unique firms between 1990 and 1999.

Table E.2: Distribution of base wages and skill premia across firms

	Sample period: 2008–2017				
	mean	sd	p5	p50	p95
θ_j	0.00	0.04	-0.06	0.00	0.07
$\Delta_j(L,m)$	0.09	0.03	0.04	0.09	0.17
$\Delta_j(L,h)$	0.19	0.04	0.12	0.18	0.25
$\Delta_j(M,l)$	0.09	0.03	0.06	0.08	0.18
$\Delta_j(M,m)$	0.28	0.03	0.26	0.27	0.38
$\Delta_j(M,h)$	0.34	0.04	0.28	0.33	0.45
$\Delta_j(H,l)$	0.19	0.07	0.06	0.20	0.31
$\Delta_j(H,m)$	0.33	0.04	0.27	0.33	0.45
$\Delta_j(H,h)$	0.51	0.05	0.43	0.51	0.62

Notes: Distributions of estimated base wages and skill premia across firms. The sample statistics displayed are the mean, standard deviation, and the 5th, 50th, and 95th percentiles. The sample covers 22,015 unique firms between 2008 and 2017.

Table E.3: Base wages and skill premia: industry × occupation controls

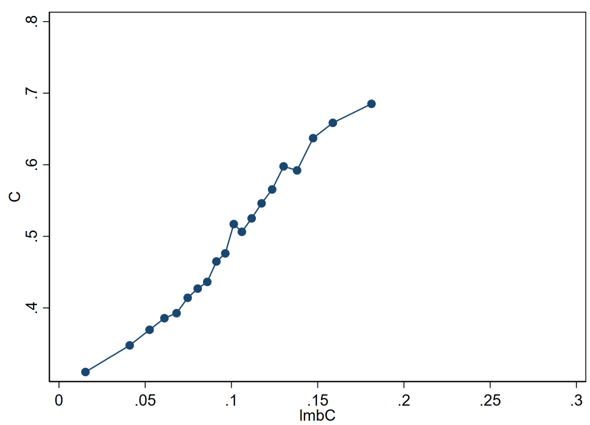
	(sample period: 1999–2008)				
	mean	sd	p5	p50	p95
θ_j	0.00	0.13	-0.24	0.01	0.19
$\Delta_j(L,m)$	0.07	0.03	0.03	0.07	0.13
$\Delta_j(L,h)$	0.11	0.05	0.02	0.11	0.17
$\Delta_j(M,l)$	-0.04	0.03	-0.09	-0.05	0.03
$\Delta_j(M,m)$	0.14	0.03	0.11	0.13	0.22
$\Delta_j(M,h)$	0.18	0.04	0.12	0.17	0.25
$\Delta_j(H,l)$	0.06	0.05	-0.04	0.07	0.14
$\Delta_j(H,m)$	0.15	0.04	0.10	0.15	0.24
$\Delta_j(H,h)$	0.28	0.04	0.20	0.27	0.37

Notes: Distributions of estimated base wages and skill premia across firms adding controls $\sum_{o\in O} \Delta_{os}\mathbb{1}[s_i = s]$. The sample statistics displayed are the mean, standard deviation, and the 5th, 50th, and 95th percentiles. The sample covers 25,604 unique firms between 1999 and 2008.

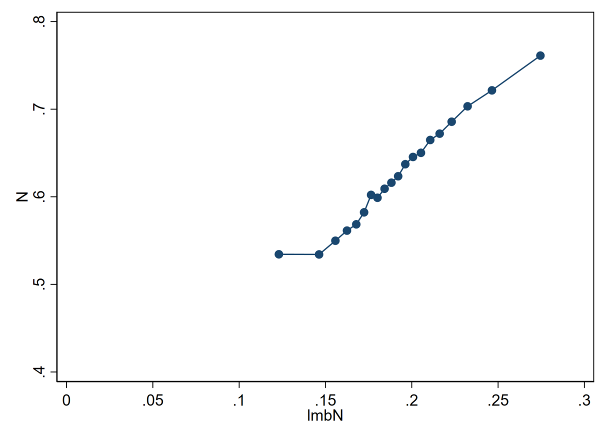
Table E.4: Jacobian matrix estimates: Average sample derivatives.

	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}	1.86	0.74	1.95	0.82	2.55	1.03
	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)
λ_j^{n}	0.70	1.36	1.15	1.46	1.25	1.70
	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)
R^2	0.653	0.514	0.717	0.600	0.596	0.555
# firms	19,634		21,755		23,809	
Estimation period	1990–1999		2008–2017		1999–2008	
Additional controls	None		None		$\sum_{o\in O} \Delta_o(s)\mathbb{1}[s_i = s]$	

Notes: Columns 1 and 2 report coefficients on δ_2 and δ_3 from estimating eqt. (6) for sample periods 1990–1999 and 2008–2017, respectively. Column 3 refers to the 1999–2008 period, but estimates eqt. (1) after adding controls $\sum_{o\in O} \Delta_o(s)\mathbb{1}[s_i = s]$. Each firm is one observation. Robust standard errors in parentheses.



(a) λ^{c} vs $\overline{c}_j$



(b) λ^{n} vs $\overline{n}_j$

Figure E.1: Average worker skills vs returns, with fine industry $\times$ occupation controls.

Notes: Binned scatterplot of firm-specific skill returns (x-axis) with average skills (y-axis) when estimating (1) in the first stage including $\sum_{o \in O} \Delta_o(s)\mathbb{1}[s_i = s]$. Estimation period only 1999–2008 as occupation information not available in 1990–1999 and industry codes change in 2010. Left: cognitive; right: noncognitive.

References

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

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.

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.

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.

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.