The Innovation Premium to Low Skill Jobs *

Philippe Aghion   Antonin Bergeaud

Richard Blundell   Rachel Griffith

October 1, 2017

Abstract

This paper uses matched employee-employer data from the UK that we augment with information on R&D expenditures, to analyze the relationship between innovativeness and average wage income for different skills across firms. We first show that more R&D intensive firms pay higher wages on average. Our second and main finding is that the premium to working in more R&D intensive firms is higher for low-skilled workers than for high-skilled workers. As technology advances, demand for high skilled workers increases and they do better overall, but low skilled workers in innovative firms do better than other low-skilled workers. To account for these findings, we develop a simple model of the firm where the complementarity between high-skill occupation and low-skill occupation employees within the firm increases with the firm's degree of innovativeness. An additional prediction of the model, which is also confirmed by the empirical analysis, is that low-skilled workers stay longer in more innovative firms.

*Aghion: College de France and LSE; Bergeaud: Banque de France and PSE; Blundell: UCL, IFS and CEPR; Griffith: University of Manchester, IFS and CEPR. The author wish to thanks, without implicating, participants at PSE lunch seminar and at the Income Distribution and Macroeconomic session of the NBER SI 2017. We also thank the Centre for the Microeconomic Analysis of Public Policy (CPP) under grant number RES-544-28-0001 and the UK Data Service Team for their support. All results, tables and figures have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

High and persistent income inequality in developed countries has attracted considerable attention, including most notably in the US and the UK (e.g. see Deaton, 2013, Atkinson, 1997 and Piketty, 2014); according to the OECD, workers in the top 10% of the wage distribution in the UK earn on average over 10 times more than workers in the bottom of the wage distribution; in the US it is almost 19 more. A leading explanation for the acceleration in income inequality over the past two decades, is that there has been an acceleration in the rate of skill-biased technical change (e.g. see Krusell et al, 2000; Acemoglu, 2002).

In this paper, however, we shall argue that while the prediction of a premium to skills holds at the macroeconomic level, it misses important aspects of the internal organization of innovating firms. More specifically, we use matched employer-employee data from the UK, which we augment with information on R&D expenditures, to analyze the relationship between innovativeness and average wage income at different skill levels across firms. Our main finding is that lower-skilled workers benefit more from working in more R&D intensive firms, relative to working in a firm which does no R&D) than higher-skilled workers. This finding is summarized by Figure 1. In that Figure, we first see that higher-skilled workers earn more than lower-skilled workers in any firm no matter how R&D intensive that firm is (the high-skill wage curve always lies strictly above the middle-skill curve which itself always lies above the lower-skill curve). But more interestingly the lower-skill curve is steeper than the middle-skill and higher-skill curve. But the slope of each of these curves precisely reflects the premium for workers with the corresponding skill level to working in a more innovative firm.

To rationalize the above finding, we propose a model in which more innovative firms display a higher degree of complementarity between low-skill workers and the other production factors (capital and high-skill labor) within the firm. Another feature of the model is that high-occupation employees' skills are less firm-specific than low-skill workers: namely, if the firm was to replace a high-skill worker by another high-skill worker, the downside risk would be limited by the fact that higher-skill employees are typically more educated employees, whose market value is largely determined by their education and accumulated reputation, whereas low-occupation employees' quality is more firm-specific. This model is meant to capture the idea that low-occupation workers can have a potentially more damaging effect on the firm's value if the firm is more innovative.

In particular an important difference with the common wisdom, is that here innovativeness impacts on the organizational form of the firm and in particular on complementarity or substitutability between workers with different skill levels within the firm, whereas the common wisdom view takes this complementarity or substitutability as given. Think of a low-occupation employee (for example an assistant) who shows outstanding ability, initiative and trustworthiness. That employee performs a set of tasks for which it might be difficult or too costly to hire a high-skill worker; furthermore, and perhaps more importantly, the low-occupation employee is expected to stay longer in the firm than higher-skill employees, which in turn encourages the firm to invest more in trust-building and firm-specific human capital and knowledge. Overall, such low-occupation employees can make a big difference to the firm's performance.

The model delivers additional predictions. In particular it also predicts that job turnover should be lower (tenure should be higher) amongst low-skilled workers who work for R&D-intensive firms than for low-skilled workers who work for non-R&D intensive firms, whereas the turnover difference should be less between high-skilled workers employed by these two types of firms. This additional prediction is confronted to the data in the last part of the paper.

The paper relates to several strands of literature. First, there is the labor and wage literature, starting with the seminal work of Abowd et al. (1999); this literature has agreed that firms' heterogeneity play a large role in explaining wage differences across workers; however, there is no consensus in explaining which features of the firm account for such variation.[1] Other studies report a link between productivity and wage policy (Cahuc et al., 2006 and Barth et al., 2014 among others). Song et al. (2015) cite outsourcing as a potential explanation for the raise of between firm inequality. We argue that a source of variation in firm's propensity to pay higher wages than other has to do with innovation intensity. This result echoes those of Van Reenen (1996), who showed that innovative firms pay higher wages on average, using information on public listed UK firms.

Second, there is the literature on wage inequality and skill-biased technical change (e.g. see Acemoglu, 2002; Goldin and Katz, 2010, Acemoglu and Autor, 2011). While this literature focuses on explaining the accelerated increase in the skill premium, we focus on the relationship between innovation and between-firm wage inequality, with

[1] For example, Card et al. (2016) assume that firm heterogeneity arises through TFP, but do not model what drives these differences in TFP

the surprising finding that the premium to working in a more innovative firm is higher for lower skilled workers.

Third, there is the recent empirical literature on innovation, inequality and social mobility (e.g. see Bell et al., 2016, Aghion et al., 2015 and Akcigit et al., 2017). We contribute to this literature by introducing firms into the analysis and focusing on the relationship between innovation and between-firm income inequality.

Fourth, and more closely related to our paper is the literature linking the aggregate dispersions in wages to productivity dispersion across firms (Barth et al., 2014, Dunne et al., 2004). Part of this literature uses matched worker-employee data (see Card et al., 2016 for a review) to investigate whether this correlation represents differences in workers selected into different firms, or the same type of worker being paid a different wage depending on the firm they work in. Abowd et al. (1999) pioneered the use of the two-way fixed effect model (firm and worker fixed effects) to study the effect on wages when a worker moves between firms. In a related literature that tries to measure rent-sharing elasticities, Card et al. (2016) report that, *"most studies that control for worker heterogeneity find wage-productivity elasticities in the range 0.05-0.15."* And most closely related to our analysis is Song et al. (2015) which finds find that *"between firm inequality accounts for the majority of the total increase in income inequality"* between 1981 and 2013 in the US. We contribute to this literature by bringing innovation into the picture, and by analyzing the relationship between innovation, wage income and occupation across firms.

Finally, we draw on the literature on wage inequality and the organization of the firm (e.g. see Kremer, 1993, Garicano and Rossi-Hansberg, 2006 and Garicano, 2000). We contribute to this literature by linking wage inequality, the organization of the firm, and its degree of innovativeness.

The remaining part of the paper is organized as follows. In Section 2 we present our data and empirical methodology, and we establish our main empirical findings, namely that more innovative firms pay higher wages and that the premium to working in more innovative firms is higher for low skilled workers. In Section 3 we develop a simple model to account for these findings and list a few additional predictions from this model. In Section 4 we test those predictions. Section 5 collects our concluding remarks.

Table 1: Variance decomposition

	Variance		
	Overall	Within-firm	Between-firm
All	0.319	0.156	0.162
Low skill (1+2)	0.136	0.064	0.071
Intermediate skill (3+4)	0.209	0.112	0.170
High skill (5+6)	0.274	0.170	0.103

Notes: This table shows the between-firm and within-firm variance of the logarithm of hourly wage, calculated for each year from 2004 to 2014 and averaged over years. The decomposition of the overall variance is described in Appendix B. The data are matched employee-employer data from the UK; the sample is described in Appendix A, and includes 572,791 Workers in private corporation with at least 400 employees. Construction of skill levels is explained in Appendix A.2.3.

2 The Basic Evidence on Wages and Innovation

In this section we present our main empirical evidence on how average wage and average wage per type of occupation in the firm, depend upon the firms' innovativeness measured by R&D intensity. We start from the following very basic fact: Table 1 shows what is a well document fact in many countries: in the UK over the last decade (2004-2014) the variance in wages *between* firms is at least as important in explaining wage inequality as the variance *within* firms.

Note that the literature has been relatively silent so far on why some firms pay higher wages than others for workers that appear similar. In a competitive labor market we would expect wages for similar workers to be the same across firms; heterogeneity in firm level technology might influence who is hired, but not the wages of any specific worker, since wages are taken as given by the firm. However, wages might deviate from marginal cost in imperfectly competitive markets. From the endogenous growth literature (e.g. see Romer, 1990 and Aghion and Howitt, 1992), where innovation-led growth is motivated by the prospect of rents, it seems that innovation would be a prime candidate, and recent papers show the effect of innovation on income inequality (e.g. Aghion et al., 2015 and Akcigit et al., 2017).

2.1 The data

Hence our focus on the relationship between wages and innovativeness across firms. We document the correlation between R&D expenditure and wages using novel matched employer-employee data that also contains information on R&D expenditure for the period 2004 to 2014. The employee data come from Annual Survey of Hours and Earnings (ASHE), which is a random sample of 1% of the UK working population, matched to the Business Expenditure on Research and Development (BERD) survey. The data are longitudinal, we follow the same workers over time, and is recorded at the establishment level, with information on related establishments in the same firm. We focus on private companies (excluding the public sector, charities, etc) with more than 400 employees. We have information on around 50,000 employees who work in around 6,300 firms, giving us a total of around 580,000 observations. Further details on the data are given in Appendix A.

2.2 More innovative firms pay higher wages

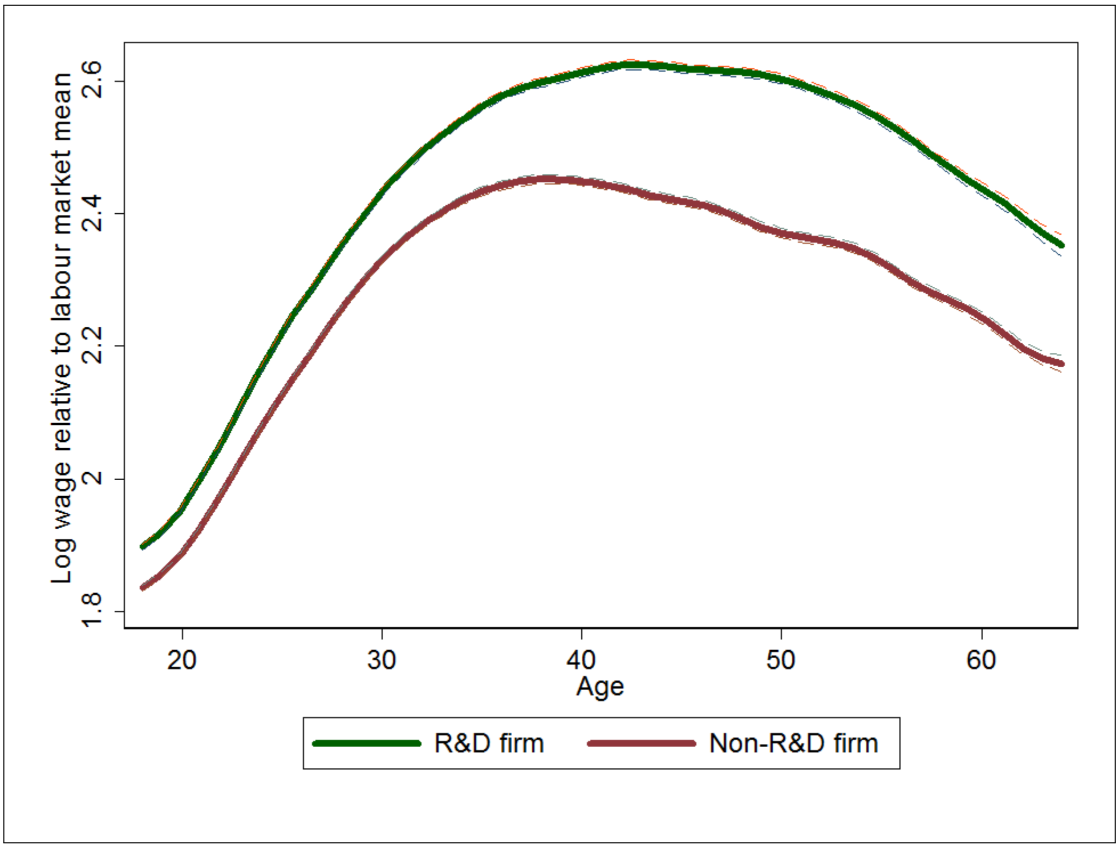
There are significant differences in the wages paid to workers in innovative firms compared to those working in non-innovative firms at all age, even after controlling for differences over time and within geographically separate labor markets (identified by travel to work areas). Figure 1 shows the mean wage of workers in all occupations split by whether the firm that they work for does any R&D or not.

We also see this if we look at the share of workers that work in a firm that does any R&D across the wage distribution. In Figure 2 we see that the share of workers that work in a firm that does any R&D increases from just over 20% for workers at the bottom of the wage distribution, to over 55% after the 80^{th} percentile of the distribution where it plateaus. The share falls right at the top, where workers in the (low innovative) financial sector are heavily represented. This effect holds within innovative firms. The average wage in a firm increases with the firm's R&D intensity,[2] as shown in Figure 3.

Of course, workers in R&D firms might have different characteristics to those working in non-R&D firms. Table 2 shows that they are indeed more likely to be male, work full-time and have longer tenure within the firm. R&D firms also differ from non-R&D firms in that they are larger (have a larger workforce).

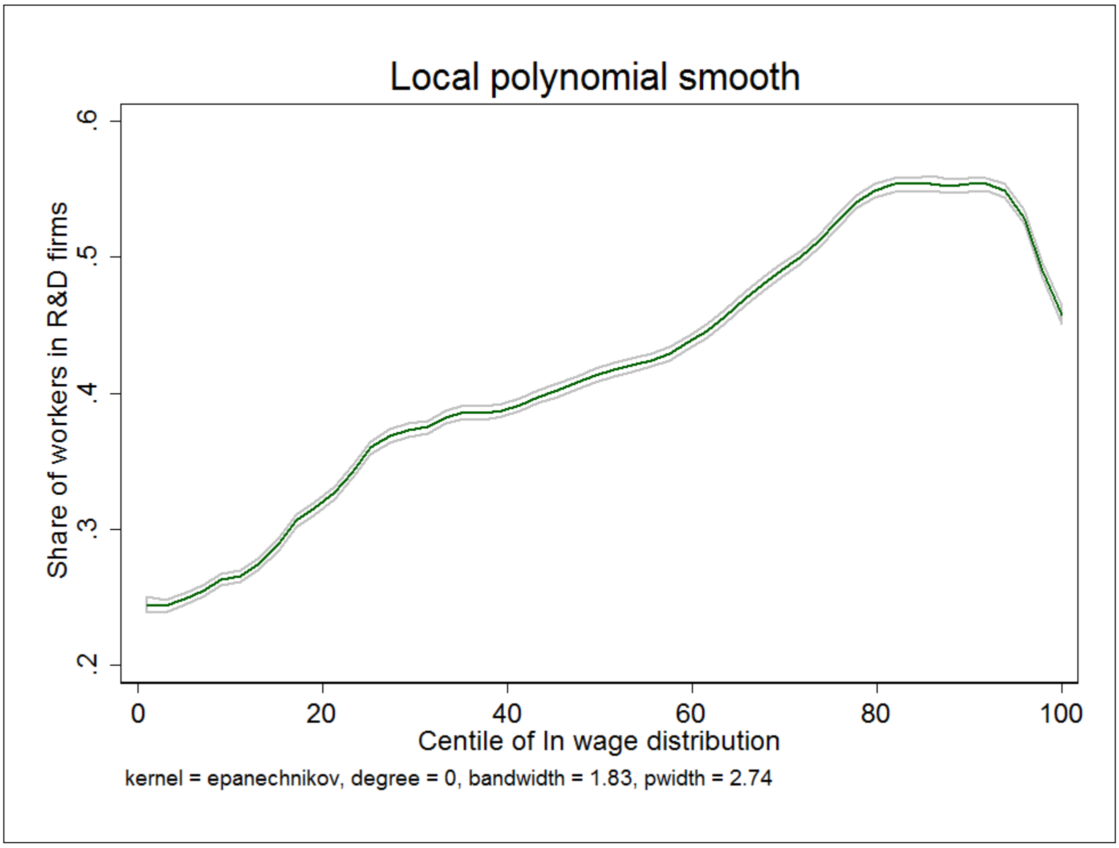
[2] In all the following, we will refer to R&D intensity as the ratio of total R&D expenditures divided by employment (see Appendix A.1).

Figure 1: *Log hourly wage, by age*



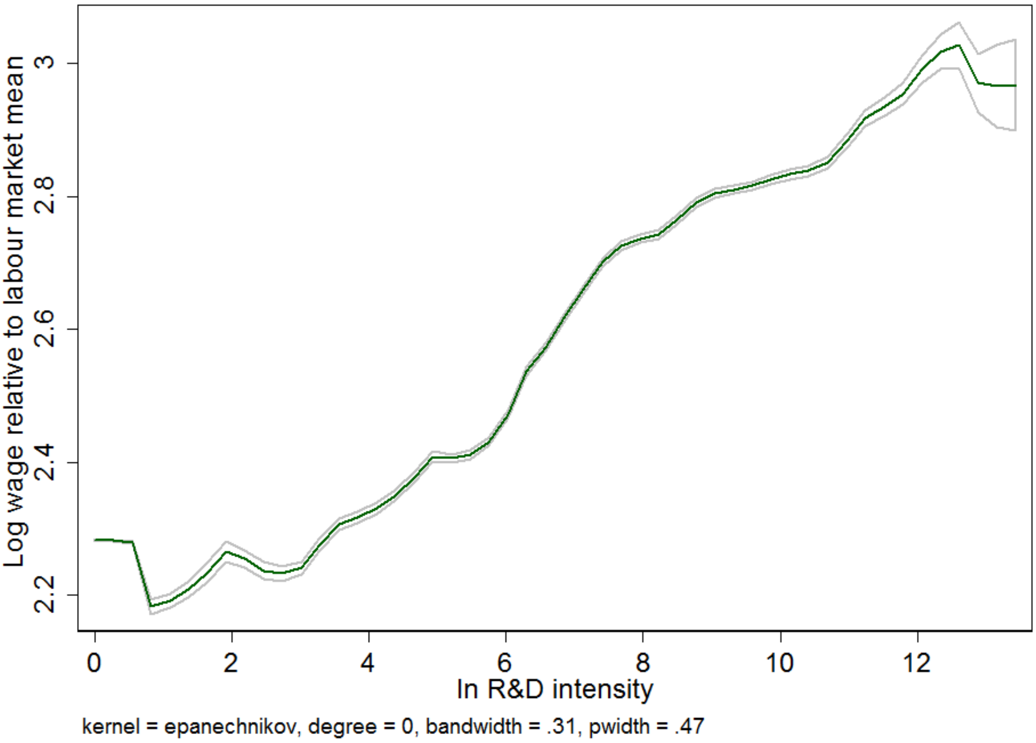
Notes: *This figure plots age dummies from a regression of log hourly wage, controlling for separate year effects for each travel to work area (there are around 240 travel to work areas). The lower curve is for workers in non-innovative firms, the upper curve for workers in innovative firms. Innovative firms are defined as firm that have declared at least one pound in R&D expenditures over the period. 95% confident intervals are included.*

Figure 2: *Share of workers in R&D firms at each percentile of the overall wage distribution*



Notes: *This figure plots the share of workers from innovative firms (defined as firms reporting a positive amount of R&D expenditure since 2000) at each percentile of the overall hourly wage distribution. All observations from our Final Sample from 2004 to 2014 are considered independently.*

Figure 3: *Log hourly wage and R&D intensity*



Notes: *This figure plots the logarithm of total hourly income against the logarithm of total R&D expenditures (intramural + extramural) per employee (R&D intensity). The x-variable is divided into 20 groups of equal size and one larger group of firms with no R&D (x-axis value set to 0). Groups of firms are computed yearly on the sample of private firms of more than 400 employees. See Tables A5 and A6 for more details.*

Table 2: Comparison of R&D and non R&D firms

	Innovative firm		Current R&D firms	
	Yes	No	Yes	No
Employment	2,828	2,221	2,491	2,401
Hourly wage (£)	15.7	12.5	15.9	12.8
Share of male (%)	68	57	71	58
Share of full-time (%)	90	76	92	77
Share of high skilled workers (%)	30	18	31	19
Share of low skilled workers (%)	51	65	50	63
Age	40.5	38.1	41.1	38.3
Tenure	8.9	5.7	9.5	5.9
Firm-years	11,463	23,369	7,684	27,148
Observations	238,994	334,305	144,272	429,027

Notes: Employment is number of workers in the firm averaged over years, hourly wage is measured by total weekly earning divided by total paid hours (including overtime), high skilled workers include categories 5 and 6 (see Appendix A.2.3), low skilled include categories 1 and 2. Innovative firms are firms that report at least one pound of total R&D expenditure over the period, current R&D firms are those that report a positive amount of R&D expenditure in that period. A Student's test on the equality of each coefficient of column 1 (resp. 3) and column 2 (resp. 4) always reject the null hypothesis.

To investigate whether these correlations hold up to controlling for other individual and firm characteristics we estimate the following equation:

$$ln(w_{ijkft}) = \mathbf{X}'_{ift}\beta_1 + \mathbf{Z}'_{ft}\beta_2 + \beta_3 ln(1 + R_{ft}) + \epsilon_{ijkft}, \quad (1)$$

where i indexes individual, j occupation, k labor market, f firm and t years. The variable ϵ_{ijkft} includes fixed effects at differing levels depending on the specification (see details in the results tables) plus an idiosyncratic error. A labor market is defined as a travel to work area and there are around 240 such areas in the UK (see Appendix A.3). w_{ijkft} is mean hourly earnings, X_{ift} are individual-firm level variables including age, gender, whether the job is full-time and tenure in the firm, Z_{ft} are firm characteristics including number of employees. R_{ft} is R&D intensity (R&D expenditure divided by number of employees). We use $ln(1 + R_{ft})$ to accommodate values of 0 in firms that do not do any R&D; it is almost always equal to $ln(R_{ft})$ given the magnitude of R&D expenditure, so we can interpret β_3 as the elasticity of wage with respect to R&D intensity. In Appendix D we show robustness of our results to alternative functional forms and alternative measures of R&D. Tables A1 and A8 in the Appendix gives descriptive statistics of the key variables.

We estimate equation (1) using a fixed effect estimator. Card et al. (2014) suggest that, in a similar wage regression on a firm measure of rent, a bias in the estimated coefficient is expected because of small fluctuations in the firm level measurement of rent. They use an instrumental variables estimation. This problem mostly arises through short-term changes in sales and materials that influence the value added per employee which is their measure of rent. Our measure of rent is R&D expenditure which we argue is less likely to be affected by such accounting definitions. In addition, we show in Appendix D.3 that using the number of workers directly involved in R&D activities (a measure even less likely to be influenced by accounting definitions) does not affect our findings.

The estimated coefficients are shown in Table 3. In column (1) we use year-labor market fixed effects, in column (2) year-labor market-occupation fixed effect, in column (3) individual fixed effect and year effects and in column (4) firm fixed effect and year effects. The coefficient on the R&D variable is always positive and significant; it decreases when firm or individual fixed effects are included.

What we see is that the correlations found in Figures 3 are robust to including a number of control variables that are likely to influence variation in income (age, experience, gender...). The positive correlation of R&D and income is also robust to

Table 3: Correlation between income and R&D intensity.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
ln R&D int	0.030***	0.016***	0.006***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.057***	0.034***	0.085***	0.043***
	(0.000)	(0.000)	(0.000)	(0.000)
Age^2	-0.001***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
$Tenure$	0.022***	0.015***	0.008***	0.015***
	(0.000)	(0.000)	(0.000)	(0.000)
$Tenure^2$	-0.00028***	-0.00022***	-0.00009***	-0.00019***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Firm Size	-0.033***	-0.010***	-0.007***	-0.023***
	(0.000)	(0.000)	(0.000)	(0.002)
Full-Time	0.304***	0.120***	-0.001	0.186***
	(0.001)	(0.001)	(0.001)	(0.001)
Fixed Effects	(k,t)	(k,j,t)	i+t	f+t
R^2	0.306	0.062	0.025	0.145

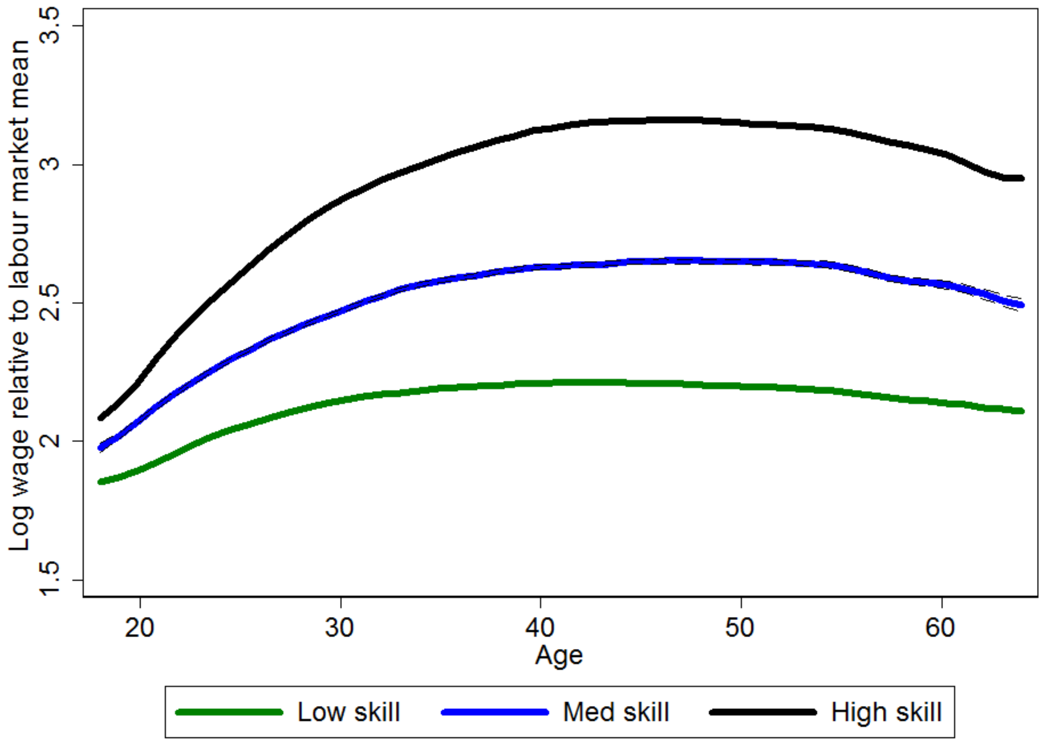
Notes: 572,786 observations. The dependent variable, log of wage, is measured by the gross hourly earning. Variables definitions are given in Table A7. Column 1 includes year-labor market fixed effects, column 2 includes year-labor market-occupation fixed effects, column 3 includes year and individual fixed effects and column 4 includes year firm fixed effects. Heteroskedasticity robust standard errors clustered at the individual level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

including various combination of fixed-effect and its magnitude decreases a lot when moving from column (1) to (4). Note that including additive firm and individual fixed effects do not alter this finding.

2.3 Our main empirical finding

The literature on skill-biased technical change (see for example Goldin and Katz, 2010) suggests that innovation drives inequality by driving up wages at the top end of the distribution. We add to this literature by looking at how the returns to working in a better (higher paying) firm vary between workers with different skill levels. We use a definition of skill based on a match between qualifications and occupations, defined in Appendix A.2.3. We consider three skill groups. Low skilled occupations

Figure 4: *Log hourly wage, by skill and age*



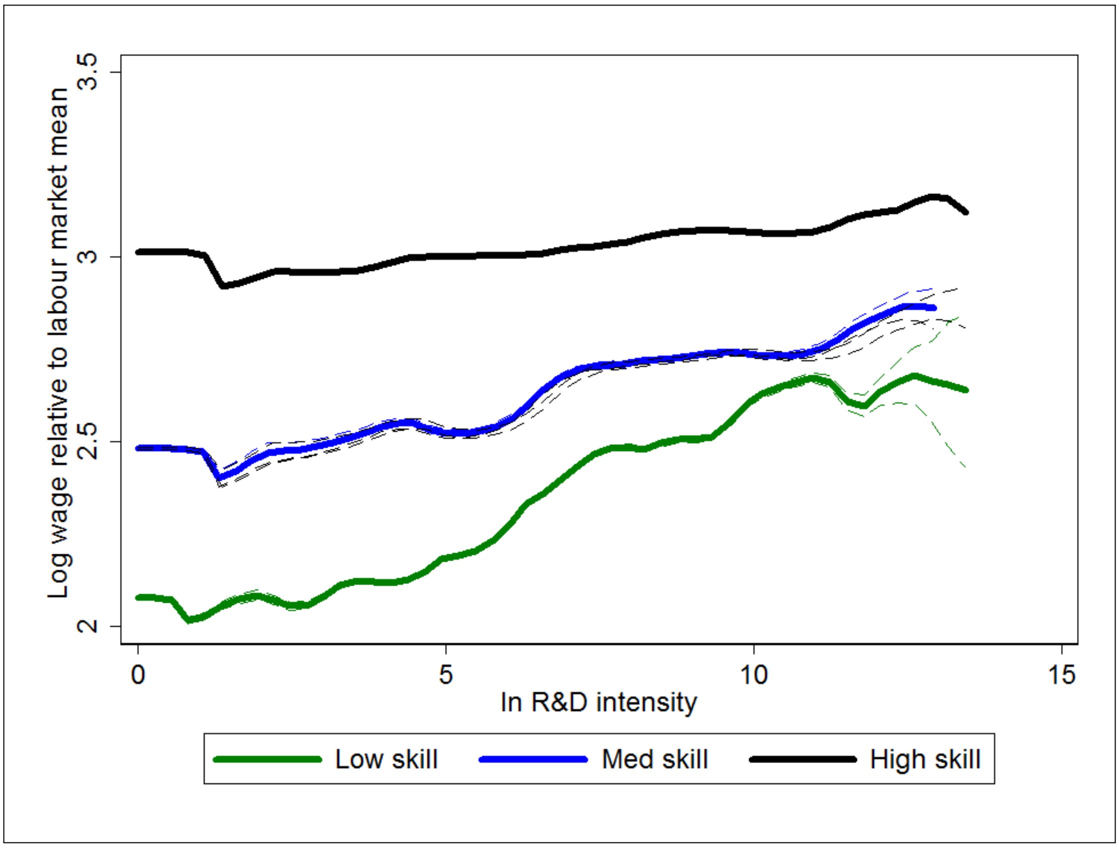
Notes: *This figure plots age dummies from a regression of log hourly wage, controlling for separate year effects for each travel to work area (there are around 240 travel to work areas). The lower curve is for workers in non-innovative firms, the upper curve for workers in innovative firms. Innovative firms are defined as firm that have declared at least one pound in R&D expenditures over the period. 95% confident intervals are included.*

include manufacturing basic occupations, housekeeping, telephone sales. Medium skilled occupations include trades, specialist clerical, associate professionals. High skilled occupations include engineers and managers. Figure 4 shows that throughout their life cycle, workers in high skill occupations receive significantly higher wage than other workers.

Surprisingly, going back to Table 1, when we look by skill category we see that the within-skill group variance of wages across firms is relatively more important for low skill workers than high skill workers.

Our focus in this paper is on the impact of innovativeness on between-firm wage inequality, in other words we want to look at how the returns to working in more innovative versus less innovative firms varies across the skill distribution.

Figure 5: *Average log wage, by skill group*



Notes: *Vertical axis show the average of the logarithm of total hourly income of workers (standardized to have mean 0 across all skill categories when there is no R&D). Horizontal axis the quantile of R&D intensity of the firm, with 20 quantiles and an additional one indicating zero R&D as quantile 0. The bottom curve shows mean wage for low skilled workers, the middle line for intermediate skill and the top line for high skilled workers (see section A.2.3). 95% confident intervals are included.*

Figure 5 replicates Figure 3 but splits workers by skill level. Workers in the highest skill categories (5+6) earn the highest wages, and these wages are on average similar across firms that do more or less (include zero) R&D. In contrast, workers in low skilled occupations earn substantially more if they work in a firm that has higher R&D intensity. The wage gradient with respect to R&D intensity is largest for low-skilled workers.

Highly innovative firms also hire fewer low-skilled workers. Table A9 in the Appendix shows that moving from the first vintile to the last one in terms of R&D intensity increases the share of high skilled workers (categories 5+6) from 13.7% to 63.8%.

Table 4: R&D intensity and hourly earnings at different skill levels.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
Skill category	low (1+2)	intermediate (3+4)	high (5+6)	All
ln R&D int	0.008***	0.002***	0.000	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)
*low skill				0.006***
				(0.000)
*med skill				0.002***
				(0.000)
Age^2	-0.000***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.009***	0.006***	0.000	0.007***
	(0.000)	(0.001)	(0.000)	(0.000)
$Tenure^2$	-0.00012***	-0.00009***	0.00003**	-0.00009***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Firm Size	-0.005***	0.003*	0.005***	-0.006***
	(0.000)	(0.002)	(0.001)	(0.000)
Full-Time	-0.014***	-0.097***	-0.117***	-0.008***
	(0.001)	(0.004)	(0.005)	(0.001)
skllow				-0.147***
				(0.002)
sklmed				-0.065***
				(0.002)
N	371815	95473	105482	572786
R^2	0.124	0.166	0.082	0.042

Notes: Definition of all variables is given in Table A7. Individual and year fixed effects are included in all columns. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

In order to see if this correlation is robust to controlling for differences between workers we estimate our preferred specification with individual fixed effects (column 3 of Table 3) separately for workers of different skill levels. In Table 4 column (1) we show results for low skilled workers (skill categories 1 and 2), in column (2) for intermediate skills (skill categories 3 and 4) and in column (3) for high skills (skill categories 5 and 6). The positive coefficient on R&D only holds for low and intermediate skill categories and is strongest for the former. In column 4 we pool all skill categories and allow the intercept and coefficient on R&D intensity to vary with the skill category. We see that compared to skill level 1, the interacted terms is always negative and is larger in absolute value as we move to higher skill levels.

One concern could be that high skilled workers receive a large part of their wage in the form of lump-sum bonuses at the end of the year and that these bonuses are not well captured by measures of weekly wages. This would particularly be an issue if high skilled workers receive larger bonuses in more R&D intensive firms. In Appendix D.1 we show that using yearly wages instead of weekly wages and including or excluding incentive payments does not affect our results.

The finding that the premium to working in a more innovative firm, is larger for lower skill workers, may look somewhat counter-intuitive and at odds with the literature on skill-biased technical change,[3]. In the next section we show how this finding can be rationalized. More specifically, we propose a model in which a firm's innovativeness is reflected in the degree of complementarity between low skill and high skill workers.

3 A Simple Model

We develop a model where the complementarity between workers in "high-skilled" and "low-skilled" occupations within a firm increases with the firm's degree of innovativeness. Another feature of the model is that the skills of workers in high-skilled occupations are less firm-specific (e.g. those are typically more educated employees, whose market value is largely determined by their education and accumulated reputation), whereas the skills of workers in low-skilled occupations are more firm-specific. Low-skill workers draw bargaining power from the fact that they can shed on their quality potential and underperform, which in turn reduces the firm's output more when low-skill workers are more complementary to high-skill workers.

The model is meant to capture the idea that low-skilled workers can have a potentially more damaging effect on the firm's value if the firm is more technologically advanced. This idea is in line with Garicano and Rossi-Hansberg (2006) where low-skilled employees draw problems and select between the easy questions which they solve themselves and the more difficult questions which they pass on to upper layers of the hierarchy. Presumably, the more innovative the firm, the harder difficult questions are to solve, therefore the more valuable high-skilled employees' time is, and therefore the more important it is to have high-ability low-skilled employees so as to make sure that the high-skilled employees within the firm concentrate on the most difficult tasks. Another interpretation of the higher complementarity between

[3] Similar findings have been derived by Matano and Naticchioni (2017) using Italian data.

low-skilled and high-skilled employees in more innovative firms, is that the potential loss from unreliable low-skilled employees is bigger in such firms: hence the need to select out those low-skilled employees which are not trustworthy.

3.1 Production technology

Suppose that the firm must employ one high-skilled and one low-skilled worker,[4] with the following partial O-Ring production function (Kremer, 1993), where the high skilled worker has quality level (quality potential) Q and the low skilled worker has quality level (quality potential) q:

$$F(Q, q, \lambda) = \theta \left[\lambda Q q + (1 - \lambda)(Q + q) \right],$$

where $\lambda \in (0, 1)$ reflects the extent to which the firm is "innovative" (or "O-Ring" in Kremer, 1993's terminology). We know from Caroli and Van Reenen (2001) and Bloom et al. (2014) that more innovating firms tend to have flatter internal organization, with more strategic complementarity between firm's employees. In this version of the model, the value of λ is assumed to be exogenous and known by the firm. The timing of moves is as follows. First, the firm decides about the qualities potential (q, Q) of the two workers it hires. Then the firm hires the workers and negotiate separately with each of them. We solve the model by backward induction, starting with the wage negotiation and then moving back to the choice of qualities.

3.2 Wage negotiation

The firm engages in separate wage negotiations with each of the two workers. This negotiation will lead to the equilibrium wages $w^L(Q, q, \lambda)$ for the low skilled worker and $w^H(Q, q, \lambda)$ for the high skill worker. In its negotiation with its two workers, the firm takes into account the fact that if the negotiation with the low-skilled worker fails, then the firm must fall back on a substitute low-skilled worker with quality q_L[5]; similarly, if its negotiation with the high skilled worker fails, the firm must look for a substitute high skilled worker of quality Q_L. We assume that:

$$Q > Q_L > q > q_L > 1. \tag{A1}$$

[4] In Appendix C we extend the model to more high-skilled and low-skilled workers.

[5] Or equivalently accept that the current worker underperform at quality level q_L.

We also assume that it is relatively easier for the firm to find a substitute for the high skilled worker than to find a substitute for the low-skilled worker. The rational for this assumption is that the ability of a low-skilled worker is harder to detect ex-ante, e.g. because there is less information the firm acquires ex ante based on the employee's CV (education, reputation). On the other hand, a high-skilled employee can show that she graduated from a leading university (Russell group, Ivy League etc.) or acquired a reputation.[6]

We thus assume that:

$$Q - Q_L < q - q_L. \tag{A2}$$

Substitute low-skilled and high-skilled workers are paid wages w_L and w_H respectively, which we assume to be exogenous. Similarly, the low-skilled and high skilled incumbent workers have outside option $\bar{w}^X$ with $X = H, L$ which are also exogenous. We assume: $w_L < w_H$ and $\bar{w}^L < \bar{w}^H$.

3.2.1 Equilibrium low skill wage

The firm's net surplus from employing the current low-skilled worker, is equal to:

$$S^F = \theta \left[\lambda Q + (1 - \lambda)\right](q - q_L) - w^L(Q, q, \lambda) + w_L,$$

whereas the low-skilled worker's net surplus is equal to

$$S^{LS} = w^L(Q, q, \lambda) - \bar{w}^L.$$

Assuming that the fraction β^L of the firm's net surplus goes to the low-skilled worker, with $\beta^L < 1$, or more formally:

$$S^{LS} = \beta^L S^F,$$

we immediately the following expression for the equilibrium wage of the low-skilled worker:

$$w^L(Q, q, \lambda) = \frac{\theta \beta^L}{(1 + \beta^L)}(q - q_L)(\lambda(Q - 1) + 1) + \frac{w_L \beta^L + \bar{w}^L}{(1 + \beta^L)} \tag{2}$$

[6]Equivalently, the current high-skill worker, if kept by the firm, will not underperform much for reputational reasons.

3.2.2 Equilibrium high skill wage

Replicating the same argument for the high-skilled worker, under the assumption that a fraction β^H of the firm's net surplus accrues to the high-skilled worker, with $1 > \beta^H \geq \beta^L$, we obtain the following expression for the equilibrium wage of the high-skilled worker:

$$w^H(Q, q, \lambda) = \frac{\theta\beta^H}{(1+\beta^H)}(Q - Q_L)(\lambda(q-1)+1) + \frac{w_H\beta^H + \bar{w}^H}{(1+\beta^H)} \tag{3}$$

Since $\beta_H > \beta_L$ and $\frac{w_H\beta^H+\bar{w}^H}{(1+\beta^H)} > \frac{w_L\beta^L+\bar{w}^L}{(1+\beta^L)}$ and since from (A1) and (A2) that $(q - q_L) > (Q - Q_L)$ and $(Q - 1) > (q - 1)$, then we clearly have $w^H(Q, q, \lambda) > w^L(Q, q, \lambda)$ for all $\lambda \in (0, 1)$ and (q, Q) satisfying (A1) and (A2).

3.2.3 How innovativeness affects equilibrium wages

Taking the derivative of equilibrium wages with respect to λ yields:

$$\begin{aligned}\frac{\partial w^H(Q,q,\lambda)}{\partial\lambda} &= \frac{\theta\beta^H}{1+\beta^H}(Q-Q_L)(q-1)\\ \frac{\partial w^L(Q,q,\lambda)}{\partial\lambda} &= \frac{\theta\beta^L}{1+\beta^L}(q-q_L)(Q-1)\end{aligned} \tag{4}$$

Our baseline case is one where there is no difference in bargaining powers between high-skilled and low-skilled workers: this will the case for example if the net surplus from employing each worker, is equally split between that worker and the firm. Then we have: $\beta^L = \beta^H$, which, together with Assumptions (A1) and (A2), immediately implies that:

$$\frac{\partial w^L(Q,q,\lambda)}{\partial\lambda} > \frac{\partial w^H(Q,q,\lambda)}{\partial\lambda}.$$

In other words the low-skilled equilibrium wage increases more with λ (and thus with innovativeness) than the equilibrium wage of the high skill worker.

More generally, when $\beta^H \geq \beta^L$, this above result will hold whenever the following condition (C1) is satisfied:

$$\frac{\beta^H(1+\beta^L)}{\beta^L(1+\beta^H)} < \frac{(q-q_L)(Q-1)}{(Q-Q_L)(q-1)} \tag{C1}$$

This condition is in turn automatically satisfied when Q is sufficiently large and/or when Q_L is sufficiently close to Q, i.e. when high-skilled workers are sufficiently easy to replace with a substitute high-skilled worker.

Optimal choice of q

Having determined the equilibrium wages $w^H(Q,q,\lambda)$ and $w^L(Q,q,\lambda)$ for given q, Q and λ, we now move back and look at the firm's choice of qualities (q,Q). We assume that the firm can choose any value of q and Q at no cost. The firm choice will maximize the firm's ex ante profit:

$$F(Q,q,\lambda) - w^H(Q,q,\lambda) - w^L(Q,q,\lambda),$$

with respect to $q > 1$ and $Q > 1$.

Assuming that $q \in [\underline{q}, \overline{q}]$ and $Q \in [\underline{Q}, \overline{Q}]$, this optimization problem immediately yields the equilibrium quality choice:

$$\begin{aligned} q &= \overline{q}; \\ Q &= \overline{Q}. \end{aligned}$$

More generally, suppose that the firm have needs to train the low-skilled worker to bring her from q_L to q at a convex cost $C(q-q_L) = \frac{1}{2}(q-q_L)^2$, and that training occurs before the wage negotiation. For simplicity, we consider the case where the bargaining surplus is split equally between the firm and each worker ($\beta_H = \beta_L = 1$). Then the firm chooses (q,Q) so as to:

$$(q^*,Q^*) = \underset{q_L<q<\overline{q}\ Q_L<Q<\overline{Q}}{\operatorname{argmax}} \left\{ F(Q,q,\lambda) - w^H(Q,q,\lambda) - w^L(Q,q,\lambda) - \frac{C}{2}(q-q_L)^2 \right\}$$

With respect to Q, the problem remains linear which again leads to the corner solution $Q^* = \overline{Q}$.

With respect to q, the problem is concave so that by first order condition we obtain:

$$q^* = q_L + \frac{\theta}{2C}\left[\lambda(Q_L - 1) + 1\right],$$

where we implicitly assume that this value if lower than $\overline{q}$.

Note that $q^*$is increasing with λ: that is, more training is invested in low-skilled workers in more innovative firms.

Next, we compute the equilibrium wage of low-skilled workers, which up to a constant is equal to:

$$w^L(\lambda) \equiv w^L(Q^*, q^*, \lambda) = \frac{\theta^2}{4C}\left(\lambda(Q_L - 1) + 1\right)\left(\lambda(\overline{Q} - 1) + 1\right),$$

so that:

$$\frac{dw^L(\lambda)}{d\lambda} = \frac{\theta^2}{2C}\left[\left(\overline{Q} - 1\right)\left(Q_L - 1\right)\lambda + \frac{\overline{Q} + Q_L - 2}{2}\right],$$

On the other hand,

$$w^H(\lambda) \equiv w^H(Q^*, q^*, \lambda) = \frac{\theta}{2}\left(\overline{Q} - Q_L\right)\left[\lambda\left(q_L + \frac{\theta}{2C}\left(\lambda(Q_L - 1) + 1\right) - 1\right) + 1\right],$$

so that:

$$\frac{dw^H(\lambda)}{d\lambda} = \frac{\theta}{2}(\overline{Q} - Q_L)\left[(q_L - 1) + \frac{\theta\lambda}{C}(Q_L - 1) + \frac{\theta}{2C}\right],$$

Then the inequality

$$\frac{dw^L(\lambda)}{d\lambda} > \frac{dw^H(\lambda)}{d\lambda}$$

boils down to:

$$2(q^* - q_L)(Q_L - 1) > (\overline{Q} - Q_L)(q_L - 1),$$

which is true from (A1) and (A2).

3.2.4 The effect of product market competition

One can augment the above model by introducing product market competition. One channel whereby competition might interact with the main effect of innovativeness on premium to a low skilled worker, is that a firm having to hire a low-skill worker with quality q_L may be driven out of the market with positive probability by a competing firm. This will obviously increase the bargaining power of a low-skill worker. And it do so to a larger extent than it increases the bargaining power of a high-skill worker when $Q - Q_L << q - q_L$.

Predictions

The main predictions of the model so far can be summarized as follows:

Prediction 1: Low-skilled workers that remain in a firm benefit more from an increase in $R\&D$ of the firm (equivalent to an increase of λ) than high-skilled workers in that firm.

But in addition, the model generates the following predictions:

Prediction 2: Low-skilled workers stay longer in more innovative firms (as more time and money is invested in them to getting them from q_L to q^*);

Prediction 3: The main effect is stronger the lower the quality of potential replacements to a low-skilled worker (i.e. the lower q_L);

Prediction 4: The main effect is stronger in more competitive sectors if the quality of potential replacements to a low skilled worker is sufficient low;

3.3 Outsourcing

The model also speaks to the relationship between innovativeness and outsourcing: namely, more innovative firms tend to outsource low skill tasks more than less innovative firms. To see how we can generate this additional prediction, consider the following extension of the above model. There are now two low skill workers with quality q_1 and q_2. Each firms has two kinds of tasks, one that require complementarity between the low and high skill workers and one that does not. Like previously, there is one high skill workers of skill Q. The production function is now:

$$F(q_1, q_2, Q, \lambda) = \theta \left[\lambda q_1 Q + (1-\lambda)(q_2 + Q)\right].$$

As before, workers engage in separate wage negotiations. But this time, if negotiations failed the task is outsourced to a worker with skill level q_L.

We also add to the model by assuming that the firm has to train each low skill worker to get more quality from her. However, the firm faces an overload constraint in its overall training activities. We model this overload constraint by assuming the following quadratic training costs function:

$$C(q_1, q_2) = C((q_1 - q_L) + (q_2 - q_L))^2.$$

Following wage negotiation, we have:

$$w^{L1} = \frac{\lambda Q\theta}{2}(q_1 - q_L) + w_L \quad (5)$$

$$w^{L2} = \frac{(1-\lambda)\theta}{2}(q_2 - q_L) + w_L \quad (6)$$

$$w^{H} = \frac{\theta}{2}\left[\lambda q_1 + (1-\lambda)\right](Q - Q_L) + w_H \quad (7)$$

The firm's ex ante payoff function thus becomes:

$$\frac{\theta}{2}\left[\lambda q_1 Q_L + (1-\lambda)(q_2+Q_L)\right] - 2w_L - w_H + \frac{q_L\theta}{2}\left[\lambda Q + (1-\lambda)\right] - C(q_1,q_2).$$

Maximizing this payoff function with respect to (q_1, q_2) yields a corner solution with $q_2 = q_L$ whenever:

$$\lambda Q_L > (1-\lambda),$$

or:

$$\lambda > \frac{1}{Q_L+1},$$

in which case q_1 is optimally chosen at

$$q_1 = q_L + \frac{\lambda Q_L}{4C}.$$

We interpret this as a firm's decision to outsource task 2 so as to focus on training the worker assigned to task 1.

Prediction 5: A highly innovative (high-λ) firm will prefer to outsource the task that involves less complementarity between low and high skill workers.

4 Further empirical evidence

In this section we confront specific predictions of our model to the data.

4.1 Characteristics of occupations

Our model links the fact that low skill workers employed in more frontier (or higher R&D) firms get a higher wage premium than high skill workers, to the idea that low skilled workers become indispensable to the firm's success - i.e. more complementary with the firm's assets and higher skilled workers.

Measuring complementarity at the individual level is not straightforward. To show support for this prediction, we match in the O*NET data. The O*NET data provides detailed information on the characteristics of occupations in the US, which we assume are still relevant for the UK (more detailed are given in Appendix A.6).

We summarize the responses to four questions which provide evidence to the effect that low skilled workers are more complementary to other workers in high R&D firms than in low R&D firms.

Table 5: *Consequence of an error*

	Tercile of R&D intensity			
Skill level	None (1)	Low (2)	Middle (3)	High (4)
Low	1.00	1.02	1.12	1.14
Intermediate	1.00	1.00	1.02	1.03
High	1.00	1.02	1.00	0.99

Notes: R&D firms are split in three groups of equal size based on the value of their R&D expenditure per employee. Data are taken from O*NET and report the average of the score for the question "What are the consequences of you making an error?" across our final sample standardized to be equal to one for non R&D firms at each skill level.

1. What are the consequences of your making an error (1 = no consequences; 2, 3, 4, 5 = very large consequences)

2. What is the impact of decisions you make (1 = no impact; 2, 3, 4, 5 = very large impact)

3. On-site or in-plant training (none, up to 6 months, between 6 months and a year, a year or more)

4. On-the-job training (none, up to 6 months, between 6 months and a year, a year or more)

Consequence of error

Workers are asked to estimate the consequences of their making an error. They provide a grade between 1 (no consequence) and 5 (very large consequence) as spelled out above. In Table 5, we provide, for each skill level, the average values of the response in our sample across firms with three levels of R&D compared to firms with no R&D. The consequences of a worker in a low-skilled occupation making an error are larger in a higher-R&D firm than in a lower-R&D firm.

Impact of decision

Similarly, workers are asked to evaluate the impact of the decision they make. They provide grades reflecting their estimated impact, as specified above. We report the

Table 6: *Impact of decision*

	Tercile of R&D intensity			
Skill level	None (1)	Low (2)	Middle (3)	High (4)
Low	1	1.00	1.00	1.01
Intermediate	1	0.99	0.98	0.98
High	1	1.00	0.98	0.97

Notes: R&D firms are split in three groups of equal size based on the value of their R&D expenditure per employee. Data are taken from O*NET and report the average of the score for the question "What is the impact of decisions that you make?" across our final sample standardized to be equal to one for non R&D firms at each skill level.

average values of the response across firms with three levels of R&D compared to firms with no R&D. The results are shown in Table 6. In particular, we see that the impact of decisions of a worker in a low-skilled occupation, is larger in a high-R&D firm than in a low-R&D firm. The difference is small, but yet it is statistically significant.

Training

The last two questions spelled out above consider the duration of training low-skill workers receive across firms with different levels of R&D. Table 7 shows that in the highest R&D intensive firms, from 14.3% to 16.2% of low skill workers report having received training for more than one year, whereas only 6.4% to 7.2% of low skill workers report having received training for more than one year in no-R&D firms.

All these results are in line with the assumptions of our model, namely that: (i) low-skill workers are dedicated to tasks that involve more complementarity with other tasks in more R&D intensive firms (in other words, we vindicate the link between λ and the firm's innovativeness); (ii) low-skill workers in more R&D-intensive firms have a higher need to develop firm-specific skills than they do in less R&D intensive firms and therefore they are in higher need to be trained (this is captured by the difference $q - q_L$ which increases with λ in our model).

Table 7: *On the job and on-site training*

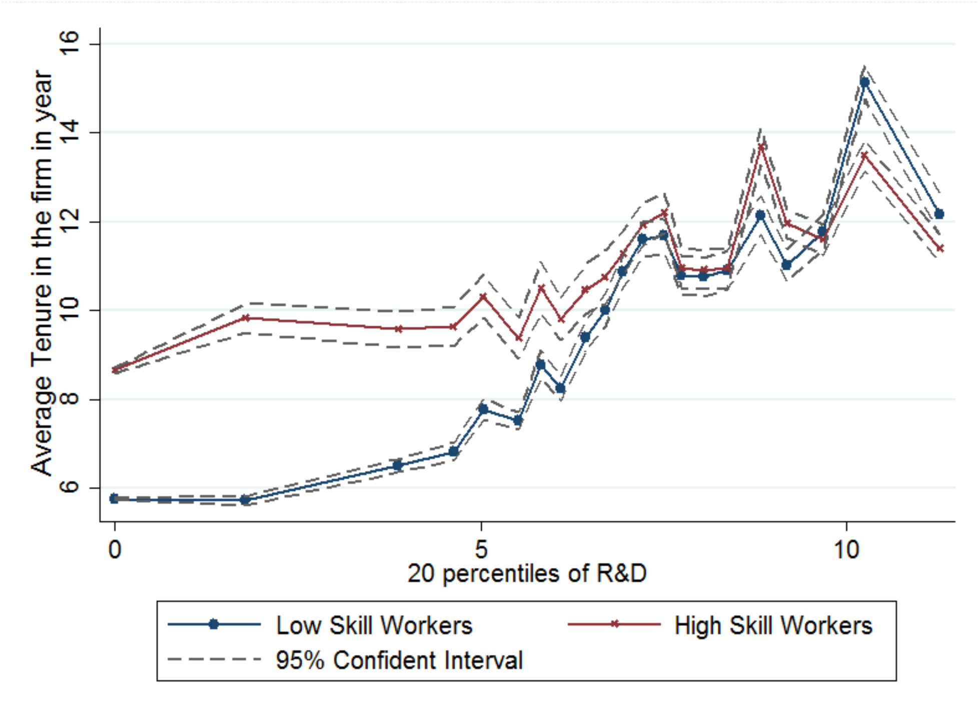
	Tercile of R&D intensity			
	None (1)	Low (2)	Middle (3)	High (4)
On-site or in-plant				
none	20.3	19.7	18.6	18.5
up to 6 months	65.6	64.3	59.6	54.4
6 months - 1 year	7.7	8.4	10.9	12.9
a year or more	6.4	7.6	10.9	14.3
On-the-job				
none	10.1	10.0	9.3	9.1
up to 6 months	74.8	72.5	66.1	59.9
6 months - 1 year	7.9	9.0	12.5	14.9
a year or more	7.2	8.5	12.1	16.2

Notes: R&D firms are split in three groups of equal size based on the value of their R&D expenditure per employee. Data are taken from O*NET and report the share of low-skill workers reporting having been trained for different durations.

4.2 Tenure of low-skilled workers

Our model predicts that lower skilled workers require more firm-specific training than higher-skilled workers, particularly in more innovative firms.[7] Hence our Prediction 2 that low-skilled workers should stay longer in more innovative firms. On the other hand, there should be a smaller effect for innovativeness on high skilled workers turnover. This is indeed what we see from Figure 6.

Figure 6: *Average tenure for workers in low skill and high skilled occupations by quantile of R&D*



***Notes**: Vertical axis show the average of the number of year spent in the firm. Horizontal axis the quantile of R&D intensity of the firm, with 20 quantiles and an additional one indicating zero R&D as quantile 0. The bottom curve shows mean tenure for low skilled workers and the top line for high skilled workers (see section A.2.3). 95% confident intervals are included.*

[7] Note that in our model, low-skill workers in innovative firms will share some rents from firm-specific human capital investments in training. They draw bargaining power from the fact that they can shed on their quality potential and under perform, which in turn reduces the firm's output more when low-skill workers are more complementary to high-skill workers.

5 Summary and conclusion

In this paper we used matched employee-employer data from the UK that we augment with information on R&D expenditures, to analyze the relationship between innovation and between-firm inequality. Our first finding is that more R&D intensive firms pay higher wages on average. Our second finding is that workers in low-skilled occupations benefit more from working in more R&D intensive firms than workers in high-skilled occupations. To account for these findings, we developed a simple model of the firm where the complementarity between employees in "high-skilled occupation" and "low-skilled occupation" within the firm increases with the firm's degree of innovativeness. An additional prediction of the model, which we also confronted to the data, is that workers in low-skilled occupations stay longer in more innovative firms.

Our analysis can be extended in several directions. One would be to look at whether, as our model predicts, the (low-skilled) occupations that yield more return to innovativeness (i.e. for which wage increases more with innovativeness) are more "relational". A second idea is to explore further whether more innovative firms provide more training to workers in low-skilled occupations. Third, our model predicts that our main effect (namely that workers in low-skilled occupations benefit more from working in a more innovative firm) is stronger in more competitive sectors or in areas where potential replacements for incumbent workers in low-skilled occupations are of lower quality: these predictions can be tested using our data. Fourth, we used R&D investment as our measure of innovativeness, and one could use other measures such as patenting. Finally, one may want to look at subgroups of agents within the high- and low-skilled occupation categories. In particular we should look at whether the premium to working in a more innovative firm, is not larger at the very top end of the occupation distribution. One first place to look at, are CEOs, taking into account their total revenues (wage income plus capital income). These and other extensions of the analysis in this paper await further research.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, March 1999, *67* (2), 251–334.

Acemoglu, Daron, "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, mar 2002, *40* (1), 7–72.

_ **and David H. Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in "Handbook of Labor Economics" number 4. In 'Orley Ashenfelter and David E. Card (eds.),.', Amsterdam: Elsevier, 2011, pp. 1043–1171.

Aghion, Philippe and Peter Howitt, "A Model of Growth through Creative Destruction," *Econometrica*, March 1992, *60* (2), 323–351.

_ **, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous**, "Innovation and Top Income Inequality," Technical Report jun 2015.

Akcigit, Ufuk, John Grigsby, and Tom Nicholas, "Immigration and the Rise of American Ingenuity," Technical Report feb 2017.

Atkinson, A B, "Bringing Income Distribution in from the Cold," *Economic Journal*, March 1997, *107* (441), 297–321.

Barth, Erling, Alex Bryson, James Davis, and Richard Freeman, "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S," IZA Discussion Papers 8437, Institute for the Study of Labor (IZA) 2014.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "The Lifecycle of Inventors," 2016.

Bell, Brian D. and John Van Reenen, "Extreme Wage Inequality: Pay at the Very Top," *American Economic Review*, May 2013, *103* (3), 153–57.

_ **and** _ , "Bankers and Their Bonuses," *Economic Journal*, 02 2014, *124* (574), F1–F21.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen, "The Distinct Effects of Information Technology and Communication Technology on Firm Organization," *Management Science*, December 2014, *60* (12), 2859–2885.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin, "Wage bargaining with on-the-job search: Theory and evidence," *Econometrica*, 2006, *74* (2), 323–364.

Card, David, Ana Rute Cardoso, Jorg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," NBER Working Papers 22850, National Bureau of Economic Research, Inc November 2016.

__ **, Francesco Devicienti, and Agata Maida**, "Rent-sharing, Holdup, and Wages: Evidence from Matched Panel Data," *Review of Economic Studies*, 2014, *81* (1), 84–111.

Caroli, Eve and John Van Reenen, "Skill-Biased Organizational Change? Evidence from A Panel of British and French Establishments," *The Quarterly Journal of Economics*, 2001, *116* (4), 1449–1492.

Deaton, Angus, *The Great Escape: Health, Wealth, and the Origins of Inequality*, Princeton University Press, 2013.

Dunne, T, L. Foster, H. Haltiwanger, and K.R. Troske, "Wage and Productivity Disperion in United States Manufacturing: The Role of Computer Investment," *Journal of Labor Economics*, 2004, *22* (2), 397–429.

Elias, Peter and Kate Purcell, "SOC (HE): A classification of occupations for studying the graduate labour market," Research Paper 6, Research Graduate Careers Seven Years On 2004.

__ **and** __ , "Classifying graduate occupation for the knowledge society," FutureTrack Working Paper 5 2013.

Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, October 2000, *108* (5), 874–904.

__ **and Esteban Rossi-Hansberg**, "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 2006, *121* (4), 1383–1435.

Goldin, Claudia and Lawrence F. Katz, *The Race between Education and Technology*, Belknap Press, 2010.

Kremer, Michael, "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 1993, *108* (3), 551–575.

Matano, Alessia and Paolo Naticchioni, "The Extent of Rent Sharing along the Wage Distribution," IZA Discussion Papers 10570, Institute for the Study of Labor (IZA) February 2017.

Office for National Statistics, "Annual Survey of Hours and Earnings, 1997-2015: Secure Access. 8th Edition," Technical Report SN:6689, UK Data Service 2016.

__ , "Business Expenditure on Research and Development, 1994-2014: Secure Access. 5th Edition," Technical Report SN:6690, UK Data Service 2016.

Piketty, Thomas, *Capital in the twenty-first century*, Belknap Press of Harvard Univ. Press, 2014.

Romer, Paul M, "Endogenous Technological Change," *Journal of Political Economy*, October 1990, *98* (5), 71–102.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter, "Firming Up Inequality," 2015.

Stole, Lars and Jeffrey Zwiebel, "Intra-firm Bargaining under Non-binding Contracts," *Review of Economic Studies*, 1996, *63* (3), 375–410.

Van Reenen, John, "The Creation and Capture of Rents: Wages and Innovation in a Panel of U. K. Companies," *The Quarterly Journal of Economics*, 1996, *111* (1), 195–226.

A Data construction and additional description

This appendix describes the construction of our main sample which results from the merge of two datasets provided by the ONS: the Annual Survey of Hours and Earnings (ASHE) and the Business Expenditures on Research and Development (BERD).

A.1 Business Expenditures on Research and Development

The Business Expenditures on Research and Development (BERD, Office for National Statistics, 2016b) is an annual survey conducted by the Office of National Statistics (ONS) that collects information on R&D activities of businesses in the United Kingdom. It is a stratified random sample from the population of firms that conduct R&D. The selected firms then receive a form asking them to detail their innovative activities in accordance to the OECD's Frascati Manual guidelines. The stratification scheme has changed over time, but includes a census of firms with over 400 employees. These are the firms we are interested in. The BERD data is available from 1994-2014 with a coverage that is consistent since 2000.

BERD records expenditure at the level of the firm, the product that the R&D is related to, and the establishment carrying out the R&D. We also know whether R&D was carried out in house (intramural) or outsourced (extramural). Product is recorded at the level of 33 categories. We know the split between civil and defense. More than 99% of the sampled firms report R&D for only one product, representing 75% of total intramural expenditures and 69% of extramural expenditures. 88.2% of intramural R&D expenditure and 96.5% of extramural R&D is civilian; 10% of firms that report doing some R&D do at least some defense R&D. Total R&D expenditures are the sum of intramural and extramural R&D at the firm level. In the paper, we refer to the level of R&D "R&D expenditures" and the level of R&D divided by the number of employees in the firm as "R&D intensity". Including extramural R&D is important as many large firms outsource a large part of their R&D activities, see Figure A1, and this varies across industries.

Table A1 reports the average amount of intramural and extramural R&D across 20 quantiles of the distribution of total R&D intensity.[8] The distributions of both intra and extramural R&D are highly skewed, firms in the highest vintile are very different from others.

[8]Quantiles of R&D are computed each year, so firms can move between quantiles.

Figure A1: *Share of total R&D expenditures that is outsourced (extramural) for 20 quantiles of total R&D intensity. Source: BERD.*

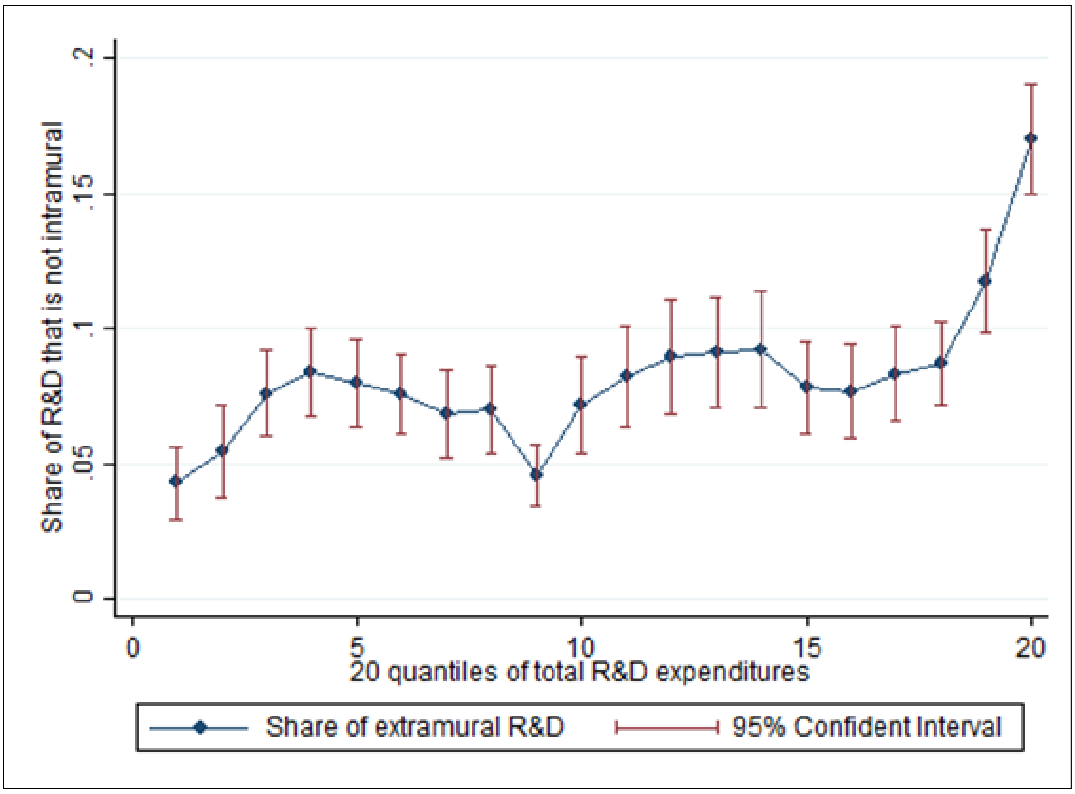


Table A1: Distribution of employment and R&D

Quantile of R&D	Employment	Intramural R&D	Extramural R&D	Number of firms
0 (no R&D)	2,401	0	0	27,183
1	8,172	71	5	390
2	4,480	215	14	384
3	2,932	282	23	383
4	2,521	338	59	387
5	2,829	638	73	383
6	1,643	512	55	384
7	1,963	814	72	386
8	1,749	1,015	98	384
9	1,349	1,008	110	384
10	1,727	1,609	218	381
11	1,629	2,012	231	387
12	1,888	3,136	387	383
13	1,523	3,249	335	385
14	1,455	4,328	387	386
15	1,629	6,749	435	382
16	2,471	16,163	840	386
17	2,668	24,990	1489	386
18	2,314	35,573	2903	383
19	2,513	62,948	9973	384
20	2,290	140,127	70213	380

Notes: This table presents the average number of employees, average expenditures in intramural R&D (in thousand pounds) and average expenditures in extramural R&D (in thousand pounds) for 20 quantiles of R&D intensity (defined as the sum of intramural and extramural R&D expenditures per employee). The first categories "0 (no R&D)" corresponds to firm that do not report R&D in the current year. Quantiles of R&D are computed each year on the sample of firms that have been matched to ASHE and that contains more than 400 employees (see subsection A.4.

A.2 Annual Survey on Hours and Earnings (ASHE)

The Annual Survey of Hours and Earning (ASHE, Office for National Statistics, 2016a) is a 1% random sample of the UK workforce based on the last two digits of the national insurance numbers. We use data from 2004 to 2014.[9] The level of observation in ASHE is the individual job, however, over 98% of individuals have only one job at any point in time, so appear only once per year in the dataset. We have a total of over 1,850,000 observations on around 340,000 individuals working in around 158,000 enterprises.[10]

A.2.1 Cleaning

We clean the data and remove observations: with a missing individual identifier (variable *piden*), with a missing firm identifier (variable $entref$) or those not coded with an adult rate marker (variable *adr*), which mostly involves removing trainees from the sample. We keep only the main job for each individual. This cleaning removes 4.2% of observations. The version of ASHE we use is a panel where individuals are uniquely identified by their (transformed) national insurance number. However, a problem occurs with temporary national insurance number that are reused for different people. We drop all individuals that change gender or change birth dates: 1.2% of observations are affected and dropped. We delete individuals where the data are clearly miscoded, e.g. their age that is less than their tenure in the firm, and we drop individuals aged less than 18 or older than 64 (around 2% of total observations). The outcome of this cleaning is a database of more than 1,650,000 observations on around 320,000 individuals working in 140,000 enterprises. We call this database "Clean ASHE".

A.2.2 Individual income

There are various measures of income in ASHE. Gross weekly wage is collected during the survey period (from one to four weeks) in April of each year. This is reported by the employer and is considered to be very accurate. The gross weekly wage can be broken down into basic pay, incentive pay, overtime pay, additional premium payment for shifts that are not considered overtime and additional pay for other reasons. The gross weekly wage does not include any capital income such as stock-options (reported

[9] There is a discontinuity in ASHE in 2004.

[10] An enterprise can be a private corporation, public company, government agency, non profit organisation, etc.

"incentive pay" includes profit sharing, productivity, performance and other bonus or incentive pay, piecework and commission.). The number of hours worked are reported, split between overtime and basic paid hours. ASHE also provides data on gross annual earnings, as well as the share of this earning that is an incentive payment.

We define hourly income as the ratio of gross weekly wage divided by total number of paid hours (including overtimes). This is the measure of income we will consider as a baseline but we also show descriptive statistics for gross annual earnings. Including other types of income and incentive payments is arguably relevant especially in the case of very high incomes as shown by Bell and Van Reenen (2013, 2014). We study the sensitivity of our results to including or excluding additional types of income in the basic pay in section D.1.[11]

A.2.3 Skills classification

We use a classification based on a match between the National Qualification Framework (NQF) and the Standard Occupation Code (SOC).[12] The NQF defines 8 levels of skill based on the required qualification from PhD (level 8) to Entry level (less than GCSE grade D-G). The current UK immigration rules use 6 groups (see Table A2).[13]

Note that there is another possible classification of skills, based on the standard occupational classification (SOC). Skills here are defined as "the length of time deemed necessary for a person to become fully competent in the performance of the tasks associated with a job". Level 4 corresponds to the highest skill level and includes Corporate Managers, Science and technology professionals, Health professionals, Teaching and research professionals and Business and public service professionals. Level 1 corresponds to the lowest skill level and includes elementary trades, plant and storage related occupations and elementary administration and service occupations.

This classification relies on the first two digits of the 4-digit SOC code. Its main advantage is that it is very straightforward to implement and it is consistent in time. Indeed, although the SOC changed its classification in 2000 and 2010, the first two digits remain unchanged. However, one disadvantage is that relying on the first two

[11] The share of incentive pay increases strongly with earnings, while the share of overtime pay is stable around 5% for the first three quarters of the income distribution, and decreases with wage for the remaining top quarter.

[12] See for example the "code of practice for skilled work, Immigration Rule Appendix J".

[13] A few specific occupations, which we don't use in our analysis, are unclassified: clergy, military, elected officers, sports players and coaches and prison service officers.

Table A2: Skill classification

Skill category	Description
Low skill	
Skill cat 1	Lowest skill occupations, includes many manufacturing basic occupations, child-care related education, housekeeping, telephone salespersons
Skill cat 2	corresponds to a NQF below 3 but not considered as an entry level. Occupations such as pharmaceutical dispensers, greenkeepers, aircraft maintenance technician
Intermediate skill	
Skill cat 3	Requires a NQF of 3 which corresponds to a Level of Advanced GCE (A-level). This category spans many different occupations from Fitness instructors to Legal associate professionals.
Skill cat 4	Requires a NQF of 4 and above which corresponds to a Certificate of Higher Education. It includes many technical occupations like Medical technicians or IT operations technicians and some managerial occupations.
High skills	
Skill cat 5	Includes most managerial and executive occupations as well as engineers. These occupations require at least a NQF of 6 which corresponds to a Bachelors degree or a Graduate Certificate.
Skill cat 6	Corresponds to occupational skilled to PhD-level and include most scientific occupations like Chemical scientists, Biological scientists, Research and development manager but also Higher education teaching professionals.

Notes: This table describe the education requirement for each of our six skill categories. These requirements have been taken from the "code of practice for skilled work, Immigration Rule Appendix J".

Table A3: Demographics by skill level

	Obs.	Hours	% Work full-Time	% Male	Age	Tenure
Low skill						
Skill cat 1	338,102	30.2	60	49	37.3	6.2
Skill cat 2	35,959	35.5	83	68	39.2	8.2
Intermediate skill						
Skill cat 3	71,231	36	88	60	39.1	9.3
Skill cat 4	24,740	36.4	93	60	39.5	9
High skill						
Skill cat 5	102,539	36.4	95	70	40.7	9.8
Skill cat 6	3,284	35.8	92	62	39.3	10.4
Total	575,855	32.6	73	56	38.4	?

Notes: Skill categories are based on occupation codes as described in A.2.3.

digit is not accurate enough to distinguish between, for example, a restaurant manager (SOC2010 code 1223) and a natural environment and conservation manager (SOC2010 code 1212). But according to the work of Elias and Purcell (2004), the former group counts 9.5% of people aged 21-35 and holding a first degree or higher whereas the latter counts 72% of them. This analysis uses on the labor Force Survey 2001-2003. In another article, Elias and Purcell (2013), they advocate the use of another classification and consider the restaurant manager group as a “non graduate group’ and the natural environment manager as an “expert group”.

Tables A3 and A4 show that these workers have different labor market participation behaviour and different outcomes in the labor market.

Table A4: Pay by skill categories

Skill	Hourly pay	Weekly pay	% incentive	% overtime	Annual earnings
Low skill					
Skill cat 1	8.58	285.29	2.59	5.66	13,659
Skill cat 2	11.54	444.87	2.23	5.45	21,948
Intermediate skill					
Skill cat 3	13.52	504.32	5.23	3.61	25,840
Skill cat 4	16.83	625.04	5.23	2.19	32,904
High skill					
Skill cat 5	25.45	931.56	7.67	1.46	53,978
Skill cat 6	22.25	804.11	6.24	1.10	43,542
Total	12.82	455.98	?	?	23,900

Notes: Skill categories are based on occupation codes as explained in subsection A.2.3.

A.3 Travel to work areas

A labor market is defined as a travel to work area and there are around 240 such areas in the UK depending on the year.[14] Since 2011, there are exactly 228 travel to work areas (TTWAs) in the UK with 149 in England, 45 in Scotland, 18 in Wales, 10 in Northern Ireland and 6 cross-border. This is a tool widely used by geographers and statisticians although they have no legal status. They are defined as a form of Metropolitan Area and intent to group urban areas and their commuters hinterland. London, Bristol and Manchester are examples of Travel To Work Areas.

A.4 Matching BERD and ASHE

We match the individuals in "Clean ASHE" with the firms they work for in BERD; we restrict attention to private corporations (dropping public corporations, charities, unincorporated firms, etc). We start with all individuals in "Clean ASHE" who work for a firm with 400 or more employees and match them to the population of firms

[14]Definition of travel to work areas change in time. For this reason, we never use a travel to work area continuously in time.

Table A5: Construction of the sample

ASHE	Observations	Individuals	Mean wage	Sd wage
Raw ASHE	1,841,495	341,463	13	43.1
Clean ASHE	1,655,627	323,409	13.3	14.3
Private Corporations	977,236	230,501	12.9	16.3
Final Sample	573,299	148,503	12.8	16.7
BERD	Observations	Firms	% intramural R&D	% extramural R&D
Raw BERD	216,957	48,554	100	100
400+ Employees	8,086	1,782	75.1	84.0
Final Sample	7,703	1,767	66.1	77.9

Notes: This table presents the evolution of the two databases ASHE and BERD across the successive steps conducted to match them. **ASHE**: Raw data corresponds to the standard ASHE database 2004-2014. Clean ASHE corresponds to the database "Cleaned ASHE" as described in subsection A.2.1. Private corporation corresponds to "Clean ASHE" restricted to private corporations and Final corresponds to "Clean ASHE" restricted to private corporations with more than 400 employees. Mean wage is measured as the average total weekly earning. **BERD**: Raw data corresponds to the standard BERD database 2004-2014. 400+ employees corresponds to this database restricted to firm with more than 400 employees and Final corresponds to firms of more than 400 employees that matched the final version of ASHE. % of intramural and extramural R&D are measured with respect to Raw data.

in BERD with 400 or more employees. Our final sample includes around 580,000 observations on around 150,000 individuals working in around 6,300 different firms; there are around 31,000 firm-year combinations. The implication of the matching and exact numbers can be found in Table A5 and the outcome of merging the subsample of firms in BERD over 400 employees and private firms in ASHE over 400 employees is presented in Table A6.

We use information on firms with more than 400 employees. These firms differ from smaller ones in some ways that are shown in Table A5. However, the distribution of wage in this sample is very similar to the one in the total sample, as seen in Figure A2. The geographical coverage of these firms is also very similar.

A.5 Descriptive statistics

Table A7 gives description of the variables used in the regressions throughout the paper while A8 shows statistical moments of the main variables of interest at the individual level. Low skill workers represent the majority of workers in our sample

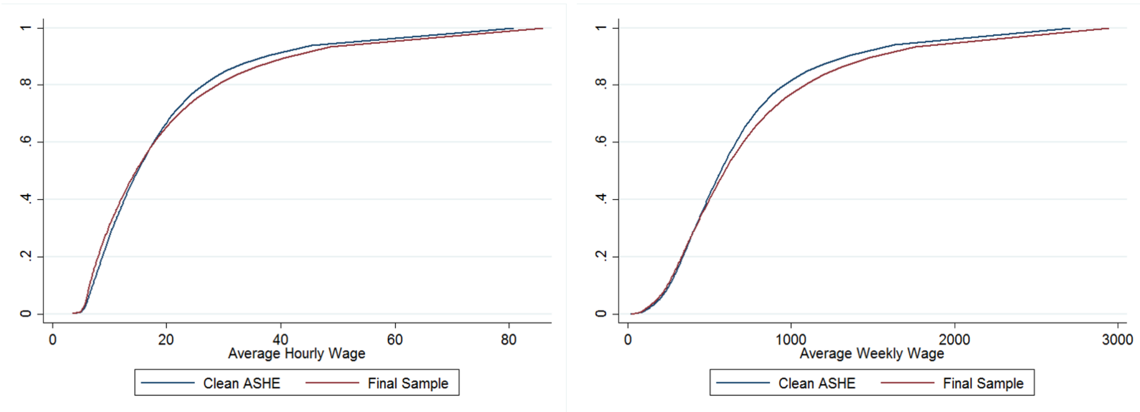
Table A6: Matching results at the firm-year level

Year	in BERD not in ASHE	in ASHE not in BERD	in BERD and ASHE
2004	102	2,406	670
2005	91	2,377	808
2006	91	2,339	956
2007	102	2,372	757
2008	96	2,408	628
2009	75	2,370	798
2010	86	2,322	696
2011	97	2,372	708
2012	97	2,435	781
2013	108	2,488	799
2014	109	2,612	844

Notes: This table presents the number of firms that did not match because they are in BERD but not in ASHE (column 1) or because they are in ASHE but not in BERD (column 2) and the firms that are both in BERD and ASHE (column 3).

Figure A2: *Cumulative distribution function of log wage*

(a) Hourly Wage (b) Weekly Wage



Notes: *This figure plots the empirical cumulative distribution function for two samples: Clean ASHE, corresponding to the 1% random sample of the British population without restriction (other than some cleaning described in Appendix A.2 and Final Sample corresponding to workers of private companies with more than 400 employees.*

Table A7: Variable description

Variable name	Description
Age	Age of the individual at the time of the survey in year
Tenure	Number of year spent in the firm by the individual
Male	Dummy for being a male
Full Time	Dummy for working more than 25 hours a week on average
Age2	Age squared
Tenure2	Tenure squared

Notes: This table presents the description of the main variables used in the regressions.

Table A8: Descriptive statistics of wage variables

Variable	Mean	sd	p10	p25	p50	p75	p90	p99
Total hourly wage (£)	13.5	14.5	6	7.1	10	15.5	24.1	57.6
Weekly wage (£)	493	505	130	254	390	606	911	2,080
Weekly incentive pay (£)	9.3	66.3	*	*	*	*	0.6	220.9
Weekly overtime pay (£)	19	60	*	*	*	*	60.8	280.5
Annual wage (£)	26,024	57,481	4,197	10,937	19,231	30,671	47,000	132,000
Basic paid hours	34.4	10.3	18	34.9	37.5	39.8	42	54.8
Paid overtime hours	1.5	6	*	*	*	*	5.3	20.5
Tenure in years	6.8	7.7	1	1	4	9	17	35
Age	38.9	12	23	29	38	48	56	63

Notes: This table presents some moments (mean, standard deviation and different percentile thresholds) for a set of key variables. Tenure is the number of year an individual has been working in its current firm.

(59%)[15], see Table A3. Workers with higher skill level earn higher wages with the exception of skill category 6 (researchers and professors), where the average is below the average for category 5. We also see from Table A4 that more innovative firms have on average a larger proportion of high skilled workers.

A.6 O*NET data

The O*NET dataset is a database aiming at providing an accurate definition of each occupations in the US at a very detailed level. Information include the type of tasks,

[15]This is a slightly larger proportion than when considering the share of low skilled worker in the whole "clean ASHE" dataset (48%).

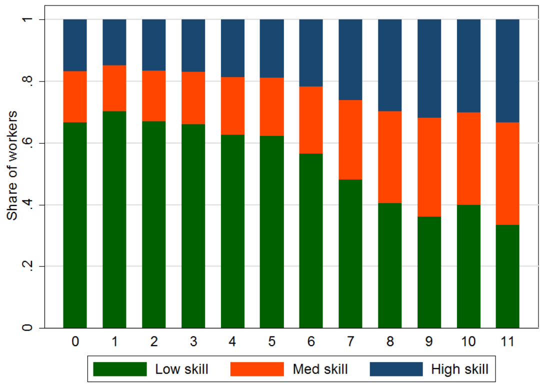
Table A9: Share of workers at each skill category and quantiles of R&D

	Skill category						
	Low		Intermediate		High		
Quantile of R&D	1	2	3	4	5	6	Obs.
0 (no R&D)	63.5	5.6	11.7	3.8	15	0.3	432,029
1	65.8	7.4	10.2	2.8	13.5	0.2	20,654
2	63.2	8.1	10.2	3.2	14.7	0.5	11,962
3	56.0	9.6	11.2	4.2	18.4	0.6	8,271
4	55.7	6.1	14.8	3.6	19.2	0.7	6,884
5	60.9	4.6	14.2	3.3	16.7	0.4	8,382
6	54.0	6.0	15.0	4.2	19.9	0.9	4,855
7	51.9	9.0	12.2	5.0	21.4	0.6	5,895
8	48.6	8.3	14.4	5.2	22.7	0.7	5,012
9	51.4	8.4	11.6	4.5	23.3	0.7	4,037
10	43.5	9.3	12.7	5.1	28.6	0.8	5,176
11	36.3	10.4	15.6	5.8	31.2	0.7	5,265
12	35.8	9.2	15.6	6.2	32.2	1.0	5,993
13	36.0	7.5	15.1	5.7	35.0	0.8	4,583
14	30.2	9.7	12.9	6.7	39.3	1.0	4,415
15	30.8	8.3	18.9	8.7	31.8	1.4	4,816
16	23.2	7.5	19.9	10.4	37.7	1.3	7,453
17	22.1	6.2	21.0	12.3	37.5	0.9	8,600
18	25.1	7.8	18.7	9.3	37.1	2.0	7,245
19	22.9	13.1	15.6	6.2	39.5	2.8	8,468
20	19.2	6.1	14.5	6.6	41.5	12.2	7,007

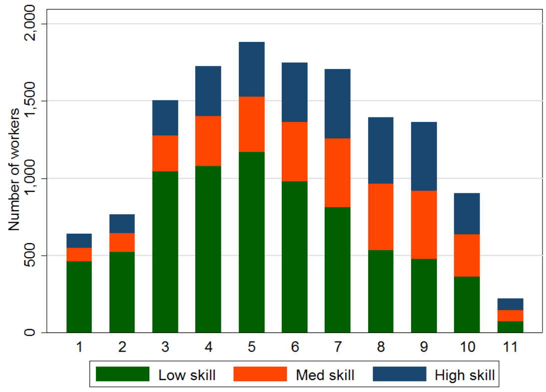
Notes: This table presents the average proportion of each skill group by quantile of R&D intensity. Skill groups are defined in Appendix A.2.3. Quantiles are the same as in Table A1.

Figure A3: *Distribution of workers by skill category and R&D intensity of firm*

(a) Share



(b) Number



the skills and abilities usually required and many characteristics such as, for example, the level of exposition to noise.

The database is freely available from the dedicated website[16] and we use the version 21.1 Database - November 2016 Release.

The information have been gathered either from interviewing workers or from experts descriptions. Although the O*NET data is only based on US workers, we believe that the job descriptions are very similar to those of the UK. To match the different occupation classification we match O*NET data to UK data via isco08.

[16]http://www.onetcenter.org/database.html

B Decomposition of variance

We decompose the variance as presented in Song et al. (2015) among others. More specifically, let $w_{i,f}$ be a measure of the log of income of the individual i (we drop time dependence but in practice, all this is computed for one given year) working in firm f. Let $\bar{w}_f$ be the average wage within this firm and $\bar{w}_A$ be the average value of $w_{i,f}$ across all N observations. We have:

$$[w_{i,f} - \bar{w}_A] = [\bar{w}_f - \bar{w}_A] + [w_{i,f} - \bar{w}_f] \,.$$

We take this equality to square and sum over all N individual. By construction, the covariance quantity is equal to 0 and this yields:

$$\mathrm{Var}(w_{i,f}) = \underbrace{\sum_{f=1}^{F} \frac{N_f}{N} \left[\bar{w}_f - \bar{w}_A\right]^2}_{\text{Within-firm variance}} + \underbrace{\sum_{f=1}^{F} \frac{N_f}{N} \mathrm{Var}(w_{i,f} \mid f)}_{\text{Between-firm variance}}$$

C Extending the model

Extension to more skilled and unskilled workers

We now consider the more general case with $n \geq 1$ low-occupation workers and $m \geq 1$ high-occupation workers. To determine the equilibrium wages resulting from ex post negotiation, we rely on Stole and Zwiebel (1996). In their framework, if the n^{th} low-occupation worker refuses the wage offer w_n^L, then the remaining $n-1$ low-occupation workers renegotiate a wage w_{n-1}^L. By induction, this provides a generic expression for the two equilibrium wages $w_{n,m}^L(Q,q,\lambda)$ and $w_{n,m}^L(Q,q,\lambda)$ (up to a constant in q, Q and λ):

$$\begin{aligned} w_{n,m}^L(Q,q,\lambda) &= \frac{(q-q_L)\lambda\theta}{n(n+1)}\sum_{i=0}^{n} iQ^m q^{i-1} - \frac{\theta(1-\lambda)}{2}(q-q_L) \\ w_{n,m}^H(Q,q,\lambda) &= \frac{(Q-Q_L)\lambda\theta}{m(m+1)}\sum_{i=0}^{m} iq^n Q^{i-1} - \frac{\theta(1-\lambda)}{2}(Q-Q_L), \end{aligned} \tag{C1}$$

when assuming equal bargaining powers for high- and low-occupation workers. Note that this extension nests the baseline version of the model since taking $n=1$ and $m=1$ yields the same results as above.

The case where $n = 1$ and $m = 2$

In this case, we have:

$$\frac{\partial w_{1,2}^L(Q,q,\lambda)}{\partial\lambda} = \frac{\theta(q-q_L)(Q^2-1)}{2} \text{ and } \frac{\partial w_{1,2}^H(Q,q,\lambda)}{\partial\lambda} = \frac{\theta(Q-Q_L)\left(\frac{q(1+2Q)}{3}-1\right)}{2},$$

and we can show that[17] $\frac{q(1+2Q)}{3}-1 < Q^2-1$, which, combined with the assumption that $(Q-Q_L) < (q-q_L)$, immediately implies that:

$$\frac{\partial w_{1,2}^L(Q,q,\lambda)}{\partial\lambda} > \frac{\partial w_{1,2}^H(Q,q,\lambda)}{\partial\lambda}.$$

[17] This is straightforward since $Q > q$ implies that: $q(1+2Q) < Q(1+2Q) < Q(Q+2Q)$ (recall $Q > 1$).

The case where $n = 2$ and $m = 1$

In this case, we have:

$$\frac{\partial w^L_{2,1}(Q,q,\lambda)}{\partial \lambda} = \frac{\theta(q-q_L)(Q+2qQ)}{6} - \frac{q-q_L}{2} \text{ and } \frac{\partial w^H_{2,1}(Q,q,\lambda)}{\partial \lambda} = \frac{\theta(Q-Q_L)(q-1)}{2},$$

Then a sufficient condition for $\frac{\partial w^L_{2,1}(Q,q,\lambda)}{\partial \lambda} > \frac{\partial w^H_{2,1}(Q,q,\lambda)}{\partial \lambda}$ is that $Q + 2qQ > 3q$ which in turn is always true since $Q > q > 1$.

The case where $n = m$

For a given $n \geq 2$, a sufficient condition for $\frac{\partial w^L_{n,n}(Q,q,\lambda)}{\partial \lambda} > \frac{\partial w^H_{n,n}(Q,q,\lambda)}{\partial \lambda}$ is:

$$\frac{1}{n(n+1)} \sum_{i=0}^{n} iQ^n q^{i-1} > \frac{1}{n(n+1)} \sum_{i=0}^{n} iq^n Q^{i-1},$$

which is equivalent to:

$$\sum_{i=0}^{n} \frac{i}{q^{n-i+1}} > \sum_{i=0}^{n} \frac{i}{Q^{n-i+1}},$$

which is automatically true as long as $n \geq 2$.

The case where $n < m$

By induction, for a given $n > 2$, if we assume that $\frac{\partial w^L_{n,m}(Q,q,\lambda)}{\partial \lambda} > \frac{\partial w^H_{n,m}(Q,q,\lambda)}{\partial \lambda}$, then it is easy to show that:

$$\frac{1}{n(n+1)} \sum_{i=0}^{n} iQ^{m+1} q^{i-1} > \frac{1}{(m+1)(m+2)} \sum_{i=0}^{m+1} iq^n Q^{i-1},$$

and therefore that

$$\frac{\partial w^L_{n,m}(Q,q,\lambda)}{\partial \lambda} > \frac{\partial w^H_{n,m+1}(Q,q,\lambda)}{\partial \lambda}.$$

This case is all the more important since we know that most innovative firms have more high-occupation workers than low-occupation workers.

Finally, note that the case $n < m$ corresponds to more R&D intensive firms as we document in the empirical part of the paper.

D Additional specifications

D.1 Different measures of income

In our baseline results, we have chosen to use the average total labor income received per week during the time of the survey divided by the average total number of hours worked. As explained in subsection A.2.2, the numerator includes a fixed salary and additional variable incomes (incentive, overtime and other pay). In this section, we test the sensitivity to our main result to using other measures of income. Results are presented in Table D1 when the usual set of control variables are included and individual and year fixed effects are added. Column 1 uses the baseline measure (logarithm of total earning per hours) as a reference. Column 2 uses the same measure but restricting to fixed salary and excluding overtime. Column 3 uses the total weekly earnings and column 4 and 5 use total annual earnings including (resp. excluding) bonuses. One concern with our results is that high occupation workers receive most of their income from incentive paid at the end of the year and hence not well captured by our baseline measure of income (based on a standard week in April). This could potentially drive our result if in turns, high occupation workers receive a larger share of their income as incentive in innovative firms. In fact, the average share of bonus in yearly income is 8.8% for non R&D firms against 6.5% for non R&D firms. Finally, comparing column 4 and 5 of Table D1 shows no substantial differences when bonus are included or excluded.

D.2 Different functions of R&D

In this section we show that our main results hold using alternative function of R&D. We consider: $\frac{R\&D}{L}$, $ln(1+\frac{R\&D}{L})$, Hyperbolic with R&D, Hyperbolic with $\frac{R\&D}{L}$, $ln(1+R\&D)$, $R\&D>0$ and $R\&D>0$. Results are shown in Table D2.

Next, we allow the coefficient to adjust at different point in the R&D distribution. To do so, we include a binary variable for each of the twenty quantile of R&D:

$$ln(w_{ijkft}) = x'_{ift}\beta_1 + z'_{ft}\beta_2 + \sum_{l=1}^{20}\beta_{3l}R_{ftl} + \nu_w + \epsilon_{it} \tag{D1}$$

Where R_{ftl} is equal to 1 if firm f belongs to quantile l in year t. The resulting estimated coefficients β_{3l} on each of these binary variables are presented in Table D3. We see that the coefficients are positive and increase with the quantile of R&D for

Table D1: Robustness to using different measures of income.

	Dependent variable: $ln(w_{ijkft})$			
Income	Total hourly pay (1)	Fixed hourly pay (2)	Total pay (3)	Fixed pay (4)
$ln(R_{ft}+1)$	0.006***	0.005***	0.008***	0.012***
	(0.000)	(0.000)	(0.000)	(0.001)
Age^2	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.008***	0.007***	0.006***	0.069***
	(0.000)	(0.000)	(0.000)	(0.001)
$Tenure^2$	-0.000***	-0.000***	-0.000***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.007***	-0.010***	-0.016***	-0.026***
	(0.001)	(0.001)	(0.001)	(0.002)
Full-Time	-0.001	0.013***	0.659***	0.495***
	(0.002)	(0.002)	(0.004)	(0.006)
Fixed Effects	i+t	i+t	i+t	i+t
R^2	0.888	0.907	0.888	0.800
N	572,786	572,573	575,859	571,052

Notes: This table presents results from estimating equation 1 using different measures of income. Column 1 uses the logarithm of total hourly earnings, column 2 uses the logarithm of the basic (fixed) hourly income, column 3 uses the logarithm of the total weekly earning and column 4 uses the logarithm of annual gross earnings. Control variables definition and construction are given in Table A7. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table D2: Testing different function of R&D

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
$\frac{R\&D}{L}$	0.00415***	0.00216***	0.000455***	0.000170*
$ln(1+\frac{R\&D}{L})$	0.117***	0.0649***	0.0286***	0.0101***
Hyperbolic with R&D	0.0198***	0.0105***	0.00400***	0.000963***
Hyperbolic with $\frac{R\&D}{L}$	0.0979***	0.0541***	0.0238***	0.00819***
$ln(1+R\&D)$	0.0215***	0.0114***	0.00438***	0.00111***
$R\&D > 0$	0.147***	0.0751***	0.0265***	0.00224
$R\&D$	0.282***	0.120***	0.0531***	0.0154**
Fixed Effects	(k,t)	(k,j,t)	i+t	f+t
Observations	572,799	572,799	572,799	572,799

Notes: This table presents the coefficient on the function of R&D intensity when estimating equation 1 but replacing the log of R&D by alternative functions. The set of control variables and fixed-effects are the same as in Table 3. Each line corresponds to a different functional form. Hyperbolic function is $H(x) = ln(x+\sqrt{x^2+1})$. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

almost all quantiles except for the first ones. The only exception occurs when we use firm fixed effects (column 4) where the coefficients become positive only for the very high quantiles.

D.3 Other measures of innovation

In this section, we run our baseline regression using different proxies for the intensity of R&D. As seen in Table D4, the effect of the intensity of R&D is always positive and significant.

D.4 Other robustness

In this last section we test three additional robustness checks. First, as seen in Table A1, firms from the highest quantile of R&D are very different from others. We thus check that our results are not mainly driven by these firms by removing observations associated with total R&D expenditures higher than 293,634,000 pounds. Results are shown in Table D5. Second, we run our main regressions restricting on firms with positive expenditures in R&D in the current year. We change the measure of R&D to $ln(R_{ft})$ with R_{ft} being the total expenditures in R&D of firm f during year t. Results are presented in Table D6. Third, we test the robustness of our results regarding the different effects of R&D to income by skill to using an alternative definition of skill level as defined in subsection A.2.3. Results are robust in the sense that there is no effect of R&D expenditures on income for high occupation workers as presented in Table D7 where each column corresponds to a different skill level (1 for the lowest and 4 for the highest).

Table D3: 20 QUANTILES OF R&D BASED ON LEVEL OF TOTAL R&D EXPENDITURES

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
Quantile 1	-0.0233***	-0.0172***	-0.00557	-0.0196***
Quantile 2	0.0471***	0.00118	0.0150***	-0.00454
Quantile 3	-0.0170**	-0.0267***	0.00512	0.000841
Quantile 4	-0.0226***	-0.00101	0.0153***	-0.00538
Quantile 5	0.0502***	0.0376***	0.0187***	-0.00229
Quantile 6	0.0267***	0.00483	0.0109***	0.00622
Quantile 7	0.00729	0.0101	0.00132	-0.0362***
Quantile 8	0.0478***	0.0341***	0.00461	-0.0290***
Quantile 9	0.0531***	0.0356***	0.0228***	-0.0137**
Quantile 10	0.0733***	0.0522***	0.0281***	-0.000501
Quantile 11	0.0904***	0.0513***	0.0161***	-0.0181***
Quantile 12	0.0439***	0.0341***	0.0337***	0.00846
Quantile 13	0.0704***	0.0398***	0.0270***	-0.0190***
Quantile 14	0.0745***	0.0483***	0.0269***	0.0168***
Quantile 15	0.146***	0.0961***	0.0330***	0.00276
Quantile 16	0.167***	0.0997***	0.0366***	0.0192***
Quantile 17	0.234***	0.109***	0.0440***	0.0241***
Quantile 18	0.271***	0.141***	0.0492***	0.0249***
Quantile 19	0.248***	0.149***	0.0607***	0.0500***
Quantile 20	0.380***	0.197***	0.0844***	0.0208**
Fixed Effects	(k,t)	(k,j,t)	i+t	f+t
Observations	572,799	572,799	572,799	572,799

Notes: This table presents the coefficient on each of the 20 quantiles of total R&D expenditure when estimating equation D1. The set of control variables and fixed-effects are the same as in Table 3. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table D4: Robustness to using different measures of R&D.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
R&D	0.0286***	0.0300***	0.0123***	0.239***
	(0.002)	(0.002)	(0.003)	(0.024)
Age^2	-0.000590***	-0.000590***	-0.000593***	-0.000592***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.00777***	0.00777***	0.00787***	0.00787***
	(0.000)	(0.000)	(0.000)	(0.000)
$Tenure^2$	-0.0000870***	-0.0000867***	-0.0000872***	-0.0000885***
	(0.000)	(0.000)	(0.000)	(0.000)
$ln(emp)$	-0.00721***	-0.00722***	-0.00739***	-0.00712***
	(0.001)	(0.001)	(0.001)	(0.001)
Full Time	-0.000678	-0.000666	0.000379	0.000118
	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effects	i+t	i+t	i+t	i+t
N	572,799	572,799	572,799	572,799
R^2	0.888	0.888	0.888	0.888

Notes: This table presents results from estimating the effect of R&D intensity on income. Column 1 uses total R&D expenditures per number of employees, column 2 and 3 uses respectively intramural and extramural R&D expenditures per number of employees and column 4 uses the share of workers involved in R&D activities. All these measures are transformed with a function $ln(1+x)$. Control variables definition and construction are given in Table A7. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table D5: ROBUSTNESS: REMOVING FIRMS FROM THE HIGHEST QUANTILE OF R&D EXPENDITURES.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
$ln(R_{ft}+1)$	0.123***	0.0694***	0.0295***	0.0143***
	(0.003)	(0.002)	(0.002)	(0.003)
Age	0.0584***	0.0340***		0.0446***
	(0.001)	(0.000)		(0.001)
Age^2	-0.000703***	-0.000393***	-0.000579***	-0.000523***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.0235***	0.0152***	0.00792***	0.0160***
	(0.000)	(0.000)	(0.000)	(0.000)
$Tenure^2$	-0.000316***	-0.000224***	-0.0000933***	-0.000232***
	(0.000)	(0.000)	(0.000)	(0.000)
$ln(emp)$	-0.0315***	-0.00829***	-0.00743***	-0.0237***
	(0.001)	(0.001)	(0.001)	(0.003)
Male	0.162***	0.145***		0.159***
	(0.003)	(0.002)		(0.003)
Full Time	0.250***	0.0740***	0.000981	0.143***
	(0.002)	(0.002)	(0.002)	(0.002)
Fixed Effects	(k,t)	(k,j,t)	i+t	f+t
N	546,556	546,556	546,556	546,556
R^2	0.368	0.614	0.884	0.550

Notes: This table presents estimates of the effect of R&D as measured by the logarithm of 1 + total R&D expenditures divided by employment in the year, on the logarithm of income as measured by the gross hourly earnings (in log). Firm with R&D expenditures higher than 293,634,000 pounds in the current year are excluded (top vintile). Control variables definition and construction are given in Table A7. Column 1 uses labor market interacted with year fixed effect, column 2 uses labor market interacted with year and occupation fixed effects, column 3 uses firm fixed effects and column 4 uses individual fixed effects. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table D6: ROBUSTNESS: REMOVING FIRMS WITH NO R&D EXPENDITURES.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
$ln(1+R_{ft})$	0.0504***	0.0319***	0.00532***	0.00164
	(0.001)	(0.001)	(0.001)	(0.001)
Age	0.0650***	0.0407***	0	0.0560***
	(0.001)	(0.001)	(.)	(0.001)
Age^2	-0.000745***	-0.000450***	-0.000546***	-0.000635***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.0139***	0.0108***	0.00528***	0.0122***
	(0.001)	(0.001)	(0.001)	(0.001)
$Tenure^2$	-0.000198***	-0.000184***	-0.0000765***	-0.000186***
	(0.000)	(0.000)	(0.000)	(0.000)
$ln(emp)$	-0.0137***	-0.0101***	-0.00132	-0.0326***
	(0.002)	(0.001)	(0.003)	(0.006)
Male	0.177***	0.161***	0	0.166***
	(0.005)	(0.005)	(.)	(0.005)
Full Time	0.200***	0.0318***	-0.0860***	0.137***
	(0.006)	(0.005)	(0.008)	(0.006)
Fixed Effects	(k,t)	(k,j,t)	i+t	f+t
N	144,205	144,205	144,205	144,205
R^2	0.407	0.631	0.917	0.512

Notes: This table presents estimates of the effect of R&D as measured by the logarithm of total R&D expenditures divided by employment in the year, on the logarithm of income as measured by the gross hourly earnings (in log). Firm with 0 R&D expenditures are excluded. Control variables definition and construction are given in Table A7. Column 1 uses labor market interacted with year fixed effect, column 2 uses labor market interacted with year and occupation fixed effects, column 3 uses firm fixed effects and column 4 uses individual fixed effects. Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table D7: ROBUSTNESS: ALTERNATIVE MEASURE OF SKILL.

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
$ln(R_{ft}+1)$	0.0359***	0.0339***	0.00985***	-0.00117
	(0.007)	(0.003)	(0.003)	(0.002)
Age^2	-0.000208***	-0.000361***	-0.000613***	-0.000875***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.00733***	0.00932***	0.00342***	0.00144**
	(0.001)	(0.001)	(0.001)	(0.001)
$Tenure^2$	-0.000124***	-0.000151***	-0.0000538***	-0.00000546
	(0.000)	(0.000)	(0.000)	(0.000)
$ln(emp)$	0.00360*	-0.00645***	0.000285	0.00625**
	(0.002)	(0.001)	(0.003)	(0.003)
Full Time	-0.0428***	-0.0159***	-0.120***	-0.118***
	(0.006)	(0.003)	(0.011)	(0.013)
Skill Level	1 (low)	2	3	4 (high)
Fixed Effects	i+t	i+t	i+t	i+t
N	92,305	268,760	104,647	107,087
R^2	0.701	0.784	0.870	0.900

Notes: This table presents estimates of the effect of R&D as measured by the logarithm of 1 + total R&D expenditures divided by employment in the year, on the logarithm of income as measured by the gross hourly earnings (in log). Control variables definition and construction are given in Table A7. Column 1 restricts to lowest skill workers (skill level 1) with the alternative definition of skill presented in subsection A.2.3. Column 2 restricts to skill level 2, etc... Ordinary Least Square regression. Heteroskedasticity robust standard errors clustered at the individual level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.