Skill-Biased Firms and the Distribution of Labor Market Returns

PRELIMINARY AND INCOMPLETE

Michael J. Böhm
University of Bonn and IZA
mboehm1@uni-bonn.de

Khalil Esmkhani
UBC, Sauder School of Business
khalil.esmkhani@sauder.ubc.ca

Giovanni Gallipoli
UBC, CEPR and HCEO
gallipol@mail.ubc.ca

THIS DRAFT: FEBRUARY 15, 2019

Abstract. How much do firms contribute to variation in labor market returns? To address this question we begin by documenting a historical episode that saw the economy-wide return to cognitive skills in Sweden surge in the 1990s and decline after 2000. Both upswing and reversal can be accounted for by between-firm variation. Motivated by this observation we develop and estimate an equilibrium model of skills demand with multi-dimensional firm heterogeneity. The model delivers sorting due to firm-specific skill-biases, which we estimate using rich population data on worker abilities. We find that: (i) the return to cognitive skills varies considerably across firms; (ii) more able workers sort into firms with higher returns; (iii) firm-specific wage premia, independent of workers' skills, exist even after controlling for worker sorting. A quantitative evaluation of the model shows that firm-level heterogeneity accounts for more than 2/3 of aggregate wage variation, with most of the impact due to skill-independent wage premia. Firm-level skill-biases and non-pecuniary returns play a critical role in workers' sorting and in the evolution of the average return to skills.

Keywords: Firm Heterogeneity, Cognitive Skills, Occupations, Wage Inequality
JEL Classification: D3, J23, J24, J31

---

* We received valuable feedback from numerous individuals, discussants and participants at conferences and seminars. We are grateful to Chris Boehm, Matias Cortes, Gino Gancia, Georg Graetz and Magne Mogstad for their comments on earlier drafts. We also thank 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. We alone are responsible for errors and interpretations.

1 Introduction

The distribution of labor market returns is arguably related to heterogeneity among both workers and firms, and to the prevailing patterns of sorting between them. This paper presents direct evidence on some of the forces that drive worker sorting across firms and their effect on the distribution of returns, both pecuniary and non-pecuniary. In particular, we identify significant heterogeneity in the returns to measurable cognitive ability due to firm-level differences in skill-bias.[1]

We begin by documenting strong sorting of observable skills to high-paying firms. We do this in the context of a historical episode that saw labor market returns to cognitive skills among Swedish workers rise significantly between 1990 and 2001, and decline thereafter (see also Edin et al., 2017). This roller-coaster ride is similar in quality, if not timing, to what has been documented in the US and challenges the view of inexorably rising demand for cognitive skills. We show that sorting based on observable cognitive measures can descriptively explain a substantial part of the total (level) return to skills and most of its run-up during the 1990s. Our analysis exploits direct test scores for cognitive skills from military enlistment records that are available for most men born between 1951 and the early 1980s. These ability measures cover a consistently high fraction of Swedish males and, for the cohorts we consider, exhibit a stable distribution in the population.[2] This assuages concerns about the changing composition of these skills (see, for example, Carneiro and Lee, 2011; Bowlus et al., 2017).

Three facts stand out from our descriptive analysis: (i) firms that employ many high-skill

[1]The evolution of the return to skills has attracted much interest in the debate on the causes of growing income inequality. Valletta (2017) documents a slowdown in the long-term growth of higher education wage premia in the US. Castex and Kogan Dechter (2014) compare the 1979 and 1997 waves of the NLSY to show that cognitive skills had a larger effect on US wages in the 1980s than in the 2000s. Deming (2016) highlights the role of social skills, showing that while employment and wages have declined in some occupations requiring cognitive skills, they did not in those that require both social and cognitive skills. Cortes et al. (2018) argue that demand for women labor and skills, a large part of which they measure as social skills, also increased. Beaudry et al. (2016) show that demand for skilled labor in the US outpaced supply in cognitive-intensive tasks up until the year 2001, while the opposite is true since the turn of the century. Gallipoli and Makridis (2018) also find that the growth of earnings premia for college-educated and cognitive/non-routine workers has tapered off in the US after 2000. However, they show that wage premia continued growing in IT intensive jobs. Ashworth and Ransom (2018) analyze five major labor market surveys in the US and also find that the increase in the college wage premium ceased or partially went into reverse for cohorts born after the 1970s.

[2]When considering also females we use skill proxies based on percentiles in the final grade distribution of each high-school graduation cohort as well as a measure of predicted cognitive skills constructed using the relationship between grades and cognitive ability among males.

workers increased their hiring during the 1990s, compared to less skill-intensive firms; (ii) skill-intensive firms exhibit higher wage premia on average; and (iii) during the 1990s wage premia of skill-intensive firms increased. Through an accounting exercise we show that a hiring spree by skill-intensive firms in the 1990s was the main driver of the economy-wide surge in the return to skills. The process went into reverse after 2001, with labor demand shifting towards firms with lower skill returns. This novel evidence is consistent with the view that shifts in the employment structure across firms can shape the evolution of the returns to skills in the economy. It also leads to questions about the contribution of firms characteristics to aggregate wage dispersion (Card et al., 2013, 2018; Song et al., 2018), and about the mechanism sorting high-skilled workers into high-paying firms.

To examine these issues we develop an equilibrium model of the labor market with two-sided heterogeneity and endogenous sorting which builds on the stationary monopsonistic framework popularized by Manning (2011) and Card et al. (2018). The model features firms that differ in three dimensions: wage premia, non-pecuniary amenities and skill-biases in production. The latter dimension of heterogeneity introduces a motive for skill sorting and, as we show, can be measured using population data about employer-employee pairs. Our theoretical structure is simple, as it assumes only two skill levels; this turns out to be helpful when examining the role of sorting in a labor market with two-sided heterogeneity, a topic that has attracted increasing attention in both the empirical and theoretical literature on matching (see, among others, Abowd et al., 1999; Eeckhout and Kircher, 2011; Hagedorn et al., 2017; de Melo, 2018; Borovickova and Shimer, 2018). Despite its simplicity, the equilibrium exhibits realistic features: for example, it allows for state-dependent firm sizes, it generates non-uniform levels of within-firm wage variation in the cross section of employers, and it accommodates the possibility of worker-firm sorting due to non-pecuniary preferences for firms (e.g., Sorkin, 2018).

The model delivers a log-linear wage specification that explicitly accounts for heterogeneity in workers' ability and firm-specific skill biases. Heterogeneous skill returns are compatible with, but do not require, the presence of additively separable worker fixed effects.[3]

[3]Information about workers skills allows us to account for non-random sorting and to estimate heteroge-

Estimates based on this flexible wage specification indicate the presence of significant and persistent differences in within-firm returns to skills. To gauge the robustness of these findings, and to account for the possibility of selection-bias and measurement error, we perform a variety of micro-empirical checks (i.e., we estimate two-sided fixed effect specifications, experiment with various grouping techniques, and perform event studies). These exercises consistently corroborate the baseline finding of meaningful and persistent skill-bias heterogeneity between firms.

While not focusing on skill heterogeneity, several studies based on US and European data (Card et al., 2018; Sorkin, 2018; Mogstad et al., 2018) demonstrate the empirical relevance of worker sorting in settings where firms operate in frictional or non-competitive labor markets. In related research, Bonhomme et al. (2018) develop a statistical approach that bridges the gap between structural models (e.g., Hagedorn et al., 2017) and double fixed effect specifications (Abowd et al., 1999; Card et al., 2013). This encompassing approach allows for complementarity between worker and firm types, providing a rationalization for sorting. Our work can be related to these contributions as it presents direct microdata evidence about a specific sorting mechanism.[4] As in Bonhomme et al., we find a smaller role for firm heterogeneity than is commonly estimated in AKM-type specifications (e.g., Card et al., 2013) and a larger role for the correlation between firm and worker types. This covariation turns out to be particularly important for the economy-wide return to skills.

We also test ancillary model restrictions about worker sorting, heterogeneity in returns and skill-independent wage premia. We find strong support for the hypothesis of positive assortative matching based on heterogeneous returns to skills. Firms that hire many high-skill workers also have higher estimated returns to skills. This fact is especially apparent when we look across broad industry sectors, with highly-skilled industries such as information

neous cognitive skill gradients; the ability to track workers across matches with different employers allows us to control for additively separable worker fixed effects, above and beyond cognitive returns. Cognitive gradients are identified by the interaction of skill measures and firm identifiers.

[4] As in Bonhomme et al., we restrict some dimensions of heterogeneity to make it tractable. We split workers into high- and low-skill and let firms differ across percentiles of skill-bias in production. Subsuming workers heterogeneity through direct measures of their cognitive skills makes it possible to map theory to data. We also group firms in terms of production heterogeneity, which reduces the computational burden and increases precision, while avoiding biases caused by the large number of firm-specific parameters.

technology, accounting, consulting and law exhibiting substantially higher returns to cognitive skills than other sectors. Nonetheless, we also continue to find evidence of large differences in skill-independent wage premia between-firms; this remains true after controlling for observed and unobserved worker heterogeneity.

In the last part of the paper we examine a numerical implementation of the model and quantify the role of firm-level heterogeneity for the equilibrium distribution of wages. The model's mechanism can qualitatively and quantitatively account for most of the time variation in the aggregate return to cognitive skills. When firms with strong skill-bias in production receive positive demand shocks, the economy-wide return to skills grows for three reasons. First, skill-biased firms grow in size by hiring more skilled workers. This results in more intense sorting of high-skilled workers into a subset of firms where the returns to skills are higher. Second, if skill-intensive firms also exhibit higher wage premia, as we document in the data, the employment shift of skilled workers towards skill-intensive firms reinforces their wage gains. Third, for any given worker-firm matching configuration, wage premia in skill-biased firms rise following a positive demand shock, which disproportionately benefits high-skilled workers. The model also generates a good approximation of both level and evolution of skill composition across firms as well as of overall wage inequality and within-firm wage dispersion. Our quantitative exercises suggest that firm heterogeneity has a sizable role, accounting for over 2/3 of total wage variation. Both pecuniary and non-pecuniary returns help shape the wage distribution, as their interaction drives the average return to skills and sorting patterns in the economy.

The paper proceeds as follows. Section 2 introduces the data and documents the rise and fall in the return to cognitive skills, breaking down these changes in within-firm and between-firms components. In Section 3 we introduce an equilibrium model of the labor market with two-sided heterogeneity and derive several analytical results. Section 4 examines the empirical content of the model. Section 5 reports results from the numerical implementation of our model and describes its quantitative implications, examining counterfactual scenarios and measuring the relative contribution of different layers of heterogeneity to sorting and wage dispersion. Section 6 concludes. In the Appendix we report additional empirical findings and several robustness checks.

2 Data and Stylized Facts

2.1 Data

Our main data source are administrative records for the whole Swedish population aged 16 and older. This "LISA" database runs from 1990 to 2014 and contains demographic (e.g., gender, age, education), employment (workplace, industry, and in some years occupation), and "wages" (e.g., annual labor income by employer). We match to this the records from military enlistment tests, which include measures of cognitive and non-cognitive ability (from an IQ test and interview with a psychologist, respectively). An advantage of using formal cognitive test scores over alternative skill measures is that they are very stable over time and not sensitive to compositional changes. They are also ranked one-dimensionally and relatively fine-grained, so that we can zoom-in to specific sections of the ability range.

Our main sample is restricted to males aged 30–40 because, due to high rates of service, cognitive ability is consistently available for 90 percent of the respective birth cohorts over 1990-2014. We further restrict the sample to workers employed in the private and non-primary sector who annually earn above the minimum for social security contributions (e.g., in 1998 this is 36,400 SEK$\approx$4,500 USD).

Finally, LISA reports a unique identifier for an individual's company of employment, a so-called organization number, and a workplace identifier, which is the combination of organization number and employment location. Following Card et al. (2013), among others, and consistent with our main income measure, we use workplaces with the highest earnings in a given year as our "firms" in the main analyses.[5] Details of the data sources and sample construction are in Appendix B.

[5]We have computed all of the results in the following using organizations, too, and obtained very similar results. When we relate firm wage premia to performance measures such as sales per employee, we aggregate workplaces to the organization level for which accounting variables are reported. Further robustness check that are shown in the appendix include wider age ranges, adding females by using high-school test scores, different measures of earnings (actual hourly wages), and alternatively non-cognitive skills and education measures.

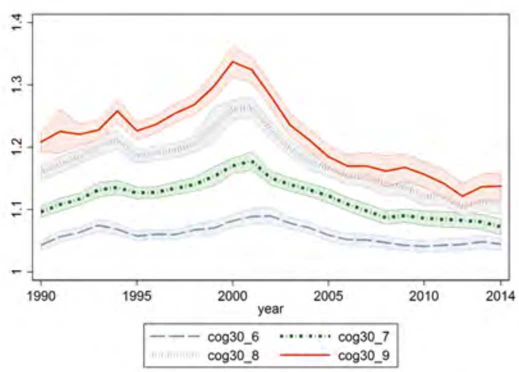
2.2 Firm Heterogeneity and the Return to Cognitive Ability

2.2.1 Trends in the Return to Skills

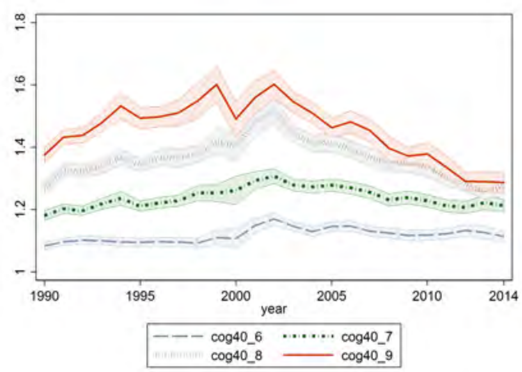
Figure 1 shows non-parametric estimates of both levels and trends in the return to skills for thirty and for forty year old males. Each line describes average earnings for a different cognitive ability group, normalized by the earnings in the mean/median ability group (stanine group 5). Relative earnings line up in the way that one would expect: top cognitive 9 individuals (solid red line; top 4 percent of the population) have the highest relative earnings, followed by groups 8, 7, and 6. Individuals with cognitive ability of 4 or lower have, on average, lower labor income than those with mean/median cognitive ability.

Figure 1: Relative Earnings by Males' Cognitive Ability Levels

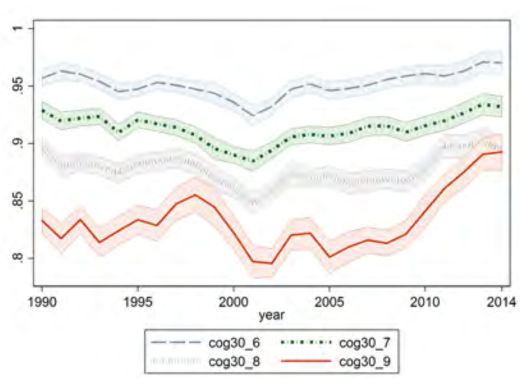
(a) Cognitives, 30 Year Olds (High)



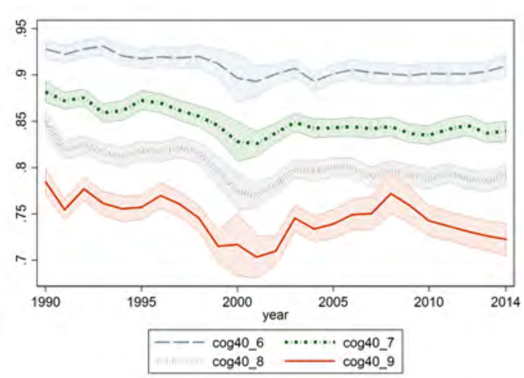
(b) Cognitives, 40 Year Olds (High)



(c) Cognitives, 30 Year Olds (Low)



(d) Cognitives, 40 Year Olds (Low)



Relative differences between ability groups are more pronounced at age forty than thirty, which is consistent with steeper life-cycle earnings profiles by skill and gradual revelation of skills over the life-cycle (Altonji and Pierret, 2001). For example, cognitive 9's wage premium is 20 percent in 1990 for thirty year olds, but almost 40 percent for forty year olds. The precision of the relative income estimates is indicated by the shaded 95 percent confidence intervals around the lines.

Next, we focus on the changes in the return to skills: the top panels of Figure 1 show how the relative earnings of top cognitive ability individuals have risen during the 1990s, with cognitive 9's premium increasing from 20 to 35 percent by the early 2000s among thirty year old workers (39% to 60% for forty year olds). After 2000 the trend strongly reverted and in 2014 the premia were lower than at the beginning of our sample period. The bottom panels of the figure report the evolution of relative earnings among males with below-average ability scores. While changes are less pronounced than at the top of the ability distribution, one can detect a downward pattern for during the 1990s for certain groups.

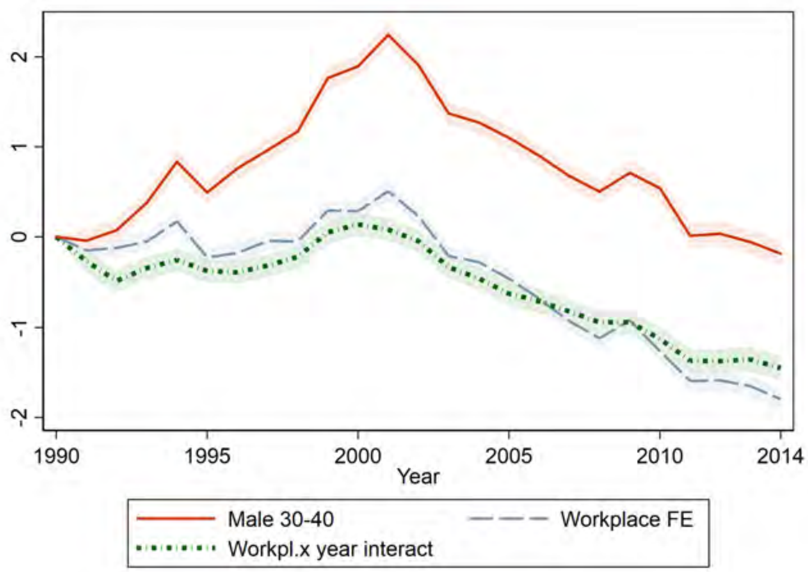
The fact that the reversal and its timing are qualitatively similar for 30 and 40 year olds suggests that these patterns are not simply due to changes in the quality of the cognitive ability measure over time or to cohort effects.[6] The reversal cannot plausibly be an artifact of different shares of low-earners entering the sample in different years (due, for instance, to the income cutoff) because the change in trend is strongest at the top of the ability and earnings distributions, which should be the least affected by this explanation. The striking patterns in Figure 1 are therefore indication of genuine variation in the skills premium, a fact that is broadly consistent with evidence available for the U.S. (Castex and Kogan Dechter, 2014; Beaudry et al., 2016; Deming, 2016; Valletta, 2017).

To examine the role that firms play in the dynamics of the return to cognitive ability we use fixed effect estimators and measure the intensity of skills' sorting across firms. Pooling together 30–40 year old males we estimate a the following wage regression:

$$\ln w_{ijt} = \alpha_t + \theta_{jt} + \lambda_t ST_i + controls_t + \varepsilon_{ijt} \tag{1}$$

[6] Böhm et al. (2018) show that the relationship between cognitive (and non-cognitive) ability measures and high-school grades has not changed much over time. The reversal in the aggregate return to cognitive skills is also robust to the inclusion of non-cognitive measures, as we show among other robustness tests in Appendix C.

Figure 2: Returns to Cognitive Ability With and Without Firm Intercepts



Ages 30–40 are pooled. In all specifications controls are year and cognitive ability interacted with age. The outcome variable is in log points.

where $\ln w_{ijt}$ are $100\times$ log annual earnings in sample year t and ST_i contains individual i's ability score in the Stanine cognitive scale. The controls include age interacted separately with year and cognitive ability to account for heterogeneity in life-cycle earning profiles and in returns to skills by age. The θ parameters in (1) are individual firm j intercepts, which can be time-invariant (i.e. $\bar{\theta}_j$) or time-varying (θ_{jt}) depending on the specification.

Figure 2 plots estimated returns to cognitive ability λ_t for each year, normalized to zero in 1990. In the specification with no firm-specific intercepts (solid red line) we replicate the result from the previous section; that is, rising returns to skills during the 1990s followed by a sharp decline after 2001. Baseline returns to skills in 1990, which are normalized in the figure, are 3.6 log points. This implies that a one standard deviation (2 Stanine scores) higher cognitive ability is associated with a 7.2 log points (7.5 percentage points) higher wage on average for 30–40 year olds.[7] This rises to peak at 5.9 log points return (i.e. 11.8 log points per standard deviation) in 2001 before it drops to 3.1 log points in 2014.

[7]Variation in cognitive ability explains roughly 10 percent of the variation in permanent wage differences (that is, in average wages for each individual over time, controlling for age and year). This is a substantial contribution for predetermined and time-invariant ability measures.

In stark contrast, when we introduce time-invariant firm intercepts (dashed blue line), there is hardly any increase in the return to skills during the 1990s apart from a small hump around 2000. When we allow for time-varying firm intercepts, there is in fact zero increase in the return to skills during the 1990s. The results for the post 2001 period are less sensitive to the introduction of firm-specific intercepts, but also in this period roughly $1/3$ of the decline in the return to cognitive ability is accounted for by firm heterogeneity (compare the drop of 2.3-(-.4)=2.7 for the solid red line to the change of 0-(-1.5)=1.5 for the dotted green line).

Figure 2 documents that most of the variation in the aggregate return to cognitive ability during the 25 years sample period occurred because of heterogeneity *between* firms, as opposed to changes in the average *within*-firm return over time. Moreover, the Table 1 shows that also much of the level of returns to skills is between firms. For example, in the first four sample years 1990–93 (i.e., top row in the left panel) firm fixed effects reduce the remaining return to cognitive ability by more than half.

2.2.2 The Role of Skill Sorting into Firms

Figure 2 documented that cross-firm variation in average wages explains much of the level and trend in the return to cognitive ability. This finding implies that firm wage premia and workers' sorting across firms must play a key role in the evolution of the aggregate return to skills. In this section we study the firm fixed effects estimated from Equation (1) and their relation to workers' cognitive skills.

Panel (a) of Figure 3 plots the variance of estimated time-varying firm effects θ_{jt}, normalized to zero in 1990.[8] This variance consists of three components: the variance of time-invariant firm fixed effects $\bar{\theta}_j$, the variance of deviations $\theta_{jt} - \bar{\theta}_j$ and a covariance term:

$$Var_t(\theta_{jt}) = Var_t(\bar{\theta}_j) + Var_t(\theta_{jt} - \bar{\theta}_j) + 2Cov_t(\bar{\theta}_j, \theta_{jt} - \bar{\theta}_j).$$

Our plot shows that the variance of firm fixed effects θ_{jt} surges in the first half of the sam-

[8] We weight by firm employment in each year. Figure A5, Panel (a), shows the corresponding standard deviations without the normalization.

Table 1: Sorting across firms and the returns to cognitive ability

	Returns to Cognitive Skills			**Between Firm Effect**		
	No FE	Firm FE	Firm-Yr FE	Overall (A)	Sorting (B)	Time-varying (C')
Initial period (1990-93)	**3.58**	**1.39**	**1.29**	**2.30**	**2.17**	**0.13**
	Δ (relative to 1990–93)			Δ (relative to 1990–93)		
difference in 1994-97	0.68	0.05	-0.06	0.71	0.64	0.07
difference in 1998-01	1.67	0.35	0.29	1.37	1.32	0.06
difference in 2002-05	1.30	-0.10	-0.09	1.38	1.39	-0.01
difference in 2006-09	0.56	-0.84	-0.56	1.15	1.40	-0.25
difference in 2010-14	-0.03	-1.44	-1.03	1.01	1.43	-0.42
N	9,477,674	9,477,674	9,477,674	9,477,674	9,477,674	9,477,674
adj. R-sq	0.231	0.496	0.514	0.028	0.039	0.000

Notes: The left panel are the three different versions of wage regression (1) by 4-year period, without and with firm fixed effects (FE). The right panel reports the decomposition according of Equation (2). No FE minus Firm-Yr FE $\approx$ Overall (A). No FE minus Firm FE $\approx$ Sorting (B). Firm FE minus Firm-Yr FE $\approx$ Time-varying (C).

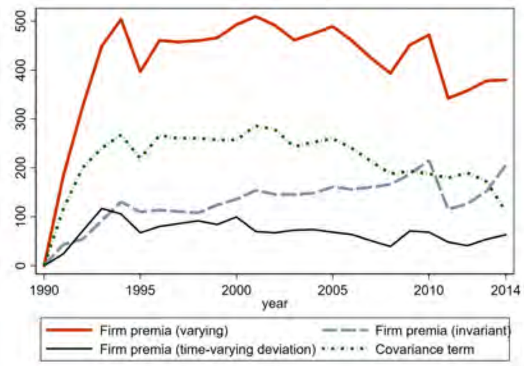
ple and modestly declines thereafter, suggesting that during the 1990s firms have become a substantially more important determinant of wages, and possibly of returns to skills. Time-invariant firm fixed effects $\bar{\theta}_j$ have also become more dispersed over those years,[9] consistent with the estimated contribution in Equation (1). However, an even larger increase in the dispersion of θ_{jt} dispersion is due to the covariance term. Notably, the variance of $\theta_{jt} - \bar{\theta}_j$ only budges at the beginning of the 1990s.

Panel (b) of Figure 3 documents the widening gap in the distribution of firm heterogeneity. For each sample year we show the share of employment for three groups of firms: those with time-varying fixed effects θ_{jt} within two standard deviations symmetrically around the

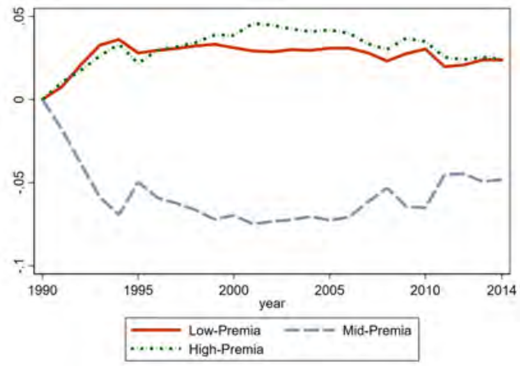
[9]This may be because of changing employment sizes of firms by $\bar{\theta}_j$ or by firm entry and exit. In the appendix we show that entry and exit is not driving the majority of this.

Figure 3: Evolution of and Sorting into Firm Wage Effects

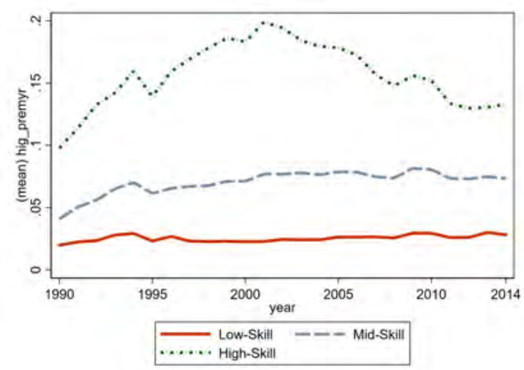
(a) Variance of θ_{jt}, $\bar{\theta}_j$, and $\theta_{jt} - \bar{\theta}_j$



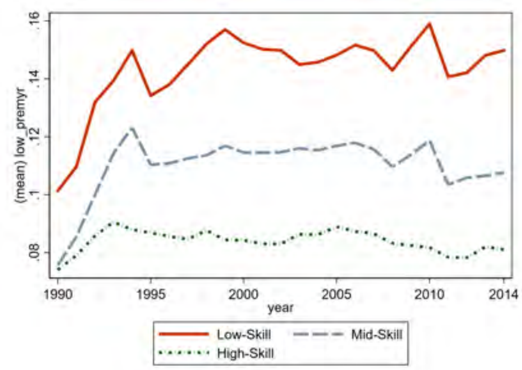
(b) Firm Effects θ_{jt} Employment Trends



(c) Sorting into High-Wage θ_{jt} Effects



(d) Sorting into Low-Wage θ_{jt} Effects



Panel (a) shows the variance of firm wage effects (i.e. $var(\theta_{jt})$ and $var(\bar{\theta}_j)$) over time. Panel (b) shows employment trends in high-wage (one $sd(\theta_{jt})$ above mean in respective year), low-wage (one $sd(\theta_{jt})$ below), and middle-wage (all other) firms where $sd(\theta_{jt})$ is calculated over all years. The bottom panels show the unconditional probability for high-skill (cognitive 7–9), middle-skill (4–6), and low-skill (1–3) individuals to work in high- and low-wage firms.

cross-sectional mean, and those whose θ_{jt} is beyond two S.D. from the mean. The standard deviation of θ_{jt} is calculated for each year. This classification provides a proxy for the share of total employment in high-, middle-, and low-paying firms, respectively. The employment structure seems to significantly polarize during the 1990s, that is, more and more individuals are working in either high- or low-paying firms, and fewer in middle-paying firms.[10]

A similar pattern can be seen for the time-invariant firm effects $\bar{\theta}_j$ in Figure A5, Panel

[10]During the 1990s the share of overall employment that was in mid-wage firms dropped by six percentage points (from a baseline of 80 percent), while the employment share of high- and low-wage firms increased by about three percentage points each (from a baseline of about 10 percent). After the 1990s this trend abated somewhat, so that the overall picture is consistent with the rising dispersion of firm wage effects in Figure 3(a).

(b). This underscores the potential for firms to drive wage changes over time. However, as we show below, the quantitatively important factor for the evolution of the return to skills is the systematic sorting of workers with different skills across firms. The lower panels of Figure 3 illustrate this sorting mechanism by plotting the share of high- (cognitive 7-to-9), middle- (cognitive 4-to-6), and low-skill (cognitive 1-to-3) employees in high and low fixed effect θ_{jt} firms, respectively. These shares change systematically during the 1990s: high-skill individuals become much more likely to work at high-paying firms while low- and middle-skill individuals sort increasingly into low-paying firms. Patterns are similar for time-invariant firm effects $\bar{\theta}_j$ (see Figure A5, Panels b–d).

The evidence presented in Figures 2 and 3 – reporting the coefficients from the wage regression (1) and the changes in the distribution of θ_{jt} and $\bar{\theta}_j$ – indicates that workers' sorting across firms plays a key role in the evolution of the aggregate return to cognitive skills. To gauge this mechanism more formally, we contrast the return to skills estimated with no firm effects to the one estimated after adding firm-specific intercepts to Equation (1). When comparing the two specifications, one can express the "omitted variable bias" as the sum of two elements. Equation (2) shows that the contribution of firm fixed effects to the cognitive skills return (A) is due to the sorting of skilled workers into high-wage firms, that is, the covariance between ST_i and θ_{jt} as well as to its changes over time. Importantly, the denominators in Equation (2) are time-invariant, since the distribution of the ability measures remains virtually unchanged over the sample period.

$$\underbrace{\frac{cov_t(\theta_{jt}, ST_i)}{Var(ST_i)}}_{\text{Between firm wage effect (A)}} = \underbrace{\frac{cov_t(\bar{\theta}_j, ST_i)}{Var(ST_i)}}_{\text{Sorting effect (B)}} + \underbrace{\frac{cov_t(\theta_{jt} - \bar{\theta}_j, ST_i)}{Var(ST_i)}}_{\text{Time-varying firm wage effect (C)}} \tag{2}$$

Equation (2) decomposes the between firm effect into a covariance between skills and firms' permanent wage premia (term B), and the covariance between skills and the deviations of firm wage premia from their average average (i.e. temporary boosts in firm-specific wages), corresponding to term (C). As shown in Figure 2, the sorting term (B) is the largest component but also the time-varying deviations (term C) play a role, especially during the late 1990s and early 2000s.

Table 1 summarizes the findings based on the decomposition (2). The top row of the left panel shows the return to cognitive ability during the initial four-year period 1990–1993. The rows below report the changes from the baseline period for different specifications of Equation (1) above: Column 2 reports results using time-invariant firm fixed effects, while Column 3 reports the results from a specification with firm-year fixed effects. The estimates in Columns 2 and 3 can be interpreted as measures of average within-firm skill premia. High-skill workers sort into high-wage firms and more than half of the initial (1990–93) return to skills is accounted for by variation across firms, as the estimated return to cognitive skills drops from 3.6 to 1.3. The left panel of Table 1 clearly documents the reversal in the overall return to cognitive ability over the four-year periods (Column 1), and shows how the time-invariant firm wage effect (Column 2) and even more so the time-varying effects (Column 3) can account for most of the increase, and a good part of the subsequent slump.

The right panel of Table 1 reports results from the decomposition (2) of the between-firm effect on the estimated returns to cognitive skills. Specifically, Column 4 shows that between firm effects account for a non-negligible share of the decline in the 2000s, most of which driven by time-varying firm effects reported in Column 6 (consistent with Panel (a) of Figure 3, where the dispersion of θ_{jt} declined after 2000).

To summarize, we highlight the following facts about the role of firms heterogeneity in the evolution of the return to skills. First, firm-specific heterogeneity is responsible for much of the economy-wide return to cognitive ability. Second, increased sorting high-skill into high-wage firms fully accounts for the run-up in skill returns during the 1990s and a substantial part of the decline after 2001. Finally, the dispersion of firm wage effects in the economy rises strongly during the 1990s and wage premia grow especially for firms that had already high wages.

3 A Model of Skill Demand by Heterogeneous Firms

In what follows we develop a model in which firms are heterogeneous in: (i) skill-bias in their production function; (ii) the demand for their products in the output market, where they have varying degrees of monopoly power; and (iii) the cost of labor in their input market, which is driven by differences in firm-specific non-pecuniary characteristics valued by employees.

Allowing for monopoly power in the output market enables us to define a simple notion of firms' surplus, whose distribution can be estimated using cross-sectional firm-level data. On the other hand, the presence of firm-specific labor supply curves (input market heterogeneity) implies that cross-sectional differences in firms' returns are reflected, in equilibrium, in firm-specific wage premia. As we show below, these assumptions are sufficient to characterize the distribution of firm-specific wage premia as a function of firms' monopoly power in the output market and firms' monopsony power in the labor market.

3.1 Workers' Skills and Labor Supply

For convenience we assume that cognitive skills can be grouped into two types $s_H > s_L$ and we examine a static skill demand problem. There is a finite set of firms $\{1, ..., J\}$ and each firm $1 \leq j \leq J$ posts a pair of skill specific wages w_{js}, $s \in \{s_H, s_L\}$, that workers observe. Worker skills are also observable and firms hire any worker of appropriate quality who is willing to accept a job at the posted wage. Crucially, each firm has idiosyncratic attributes, over which workers have heterogeneous preferences. That is, different workers derive different non-pecuniary returns from the same firm. For worker i with skill $s \in \{s_H, s_L\}$ we write the indirect utility of working at firm j as

$$u_{ijs} = \tilde{\beta}_s \ln(w_{js}) + \ln(\tilde{a}_{js}) + \tau_s \varepsilon_{ijs},$$

where $\tilde{a}_{js}$ is a firm-specific amenity, drawn from the probability distribution F_{as}, common to all workers in group s and ε_{ijs} captures idiosyncratic preferences for working at firm j with scaling parameter τ_s, arising for example from non-pecuniary match factors such as distance to work or interactions with coworkers and supervisors. We assume that shocks ε_{ijs} are independent draws from a type I Extreme Value distribution and that, given posted wages, workers are free to work at any firm they wish. Hence, using standard arguments (McFadden, 1974), the probability that workers choose firm $j \in \{1, ..., J\}$ has a logit form

$$p_{js} \equiv P(\text{argmax}_k\{u_{ijk}\} = \{j\}) = \frac{\exp[\beta_s \ln(w_{js}) + \ln(a_{js})]}{\sum_{k=1}^{J} \exp[\beta_s \ln(w_{ks}) + \ln(a_{ks})]},$$

with $\beta_s \equiv \tilde{\beta}_s/\tau_s$ and $\ln(a_{js}) \equiv \ln(\tilde{a}_{js})/\tau_s$. We abstract from strategic interactions in wage-setting and assume that, as in the actual data, the number of firms J is large, in which case logit probabilities are well approximated by exponential probabilities:

$$p_{js} \approx \xi_s \exp[\beta_s \ln(w_{js}) + \ln(a_{js})], \tag{3}$$

for $s \in \{s_H, s_L\}$, where ξ_{s_L}, ξ_{s_H} are constants common to all firms. Thus, for large J, the firm-specific labor supply functions take the form:

$$\ln(q_{js}) = \ln(N_s \xi_s) + \beta_s \ln(w_{js}) + \ln(a_{js}), \tag{4}$$

where N_s is the total number of s-type workers in the economy and q_{js} is the quantity of skill type s supplied to firm j. β_s denotes the labor supply elasticity to the offered wage. In what follows we assume that $\beta_{s_H} = \beta_{s_L} = \beta$.[11]

3.2 The Problem of the Firm and the Demand for Skills

A firm j produces according to a linear production function using different skill labor input q_{js_L} and q_{js_H}:

$$y_j(q_{js_L}, q_{js_H}) = T_j \left[s_L^{\lambda_j} q_{js_L} + s_H^{\lambda_j} q_{js_H} \right], \tag{5}$$

where the parameter λ_j captures the firm-specific skill-bias and T_j subsumes firm productivity. The main role of λ_j is to determine high-skill workers' comparative advantage in production

[11] Most of the theoretical results hold also for $\beta_{s_H} \neq \beta_{s_L}$. We present these results in Appendix A.1.

across firms. More generally, λ_j affects absolute advantage of both high- and low-skill workers, and therefore it may be reflected in measured firm-specific productivity. To transparently see both of these implications of λ_j, one can factor out $s_L^{\lambda_j}$ from the production function:

$$y_j(q_{js_L}, q_{js_H}) = T_j s_L^{\lambda_j} \left[q_{js_L} + \left(\frac{s_H}{s_L} \right)^{\lambda_j} q_{js_H} \right].$$

If $\frac{s_H}{s_L} > 1$ (and constant across firms), then a larger λ_j means that an additional unit of q_{js_H} raises output *relatively* more than an extra unit of q_{js_L}. This is comparative advantage, and skill-bias, associated with higher λ_j. In addition, for given q_{js_L} and q_{js_H}, a firm's output grows with λ_j.[12] This is the absolute advantage of having higher λ_j. Intuitively this means that firms where high-skill workers are more productive may exhibit generally higher labor productivity.

Aside from its implications for absolute advantage, the model imposes no particular correlation between the return to cognitive skills λ_j and the firm-specific wage premium θ_j. In practice one cannot easily distinguish the narrowly defined productivity T_j from absolute advantage due to higher λ_j, just as one cannot separately identify heterogeneity in the demand for the firm's output (ϕ_j) in the empirical implementation: all these elements are effectively subsumed into, and estimated as, the wage premium term θ_j. Nonetheless, estimation results in Section 4.2 clearly indicate that λ_j and θ_j are positively and significantly correlated in each cross-section of the data.

Firm output is used as an intermediate input in the production of a final good. Specifically, we posit that there is a representative firm using a CES technology to produce final good Y as follows:

$$Y = \left[\sum_{j=1}^{J} \phi_j y_j^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{6}$$

where $\sigma > 1$ is the elasticity of substitution of the different intermediates for the final output good. The firm-specific share parameter ϕ_j describes the marginal contribution of y_j to the value of output Y. Given that the technology to produce the final good is homogeneous of degree one, one can interpret ϕ_j as a demand parameter summarizing the relative 'value' of

[12]This holds if $s_H > s_L \geq 1$.

firm j's intermediate output and, therefore, the relative market power of that firm.

Denoting the market price for firm j's output as p_j, the inverse demand function for that firm's output is

$$p_j = \frac{\partial Y}{\partial y_j} = \phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}}. \tag{7}$$

Hence, higher values of the demand share ϕ_j imply that, ceteris paribus, firm j can charge more for any given amount of output y_j. Again, we abstract from strategic interactions via Y by assuming that the number of firms J is large and each firm's output is a sufficiently small fraction of the total. Profit maximization then gives marginal revenue:

$$MR_j = \frac{\partial R_j}{\partial y_j} = \frac{\sigma - 1}{\sigma} \phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}}. \tag{8}$$

The firm's problem in the factor market is to post a pair of skill-specific wages that minimize the cost of labor services given the labor supply functions (4) and some level of output y_j. The resulting cost function can then be used in a second step to determine the optimal firm output given the demand curve (7). Firms do not observe workers' preferences $\{\varepsilon_{ijs}\}$, which prevents them from perfectly discriminating among workers by offering their idiosyncratic reservation wages. The firm's optimal wage choices solve the cost-minimization problem:

$$\begin{aligned} \min_{w_{js_H}, w_{js_L}} \quad & q_{js_L} w_{js_L} + q_{js_H} w_{js_H} \\ s.t. \quad & y_j = T_j \left[s_L^{\lambda_j} q_{js_L} + s_H^{\lambda_j} q_{js_H} \right] \\ & \ln(q_{js}) = \ln(N_s \xi_s) + \beta \ln(w_{js}) + \ln(a_{js}) \end{aligned} \tag{9}$$

The first order condition for $w_{j,s}$ is

$$q_{js} + w_{js} \frac{\partial q_{js}}{\partial w_{js}} = \chi_j \frac{\partial y_j}{\partial q_{js}} \frac{\partial q_{js}}{\partial w_{js}} \tag{10}$$

where χ_j is the shadow value of each marginal unit produced by firm j. The firm sets wages that equalize the marginal cost of producing an extra unit of output to its marginal value. The value of χ_j corresponds, at the optimal choice, to marginal revenue (8). This in turn depends, at any given time, on the relative demand for firm j's output.

Proposition 1 (Firm's optimal wage choices): *For $s \in \{s_H, s_L\}$, firms' optimal wage w_{js} is equal to:*

$$w_{js} = \frac{\left(\frac{\beta}{1+\beta}\right)^{\frac{\sigma}{\sigma+\beta}} s^{\lambda_j} \times \left(\frac{\sigma-1}{\sigma}\phi_j\right)^{\frac{\sigma}{\sigma+\beta}} Y^{\frac{1}{\sigma+\beta}} T_j^{\frac{\sigma-1}{\sigma+\beta}}}{\left[s_L^{\lambda_j(1+\beta)} N_{s_L}\xi_{s_L} a_{js_L} + s_H^{\lambda_j(1+\beta)} N_{s_H}\xi_{s_H} a_{js_H}\right]^{\frac{1}{\sigma+\beta}}} \tag{11}$$

Proof. Using $\chi_j = MR_j = \frac{\partial R_j}{\partial y_j} = \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}}$ from profit maximization and (10) we get

$$\begin{aligned} q_{js} + w_{js}\frac{\beta q_{js}}{w_{js}} &= \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}} \times T_j s^{\lambda_j} \times \frac{\beta q_{js}}{w_{js}} \\ \Longrightarrow w_{js} &= \frac{\beta}{1+\beta}\frac{\sigma-1}{\sigma}\phi_j \left(\frac{Y}{y_j}\right)^{\frac{1}{\sigma}} T_j s^{\lambda_j} \end{aligned} \tag{12}$$

The right hand side represents the marginal revenue from labor type s, while the left hand side represents its marginal cost. From (12) we can write the ratio of wages for different skill groups as follows

$$\frac{w_{js_H}}{w_{js_L}} = \left(\frac{s_H}{s_L}\right)^{\lambda_j} \tag{13}$$

Using the labor supply Equation (4) and substituting (13) into (12) for s_L we get

$$w_{js_L} = \frac{\frac{\beta}{1+\beta} s_L^{\lambda_j} \times \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} T_j^{\frac{\sigma-1}{\sigma}}}{\left[s_L^{\lambda_j} N_{s_L}\xi_{s_L} a_{js_L} w_{js_L}^{\beta} + s_H^{\lambda_j} N_{s_H}\xi_{s_H} a_{js_H}((\frac{s_H}{s_L})^{\lambda_j} w_{js_L})^{\beta}\right]^{\frac{1}{\sigma}}}$$

By solving this equation for w_{js_L} we get:

$$w_{js_L} = \frac{\left(\frac{\beta}{1+\beta}\right)^{\frac{\sigma}{\sigma+\beta}} s_L^{\lambda_j} \times \left(\frac{\sigma-1}{\sigma}\phi_j\right)^{\frac{\sigma}{\sigma+\beta}} Y^{\frac{1}{\sigma+\beta}} T_j^{\frac{\sigma-1}{\sigma+\beta}}}{\left[s_L^{\lambda_j(1+\beta)} N_{s_L}\xi_{s_L} a_{js_L} + s_H^{\lambda_j(1+\beta)} N_{s_H}\xi_{s_H} a_{js_H}\right]^{\frac{1}{\sigma+\beta}}}$$

Substituting w_{js_L} into (13) concludes the proof. $\square$

Proposition 1 implies $w_{js} > 0$ for all s and j; in turn, this implies non-zero employment of all skill types in all firms. This is consistent with the empirical observation that both high and low skills are, to some degree, present across different firms. However, this model also implies skill-to-firm sorting in equilibrium.

3.3 Wages: Firm-Specific Premium and Skill Bias

It is useful to overview some properties of the wage distribution in the stationary equilibrium of the model. A formal definition of the equilibrium is in Appendix A.2. In what follows we focus on the role of output demand heterogeneity[13] (ϕ_j) as well as on firm-specific skill-bias (λ_j). A firm's optimal behavior implies:

$$w_{js} = \underbrace{\frac{\beta}{1+\beta}}_{\text{Monops.Markdown}} \times \underbrace{\frac{\sigma-1}{\sigma}\phi_j T_j \left(\frac{Y}{y_j}\right)^{\frac{1}{\sigma}}}_{\text{Marg.Revenue}} \times \underbrace{s^{\lambda_j}}_{\text{Skill Productivity}} \tag{14}$$

The expression above illustrates how the wage paid by a firm j reflects different aspects of market structure. The monopsonistic firm sets wages at a fraction $\frac{\beta}{1+\beta}$ of the marginal revenue generated by the worker, where this depends on the worker's labor supply elasticity β. In turn, the marginal revenue is an increasing function of the firm's output market demand share ϕ_j. Moreover, the marginal revenue from an extra unit of skill s depends on the firm's return to skills λ_j, and its total factor productivity T_j.

Taking logs, one can show that at any point in time the wage can be represented as the sum of a common level effect, a firm-specific fixed effect, and a skill-specific return. That is, we can write:

$$\ln(w_{js}) = \alpha + \theta_j + \lambda_j \ln(s). \tag{15}$$

The intercept $\alpha \equiv \ln\left(\frac{\beta}{1+\beta}\frac{\sigma-1}{\sigma}Y^{\frac{1}{\sigma}}\right)$ is common across firms and skills, while $\theta_j \equiv \ln\left(\phi_j T_j y_j^{-\frac{1}{\sigma}}\right)$ is a firm *wage premium* which does not vary across skills, and $\lambda_j \ln(s)$ is a *firm-specific return*

[13] Demand shocks are not separately identified from firm productivity (T_j).

to skill which varies at both skill and firm level. This representation implies that, under the model's null hypothesis, within firm wage variation is a function of workers' skill differences.

Proposition 2 (Within-Firm Wage Premia): *Within-firm wage skill premia depend on relative skill differences between workers. Ceteris paribus, firms with higher λ_j exhibit higher skill premia. That is,*

$$\ln(w_{js_H}) - \ln(w_{js_L}) = \lambda_j \left[\ln(s_H) - \ln(s_L)\right]$$

Proof. Simply take natural logarithm of (13). □

It follows that one can identify firm-specific returns to skills if cross-sectional measurements of wages and workers' skill identifiers are available. We summarize this result in the following Corollary.

Corollary 1 (Identification of Firm Skill-Bias): *If direct measures of skills are available one can identify firm-specific skill-bias λ_j (up to a normalization) from the cross-section of wages.*

To obtain identification of firm-specific skill returns, one only needs to observe a large enough cross-section of workers from different skill groups within each firm. From Equation (14) we know that $w_{js} > 0$ for all j and s. Therefore, from labor supply (4), $q_{js} > 0$ for all j and s and this requirement is satisfied.[14] The presence of different skill identifiers for workers within the firm is sufficient to guarantee identification of λ up to a normalization (that is, it is not necessary to assign a cardinal value to workers' skill differences in order to identify the cross-firm heterogeneity in the proportional returns to skills).

Equation (14) can also be used to characterize firm-specific wage premia. Consider two workers within the same skill group s who are employed at two different firms j and k. The ratio of the wages of these identical workers is

$$\frac{w_{js}}{w_{ks}} = \frac{\phi_j}{\phi_k}\frac{T_j}{T_k}\left(\frac{y_j}{y_k}\right)^{-\frac{1}{\sigma}} s^{\lambda_j-\lambda_k} = \frac{\theta_j}{\theta_k} s^{\lambda_j-\lambda_k}. \tag{16}$$

[14] Empirical implementation of this result in samples with many firms relies on partitioning firms into groups such that workers' skill heterogeneity within each group is sufficient to identify returns.

Proposition 3 (Between-Firm Heterogeneity in Wage Premia): *There exist between-firm differences in average wage premia. For given worker skill s and firm skill-bias λ_j, firms with higher output demand ϕ_j (and/or higher productivity T_j) pay higher wages.*

Proof. This is clear from (11) and assumption $\sigma > 1$. □

Incidentally, we note that the model does not restrict the wage premium to be the same across different skill groups. That is, the condition $\frac{w_{js_H}}{w_{ks_H}} = \frac{w_{js_L}}{w_{ks_L}}$ does not have to hold. However, this equality holds if firms j and k have the same λ.

Proposition (3) also allows us to identify between-firm heterogeneity in wage premia:

Corollary 2 (Identification of Firm Wage Premia): *Firm-specific wage premia θ_j can be identified (up to a normalization) from variation in log wages across firms, conditional on employees' skill measures.*

Together, Propositions (2) and (3) imply that we can use Equation (15) to estimate log wage specifications over repeated cross-sections of firms to recover both firm-specific wage premia and skill-biases. Panel variation is not required to estimated firm premia if we have direct measures of workers' skills. However, time variation is still helpful to identify time fixed effects as well as the evolution of firm-specific wage premia (and to control for differences in unobservable skills using worker fixed effects).

The model also implies that firms with a larger share in the final good production (higher demand shares) are larger.

Proposition 4 (Firm Size): *The output of a firm y_j and its total employment increase – ceteris paribus (i.e. for the same λ_j and a_{js}) – in the firm's aggregate demand share ϕ_j (and/or productivity T_j).*

Proof. From Proposition (3), $\frac{\phi_j}{\phi_k} > 1$ implies $\frac{w_{js}}{w_{ks}} > 1$ for both s_L and s_H. Given labor supply (4), this implies $q_{js_H} > q_{ks_H}$ and $q_{js_L} > q_{ks_L}$, respectively, and thus $y_j > y_k$. □

From Propositions (3) and (4) we learn that firms whose output is in higher demand produce (ceteris paribus) more output, employ more workers, and pay higher wages. By extension, if firm j experiences a positive demand shock $\Delta\phi_j > 0$, its output, employment, and wages should increase at the same time. That is, j's labor demand curve shifts upward and the new optimal employment is located further up along the firm-specific labor supply curve.

Proposition 5 (Within-Firm Skill Composition): *Workers with higher skills sort proportionally more into firms with higher λ. Hence, the average skill of workers within a firm increases with that firm's λ.*

Proof. Using (13) and the labor supply Equation (4), we get:

$$\frac{q_{js_H}}{q_{js_L}} = \left(\frac{s_H}{s_L}\right)^{\beta\lambda_j} \frac{N_{s_H}\xi_{s_H}a_{js_H}}{N_{s_L}\xi_{s_L}a_{js_L}}$$

which is increasing in λ_j. $\square$

Proposition (5) implies that the relative hiring of high- versus low-skilled workers between firms is determined by firms' skill bias. Hence, for two firms $j \neq k$ we can write:

$$\ln(\frac{q_{js_H}}{q_{ks_H}}) - \ln(\frac{q_{js_L}}{q_{ks_L}}) = \beta(\lambda_j - \lambda_k)\left[\ln(s_H) - \ln(s_L)\right] + \ln(\frac{a_{js_H}}{a_{ks_H}}) - \ln(\frac{a_{js_L}}{a_{ks_L}}). \tag{17}$$

Corollary 3 (Ranking of Firms' Skill-Bias): *If the probability measures F_{as} and F_λ are independent for $s \in \{s_L, s_H\}$ then we can rank firms according to their skill bias using observed data on their relative employment* $\ln(q_{js_H}) - \ln(q_{js_L})$.

$$\mathbb{E}\left(\ln(\frac{q_{js_H}}{q_{ks_H}}) - \ln(\frac{q_{js_L}}{q_{ks_L}})|\lambda_j, \lambda_k\right) = \beta(\lambda_j - \lambda_k)\left[\ln(s_H) - \ln(s_L)\right] \tag{18}$$

Proof. From (17) and taking conditional expectation we get:

$$\begin{aligned}
\mathbb{E}\left(\ln(\frac{q_{js_H}}{q_{ks_H}}) - \ln(\frac{q_{js_L}}{q_{ks_L}})|\lambda_j, \lambda_k\right) =& \mathbb{E}\left(\beta(\lambda_j - \lambda_k)\left[\ln(s_H) - \ln(s_L)\right]|\lambda_j, \lambda_k\right) \\
&+ \mathbb{E}\left(\ln(\frac{a_{js_H}}{a_{ks_H}}) - \ln(\frac{a_{js_L}}{a_{ks_L}})|\lambda_j, \lambda_k\right) \\
=& \beta(\lambda_j - \lambda_k)\left[\ln(s_H) - \ln(s_L)\right] + \mathbb{E}\left(\ln(\frac{a_{js_H}}{a_{ks_H}}) - \ln(\frac{a_{js_L}}{a_{ks_L}})\right) \\
=& \beta(\lambda_j - \lambda_k)\left[\ln(s_H) - \ln(s_L)\right].
\end{aligned}$$

□

Corollary 3 states that, as long as firm-specific amenities are independent of firm skill-bias, one can order groups of firms in terms of their (expected) skill-bias by ranking the skill intensity of their workers (that is, the share of skilled workers). Firms with higher skill-bias tend to hire a larger share of high-skilled workers on average. It is important to note that this ranking is robust to some violations of the independence assumption: while a sufficient condition for this corollary to hold is that amenities and firm skill biases are independent of each other, the ranking result continues to hold as long as the correlation is not negative. That is, if skills-biased firms tend to exhibit better non-pecuniary returns for high-skill than low-skill workers, the ranking of firms based on their employment structure remains valid. This seems reasonable, since firms have an incentive to create amenities for the type of worker that is relatively desirable for them (just look at what google does in terms of work environment).

4 Estimation and Results

Our equilibrium model of the labor market delivers a set of testable restrictions and, in particular, a richer specification of the wage equation featuring firm-specific skill premia, as well as additively separable firm and worker heterogeneity. In what follows we discuss our estimation approach and present the key findings.

4.1 Empirical Implementation

The log wage equation (15) delivered by the model posits heterogeneous skill bias as well as wage premia across firms. With circa one million firm-years in our dataset, this requires a computationally demanding number of parameters to be estimated. It also introduces potentially severe measurement error in the parameters, especially for small firms.[15]

[15] In particular, prior literature has shown substantial measurement error due to incidental parameters (sometimes referred to as limited mobility bias), which biases upward the variance of firm and worker fixed effects in Abowd et al. (1999, AKM)-style estimations and downward-biases their covariance (e.g., Andrews et al., 2008). This problem may be especially severe when additionally estimating firm skill-bias parameters and with our sample of only 30–40 year old males.

We therefore restrict the sample to firms which exist for at least two years and on average have at least two employees, respectively. We further group firms into 100 bins according to their skill composition and average wages. That is, we rank firms into percentile groups, denoted as $g \in \{1, 2, .., 100\}$, according to their workers' average cognitive ability. This binning is consistent with our theoretical analysis, which indicates that firms with similar skill premia should exhibit similar skill composition in their workforce (Corollary 3). We also group firms with respect to estimating the wage premia. In particular, we rank firms into percentile groups of average wages conditional on age and year to estimate wage premia θ_g that are specific to each of these groups. Bonhomme et al. (2018) show that in such environments grouping firms into a smaller number of bins largely alleviates biases while the correct average parameters λ_g and θ_g for each group can still be recovered. Nonetheless, our qualitative results in the following are the same if we do not group for the firm wage premia and estimate individual θ_js (see the Online Appendix).

Unlike the wage specification (1), the resulting estimation approach introduces heterogeneous λ_g gradients:[16]

$$\ln(w_{i,j\in g,t}) = \theta_g + \mu_i + \lambda_g \ln(s_i) + X_{i,t}b_t + e_{i,j,t}, \tag{19}$$

where $\ln(s_i)$ is a dummy for high cognitive skill (Stanine scores 7–9; ≈25% of individuals), λ_g the firm-bin specific return to that skill, θ_g the firm-bin specific wage premium that is paid independent of skill, and μ_i a worker fixed effect representing individuals' unobservable components of skill. The controls $X_{i,t}b_t$ contain age and skill interactions as well as time-varying returns to those characteristics separately, and $e_{i,j,t}$ is an orthogonal regression error.

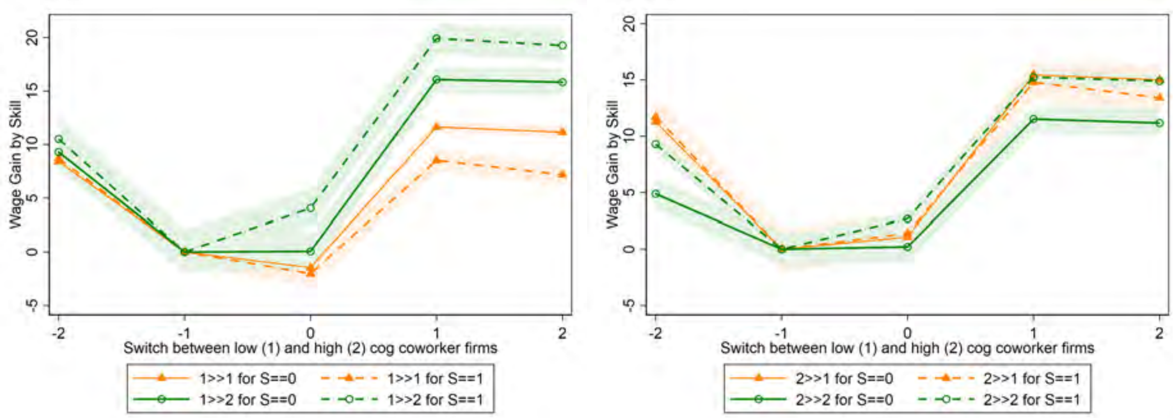
Specification (19) includes firm as well as worker fixed effects like a standard AKM model. The latter is important because μ_i accounts for the possibility of unobserved heterogeneity not captured by the cognitive skill measures, which would lead to self-selection bias. The panel

[16]Using the same index g for the firm group fixed effects θ_g and slopes λ_g is an abuse of notation for our main results in the following as in fact these a two different groupings. However, in the Online Appendix we show various other groupings, including full interactions of coarser firm average wage and cognitive ability quantiles and clustering based on the k-means algorithm (as recommended by Bonhomme et al., 2018), where each firm is allocated to a unique bin. The results with these alternative groupings are qualitatively the same.

structure of the data nonetheless allows estimation of firm-specific skill returns, by relying on information about workers' transitions across bins with different λ_g returns. This identification variation assuages concerns that bin-specific returns to cognitive skills may simply reflect the differing unobserved skill endowment within a given group of firms. The presence of cross-bin employment transitions means that our estimates convey the return to identical workers' skills in different groups of firms.

We estimate two variants of the wage equation (19): one with time-varying firm effects $\theta_{g,\tau}$ (with four-year periods as in Table 1) and one with time-invariant θ_g. In Section 4.3, when we decompose the economy-wide evolution of inequality and returns to skills over time, following Card et al. (2013) and many others, the model is estimated separately in every period. That is, different θ_g and λ_g are estimated for each four-year period.

Figure 4: Event study evidence of wage gains by skill when transitioning between firm bins



Before discussing the estimation results we show event study evidence for the heterogeneity of skill returns across firms that our λ_g parameters will represent. Figure 4 depicts the wage changes in log points of low-skilled (Stanine 1–3, S==0) versus high-skilled (Stanine 7–9, S==1) workers when transitioning between low (below median, coded 1) and high coworker cognitive ability (above median, coded 2) firms.[17] We see in the left panel that the gain from

[17] We use a balanced sample in which the firm identifier switches between $t = -1$ and $t = 0$ and is constant otherwise. Average coworker cognitive ability is measured in current year and 95 confidence intervals are plotted around the lines. The 'Ashenfelter Dip' at $t \in \{-1, 0\}$ is likely because of non-working periods around switches and our annual earnings measure. Card et al. (2013) and Song et al. (2018), among others, provide event study

moving between low coworker cognitive ability firms is decidedly larger for low-skilled workers than high-skilled workers. The gains strikingly turn around when considering transitions from low to high coworker ability firms, with now high-skilled workers having a significantly higher return.

These results are qualitatively robust for moves between or from high coworker ability firms in the right panel of Figure 4. In particular, high-skill worker's gains are statistically significantly larger when transitioning between high coworker ability firms whereas transitions from such firms to low coworker ability firms yield indistinguishable gains by worker skill. In both panels post-transition (i.e., from $t = 1$ to $t = 2$) gains are flat, which is consistent with a constant skill return difference between firm types. We therefore conclude from Figure 4 that the event study evidence indeed supports significantly different skill-biases across firms.

4.2 Estimated Firm Heterogeneity

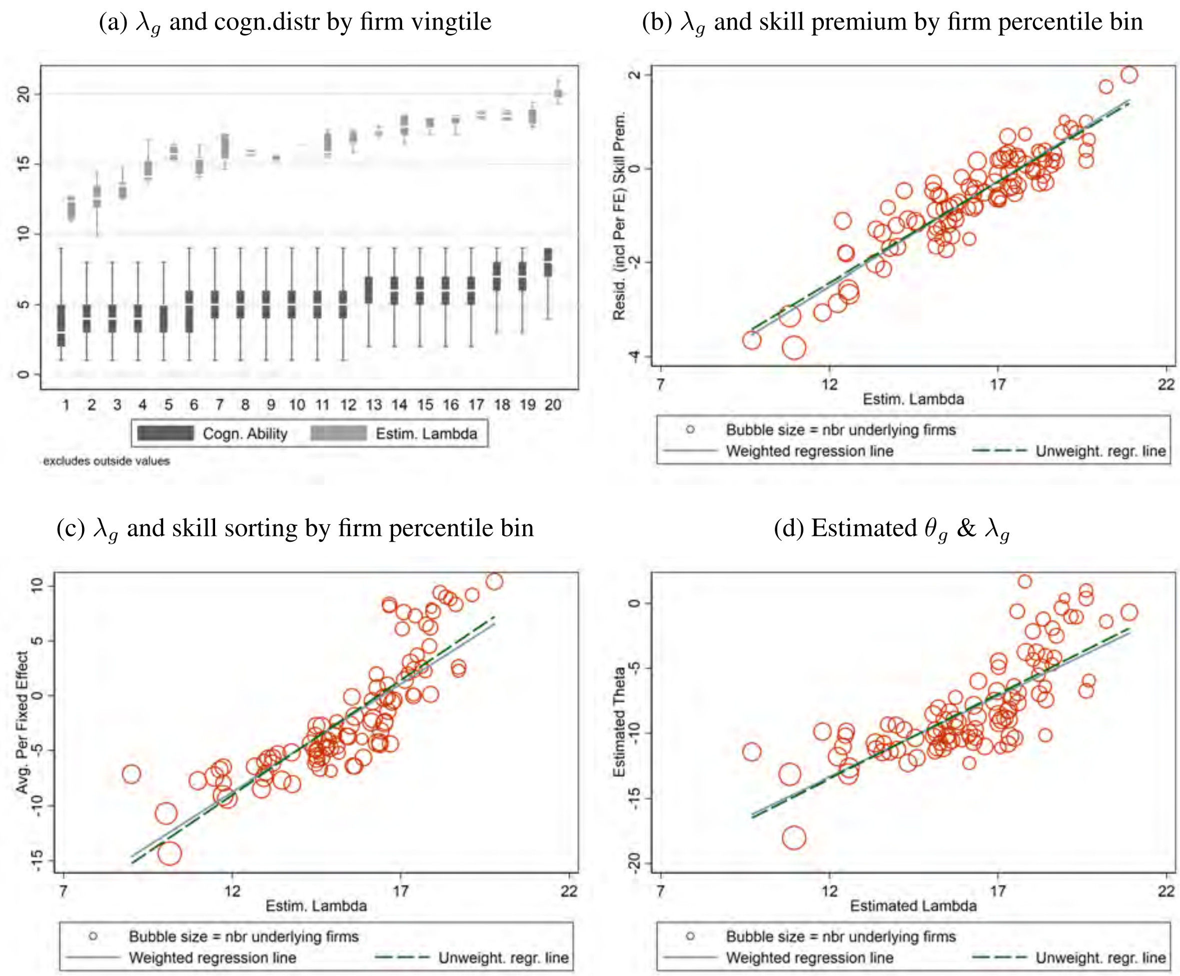
Figure 5 summarizes the estimation results with respect to firms' skill biases. In Panel (a) we report a box-plot of the distribution of estimated skill returns showing the average λ_g for 20 firm groups (cognitive ability quantiles). The same graph also shows the distribution of cognitive ability in the same vingtile (in darker shade). From the latter we can see that there is sufficient variation in skills within each quantile bins to identify the λ_g. The estimates of the λ_g clearly rise by vingtile. Moreover, average firm-level ability also rises, which reflects the grouping. Notice, however, that the scale of the two variables (λ_g and cognitive scores) are different and not comparable. In fact, the underlying correlation between cognitive percentile and estimated λ_g is around 0.7. This indicates a role for non-pecuniary valuations, which we turn to below in the model's numerical implementation.

Panel (b) of the Figure 5 shows the estimated skill bias together with a simple measure of the skill premium (the wage difference between $\ln(s_i) = 1$ and $\ln(s_i) = 0$ workers) for each of the 100 λ_g firm bins. To correct for systematic observable and unobservable selection, we residualize workers' wages by regressing them on age$\times$year$\times$cognitive skill interacted and

evidence in favor of heterogeneity in true firm wage premia (θ_g for our firm bins).

include individual fixed effects. This composition-corrected skill premium rises strongly with λ_g, providing support for Proposition 2 which states that wage differences by skill should be larger within higher-λ_g firms. Of course, this is at the same time the type of variation that empirically identifies the firm skill biases.[18]

Figure 5: Estimation Results for Firm Skill-Bias



(a) λ_g and cogn.distr by firm vingtile

(b) λ_g and skill premium by firm percentile bin

(c) λ_g and skill sorting by firm percentile bin

(d) Estimated θ_g & λ_g

Notes: Panel (a) shows boxplots of the distribution of workers' cognitive ability in the different firm bin vingtiles (i.e., the variation we use for identification) and of the estimated λ_gs in the vingtiles (five each). Panel (b) plots the estimated λ_gs against the average person fixed effect in each of the 100 firm bins and Panel (c) plots the (residualized) skill premium. Panel (d) shows average estimated firm effect by skill-bias for each of these bins.

Panel (c) of Figure 5 examines another proposition of our theoretical model, namely that

[18] As a robustness check, Figure A6 in the Appendix also shows wage inequality within firm bins, both residualized and not. The results are qualitatively similar.

the workforce average skill level should increase with the firm's skill bias. We have already seen this in the box-plot for the cognitive ability of our estimation. Here we show independent evidence (the estimation ensures that μ_i is orthogonal to $\ln(s_i)$) in favor of Proposition 5:[19] average worker fixed effects rise strongly with estimated λ_gs. Figure 5c also underscores the importance of including worker fixed effects in the estimation (19) to account for selection based on unobserved heterogeneity. Since λ_g and μ_i turn out strongly correlated, sorting into firms with higher cognitive skill returns occurs also on unobserved dimensions of heterogeneity.

The appendix reports detailed results for the estimated θ_g. We show that, as in previous literature, there is substantial variation in wage premia across firm(-bin)s and that they vary systematically with key observable characteristics. In particular, Proposition 3 states that more productive firms should pay higher wage premia. Figure A7 documents that firms with higher θ_g exhibit higher sales per worker and higher profits per worker. It also supports Proposition 4, which implies that higher-θ_g should be larger in terms of employment size.[20]

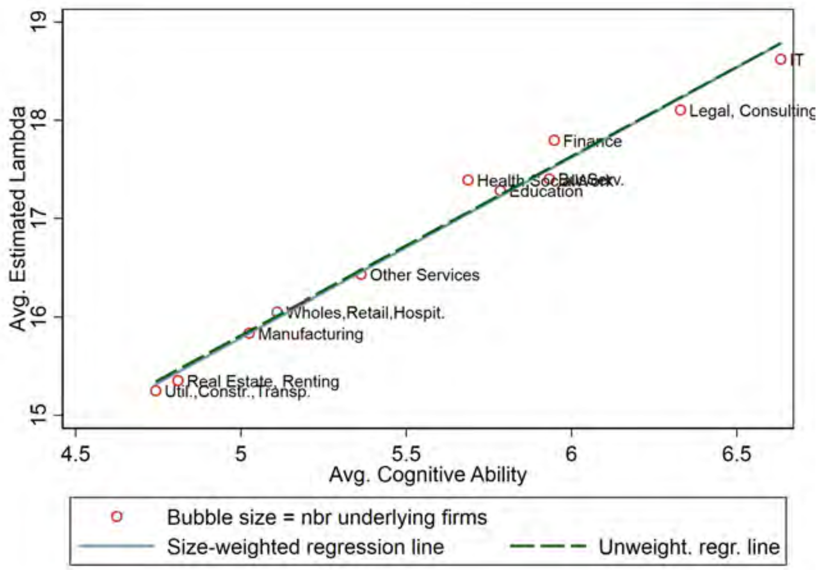
Heterogeneity in skill-biases and wage premia needs to be large in order for firms to have the strong role in generating levels and returns to skills documented in Section 2. We see in Figure 5 that skill-bias varies substantially across the 100 firm bins with λ_g ranging from below 10 to above 20 log points skill premia. The large variation in θ_g has been well-documented in the literature and can best be seen in Section 5 below. However, for firm premia to drive the returns to skills, they have to be systematically related to λ_g which affect the skill sorting (see the discussion in relation to the decomposition (2)). The last panel of Figure 5 shows λ_g and θ_g for the respective 100 firm bins together and, indeed, they are clearly positively related. This is consistent with our theoretical analysis (discussion in Section 3.2) and it will generate the skill sorting into high-paying firms that is so important for the return to skills (in levels and trends) but also affects wage inequality, as we will see in the numerical implementation

[19]The firm bins are also ranked according to observed cognitive skills, so there is no induced mechanical relationship between μ_i and λ_g.

[20]Firm accounting information including sales and profits are available from 1998 onward in our data and the sample is restricted accordingly. For firm size we use our male employees 30–40 but we also do this exercise with total employment from alternative accounting obtain similar results. The relationships persist when we look at first differences (growth rates) using the panel dimension of the firms' data set.

of the model.[21]

Figure 6: Average Cognitive Ability by Sector and Estimated Skill-Bias



Notes: Estimated λ_g on the y-axis against average cognitive ability of 1-digit sectors on the x-axis.

Finally, Figure 6 illustrates our estimation results in a dimension that one might have ex ante predictions about. In particular, we collapse firm bins by average cognitive ability and estimated firm skill biases into broad one-digit sectors. The resulting scatter-plot turns out quite striking in the sense that average skills and their within-sector returns align almost perfectly. For example, information technology ('IT') is by far the most cognitively skilled sector and we estimate that indeed its skill bias in production is also clearly the highest. Notice that there is nothing mechanical about this estimate as skill-biases are identified from wage returns to cognitive ability as opposed to average cognitive ability, which is identified from relative quantities of skilled workers. Accordingly, 'Legal, Consulting, and Accounting', 'Finance',

[21] Table A5 in the appendix summarizes the estimation results, for selected groups of firms (we put firms into decile bins, according to the rank of their average cognitive ability). The skill gradients λ_g are distributed quite tigthly within each bin and increase with firm cognitive decile, confirming the presence of assortative matching in the labor market. Interestingly, a higher degree of skill bias λ_g is positively associated with higher firm-level wage premia θ_g (we consider both time-varying and stationary wage premia). Person fixed effects also exhibit a positive correlation with cognitive skills and firm premia. Matching along observable and unobservable heterogeneity dimensions is prevalent in the labor market, with cognitive skills providing a good anchor to establish the direction and intensity of workers' sorting.

'Health', and 'Education' are among the other skilled and high skill return sectors while low-skilled and low return sectors include 'Utility, Construction, and Transport' as well as 'Real Estate and Renting'. We conclude that this, in addition to the event studies and relationships with the theoretical propositions, constitutes substantive evidence in support of our model of heterogeneously skill-biased firms and the resulting estimation approach.[22]

4.3 Accounting for Skill Returns and Wage Dispersion Over Time

Using our estimates for model (19) we can assess the contribution of firm-specific skill returns and wage premia to observed trends in wage inequality and in the return to skills. The first line of Table 2 shows the variance of log wages (normalized to 100) for the period 1990-1993. The following rows report changes (relative to the 1990–93 baseline) in successive four-year periods.

From the first column of Table 2 we recognize that a boom-bust pattern (similar to that of the average return to skills) occurred also in the variance of wages. This variance rises during the 1990s, peaks in the early 2000s and declines somewhat afterward.[23] The last column in the table report the variance of wages that is explained by the empirical wage model (19). Again, also this variance exhibits a run-up and decline, accounting for a large share of the overall changes in the variance of wages between 1990 and 2014. Aside from the variance of worker fixed effects, much of the remarkable predictive power of the model appears to be due to sorting of highly skilled workers to firms paying high wages: the contribution of the covariance of $\lambda_g \ln(s_i)$ and μ_i with $\theta_{g,t}$ are presented in Columns (6) and (7) of the table, which both are move strongly with the evolution of workers' wage inequality.

The variance of the returns to skills across firm bins, $Var(\lambda_g \ln(s_i))$, contributes a smaller share to the overall wage variance (see Column (5)) and quantitatively does not affect the

[22]Figure A8 in the Appendix further supports these findings by depicting the same relationship for 605 five-digit industries. Figure A9 examines research and development expenses of firms by estimated skill-bias. Intramural (or in-house) R&D per worker clearly rises with the firm's λ_g whereas extramural R&D does not. We think it is intuitive that high skill-biased firms also conduct more in-house R&D.

[23]To interpret these numbers, note that the corresponding standard deviation of log wages in 1990–1994 was 0.462, which rose to 0.514 in 2000–2004 and then declined to 0.485 towards the end of the sample period.

Table 2: Decomposition of the variance of log wages. Baseline levels in 1990-1993 and changes relative to baseline in different four-year periods. All magnitudes expressed as a share of logwage in 1990-93.

	logwage (1)	theta (2)	perFe (3)	controls (4)	skillrtrn (5)	skill sorting (6)	perFE sorting (7)	model (8)
				BASELINE LEVELS				
Initial prd 1990-93	**100**	**8.7**	**52.3**	**1.0**	**4.0**	**0.6**	**12.5**	**79.5**
				CHANGES FROM BASELINE (1990-93)				
diffrnce in 1994-97	15.2	1.0	6.8	1.6	0.0	0.4	3.6	13.3
diffrnce in 1998-01	27.4	2.1	10.8	1.0	0.4	1.1	6.0	21.1
diffrnce in 2002-05	25.6	2.0	12.6	-0.1	0.2	1.2	6.1	21.9
diffrnce in 2006-09	23.0	0.2	12.3	0.7	-0.9	0.6	5.9	18.7
diffrnce in 2010-14	17.1	-1.1	9.2	0.8	-1.8	0.3	4.5	11.9

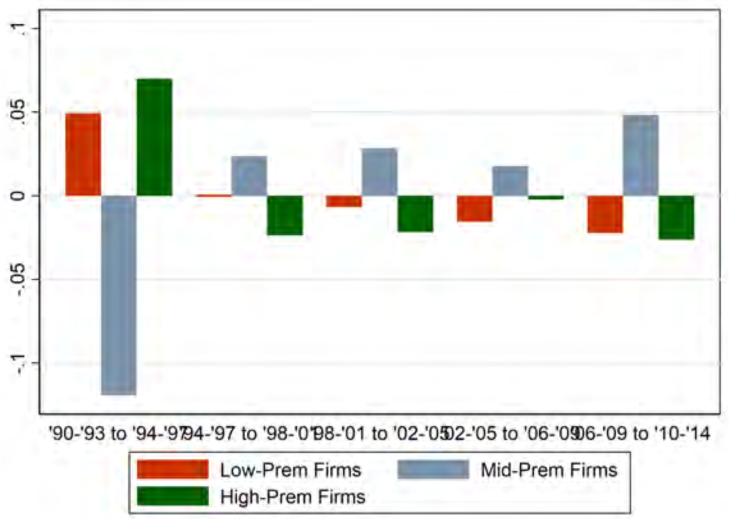
Notes: Variances of the following variables. *logwage* is $\ln(w_{i,j\in g,t})$ in regression (19), *model* the model predicted wage, *skillrtrn* the wage variation do to skill returns $\lambda_g \ln(s_i)$, *theta* the firm wage premium θ_g, *skill sorting* the wage variation due to interaction between (observable) skills and firm wage premium $2\times Cov(\lambda_g \ln(s_i), \theta_g)$, and *perFE sorting* the wage variation due to interaction between person fixed effects (unobservable skills) and firm wage premium $2\times Cov(\mu_i, \theta_g)$. *controls* $X_{i,t}b$ is the effect of fully interacted age and skills as well as age and year dummies. Circa 1.3 million observations per four-year period.

overall trend in wage inequality over time (though qualitatively it co-moves with the trend). Nonetheless, to the extent that cognitive skills dictate part of the sorting in equilibrium, cognitive skills also quantitatively contribute to the matching of skilled workers to high premium firms.

Firm fixed effects (column 2) on the other hand do matter for the variance of wages in the initial cross-section, almost accounting for ten percent of the overall variance in wage inequality. This is less than the approximately 20 percent estimated in Card et al. (2013, 2018), but it is very consistent with Bonhomme et al. (2018)'s findings after firm grouping. In addition, we find that the variance $\theta_{g,t}$ rises during the run-up of the 1990sand then declines thereafter. This confirms that sorting as well as changing firms heterogeneity is responsible for the run up in the variance of wages.

Figure 7 graphically illustrates this increase in firms heterogeneity in the first part of our sample. In particular, similar to Figure 3(b) in the descriptive analysis, we see a polarization in the firm employment structure during 1990–97 whereby high- as well as low θ_g firms

Figure 7: Firm employment polarization during first sample years



The plot shows employment share change within between four-year estimation periods of high-(above 75th percentile), low-(below 25th percentile), and middle-premium (all the others) firms. The quantiles of the thetas are computed over the whole sample.

grow strongly and middle premium firms employment share plummets. After that, the firm employment structure is more or less constant until it narrows again at the end of the sample (2006–14).

Changes in the return to cognitive skills: different components. What accounts for the evolution of the economy-wide return to skills? Table 3 reports changes in the estimated average return to skills from our empirical specification (19). The first column documents the evolution of average skill returns, confirming its fast increase in the 1990s, plateauing in the early 2000s, followed by a reversal. The average return estimated λ_g across firm bins reported in Column (2) accounts for much of the baseline level of the aggregate return to skills in the economy, with non-trivial residuals due to systematic sorting of workers by skill s_i across firms with different skill bias λ_g (Column (3)) and wage premia θ_g (Column (4)).

Consistent with the descriptive evidence of Section 2.2, the sorting of workers to firms plays a key role for the changing trends in the aggregate return to skills. The sorting effect $\frac{cov_t(\theta_{jt},\ln(s_i))}{Var(\ln(s_i))}$ of Column (4) rises by 2.5 log points during 1990–93 to 1998–01, accounting for two thirds of the increase in the return to cognitive skills to 1998–01 and almost 85% of the

Table 3: Decomposition of skill returns. Baseline levels in 1990-1993 and changes relative to baseline in different four-year periods.

	overall return (1)	average lambda (2)	lambda sorting (3)	theta sorting (4)
		BASELINE LEVELS		
Initial period 1990-93	**20.7**	**18.0**	**1.1**	**1.5**
		CHANGES FROM BASELINE (1990-93)		
difference in 1994-97	1.0	-0.2	0.3	1.0
difference in 1998-01	3.6	1.2	-0.1	2.5
difference in 2002-05	3.1	-0.1	0.6	2.6
difference in 2006-09	-0.1	-2.4	0.3	1.9
difference in 2010-14	-3.2	-5.0	0.5	1.3

Notes: Returns to high-skill dummy in log points, i.e., column (1) is the result from regression of 100 times log wage [equivalently regression model (19) predicted log wage] on the skill dummy. *average lambda* is the average estimated λ_g across the 100 cognitive firm bins, *lambda sorting* is the contribution of high-skill workers' sorting into high-λ_g firm bins, *theta sorting* the contribution of sorting into high-θ_g firm bins $\frac{cov(\theta_g, \ln(s_i))}{Var(\ln(s_i))}$. Circa 1.3 million observations per four-year period.

increase to 2002–05. After hitting the peak at the turn of the century, the rate of growth abates and then reverts back to half of the maximum increase.

The other factors contribute to the run-up in the returns to skills only to a limited extent. For 1998–01, estimated average λ_g increased but this is zero for 2002–05. Vice versa for the sorting of skill into sorting into high-λ_g firm bins, which is zero in 1998–01 and but makes a contribution of 0.6 to the increase in 2002–05.

As we have seen in Section 2.2, and contrary to what occurred during the 1990s, the run-down in the return to skills can only partially be explained by sorting into firms effects. In particular, the weakening of θ_g-sorting contributes 1.3 out of the 5.0 overall decline. The remainder is a genuine downward trend in average estimated λ_g across the 100 cognitive firm bins.

Cognitive skills and unconditional firm-level wage premia: sorting into high-paying firms. Table 4 provides additional evidence of skill sorting to firms with different wage premia. In particular, the table shows: (i) the shares of workers with high and low skills who, in 1990–93, were employed by firms with either low, middle or high wage premia; (ii) the changes over time in these shares within adjacent four-years periods.

Table 4: Sorting into low, middle, and high (top and bottom 25 percentile) premium firms

	(1)	(2)	(3)	(4)	(5)	(6)
	Low Premium		Mid Premium		High Premium	
	S=0	S=1	S=0	S=1	S=0	S=1
	INITIAL BASELINE LEVELS					
Estimation Period ’90-’99	**0.22**	**0.22**	**0.58**	**0.50**	**0.20**	**0.28**
	CHANGES WITHIN EACH ESTIMATION PERIOD					
Chg. betw. ’90-’93 and ’94-’97	0.06	0.02	-0.12	-0.11	0.06	0.09
Chg. betw. ’94-’97 and ’98-’01	0.07	0.00	-0.09	-0.11	0.02	0.11
Chg. betw. ’98-’01 and ’02-’05	0.06	0.00	-0.06	-0.09	0.00	0.09
Chg. betw. ’02-’05 and ’06-’09	0.04	-0.02	-0.05	-0.06	0.00	0.08
Chg. betw. ’06-’09 and ’10-’14	0.02	-0.04	-0.01	0.02	-0.01	0.03

Notes: this table reports the baseline and subsequent changes in the share of workers, by skill type, who work in high-premium (above the 75th percentile in each year), low-premium (below the 25th percentile), and middle-premium (all other) firms. The percentiles are estimated over the whole sample, pooling all periods.

We find strong evidence of positive assortative matching in the baseline shares, with high cognitive skill workers (S=1) being over-represented in high wage firms, while low skill workers (S=0) tend to be more present in middle and low paying firms. Even more striking, it appears that sorting has increased significantly over the 1990-2001 period, with high cognitive skills becoming a more accurate predictor of employment within high paying firms. Employment in firms with extreme θ_g also generally increased during this time, as also depicted in the firm polarization Figure 7 above. These effects subsided in the early 2000s and then reverted. Hence, sorting explains the run up in the average return to cognitive skills as well as part of the subsequent decline. The majority however appears to be due to the underlying negative drift in the economy wide average return, net of employment composition effects.

5 A Quantitative Evaluation

The model presented in Section 3 has a variety of implications for the equilibrium distribution of labor market returns. In what follows we discuss a numerical implementation of the model, in which we solve for a sequence of stationary equilibria corresponding to different periods in our sample. This allows us to (i) gauge model's performance in explaining historical patterns and (ii) quantify the role that different layers of heterogeneity have in shaping wage dispersion.

5.1 Model Parametrization

To parameterize the model we proceed sequentially. First, we normalize some parameters or set them to externally estimated values. Then, we make parametric assumptions about the probability distribution of the remaining structural parameters, which dictates the way they are drawn in the cross-section of firms. Finally, we use a generalized method of moments algorithm to estimate the shape of the parametric distribution functions, targeting a select set of data moments.

Skill endowments. We begin by normalizing the value of the two skill endowments so that $s_L = 1$ and $s_H = \exp(1)$. That is, we normalize s_L to zero and assume that the log of the ratio s_H/s_L is equal to one. This parametric choice is convenient because it allows us to identify firm-specific skill premia as returns to a high skill dummy (measured in logs). This approach is also consistent with the analysis in Section 4 where we identify firm premia using a binary classification for workers' skills[24] and, intuitively, it captures the average return of being a high skill worker in a given firm.

Elasticity of labor supply to wages. Next, we posit that the elasticity of labor supply to wages is homogeneous across skills (as we assumed throughout Section 3) and equal to 0.1. That is, we maintain that $\beta_{s_H} = \beta_{s_L} = \beta$ with $\beta = 0.1$ (see, for example, Dube et al., 2018). In robustness checks we consider several alternative values for β.

[24] In Section 4 we classify workers as high or low skills. High corresponds to a Stanine score of 7 or above.

Elasticity of substitution between firm-specific goods. The elasticity of substitution between the output produced by different firms determines the pass-through from the heterogeneity in the demand for firm-specific output to the demand for labor. We set σ in equation 6 to 3.5, which is within the range of estimates used for numerical implementations of CES aggregate demand specifications (see also Broda and Weinstein, 2006).

Number of workers and number of firms. We get the number of workers in each skill group s directly from micro data and impose them in the model, by year. We set the number of firm to $J = 50,000$, which is close to the long-term average number of firms in our sample.

Firm-level heterogeneity. In the model we allow for three dimensions of firm-level heterogeneity: skill-biases, demand/productivity differences as reflected in the aggregate demand equation 6, and non-pecuniary returns. This multidimensional heterogeneity is summarized by the following vector of firm-specific parameters,

$$\chi = (\lambda, \log(\phi), \log(a_{s_H}), \log(a_{s_L})).$$

We assume that each quadruplet χ is an i.i.d. draw from a joint Normal distribution with a mean vector μ, a standard deviation vector $\varSigma$ and correlation matrix ρ. Hence, the joint distribution can be described by the following vectors of means and standard deviations,

$$\begin{aligned} \mu &= (\mu_\lambda, \mu_{\log(\phi)}, \mu_{\log(a_{s_H})}, \mu_{\log(a_{s_L})}) \\ \varSigma &= (\varSigma_\lambda, \varSigma_{\log(\phi)}, \varSigma_{\log(a_{s_H})}, \varSigma_{\log(a_{s_L})}). \end{aligned}$$

Moreover, the correlation structure is parametrically summarized by the following 4×4 matrix:

$$\rho = \begin{bmatrix} 1 & \rho_{\lambda,\log(\phi)} & \rho_{\lambda,\log(a_{s_H})} & \rho_{\lambda,\log(a_{s_L})} \\ \rho_{\log(\phi),\lambda} & 1 & \rho_{\log(\phi),\log(a_{s_H})} & \rho_{\log(\phi),\log(a_{s_L})} \\ \rho_{\log(a_{s_H}),\lambda} & \rho_{\log(a_{s_H}),\log(\phi)} & 1 & \rho_{\log(a_{s_H}),\log(a_{s_L})} \\ \rho_{\log(a_{s_L}),\lambda} & \rho_{\log(a_{s_L}),\log(\phi)} & \rho_{\log(a_{s_L}),\log(a_{s_H})} & 1 \end{bmatrix}$$

Without loss of generality, we normalize $\mu_{\log(\phi)}$, $\mu_{\log(a_{s_H})}$ and $\mu_{\log(a_{s_L})}$ to zero.[25] Other than these three normalized values, all remaining parameters in μ, Σ and ρ are free parameters that we use to match a set of targeted data moments.

Intuitively, we want to sample a cross-section of firm parameters to generate simulated moments that are "close" to the corresponding moments estimated from microdata.[26] For each subperiod in our sample, we target the following data moments:[27]

1. mean of skill returns λ (weighted using firm sizes);
2. variance of skill returns λ (weighted using firm sizes);
3. variance of firm-specific wage-premia θ (weighted using firm sizes);
4. standard deviation of firm sizes $\log(q_H + q_L)$;
5. variance of wages $\log(w)$ (weighted using firm sizes);
6. within variance of wages $\log(w)$ (weighted using firm sizes);
7. standard deviation of firm-specific skill intensities $\log(q_H/q_L)$;
8. cross-sectional correlation between λ and θ;
9. cross-sectional correlation between λ and q_H and between λ and q_L;
10. cross-sectional correlation between θ and q_H and between θ and q_L.

We perform the moment-matching exercise separately for each sub-period in our sample, so as to characterize a sequence of stationary equilibria. In Appendix E we describe all the steps necessary to obtain a numerical solution of the model.

5.2 Model Fit

The model fit to targeted moments is consistently good, as all simulated moments lie within a 5 percent margin of their data counterparts for every subperiod. The plots in Figure 8 juxtapose

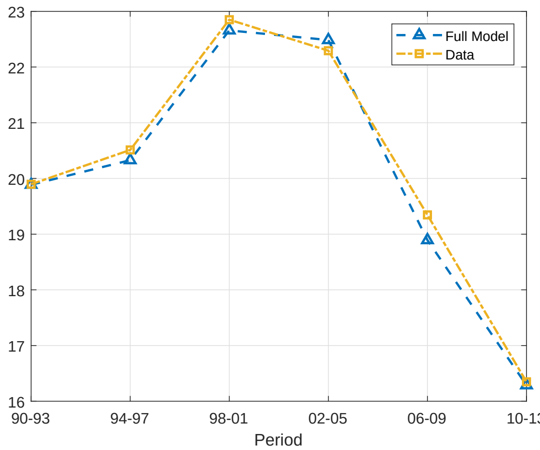
[25] $\mu_{\log(\phi)}$, $\mu_{\log(a_{s_H})}$, and $\mu_{\log(a_{s_L})}$ are level shifters and do not affect firms' hiring decisions in the labor market. Only the dispersion of the associated parameters matters. Moreover, we normalize $T_j = 1$ in equation 5, for all firms. Also this normalization is without loss of generality because if $T_j \neq 1$ we can redefine ϕ_j so that the model delivers the exact same solution.

[26] To measure the distance between simulated and data moments, we first solve the model for a given sample of firm-specific parameters i.e. $\{(\lambda_j, \log(\phi_j), \log(a_{js_H}), \log(a_{js_L}))\}_{j=1}^{J}$; then, we compute model outcomes that are directly observable in the microdata such as firm sizes q_{js}, firm-specific wage premia θ_j, and so on. Finally, we compare a set of targeted data moments with their model counterparts.

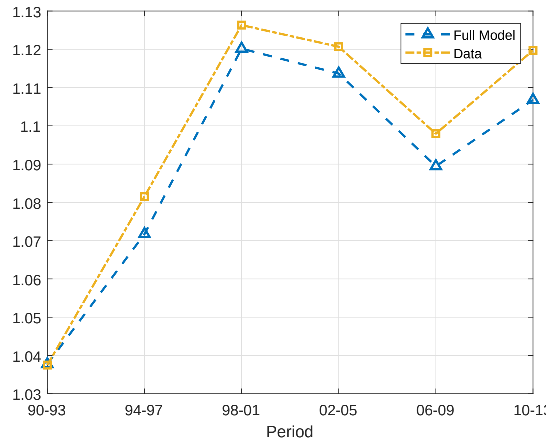
[27] In the GMM implementation we minimize the sum of squared percentage differences between model and data moments for the listed moments.

Figure 8: Model fit: baseline model at different dates

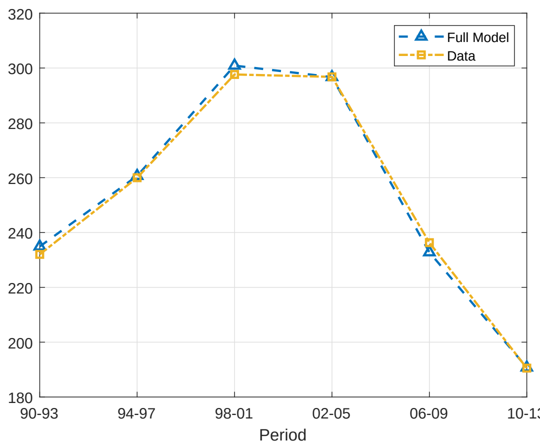
(a) Average (economy-wide) return to skills



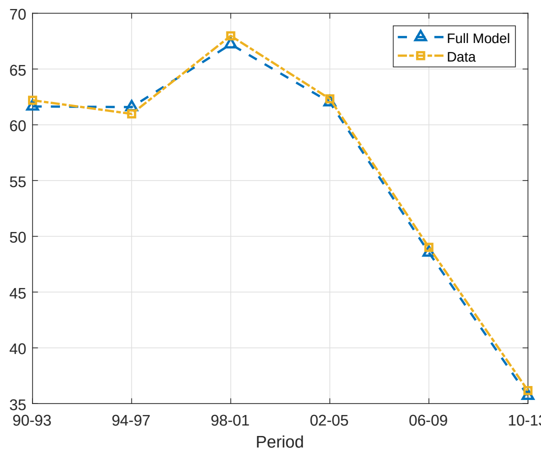
(b) Dispersion of firm-specific skill intensity: standard deviation of $\log(q_H/q_L)$



(c) Total variance of log wages, weighted by firm size



(d) Within-firm variance of log wages, weighted by firm size



four model and data moments for each of the targeted subperiods. Panel (a) shows that the model can reproduce with remarkable accuracy both size and evolution of the average return to skills in the economy; panel (b) documents that the model generates patterns of skill intensity across firms that are consistent with the dispersion observed in actual data; finally, panels (c) and (d) documents an extremely good fit to both the total and the within-firm variance of log wages, suggesting that despite its simplicity the model can rationalize patterns of wage dispersion between and within firms.

5.3 Counterfactual Experiments

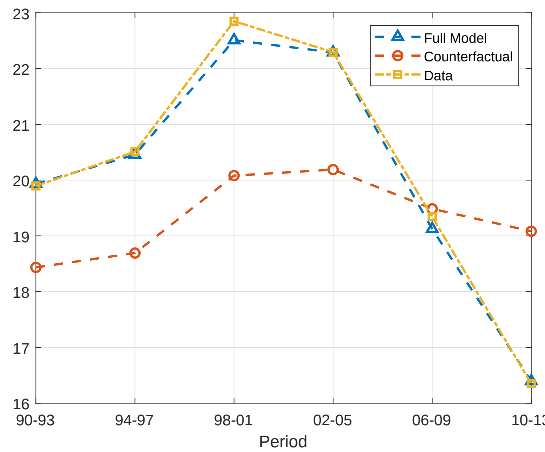
Having established that the model provides a reliable approximation of key data moments over time, we use it to draw inference about the quantitative importance of different parameters. To this purpose we perform counterfactual experiments in which we shut down alternative sources of firm-level heterogeneity.

Shutting down variation in skill biases (homogeneous λ). First, we compute counterfactual equilibria in which we restrict the skill biases of firms to be the same both cross-sectionally and over time. Specifically, we assume that the skill premium λ is equal to the average λ estimated by pooling all firms in the baseline specifications together (that is, averaging the return to skills in both the cross section and the time series of all firms). Figure 9 reports the resulting equilibrium moments, and compares them to the outcomes from the baseline parameterization. Removing skill-bias heterogeneity has a significant impact on the average return to cognitive ability and wage dispersion. The economy-wide average return to skills is significantly lower than the baseline over all periods before 2006, and higher in the 2010-12 interval. This change is consistent with microdata evidence that employment composition effects have played a central role in the roller-coaster ride of the return to skills. The magnitudes of the changes are large, with deviations in excess of 10 percentage points. Panel (b) in Figure 9 documents the large shifts in skill composition that are associated to the elimination of skill biases, as the standard deviation of skill intensity across firms drops in all time periods. These changes are reflected also in the variance of log wages which drops by roughly 10% from baseline: unsurprisingly, most of this drop is driven by changes in the within firm dispersion, which becomes flatter over time and exhibits large deviations from historical values, between 15 and 30 percentage points. Overall, it appears that skill biases play a quantitatively non-trivial role for the dynamics of both return to skills and within-firm wage dispersion.

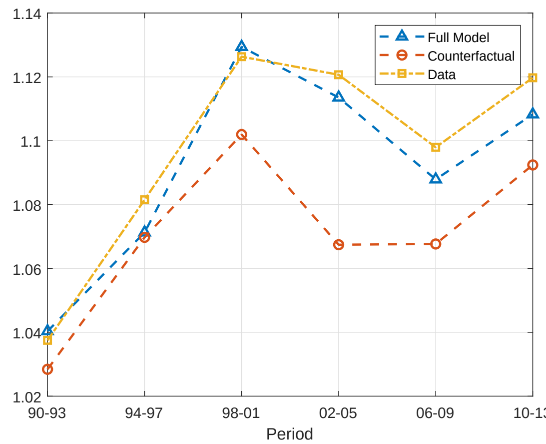
Shutting down variation in firm-specific output demand (homogeneous ϕ). In our second counterfactual experiment we examine a scenario in which all firms have identical shares (ϕ) in aggregate demand. This can be interpreted as a homogeneous demand share. This

Figure 9: Counterfactual with no heterogeneity in returns to skills: setting λ to the baseline average value across firms and over time.

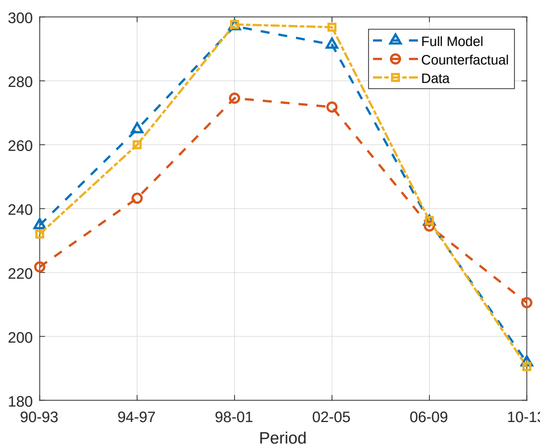
(a) Average (economy-wide) return to skills



(b) Dispersion of firm-specific skill intensity: standard deviation of $\log(q_H/q_L)$



(c) Total variance of log wages, weighted by firm size



(d) Within-firm variance of log wages, weighted by firm size

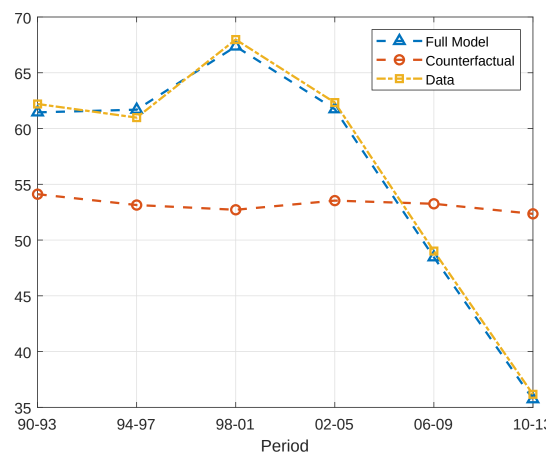
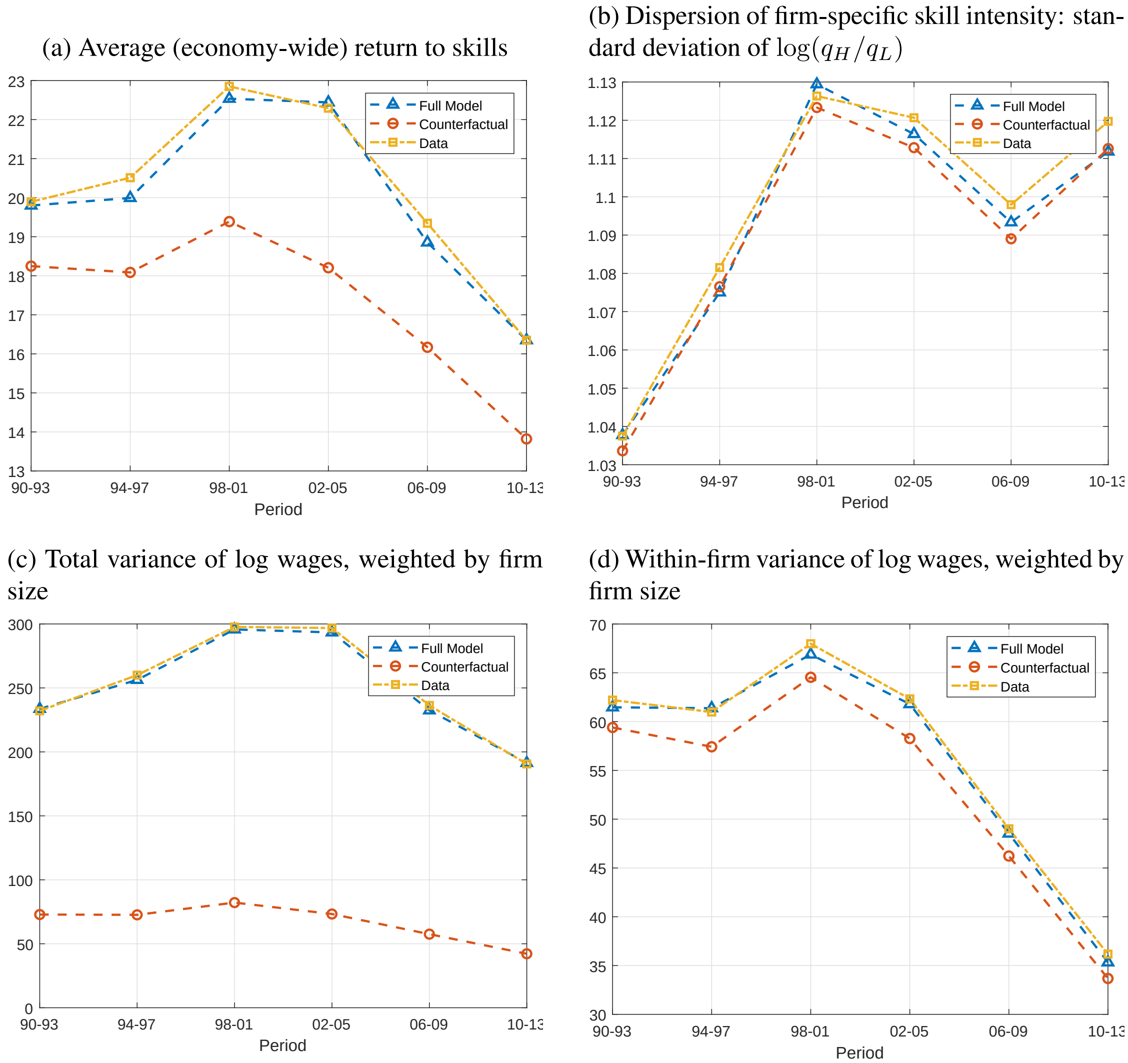


Figure 10: Counterfactual with no heterogeneity in firm-specific output demand: setting ϕ to be equal across firms.



(a) Average (economy-wide) return to skills

(b) Dispersion of firm-specific skill intensity: standard deviation of $\log(q_H/q_L)$

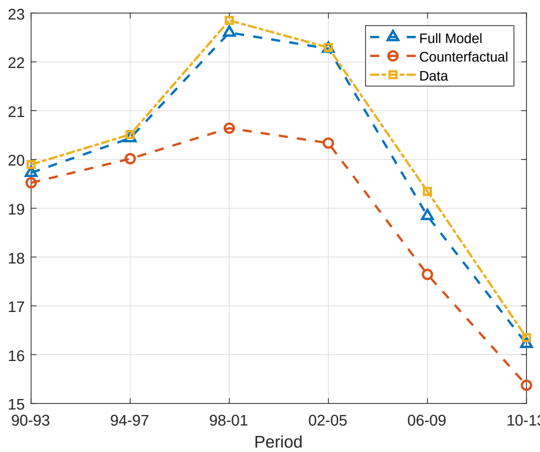
(c) Total variance of log wages, weighted by firm size

(d) Within-firm variance of log wages, weighted by firm size change induces a drop in the average return to skills (panel a of Figure 10) which is even larger than in the case of homogeneous skill biases λ. However, it is also apparent that the evolution of the economy-wide return to skills is effectively unchanged, albeit the levels are lower. The effects on the dispersion of firm-specific skill intensity and on within-firm wage dispersion (panels b and d) are also fairly small. This suggests that heterogeneity in firm-specific output demand has played a quantitative important role for the intensity of the boom-bust evolution of the return to cognitive skills but its impact on skill sorting across firms has been limited. That is, shifts in the composition of firms' hiring have contributed to changes in the *average* firm size of firms, but have had a muted impact on the composition of skill sorting with firms. Crucially, the large changes in firm sizes associated to shifts in the distribution of ϕ reflect strongly on the total variance of log wages (panel c), which experiences a counterfactual drop between 50 and 75 percentage points; this large drop is due to the fact that, despite the relative stability of within firm wage dispersion, the between firm wage differences are driven by skill-independent wage premia. These premia collapse in the absence of variation in firm-specific demand or productivity fixed effects ϕ. The magnitude of the drop in the variance of log wages confirms the quantitative importance of firm heterogeneity in skill-independent wage premia.

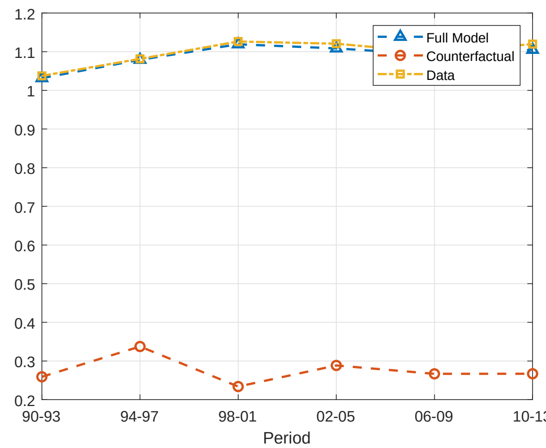
Shutting down variation in non-pecuniary returns (homogeneous a_{s_H} and a_{s_L}). Next, we explore the quantitative importance of non-pecuniary returns by making them identical across firms. That is, we set non-pecuniary preference parameters a_{s_H} and a_{s_L} to be the same across all firms (note that their absolute value is non consequential in the model). By removing selection on non-pecuniary returns, we amplify the relative importance of selection on skills. This is evident in panel (d) of Figure 11, which documents a counterfactual jump in the variance of within-firm wages, in excess of 25 percentage points in many periods. The counterpart of this change in variance is the massive drop (in excess of 80 percentage points) in the dispersion of firm-specific skill intensities (panel b of Fig. 11). The removal of non-pecuniary incentives to sort across firms results in a pattern of matching in which most high skill workers go to high skill-bias firms (and viceversa). This in turn makes skill intensity fairly homogeneous within each firm and is reflected in a large drop in the cross sectional

Figure 11: Counterfactual with no heterogeneity in firm-specific non-pecuniary returns: setting a_{s_H} and a_{s_L} to be the same across all firms.

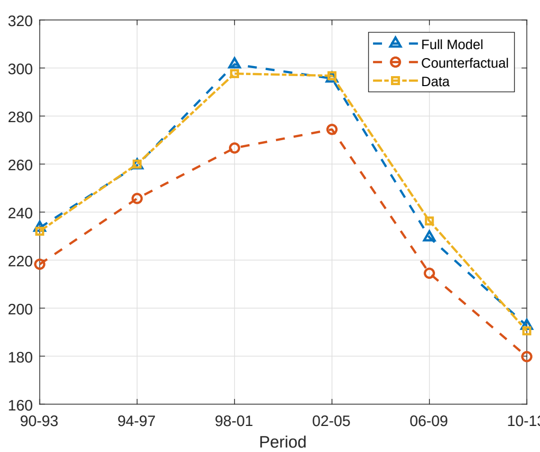
(a) Average (economy-wide) return to skills



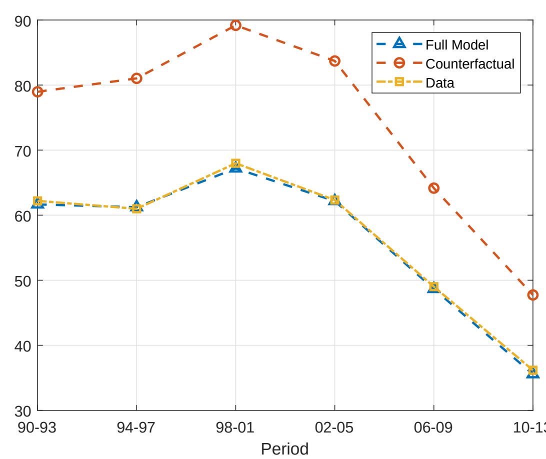
(b) Dispersion of firm-specific skill intensity: standard deviation of $\log(q_H/q_L)$



(c) Total variance of log wages, weighted by firm size



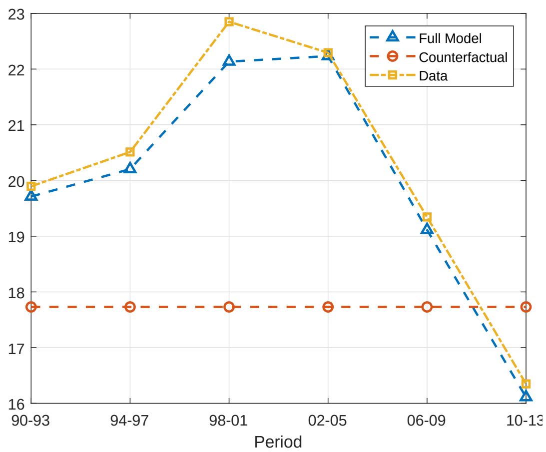
(d) Within-firm variance of log wages, weighted by firm size



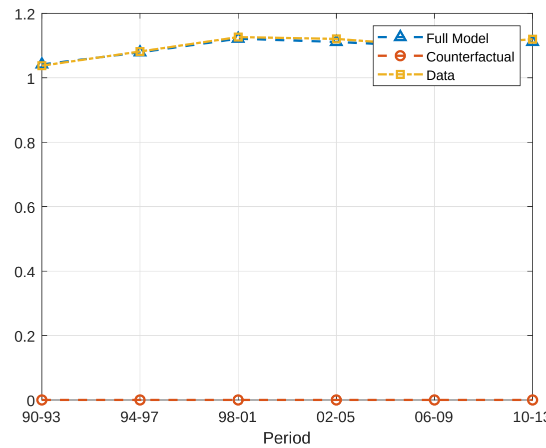
dispersion of within-firm headcount differences. The effect of non-pecuniary heterogeneity on the economy-wide return to skill is fairly limited, with only a downward shift in average returns and non change in the evolution patterns over time.

Figure 12: Counter factual results where firms are completely homogeneous in all dimensions.

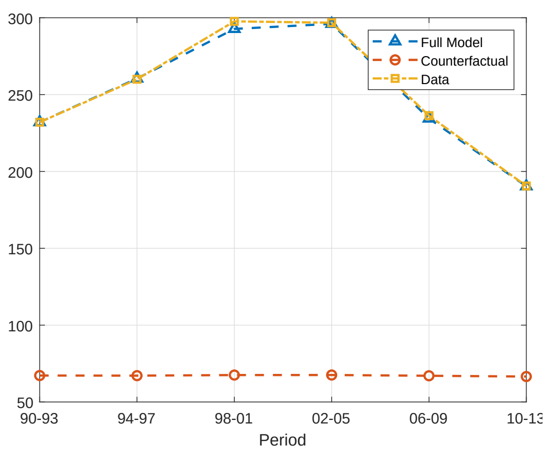
(a) Average (economy-wide) return to skills



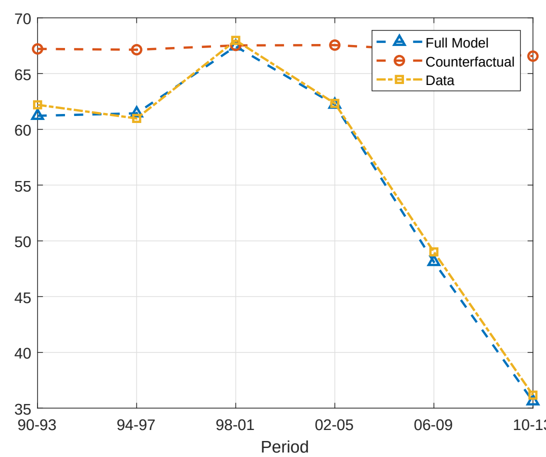
(b) Dispersion of firm-specific skill intensity: standard deviation of $\log(q_H/q_L)$



(c) Total variance of log wages, weighted by firm size



(d) Within-firm variance of log wages, weighted by firm size



Shutting down all idiosyncratic variation across firms. What happens when we shut down all sources of firm-level heterogeneity? This counterfactual exercise entails making the firms completely homogeneous by removing all differences in skill biases, output demand and non-pecuniary returns.

Figure 12 reports the results from this experiment. As expected, the average (economy-wide) return to skills becomes constant over time and corresponds exactly to the assumed λ value we impose (namely, 17.71, which correspond to the average lambda estimated across all establishments over our sample period). At the same time, the counterfactual dispersion of firm-specific skill intensities drops to zero: removing all firm-level heterogeneity results in random assignment of workers to firms, which in a large enough sample implies no ex-post variation in skill intensities.

The within-variance of log wages (panel d in Fig. 12) is also constant over time, as random assignment of workers and homogeneous λ returns imply, on average, an identical wage dispersion with firms. Finally, and perhaps most remarkably, panel (c) of figure 12 documents a very large drop in the total variance of wages, between 65 and 80 percentage points, depending on time period. This drop is informative about the overall role that firm heterogeneity plays for the distribution of labor market returns: removing all differences between firms results in a drop of at least 2/3 in wage dispersion over our sample period. Most of this drop is due to the removal of skill-independent wage premia, which appear to account for a large share of inequality in pecuniary returns.

6 Discussion and Conclusions

We investigate the importance of firm-level heterogeneity for the distribution of labor market returns. Using population data from Sweden, we document large swings in the economy-wide return to cognitive skills over a 25 year period starting in 1990, with a big run-up followed by a sudden reversal.[28] We document several aspects of this historical episode. In particular, we highlight the crucial role of cross-firm employment shifts and of heterogeneity in hiring behaviors. Two key facts emerge from our descriptive analysis: (i) firms that employ many high-skill workers increased their hiring during the 1990s, compared to less skill-intensive firms; (ii) skill-intensive firms also exhibit higher wage premia on average. A simple account-

[28] The presence of sudden swings in the labor market return to cognitive ability was not specific to Sweden, as similar patterns have been detected in the US. The sudden drop in the return to cognitive skills after the year 2000 runs counter to the view of ever increasing labor market returns to ability.

ing exercise shows that a hiring spree by skill-biased firms in the 1990s was the main driver of the economy-wide surge in the return to skills.

To make sense of the interaction of different forces in shaping observed outcomes, we develop an equilibrium model of the labor market with two-sided heterogeneity. The model features workers with heterogeneous skills and firms that differ in terms of their skill-bias, market power and non-pecuniary returns. The equilibrium exhibits positive assortative matching of high skill workers to high return firms, as well as differences in both between-firms and within-firm wage inequality.

The model delivers a log-linear wage function that explicitly accounts for heterogeneity in workers' ability and firm-specific skill biases. Taking advantage of matched employer-employee data, and of direct measures of workers' skills, we estimate different components of the model and assess its empirical content. We document significant changes in wage dispersion and assortative matching over our sample period and we find strong evidence that both skill returns and overall inequality were driven up by increasing sorting of high skill workers to high return firms during the 1990s and early 2000s. This phenomenon petered out at the turn of the century and then somewhat reversed. We find that most of the post-2000 downward trend in the economy-wide return to skills is due to the increasingly fast decline in the within-firm average return to cognitive skills (that is, changes in technology rather than composition effects). Our analysis suggests that the average return to skills, net of employment composition effects, has hardly risen at any point between 1990 and 2014. In the first half of this period, this has been masked by employment shifts towards high return firms, which has resulted in the rise of the economy-wide return to skills documented in the first part of the paper. However, as soon as these composition effects started to subside, the average return to cognitive skills stopped growing and, rather, visibly declined as the secular trend towards lower returns took hold.

These results indicate that economy-wide phenomena, such as swings in income inequality and returns to skills, are in fact the result of different and offsetting forces. These conflicting effects may not always be obvious. To quantify the influence of alternative factors on the distribution of labor returns, we perform a quantitative analysis of the equilibrium outcomes of our model. The model appears to accurately reproduce the evolution of the economy-wide return to skills, as well as key patterns of sorting and wage inequality. Through counterfactual experiments we show that over 2/3 of overall wage inequality can be attributed to between-firm heterogeneity, with skill-independent wage premia accounting for most of that variation. Firm-specific skill-biases and non-pecuniary returns remain, however, the main driver of the evolution of the economy-wide return to skills and of workers' sorting across firms.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Altonji, J. G. and C. R. Pierret (2001): "Employer learning and statistical discrimination," *The Quarterly Journal of Economics*, 116, 313–350.

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.

Ashworth, J. and T. Ransom (2018): "Has the College Wage Premium Continued to Rise? Evidence from Multiple US Surveys," .

Beaudry, P., D. A. Green, and B. M. Sand (2016): "The great reversal in the demand for skill and cognitive tasks," *Journal of Labor Economics*, 34, S199–S247.

Böhm, M. J., D. Metzger, and P. Strömberg (2018): "'Since You're So Rich, You Must Be Really Smart': Talent and the Finance Wage Premium," .

Bonhomme, S., T. Lamadon, and E. Manresa (2018): "A Distributional Framework for Matched Employer Employee Data," .

Borovickova, K. and R. Shimer (2018): "High Wage Workers Work for High Wage Firms," Mimeo.

Bowlus, A., E. Bozkurt, L. Lochner, and C. Robinson (2017): "Wages and Employment: The Canonical Model Revisited," Working Paper 24069, National Bureau of Economic Research.

Broda, C. and D. E. Weinstein (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541–585.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

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

Carneiro, P. and S. Lee (2011): "Trends in Quality-Adjusted Skill Premia in the United States, 1960-2000," *American Economic Review*, 101, 2309–2349.

Castex, G. and E. Kogan Dechter (2014): "The changing roles of education and ability in wage determination," *Journal of Labor Economics*, 32, 685–710.

CORTES, G. M., N. JAIMOVICH, AND H. E. SIU (2018): "The End of Men and Rise of Women in the High-Skilled Labor Market," Working Paper 24274, National Bureau of Economic Research.

DAL BÓ, E., F. FINAN, O. FOLKE, T. PERSSON, AND J. RICKNE (2016): "Who Becomes a Politican?" .

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

DEMING, D. (2016): "The Growing Importance of Social Skills in the Labor Market," .

DUBE, A., J. JACOBS, S. NAIDU, AND S. SURI (2018): "Monopsony in Online Labor Markets," Tech. rep., National Bureau of Economic Research.

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

EDIN, P.-A., P. FREDRIKSSON, M. NYBOM, AND B. OCKERT (2017): "The Rising Return to Non-Cognitive Skill," .

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

FLYNN, J. R. (2000): "IQ Trends over Time: Intelligence, Race, and Meritocracy," in *Meritocracy and Economic Inequality*, ed. by K. Arrow, S. Bowles, and S. Durlauf, Princeton University Press, 35–60.

GALINDO DA FONSECA, J., G. GALLIPOLI, AND Y. YEDID-LEVI (2017): "Match Quality, Contractual Sorting and Wage Cyclicality," Vancouver School of Economics working paper.

GALLIPOLI, G. AND C. A. MAKRIDIS (2018): "Structural Transformation and the Rise of Information Technology," *Journal of Monetary Economics*.

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

HANSEN, K. Y., J. J. HECKMAN, AND K. J. MULLEN (2004): "The effects of schooling and ability on achievement test scores," *Journal of Econometrics*, 121, 39–98.

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

HETHEY, T., J. F. SCHMIEDER, ET AL. (2010): "Using worker flows in the analysis of establishment turnover–Evidence from German administrative data," *FDZ Methodenreport*, 6, 43.

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.

MANNING, A. (2011): "Imperfect competition in the labor market," in *Handbook of labor economics*, Elsevier, vol. 4, 973–1041.

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

MOGSTAD, M., T. LAMADON, AND B. SETZLER (2018): "Imperfect Competition and Rent Sharing in the U.S. Labor Market," *Mimeo*.

SKANS, O. N., P.-A. EDIN, AND B. HOLMLUND (2009): "Wage dispersion between and within plants: Sweden 1985-2000," in *The structure of wages: an international comparison*, University of Chicago Press, 217–260.

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

SORKIN, I. (2018): "Ranking Firms Using Revealed Preference*," *The Quarterly Journal of Economics*, qjy001.

SUNDET, J. M., D. B. BARLAUG, AND T. M. TORJUSSEN (2004): "The End of the Flynn Effect? A Study of Secular Trends in Mean Intelligence Test Scores of Norwegian Conscripts During Half a Century," *Intelligence*, 32, 349–362.

VALLETTA, R. G. (2017): "Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both?" in *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, University of Chicago Press.

Appendix

A Theory and Equilibrium

A.1 Allowing for Heterogeneous Labor Supply Elasticities

Here we present the model results for the case $\beta_{s_L} \neq \beta_{s_H}$.

Proposition 6 (Firm's optimal wage choices): *For $s \in \{s_H, s_L\}$, firms' optimal wages w_{js} are unique and solve the following equations:*

$$w_{js_L} = \frac{\beta_{s_L}}{1+\beta_{s_L}} s_L^{\lambda_j} \times \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} T_j^{\frac{\sigma-1}{\sigma}} \left[s_L^{\lambda_j} N_{s_L} \xi_{s_L} a_{js_L} w_{js_L}^{\beta_{s_L}} + s_H^{\lambda_j} N_{s_H} \xi_{s_H} a_{js_H} (\Omega w_{js_L})^{\beta_{s_H}} \right]^{-\frac{1}{\sigma}} \tag{20}$$

$$w_{js_H} = \frac{\beta_{s_H}}{1+\beta_{s_H}} s_H^{\lambda_j} \times \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} T_j^{\frac{\sigma-1}{\sigma}} \left[s_L^{\lambda_j} N_{s_L} \xi_{s_L} a_{js_L} \left(\frac{w_{js_H}}{\Omega}\right)^{\beta_{s_L}} + s_H^{\lambda_j} N_{s_H} \xi_{s_H} a_{js_H} w_{js_H}^{\beta_{s_H}} \right]^{-\frac{1}{\sigma}} \tag{21}$$

where $\Omega = \frac{\beta_{s_H}}{1+\beta_{s_H}} \frac{1+\beta_{s_L}}{\beta_{s_L}} \left(\frac{s_H}{s_L}\right)^{\lambda_j}$ is a constant.

Proof. Using $\chi_j = MR_j = \frac{\partial R_j}{\partial y_j} = \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}}$ from profit maximization and (10) we get

$$\begin{aligned} q_{js} + w_{js}\frac{\beta_s q_{js}}{w_{js}} &= \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} y_j^{\frac{-1}{\sigma}} \times T_j s^{\lambda_j} \times \frac{\beta_s q_{js}}{w_{js}} \\ \implies w_{js} &= \frac{\beta_s}{1+\beta_s}\frac{\sigma-1}{\sigma}\phi_j \left(\frac{Y}{y_j}\right)^{\frac{1}{\sigma}} T_j s^{\lambda_j} \end{aligned} \tag{22}$$

The right hand side represents the marginal revenue from labor type s, while the left hand side represents its marginal cost.

$$\implies \frac{w_{js_H}}{w_{js_L}} = \frac{\frac{\beta_{s_H}}{1+\beta_{s_H}}}{\frac{\beta_{s_L}}{1+\beta_{s_L}}} \left(\frac{s_H}{s_L}\right)^{\lambda_j} \equiv \Omega \tag{23}$$

Using the labor supply equation (4) and substituting (13) into (12) we get

$$w_{js_L} = \frac{\beta_{s_L}}{1+\beta_{s_L}} s_L^{\lambda_j} \times \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} T_j^{\frac{\sigma-1}{\sigma}} \left[s_L^{\lambda_j} N_{s_L} \xi_{s_L} a_{js_L} w_{js_L}^{\beta_{s_L}} + s_H^{\lambda_j} N_{s_H} \xi_{s_H} a_{js_H} (\Omega w_{js_L})^{\beta_{s_H}} \right]^{-\frac{1}{\sigma}}$$

$$w_{js_H} = \frac{\beta_{s_H}}{1+\beta_{s_H}} s_H^{\lambda_j} \times \frac{\sigma-1}{\sigma}\phi_j Y^{\frac{1}{\sigma}} T_j^{\frac{\sigma-1}{\sigma}} \left[s_L^{\lambda_j} N_{s_L} \xi_{s_L} a_{js_L} \left(\frac{w_{js_H}}{\Omega}\right)^{\beta_{s_L}} + s_H^{\lambda_j} N_{s_H} \xi_{s_H} a_{js_H} w_{js_H}^{\beta_{s_H}} \right]^{-\frac{1}{\sigma}}$$

Note that the right hand side of the above equations are strictly decreasing in w_{js_L} and w_{js_H} respectively, going from infinity to zero as w_{js} goes from zero to infinity. Hence the intersections of these functions with the 45 degree line give us the unique optimal wages. □

Proposition 7 (Within-Firm Wage Premia): *For given λ_j, within-firm wage skill premia depends only on the difference of monopsonistic markdown for different skill groups and relative skill differences between workers. Moreover, firms with higher λ_j exhibit higher skill premia.*

Proof. Simply use (13) to write the within-firm skill premium as:

$$\ln(w_{js_H}) - \ln(w_{js_L}) = \ln(\frac{\beta_{s_H}}{1+\beta_{s_H}}) - \ln(\frac{\beta_{s_L}}{1+\beta_{s_L}}) + \lambda_j\left[\ln(s_H) - \ln(s_L)\right]$$

□

Proposition 8 (Between-Firm Heterogeneity in Wage Premia): *Given optimal choices, there exist between-firm differences in average wage premia. Ceteris paribus (for same worker skill s, same firm skill-bias λ_j), firms with higher demand ϕ_j (or productivity T_j) pay higher wages.*

Proof. From proposition (6) we know that w_{js_L} is the unique intersection of the right hand side function of (20) with the 45 degree line. Increasing ϕ_j shifts the right hand side function upward moving the intersection to higher w_{js_L}. Same argument works for w_{js_H}. □

Proposition 9 (Firm Size): *The output of a firm y_j and its total employment are—ceteris paribus (i.e. for the same λ_j and a_{js})—an increasing function of its aggregate demand share ϕ_j (or productivity T_j).*

Proof. From Proposition (8), $\frac{\phi_j}{\phi_k} > 1$ implies $\frac{w_{js}}{w_{ks}} > 1$ for both s_L and s_H. Given labor supply (4), this implies $q_{js_H} > q_{ks_H}$ and $q_{js_L} > q_{ks_L}$, respectively, and thus $y_j > y_k$. □

Proposition 10 (Within-Firm Skill Composition): *If $|\beta_{s_L} - \beta_{s_H}|$ is small enough, workers with relatively higher skills sort proportionally more into firms with relatively higher λ. Hence, the average skill of workers in a firm is an increasing function of that firm's λ.*

Proof. Using (23) and the labor supply equation (4), we get:

$$\left(\frac{q_{js_H}}{q_{js_L}}\right)^{\frac{1}{\beta_{s_L}}} \approx \left(\frac{q_{js_H}}{q_{js_L}}\right)^{\frac{1}{\beta_{s_L}}} q_{js_H}^{\frac{\beta_{s_L}-\beta_{s_H}}{\beta_{s_H}\beta_{s_L}}} = \left[\frac{\frac{\beta_{s_H}}{1+\beta_{s_H}}}{\frac{\beta_{s_L}}{1+\beta_{s_L}}}\left(\frac{s_H}{s_L}\right)^{\lambda_j}\frac{(N_{s_H}\xi_{s_H}a_{js_H})^{\frac{1}{\beta_{s_H}}}}{(N_{s_L}\xi_{s_L}a_{js_L})^{\frac{1}{\beta_{s_L}}}}\right]$$

which is increasing in λ_j. □

A.2 Equilibrium Definition

For the environment described in Section (3), an equilibrium consists of a set of wage functions w^*_{js} and firm choices j^*_{is} such that:

- Given wages, workers choose to work for firms where their utility is highest, i.e. given $\{w^*_{js}\}_{j=1}^J$, we have $j^*_{is} \in \arg\max_j u_{ijs}$ for all i and s.

- Given a level of aggregate production Y, demand for their output, and labor supply equation $q_{js}(w_{js})$, firms choose wages optimally to maximize their profit i.e. $w^*_{js} \in \arg\max_{w_{js}}\{d_j(y_j)y_j - w_{js_H}q_{js_H} - w_{js_L}q_{js_L}\}$ where for all j and s:

  a. $d_j(y) = \phi_j Y^{1/\sigma} y^{-1/\sigma}$;

  b. $q_{js} = q_{js}(w_{js}|\{w^*_{ks}\}_{k\neq j}) = N_s Pr(j \in \arg\max_k u_{iks})$;

  c. $y_j = T_j[s_H^{\lambda_j} q_{js_H} + s_L^{\lambda_j} q_{js_L}]$.

- The output market clears, that is $Y = \left[\sum_{j=1}^{J} \phi_j y_j^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$.

- The labor market clears or $q_{js}(w^*_{js}|\{w^*_{ks}\}_{k\neq j}) = N_s Pr(j \in \arg\max_k u_{iks})$ for all j and s. Hence, $\sum_{j=1}^J q_{js}(w^*_{js}|\{w^*_{ks}\}_{k\neq j}) = N_s$.

B Data Sources

The main data source for our analysis is the *Longitudinal Integrated Database for Health Insurance and Labor Market Studies* (LISA), provided by Statistics Sweden (SCB). LISA contains employment information (such as employment status, firm and establishment identifiers, occupation), tax records (including labor and capital income) and demographic information (such as age, education, family composition) for all individuals 16 years of age and older domiciled in Sweden, starting in 1990. Our baseline measures of pay are annual labor income from the employer with the highest earnings in that year (in hundreds of Swedish kronors).[29] These include end of year bonuses and performance pay, which likely account for a substantial part of high-ability workers' compensation (Galindo da Fonseca et al., 2017). Earnings are not top-coded.[30]

To arrive at our final sample, we restrict the dataset to workers whose total annual labor earnings exceed the *Prisbasbelopp* (the minimum amount of earnings that qualifies for the earnings-related part of the public pension system), as in Edin and Fredriksson (2000).[31] In 1998, this amount was 36,400 SEK per year, or approximately 4,500 USD in that year. We drop all observations with incomplete data (e.g., missing gender information, age, or firm ID) and, following Beaudry et al. (2016), restrict the initial sample to 25–54 year old prime age

[29]We have also conducted the analyses with the main employer, which is mostly but not always the employer with the highest earnings, and annual labor income from all sources. The results were very similar.

[30]We document the robustness of our basic results for hourly wages and, throughout the paper, we often use the term wages for annual earnings.

[31]We verify that inclusion or exclusion of workers at the lower end of the earnings distribution based on the Prisbasbelopp does not change the main results.

individuals. Finally, we keep only dependent workers employed in the private and non-primary sector (although including self-employed or public sector workers does not change much). This selection process results in a sample of about 59 million individual-year observations (descriptives in Panel (a), Table A1).

One strength of our dataset is that we obtain information about workers' inherent skills, in particular cognitive ability. This information comes from military enlistment tests, which were mandatory for Swedish male citizens before 2007 and typically taken between age 18 and 19. Lindqvist and Vestman (2011) and Dal Bó et al. (2016) provide further details on these data.[32] Cognitive ability is largely comparable to what is measured through an IQ test and was assessed through four different subtests (logic, verbal, spatial, and technical comprehension), which were then aggregated into an overall score. This 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.[33]

An advantage of using formal cognitive test scores over alternative skill measures is that they appear to be stable over time and not sensitive to compositional changes.[34] This is in notable contrast to education, for which increased attainment leads to a substantial worsening of ability within higher education groups over time (see Böhm et al. (2018) for evidence from Swedish data, or Carneiro and Lee (2011) and Bowlus et al. (2017) for evidence from the U.S.). In addition, military enlistment scores are largely exogenous and predetermined with respect to individuals' career choices. While cognitive ability may not be completely innate (Hansen et al., 2004; Heckman et al., 2006), it remains difficult for an individual to manipulate. Finally, cognitive ability is ranked one-dimensionally and is relatively fine-grained, so that we can zoom in to specific sections of the ability range (e.g. the top 4 percent).

For our main analysis we focus on males aged 30-40, because cognitive ability is most

[32] In the early 2000s, Sweden started requiring progressively fewer males to do military service. Mandatory service was abolished in 2010. Up until 2006, however, all males were required to take the military enlistment tests. To ensure that we have ability measures for a representative sample of the male population, when using these measures we restrict our analysis to individuals born before 1985. One might worry that certain individuals could deliberately perform badly on these tests to avoid military service. However, there seems to be little basis for this concern. Military test scores have been shown to significantly predict future wages and incidence of unemployment, see e.g., Lindqvist and Vestman (2011). Moreover, potential employers usually put considerable weight on military service performance and anecdotal evidence suggests that success in obtaining various jobs increases with military service experience. Finally, a large fraction of individuals working in high positions in Swedish society went through military service.

[33] A score of 5 is reserved for the middle 20 percentiles of the population taking the test. The scores of 6, 7, and 8, are given to the next 17, 12, and 7 percentiles, and the score of 9 to the top 4 percent of individuals (scoring below 5 is symmetric). See Dal Bó et al. (2016).

[34] Figure A2 shows that the distribution of the skill measures is stable over the period 1990-2014 and thus comparable over time. This allows one to select a percentile of interest in a given year and relate it to measurements in other time periods. The stability of the military test scores is partly due to standardization by the enlistment authority, but the underlying distribution of cognitive ability is arguably fairly stable over time as well. While Flynn (2000) reports improvements in average intelligence during the mid-20th century, these gains are not apparent for most of the individuals in our sample. Sundet et al. (2004) find that 18-year-old Norwegian male conscripts born after the mid-1950s had rapidly decreasing gain rates with a complete cessation of the Flynn effect for birth cohorts after the mid-1970s. Similar findings exist for Danish conscripts and for Swedish 13 year olds born 1947-1977 including girls.

Table A1: Descriptive Statistics for the Two Sample Definitions Used

Both Sexes, Age 25–54	**obs (tsd)**	**mean**	**sd**	**p5**	**p25**	**p50**	**p75**	**p95**
Age	58,900	39.6	8.5	26	32	40	47	53
Female	58,900	0.47	0.50	0	0	0	1	1
Earnings (Tsd 2014SEK)	58,900	287.9	209.1	80.9	188.3	261.1	342.2	571.4
Cognitive Ability	21,500	5.2	1.9	2	4	5	7	8
Non-cog Ability	20,800	5.1	1.7	2	4	5	6	8
High-School GPA	31,500	50.6	28.6	5.7	26.2	50.7	75.3	95.1
Predict. Cogn. Ability	31,500	50.7	28.7	5.7	26.1	50.8	75.6	95.1
Years of School	58,600	12.1	2.7	9.0	10.5	12.0	13.5	16.0
Post-Second. Degree	58,600	0.37	0.48	0.00	0.00	0.00	1.00	1.00
University Degree	58,600	0.22	0.41	0.00	0.00	0.00	0.00	1.00
PhD Degree	58,600	0.01	0.10	0.00	0.00	0.00	0.00	0.00
Males, Age 30–40								
Age	11,800	35.0	3.2	30	32	35	38	40
Female	11,800	0.0	0.0	0	0	0	0	0
Earnings (Tsd 2014SEK)	11,800	323.2	206.1	102.2	228.8	294.0	379.4	603.3
Cognitive Ability	9,700	5.2	1.9	2	4	5	7	8
Non-cog Ability	9,400	5.1	1.7	2	4	5	6	8
High-School GPA	8,000	50.4	28.6	5.6	26.0	50.4	75.0	95.0
Predict. Cogn. Ability	8,000	50.3	28.8	5.6	25.5	50.1	75.5	95.0
Years of School	11,700	12.1	2.5	9.0	10.5	12.0	13.5	16.0
Post-Second. Degree	11,700	0.35	0.48	0.00	0.00	0.00	1.00	1.00
University Degree	11,700	0.20	0.40	0.00	0.00	0.00	0.00	1.00
PhD Degree	11,700	0.01	0.12	0.00	0.00	0.00	0.00	0.00

consistently available in their birth cohorts over our 1990–2014 sample period (circa 10 million observations overall; Panel (b) of Table A1). In robustness checks we use alternative measures of worker heterogeneity, namely non-cognitive ability (also Stanine scaled and from military enlistment) and GPA in high-school (in percentiles, available for both males females).

Finally, LISA reports a unique identifier for an individual's company of employment, a so-called organization number, and a workplace identifier, which is the combination of organization number and employment location. Following Card et al. (2013), among others, and consistent with our main income measure, we use workplaces with the highest earnings in a given year as our "firms" in the main analyses. We have computed all of the results in the following using organizations, too, and obtained very similar results. When we relate firm wage premia to performance measures such as sales per employee, we aggregate workplaces to the organization level for which accounting variables are reported.

C Returns to Skills: Details and Robustness

To establish the quantitative significance of the changes in the return to cognitive ability over our sample period we estimate linear earnings regressions on individual ability. We do this for five-year periods using this specification:[35]

$$\ln w_{i\tau} = \alpha_\tau + \lambda_\tau ST_i + \varepsilon_{i\tau}. \tag{24}$$

the five year window is indexed by $\tau \in \{1990–94, 1995–99, 2000–04, 2005–09, 2010–14\}$, while $\ln(w_{i\tau})$ are $100 \times log$ annual earnings and ST_i contains individual i's ability score in the Stanine cognitive scale.

Table A2 reports results. Each column reports the difference relative to the baseline return in 1990-94 (in bold). The estimates in Table A3 confirm that the return to cognitive ability grew significantly up to the early 2000s, after which it dropped below its 1990-94 levels. The reversal in the aggregate return to cognitive skills is robust to the inclusion of non-cognitive measures, as we show below.

Finally, Section C.3 examines changes in the return to education, relative to baseline returns for high-school graduates, by age and gender. This allows us to consider also older cohorts for which we do not have cognitive ability measures in all years. While we do not know how much of the overall trend is due to deteriorating ability selection in these higher educational levels and how much to declining market prices holding ability constant, our estimates of the returns to education clearly exhibit a downward kink around 2001.

Figure A3 describes trends in male and female high-school grades and in *predicted* cognitive ability, documenting similar patterns.[36] We also verify that the results hold when using hourly wages from a survey of employers covering half of all private sector employment (for similar evidence on hourly wages see Edin et al., 2017, Fig 3b). All these alternative samples and proxies consistently indicate an increase, and then a reversal, in the return to skills after 2001 in Sweden.

These empirical findings are robust to a wide battery of tests, confirming the presence of a boom-bust episode in the economy-wide return to skills. Also the role that firms play is robust

[35]Effects are in log points. For example, in column 2 in the base period 1990-1994, a one standard deviation (two scores) increase in cognitive ability is associated with fifteen log points higher wages (approx. 2 x 7.42).

[36]For this, we regress cognitive ability of males on a third order polynomial of high-school grades interacted with school track and the age at graduation for each graduation year. We then use the parameters from this regression to predict individual cognitive ability for both genders. This predicted talent measure alone explains more than 40 percent (corr=0.644) of the variation in males actual cognitive score in the analysis sample (including school dummies gives similar results). Finally, we standardize the measure to percentiles (1 to 100) within graduation year and gender to account for possible grade inflation and for the fact that females on average have better grades in high school.

Table A2: Average return to cognitive ability in different five-year periods (men, age 30 and 40). Standard errors in parenthesis.

	(1) **Age 30**	(2) **Age 40**
Rtrn to cognitive ability (1990-94)	**4.77**	**7.42**
	(0.00)	(0.00)
difference in 1995-99	0.05	0.50
	(0.58)	(0.00)
difference in 2000-04	0.57	1.71
	(0.00)	(0.00)
difference in 2005-09	-0.78	0.83
	(0.00)	(0.00)
difference in 2010-14	-1.51	-0.01
	(0.00)	(0.91)
N	808,213	832,946
R-sq	0.190	0.273
Sample	Men	Men

Notes: p-values are under the coefficients.

to these checks, as we see in the following.

C.1 Restricting the Samples

First, one might argue that plants or establishments, which are subunits of a firm with several locations of production, might provide a finer view of employer-level variation. In the estimation of Equation (1) the index j corresponds to the employer ID numbers from the Swedish organization registry. In Figure A4, Panel (a), we report estimates of the same returns using establishment level employer effects (only available from 1991) and we find very similar results (in fact, even more of the decline after 2001 can be accounted for through between-establishment variation). Results are also unchanged when the public and the primary sector are included, see Figure A4, Panel (b).

Second, returns to skills within small firms could be mis-measured if there is little signal (that is, small variation in workers' skills) relative to noise (potential measurement error in the skill variable). In fact, firms with only one worker are completely absorbed. In Figure A4 (Panels C and D) we therefore restrict the sample to large firms with either 5-to-19 or more than 20 workers on average, per year (males aged 30-40). These make up 19% and 53% of employment, and 12% and 4% of firms, respectively. The results are similar, suggesting that the inclusion of small firms does not to introduce much measurement error.

A recurring concern when using employer-employee data relates to spurious firm entry. For example, Hethey et al. (2010) use clusters of worker movements to identify true, as opposed to spurious (e.g. due to recoding), establishment entry and death in German administrative data. They find that only 40 percent of ID changes are true entries or exits. Skans et al. (2009), using Swedish data different from ours, only includes plants that exist in two consecutive years when studying changes in wage inequality over time. To examine whether spurious firm ID changes may be driving our results we restrict the sample only to firms which have neither entered in the current period nor exit in the subsequent period, see Figure A4, Panel (e). The key findings hardly change at all. In Figure A4, Panel (f), we also restrict the sample to firms which existed continuously between 1991 and 2013. This more than halves the sample size, from 9.1 to 4.1 million observations, but the overall dynamics of the return to cognitive skills remain fairly similar.

We also examine the extent to which broad economic sectors and detailed industries capture the effects of firm-level variation. In Figure A4, Panel (g), we show that fixed effects for 10 broad sectors hardly explain any of the reversal.[37] These broad sectors include information technology (IT): one might expect IT employment to be a factor in the run-up and reversal as it is fairly large and Sweden had a strong tech growth episode and subsequent bust.[38] Figure A4, Panel (h), displays the results when controlling for 5-digit industry fixed effects (amounting to 600 unique identifiers). Variation between detailed industries, and varying industry wages over time, account for half of the increase in the return to skills. This is not surprising, since industries are in fact aggregates of firms; this fine industry decomposition is consistent with the viewthat demand or productivity shocks at the firm-level may be correlated within an industry. However, comparing Figure 2 and Panel (h) of Figure A4 much variation remains unexplained by fine industry heterogeneity, confirming the importance of firm-level differences that go beyond even the very fine level of industry classification.

Finally, Figure 2, Panel J, shows the returns to non-cognitive skills. While firm-specific intercepts do not account for all of the run-up, they do explain about two thirds of its variation. Also, they capture between one third and one half of the decline after 2001. Therefore, variation between firms is important when accounting for changing returns to different skills over time.[39]

[37] The sectors are 'Manufacturing'; 'Health and Social Work'; 'Utilities, Construction, and Transportation'; 'Wholesale, Retail, and Hospitality'; 'Information Technology'; 'Finance'; 'Professional Services (Law, Consulting, Accounting)'; 'Other Business Services'; and 'Real Estate'.

[38] Figure A4 (i) shows the IT sector separately, and indeed we see that the return to skill in that sector exhibited a similar dynamic as the overall return to skill in the economy, but it was not the driver of the economy-wide changes (we have also removed IT from the sample and found similar results).

[39] Panels K and L of Figure 2 consider the role of occupation fixed effects and show that permanent between-occupation effects do not account for much of the run-up and reversal, while time-varying occupation effects appear more important.

C.2 Returns to Non-cognitive Ability

Controlling for returns to non-cognitive skills does not significantly change the baseline estimate of the return to cognitive ability. Table A3 reports results from linear earnings regressions for five-year periods on individual ability of the following form:[40]

$$\ln w_{i\tau} = \alpha_\tau + \lambda_\tau ST_i + \varepsilon_{i\tau}. \tag{25}$$

Here $\tau\epsilon\{1990–94, 1995–99, 2000–04, 2005–09, 2010–14\}$ indexes the five-year interval, while $\ln(w_{i\tau})$ are $100 \times log$ annual earnings and ST_i contains individual i's ability score in the Stanine scale (cognitive or non-cognitive, respectively).

Each column reports the difference relative to the baseline return in 1990-94 (in bold). The first two columns of Table A3 confirm that the return to cognitive ability was significantly up to the period 2000–04, after which it became lower than in 1990-94. Columns 3 and 4 consider the return to non-cognitive ability. In level terms, this is roughly similar to that on cognitive ability. Also this return rises up to the early 2000s, then declines somewhat.

The reversal in the aggregate return to cognitive skills is robust to the inclusion of non-cognitive measures. Columns 5 and 6 report estimated returns to cognitive and non-cognitive ability when both proxies are accounted for. The increase in the return to cognitive ability up to 2000–04 is somewhat weaker but still sizable and significant. The first row of Appendix Figure A3 plots the unconditional relative earnings by non-cognitive ability group. As in Figure 1 a hump-shaped is apparent. The reversal after the early 2000s is weaker, especially for forty year olds, but there is clearly an inflection point in the return to non-cognitive skills. Overall, the returns to non-cognitive skills have increased over the sample period. To the extent that non-cognitive measures contain a social skill component, this is consistent with evidence for the US by Deming (2016). However, also for non-cognitive skills we note an increase during the 1990s and a later slowdown.

C.3 Returns to Education

Figure A1 depicts the relative wages of four education levels relative to high-school graduates by age and gender over time. Individuals with less than high-school have earnings lower than those with a high-school degree in all age groups and for both genders. Individuals with some college education, a university degree, and a PhD earn more than high-school graduates and the differential rises with age. Notable facts about earnings levels are that male PhD graduates at younger ages do not earn more than university graduates and that the premium for "some college" workers is small and declines to zero for thirty year old males in the later sample years. Since education information is available for all individuals and sample years regardless of birth cohort, the bottom row of Figure A1 graphs relative earnings of 50 year olds by education level. Again the level differences are larger in this relatively old group of workers.

Turning to the trends in relative earnings by education level, these are in parts as striking

[40] Effects are in log points. For example, in column 2 in the base period 1990-1994, a one standard deviation (two scores) increase in cognitive ability is associated with fifteen log points higher wages (approx. 2 x 7.42).

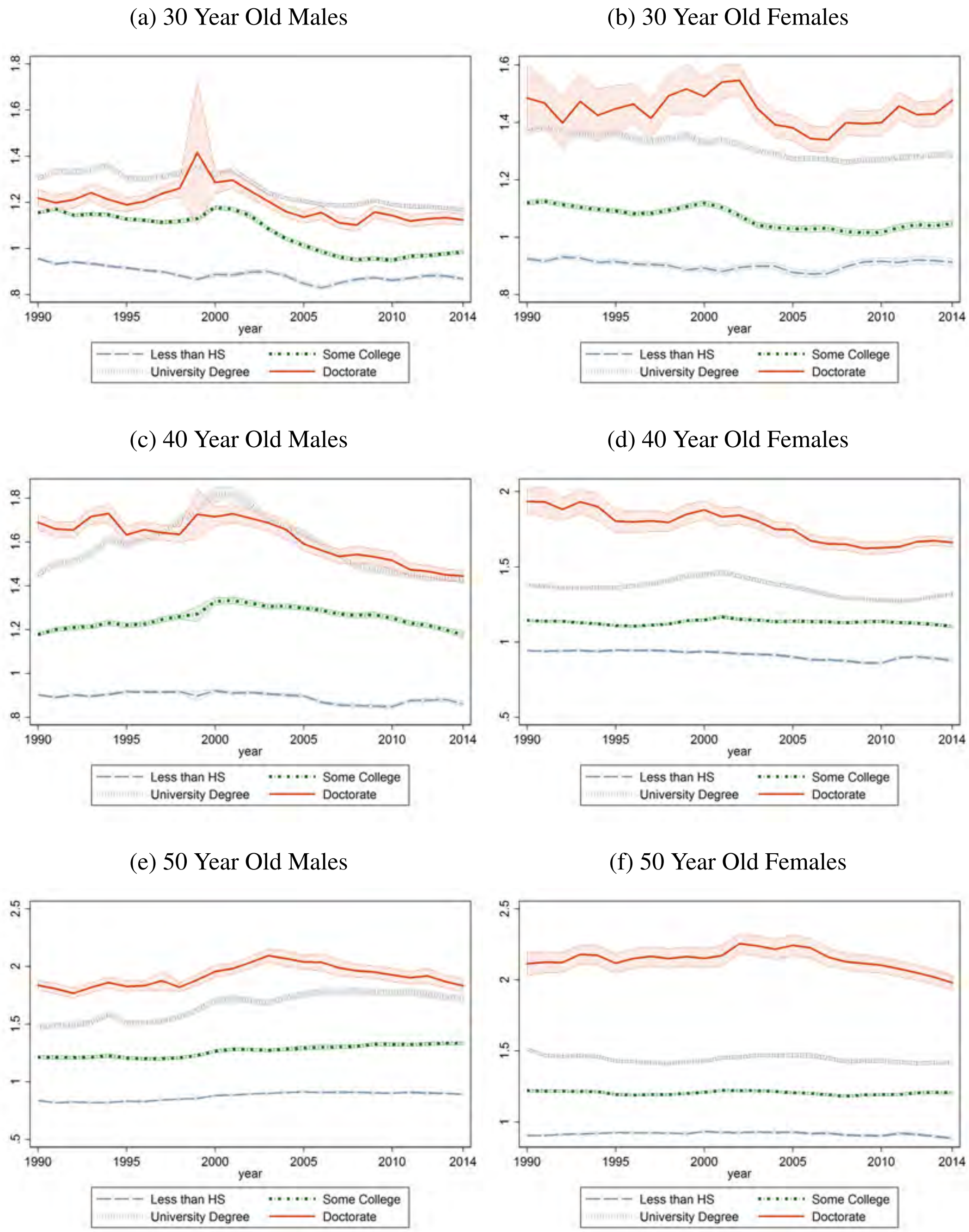
Table A3: Average return to ability in different five-year periods (men, age 30 and 40). Standard errors in parenthesis.

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

Notes: p-values are under the coefficients.

as the trends in returns to cognitive ability. Figure A1 shows a clear and strong decline after the early 2000s in the returns to all postsecondary education levels for thirty and forty year old males. This is strongest for the relative wages of forty year old university graduates, which first increase from 40 to 80 percent during 1990–2001 and then precipitously decline back to

Figure A1: Relative Earnings by Education Levels



40 percent in 2014. Such a “hump-shaped” evolution over time can also clearly be seen for 40 year old female university graduates and 40 year old males with some college.

Instead of an overall “hump-shape”, all thirty year old postsecondary education levels’ relative earnings are better described as relatively flat during most of the 1990s, a small temporary budge around the 2000s, and a decline after that. Since attainment rises across the board and consequently the average ability of graduates declines (e.g., Böhm et al., 2018, for similar Swedish data), it is not clear in how far the flatness during the 1990s reflects the counteracting forces of rising demand for skill versus falling quality of workers by skill level. This is a key advantage of focusing on the ability measures in this paper.

In case of 50 year olds, the relative wages of male as well as female PhD graduates start declining around 2003, while the relative wage of male university graduates flattens out after rising during the end of the 1990s. The remaining relative wages appear largely constant over the sample period. One explanation of these facts could be that on the one hand the demand for skilled older workers also reverts after the turn of the century, but that on the other hand consecutive cohorts of 50 year old educated workers retain some of the relative earnings gains they have made during the previous period so that overall there is only a small decline or a flattening out of relative wages.

Overall, however, the accelerated decline in 30 and 40 year old educated workers' relative wages after the turn of the century indicates that the returns to education have reversed, too. This seems somewhat in contrast to Castex and Kogan Dechter (2014)'s findings of rising returns to education in the U.S..

Table A4 reports results from linear earnings regressions on ability and education controlling for the respective other measure, again using five-year periods and the number of years of schooling as a linear summary measure for an individuals' educational attainment.[41] In the first two columns of Table A4, the returns to cognitive ability conditional on schooling rise and then decline in the 2000s as in Table A3, though the reversal is weaker for forty year olds. This is qualitatively similar for the conditional returns to non-cognitive ability, but the increase in the 1990s is stronger for forty year olds and the decrease in the 2000s is weaker for thirty year olds and there is no overall decrease for forty year olds. At the same time, the return to schooling perform worse in both decades when controlling for non-cognitive than for cognitive ability. This suggests that, if education as a skill contains cognitive and non-cognitive components, the return to the cognitive component (i.e., controlling for non-cognitives) has performed worse over time than the return to the non-cognitive component (i.e., controlling for cognitives).

Columns 5–6 of Table A4 report the returns to schooling controlling for cognitive and non-cognitive ability at the same time. As in Columns 1–4 and in Table A3, the returns to cognitives perform worse than the returns to non-cognitives, which do not even decline for forty year olds. As in the previous columns, the returns to schooling decline overall, but they decline during the 1990s mainly for thirty year olds and during the 2000s mainly for 40 year olds. Part of this is likely a selection effect whereby education attainment rises (and ability declines) more strongly for birth cohorts that are around 30 years old in the 1990s and around

[41] Regression Equation (24) becomes

$$logearn_{i\tau} = \alpha_\tau + \beta_\tau ability_i + \gamma_\tau yrsofschl_i + \varepsilon_{i\tau},$$

with $yrsofschl_i$ being a measure of the number of individual i's years of schooling based on their highest degree reported in the LISA data.

Table A4: Returns to Education and Ability in Different Five-Year Periods

	1	2	3	4	5	6	7	8
	Men,30	**Men,40**	**Men,30**	**Men,40**	**Men,30**	**Men,40**	**Men,30**	**W'en,30**
Return to yrs of school (1990-94)	**3.57**	**4.58**	**3.72**	**5.42**	**3.04**	**4.31**	**3.92**	**4.51**
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
difference in 1995-99	-0.10	1.39	0.02	1.10	-0.14	1.13	-0.48	-0.08
	0.24	0.00	0.76	0.00	0.13	0.00	0.00	0.50
difference in 2000-04	-1.27	1.57	-0.99	1.40	-1.40	1.13	-1.63	-0.72
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
difference in 2005-09	-1.20	0.19	-1.38	-0.04	-1.24	-0.32	-2.94	-1.13
	0.00	-0.01	0.00	0.56	0.00	0.00	0.00	0.00
difference in 2010-14	-1.13	-0.44	-1.41	-0.89	-1.00	-0.92	-3.56	-1.18
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Rtrn to cognitive ability (1990-94)	**2.47**	**4.12**			**1.63**	**3.22**		
	0.00	0.00			0.00	0.00		
difference in 1995-99	0.05	-0.04			0.23	-0.16		
	0.66	0.66			0.03	0.11		
difference in 2000-04	1.30	0.86			1.15	0.32		
	0.00	0.00			0.00	0.00		
difference in 2005-09	0.05	0.81			-0.20	0.31		
	0.65	0.00			0.06	0.00		
difference in 2010-14	-0.59	0.38			-0.87	-0.11		
	0.00	0.00			0.00	0.28		
Rtrn to non-cogn. ablty (1990-94)			**3.90**	**4.30**	**3.48**	**3.60**		
			0.00	0.00	0.00	0.00		
difference in 1995-99			0.08	0.33	0.13	0.26		
			0.45	0.00	0.24	0.01		
difference in 2000-04			1.25	2.30	1.08	2.04		
			0.00	0.00	0.00	0.00		
difference in 2005-09			0.72	3.18	0.80	3.06		
			0.00	0.00	0.00	0.00		
difference in 2010-14			0.24	2.49	0.46	2.48		
			0.03	0.00	0.00	0.00		
Return to HS GPA pctile (1990-94)							**0.15**	**0.08**
							0.00	0.00
difference in 1995-99							0.04	0.05
							0.00	0.00
difference in 2000-04							0.12	0.14
							0.00	0.00
difference in 2005-09							0.15	0.18
							0.00	0.00
difference in 2010-14							0.14	0.19
							0.00	0.00
N	807,590	832,104	807,590	832,104	807,590	832,104	789,284	701,580
R-sq	0.200	0.310	0.211	0.323	0.214	0.332	0.203	0.192
Sample	Men	Men	Men	Men	Men	Men	Men	Women

Notes: p-values are under the coefficients.

forty years old in the 2000s (again, see Böhm et al., 2018, for this evidence in Swedish data).

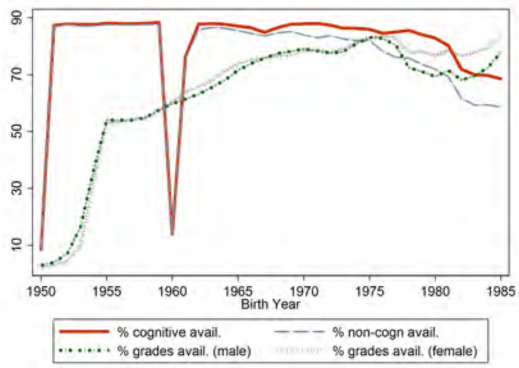
The last two colums of Table A4 report the returns to grade percentiles and schooling conditioning on each other. Contrary to Table A3, the returns to grades now rise throughout the sample, for women as well as for men, although the increase slows down after 2000–04. The conditional return to schooling, on the other hand, declines throughout and very strongly for men so that it is approximately zero in the end of the sample.

Overall, the evidence in this section points to the following conclusions. The returns to cognitive skills *for males* seem to indeed have completed a cycle during the period 1990–2014, first increasing and then reverting. The returns to non-cognitive skills seem to have increased overall, albeit most of this occurred during the 1990s and the 2000s were largely flat. The returns to high-school grades are somewhere in between the two. The returns to education have also experienced a reversal, with the overall performance stronger for its non-cognitive than its cognitive component. Finally, the returns to skills *for females* performed substantially better than for males, but again there was an inflection point toward weaker increases at the turn of the century.

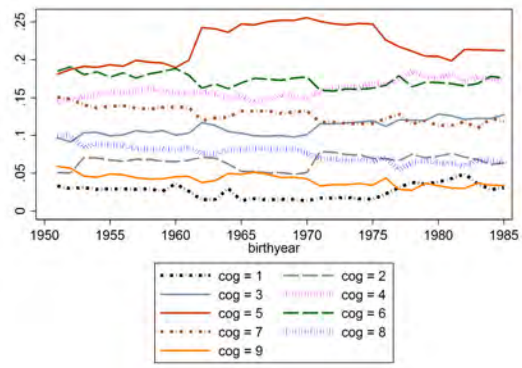
D Additional Figures and Tables

Figure A2: Skill Availability by Birth Year and Distribution by Birth Year

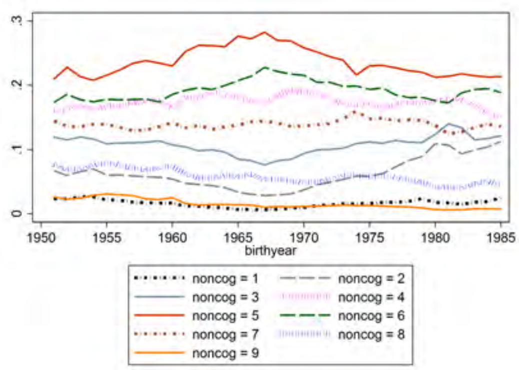
(a) Skill Availability



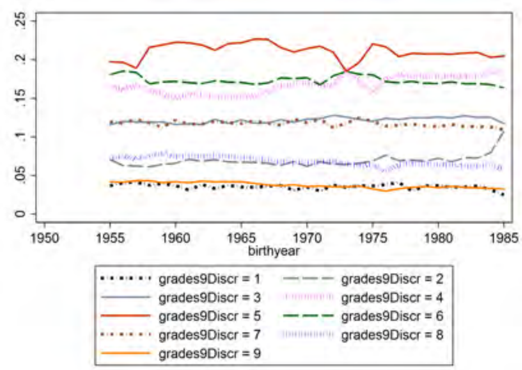
(b) Males' Cognitives Distrib.



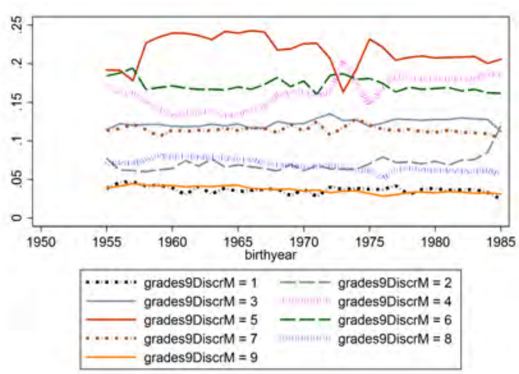
(c) Males' Non-Cogn. Distrib.



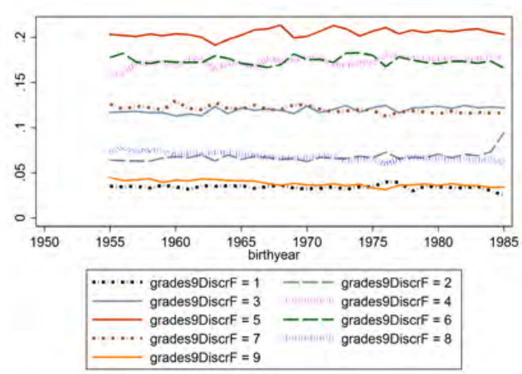
(d) Grades Distr (Both Genders)



(e) Males' Grades Distr



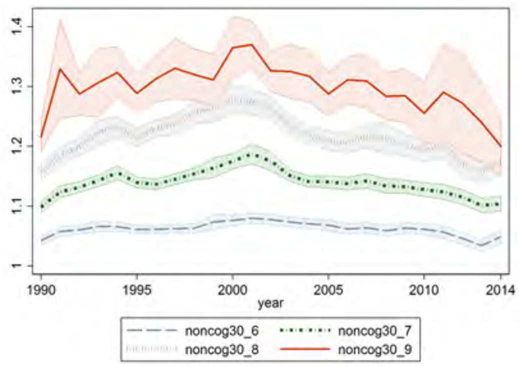
(f) Females' Grades Distr



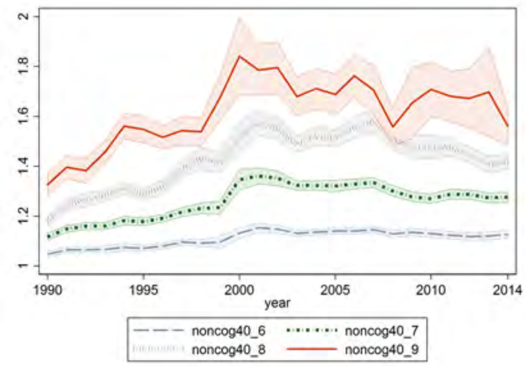
For the purpose of the graph, females' high-school grade percentiles and predicted cognitive skill percentiles were discretized to a Stanine distribution, too.

Figure A3: Returns to Non-Cognitives, Grades, and Predicted Cognitives

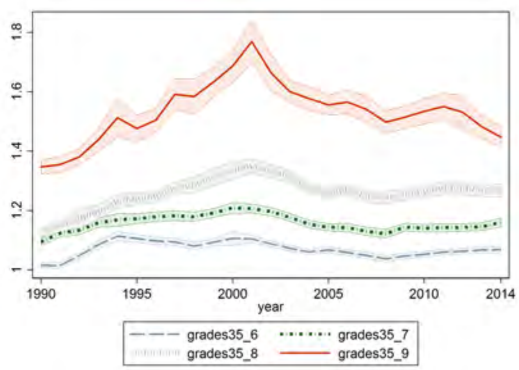
(a) Non-Cognitives, 30 Yr Old Males (High)



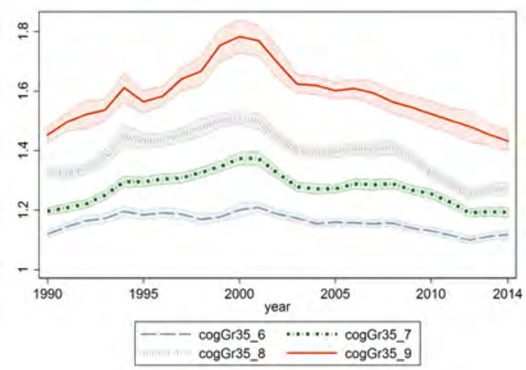
(b) Non-Cognitives, 40 Yr Old Males (High)



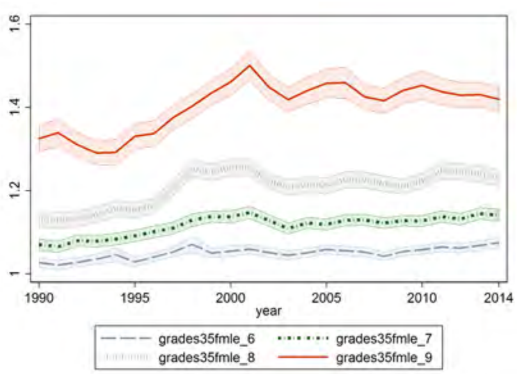
(c) HS Grades, 35 Yr Old Males (High)



(d) Predicted Cogn., 35 Yr Old Males (High)



(e) HS Grades, 35 Yr Old Females (High)



(f) Predicted Cogn., 35 Yr Old Females (High)

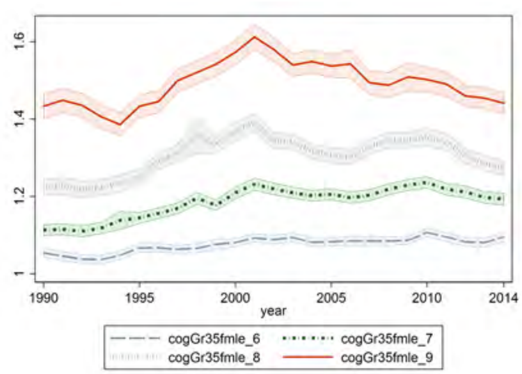
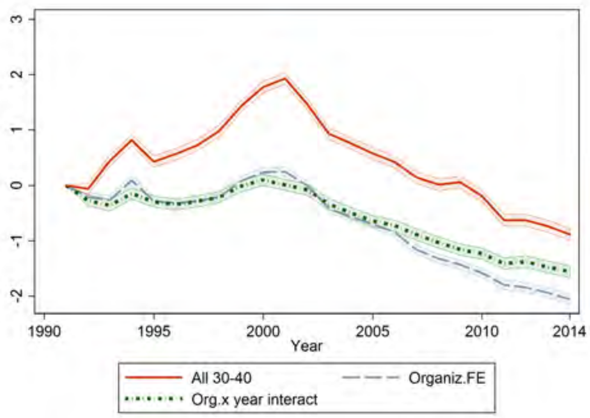
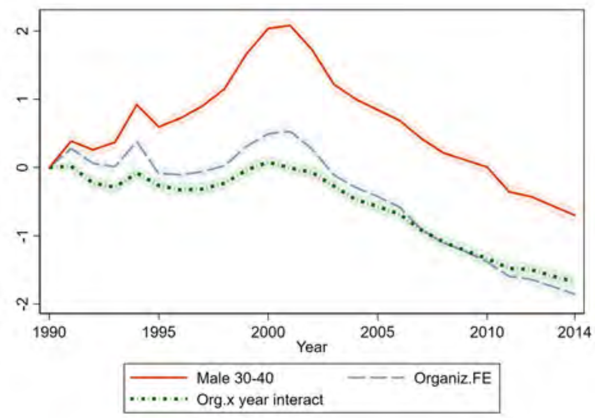


Figure A4: Robustness of Firm Intercept Result

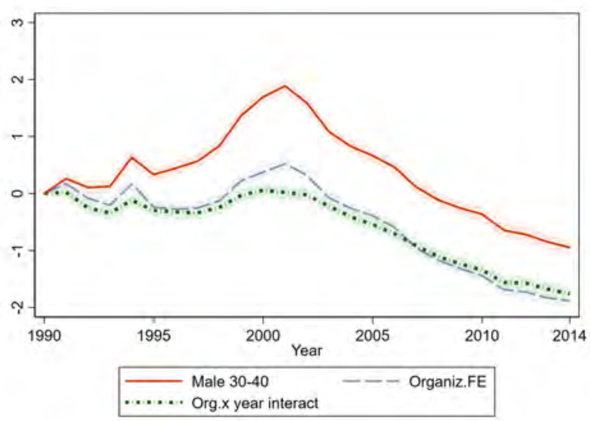
(a) Establishment Fixed Effects



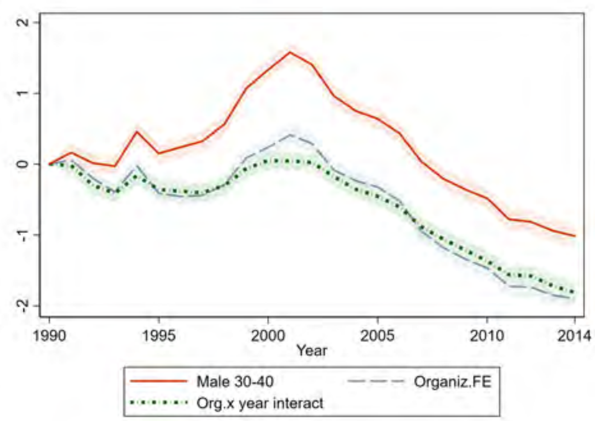
(b) With public&primary sector



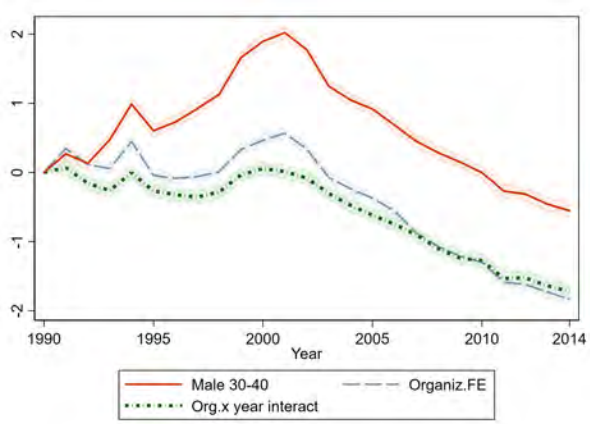
(c) Firms between 5 and 19 workers



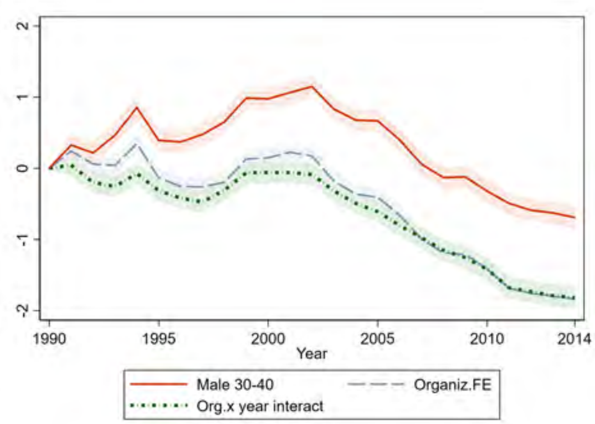
(d) Firms with at least 20 workers



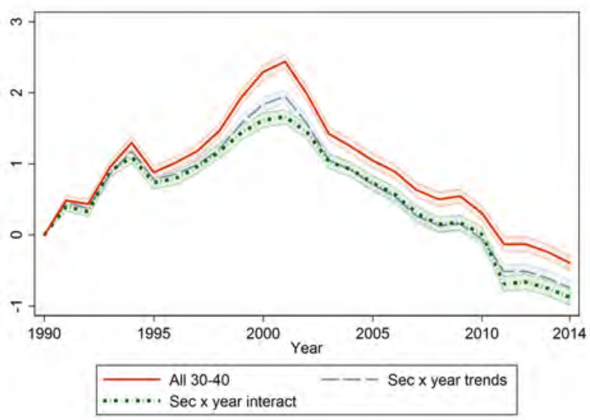
(e) Firms no entry/exit in current year



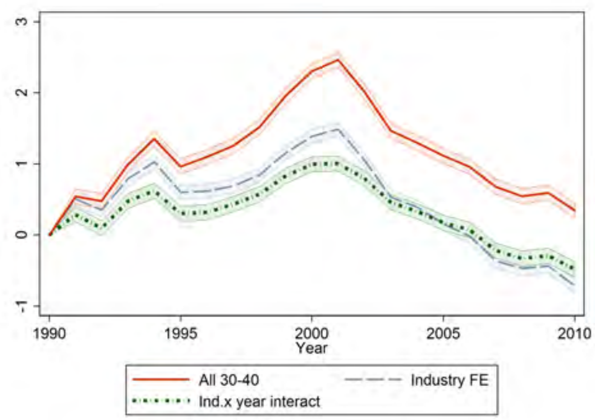
(f) Firms exist since 1991 and until 2013



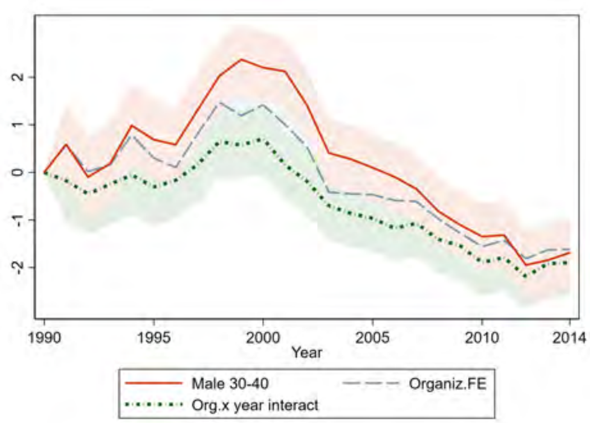
(g) Broad Sector Fixed Effects



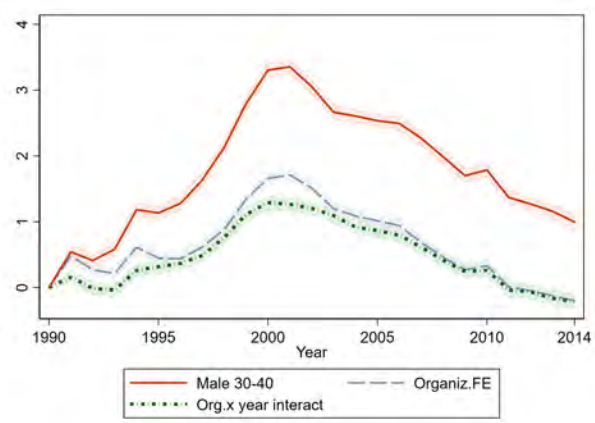
(h) Detailed Industry Fixed Effects



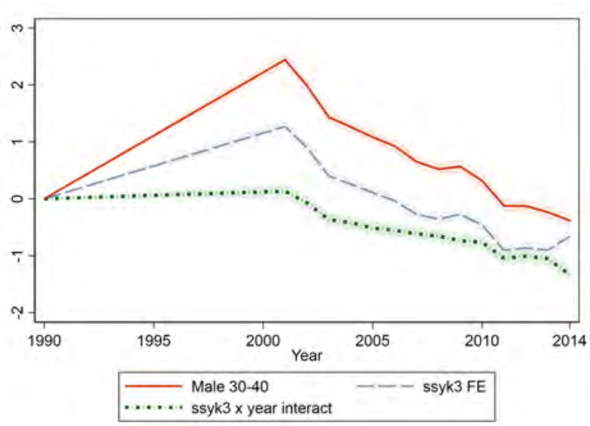
(i) Only IT Sector



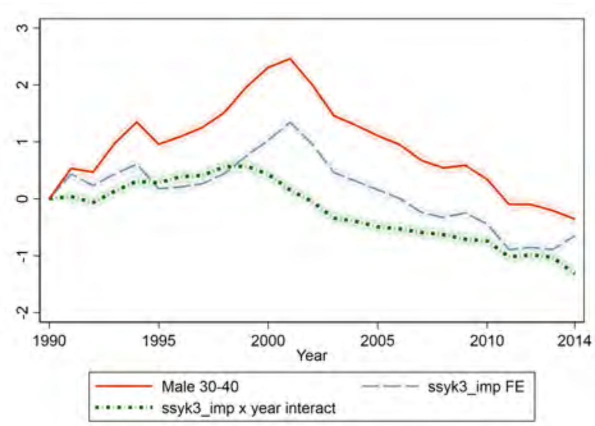
(j) Return to non-cognitive skills



(k) Occupation Fixed Effects



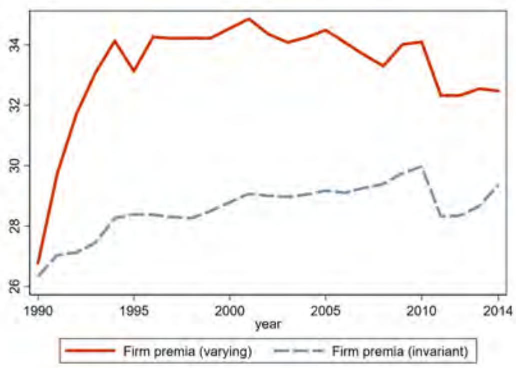
(l) Imputed Occupations Fixed Effects



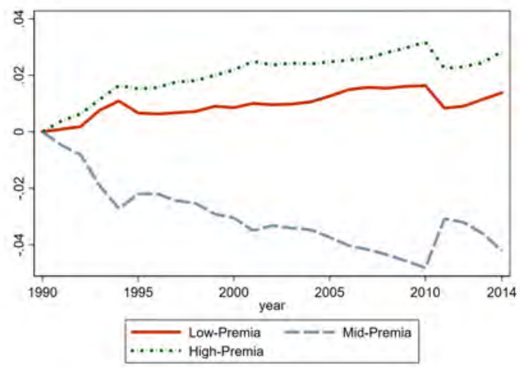
Ages 30–40 are pooled. In all specifications controls are year and (non-)cognitive ability interacted with age. The outcome variable is in log points.

Figure A5: Evolution and Sorting into Firm Wage Effects (Additional Graphs)

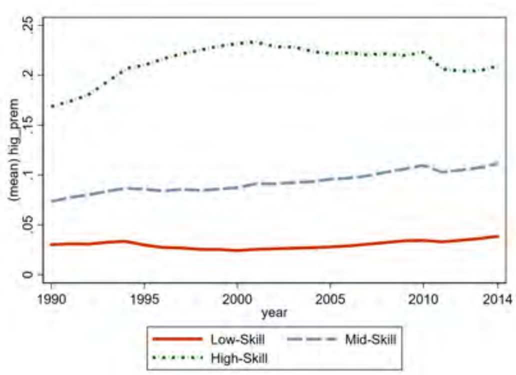
(a) Dispersion of Firm Effects θ_{jt} and $\bar{\theta}_j$



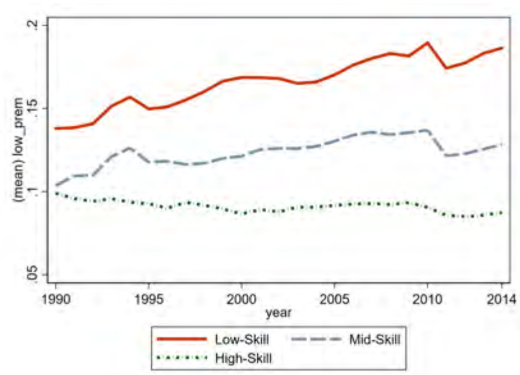
(b) Firm Effects $\bar{\theta}_j$ Employment Trends



(c) Sorting into High-Wage $\bar{\theta}_j$ Effects



(d) Sorting into Low-Wage $\bar{\theta}_j$ Effects



Panel (a) shows the variance of firm wage effects θ_{jt}, $\bar{\theta}_j$, and $\theta_{jt} - \bar{\theta}_j$ over time. Panel (b) shows employment trends in high-wage (one $sd(\bar{\theta}_j)$ above mean in respective year), low-wage (one $sd(\bar{\theta}_j)$ below), and middle-wage (all other) firms where $sd(\bar{\theta}_j)$ is calculated over all years. The bottom panels show the unconditional probability for high-skill (cognitive 7–9), middle-skill (4–6), and low-skill (1–3) individuals to work in high- and low-wage firms.

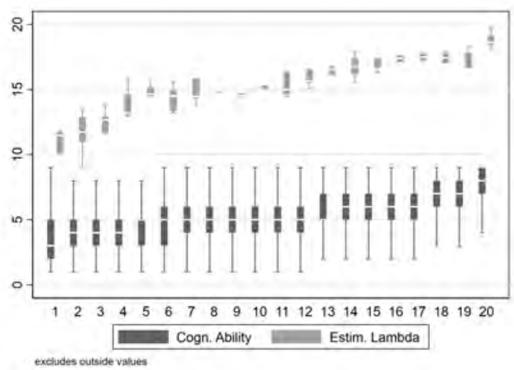
Table A5: Cross-Sectional Estimation Results by Firm Cognitive Ability Decile

cog decile	% empl	nbrempl	lambda	theta	theta(inv)	per fe	avg cog
1	10.0	2.6	11.4	-10.5	-11.3	-10.5	3.6
		7	*2*	*35*	*31*	*26*	*1*
2	10.1	5.9	15.4	-5.4	-4.8	-6.0	4.3
		23	*1*	*30*	*19*	*19*	*1*
3	10.1	6.9	16.7	-4.5	-3.5	-3.5	4.6
		36	*1*	*29*	*18*	*18*	*1*
4	9.9	6.2	16.8	-4.7	-4.9	-2.6	4.8
		49	*1*	*29*	*20*	*19*	*1*
5	10.2	5.5	17.6	-4.2	-4.1	-2.9	5.0
		46	*1*	*31*	*22*	*19*	*1*
6	10.0	8.7	17.4	-3.7	-1.1	-2.2	5.2
		80	*1*	*31*	*19*	*18*	*1*
7	9.9	7.5	18.7	-1.5	-0.4	-0.7	5.5
		54	*1*	*33*	*20*	*20*	*1*
8	10.0	6.5	19.7	5.2	5.0	5.6	5.8
		43	*1*	*35*	*21*	*21*	*1*
9	9.9	8.4	19.1	10.0	9.8	10.6	6.2
		55	*1*	*36*	*20*	*22*	*0*
10	9.9	5.5	20.1	18.0	14.8	10.0	7.0
		47	*1*	*37*	*25*	*25*	*1*
Average	**10.0**	**5.7**	**17.3**	**0.0**	**0.0**	**-0.1**	**5.2**
		43	*3*	*34*	*23*	*22*	*1*

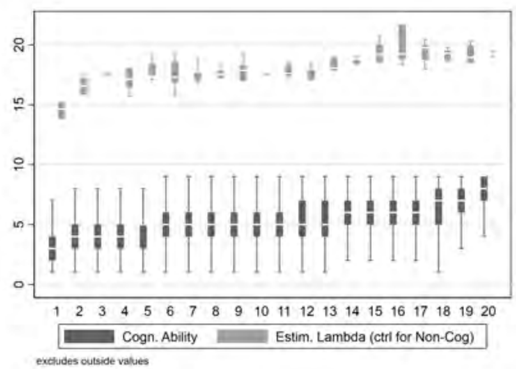
Notes: The Table summarizes the main results (levels on top and standard deviations below) of estimating Equ. (19) in the cross-section by cognitive ability bin, collected into groups of ten deciles. The first two columns show the share of total employment and average number of employees per firm in each bin. The remaining columns report the (firm employment weighted) estimates of λ_g, $\theta_{j,t}$, $\bar{\theta}_j$ (estimation with time-invariant firm effects), person effect μ_i, and average cognitive ability per firm, respectively.

Figure A6: Estimation Results for Firm Skill-Bias

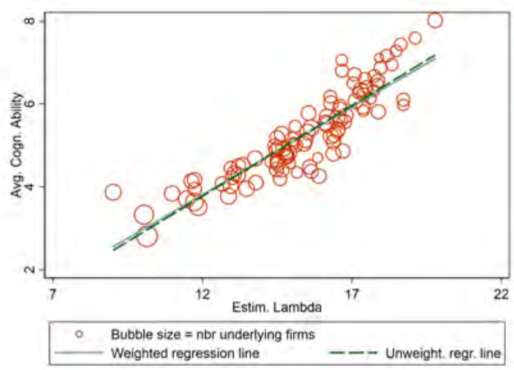
(a) λ_g & cogn.distr (invariant θ_g)



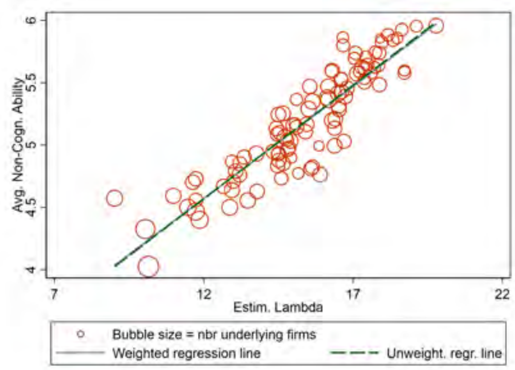
(b) λ_g & cogn.distr (noncog cntrl)



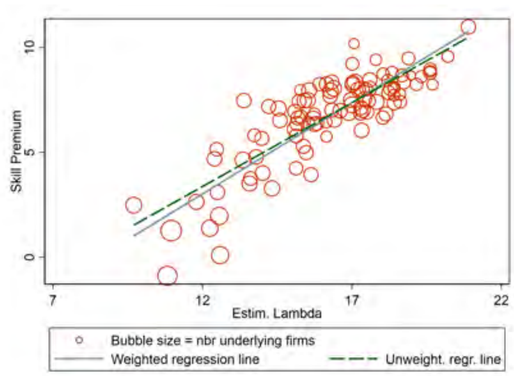
(c) λ_g and skill sorting (cognitive)



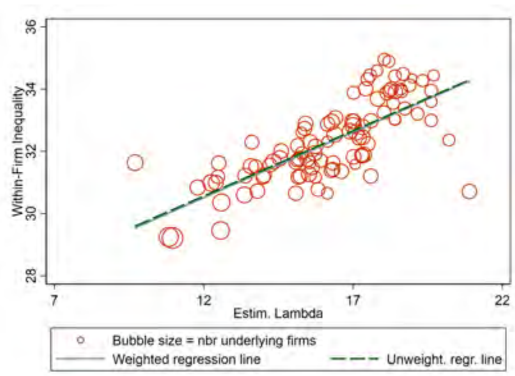
(d) λ_g and skill sorting (non-cogn)



(e) λ_g & skill premium (non-residlzd)



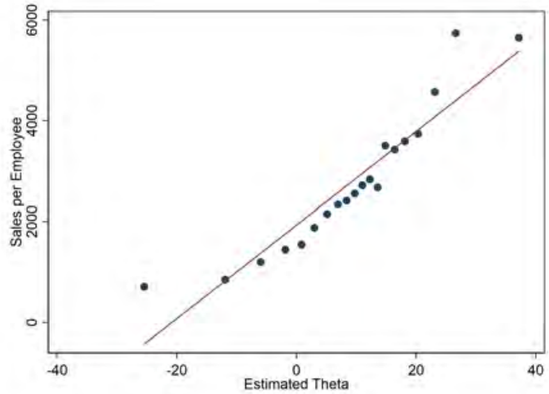
(f) λ_g & wage inequality



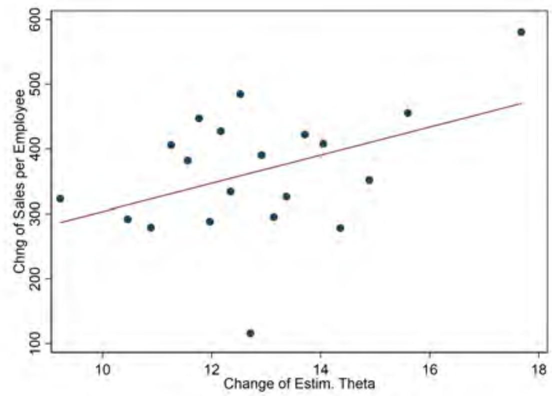
Notes: The top panels show boxplots of the distribution of workers' cognitive ability in the different firm bin vingtiles (i.e., the variation we use for identification) and of the estimated λ_gs in the vingtiles (five each). Panel (a) is estimated without person fixed effects in (19), Panel (b) with person fixed effects and including non-cognitive skills interacted with firm bin. The middle panels plot the estimated λ_gs in our main specification (person FE, no non-cog controls) against the average cognitive and non-cognitive ability in each of the 100 firm bins. The bottom panels plot the non-residualized skill premium and wage inequality agains λ_gs in each of the bins.

Figure A7: Estimation Results for Firm Wage Effects (Levels and Four-Year Differences)

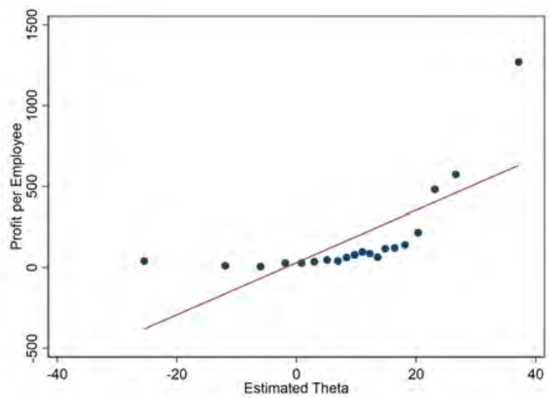
(a) $\theta_{g,\tau}$ and firm sales / worker



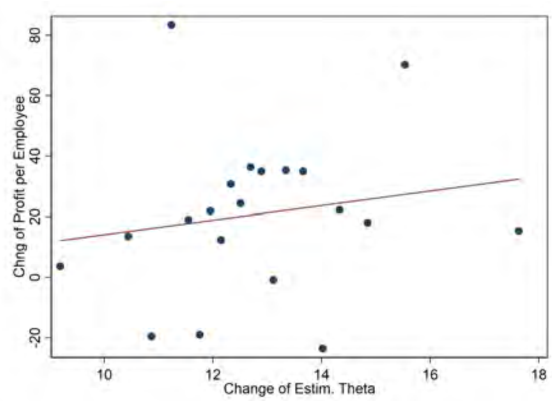
(b) $\theta_{g,\tau}$ and firm sales / worker (4-year diff.)



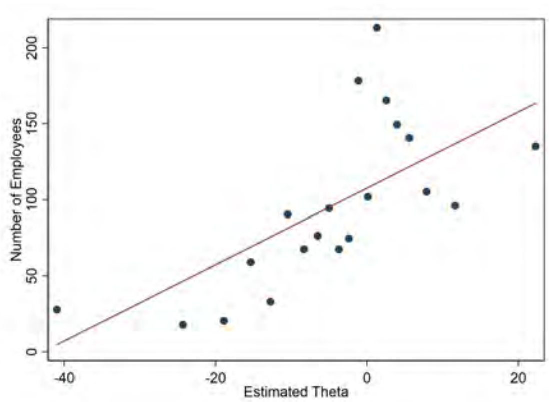
(c) $\theta_{g,\tau}$ and firm profits / worker



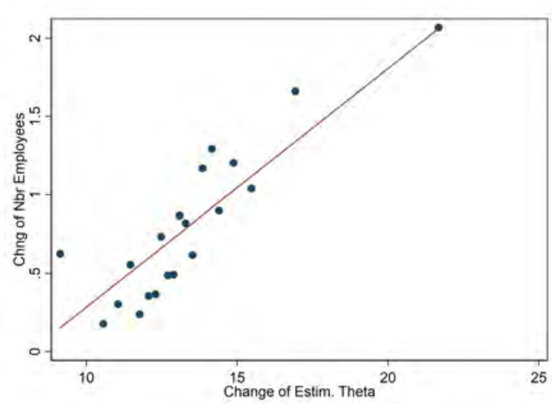
(d) $\theta_{g,\tau}$ and firm profits / worker (4-year diff.)



(e) $\theta_{g,\tau}$ and number of workers



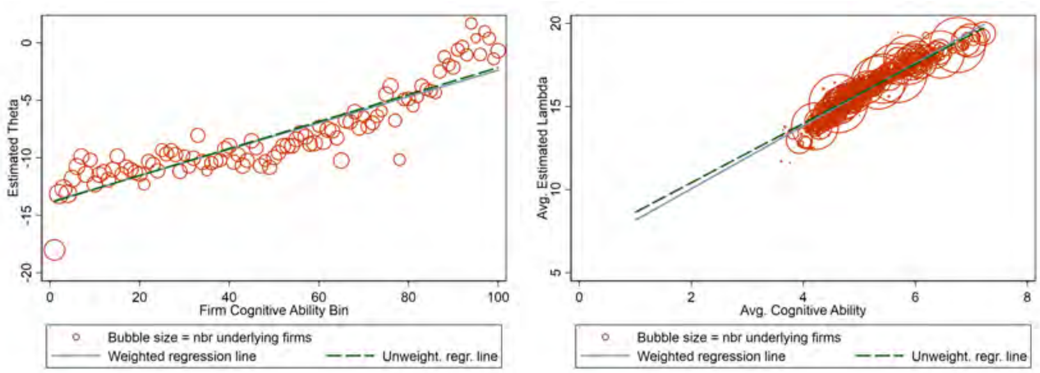
(f) $\theta_{g,\tau}$ and number of workers (4-year diff.)



Notes: The left panels show bin scatter plots of the change in average estimated $\theta_{g,\tau}$ together with the change in firm characteristics. The right panels show this in differences between four-year periods (1990–93, 1994–97, 1998–2001, 2002–05, 2006–2009, 2010–2014). These firm characteristics sales/employee, profits/employee, and number of employees. We weight by firms' employment numbers, and remove outliers above top and below the bottom one percent for each variable and only consider firms present in all periods.

Figure A8: Sources of Sorting and Average Skill by Industry

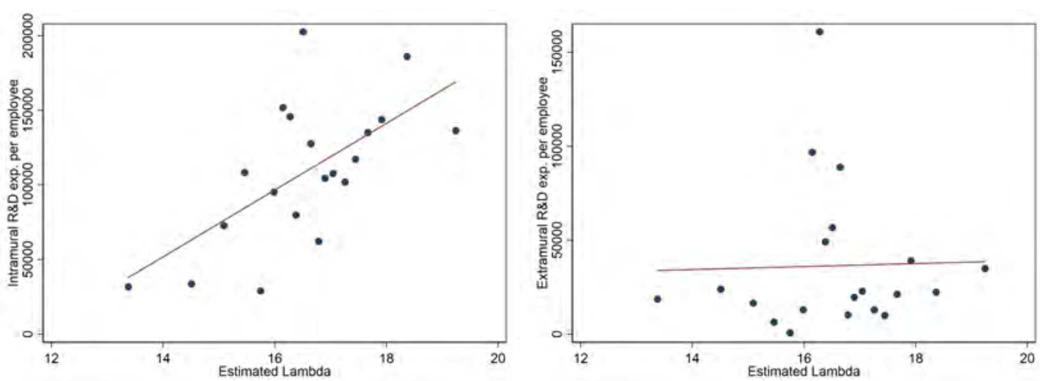
(a) θ_g & Firm Cognitive Ability Bin (b) λ_g and skill sorting by 5-digit industry



Notes: Left panel shows the average estimated firm effect θ_g by firm cognitive ability bin rank. Right panel estimated λ_g against average cognitive ability by 5-digit industry (605 unique codes). A linear regression line is fit through these average, weighted and unweighted by the underlying number if firms.

Figure A9: Skill Returns λ_g and R&D per worker

(a) θ_g & Firm Cognitive Ability Bin (b) λ_g and skill sorting by 5-digit industry



Notes: Left plot are estimated λ_g and *intramural* $R\&D$ per worker, right plot *extramural* $R\&D$. The hypothesis is that skill intensive firm characteristics (such as *intramural* $R\&D$) are *differentially* strongly related to firm skill bias λ_g. All correlations are conditional on year, sector fixed effects, and θ_g.

E Numerical Solution of the Model: Sequence of Steps

This appendix lays out the algorithm used to solve the model given (i) a sample of firm-specific parameters $\chi_j = \{(\lambda_j, \log(\phi_j), \log(a_{js_H}), \log(a_{js_L}), T_j)\}_{j=1}^{J}$, where j is a firm identifier; and (ii) a set of aggregate parameters β, σ, s_H, s_L, N_{s_H}, and N_{s_L}.

The steps are as follows:

1. guess ξ_{s_H}, ξ_{s_L}, and Y.
2. for each j, calculate w_{js_H} and w_{js_L} using equation (11).
3. for each j, calculate q_{js_H} and q_{js_L} using equation (4).
4. for each j, calculate y_j using equation (5).
5. update Y using equation (6); also update ξ_{s_H} and ξ_{s_L} using equation (3).
6. Check for convergence in ξ_{s_H}, ξ_{s_L}, and Y; if convergence is still higher than tolerance level, go to step 2; otherwise a solution is achieved.