The Innovation Premium to Low-Skill Jobs *

Philippe Aghion   Antonin Bergeaud

Richard Blundell   Rachel Griffith

February 23, 2018

Abstract

This paper uses matched employee-employer data from the UK augmented with information on R&D expenditures, to analyze the relationship between innovativeness of the firm and the wages of workers in occupations with different levels of skills. We show that more R&D intensive firms pay higher wages on average and more importantly that the premium to working in an R&D intensive firm is higher for workers in low-skilled occupations than for workers in high-skilled occupations. To rationalize these findings we develop a simple model of the firm where the complementarity between workers in high-skilled and low-skilled occupations increases with the firm's innovativeness. Additional predictions of the model, also confirmed by the empirical analysis, is first, that the tenure of workers in low-skilled occupations is longer in more innovative firms than in less innovative firms, and second that more innovative firms outsource a higher fraction of tasks.

*Aghion, College de France, LSE and CEPR, p.aghion@lse.ac.uk; Bergeaud, Banque de France, antonin.bergeaud@banque-france.fr; Blundell, UCL, IFS and CEPR, r.blundell@ucl.ac.uk; Griffith, University of Manchester, IFS and CEPR, rgriffith@ifs.org.uk. We thank the Centre for the Microeconomic Analysis of Public Policy (CPP) for funding under grant number ES/M010147/1.

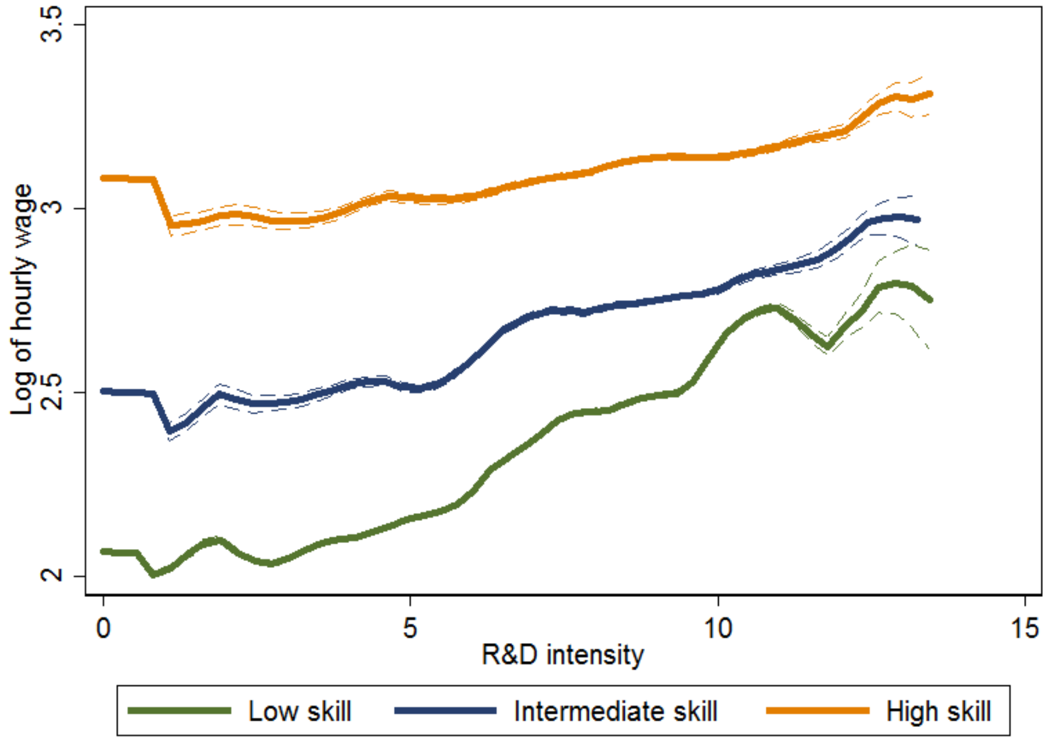
1 Introduction

This paper has come out of a surprising finding. Looking at the wage premia to working in more innovative firms, our prior belief, based on what we have learned about skill-biased technical change, would be that such premia should be higher for high-occupation workers than for low-occupation workers. However Figure 1, drawn from our data, shows the opposite: namely, that the wage premium from working in a more R&D intensive firm is higher for workers in low-skilled occupations. In this figure we see that workers in high-skilled occupations earn more than workers in low-skilled and intermediate-skilled occupations regardless of where they work (the high-skill wage curve always lies strictly above the intermediate-skill curve). Workers in intermediate-skilled occupations earn more than workers in low-skilled occupations conditional on the level of R&D intensity of the firm they work for, but workers in intermediate-skilled occupations working in low R&D intensity firms earn less than workers in low-skilled occupations who work in high R&D intensity firms. More interestingly the wages of workers in lower-skilled occupations increases more steeply; workers in all occupations obtain some wage premium from working in a more innovative firm, but that premium (the slope of the curve) is greater for workers in low-skilled occupations.

Our contribution in this paper is twofold. First, we use matched employer-employee data from the UK, augmented with information on R&D expenditures, to show that the wage premium from working in a more R&D intensive firm, relative to working in a less R&D intensive firm, is indeed higher for workers in low-skilled occupations than those in high-skilled occupations. Second, we propose a model in the spirit of Kremer (1993) and Garicano and Rossi-Hansberg (2006) to rationalize this finding, in which more innovative firms display a higher degree of complementarity between workers in low-skill occupations and other factors of production (capital and high-skill labour) within the firm; we show that this model has additional empirical implications, in particular on tenure, training and outsourcing of low versus high occupation workers, which are consistent with the data.

The underlying idea is that workers in high-skilled occupations typically have observable qualifications, and their market value is largely determined by their education and accumulated reputation, meaning that the skills of workers in high-skilled occupations are less firm-specific than workers in low-skilled occupations. A firm can replace a worker in a high-skilled occupation by another similar worker with limited downside risk, because their quality is observable. In contrast, the qualities of work-

Figure 1: Wage by occupation skill group



Notes: This figure plots the log of hourly wages of workers against R&D intensity of the firm for which they work. Data have been fitted using local polynomial filtration. Wages are defined in Appendix A.2.2 and include overtime and bonuses. Skill level is defined in Appendix A.2.3. R&D intensity is defined in Appendix A.1. 95% confident intervals are shown by dashed lines.

ers in *some* low-skilled occupations can be firm-specific and the potential to develop these skills can be difficult to observe, so difficult to replace. If these workers are complementary in production to the workers in high-skilled occupations, then the costs to the firm can be high, and these works can command a higher wage.

An important insight from this model is that innovativeness impacts on the organizational form of the firm, and in particular on the complementarity or substitutability between workers in occupations with different skill levels within the firm. Think of a worker in a low-skilled occupation, for example an assistant, who shows outstanding ability, initiative and trustworthiness. That worker performs tasks for which it might be difficult to find a high-skill worker, and which complements the tasks performed by high-skill employees within the firm in the sense that mistakes by the low-occupation employee can be damaging to the firm's overall performance.

The model has additional empirical implications that we bring to data. First, it suggests that training should be higher amongst workers in low-skilled occupations who work for R&D intensive firms compared to similar workers who work for non-R&D intensive firms, as R&D intensive firms have a higher incentive to train the low-occupation employees they decide to retain than less R&D intensive firms. Moreover, in R&D intensive firms tenure should be higher for low-occupation workers than for the high-occupation workers to the extent that the former rely more on firm-specific training whereas the latter rely more on their prior education. The model also predicts that more R&D intensive firms will prefer to outsource workers in low-skilled occupations where these is less complementarity with high-skilled workers (consistent with Goldschmidt and Schmieder, 2017). We find support for both predictions.

The paper relates to several strands of literature. First, 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), Krusell et al., 2000, Acemoglu, 2002). Our finding that the premium to working in more innovative firms is higher for low-occupation workers, may appear at odds with the view that technical change has become increasingly skill-biased over the past thirty five years. However, we also find that more innovative firms outsource a higher fraction of low-occupation tasks which reconciles our main finding with the fact that on average in the overall economy high skill workers perform better as technology advances.

Second, there is the labour and wage literature (Gibbons and Katz, 1992; Groshen, 1991 and Abowd et al., 1999 among others); 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. For example, Card et al. (2016) assume that firm heterogeneity arises through TFP, but do not model what drives these differences in TFP. Other studies report a link between productivity and wage policy (Cahuc et al., 2006 and Barth et al., 2016 among others) and Song et al. (2015) consistently find that "between firm inequality" accounts for the majority of the total increase in income inequality between 1981 and 2013 in the US. A recent trend of this literature is to link the aggregate dispersions in wages to productivity dispersion across firms (Barth et al., 2016, Dunne et al., 2004). Matched worker-employee data are often used (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."*. We contribute to this literature by bringing innovation into the picture, and by analysing the relationship between innovation, wages and occupation across firms.

Third, we should mention two recent papers, namely Kline et al. (2017) and Aghion et al. (2018), which use individual fiscal data merged with patent data respectively in the US and in Finland to look at the individual returns from innovation to the inventors and to their co-workers. Both paper find significant returns to innovation, most of which accrue to other employees or stakeholders within the inventor's firm.[1] We contribute to this literature by focusing on the comparison between high- and low-occupation workers in more versus less innovative firms, and on how innovativeness affects the degree of complementarity between high occupation workers and low occupation workers.

Finally, we draw on the literature on wage inequality and the organization of the firm (e.g. see Kremer, 1993, Kremer and Maskin, 1996, 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 finding, namely that more innovative firms pay higher wages and that the premium to working in more innovative firms is higher for workers in low-skilled occupations. In Section 3 we develop a model to account for these findings and derive two additional predictions from this model. In Section 4 we test these additional predictions and discuss the robustness of our main findings. Section 5 collects our concluding remarks.

2 The main findings

In this section we present our data and empirical methodology, and establish our main empirical findings, namely that more innovative firms pay higher wages and that the

[1] Thus Kline et al. (2017) find that workers capture 29 cents of every dollar of patent-induced operating surplus. Aghion et al. (2018) find that inventors get only 7.9% of the total gains; second, entrepreneurs get over 44.5% of the total gains; and finally, blue-collar workers get about 25.7% of the gains.

premium to working in more innovative firms is higher for workers in low-skilled occupations.

The literature has been relatively silent as to why some firms pay higher wages than others for workers that appear similar. In a competitive labour 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).

Here we focus on the relationship between the wages of workers and the R&D intensity of the firms they work for. We use novel matched employer-employee data for the UK that also contains information on R&D expenditure for the period 2004 to 2015. The employee data come from the Annual Survey of Hours and Earnings (ASHE), which is a random sample of 1% of the UK working population. We match this to the Business Expenditure on Research and Development (BERD) survey, which is a census for firms with 400+ employees. The data are longitudinal, we follow the same workers over time, and is recorded at the establishment level, with information on which establishments are part of the same firm. We focus on private companies (excluding the public sector, charities, etc) that have 400 or more employees. We use information on *186,000* employees who work in around *7,370* firms, giving us a total of *626,722* observations. Further details on the data are given in Appendix A.

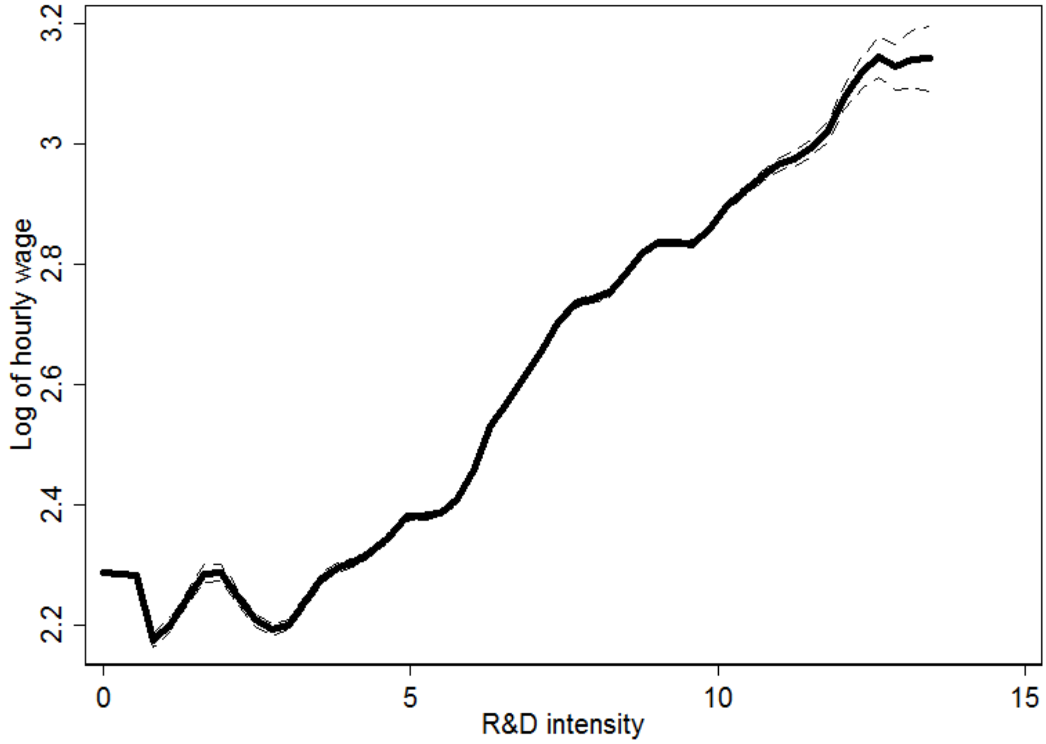
2.1 First finding: wages are higher in more innovative firms

There are significant differences in the wages paid to workers in innovative firms compared to those working in non-innovative firms at all ages and after controlling for a range of worker and firm characteristics. Figure 2 shows that the average wage of workers increases with the firm's R&D intensity; average wages are around 150% higher in the most R&D intensive firms compared to firms that do no R&D.[2] 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. Another way

[2] This is $(exp(3.2) - exp(2.287))/exp(2.287) = 1.49$.

to see this is by looking at the share of workers that work in a firm that does any R&D across the wage distribution; this 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 (see Figure A4 in Appendix A.5).

Figure 2: Wages and R&D intensity



Notes: This figure plots a local polynomial of log hourly wages on log R&D intensity. Wages are defined in Appendix A.2.2. R&D intensity is defined in Appendix A.1.

Workers in more R&D intensive firms might have different characteristics to those working in less R&D intensive firms. Table 1 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), all of which might affect the wages of workers in these firms. In Appendix A we give further descriptive statistics of the key variables.

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

Table 1: Comparison of R&D and non R&D firms

	Innovative firm		Current R&D firms	
	Yes	No	Yes	No
Employment	2,784	2,213	2,543	2,365
Hourly Wage (£)	15.8	12.5	16.1	12.9
Share of Male (%)	68	56	70	57
Share of full-time	90	76	92	77
Workers in high-skill occupations (%)	30	18	31	19
Workers in low-skill occupations (%)	51	65	50	63
Age	40.4	38.1	41.1	38.3
Tenure	8.8	5.7	9.5	5.9
Workers	72,718	113,181	52,617	135,551
Firms	2,332	5032	1,877	5,939
Firms-years	12,871	25,481	8,542	29,810
Worker-firm-year	263,447	363,275	162,764	463,958

Notes: Innovative firms are those that report any R&D expenditure over the period, current R&D firms are those that report a positive amount of R&D expenditure in that period. Employment is the average number of workers in the firm over all years. Wages are defined in Appendix A.2.2. Skill level is defined in Appendix A.2.3.

$$ln(w_{ijkft}) = \beta_1 a_{it} + \beta_2 a_{it}^2 + \beta_3 T_{ift} + \beta_4 T_{ift}^2 + \beta_5 ft_{ift} + \beta_6 S_{ft} + \beta_7 \tilde{R}_{ft} + \gamma_i + \eta_t + e_{ijkft}, \tag{1}$$

where i indexes individual, j occupation, k labor market, f firm and t years. w_{ijkft} is hourly wages, a_{it} is age, T_{ift} is tenure in the firm, $ft_{it} = 1$ if the job is full-time (as opposed to part-time), S_{ft} is number of employees in the firm. $\tilde{R}_{ft} = ln(1 + R_{ft})$ is R&D intensity.[3] η_t capture common time effects.

Our main specification, column (3) in Table 2, includes individual worker effects (γ_i). These are important. Higher quality workers might select into higher quality firms. In column (1) we include only labour market (defined as a travel to work area;

[3] R&D expenditure divided by number of employees, we use $ln(1 + R_{ft})$ to accommodate values of zero 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_7 as the elasticity of wage with respect to R&D intensity. In Section 4.3 we show robustness of our results to alternative functional forms and alternative measures of R&D.

there are around 240 such areas in the UK, see Appendix A.3) and time effects. The coefficient estimate of 0.029 suggest that workers in the most R&D intensive firms earn nearly 50% more than workers in firms that do no R&D,[4] controlling for these characteristics accounts for a substantial part of the differences we saw in the raw data.

Table 2: Relationship between wages and R&D intensity

	Dependent variable: $ln(w_{ijkft})$		
	(1)	(2)	(3)
$\tilde{R}_{ft}$	0.029***	0.016***	0.006***
	(0.002)	(0.001)	(0.001)
Age	0.058***	0.034***	
	(0.003)	(0.002)	
Age Sq.	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)
Tenure	0.023***	0.015***	0.008***
	(0.001)	(0.001)	(0.000)
Tenure Sq.	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)
Firm Size	-0.032***	-0.010***	-0.008***
	(0.006)	(0.004)	(0.002)
Gender	0.156***	0.143***	
	(0.006)	(0.004)	
Full-Time	0.244***	0.070***	0.004
	(0.014)	(0.007)	(0.005)
Fixed Effects	(k,t)	(k,j,t)	i+t
R^2	0.385	0.624	0.887

Notes: 626,210 observations. The dependent variable is log of wage which is defined in Appendix A.2.2. $\tilde{R}_{ft} = ln(1 + R_{ft})$. Other covariates definitions are given in Table A7. Column 1 includes year-labour market fixed effects, column 2 includes year-labour market-occupation effects, column 3 includes year and individual fixed effects. The specification in column 3 can't identify Age and the Gender dummy because of additive worker and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

[4] This is (exp(predicted wage at $max(\tilde{R}_{ft})$) minus exp(predicted wage at $min(\tilde{R}_{ft})$))/exp(predicted wage at $min(\tilde{R}_{ft})$)=(exp(2.678)-exp(2.287))/exp(2.287)=0.48, where the predictions use the coefficient estimates from column (1) of Table 2.

In column (2) we add occupation effects at the two-digit level (25 occupations). This reduces the coefficient on R&D intensity by about half, the coefficient estimate of 0.016 suggest that workers in the most R&D intensive firms earn around 24% more than workers in firms that do no R&D.[5] In column (3) we add worker effects. We drop occupation and labour market effects as we do not observe many workers who move across occupations or labour markets. This reduces the coefficient on R&D intensity to 0.006, which implies that workers in the most R&D intensive firms earn around 8% more than workers in firms that do no R&D.[6] Compared to the estimates in column (3) the estimates without worker effects considerably over-estimate the impact of R&D intensity on wages.

In column (3) we include worker effects, so identification comes off individuals who move jobs between firms that do more or less R&D, and individuals working in firms that increase or decrease their R&D intensity. We face a potential endogeneity problem if better workers will get better matches and remain in the labour market longer and if the worker effect (γ_i) does not account for this. This would imply that

$$e_{ijkft} = \xi_{ift} + u_{ijktf}$$

where ξ_{ift} is an unobserved work-firm-time varying productivity, and that this is correlated with firm level R&D, $E[\xi_{if0}, \tilde{R}_{ft}] > 0$.

In section 4 we show that this is not the only thing going on because, for example, workers in low-skilled occupations get more training in more R&D intensive firms than in less R&D intensive firms (see Table 6). Firms invest in increasing ξ_{ift} of workers in some low-skilled occupations, because it also improves productivity of workers in high-skilled occupations when they are complements.

2.2 Second finding: higher returns to innovativeness for low-occupation workers

We use a definition of skill based on a match between qualifications and occupations, see details in Appendix A.2.3. We consider three skill groups. Low-skilled occupations are those that require on minimal formal education and training, but might require considerable on-the-job knowledge, such as maintenance technicians, telephone sales,

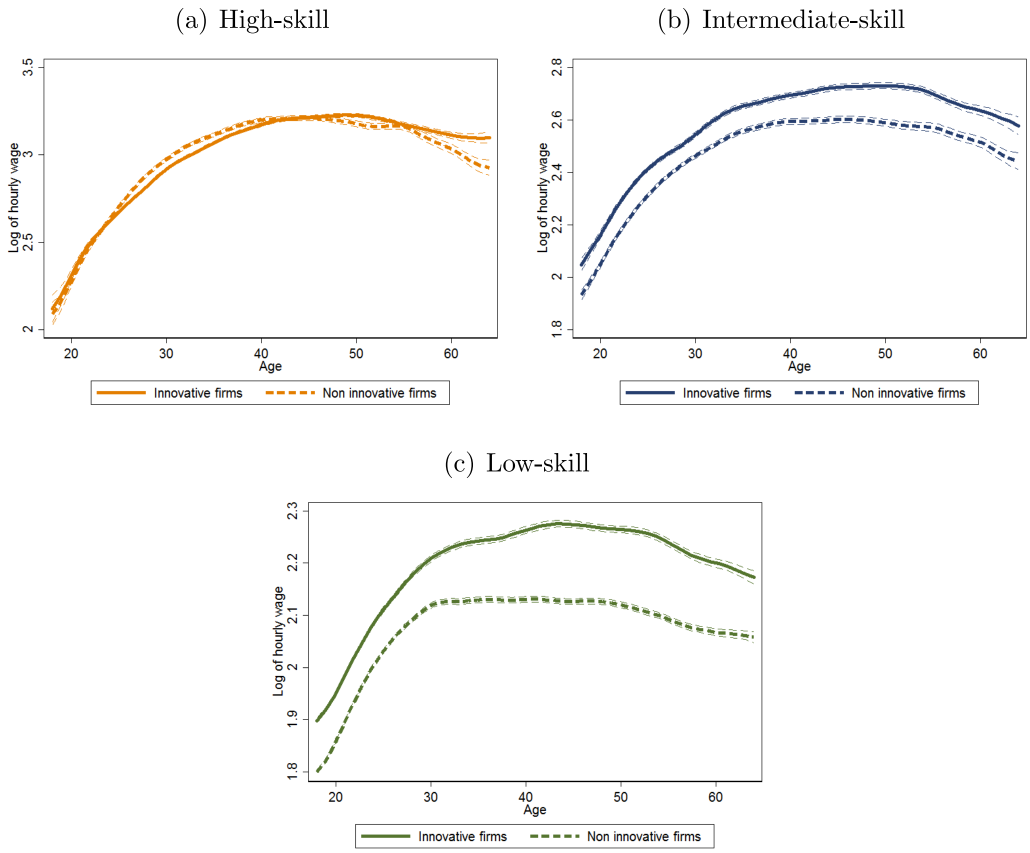
[5]As footnote 4, using the coefficient estimates from column (2) of Table 2, =exp(2.506)-exp(2.287))/exp(2.287)=0.24.

[6]As footnote 4, using the coefficient estimates from column (3) of Table 2,=(exp(2.368)-exp(2.287))/exp(2.287)=0.0.08.

as well as cleaning and other jobs that will not require any job-specific knowledge. Intermediate-skilled occupations typically require the equivalent of a high-school education and include trades, specialist clerical, associate professionals. High-skilled occupations typically included advance training or a university degree and include engineers and managers.

Workers in high-skill occupations receive higher wage than workers in lower-skilled occupations at all ages. Our focus in this paper is on the impact of innovativeness on between-firm wage inequality, in other words how the returns to working in a more innovative firm varies across the skill distribution. Figure 3 shows that workers earn higher wages when the firm that they work for is innovative (does any R&D). When we look by skill category we see that the within-skill group variance of wages across firms is relatively more important for workers in low-skill occupations than workers in high-skill occupations. Workers in higher skill occupations earn the highest wages, and these wages are on average similar across firms that are more or less R&D intensive. 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 workers in low-skilled occupations.

Figure 3: Wage, by skill, R&D intensity and age



Notes: This figure plots age dummies from a regression of log hourly wage on year-travel to work area effects. The lower curve is for workers in non-innovative firms, the upper curve for workers in innovative firms. Innovative firms are firm that report at least £1 in R&D expenditures over the period. 95% confident intervals are included.

Highly innovative firms also hire fewer workers in low-skilled occupations. Table A9 in the Appendix shows that moving from the least to the most R&D intensive firm increases the share of workers in high-skilled occupations from 13.7% to 53.8%.

In order to see if the wage premium shown in Figure 3 is robust to controlling for other difference in workers and firms we re-estimate our preferred specification with individual fixed effects (column 3 of Table 2) separately for workers in occupations of different skill levels. We show results for workers in low-skilled occupations in column 1 of Table 3, intermediate-skilled occupations in column 2 and high-skilled in column 3. The positive coefficient on R&D intensity holds for low and intermediate-skill categories and is strongest for the low-skilled occupations. 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 workers in high-skilled occupations (the omitted category), workers in intermediate and low-skilled occupations earn more from working in an R&D firm.

The estimates in column 4 suggest that workers in low-skilled occupations in the most R&D intensive firms earn 12% more than workers in firms that do no R&D,[7] for workers in intermediate-skilled occupations they earn 6% more and for workers in high-skilled occupations 3% more, once we condition on worker effects and other observables.

The finding that the premium to working in a more innovative firm is larger for workers in lower skilled occupations may look somewhat counter-intuitive and at odds with the literature on skill-biased technical change. 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 workers in low-skill and high-skill occupations.

[7] As footnote 4, using the coefficient estimates from column 4 of Table tab:regskill, (exp(2.174))-exp(2.065))/exp(2.065)=0.12.

Table 3: R&D intensity and hourly wages at different skill levels

	Dependent variable: $ln(w_{ijkft})$			
Skill Category	Low (1)	Intermediate (2)	High (3)	All (4)
ln R&D int	0.007*** (0.001)	0.003*** (0.001)	-0.000 (0.001)	0.002*** (0.001)
*med skill				0.002*** (0.001)
*low-skill				0.006*** (0.001)
Age Sq.	-0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Tenure	0.009*** (0.001)	0.006*** (0.001)	0.001 (0.001)	0.007*** (0.000)
Tenure Sq.	-0.000*** (0.000)	-0.000*** (0.000)	0.000 (0.000)	-0.000*** (0.000)
Firm Size	-0.005** (0.002)	0.002 (0.003)	0.004 (0.002)	-0.006*** (0.002)
Full-Time	-0.011* (0.006)	-0.089*** (0.014)	-0.109*** (0.014)	-0.004 (0.005)
low-skill				-0.136*** (0.043)
med-skill				-0.052 (0.043)
high-skill				0.021 (0.044)
R^2	0.774	0.851	0.885	0.889
Observations	407,341	104,318	114,535	626,210

Notes: The dependent variable, log of wage, is defined in Appendix A.2.2. The main regressor, ln R&D int, is defined as the logarithm of total R&D expenditures divided by total employment at the firm level. Other covariates definitions are given in Table A7. Ordinary Least Square regression including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

3 A 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. Workers in low-skilled occupations 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 workers in low-skilled occupations are more complementary to workers in high-skilled occupations.

The model is meant to capture the idea that workers in low-skilled occupations 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 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 Model setup

Production function

We consider a representative firm with a continuum of tasks indexed by $\lambda \in [0,1]$, where the output of each task is produced according to a partially O'Ring production function (see Kremer, 1993). Each task λ combines one high skill occupation associated with quality Q and one low skill occupation associated with quality $q(\lambda)$. We take the quality Q of high-skill workers to be the same on all tasks. More formally, for all $\lambda \in [0,1]$:

$$f(\lambda, q, Q) = \lambda q Q + (1-\lambda)(q+Q).$$

The firm's total output is then taken to be a weighted sum of the outputs on the individual tasks. Formally, if $\phi(\lambda)$ denotes the weight function on tasks, we let:

$$F(\vec{q}, Q) = \int_0^1 f(\lambda, q(\lambda), Q)\phi(\lambda)d\lambda.$$

where

$$\vec{q} = (q(\lambda))_{\lambda\in[0,1]} \text{ and } \int_0^1 \phi(\lambda)d\lambda = 1.$$

Wage negotiation

For each task λ, the firm engages in separate wage negotiations with the high and low skill workers on that task. This negotiation will lead to the equilibrium wage $w_q(\lambda)$ for the low-occupation worker and to w_Q for the high-occupation worker. We denote by β^L (resp. β^H) the fraction of the firm's net surplus that accrues to the low-occupation (resp. high occupation) worker, where we assume: $\beta^L \leq \beta^H < 1$.

Wages within the firm are determined by Nash bargaining following Stole and Zwiebel (1996). In this bargaining, the firm has the outside option of replacing a low-occupation worker within the firm with quality $q(\lambda)$ and paid at wage $w_q(\lambda)$ by an outside worker with reservation quality q_L at reservation wage w_L.[8] Similarly, if its negotiation with the high-occupation worker fails, the firm must look for a substitute high-occupation worker of quality Q_L. We assume that it is easier for the firm to find a substitute for the high-occupation worker than for the low-occupation worker. The underlying idea is that the ability of a low-occupation worker is harder to detect ex-ante, in particular because less information is acquired by the firm ex ante based on the employee's CV (education, reputation); in contrast, a high-occupation worker is more likely to have built a reputation or a CV which informs about her level of ability.[9] Formally, this leads us to assume that:

$$Q - Q_L < q(\lambda) - q_L.$$

we shall also assume that

$$Q > Q_L >> q(\lambda) > q_L > 1.$$

[8]An alternative interpretation is that absent a wage agreement the low-occupation worker chooses to underperform at quality level q_L.

[9]An alternative interpretation is that the high-occupation worker is less likely to underperform in the absence of an agreement as she still wants to preserve her reputation.

Substitute workers in low-skilled and high-skilled occupations 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 $\bar{w}^L$ and $\bar{w}^H$, which are also exogenous. We assume: $w_L < w_H$ and $\bar{w}^L << \bar{w}^H$.

The firm's total wage bill is then equal to

$$W(\vec{q}) = \int_0^1 w_q(\lambda)d\lambda + w_Q,$$

Training and profits

The firm's ex post profit is equal to:

$$\tilde{\Pi}(\vec{q}) \equiv F(\vec{q}) - W(\vec{q}).$$

We assume that prior to wage negotiation, the firm can train the low-occupation worker on each task λ from q_L to some higher quality level $q(\lambda)$ at a quadratic cost. The firm's ex ante training investment will seek to maximize:

$$\tilde{\Pi}(\vec{q}) - \int_0^1 C(\lambda)\left(q(\lambda) - q_L\right)^2 d\lambda,$$

with respect to $\vec{q} = (q(\lambda))_\lambda$.

Innovativeness and task distribution

We know from Caroli and Van Reenen (2001), Acemoglu et al. (2007) and Bloom et al. (2014) that more frontier firms are also more innovative and tend to exhibit flatter organizational forms, with more strategic complementarity between firm's employees. We capture this idea in our model by assuming that more frontier firms focus on tasks with higher complementarity between high- and low-occupation workers on average. More formally, we assume that a more frontier firm is characterized by a higher level of

$$\mathbb{E}_\phi(\lambda) = \int_0^1 \lambda\phi(\lambda)d\lambda.$$

We shall parametrize $\phi(\lambda) = \phi(\lambda, z)$ by the firm's level of innovativeness z. We assume that ϕ is increasing both in z and in λ. Tractable cases which will allow us to nicely develop our intuitions, are:

Example 1. *Suppose that* $\phi(\lambda, z) = (z+1)\lambda^z$*. In that case we have:*

$$\mathbb{E}_\phi[\lambda] = 1 - \frac{1}{z+2},$$

which increases with the innovation intensity z*.*

Example 2. *An even simpler case which we shall refer to as the "toy case", is where* $\phi(\lambda, z)$ *is equal to 1 only for* $\lambda = \lambda_z \equiv \frac{z}{z_{max}}$ *(where* z_{max} *denotes the maximum value* z *can take so that* $\lambda_z \in [0,1]$*) and zero otherwise. In that case:*

$$\mathbb{E}_\phi[\lambda] = \frac{z}{z_{max}},$$

3.2 Solving the model

From this point, we shall simplify the analysis by assuming that the bargaining surplus is split equally between the firm and each worker ($\beta_H = \beta_L = 1$) and that the training cost parameter C is independent of the task.

3.2.1 The toy case

Here we assume that $\phi(\lambda, z) = 1$ if $\lambda = \lambda_z$ and 0 otherwise. In this case, the firm with innovativeness level z has only one task $\lambda = \lambda_z$ performed (other tasks are irrelevant to the firm since they have no impact on its production).

Equilibrium low-occupation wage The firm's net surplus from employing a low-occupation worker with quality q, is equal to:

$$[\lambda_z Q + (1-\lambda_z)](q - q_L) - w_q + w_L,$$

whereas the net surplus of the worker in the low-skilled occupation is equal to

$$S^{LS} = w_q - \bar{w}^L,$$

where $\bar{w}^L$ is the worker's outside option. We immediately obtain the following expression for the equilibrium wage of the low-occupation worker:

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

Equilibrium high-occupation wage Replicating the same argument for the high-occupation worker, we obtain the following expression for the equilibrium wage of the high-occupation worker:

$$w_Q(\lambda_z, q, Q) = \frac{Q - Q_L}{2}\left(\lambda_z(q-1)+1\right) + \frac{w_H + \bar{w}^H}{2} \tag{3}$$

Optimal training decision Having determined the equilibrium wages w_Q and w_q for given q, Q and z, we now move back and consider the firm's optimal choice of qualities (q^*, Q^*), where we impose

$$q \in [q_L, \overline{q}]; Q \in [Q_L, \overline{Q}].$$

Then the firm chooses (q^*, Q^*) by solving:

$$(q^*, Q^*) = \underset{q_L<q<\overline{q}\ Q_L<Q<\overline{Q}}{\text{argmax}} \left\{ f(\lambda_z, q, Q) - w_Q(\lambda_z, q, Q) - w_q(\lambda_z, q, Q) - C(q - q_L)^2 \right\}$$

With respect to Q, the problem is linear which leads to the corner solution $Q^* = \overline{Q}$. With respect to q, the problem is concave so that by first order condition we obtain:

$$q^*(\lambda_z) = q_L + \frac{1}{4C}\left[\lambda_z(Q_L - 1) + 1\right],$$

where we implicitly assume that this value if lower than $\overline{q}$.[10] Note that q^* is increasing with λ_z, and therefore with z: that is, the optimal level of training of a low occupation worker is higher in a more innovative firm.

Innovativeness and high versus low occupation wages The equilibrium wage of the low-occupation worker on task z, up to a constant, is equal to:

$$w_q(z) \equiv w_q(\lambda_z, q^*(z), Q^*)$$

or

$$w_q(z) = \frac{1}{8C}\left(\lambda_z(Q_L - 1) + 1\right)\left(\lambda_z(\overline{Q} - 1) + 1\right)$$

Similarly, the equilibrium wage of the high-occupation worker on task z, up to a constant, is equal to:

$$w_Q(z) \equiv w_Q(\lambda_z, q^*(z), Q^*) = \frac{(\overline{Q} - Q_L)}{2}\left[\lambda_z\left(q_L - 1 + \frac{\lambda_z(Q_L - 1) + 1}{2C}\right) + 1\right]$$

[10] A sufficient condition is that $\overline{q} > q_L + \frac{Q_L}{4C}$. Note that we must have $Q_L >> \overline{q}$, which is true as long as training costs are large enough.

or

$$w_Q(z) = \frac{(\overline{Q} - Q_L)}{2}\left[\lambda_z\left(q_L - 1 + \frac{\lambda_z(Q_L - 1) + 1}{2C}\right) + 1\right].$$

Proposition 1. *The premium to working in a more innovative task-firm is higher for low-occupation workers than for high-occupation workers:*

$$\frac{dw_q(z)}{dz} > \frac{dw_Q(z)}{dz}$$

Proof. The proof is detailed in Appendix C.1. □

The proposition immediately results from the fact that in more innovative firms the complementarity is higher between the low-occupation and the high-occupation workers, and that the optimal training of low-occupation workers is also higher in more innovative firms so that replacing the current low-occupation worker by an outside worker is more damaging for such a firm.

3.2.2 The continuous task distribution case

We now consider the case where firms cover a whole range of tasks with a continuous density distribution ϕ over tasks λ. We assume:

$$\phi(\lambda) = \phi(\lambda, z)$$

where ϕ is increasing in both λ and the innovativeness level z.

Nash bargaining and wage determination From the wage negotiation with Nash bargaining, we obtain the equilibrium wage for a low-occupation worker on task λ, for all $\lambda \in [0, 1]$:

$$w_q(\lambda) = \frac{\phi(\lambda, z)}{2}\left[f(\lambda, q(\lambda), Q) - f(\lambda, q_L, Q)\right] + \frac{(\bar{w}^L + w_L)}{2},$$

and similarly for the high occupation worker:

$$w_Q = \int_0^1 \left[f(\lambda, q(\lambda), Q) - f(\lambda, q(\lambda), Q_L)\right]\frac{\phi(\lambda, z)}{2}d\lambda + \frac{(\bar{w}^H + w_H)}{2}.$$

Profits The total wage bill is equal to

$$W(\vec{q}) = \int_0^1 w_q(\lambda)d\lambda + w_Q$$

so that total the firm's ex post profit $\tilde{\Pi}(\vec{q}) \equiv F(\vec{q}) - W(\vec{q})$ is equal:

$$\begin{aligned}\tilde{\Pi}(\vec{q}) &= \int_0^1 \left[\frac{f(\lambda, q_L, Q)}{2} + \frac{f(\lambda, q(\lambda), Q_L)}{2}\right]\phi(\lambda, z)d\lambda \\ &- \frac{\bar{w}^L + w_L}{2} - \frac{\bar{w}^H + w_H}{2}\end{aligned}$$

Optimal training Ex ante the firm chooses $\overrightarrow{q} = (q(\lambda))_\lambda$ to maximize profits net of training costs

$$\tilde{\Pi}(\vec{q}) - \int_0^1 C\left(q(\lambda) - q_L\right)^2 d\lambda.$$

This yields:

$$q^*(\lambda) = q_L + \phi(\lambda, z)\frac{\lambda(Q_L - 1) + 1}{4C},$$

Note in particular that here again $q^*(\lambda)$ is increasing with the innovation intensity level z and decreasing in the training cost parameter C.

Then the equilibrium wages for low and high occupation workers on task λ can be then rewritten (up to a constant):

$$w_q(\lambda, z) = \frac{\phi(\lambda, z)^2}{8C}\left(\lambda(Q_L - 1) + 1\right)\left(\lambda(\overline{Q} - 1) + 1\right)$$

and

$$\begin{aligned}w_Q(\lambda, z) &= \left(\overline{Q} - Q_L\right)\int_0^1 \lambda\frac{\phi(\lambda, z)^2}{8C}\left[\lambda(Q_L - 1) + 1\right]d\lambda \\ &\quad + \left(\overline{Q} - Q_L\right)\int_0^1 \frac{\phi(\lambda, z)}{2}\left[\lambda(q_L - 1) + 1\right]d\lambda\end{aligned}$$

Total effect of innovation We want to look at the effect of an increase in frontierness z on the average equilibrium wage of low versus high occupation workers. Let

$$\overline{w}_q(z) = \int_0^1 w_q(\lambda, z)d\lambda$$

and

$$\overline{w}_Q(z) = \int_0^1 w_Q(\lambda, z)d\lambda = w_Q$$

the average wages of low and high occupation workers respectively.

Proposition 2. *The average premium across tasks to working in more innovative firms, is higher for low occupation than for high occupation workers:*

$$\frac{d\overline{w}_q}{dz} > \frac{d\overline{w}_Q}{dz}$$

Proof. The proof is detailed in Appendix C.2. □

3.3 Outsourcing

We now extend the model by adding a capacity constraint whereby the firm only has limited time to train low occupation workers. Hence ex ante the firm maximizes

$$\tilde{\Pi}(\vec{q}) - \int_0^1 C(\lambda)\,(q(\lambda) - q_L)^2\, d\lambda$$

but now, subject to the time (or overload) constraint:

$$\int_0^1 (q(\lambda) - q_L)\, d\lambda \leq T,$$

Also, while it is straightforward to extend the proof to any function $\phi(\lambda, z)$, we shall focus attention on the case where $\phi(\lambda, z) = \lambda^z(z+1)$, where z is innovation level of the firm. In Appendix C.2.1 we hence establish:

Proposition 3. *There exists a cutoff value $\bar{\lambda}(z)$ such that for all tasks $\lambda \leq \bar{\lambda}(z)$, then $q^*(\lambda) = q_L$: in other words all tasks $\lambda \leq \bar{\lambda}(z)$ are outsourced. Moreover, we have*

$$\frac{d\bar{\lambda}(z)}{dz} > 0.$$

That is, more frontier firms outsource a higher fraction of tasks.

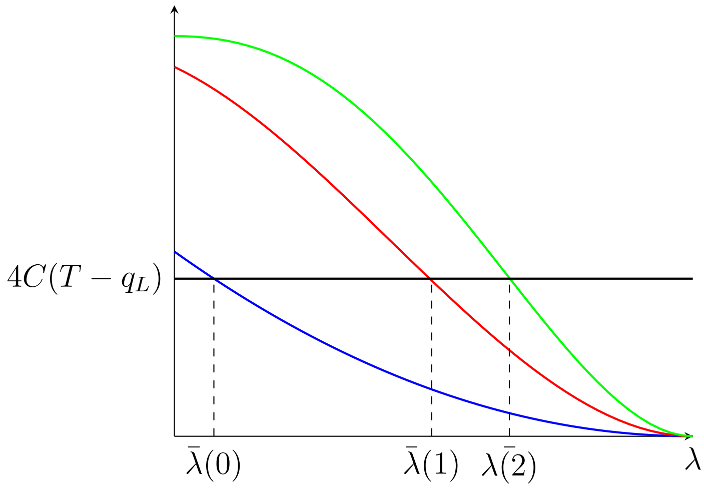
In the case where $\phi(\lambda, z) = \lambda^z(z + 1)$, it is possible to find tractable formula that defines $\bar{\lambda}$ for integer values of z (see Appendix C.2.1). Figure 4 shows the cases $z = 0, 1$ and 2.

3.4 Predictions

Prediction 1. *The premium to working in a more innovative firm is higher for low-occupation workers than for high occupation workers.*

We have already shown support for this first prediction in the previous section. The model also make the assumption that innovative firms put more weight in low-

Figure 4: $\bar{\lambda}$ as a function of z



skill occupation that require more firm-specific human capital, which is captured by the function $\phi(\lambda)$ and by the fact that $q - q_L > Q - Q_L$. This implies the following prediction:

Prediction 2. *Workers in low-skilled occupations should stay longer in more innovative firms (as more time and money is invested in them to getting them from q_L to q^*). Whereas we should not see such a difference for high skill workers.*

Finally, the model makes two additional predictions:

Prediction 3. *A more innovative firm will invest more in training its worker in low skill occupations than a non innovative firm. This is captured by the fact that $q - q_L$ is an increasing function of z in the model.*

Prediction 4. *A highly innovative firm will prefer to outsource the task that involves less complementarity between workers in low-skilled and high-skilled occupations.*

3.5 Extension

Appendix C.3 extends the model to the case where there is more than 1 worker in each occupation, and shows that the same predictions can be derived.

4 Empirical evidence

The model in the previous section yield predictions that we can take directly to the data. We already showed evidence supporting Prediction 1 in Section 2 (in particular see Figure 3 and Table 3). In this section we show empirical support for the other predictions and provide additional empirical results.

The model relies on the assumption that low-skill workers have more firm-specific human capital. We show support for this statement and for its implications, namely that workers in low-skilled occupations stay longer in more innovative firms, because more time and money is invested in increasing their productivity. We show that indeed more R&D intensive focus on low-skill tasks that are more complementary with the firm's assets and workers in high skill occupations and that in turns, require more training than other firms. We also show that workers in low-skilled occupations have longer tenure when they work in high R&D intensive firms that when they work in low R&D intensive firms.

4.1 Tenure and training in low-skilled occupations are higher in more R&D intensive firms

4.1.1 Tenure

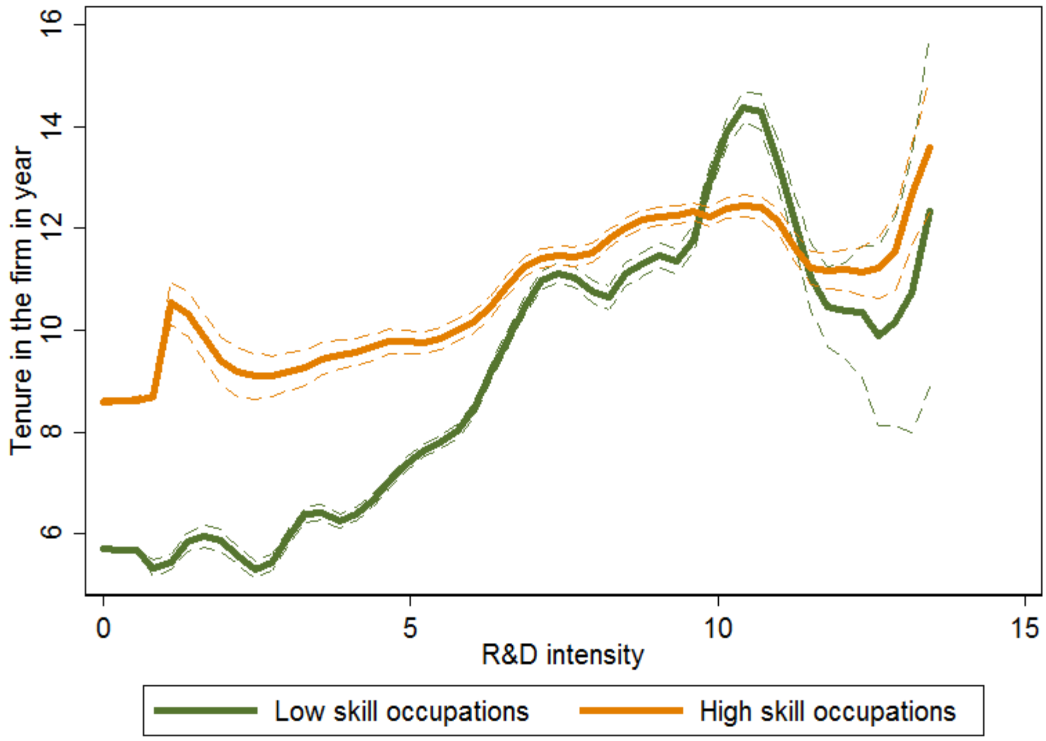
Our model predicts that workers in lower skilled occupations require more firm-specific training than workers in higher-skilled occupations, particularly in more innovative firms.[11] Hence our prediction that workers in low-skilled occupations should stay longer in more innovative firms. On the other hand, there should be a smaller effect for innovativeness on turnover of workers in high-skilled occupations. This is indeed what we see from Figure 5.

4.1.2 O*NET data

Our model links the fact that workers in low-skill occupations that are employed in more frontier (or higher R&D intensity) firms get a higher wage premium than workers in high-skill occupations, to the idea that workers in low-skilled occupations

[11]Note that in our model, workers in low-skill occupations 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 workers in low-skill occupations are more complementary to workers in high-skill occupations.

Figure 5: Tenure, by occupation and R&D intensity



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 workers in low-skilled occupations and the top line for workers in high-skilled occupations (see section A.2.3). 95% confident intervals are included.

become indispensable to the firm's success - i.e. more complementary with the firm's assets and workers in higher skilled occupations.

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 workers in low-skilled occupations are more complementary to other workers in high R&D intensive firms than in low R&D intensive firms.

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 4, we provide, for each skill level, the average values of the response in our sample across firms with three levels of R&D intensity compared to firms with no R&D. The values are standardized to be equal to 1 for no R&D firms, separately for each skill level. The consequences of a worker in a low-skilled occupation making an error are larger in a more R&D intensive firm than in a less R&D intensive firm. We see however that this pattern does not hold when R&D intensity is replaced by size (measured by total employment of the firm). The fact that R&D intensive firms do focus on tasks that are associated with a higher score for the "consequence of an error" is therefore not a size effect.

Table 4: Consequence of an error

	Tercile of R&D intensity				Quartile of employment			
	None (1)	Low (2)	Middle (3)	High (4)	Very Small (1)	Small (2)	Middle (3)	Large (4)
Skill level								
Low	1.00	1.01	1.11	1.14	1.00	0.98	0.94	0.85
Intermediate	1.00	1.00	1.02	1.03	1.00	1.01	1.01	1.02
High	1.00	1.02	1.00	0.99	1.00	1.00	1.01	1.02

Notes: R&D firms are split in three groups of equal size based on the value of their total 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 (resp. for the lowest quartile of firm employment) at each skill level.

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 average values of the response across firms with three levels of R&D intensity compared to firms with no R&D. The results are shown in Table 5 where the same standardization as in Table 4 is done. In particular, we see that the impact of decisions of a worker in a low-skilled occupation, is larger in a more R&D intensive firm than in a less R&D intensive firm. The difference is small, but yet it is statistically significant.

Table 5: Impact of decision

	Tercile of R&D intensity				Quartile of employment			
	None (1)	Low (2)	Middle (3)	High (4)	Very Small (1)	Small (2)	Middle (3)	Large (4)
Skill level								
Low	1.00	1.00	1.00	1.01	1.00	1.00	1.00	0.99
Intermediate	1.00	0.99	0.98	0.98	1.00	1.01	1.02	1.03
High	1.00	1.00	0.98	0.97	1.00	1.00	0.99	1.02

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 (resp. for the lowest quartile of firm employment) at each skill level.

Training

The last two questions spelled out above consider the duration of training workers in low-skill occupations receive across firms with different levels of R&D intensity. Table 6 shows that in the highest R&D intensive firms, from 14.3% to 16.2% of workers in low-skill occupations report having received training for more than one year, whereas only 6.4% to 7.2% of workers in low-skill occupations 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) workers in low-skill occupations 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) workers in low-skill occupations 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 6: On the job and on-site training for low skill occupations

	Tercile of R&D intensity				Quartile of employment			
	None (1)	Low (2)	Middle (3)	High (4)	Very Small (1)	Small (2)	Middle (3)	Large (4)
On-site or in-plant								
None	20.4	20.1	18.6	18.4	19.1	19.2	19.9	21.7
Up to 6 months	65.7	64.5	59.9	54.5	63.1	63.9	65.3	65.9
6 months - 1 year	7.6	8.2	10.8	12.8	9.2	8.8	8.1	6.9
A year or more	6.3	7.3	10.7	14.2	8.6	8.1	6.7	5.5
On-the-job								
None	10.2	10.1	9.4	9.1	10.5	10.3	10.4	9.4
Up to 6 months	74.9	73.1	66.3	60.0	69.5	70.8	73.6	78.2
6 months - 1 year	7.8	8.7	12.4	14.8	10.5	9.8	8.5	6.1
A year or more	7.1	8.2	11.9	16.1	9.6	9.0	7.6	6.2

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 workers in low-skilled occupations reporting having been trained for different durations whether on-site or on-the job.

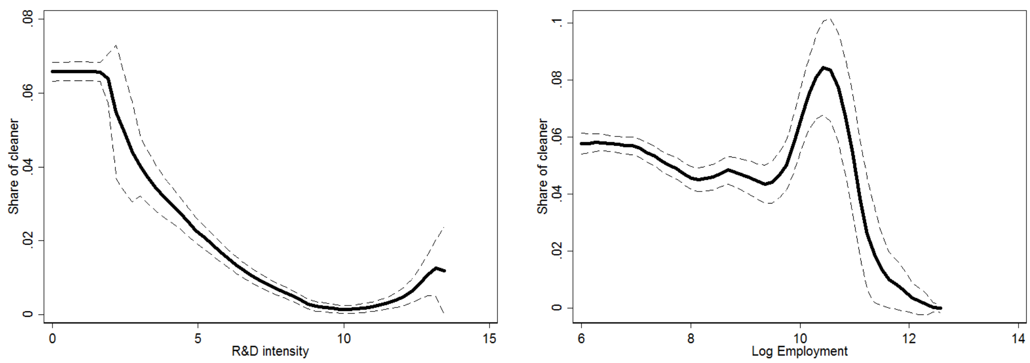
4.2 Outsourcing

Our model also predicts that more innovative firms tend to outsource low-skill tasks more than less innovative firms, and this is particularly true for low-skill occupation that have little complementarity (associated with a small λ in the model). The previous results using O*Net data have already shown that innovative firms put more weight in low-skill occupations that are associated with longer training and larger consequences in case of error. Unfortunately, it is not easy to directly measure outsourcing in our data for at least two reasons. First because outsourced workers do not necessary appear in the ASHE data, and even if they do, they won't be link to the firm that use their service. Second because we conjecture that most of the outsourcing occurred before 2004, which prevent us from following workers in low-skill occupations that are outsourced from innovative firms. We therefore proceed indirectly.

We start from the idea that all firms need the same share of cleaners which can be arguably seen as a low λ task. The only reason this share is lower than average in more innovative firms is because of the outsourcing of cleaners by those firms. In Figure 6, we plot the share of cleaners among all workers in low-skill occupations against R&D intensity in the left-hand side panel and against total employment in the right-hand side panel. This graph clearly shows that innovative firms outsource more their cleaners than non innovative firms, and here again, this is not a size effect.

Figure 6: Share of workers in low-skilled occupations that are cleaners

(a) by R&D (b) by employment



Notes: The y-axis shows the share of cleaners over the total number of workers in low-skill occupations.

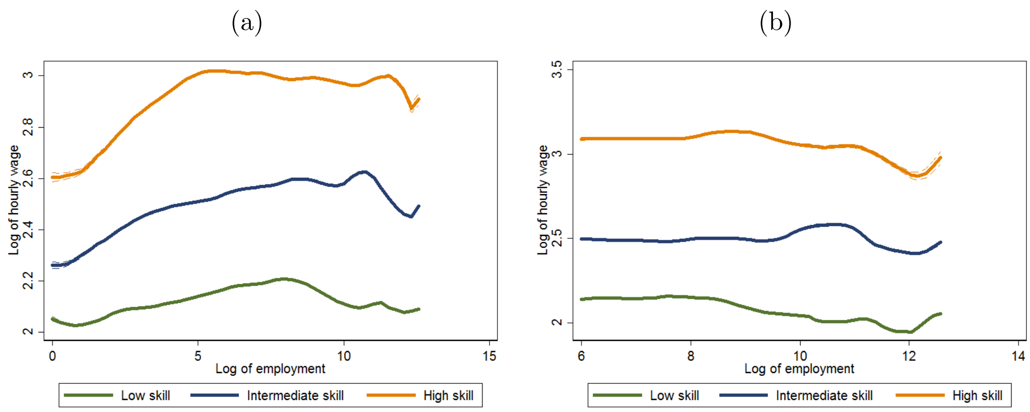
4.3 Robustness

In this section we show that our main results pass a number of robustness tests.

4.3.1 Firm size

Our empirical results estimate a negative elasticity of wage with respect to the size of the firm. However, the fact that larger firms pay higher wage is a well established fact in the labour literature (see among other Oi and Idson, 1999). This negative effect actually stems from the fact that we are focusing on large firms while the premium from working in a firm with more employees is essentially captured by relatively small firms as shown in Figure 7. Moreover, in Table 10 that we discuss further below, we see that when we estimate the effect of R&D on wage, using the whole ASHE sample (that is, without restricting to large firms), we find a positive and significant coefficient of the logarithm of total employment on wage.

Figure 7: Wages by firm size



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 workers in low-skilled occupations and the top line for workers in high-skilled occupations (see section A.2.3). 95% confident intervals are included.

4.3.2 Bonus income and other measures of wage or income

A first concern is that high-skilled workers may 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 workers in high-skilled occupations receive larger bonuses in more R&D intensive firms. In Table 7 we show that using average annual wages instead of average weekly wages and including or excluding incentive payments does not affect our results.

More generally, how are our results affected by the definition of income that we use? In our baseline results, we have chosen to use the wages measured in the week that the survey is collected. As explained in Appendix A.2.2, the numerator includes a fixed salary and additional variable earnings (incentive, overtime and other pay). Here, we test the sensitivity to our main result to using other measures of wages. Results are presented in Table 7 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 earnings from incentive paid at the end of the year and hence not well captured by our baseline measure of wages (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 earnings as incentive in innovative firms. In fact, the average share of bonus in annual earnings is 8.8% for non R&D firms against 6.5% for non R&D firms. Finally, comparing column 4 and 5 of Table 7 shows no substantial differences when bonus are included or excluded.

4.3.3 Different functions of R&D

Here we show that our main results hold using alternative function of R&D. Our baseline results use the logarithm of total R&D expenditure divided by total employment in the firm. Figure 2 shows that the relationship between the log of hourly wage and this function of R&D seems to be close to linear. Nevertheless, in this section, we see that our results hold when we consider different functional form of R&D, that can give different weight to different level of R&D intensity. Hence, in Table 8, we successively consider: $\frac{R\&D}{L}$, $ln(1+\frac{R\&D}{L})$, an hyperbolic function with R&D and with R&D per employee, $ln(1+R\&D)$, $R\&D > 0$ and $R\&D > 0$. In each case, the coefficient is positive and significant in the case of low-skill occupation workers consistently with our baseline model that is shown again in column 2.

Next, to allow for even more flexibility, we let the coefficient 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:

$$ln(w_{ijkft}) = x'_{ift}\beta_1 + z'_{ft}\beta_2 + \sum_{l=1}^{20} \beta_{3l} R_{ftl} + \nu_i + \nu_t + \epsilon_{it} \tag{4}$$

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 9, where the reference is the group of firm with no R&D. We see that the coefficients are positive and increase with the quantile of R&D for low skill occupations (column 1), is positive and significant for the highest quantiles in the case of intermediate skill occupations (column 2) and never significant in the case of high skill occupations (column 3). Column 4 shows that overall, innovation is associated with higher wages for most quantiles.

Table 7: Robustness to using different measures of wages

	Dependent variable: $ln(w_{ijkft})$				
Income	Total hourly pay	Fixed hourly pay	Total pay	Total pay (inc. incentive)	Fixed pay
	(1)	(2)	(3)	(4)	(5)
$ln(R_{ft}+1)$	0.002***	0.002***	0.003***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
* med skill	0.002***	0.002**	0.002**	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
* low-skill	0.006***	0.005***	0.007***	0.011***	0.011***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Age Sq.	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.007***	0.006***	0.005***	0.068***	0.066***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.003)
Tenure Sq.	-0.000***	-0.000***	-0.000	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.006***	-0.009***	-0.013***	-0.024***	-0.022***
	(0.002)	(0.001)	(0.003)	(0.005)	(0.005)
Full-Time	-0.004	0.009	0.658***	0.493***	0.489***
	(0.005)	(0.006)	(0.015)	(0.014)	(0.014)
low-skill	-0.136***	-0.114**	0.299	0.182	0.168
	(0.043)	(0.049)	(0.278)	(0.280)	(0.286)
med-skill	-0.052	-0.032	0.398	0.317	0.298
	(0.043)	(0.049)	(0.278)	(0.280)	(0.286)
high-skill	0.021	0.037	0.473*	0.377	0.356
	(0.044)	(0.050)	(0.278)	(0.281)	(0.287)
R^2	0.889	0.908	0.887	0.796	0.785

Notes: 626,210 observations. This table is similar to the last column of Table 3 but uses different measures of wages to construct the dependent variable. Column 1 uses the logarithm of total hourly earnings, column 2 uses the logarithm of the basic (fixed) hourly wages, 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 including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table 8: Testing different function of R&D

	Dependent variable: $ln(w_{ijkft})$							
R&D function	$\frac{x}{l}$ (1)	$log(1+\frac{x}{l})$ (2)	$H(x)$ (3)	$H\left(\frac{x}{l}\right)$ (4)	$log(1+x)$ (5)	$x>0$ (6)	x (7)	$log(\frac{x}{l})$ (8)
R&D intensity	0.000**	0.002***	0.001**	0.013***	0.001*	0.006	0.019	0.002
	(0.000)	(0.001)	(0.001)	(0.003)	(0.000)	(0.005)	(0.014)	(0.002)
* med skill	0.000*	0.002***	0.001**	0.010***	0.001**	0.011**	0.020**	0.002
	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.006)	(0.009)	(0.001)
* low-skill	0.001*	0.006***	0.003***	0.024***	0.002***	0.026***	0.072**	0.005***
	(0.000)	(0.001)	(0.001)	(0.003)	(0.001)	(0.008)	(0.031)	(0.002)
Age Sq.	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.008***	0.007***	0.007***	0.007***	0.007***	0.007***	0.008***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Tenure Sq.	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.006***	-0.006***	-0.007***	-0.006***	-0.007***	-0.007***	-0.006***	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Full-Time	-0.003	-0.004	-0.004	-0.004	-0.004	-0.003	-0.003	-0.080***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.023)
low-skill	-0.130***	-0.136***	-0.134***	-0.132***	-0.134***	-0.134***	-0.130***	-0.067***
	(0.039)	(0.043)	(0.042)	(0.040)	(0.042)	(0.042)	(0.039)	(0.007)
med-skill	-0.051	-0.052	-0.052	-0.049	-0.052	-0.052	-0.051	-0.038***
	(0.039)	(0.043)	(0.042)	(0.040)	(0.042)	(0.042)	(0.039)	(0.005)
high-skill	0.016	0.021	0.020	0.024	0.019	0.018	0.017	0.000
	(0.040)	(0.044)	(0.043)	(0.040)	(0.043)	(0.043)	(0.040)	(.)
R^2	0.889	0.889	0.889	0.889	0.889	0.889	0.889	0.917
Observations	626,210	626,210	626,210	626,210	626,210	626,210	626,210	162,696

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

Table 9: 20 quantiles of R&D based on level of total R&D expenditures

	Dependent variable: $ln(w_{ijkft})$			
Skill Category	Low (1)	Intermediate (2)	High (3)	All (4)
Quantile 1	0.004	-0.001	0.001	0.004
Quantile 2	0.017**	0.003	-0.007	0.010
Quantile 3	0.006	0.003	-0.001	0.002
Quantile 4	0.031***	-0.018	-0.008	0.012*
Quantile 5	0.036**	0.010	-0.000	0.023***
Quantile 6	0.036***	0.012	0.011	0.027***
Quantile 7	0.037***	0.009	-0.008	0.025***
Quantile 8	0.039***	0.014	0.000	0.031***
Quantile 9	0.044***	0.021*	-0.007	0.035***
Quantile 10	0.048***	0.021	-0.001	0.038***
Quantile 11	0.065***	0.029*	-0.006	0.053***
Quantile 12	0.070***	0.046***	-0.003	0.056***
Quantile 13	0.073***	0.029**	-0.013	0.051***
Quantile 14	0.073***	0.035***	0.012	0.064***
Quantile 15	0.061***	0.035***	0.012	0.064***
Quantile 16	0.096***	0.048***	-0.011	0.081***
Quantile 17	0.085***	0.022*	-0.003	0.071***
Quantile 18	0.090***	0.043***	0.007	0.082***
Quantile 19	0.114***	0.028**	-0.013	0.077***
Quantile 20	0.147***	0.020	-0.001	0.099***
R^2	0.774	0.851	0.885	0.887
Observations	407,341	104,318	114,535	626,210

Notes: This table presents the coefficient on each of the 20 quantiles of total R&D expenditure when estimating equation 4. The usual set of control variables are included but not reported. Ordinary Least Square regression including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed (but not reported) to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

4.3.4 Other measures of innovation

Our results are therefore robust to considering different functional form of R&D, but what happens when we change the definition of R&D expenditures? Table 10 shows how our results are affected compared to our baseline definition that use both intramural and extramural R&D expenditures (the baseline specification is reported in column 1). We hence estimate equation (1), allowing the coefficient on R&D and the intercept to vary across skill categories, and using different proxies for the intensity of R&D: first using only intramural R&D (column 2), then using only extramural (column 3) and then using the number of workers directly involve in R&D activities that we directly take from BERD (column 4). Results are consistent with our baseline model, that is, the effect is always stronger and significant for low-skill occupation workers. Finally, we measure R&D as the share of workers that correspond to a skill category 6 (PhD level scientific occupations, see Appendix A.2.3). This measures does not require any information from the BERD database and allows us to relax the restriction to firms of more than 400 employees. The results presented in column 5 is, here again, consistent with our baseline model.

4.3.5 Other robustness

We conclude by performing two additional robustness checks. First, as seen in Table A1 in Appendix A, 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 11.

Second, we test the robustness of our results regarding the different effects of R&D on wages by skill to using an alternative definition of skill level as defined in Appendix A.2.3. Results are robust in the sense that there is no effect of R&D expenditures on wages for high occupation workers as presented in Table 12 where each column corresponds to a different skill level (1 for the lowest and 4 for the highest).

Table 10: Robustness to using different measures of R&D.

	Dependent variable: $ln(w_{ijkft})$				
Measure of R&D	Baseline (1)	Only Intramural (2)	Only Extramural (3)	R&D workers (4)	Scientists (5)
ln R&D int	0.002***	0.002***	-0.000	0.009***	0.010
	(0.001)	(0.001)	(0.001)	(0.002)	(0.009)
* med skill	0.002***	0.002***	0.004***	0.002**	0.058***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.019)
* low-skill	0.006***	0.006***	0.008***	0.005***	0.153***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.020)
Age Sq.	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.007***	0.007***	0.007***	0.007***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure Sq.	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.006***	-0.006***	-0.006***	-0.006***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Full-Time	-0.004	-0.004	-0.004	-0.004	-0.005*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)
low-skill	-0.136***	-0.136***	-0.136***	-0.136***	-0.116***
	(0.043)	(0.043)	(0.043)	(0.043)	(0.011)
med-skill	-0.052	-0.052	-0.051	-0.052	-0.018
	(0.043)	(0.043)	(0.043)	(0.043)	(0.011)
high-skill	0.021	0.021	0.026	0.019	0.081***
	(0.044)	(0.044)	(0.043)	(0.044)	(0.011)
R^2	0.889	0.889	0.889	0.889	0.854
Observations	626,210	626,210	626,210	626,210	1,815,722

Notes: This table presents results from estimating the same model as in the last column of Table 3 but using different measure for R&D. 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, column 4 uses the share of workers involved in R&D activities taken from BERD and Column 5 uses the share of workers in occupation skill category 6 using the whole ASHE database. 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 including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table 11: Robustness: Removing firms from the highest quantile of R&D expenditures.

	Dependent variable: $ln(w_{ijkft})$			
Skill Category	Low (1)	Intermediate (2)	High (3)	All (4)
ln R&D int	0.007***	0.003***	-0.000	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
* med skill				0.002***
				(0.001)
* low-skill				0.006***
				(0.001)
Age Sq.	-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***
	(0.001)	(0.001)	(0.001)	(0.000)
Tenure Sq.	-0.000***	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.005**	0.002	0.003	-0.006***
	(0.002)	(0.003)	(0.003)	(0.002)
Full-Time	-0.011*	-0.089***	-0.111***	-0.004
	(0.006)	(0.015)	(0.014)	(0.005)
low-skill				-0.135***
				(0.043)
med-skill				-0.051
				(0.043)
high-skill				0.022
				(0.043)
R^2	0.772	0.850	0.885	0.888
Observations	405,331	102,733	110,444	618,524

Notes: This table shows results from estimating the same model as in Table 3 but removing firms belonging to the highest quantile (out of 20) in terms of R&D intensity. Ordinary Least Square regression including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

Table 12: Robustness: Alternative measure of skill

	Dependent variable: $ln(w_{ijkft})$				
Skill Category	1 (low) (1)	2 (2)	3 (3)	4 (high) (4)	All (5)
ln R&D int	0.005*** (0.001)	0.007*** (0.001)	0.002** (0.001)	-0.000 (0.001)	0.003*** (0.001)
* med-high skill					0.002** (0.001)
* med-low skill					0.005*** (0.001)
* low-skill					0.004*** (0.001)
Age Sq.	-0.000*** (0.000)	-0.000*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Tenure	0.007*** (0.001)	0.009*** (0.001)	0.004*** (0.001)	0.002*** (0.001)	0.007*** (0.000)
Tenure Sq.	-0.000*** (0.000)	-0.000*** (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)
Firm Size	0.003 (0.003)	-0.007*** (0.003)	0.000 (0.002)	0.004 (0.003)	-0.006*** (0.002)
Full-Time	-0.038*** (0.006)	-0.014** (0.007)	-0.115*** (0.014)	-0.110*** (0.014)	-0.006 (0.005)
low-skill					-0.170*** (0.006)
med-low-skill					-0.143*** (0.006)
med-high-skill					-0.049*** (0.004)
high-skill					0.000 (.)
R^2	0.706	0.782	0.872	0.901	0.889
Observations	103,136	293,543	113,802	115,729	626,210

Notes: This table shows results from estimating the same model as in Table 3 but defined skill level of occupations differently. Ordinary Least Square regression including additive individual and year fixed effects. Heteroskedasticity robust standard errors clustered at the firm level are computed to indicate the level of significance: ***, ** and * for 0.01, 0.05 and 0.1 levels of significance.

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, J. M., F. Kramarz, and D. N. Margolis (1999, March). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–334.

Acemoglu, D. (2002, mar). Technical change, inequality, and the labor market. *Journal of Economic Literature 40*(1), 7–72.

Acemoglu, D., P. Aghion, C. Lelarge, J. Van Reenen, and F. Zilibotti (2007). Technology, information, and the decentralization of the firm. *The Quarterly Journal of Economics 122*(4), 1759–1799.

Acemoglu, D. and D. H. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Number 4 in Orley Ashenfelter and David E. Card (eds.),, pp. 1043–1171. Amsterdam: Elsevier.

Aghion, P., U. Akcigit, A. Bergeaud, R. Blundell, and D. Hémous (2015, jun). Innovation and top income inequality. Technical report.

Aghion, P., U. Akcigit, A. Hyytinen, and O. Toivanen (2018). On the returns to invention within firms: Evidence from Finland. Forthocoming, AEA Papers and Proceedings.

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

Akcigit, U., J. Grigsby, and T. Nicholas (2017, feb). Immigration and the rise of american ingenuity. Technical report.

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2016). Its where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics 34*(S2), S67–S97.

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

Bell, B. D. and J. Van Reenen (2014, 02). Bankers and Their Bonuses. *Economic Journal 124*(574), F1–F21.

Bloom, N., L. Garicano, R. Sadun, and J. V. Reenen (2014, December). The Distinct Effects of Information Technology and Communication Technology on Firm Organization. *Management Science 60*(12), 2859–2885.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica 74*(2), 323–364.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2016, November). Firms and Labor Market Inequality: Evidence and Some Theory. NBER Working Papers 22850, National Bureau of Economic Research, Inc.

Caroli, E. and J. Van Reenen (2001). Skill-biased organizational change? evidence from a panel of british and french establishments. *The Quarterly Journal of Economics 116*(4), 1449–1492.

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

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

Elias, P. and K. Purcell (2013). Classifying graduate occupation for the knowledge society. FutureTrack Working Paper 5.

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

Garicano, L. and E. Rossi-Hansberg (2006). Organization and inequality in a knowledge economy. *The Quarterly Journal of Economics 121*(4), 1383–1435.

Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies 59*(3), 515–535.

Goldin, C. and L. F. Katz (2010). *The Race between Education and Technology*. Belknap Press.

Goldschmidt, D. and J. F. Schmieder (2017). The rise of domestic outsourcing and the evolution of the german wage structure. *The Quarterly Journal of Economics*, qjx008.

Groshen, E. L. (1991). Sources of intra-industry wage dispersion: How much do employers matter? *The Quarterly Journal of Economics 106*(3), 869–884.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2017). Who profits from patents? rent-sharing at innovative firms.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics 108*(3), 551–575.

Kremer, M. and E. Maskin (1996, August). Wage inequality and segregation by skill. Working Paper 5718, National Bureau of Economic Research.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Office for National Statistics (2016a). Annual survey of hours and earnings, 1997-2015: Secure access. 8th edition. Technical Report SN:6689, UK Data Service.

Office for National Statistics (2016b). Business expenditure on research and development, 1994-2014: Secure access. 5th edition. Technical Report SN:6690, UK Data Service.

Oi, W. Y. and T. L. Idson (1999). Firm size and wages. *Handbook of labor economics 3*, 2165–2214.

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

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. v. Wachter (2015). Firming up inequality.

Stole, L. and J. Zwiebel (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies 63*(3), 375–410.

Van Reenen, J. (1996). The creation and capture of rents: Wages and innovation in a panel of u. k. companies. *The Quarterly Journal of Economics 111*(1), 195–226.

A Data construction and additional description

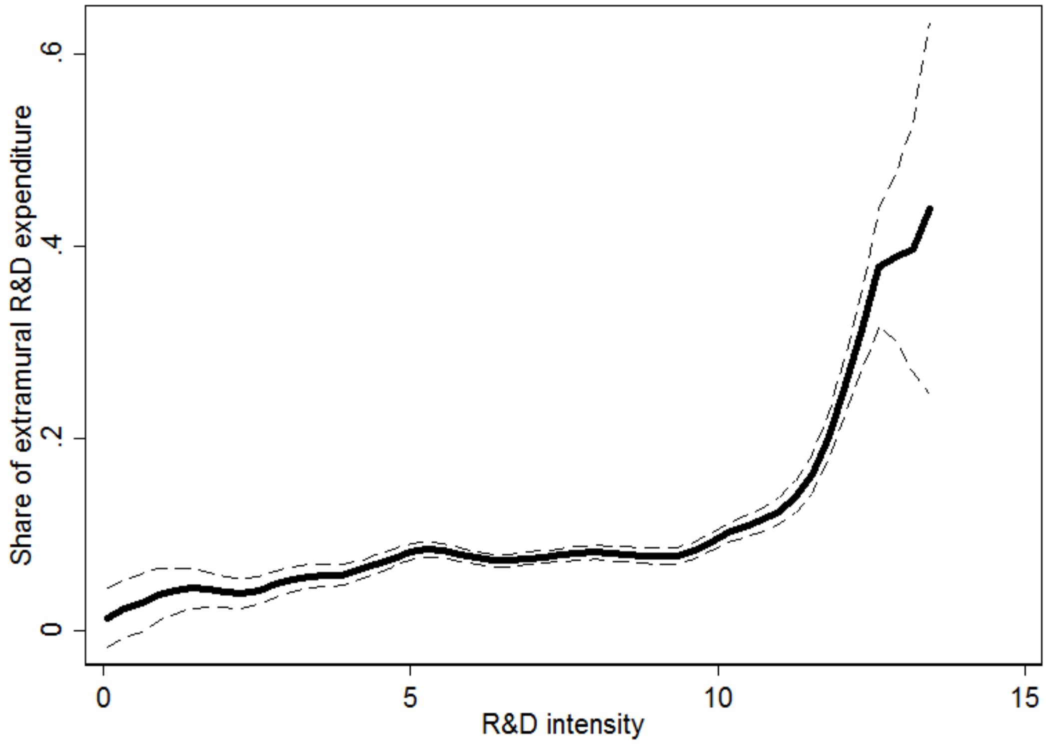
This appendix describes the construction of our main sample which results from the merge of two datasets provided by the ONS: the Annual Survey of Hours and Earnings (ASHE) and the Business Expenditures on Research and Development (BERD).

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.

Figure A1: Share of total R&D expenditures that is outsourced (extramural) against R&D intensity.



Notes: Source: BERD. R&D intensity is defined as the logarithm of total R&D expenditures divided by employment. Dashed lines correspond to the upper and lower bound of a 95% confident interval.

Table A1 reports the average amount of intramural and extramural R&D across 20 quantiles of the distribution of total R&D intensity.[12] The distributions of both intra and extramural R&D are highly skewed, in particular, firms in the highest vintile are very different from others.

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

Table A1: Distribution of employment and R&D

Quantile of R&D	Employment	Intramural R&D	Extramural R&D	Number of firm-years
0 (no R&D)	2,365	0	0	29,848
1	8,337	76	5	434
2	4,634	233	14	427
3	3,129	298	22	426
4	2,775	382	62	429
5	2,890	653	80	427
6	1,702	541	56	427
7	1,916	808	66	427
8	1,761	1,038	88	427
9	1,344	1,019	110	427
10	1,700	1,618	211	426
11	1,732	2,138	293	431
12	1,958	3,177	557	427
13	1,625	3,486	417	427
14	1,439	4,345	390	428
15	1,627	6,793	488	424
16	2,499	16,306	799	429
17	2,635	24,618	1,358	429
18	2,290	34,882	2,891	426
19	2,480	62,946	10,027	427
20	2,282	140,744	81,145	422

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.

Our measure of R&D intensity is the logarithm of total R&D expenditure (including both intramural and extramural R&D) divided by the number of employee in the firm. R&D expenditure, as well as total employment, are defined at the firm level. In practice, we use $ln(1 + R_{ft})$, where R_{ft} is the ratio of total R&D expenditure over total employment, 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.

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 2015.[13] 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.[14]

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 Wages

There are various measures of wages 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 "incentive pay" includes profit sharing, productivity, performance and other bonus or incentive pay, piecework and commission.). The number of hours worked are reported,

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

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

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 wages as the ratio of gross weekly wage divided by total number of paid hours (including overtimes). This is the measure of wage we will consider as a baseline but we also show descriptive statistics for gross annual earnings. Including other types of earnings and incentive payments is arguably relevant especially in the case of high income individuals 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 4.3.2.[15]

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).[16] 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).[17]

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 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

[15]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 wage distribution, and decreases with wage for the remaining top quarter.

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

[17]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-skill	
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	371,613	30.1	60	49	37.3	6.2
Skill cat 2	39,165	35.4	83	68	39.1	8.2
Intermediate-skill						
Skill cat 3	77,847	36	88	60	39.1	9.4
Skill cat 4	27,115	36.4	93	60	39.6	9.1
High-skill						
Skill cat 5	111,539	36.4	94	70	40.7	9.8
Skill cat 6	3,475	35.8	92	61	39.3	10.4

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

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.

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.[18] 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.

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

Table A4: Pay by skill categories

Skill	Hourly pay	Weekly pay	% incentive	% overtime	Annual earnings
Low-skill					
Skill cat 1	8.64	286	2.54	5.64	13,612
Skill cat 2	11.59	446	2.25	5.32	21,970
Intermediate-skill					
Skill cat 3	13.59	507	5.21	3.56	25,936
Skill cat 4	16.83	625	5.21	2.13	32,820
High-skill					
Skill cat 5	25.62	938	7.64	1.42	54,075
Skill cat 6	22.39	810	6.33	1.11	43,868

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

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 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.

Table A5: Construction of the sample

ASHE	Observations	Individuals	Mean wage	Sd wage
Raw ASHE	2,023,886	355,546	13.1	41.5
Clean ASHE	1,818,812	336,794	13.3	15
Private Corporations	1,074,318	243,010	12.9	17.3
Final Sample	626,722	156,966	12.9	18.7
BERD	Observations	Firms	% intramural R&D	% extramural R&D
Raw BERD	247,468	53,394	100	100
400+ Employees	9,013	1,917	75	85
Final Sample	8,562	1,884	66	80

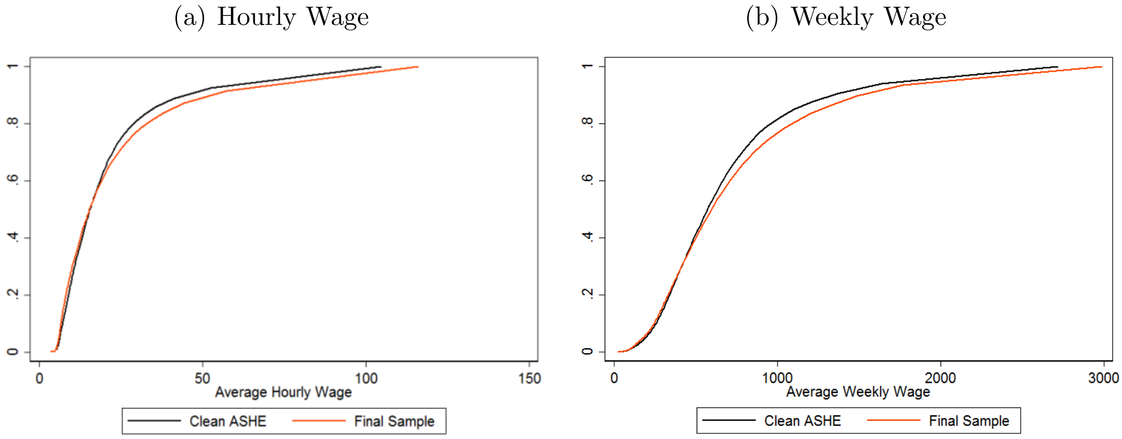
Notes: This table presents the evolution of the two databases ASHE and BERD across the successive steps conducted to match them. **ASHE**: Raw data corresponds to the standard ASHE database 2004-2015. Clean ASHE corresponds to the database "Cleaned ASHE" as described in subsection A.2.1. Private corporation 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 2004-2015. 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.

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	2394	670
2005	91	2365	808
2006	91	2329	956
2007	102	2363	757
2008	96	2400	628
2009	75	2360	798
2010	86	2310	696
2011	97	2360	708
2012	98	2419	781
2013	109	2464	800
2014	109	2592	844
2015	123	2674	890

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



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.

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 (59%)[19], 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 workers in high-skilled occupations.

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

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 (£)	14.1	13.5	6.5	7.5	10.3	16.2	25.5	61
Weekly Wage (£)	501	476	121	231	388	632	962	2202
Weekly Incentive Pay (£)	8.4	81.1	0	0	0	0	0	182.7
Weekly Overtime Pay (£)	20.1	60.9	0	0	0	3.6	62.7	292.9
Total Annual Wage (£)	26,173	44,143	4,563	9,776	19,069	32,069	49,652	135,958
Basic Paid Hours	35.8	11.8	16.8	30.8	40.6	42.8	44.2	57.8
Paid Overtime	1.7	4.5	0	0	0	0.4	5.8	22
Age	42.2	13.5	24.2	30.8	41.8	52.8	61.6	69.3
Tenure	8.2	8.9	1.1	2.2	4.4	11	20.9	39.6

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.

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)	63.7	5.6	11.7	3.9	14.8	0.3	467,207
1	66	7.7	9.9	2.8	13.4	0.2	23,168
2	63.8	7.7	10.1	3.3	14.6	0.5	13,708
3	59.7	8.2	11.5	3.8	16.2	0.5	9,983
4	59.6	5.5	14.4	3.3	16.6	0.6	8,380
5	60.6	4.9	14	3.3	16.9	0.3	9,478
6	53.7	6.2	14.8	4.3	20.2	0.9	5,613
7	53.6	8.8	11.2	4.9	20.9	0.6	6,328
8	46.9	7.9	16.8	5.5	22.2	0.7	5,730
9	51.6	8.4	11.2	4.4	23.7	0.7	4,486
10	43.1	9.2	13.1	5.4	28.3	0.9	5,755
11	36	10.1	15.1	5.8	32.2	0.7	6,090
12	37.1	8.8	15.1	6.1	32	0.9	6,612
13	34.3	8.3	15.2	6	35.4	0.8	5,473
14	31.2	9.4	13.4	6.7	38.3	1.1	4,862
15	30.7	8	18.2	8.9	32.8	1.3	5,270
16	22.9	7.8	20.4	10.1	37.4	1.3	8,442
17	21.9	6.1	20.4	11.9	38.7	0.9	9,291
18	25.1	7.8	18.1	9	38	1.9	7,858
19	22.6	13.2	15.7	6.3	39.8	2.3	9,306
20	20.1	6	13.9	6.7	41.6	11.7	7,714

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

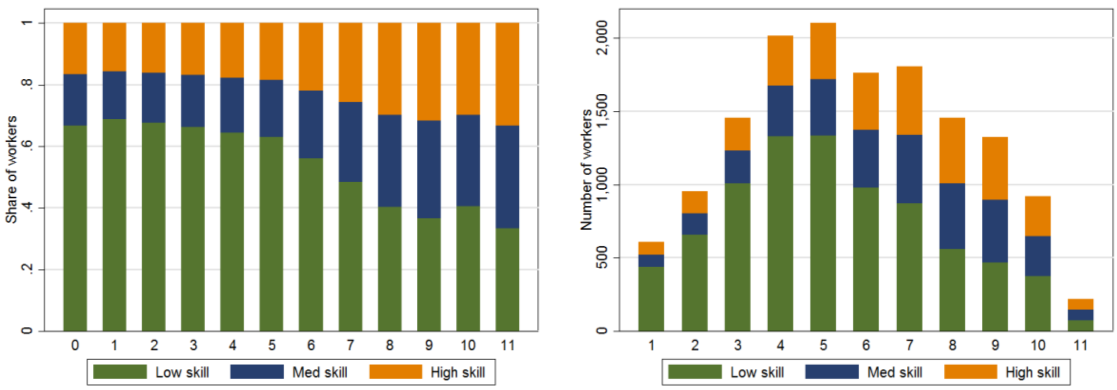
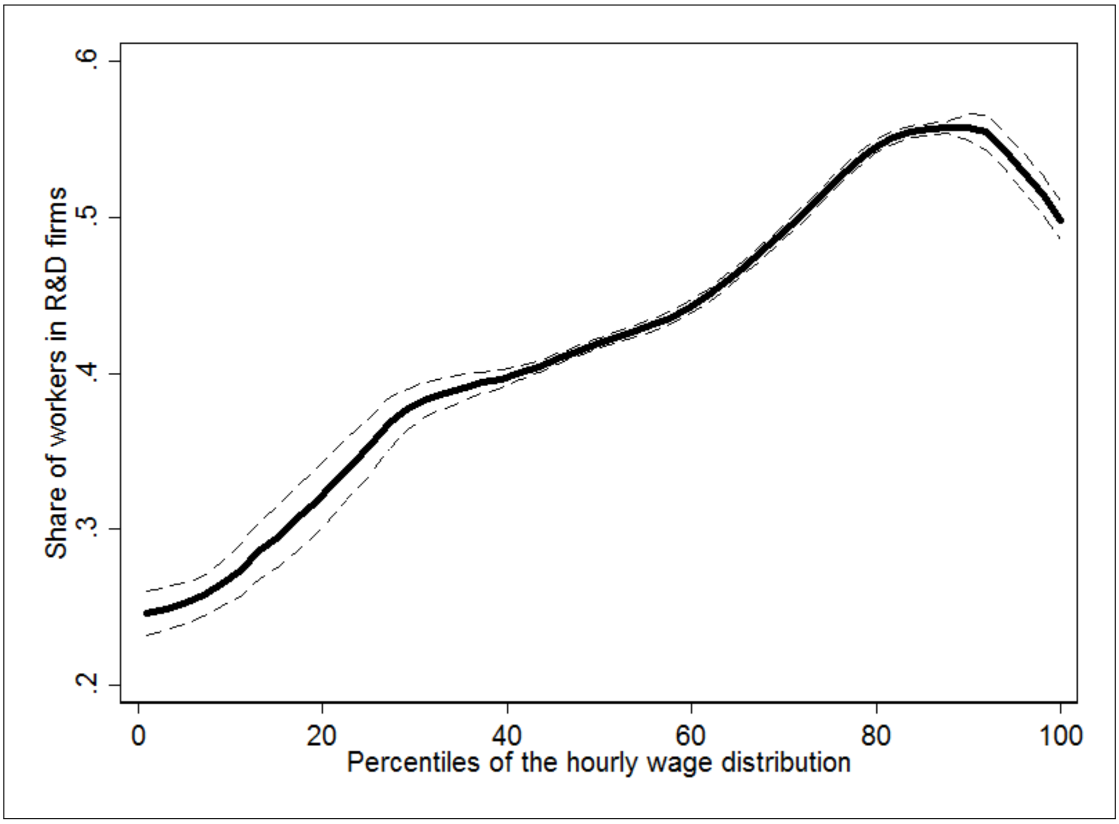


Figure A4 shows that the share of workers that work in a firm that does any R&D increases from just over 20% for workers at the bottom of the wage distribution, to over 55% after the 80^{th} percentile of the distribution where it plateaus. The share falls right at the top, where workers in the (low innovative) financial sector are heavily represented. This effect holds within innovative firms.

Figure A4: 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 2015 are considered independently.

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, 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[20] 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

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

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.

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 wages 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 Theoretical Appendix

C.1 Proof of Proposition 1

To measure the complete effect of innovation, let us consider how equilibrium wages react to a continuous increase in z (hence in λ_z wich corresponds to an upward shift of $\mathbb{E}_\phi[\lambda]$). For notation simplicity, let us consider that $z_{max} = 1$ which implies that $\lambda_z = z$.

We have:

$$\begin{aligned}
\frac{dw_q(z)}{dz} = \frac{dw_q(z)}{d\lambda_z} &= \frac{1}{4C}\left[\left(\overline{Q}-1\right)\left(Q_L-1\right)\lambda_z + \frac{\overline{Q}+Q_L-2}{2}\right] \\
&= \frac{q(\lambda_z)-q_L}{2}(\overline{Q}-1) + \frac{dq(\lambda_z)}{dz}\frac{Q-1}{2}\lambda_z
\end{aligned}$$

and:

$$\begin{aligned}
\frac{dw_Q(z)}{dz} = \frac{dw_Q(z)}{d\lambda_z} &= \frac{(\overline{Q}-Q_L)}{2}\left[(q_L-1) + \frac{\lambda_z}{2C}(Q_L-1) + \frac{1}{4C}\right] \\
&= \frac{\overline{Q}-Q_L}{2}\left[(q-1) + \lambda_z\frac{dq(\lambda_z)}{dz}\right]
\end{aligned}$$

The inequality

$$\frac{dw_q(z)}{dz} > \frac{dw_Q(z)}{dz}$$

then results from the fact that $\forall z$:

1. $(q(\lambda_z)-q_L)(\overline{Q}-1) > (\overline{Q}-Q_L)(q(\lambda_z)-1)$;
2. $\frac{dq(\lambda_z)}{dz} = \frac{dq(\lambda_z)}{d\lambda_z} > 0$;
3. $(\overline{Q}-1) > (\overline{Q}-Q_L)$.

C.2 Proof of Proposition 2

We have:

$$\frac{dw_q(\lambda,z)}{dz} = \left[\lambda(\overline{Q}-1)+1\right]\left(\frac{d\phi(\lambda,z)}{dz}(q(\lambda,z)-q_L) + \phi(\lambda,z)\frac{dq(\lambda,z)}{dz}\right)$$

and

$$\frac{dw_Q(z)}{dz} = \left(\overline{Q}-Q_L\right)\int_0^1\left(\left[\lambda(q(\lambda,z)-1)+1\right]\frac{d\phi(\lambda,z)}{dz} + \phi(\lambda,z)\lambda\frac{dq(\lambda,z)}{dz}\right)d\lambda$$

The proposition then follows from the following facts:

1. $\frac{d\phi(\lambda,z)}{\lambda} > 0$ and $\frac{d\phi(\lambda,z)}{dz} > 0$, which imply that $\frac{dq(\lambda,z)}{dz} > 0$
2. $(\overline{Q} - Q_L) < (q - q_L)$ and $\overline{Q} > \overline{q}$.

C.2.1 Proof of Proposition 3

We start from the maximization problem (recall that the optimal value of Q is $\overline{Q}$:

$$\max_{\vec{q}} \left(\tilde{\Pi}(\vec{q}) - \int_0^1 C(\lambda) \left(q(\lambda) - q_L \right)^2 d\lambda \right) \text{ s.t. } \int_0^1 \left(q(\lambda) - q_L \right) d\lambda \leq T,$$

If ν denotes the Lagrange multiplier associated with the time constraint, then the optimal value of $q(\lambda)$ is:

$$q^*(\lambda) = q_L + \max\{0, \phi(\lambda, z) \frac{\lambda(Q_L - 1) + 1}{4C} - \frac{\nu}{2C}\}.$$

Let

$$g(\lambda) = \phi(\lambda, z) \frac{\lambda(Q_L - 1) + 1}{4C}.$$

The function g is clearly increasing in λ and $g(0) = \frac{\phi(0,z)}{4C}$ and $g(1) = \frac{\phi(1,z)Q_L}{4C}$. Then for each value of z, there exists a cutoff value $\bar{\lambda}$ such that when $\lambda \leq \bar{\lambda}$, then $q^*(\lambda) = q_L$ and the firm outsources this task. This cutoff value is simply determined by:[21]

$$g(\bar{\lambda}) = \frac{\nu}{2C} \Longrightarrow \phi(\bar{\lambda}, z) \frac{\bar{\lambda}(Q_L - 1) + 1}{2} = \nu.$$

To determine the value of ν, we use the fact that the constraint is binding, so we must have:

$$T = q_L + \int_{\bar{\lambda}}^1 g(\lambda) d\lambda - \frac{\nu(1 - \bar{\lambda})}{2C} = q_L + \int_{\bar{\lambda}}^1 g(\lambda) d\lambda - g(\bar{\lambda})(1 - \bar{\lambda}).$$

From here, we will assume that $\phi(\lambda, z) = (1 + z)\lambda^z$ and $z \in \mathbb{N}$. This implies that:

$$\int_{\bar{\lambda}}^1 g(\lambda) d\lambda = \frac{z+1}{z+2} \frac{(Q_L - 1)}{4C} (1 - \bar{\lambda}^{z+2}) + \frac{(1 - \bar{\lambda}^{z+1})}{4C}$$

we get $\bar{\lambda}$ to be the solution of the equation:

$$4C(T - q_L) = \frac{z+1}{z+2}(Q_L - 1)(1 - \bar{\lambda}^{z+2}) + (1 - \bar{\lambda}^{z+1}) - \bar{\lambda}^z(1 - \bar{\lambda})(z + 1) \left(\bar{\lambda}(Q_L - 1) + 1 \right) \quad \text{(C1)}$$

[21] There is always one and only one value of $\bar{\lambda}$ for each value of z. However, this value is not necessarily bound to the $[0, 1]$ interval. If $\bar{\lambda} < 0$, then we shall consider that there is no outsourcing.

We want to show that $\bar{\lambda}$ increases with z.

Example 3. *Consider the case of two firms A and B. Firm A is not innovative: $z = 0$ whereas Firm B is innovative with $z = 1$.*

Hence in firm A the outsourcing equation (C1) yields:

$$\bar{\lambda}_A = 1 - \sqrt{\frac{8C(T - q_L)}{Q_L - 1}}$$

whereas in Firm B the outsourcing equation (C1) yields a $\bar{\lambda}_B$ which satisfies:

$$4C(T - q_L) = (1 - \bar{\lambda})^2 \left(1 + \frac{2(Q_L - 1)}{3}(2\bar{\lambda} + 1)\right)$$

Since for all $\lambda \in [0, 1]$, we have

$$1 + \frac{2(Q_L - 1)}{3}(2\lambda + 1) > \frac{Q_L - 1}{2},$$

then we must have $(1 - \bar{\lambda}_A)^2 > (1 - \bar{\lambda}_B)^2$ which implies $\bar{\lambda}_B > \bar{\lambda}_A$.

Note that a necessary condition for A to outsource is that:

$$q_L < \left(T - \frac{Q_L - 1}{8C}\right)$$

and similarly for B:

$$q_L < \left(T - \frac{1 + 2Q_L}{12C}\right) < \left(T - \frac{Q_L - 1}{8C}\right)$$

In other words, if the outside quality q_L is too low then firms won't outsource. The propensity to outsource also increases with the training cost and with the tightness of the capacity constraint is tight inversely measured by T.

More generally, and as long as $z \in \mathbb{N}$, the outsourcing condition (C1) becomes:

$$4C(T - q_L) = (1 - \bar{\lambda})^2 \underbrace{\left[\frac{z+1}{z+2}(Q_L - 1)\sum_{i=0}^{z+1} i\bar{\lambda}^{i-1} + \sum_{i=0}^{z} i\bar{\lambda}^{i-1}\right]}_{u_z(\bar{\lambda})}$$

where $u_z(\bar{\lambda})$ is increasing in z and always positive when $\bar{\lambda} \in [0, 1]$. This ensures that $\bar{\lambda}$ is increasing with z, which completes the proof.

C.3 Extension to multiple workers in the same task

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^q, then the remaining $n-1$ low-occupation workers renegotiate a wage w_{n-1}^q. By induction, this provides a generic expression for the two equilibrium wages $w_{n,m}^q(Q, q, \lambda)$ and $w_{n,m}^Q(Q, \overline{q})$ (up to a constant in q, Q and λ):

$$\begin{aligned} w_{n,m}^q(Q,q,\lambda) = &\ \lambda\phi(\lambda)\frac{q(\lambda)-q_L}{n(n+1)}\sum_{i=0}^{n} iQ^m q(\lambda)^{i-1} + \frac{(1-\lambda)}{2}(q(\lambda)-q_L)\phi(\lambda) \\ w_{n,m}^Q(Q,\overline{q}) = &\ \frac{Q-Q_L}{m(m+1)}\sum_{i=0}^{m} i\int_0^1 q(\lambda)^n Q^{i-1}\lambda\phi(\lambda)d\lambda + \frac{1-\mathbb{E}_\phi\left[\lambda\right]}{2}(Q-Q_L), \end{aligned}$$

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.

Let us now assume that we are in the toy case, that is, $\phi(\lambda) = 1$ if $\lambda = \lambda_z \equiv \frac{z}{z_{max}}$ and 0 otherwise. In that case:

$$\begin{aligned} w_{n,m}^q(Q,q,\lambda_z) = &\ \lambda_z\frac{q-q_L}{n(n+1)}\sum_{i=0}^{n} iQ^m q^{i-1} + \frac{1-\lambda_z}{2}(q-q_L) \\ w_{n,m}^Q(Q,\overline{q},\lambda_z) = &\ \lambda_z\frac{Q-Q_L}{m(m+1)}\sum_{i=0}^{m} iq^n Q^{i-1} + \frac{1-\lambda_z}{2}(Q-Q_L), \end{aligned} \tag{C2}$$

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

In this case, we have:

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

and since $Q > q$ implies $q(1+2Q) < Q(Q+2Q)$, then $\frac{q(1+2Q)}{3} - 1 < Q^2 - 1$, which, combined with the assumption that $(Q - Q_L) < (q - q_L)$, immediately implies that:

$$\frac{\partial w^q_{1,2}(Q, q, \lambda_z)}{\partial \lambda_z} > \frac{\partial w^Q_{1,2}(Q, q, \lambda_z)}{\partial \lambda_z}.$$

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

In this case, we have:

$$\frac{\partial w^q_{2,1}(Q, q, \lambda_z)}{\partial \lambda_z} = \frac{(q - q_L)(Q + 2qQ)}{6} - \frac{q - q_L}{2} \text{ and } \frac{\partial w^Q_{2,1}(Q, q, \lambda_z)}{\partial \lambda_z} = \frac{(Q - Q_L)(q - 1)}{2},$$

Then a sufficient condition for $\frac{\partial w^q_{2,1}(Q,q,\lambda_z)}{\partial \lambda_z} > \frac{\partial w^Q_{2,1}(Q,q,\lambda_z)}{\partial \lambda_z}$ 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 $\frac{\partial w^q_{n,n}(Q,q,\lambda_z)}{\partial \lambda_z} > \frac{\partial w^Q_{n,n}(Q,q,\lambda_z)}{\partial \lambda_z}$ 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 $\frac{\partial w^q_{n,m}(Q,q,\lambda_z)}{\partial \lambda_z} > \frac{\partial w^Q_{n,m}(Q,q,\lambda_z)}{\partial \lambda_z}$, 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

$$\frac{\partial w^q_{n,m}(Q, q, \lambda_z)}{\partial \lambda_z} > \frac{\partial w^Q_{n,m+1}(Q, q, \lambda_z)}{\partial \lambda_z}.$$