WORKING paper

BANQUE DE FRANCE
EUROSYSTÈME

The Innovation Premium to Soft Skills in Low-Skilled Occupations

Philippe Aghion[1], Antonin Bergeaud,[2] Richard Blundell[3] and Rachel Griffith[4]

November 2019, WP #739

ABSTRACT

Matched employee-employer data from the UK are used to analyze the wage premium to working in an innovative firm. We find that firms that are more R&D intensive pay higher wages on average, and this is particularly true for workers in some low-skilled occupations. We propose a model in which a firm's innovativeness is reflected in the degree of complementarity between workers in low-skill and high-skilled occupations, and in which non-verifiable soft skills are an important determinant of the wages of workers in low-skilled occupations. The model yields additional predictions on training, tenure and outsourcing which we also find support for in data.

Keywords: Innovation, Skill-biased Technological Change, Wage, Complementarity

JEL classification: O33, L23, J31

[1] Collège de France and LSE
[2] Banque de France, DGSEI/DEMS/SEPS and CEP
[3] UCL and IFS
[4] University of Manchester and IFS

NON-TECHNICAL SUMMARY

Using data from the United Kingdom, we show that as a result of technological change, there is an increased demand for technical skills and therefore for skilled workers. However, this effect benefits all skilled workers, regardless of their occupation or the company that employs them. Conversely, for low-skilled or unskilled workers, whether an innovative or non-innovative company employs them can make a significant difference. Figure 1a therefore compares the average hourly wages of firms that invest in R&D with those that do not, for occupations that require a minimum level of education, and Figure 1b for those that require advanced training or qualifications. For the former, the wage gap persists throughout working life, with a premium for those working in innovative companies, while no such difference seems to exist in the case of the latter.

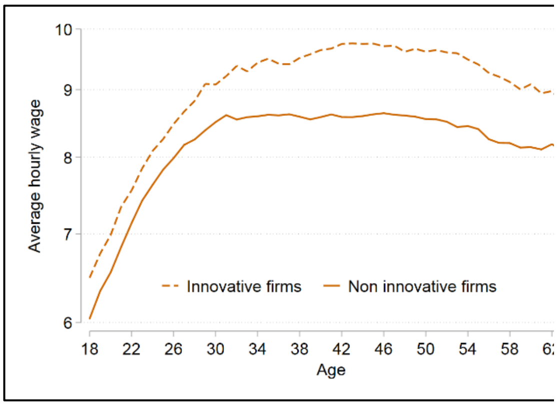


Figure 1a: Hourly wage (in log) at each age for low educated workers in innovative (resp non innovative) firms

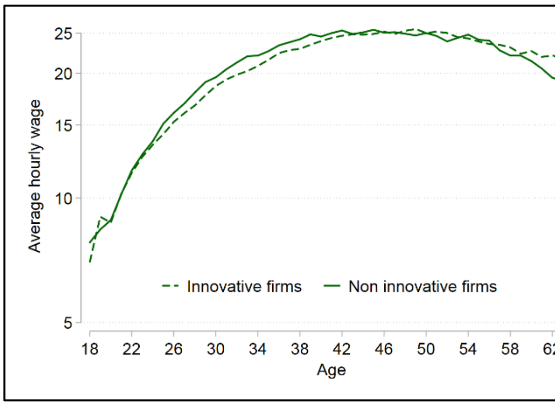


Figure 1b : Hourly wage (in log) at each age for highly educated workers in innovative (resp non innovative) firms

Based on these findings, the study shows that, even in a very advanced technological environment, the most innovative companies still value certain low-skilled tasks and therefore pay higher wages than a less innovative company would for these professions. We develop a model to rationalize this idea. Namely, the model is constructed around the view that all firms value qualified workers (usually managers, engineers, etc.) on the basis of their technical skills and reputation acquired during their careers. To some extent, these characteristics are observable and verifiable, for example by reading a CV. A company can therefore replace a skilled worker with another worker who theoretically possesses comparable skills with a relatively low risk of error. Conversely, the most innovative and technologically advanced companies tend to value certain skills of their less skilled employees more than other firms.

These firms generally have a flatter than average organizational hierarchy, which translates into increased complementarity between different workers, especially between those in low-skilled occupations and those performing more complex and technical tasks. Therefore, in these types of structures, it is extremely risky to employ people who regularly make mistakes. These companies have therefore developed a critical need for the skills of their less educated workers, such as initiative and reliability. These "soft-skills" are not normally recognized with a qualification and are therefore difficult to observe and, potentially, difficult to replace. More innovative companies are therefore willing to pay a wage premium to their employees and invest more in training their workers.

The model delivers a number of predictions that we then takes to the data. Namely: (i) workers in low-skilled occupations get a positive and sizeable premium from working in a more innovative firm; (ii) workers in low-skilled occupations exhibit on average a higher degree of complementarity with the firm's other assets (and in particular high skill workers) in more innovative firms compared to in less innovative firms; (iii) the wage premium to working in a more innovative firm for workers in low-skilled occupations increases with the complementarity between their quality and those of other workers; (iv) workers in low-skilled occupations have longer tenure in more innovative firms than in less innovative firms as more time will be spent by more innovative firms to enhance (or learn about) these workers soft skills; (v) a more innovative firm will invest more in training its workers in low-skilled occupations to increase their level of soft skills than a non-innovative firm; (vi) a more innovative firm will outsource a higher fraction of tasks which involve lower complementarity of workers in low-skilled occupations with the firm's other assets.

Le premium de salaire pour les emplois peu qualifiés des entreprises innovantes

RÉSUMÉ

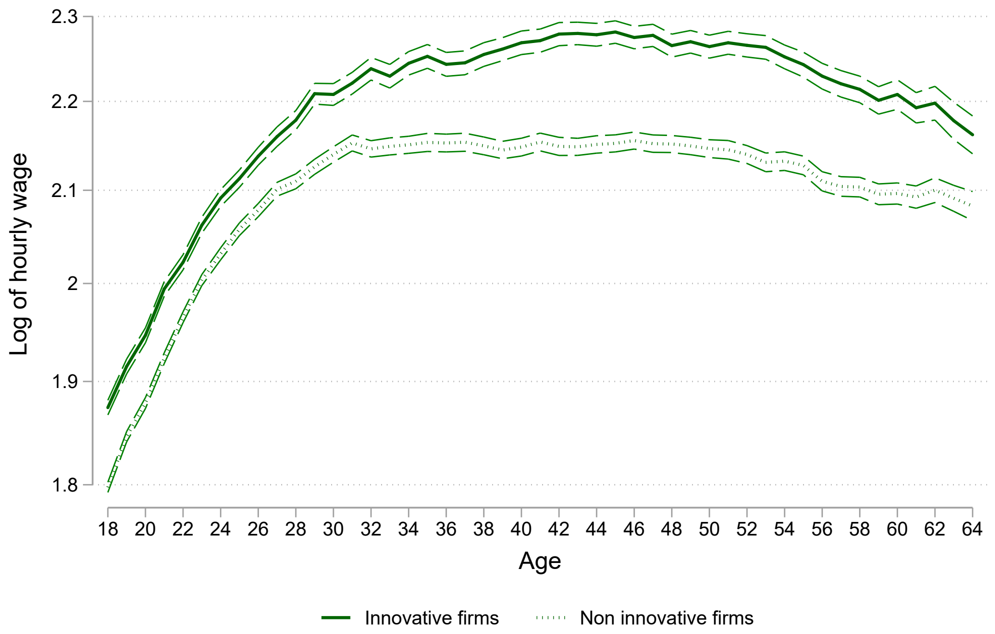
Nous utilisons des données administratives liants les employés à leurs employeurs afin d'étudier les avantages salariaux des travailleurs dans les entreprises innovantes. Nous trouvons que les entreprises effectuant beaucoup de dépenses de R&D payent des salaires en moyenne plus élevés que les autres, et cela est particulièrement vrai pour les travailleurs réalisant certaines tâches. Nous construisons alors un modèle théorique dans lequel les entreprises les plus innovantes sont caractérisées par un degré de complémentarité plus important entre les travailleurs effectuant certaines tâches peu qualifiées et les autres. Pour ces tâches peu qualifiées, les compétences dites « soft skills », typiquement mal observées et difficilement vérifiables représentent une part importante du salaire. Le modèle fournit des prévisions supplémentaires sur la formation, la durée de titularisation et la délocalisation, qui sont également étayées par les données.

Mots-clés : Innovation, Changement Technique, Salaires, Complémentarité

1 Introduction

This project came out of a surprise empirical finding. Figure 1 uses matched employer-employee data from the UK over the period 2004-2016 and plots the average hourly wage at each age for workers in low-skilled occupation, respectively in innovative and non-innovative firms.[1] We see that the average worker in low-skilled occupations obtains a significant wage premium from working in an innovative firm.

Figure 1: Average wage of workers in low-skilled occupations



Notes: Authors' calculations based on 442,916 observations (228,613 on work in non-innovative firms and 154,303 in innovative firms) in matched ASHE-BERD data; see Section 3 and Appendix A for more details about the data. The figure plots the average log of hourly wage at each age from 18 to 64 for workers in low-skilled occupations (see Appendix A.2.3). The dashed curve is for workers in firms that do not report R&D expenditure, the solid curve for workers in firms are firm that report positive R&D expenditure; see Appendix A.1. 95% confident intervals are included.

Our contribution in this paper is twofold. First, we develop a model of wage bargaining with complementarity between high and low-skilled workers, to explain why workers in *some* low-skilled occupations get a positive premium from working in a more innovative firm. The main assumptions of the model are that: (i) a worker's productivity depends upon both hard skills and soft skills; (ii) more innovative firms exhibit a higher degree of complementarity between workers in high-skilled occupation and those workers in low-skilled occupation that have a high level of soft skills; (iii)

[1]The data are described in more details in Section 3. Innovative (resp. non innovative) firms are defined as firms with positive (resp. zero) R&D investments over the period.

hard skills are largely observable whereas soft skills are less easy to detect *ex-ante*; (iv) soft skills form a larger proportion of the abilities of workers in low-skilled occupations (i.e. hard, verifiable skills are not that important in determining the wages of these workers).

These assumptions in turn imply that workers in low-skilled occupations with high soft skills command higher bargaining power in more innovative firms (as compared to similar workers in less innovative firms). Workers in higher-skilled occupations typically have observable qualifications, and their market value is primarily determined by their education and accumulated reputation, which are easily observable and verifiable. A firm can replace a worker with a high proportion of hard skills by another similar worker with limited downside risk, because their quality is observable. In contrast, the key qualities of workers in *some* low-skilled occupations are soft skills (non-cognitive) and can be difficult to observe or develop, hence difficult to replace.[2]

To gain intuition think of a worker in a low-skilled occupation, for example a maintenance worker, a personal assistant or a sales telephonist, who shows outstanding initiative and reliability. These attributes may be difficult to measure and verify. Yet, they allow the worker to perform tasks which complement the tasks performed by workers in high-skilled occupations within the firm in the sense that if performed well they can increase the productivity of the high-skilled employees, but if mistakes are made by the worker in the low-skilled occupation these can be damaging to the firm's overall performance.

Our second contribution is to use matched employer-employee data from the UK, augmented with information on R&D expenditures, to test the assumptions and main predictions of the model. Namely: (i) workers in low-skilled occupations get a positive and sizeable premium from working in a more innovative firm; (ii) workers in low-skilled occupations exhibit on average a higher degree of complementarity with the firm's other assets in more innovative firms compared to in less innovative firms; (iii) the wage premium to working in a more innovative firm for workers in low-skilled occupations increases with the complementarity between their quality and the firm's other assets; (iv) workers in low-skilled occupations have longer tenure in more innovative firms than in less innovative firms as more time will be spent by more innovative firms to enhance (or learn about) these workers soft skills; (v) a more

[2]In our model the soft skills of workers in low-skilled occupations are largely unknown to the firm at the point of hiring or they require that the firm invest in training. Our model is not therefore a simple matching set up, and tenure increases the premium for workers in low-skilled occupations more in more innovative firms.

innovative firm will invest more in training its workers in low-skilled occupations to increase their level of soft skills than a non-innovative firm; (vi) a more innovative firm will outsource a higher fraction of tasks which involve lower complementarity of workers in low-skilled occupations with the firm's other assets.

Our work 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). Our finding that the premium to working in more innovative firms is high for workers in some low-skilled occupations is not at odds with the view that technical change has become increasingly skill-biased over the past thirty five years. Indeed, we find that more innovative firms outsource a higher fraction of tasks performed by workers in low-skilled occupations, but presumably they keep those workers in low-skilled occupations with high soft skills and which are more essential to the firm. Thus Akerman et al. (2015) study the impact of the adoption of broadband internet on wages, and find that overall workers in low-skill occupations benefit less from the new technology, even though the quality of some workers in low-skill occupations and the tasks they perform remain valuable.

Second, there is the labor and wage literature (Gibbons and Katz, 1992, Groshen, 1991, Abowd et al., 1999 and Bonhomme et al., 2019 among others), which emphasizes that firm heterogeneity plays a large role in explaining wage differences across workers; however, there is little consensus in explaining which features of the firm account for such variation.[3] This literature has established that there is considerable wage inequality between seemingly similar workers, and that this inequality is in turn correlated with the firm that employs these workers (this is typically captured by a firm fixed effect).[4] Less is known, however, about what drives these cross-firm differences in wages, particularly for workers in low-skilled occupations. We highlight one channel, namely the interplay between the firm's innovativeness, and the com-

[3]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. (2018) 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 dispersion in wages to productivity dispersion across firms (Barth et al., 2016, Dunne et al., 2004). Matched employer-employee data are often leveraged 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 (see Card et al., 2016 for a review). 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.

[4]Card et al. (2016) report that, *"most studies that control for worker heterogeneity find wage-productivity elasticities in the range 0.05-0.15."*

plementarity between the (soft) skills of workers in low-skilled occupations and the firm's other assets.

Third, our paper relates to the literature on innovation and growth. That innovation should play a key role in explaining cross-firm wage differences, is in line with 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. And indeed recent papers in this literature look at the effects of innovation on income inequality using aggregate data (e.g. Aghion et al., 2018a; and Akcigit et al., 2017). Taking a more microeconomic approach, Kline et al. (2018) for the US and Aghion et al. (2018b) for Finland, use administrative tax data merged with patent data to look at the individual returns from innovation to the inventors and to their co-workers. Both papers find significant returns to innovation, most of which accrue to other employees or stakeholders within the inventor's firm.[5] We contribute to this literature by focusing on the returns to soft skills for some workers in low-skilled occupations in more versus less innovative firms, and on how innovativeness affects the degree of complementarity between these workers and the firm's other assets (including workers in high-skill occupations).

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, 2000 and Garicano and Rossi-Hansberg, 2006). We contribute to this literature by linking the firm's innovativeness to the complementarity between workers in low-skill occupations who have high soft skills and the firm's other assets.

The structure of the rest of the paper is as follows. In Section 2 we develop the theoretical framework. In Section 3 we present our data and empirical methodology, and establish that more innovative firms pay higher wages to observationally similar workers, particularly in low-skilled occupations. In Section 4 we develop our measure of the complementarity between the firm's other assets and workers in low-skilled occupations. We do this at the occupation level using the O*NET data (based on workers and firms in the US). We provide evidence consistent with the view that the wage premium of workers in low-skilled occupations from working in more innovative firms increases with the workers' soft skills and with the complementarity between these soft skills and the firm's other assets. In Section 5 we test the additional

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

predictions from the model and discuss the robustness of our main findings. Section 6 collects our concluding remarks.

2 A Model

In this section we develop a model in which: (i) workers productivity depends upon both, hard skills and soft skills and (ii) more innovative firms exhibit a higher degree of complementarity between workers in high and low-skilled occupations. A key feature of the model is that hard skills are largely observable (e.g. those are typically more educated employees, whose market value is largely determined by their education and accumulated reputation), whereas soft skills are less easy to detect *ex-ante* or require more training. Moreover, soft skills account for a larger fraction of workers' overall abilities for workers in low-skilled occupations than for workers in high-skilled occupations. Workers in low-skilled occupations with relatively high soft skills draw bargaining power for two reasons. First from the fact that they are more complementary to workers in high-skilled occupations. Second from the fact that it is hard for the firm to find alternative workers in low-skilled occupations with relatively high soft skills: instead, firms need time to find or train workers to get equal levels of soft skills. As a result, workers in low-skilled occupations with high soft skills will command a higher wage in more innovative firms. If we further assume that the firm's output suffers more from replacing a worker in a low-skilled occupation with high soft skill than from replacing a worker in a high-skilled occupation, then the wage differential between workers in low-skilled occupations in more versus less innovative firms will be higher than the wage differential between workers in high-skilled occupations in more versus less innovative firms. We now proceed to formalize our argument.

2.1 Model setup

Production function

We consider a representative firm which we model as a two-layer hierarchy with workers in high and low-skilled occupations. For simplicity we assume that there is one employee in a high-skilled occupation who monitors a continuum of tasks, each of which is performed by a different worker in a low-skilled occupation.[6] Tasks are ranked according to the degree of complementarity $\lambda \in [0, 1]$ between workers in high

[6]See Appendix C.4 for an extension with more than one worker of each type in each task.

and low-skilled occupations. If Q denotes the overall quality of the employee in the high-skilled occupation, and $q = q(\lambda)$ denotes the overall quality of the worker in the low-skilled occupation on task λ, then the output produced on that task is assumed to be determined by the following "partially O'Ring" production function (see Kremer, 1993 and Kremer and Maskin, 1996):

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

The value $\lambda = 0$ corresponds to full substitutability between the qualities of the employees in the high and low-skilled occupations. The value $\lambda = 1$ corresponds to the case where the qualities of the employees in the high and low-skilled occupations are fully complementary.

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, which we allow to vary with the degree of innovativeness z of the firm, we denote the firm's aggregate production by:

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

where:

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

Wage negotiation

For each task λ, the firm engages in separate wage negotiations with the workers respectively in the high and low-skilled occupations on that task. This negotiation leads to the equilibrium wage $w_q(\lambda)$ for the worker in the low-skilled occupation and to w_Q for the worker in the high-skilled occupation. We denote by β^L (resp. β^H) the fraction of the firm's net surplus that accrues to the worker in the low-skilled occupation (resp. high-skilled occupation) 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 opportunity of replacing the employee in the high-skilled occupation – whose quality is Q – by an outside high-skilled employee with *ex-ante* expected quality Q_L. Similarly, on each task λ, the firm has the outside option of replacing the worker in the low-skilled occupation on that task - this worker has quality $q(\lambda)$ and is paid wage $w_q(\lambda)$ - by an outside worker with reservation quality q_L and reservation wage w_L.[7]

We assume that it is easier for the firm to find a substitute for the employee in the high-skilled occupation than for the employee in the low-skilled occupation. The underlying idea is that soft skills account for a higher share of the overall quality for a workers in these low-skilled occupations than for workers in high-skilled occupations, and that soft skills are harder to detect *ex-ante* or to generate via training than hard skill. Formally, this leads us to assume that:

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

for all λ, where we also assume that $Q > Q_L >> q(\lambda) > q_L > 1$.

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 incumbent workers in the high and low-skilled occupations have outside options $\bar{w}^L$ 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

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

Net 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 the wage negotiation, the firm can learn about or train the worker in the low-skilled occupation on each task λ, so that the expected quality of the worker moves up 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 *ex-post* profit minus the training cost, namely:

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

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

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

Innovativeness and complementarity

We assume that more innovative firms display higher average complementarity between the qualities of workers in high and low-skilled occupations across tasks. More formally, we assume that

$$\mathbb{E}_{\phi}\left(\lambda, z\right) = \int_0^1 \lambda \phi(\lambda, z) d\lambda$$

increases with the innovativeness measure z.

Here are two tractable cases which will allow us to nicely develop our intuitions:

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

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

which increases with the innovation intensity z.

Example 2. *("Toy Case"): Suppose that: $\phi(\lambda, z)$ is equal to 1 only for $\lambda = \lambda_z \equiv \frac{z}{z_{max}}$ (where z_{max} denotes the maximum value z can take) and to zero for $\lambda \neq \lambda_z$. In that case:*

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

which again increases with the innovation intensity z.

2.2 Solving the model

To simplify the analysis we henceforth assume 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.

2.2.1 The toy case

Here we consider the toy case where $\phi(\lambda, z) = 1$ 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 wages in the low-skilled occupation The firm's net surplus from employing a worker with quality q in a low-skilled occupation on the unique task λ_z, is equal to:

$$S^F = \left[\lambda_z Q + (1-\lambda_z)\right](q - q_L) - w_q + w_L.$$

The surplus of the worker in the low-skilled occupation on that task is equal to

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

where $\bar{w}^L$ is the worker's outside option.

Since we assume $\beta_L = 1$, the equilibrium wage of the worker in the low-skilled occupation on the unique task λ_z is determined by equalizing the two surplus, hence:

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

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

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

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

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

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

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

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 is lower than $\overline{q}$. [8] Note that q^* is increasing with λ_z, and therefore with z: that is, the optimal level of training of a worker in a low-skilled occupation is higher in a more innovative firm.

[8] A sufficient condition is that $\overline{q} > q_L + \frac{Q_L}{4C}$, which is true as long as training costs are sufficiently large.

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

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

and similarly, the equilibrium wage of the worker in the high-skilled occupation on task z, up to a constant, is equal to:

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

We can then establish:

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

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

Proof. See Appendix C.1 □

The proposition immediately results from the facts that: (i) in more innovative firms the complementarity is higher between the worker in the high-skilled occupation and the workers in the low-skilled occupation with training (i.e. with high soft skill); (ii) optimal training of workers in low-skilled occupations is higher in more innovative firms so that replacing the current worker in the low-skilled occupation by an outside worker has a more negative impact for such firms.

2.2.2 The general case

We now consider the case where the firm covers a whole range of tasks with a continuous density distribution $\phi(\lambda, z)$ over tasks λ. We assume that $\phi(\lambda, z)$ is increasing in both λ and the innovativeness level z. We then establish:

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

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

Proof. See Appendix C.2 □

2.2.3 Outsourcing

Assume that the firm is subject to an overall time constraint (or "limited attention" constraint) for training or screening. Formally:

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

so that *ex-ante* the firm maximizes

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

subject to that constraint.

Then if the above time constraint is binding, for sufficiently low λ it is optimal for the firm to fix $q^*(\lambda) = q_L$, which we interpret as outsourcing the corresponding task. The following proposition establishes that the cutoff value of λ below which the firm outsources tasks increases with the firm's degree of innovativeness z.

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.

Proof. See Appendix C.3 □

In the case where $\phi(\lambda, z) = \lambda^z(z+1)$, it is possible to derive close-form expressions for $\bar{\lambda}(z)$ for integer values of z (see Appendix C.3). Figure 2 shows the cases $z = 0, 1$ and 2.

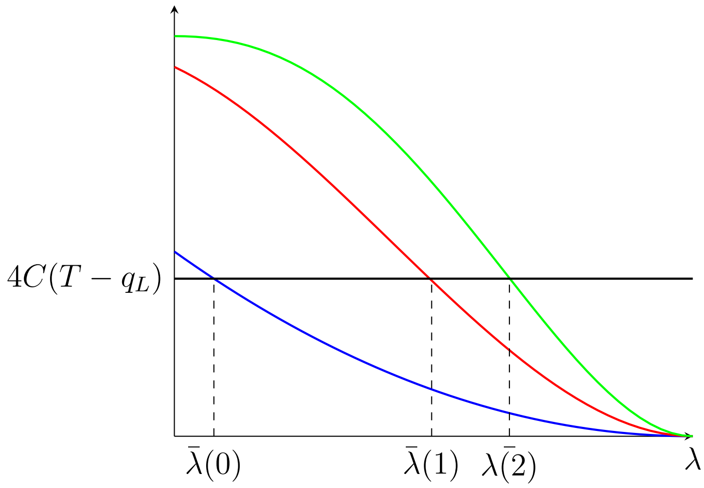
2.3 From model to data

In the next sections, we confront the following assumptions and predictions of the model with data.

Fact 1: Workers in low-skilled occupations get a positive premium from working in a more innovative firm.

Fact 2: In more innovative firms workers in low-skilled occupations exhibit on average a higher degree of complementarity with the firm's other assets.

Figure 2: $\bar{\lambda}$ as a function of z



Fact 2: There is a wage premium to working in a more innovative firm for workers in low-skilled occupations, which is driven by the complementarity between their quality and the firm's other assets.

Fact 3: Workers in low-skilled occupations should have longer tenure in more innovative firms than in less innovative firms (as more time and money is invested in getting them from q_L to q^*).

Fact 4: A more innovative firm will invest more in training its workers in low-skilled 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).

Fact 5: A more innovative firm will outsource a higher fraction of tasks which involve lower complementarity between workers in low-skilled occupations and the firm's other assets.

3 The innovation premium for workers in low-skill occupations

In this section we describe our data and empirical approach to establish that more innovative firms pay higher wages to observationally similar workers, particularly to workers in low-skilled occupations.

3.1 Data and descriptive statistics

We use novel matched employer-employee data for the UK to which we have also matched information on R&D expenditure; we consider the period 2004 to 2016. 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 expenditures on Research and Development (BERD) survey, which is a census for firms with more than 400 employees. These data are longitudinal - we follow the same workers over time - and are 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 *164,345* employees who work in *6,941* firms, giving us a total of *680,649* observations. Further details on the data are given in Appendix A.

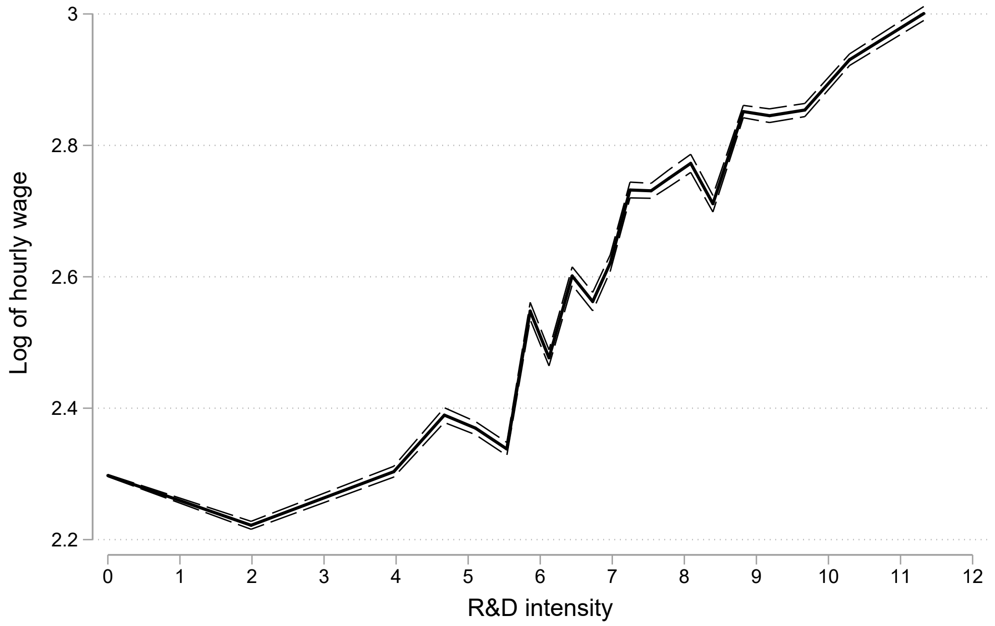
We classify occupations by the average required skill level based on qualifications, see details in Appendix A.2.3. We distinguish between low-skilled occupations, which require minimal formal education and training; intermediate-skilled occupations, which typically require the equivalent of a high-school education and include trades, specialist clericals, associate professionals; and high-skilled occupations, which typically require advanced training or a university degree and include engineers and managers.

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 even after controlling for a range of observable worker and firm characteristics. Figure 3 shows that the average wage of workers increases with the firm's R&D intensity; average wages are over twice as high in the most R&D intensive firms compared to firms that do no R&D.

This result echoes those of Van Reenen (1996), who showed that innovative firms pay higher wages on average, using information on publicly listed UK firms. Another way to see this is by comparing the share of workers who work in a firm that does R&D at different points in the wage distribution; this increases from just over 20% for workers at the bottom of the wage distribution, to over 50% at the 80^{th} percentile of the distribution (see Figure A4 in Appendix A.5).

Workers in more R&D intensive firms have different characteristics to those working in less R&D intensive firms. Table 1 shows that the former are more likely to be males, work full-time and have longer tenure within the firm. R&D firms also differ

Figure 3: Wages and R&D intensity



Notes: This figure plots the average value of the log hourly wage against the average value of R&D intensity ($log(1 + R/L)$ where R is total R&D expenditures and L is total employment in the firm. R&D intensity ranges from 0 (no R&D expenditures) to around 12. R&D intensity if plotted by dividing firm into 20 percentiles of the R&D intensity distribution, with one category for non innovative firms; the number shown on the horizontal axis is the mean R&D intensity in that percentile. Wages are defined in Appendix A.2.2. R&D intensity is defined in Appendix A.1.

from non-R&D firms in that they have a larger workforce. All of these might affect the wages of workers in these firms. In Appendix A we provide further descriptive statistics of the key variables.

3.2 The wage premia to working in more innovative firms

In this subsection we provide evidence that more innovative firms pay higher wages to observationally similar workers and, in particular, that workers in low-skilled occupations get a positive premium from working in a more innovative firm.

We estimate the following relationship:

$$ln(w_{ijkft}) = \beta \tilde{R}_{ft} + g_T(T_{ift}) + g_A(A_i) + \alpha_F F_{ift} + \alpha_S S_{ift} + \gamma_i + \eta_t + e_{ijkft}, \qquad (3)$$

where i indexes individual, j occupation, k labor market, f firm and t years; w_{ijkft} is the hourly wage, and $\tilde{R}_{ft} = ln(1 + R_{ft})$ is R&D intensity.[9] T_{ift} is the worker's tenure

[9] R_{ft} is defined as R&D expenditures divided by the 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 equal to $ln(R_{ft})$ given

Table 1: Firm characteristics

	Innovative firms	Non-innovative firms	All
Employment	2,796	2,222	2,396
Hourly Wage (£)	15.8	12.9	13.8
Share of Male (%)	68	57	60
Share of full-time (%)	90	77	81
Workers in high-skilled occupations (%)	30	18	22
Workers in low-skilled occupations (%)	52	64	60
Age (years)	40.5	38.3	38.9
Tenure (years)	8.9	5.9	6.8
Workers	71,847	135,184	164,345
Firms	2,299	6,184	6,941
Firms-years	12,711	29,158	41,869
Workers-firms-years	261,513	419,136	680,649

Notes: Innovative firms are those that report any R&D expenditures over the period. Employment is the average number of workers in the firm over all years. Wages are defined in Appendix A.2.2. Definitions of high and low-skill occupations are given in Appendix A.2.3.

(length of time working in the firm), A_i is the age of the worker, F_{ift} is an indicator of whether the job is full-time (as opposed to part-time), S_{ft} is the number of employees in the firm, and η_t represent common time effects. Finally, e_{ijkft} captures remaining idiosyncratic time varying unobservables.

Table 2 presents our estimates of equation (3). In columns (1)-(3) we include workers in all occupations; the columns differ in the controls for unobservables that we include. In column (1) we include labor market effects interacted with time effects ("Geo-Year"). Labour markets are defined as a travel to work area; there are around 240 such areas in the UK, see Appendix A.3. The coefficient estimate of 0.030 suggests that workers in the most R&D intensive firms earn nearly 40% more than workers in firms that do no R&D.[10] Compared to Figure 3, we see that controlling for workers and firms characteristics accounts for a substantial part of the differences we saw in the raw data.

the magnitude of R&D expenditure, so we can interpret β as the elasticity of wage with respect to R&D intensity. In Section 5.5 we show robustness of our results to alternative functional forms and alternative measures of R&D.

[10] This is exp(predicted wage at $max(\tilde{R}_{ft})$) minus exp(predicted wage at $min(\tilde{R}_{ft})$))/exp(predicted wage at $min(\tilde{R}_{ft})$)=0.38, where the predictions use the coefficient estimates from column (1) of Table 2, $max(\tilde{R}_{ft})$ denotes the average R&D intensity level of firms in the highest quantile (see Table A1) and $min(\tilde{R}_{ft}) = 0$ (non innovative firms).

Table 2: Relationship between wages and R&D intensity

	Dependent variable: $ln(w_{ijkft})$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Skill level	All	All	All	Low	Int	High	All
$\tilde{R}_{ft}$	0.030***	0.016***	0.006***	0.007***	0.003***	-0.000	0.003***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
× Int-skill							0.002***
							(0.001)
× Low-skill							0.006***
							(0.001)
Tenure	0.022***	0.015***	0.008***	0.009***	0.008***	0.000	0.008***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Tenure Squared	-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)
Age	0.057***	0.034***					
	(0.003)	(0.002)					
Age Squared	-0.001***	-0.000***	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.032***	-0.010***	-0.008***	-0.005**	0.001	0.003	-0.007***
	(0.006)	(0.004)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Gender	0.152***	0.140***					
	(0.006)	(0.004)					
Full-time	0.241***	0.069***	0.007	-0.009*	-0.079***	-0.102***	-0.001
	(0.013)	(0.007)	(0.005)	(0.005)	(0.013)	(0.013)	(0.005)
Low-skill dummy							-0.162***
							(0.006)
Int-skill dummy							-0.077***
							(0.004)
Geo-year	✓						
Geo-Occupation-year		✓					
Individual			✓	✓	✓	✓	✓
Year			✓	✓	✓	✓	✓
R^2	0.384	0.625	0.886	0.769	0.847	0.880	0.888
Observations	680,583	669,899	634,542	399,690	100,989	114,953	634,542

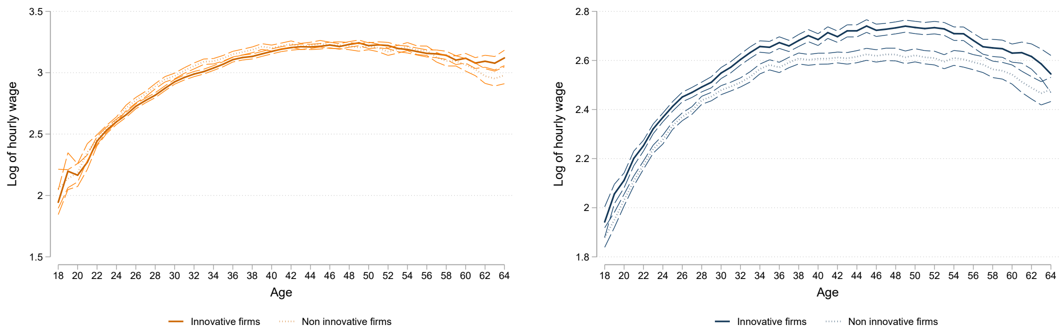
Notes: The dependent variable is log of wage which is defined in Appendix A.2.2. $\tilde{R}_{ft} = ln(1 + R_{ft})$ where R_{ft} is total R&D expenditures of firm f during year t divided by employment. Other covariates definitions are given in Table A7. Column 1 includes year-labor market fixed effects, column 2 includes year-labor market-occupation effects, column 3-7 include year and individual fixed effects. In column 3-7 Age and the Gender aren't identified 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.

In column (2) of Table 2 we add occupation-geo-year effects using the two-digit level of the occupation code (25 occupations). This reduces the coefficient on R&D intensity by about half, the coefficient estimate of 0.016 suggests that workers in the most R&D intensive firms earn around 20% more than workers in firms that do no R&D.

In column (3) we add worker effects, (γ_i). These are meant to capture permanent unobserved attributes that workers carry across firms. These are important and control for selection on unobserved permanent individual characteristics. Higher quality

Figure 4: Average wage of workers in high and intermediate-skilled occupations

(a) High-skill (b) Intermediate-skill



Notes: This figure plots the average log of hourly wage at each age from 18 to 64 for workers in different skill level occupations. The dashed curve is for workers in non-innovative firms, the solid curve for workers in innovative firms. Innovative firms are firms that report at least £1 in R&D expenditures over the period. 95% confidence intervals are included.

workers might select into higher quality firms. This specification accounts for endogenous selection and matching based on the individual effect (γ_i). We drop occupation and labor market effects as we do not observe many workers who move across occupations or labor markets (although including them has negligible effect on our results). This reduces the coefficient on R&D intensity to 0.006, which implies that workers in the most R&D intensive firms earn around 7% more than workers in firms that do no R&D, after controlling for all these differences. Compared to the estimates in column (3) the estimates without worker effects considerably over-estimate the impact of R&D intensity on wages,[11] however, the premium from working in an innovative firm remain statistically and economically significant.

Figure 1 showed that workers in low-skilled occupations earned higher wages on average in innovative firms; is that also true for workers in high and intermediate-skilled occupations? Figure 4 shows that the within-skill group variance of wages across firms is relatively less important for workers in high and intermediate-skilled occupations than for workers in low-skilled occupations. Workers in higher skilled occupations earn the highest wages, and these wages are *on average* similar across firms that are more versus 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. Hence the wage gradient with respect to R&D intensity is largest for workers in low-skilled occupations.

[11] Results using alternative fixed effects and other measures of R&D are shown in Section 5.5.

In order to see whether the wage premia shown in Figures 1 and 4 are robust to controlling for other differences in workers and firms, we separate workers by the skill level of their occupations. Column (4) of Table 2 shows estimates for the sample of workers in low-skilled occupations, column (5) for intermediate-skilled occupations and column (6) for high-skilled occupations. The positive coefficient on R&D intensity is statistically significant for low and intermediate-skill categories and is strongest for the low-skilled occupations. In column (7) we pool all skill levels and allow the intercept and coefficient on R&D intensity to vary with the skill level of the occupation. The premium is highest for workers in low-skilled occupations.

The estimates in column (7) suggest that on average workers in low-skilled occupations in the most R&D intensive firms earn 6% more than workers in firms that do no R&D; and this is more than for workers in high or intermediate-skilled occupations.

These results confirm that on average workers in low-skill occupations obtain a positive wage premium from working in a more innovative firm. This remains true after controlling for observable and non time-varying unobservable characteristics; this in turns validates Fact 1 from our model.

One remark to conclude this section: even though workers in high-skill occupations do not seem to enjoy much of a wage premium from working in a more R&D intensive firm, at the same time more innovative firms hire fewer workers in low-skilled occupations and more workers in high-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 19% to 60%. Thus overall, workers in high-skilled occupations are those who benefit the most from working in a more innovative environment. That we do not observe sizeable wage premia for workers in high-skill occupations employed in more innovative firms is explained in our model by the fact that hard skills are more easily verifiable and therefore the labor market is more competitive for these workers (we come back to this point in the next section).[12]

[12]As noted earlier, our contrast is different to that studied in Kline et al. (2018). They study innovative firms and ask which types of workers gain from the award of a patent using workers in firms where a patent is refused as a control group. They find that it is the high-skilled occupations in innovative firms that receive a larger share of the rent when the patent is approved.

4 Measuring complementarity and the return to soft skills

Our model in Section 2 relies on the distinction between hard skills and soft skills. Hard skills are reasonably easy to observe, for example, by formal qualifications. Soft skills are more difficult to observe, both for us as researchers and for employers. In our model, what drives the returns to working in an R&D firm for workers in low-skilled occupations is that *some* of these workers have soft skills and these are important for the firm, i.e. more complementary to the firm's other assets. By their very nature these soft skills are difficult to observe and measure. Note that we are not claiming that the *absolute* importance of soft skills is higher for workers in low-skilled occupations than for workers in high-skilled occupations. Instead, our model is predicated on the idea that soft skills are *relatively* more important for workers in low-skilled occupations. Consider the example of a researcher and an administrative assistant. Even if the average researcher turns out to enjoy higher soft skills than the administrative assistant, yet her income will be mostly determined by the university she graduated from and more importantly by her track record of publications and inventions: all these elements are verifiable (they are typically included in the researcher's CV). In other words, soft skills will be relatively unimportant in determining the researcher's overall income. In contrast, an administrative assistant might have lower soft skills than researchers on average. Yet these will represent a higher share of her value to the firm, especially if they are more essential to the firm, and therefore will play a more important role in determining the assistant's overall income. These will also be difficult to observe, but will be revealed progressively over time.

4.1 Measuring complementarity

How can we measure the complementarity between the firm's other assets and workers in low-skilled occupations for each task? We do so at the occupation level using the O*NET data. The O*NET data provide detailed information on the characteristics of occupations based on surveys of workers and experts in the US (more detailed are given in Appendix A.6). We work at the 3 digit SOC 2010 occupation level, where we can reasonably match US occupations to UK occupations.

The O*NET data contain a number of questions that are related to the idea of complementarity, which in the model is captured by λ. We select seven dimensions that we believe are most relevant for our purposes and which we aggregate into a single score using principle components analysis. Specifically, workers are surveyed across occupations and asked to grade each dimension from 1 (when this dimension is not relevant to the workers' occupation) to 5 (when this dimension is very relevant to the workers' occupation). We run a principal component analysis eigen decomposition and consider the first eigen axis as our measure of complementarity λ; this explains more than 57% of the total variance from all the dimensions. Table 3 presents these seven dimensions and their relative importance in the definition of λ.

Table 3: O*NET dimensions contributing to λ

O*NET Dimension	**Weight**
How important is being very exact of highly accurate in performing the job?	0.1057
How serious would be the result usually be if the worker made a mistake that was not readily correctable?	0.3262
What results do your decisions usually have on other people or the image or reputation or financial resources of your employer?	0.4371
How important is it to work with others in a group or team in this job?	0.3793
How responsible is the worker for work outcomes and results of other workers?	0.4027
How important is it to coordinate or lead others in accomplishing work activities in this job?	0.4453
How important is the following skill for your job: "Adjusting actions in relation to others' action"?	0.4325

Notes: Results from a Principal Component Analysis of the seven dimension taken from O*NET at the occupation level. The weight correspond to the decomposition of the first axis.

Table 4: Mean λ by skill level of occupation

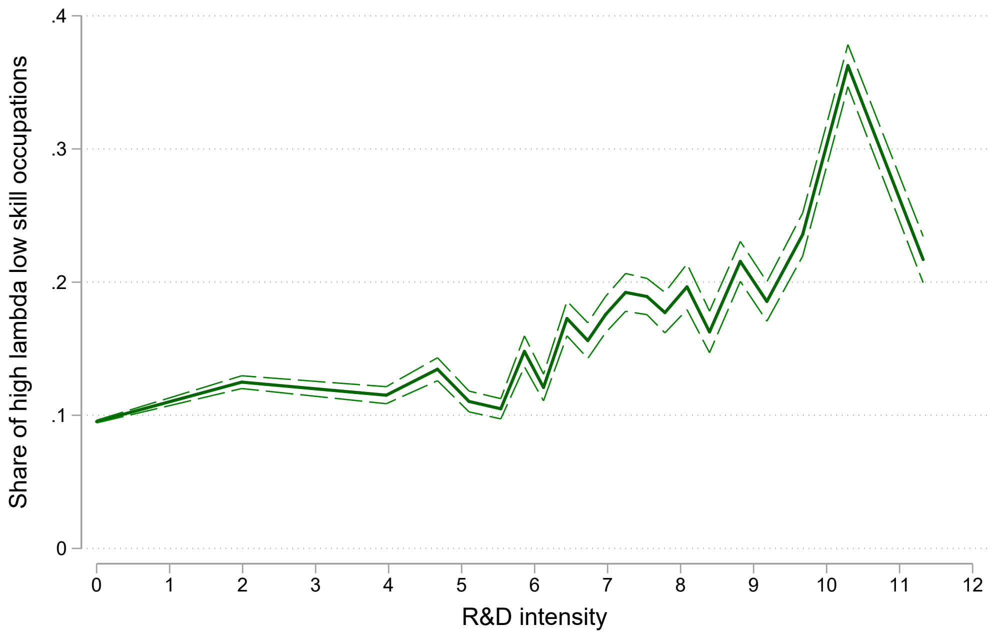
Skill level	mean (sd) λ
Low	0.34 (0.13)
Medium	0.60 (0.18)
High	0.65 (0.15)

We standardized our resulting λ measure so that it lies between 0 and 1. On average, employees in low-skilled occupations work on a task with a λ equal to 0.34 (see Table 4). In what follows, we refer to a high lambda occupation as an occupation in the top 33% of the distribution of lambdas, and similarly to a low lambda occupation as an occupation in the bottom 33%. Following that terminology, 49% of workers are in low lambda occupations, 33% on the middle lambda occupations, and 18% in high lambda occupations.

4.2 Workers in high-lambda low-skilled occupations are more essential in more innovative firms

Here we show that more innovative firms have a higher share of workers in low-skilled occupations that are associated with a high value of λ (hence displaying greater complementarity with workers in high-skilled occupations) than less innovative firms. More specifically, in Figure 5 we see that the share of high lambda occupations among low-skilled occupations essentially increases with the firm's R&D intensity; this vindicates our first conjecture (Fact 2 of the model). This share increases from around 10% for firms that do no R&D up to around one third for the most innovative firms.

Figure 5: Share of workers in low-skilled occupations that are in high lambda occupations, by R&D intensity

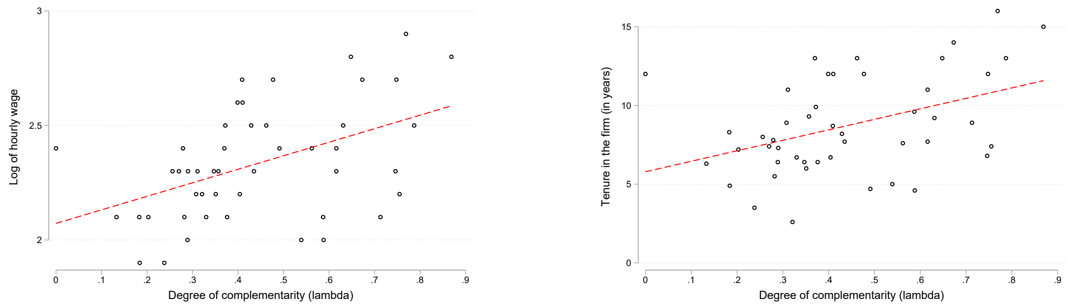


Notes: This Figure reports the average share of workers in high lambda occupations among workers in low-skilled occupations against &D intensity. R&D intensity if plotted by dividing firm into 20 percentiles of the R&D intensity distribution, with one category for non innovative firms; the number shown on the horizontal axis is the mean R&D intensity in that percentile.

In line with our model, we expect workers in these *high lambda and low-skill* occupations to have a higher wage in more innovative firms. This is indeed what Figure 6(a) hints at; this figure shows a clearly positive correlation between the log of hourly wage and the average level of complementarity of tasks in the firm, for innovative firms. Doing the same exercise but replacing the logarithm of hourly wage by the tenure in the firm also yields a positive correlation, as shown in Figure 6(b).

Figure 6: Hourly wage and tenure against lambda for low-skilled occupation workers in innovative firms

(a) Wage (b) Tenure



Notes: These figures show the average log of hourly wage (left-hand side figure) and tenure in years (right-hand side figure) for workers in low-skilled occupations who work in innovative firms against the level λ.

4.3 The return to soft skills in more innovative firms

In this subsection we provide evidence consistent with the view that the wage premium for workers in low-skilled occupations from working in more innovative firms increases with the workers' soft skills and the complementarity between these soft skills and the firm's other assets.

Suppose that the level of soft skills for individual i is represented by ψ_i. To allow for the value of soft skills to differ across firms depending on their R&D intensity we augment wage equation (1) and write:

$$ln(w_{ijkft}) = \beta R_{ft} + \phi_j(R_{ft}, T_{ift}, \psi_i) + g_T(T_{ift}) + g_A(A_i) + \alpha_F F_{ift} + \alpha_S S_{ift} + \gamma_i + \eta_t + e_{ijkft}, \quad (4)$$

where $\phi_j(R_{ft}, T_{ift}, \psi_i)$ is included to capture the marginal impact of soft skills on the wage of workers in low-skilled occupations in an innovative firm. From our model, we know that this effect increases with the degree of complementarity λ_j for occupation j and the R&D intensity of the firm. In what follows, we use the following parametric representation of ϕ_j:

$$\phi_j(R_{ft}, T_{ift}, \psi_i) = \alpha_1 \lambda_j.R_{ft}.T_{ift} + \alpha_2 \lambda_j.R_{ft} + \alpha_3 R_{ft}.T_{ift} + \alpha_4 \lambda_j + \psi_i.$$

This measures the return for worker i in occupation j with degree of complementarity λ_j, with soft skills ψ_i, with tenure T_{ift} in firm f, and where R_{ft} is the R&D intensity of firm f at time t (as before i indexes individual, j occupation, k labor market, f firm and t years).

The introduction of the tenure variable (T_{ift}) reflects the fact that soft skills are, by definition, not easily verifiable. The firm (and the worker) have to learn about them. As emphasized in the theory section, this lack of easy verification places workers with high soft skills in innovative firms in a stronger bargaining position. This is precisely what the term $\phi_j(R_{ft}, T_{ift}, \psi_i)$ is designed to capture.

We include two dimensions of unobserved worker heterogeneity – namely γ_i and ψ_i. We are interested in capturing the impact of (difficult to observed) soft skills (ψ_i) on wages. There are potentially confounding factors, for example people with high soft skills might also have other skills that are unobserved by the econometrician but observed by the firm. In principle γ_i would capture these. The difficulty is that γ_i reflects the average of the unobserved component in soft skills that is revealed while the worker is in an innovative firm during the sample period. This could lead us to underestimate the impact of soft skills. Instead we would like to condition on the level of skills of the worker at *entry* into our sample, rather than on the average over the whole sample. To do this our preferred specification is one in which we replace γ_i by a measure of the initial wage. Conveniently our data does contain pre-sample observations on wages, and we use an average of these observations on the wage for each individual. This pre-sample mean will reflect the worker's initial skill level and is not influenced by the evolution of the soft skills term during the observation period.[13]

Table 5 shows the estimates of the parameters in equation (4.3). For simplicity, when computing ϕ_j, we measure R&D by a binary variable equal to 1 if the firm is innovative and use a categorical variable for λ (using our three groups in Table 4). The columns relate to different samples based on tenure of the worker in the firm. The first column includes workers with tenure up to 5 years, the second column up to 8 years, the third up to 10 years and the fourth column up to 15 years. The three way interaction term, R&D firm $\times$Tenure $\times$high-lambda, shows that there is a wage premium to working in a more innovative firm for workers in low-skilled occupations, when these occupations have high lambda (suggesting it is driven by complementarity). This is increasing in tenure, and the more so in the earlier years (Fact 3 of the model). This in turn reflects the fact that soft skills take time to be revealed to the firm. The effect of ϕ_j is less precise when we include workers with tenure up to 10 years and disappears as we move to worker up to a 15 year tenure.

[13]This is similar to an idea developed in Blundell et al. (1999) and Blundell et al. (2002).

Table 5: R&D and hourly wages low-skilled occupations

	Dependent variable: $ln(w_{ijkft})$			
	(1)	(2)	(3)	(4)
Tenure:	$<$ 5 years	$<$ 8 years	$<$ 10 years	$<$ 15 years
R&D firm	0.044***	0.046***	0.044***	0.042***
	(0.014)	(0.014)	(0.014)	(0.013)
$\times$Med-lambda	0.048***	0.041**	0.045***	0.046***
	(0.017)	(0.017)	(0.017)	(0.017)
$\times$High-lambda	0.046*	0.048*	0.056**	0.067***
	(0.028)	(0.027)	(0.026)	(0.025)
$\times$ Tenure	0.001	0.000	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.001)
$\times$Tenure $\times$Med-lambda	-0.001	-0.001	0.000	-0.000
	(0.004)	(0.003)	(0.002)	(0.002)
$\times$Tenure $\times$High-lambda	0.010*	0.008**	0.005	0.001
	(0.005)	(0.004)	(0.003)	(0.002)
Med-lambda	0.061***	0.060***	0.058***	0.055***
	(0.008)	(0.008)	(0.008)	(0.008)
High-lambda	0.098***	0.090***	0.083***	0.074***
	(0.010)	(0.010)	(0.010)	(0.010)
Tenure	0.015***	0.018***	0.020***	0.021***
	(0.002)	(0.002)	(0.001)	(0.001)
$\times$ Med-lambda	0.003	0.002	0.002	0.002*
	(0.002)	(0.002)	(0.001)	(0.001)
$\times$ High-lambda	0.001	0.003*	0.005***	0.007***
	(0.003)	(0.002)	(0.002)	(0.002)
Tenure Squared	0.000	0.000	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.021***	0.021***	0.021***	0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.011***	-0.013***	-0.013***	-0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Gender	0.063***	0.066***	0.068***	0.069***
	(0.004)	(0.004)	(0.005)	(0.005)
Full-time	0.109***	0.109***	0.108***	0.106***
	(0.009)	(0.009)	(0.009)	(0.009)
Initial Wage	0.224***	0.248***	0.259***	0.282***
	(0.008)	(0.009)	(0.009)	(0.010)
Fixed Effects				
Geo-Year	✓	✓	✓	✓
R^2	0.325	0.351	0.365	0.392
Observations	220,214	275,751	300,202	336,596

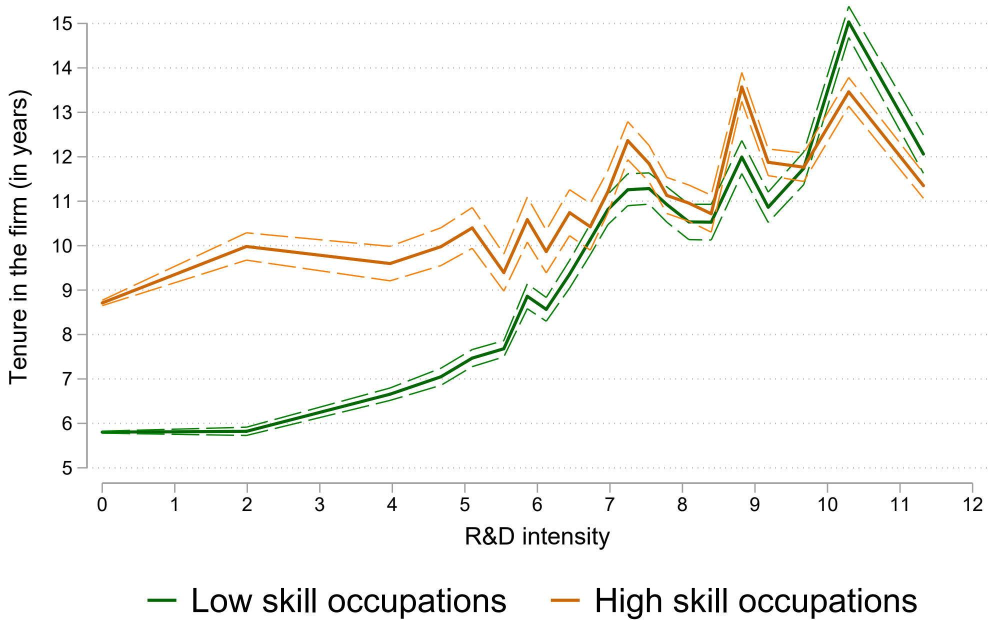
Notes: The dependent variable, log of hourly wage, is defined in Appendix A.2.2. The main regressor, R&D firm, is a dummy variable equal to 1 if the firm is reporting any positive expenditures in R&D. Other covariates definitions are given in Table A7. Ordinary Least Square regression on the sample of workers in low-skilled occupations including travel to work area times 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 Additional Results

5.1 Tenure of workers in low-skilled occupations in more innovative firms

Our model predicts that workers in low-skilled occupations should have longer tenure in more innovative firms, as more time and money is required to get them to the position where they can achieve their full potential. This is indeed what we see from Figure 7 (Prediction 4 of the model).

Figure 7: Tenure, by occupation and R&D intensity



Notes: Vertical axis shows the average of the number of years spent in the firm. R&D intensity if plotted by dividing firm into 20 percentiles of the R&D intensity distribution, with one category for non innovative firms; the number shown on the horizontal axis is the mean R&D intensity in that percentile. 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

As suggested by the theory, a firm may use on the job training to help reveal, and even develop, soft skills. This is something that we also confirm below.

5.2 Training of workers in low-skilled occupations is higher in more R&D intensive firms

We look for evidence in support of the prediction (Fact 4) that more innovative firms invest more in training their workers in low-skilled occupations. Unfortunately, we do

Table 6: On the job and on-site training for workers in low skilled occupations

	Tercile of R&D intensity			
	None (1)	Low (2)	Middle (3)	High (4)
On-site or in-plant				
None	20.3	20.0	18.1	18.4
Up to 6 months	65.8	64.2	59.0	54.8
6 months - 1 year	7.6	8.3	11.5	12.8
A year or more	6.3	7.5	11.4	14.0
On-the-job				
None	10.3	9.9	9.5	9.1
Up to 6 months	74.8	72.9	64.5	60.3
6 months - 1 year	7.8	8.8	13.3	14.8
A year or more	7.1	8.4	12.7	15.9

Notes: R&D firms are split in three groups of equal size based on the value of their R&D expenditures 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.

not have direct information on the spending on training by UK firms, and we do not know if a worker has actually received training. We thus return to the occupation level data in O*NET, exploiting two additional questions about the duration of training on-site or on-the-job. Table 6 reports the share of workers that are in occupations associated with different levels of training (in the US): none, up to 6 months, between 6 months and a year and more than a year. What Table 6 shows in columns 1 to 4 is that in the most R&D intensive firms, from 14% to 15.9% of workers in low-skilled occupations report having received training for more than one year, whereas only 6.3% to 7.1% of workers in low-skilled occupations report having received training for more than one year in no-R&D firms.

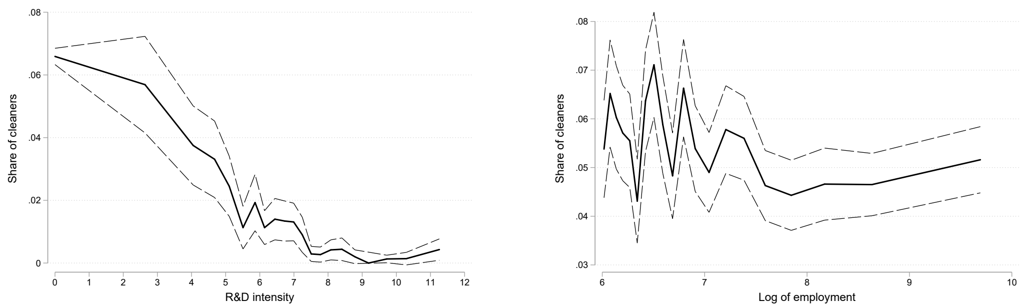
5.3 Outsourcing

Our model predicts (Fact 5) that more innovative firms tend to outsource a higher fraction of tasks than less innovative firms, in particular those tasks with lower complementarity (associated with a smaller λ in the model). The previous results using O*NET data have already shown that innovative firms put more weight on low-skilled 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 linked to the firm that use their services. Second, because we conjecture that most of the outsourcing occurred before 2004, which prevents us from following workers in low-skilled occupations that are outsourced from innovative firms as in Goldschmidt and Schmieder (2017).

We therefore proceed indirectly. We start from the assumption that, because the technology of cleaning does not vary much across firms, all firms need the same share of cleaners, which can be arguably seen as a low λ task. The only reason that the share of cleaners amongst low-skilled workers would be lower than average in some firms is because those firms outsource cleaning. In Figure 8, we plot the share of cleaners among all workers in low-skilled occupations against R&D intensity in the left-hand side panel and against total employment in the right-hand side panel. Figure 8(a) clearly shows that innovative firms employ fewer cleaners than non innovative firms. Our interpretation is that innovative firms are outsourcing them, and Figure 8(b) suggests that this is not a firm size effect.

Figure 8: Share of workers in low-skilled occupations that are cleaners

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



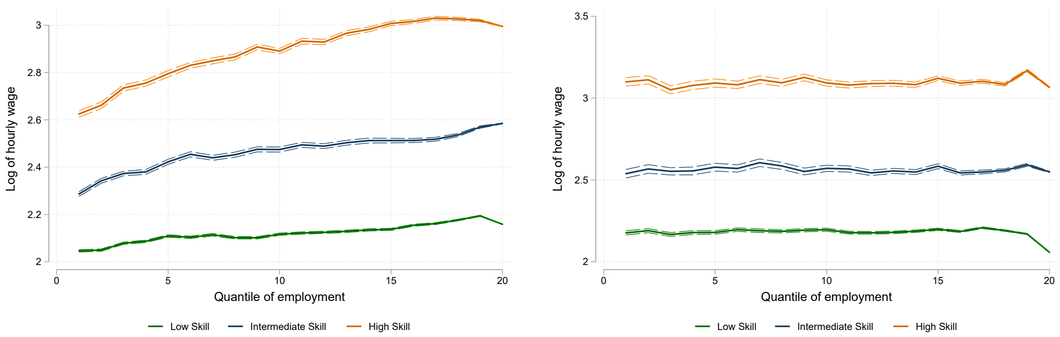
Notes: The y-axis shows the share of cleaners over the total number of workers in low-skilled occupations. R&D intensity if plotted by dividing firm into 20 percentiles of the R&D intensity distribution, with one category for non innovative firms; the number shown on the horizontal axis is the mean R&D intensity in that percentile (left-hand side panel) and the average value of employment for each quantile of employment of the firm with 20 quantiles (right-hand side panel).

5.4 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 wages is a well established fact in the labor literature (see for example Oi and Idson, 1999). This negative effect 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 9.[14]

Figure 9: Wages by firm size

(a) All firms in ASHE (b) Only firms with more than 400 employees



Notes: Vertical axis shows the average of the logarithm of hourly wage by firm. Horizontal axis shows the average value of employment for each quantile of employment of the firm, with 20 quantiles. The left-hand side panel considers all firms that are in ASHE whereas the right-hand side panel only considers firms over 400 employees and corresponds to our final sample.

5.5 Robustness

In this subsection we show that our main results are robust to a number of potential concerns. In particular, we show that the positive correlation between the wage of workers in low-skilled occupations and R&D intensity is robust to including different sets of fixed effects, namely to including firm effects, as in Abowd et al. (1999). We also show that our results are not sensitive to the way we measure wages and to considering alternative functions of R&D, both of which yield qualitatively similar results.

5.5.1 Initial wage and binary measure of R&D

As a first robustness test, we show in Appendix B how the results in Table 2 are affected when we control for initial wage and/or use a binary measure of R&D (Tables B1, B2 and B3) for consistency with what is done in Table 5. Although the difference between workers in low and intermediate-skilled occupations tend to thin down, the

[14]In results not reported here from estimating the effect of R&D on wage using the whole ASHE sample (that is, without restricting attention to large firms), we find a positive and significant coefficient of the logarithm of total employment on wage.

difference in the return to R&D between workers in low and high-skilled occupations remains significant.

5.5.2 R&D, firm effect and other fixed effects

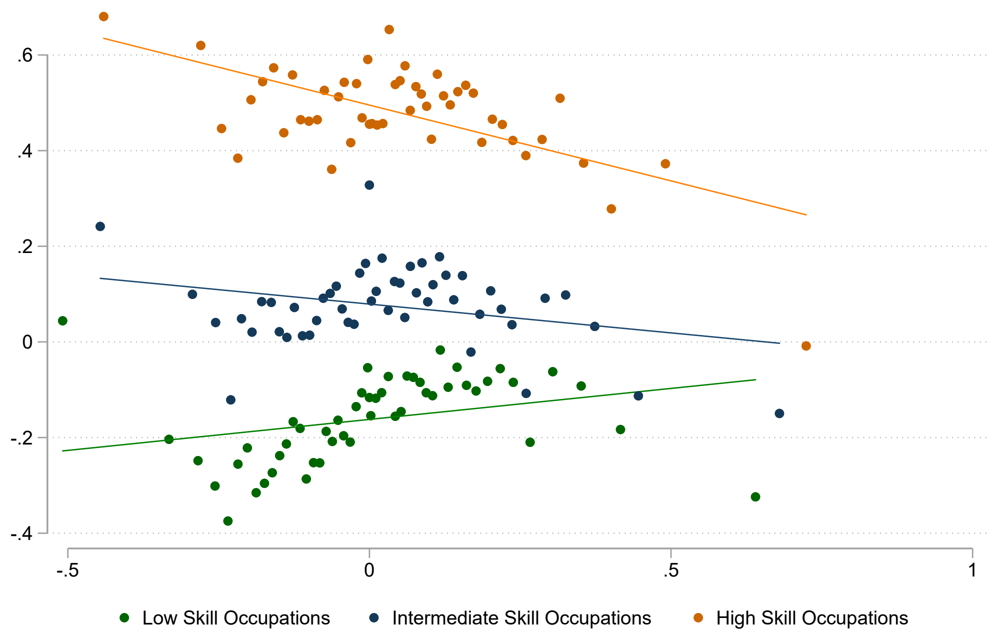
Here we show robustness of our results when different sets of fixed effects are used. From Table 2 we investigate in particular the use of additive two-way firm and worker fixed effects. As discussed in Section 4.3, we cannot use individual fixed effect in the model presented in Table 5, we nevertheless consider robustness tests when occupation fixed or firm fixed effects are included to the model.

Occupation effects: Our results shown in Column 7 of Table 2 control for individual fixed effects. In Table B4 in Appendix B, we change the set of fixed effects. Column 1 drops the individual effect and includes travel to work area - year fixed effects and column 2 includes travel to work area - occupation - year fixed effects. In both case, we include the log of the initial wage as a control to capture worker's initial skill level. Column 3 reproduces our main results using individual fixed effect for comparison. Similarly, in column 2 of Table B5 in Appendix B we show that when occupation fixed effects are added to Table 5 the marginal effect of R&D intensity interacted with tenure for workers in low-skilled occupations in high lambda occupations remains significant (this even holds for workers up to a tenure of 10 years).

Firm effects: In our analysis so far we have not controlled for potential unobservable firm factors. In columns 4 and 5 of Table B4 we include firm effects. R&D remains positive and statistically significant for the workers in low-skill occupations. Here, identification comes off changes in wages in firms that increase or decrease their R&D intensity. The full impact of R&D here is the coefficient on R&D plus the effect of R&D on the firm effect. We obtain an estimate of this from an auxiliary regression. We recover the firm fixed effect from the estimates in column (5) and regress this on mean R&D intensity. The estimated coefficient is 0.0174 with standard error of 0.0011 from a regression on 6,071 observations.

We recover the firm and individual fixed effects using the estimates in column (3) of Table B4. Figure 10 shows that they are positively correlated for workers in low-skilled occupations. Each dot in the graph is the mean effect in the relevant centiles (we split firms into 100 bins of equal size, with around 66 observations per bin) and take the average of worker fixed effects within each bin.

Figure 10: Correlation between firm and individual effects, by skill



Notes: The x-axis shows the average firm fixed effect in each centile. The y-axis shows the average individual fixed effect in each centile. Centiles are 100 equal sized bins by firm fixed effect.

These estimates suffer from a potential incidental parameter bias. We use the panel jackknife split sample procedure suggested in Dhaene and Jochmans (2015). This involves estimates of the model parameters based on splitting the sample into two non-overlapping random sub-samples. Each sub-sample contains one half of the choice occasions for each individual. Denote the estimate for the full sample $\widehat{\beta}$ and the estimate for the two subsamples $\widehat{\beta}_{(1,T/2)}$ and $\widehat{\beta}_{(T/2,T)}$. The jackknife (bias corrected) estimator is:

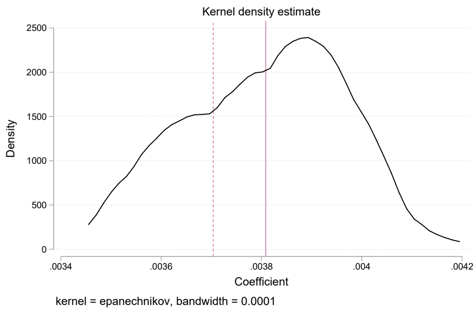
$$\widetilde{\beta}_{split} = 2\widehat{\beta} - \frac{\widehat{\beta}_{(1,T/2)} + \widehat{\beta}_{(T/2,T)}}{2}.$$

We compute the bias-corrected coefficients on R&D firm interacted with skill level of the occupation and on the firm effects. We take the average values over 50 draws. The corrected coefficient on R&D firm interacted with low-skill occupation is 0.0038, compared with the uncorrected estimate of 0.0037. Figures 11(a), 11(b) and 11(c) plots the density of the split sample corrected coefficient over the 50 draws for workers in low, intermediate and high-skilled occupations respecitvely. The solid line shows the mean of these corrected estimates, the dashed line shows the uncorrected estimate.

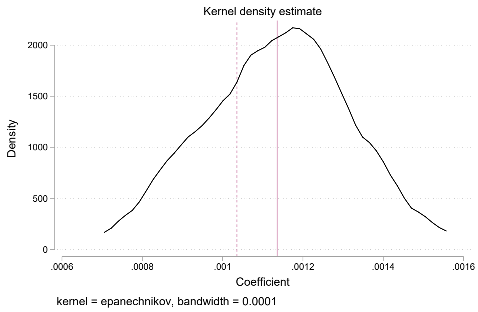
Finally, column 3 of Table B5 includes firm fixed effects to the model presented in Table 5 and shows that the main result remains significant, namely that the in-

Figure 11: Firm fixed effect with and without Jacknife correction

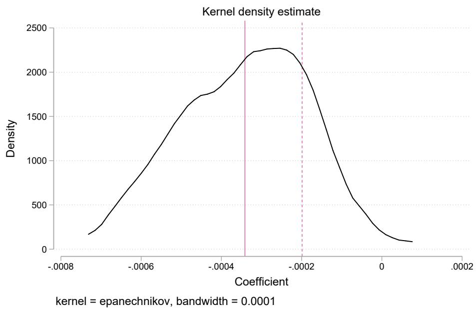
(a) Low-skilled occupations



(b) Intermediate-skilled occupations



(c) High-skilled occupations



Notes: Panel (a) shows the distribution of the coefficient on R&D intensity interacted with the low-skill occupation dummies across our 50 draws. The vertical lines represents the average value of the distribution (solid line) and the uncorrected point estimate from column (5) of Table B4 (dashed line). Panels (b) and (c) do the same but for the interaction with the intermediate skill and high skill dummies respectively.

teraction between the R&D dummy and tenure is positive for workers in low-skilled occupations in high lambda occupations.

5.5.3 Bonus income and other measures of income

We might be concerned with the fact that workers in high-skill occupations might receive a substantial part of their wage in the form of lump-sum bonuses at the end of the year, and that these bonuses might not be 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.

More generally, how are our results affected by the definition of income that we use? In our baseline results we used wages measured in the week that the survey is collected. As explained in Appendix A.2.2, this includes a fixed salary plus variable earnings from incentive, overtime and other pay. Here we test the sensitivity of the correlation between R&D intensity and wage to using other measures of income.

Results are presented in Table B6 in Appendix B where 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 and 4 use total annual earnings including (resp. excluding) bonuses. If workers in high-skilled occupations 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 results if workers in high-skilled occupations receive a larger share of their earnings as bonuses in innovative firms. In fact, the average share of bonuses in annual earnings is 8.8% for R&D firms against 6.5% for non R&D firms. Finally, comparing columns 3 and 4 of Table B6 shows no substantial differences when bonus are included or excluded.

Similarly in Columns 4 and 5 of Table B5, we show that our main results from Table 5 are unaffected when the dependent variable is replaced by the logarithm of total fixed pay per hours (i.e. when incentive and overtime pays are excluded) and the logarithm of total gross annual revenue respectively.

5.5.4 Alternative functions of R&D

Here, we show that our results from Table 2 hold using alternative functions of R&D. Our baseline results use the logarithm of total R&D expenditure divided by total employment in the firm. Figure 3 shows that the relationship between the log of hourly wage and this function of R&D seems to be close to linear. In this subsection we show that our results hold when we consider different functional forms of R&D that give different weights to different level of R&D intensity. In Table B7 we successively consider: $log(1 + \frac{R\&D}{L})$, $log(1 + R\&D)$, $\frac{R\&D}{L}$ and $H\left(\frac{x}{l}\right)$ where H is the hyperbolic function: $H(x) = log(x + \sqrt{1 + x^2})$. In each case, our focus is on the interaction coefficient between this function of R&D and the low-skill occupation dummy.

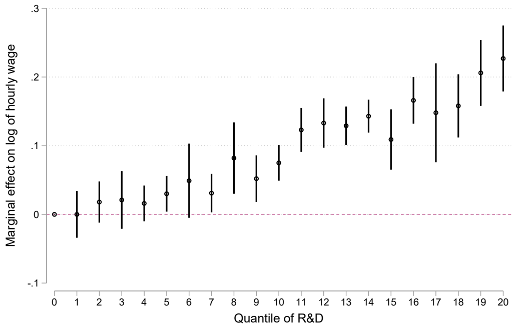
To allow for even more flexibility we let the coefficient vary across the R&D distribution by including a binary variable for each of the twenty quantiles of R&D:

$$ln(w_{ijkft}) = x'_{ift}\beta_1 + z'_{ft}\beta_2 + \sum_{l=1}^{20} \beta_{3l} R_{ftl} + \nu_i + \nu_t + \epsilon_{it}, \tag{5}$$

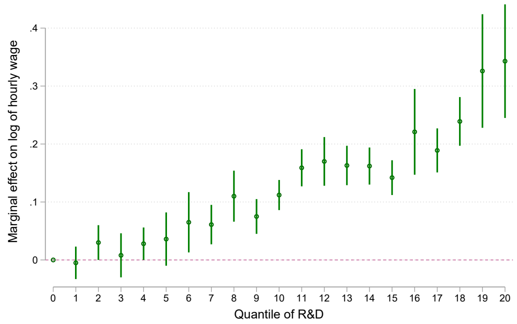
where R_{ftl} is equal to 1 if firm f belongs to quantile l in year t. The estimated coefficients β_{3l} on each of these binary variables are presented graphically in Figure 12, where the reference is the group of firms with no R&D. We see that the coefficients

Figure 12: 20 quantiles of R&D based on level of total R&D expenditures

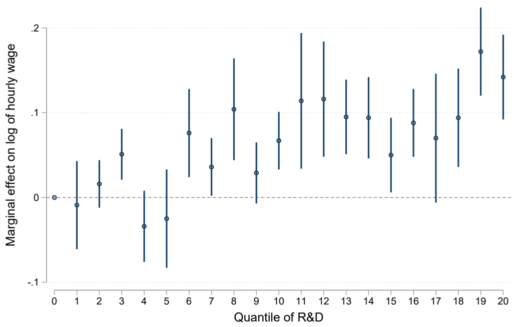
(a) All workers



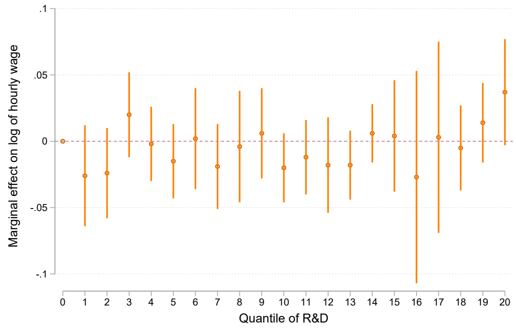
(b) Workers in low-skill occupations



(c) Workers in intermediate-skill occupations



(d) Workers in high-skill occupations



Notes: These Figures reports coefficients on each of the 20 quantiles of total R&D expenditures when estimating equation 5 (β_{3l}). The usual set of control variables are included as well as a set of travel to work area times year fixed effects. Confidence interval at 95% are reported and are computed using heteroskedasticity robust standard errors clustered at the firm level.

are positive and increase with the quantile of R&D for low skill occupations (Figure 12(b)), are positive and significant for the highest quantiles in the case of intermediate-skill occupations (Figure 12(c)) and are never significant in the case of high-skill occupations (Figure 12(d)). Figure 12(a) shows that overall, innovation is associated with higher wages for most quantiles.

6 Conclusion

In this paper, we use matched employee-employer administrative data from the UK, augmented with information on R&D expenditures, to analyze the relationship between wages and innovation. We show that more R&D intensive firms pay higher wages on average, and in particular workers in *some* low-skilled occupations benefit considerably from working in more R&D intensive firms.

We develop a simple model of the firm which generates this finding, one where the complementarity between soft skills of workers in low-skilled occupations with the other firm's assets increases with the firm's degree of innovativeness. Soft skills are an important determinant of the wages of workers in low-skilled occupations – they are a key driver of the complementarity between low-skilled and high-skilled workers – yet they are difficult to observe and verify (both by the firm and the econometrician).

Our model generates a number of additional predictions that we can take to the data. A key additional prediction from the model is that returns to tenure are higher for workers in low-skilled occupations in innovative firms compared with non-innovative firms. We find empirical support for this when we include the triple interaction between tenure, R&D intensity and a measure of the degree complementarity between workers in low-skilled occupations and the other assets of the firm. The model also predicts that workers in low-skilled occupations have longer tenure in more innovative firms than in less innovative firms, which we show is true in our data and that more effort is spent on their training, which is consistent with evidence from O*NET data. A final prediction is that more innovative firms outsource a higher fraction of tasks where there is low complementarity between the workers in low-skilled and high-skilled occupations. Although we do not have great data on outsourcing, we show that evidence from one industry, the cleaning industry, is consistent with this.

While we do not rule out other possible models or explanations for workers in *some* low-skilled occupations earning higher returns in more innovative firms, we argue that our model is an appealing way to explain these findings, and is consistent with a number of other empirical facts.

Our analysis can be extended in several directions. It would be interesting 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 further explore 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. Fourth, we used R&D investment as our measure of innovativeness, and it would be interesting to look at other measures, such as patenting.

Finally, in this work we used data on a 1% sample of employees. It is likely that workers at different parts of the earning distribution and company hierarchy are differentially affected. We could then look at subgroups of agents within the high- and low-skilled occupation categories. In particular, we would expect the premium to working in a more innovative firm to be higher at the very top end of the occupation distribution. A first place to look is at CEOs, taking into account their total revenues (wage income plus capital income). These and other extensions of the analysis in this paper await further research.

References

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

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

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

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

_ **, Ufuk Akcigit, Antonin Bergeaud, Richard Blundell, and David Hémous**, "Innovation and top income inequality," *The Review of Economic Studies*, 2018, *86* (1), 1–45.

_ **,** _ **, Ari Hyytinen, and Otto Toivanen**, "On the returns to invention within firms: Evidence from Finland," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 208–12.

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

Akerman, Anders, Ingvil Gaarder, and Magne Mogstad, "The skill complementarity of broadband internet," *The Quarterly Journal of Economics*, 2015, *130* (4), 1781–1824.

Barth, Erling, Alex Bryson, James C Davis, and Richard Freeman, "ItâĂŹs where you work: Increases in the dispersion of earnings across establishments and individuals in the United States," *Journal of Labor Economics*, 2016, *34* (S2), S67–S97.

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

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

Blundell, Richard, Rachel Griffith, and Frank Windmeijer, "Individual effects and dynamics in count data models," *Journal of econometrics*, 2002, *108* (1), 113–131.

_ **,** _ **, and John Van Reenen**, "Market share, market value and innovation in a panel of British manufacturing firms," *The Review of Economic Studies*, 1999, *66* (3), 529–554.

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, May 2019, *87* (3), 699–739.

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

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

Dhaene, Geert and Koen Jochmans, "Split-panel jackknife estimation of fixed-effect models," *The Review of Economic Studies*, 2015, *82* (3), 991–1030.

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

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

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

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

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

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

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

Goldschmidt, Deborah and Johannes F Schmieder, "The rise of domestic outsourcing and the evolution of the German wage structure," *The Quarterly Journal of Economics*, 2017, p. qjx008.

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

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar, "Who Profits from Patents? Rent-Sharing at Innovative Firms," Working Paper 25245, National Bureau of Economic Research November 2018.

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

_ **and Eric Maskin**, "Wage Inequality and Segregation by Skill," Working Paper 5718, National Bureau of Economic Research August 1996.

Krusell, Per, Lee E Ohanian, José-Víctor Ríos-Rull, and Giovanni L Violante, "Capital-skill complementarity and inequality: A macroeconomic analysis," *Econometrica*, 2000, *68* (5), 1029–1053.

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

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

Oi, Walter Y and Todd L Idson, "Firm size and wages," *Handbook of labor economics*, 1999, *3*, 2165–2214.

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

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter, "Firming up inequality," *The Quarterly Journal of Economics*, 2018, *134* (1), 1–50.

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

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

ONLINE APPENDICES
NOT INTENDED FOR PUBLICATION

A Data construction and additional description

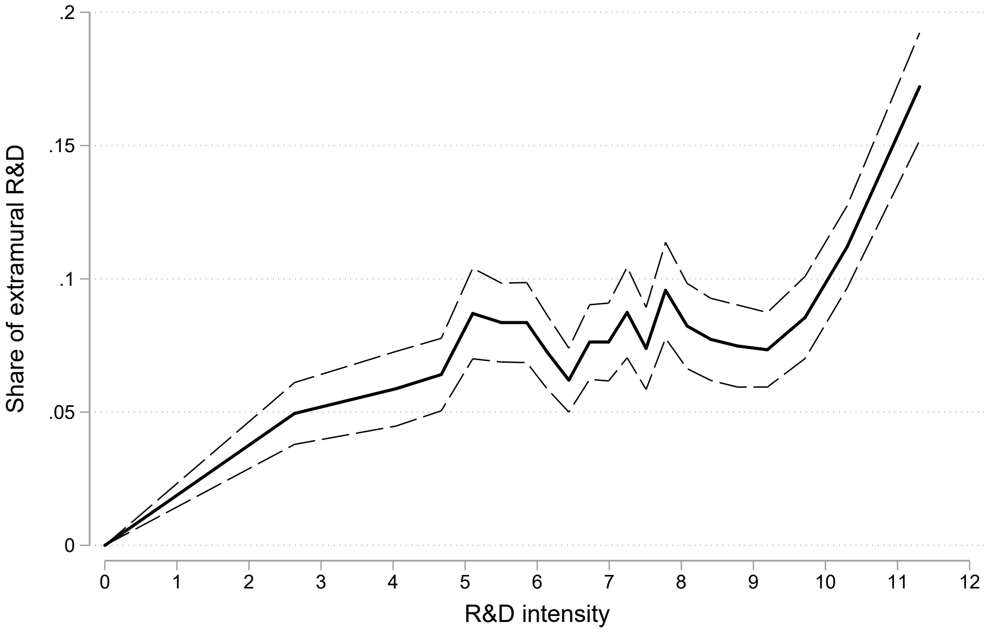
This appendix describes the construction of our main sample which results from the merge of two datasets provided by the ONS: the Annual Survey of Hours and Earnings (ASHE) and the Business Expenditures on Research and Development (BERD).

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 expenditures 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 expenditures 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.[15] The distributions of both intra and extramural R&D are highly skewed, in particular, firms in the highest vintile are very different from others.

[15]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	Total R&D	R&D intensity	Number of firm-years
0 (no R&D)	2,347	0	0	0	0	29,848
1	8,448	83	5	88	2.63	434
2	4,663	245	13	258	4.04	427
3	3,437	361	23	384	4.67	426
4	2,929	428	65	493	5.11	429
5	2,903	672	80	752	5.50	427
6	1,678	542	52	594	5.85	427
7	1,925	824	66	890	6.14	427
8	1,740	1,041	81	1,122	6.44	427
9	1,339	1,013	121	1,134	6.73	427
10	1,592	1,528	192	1,720	6.99	426
11	1,752	2,179	319	2,498	7.25	431
12	1,951	3,165	592	3,757	7.51	427
13	1,597	3,469	407	3,876	7.78	427
14	1,467	4,453	412	4,865	8.08	428
15	1,618	6,813	462	7,275	8.41	424
16	2,609	16,900	819	17,719	8.78	429
17	2,535	23,600	1,333	24,933	9.19	429
18	2,255	34,200	2,742	36,942	9.72	426
19	2,484	64,400	10,300	74,700	10.30	427
20	2,284	143,000	91,700	234,700	11.31	422

Notes: This table presents the average number of employees, average expenditures in intramural R&D (in thousand pounds), average expenditures in extramural R&D (in thousand pounds), the sum of the two and average R&D intensity (defined as the sum of intramural and extramural R&D expenditures per employee) for 20 quantiles of R&D intensity. 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 expenditures (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 expenditures 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 Earnings (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.[16] 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.[17]

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

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

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, 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 5.5.3.[18]

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).[19] 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).[20]

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 trade, 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 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 is based 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 behaviours and different outcomes in the labor market.

[18]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.

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

[20]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 technicians
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 Bachelor's 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	443,002	30.9	62	51	37.6	6.5
Intermediate-skill	113,297	36.3	90	60	39.3	9.4
High-skill	124,420	36.5	95	69	40.8	9.9

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

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

A.4 Matching BERD and ASHE

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

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

Table A5: Construction of the sample

ASHE	Observations	Individuals	Mean wage	Sd wage
Raw ASHE	2,680,485	394,012	12.7	36.5
Clean ASHE	1,988,984	348,996	13.5	14.7
Private Corporations	1,180,002	254,739	13	17
Final Sample	681,253	165,112	13	18.3
BERD	Observations	Firms	% intramural R&D	% extramural R&D
Raw BERD	279,257	57,439	100	100
400+ Employees	9,978	2,032	75	85
Final Sample	9,433	1,995	66	81

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	101	2,378	671
2005	91	2,352	808
2006	92	2,314	955
2007	102	2,343	757
2008	96	2,390	628
2009	73	2,347	800
2010	85	2,304	697
2011	95	2,351	710
2012	97	2,410	782
2013	109	2,449	800
2014	109	2,567	845
2015	123	2,663	892
2016	156	2,605	899

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).

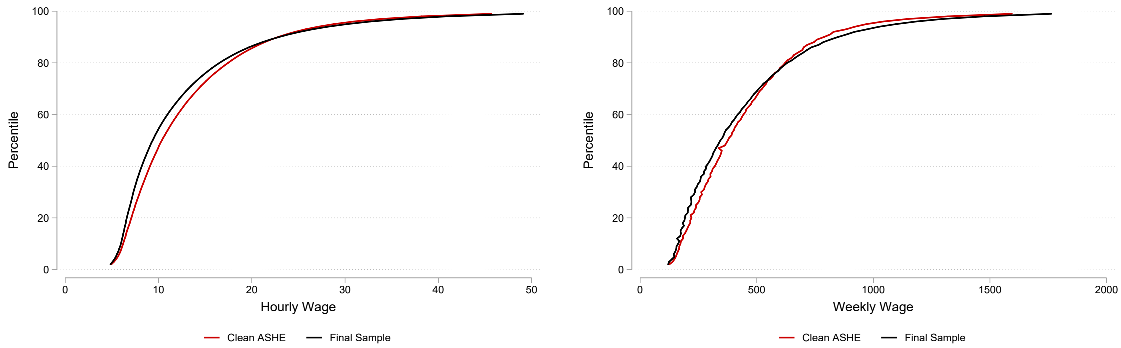
Table A7: Variable description

Variable name	Description
Age	Age of the individual at the time of the survey in year
Tenure	Number of years 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.

Figure A2: Cumulative distribution function of log wage

(a) Hourly Wage (b) Weekly Wage



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

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%)[22], 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.

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

Table A8: Descriptive statistics of wage variables

Variable	Mean	sd	p10	p25	p50	p75	p90	p99
Hourly Wage (£)	14.3	13.5	6.6	7.7	10.4	16.3	25.7	61.4
Weekly Wage (£)	508	480	126	238	394	637	973	2214
Weekly Incentive Pay (£)	8.5	85.5	0	0	0	0	0	184.3
Weekly Overtime Pay (£)	20.2	61.3	0	0	0	3.6	63.2	294
Annual Wage (£)	26,438	43,919	4,673	9,966	19,288	32,396	50,169	136,841
Basic Paid Hours	36.1	11.6	17.6	31.7	40.6	42.9	44.5	58
Paid Overtime	1.7	4.5	0	0	0	0.4	5.8	22
Age	42.3	13.6	24.2	30.8	41.8	52.8	61.6	69.3
Tenure	8.4	8.9	1.1	2.2	5.5	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 years an individual has been working in its current firm.

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

(a) Share (b) Number

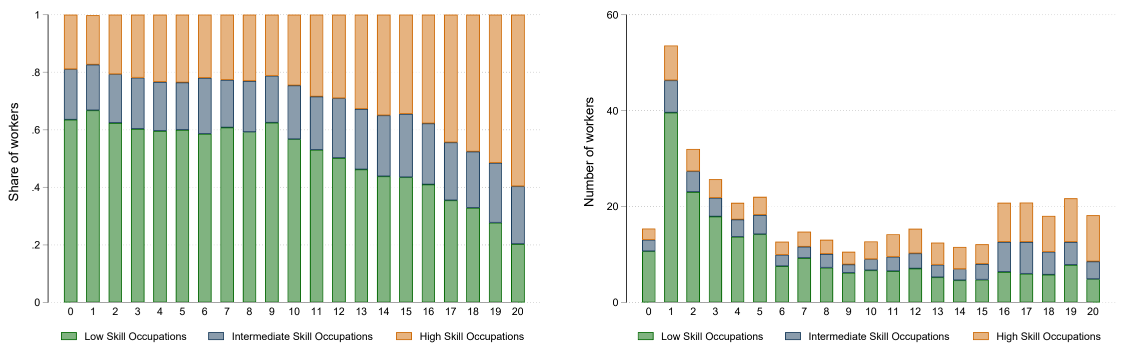


Figure A4 shows that the share of workers that work in a firm that does any R&D increases from just over 20% for workers at the bottom of the wage distribution, to over 55% after the 80^{th} percentile of the distribution where it plateaus. The share falls right at the top, where workers in the (low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

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

Quantile of R&D	Skill category: Low	Intermediate	High	Nb of Firms
0 (no R&D)	64	18	19	32467
1	67	16	17	479
2	62	17	21	470
3	60	18	22	469
4	60	17	23	470
5	60	16	24	472
6	59	19	22	471
7	61	17	23	469
8	59	18	23	471
9	62	16	21	471
10	57	19	25	469
11	53	18	28	473
12	50	21	29	468
13	46	21	33	473
14	44	21	35	470
15	44	22	34	468
16	41	21	38	473
17	35	20	44	471
18	33	20	48	470
19	28	21	51	469
20	20	20	60	466

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

Figure A4: Share of workers in R&D firms at each quantile 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 expenditures since 2000) at each quantile 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[23] and we use the version 21.1 Database - November 2016 Release. The information have been gathered either from interviewing workers or from experts descriptions. Although the O*NET data is only based on US workers, we believe that the job descriptions are very similar to those of the UK. To match the different occupation classification we match O*NET data to UK data via isco08.

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

B Additional Empirical Results

Table B1: Relationship between wages and R&D firm dummy

	Dependent variable: $ln(w_{ijkft})$						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Skill level	All	All	All	Low	Int	High	All
R&D firm	0.121***	0.066***	0.023***	0.023***	0.008	0.001	0.004
	(0.020)	(0.010)	(0.003)	(0.004)	(0.006)	(0.004)	(0.004)
× Int-skill							0.015***
							-0.005
× Low-skill							0.023***
							-0.006
Age	0.059***	0.034***					
	(0.003)	(0.002)					
Age Squared	-0.001***	-0.000***	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***
	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Tenure	0.024***	0.015***	0.008***	0.009***	0.008***	0.001	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Squared	-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	-0.039***	-0.013***	-0.009***	-0.006***	0.001	0.003	-0.008***
	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Gender	0.160***	0.141***					
	(0.006)	(0.004)					
Full-Time	0.255***	0.072***	0.008	-0.008	-0.080***	-0.102***	0.000
	(0.015)	(0.008)	(0.005)	(0.006)	(0.013)	(0.013)	(0.013)
Int-skill dummy							-0.077***
							(0.005)
Low-skill dummy							-0.158***
							(0.007)
Geo-year	✓						
Geo-Occupation-year		✓					
Individual			✓	✓	✓	✓	✓
Year			✓	✓	✓	✓	✓
R^2	0.370	0.621	0.885	0.768	0.847	0.880	0.888
Observations	680,583	669,899	634,542	399,690	100,989	114,953	634,542

Notes: This Table is the same as Table 2 but the R&D intensity variable has been replaced by a dummy equal to 1 if the firm reports some R&D expenditures in the past (R&D firm).

Table B2: Relationship between wages and R&D intensity - control for initial wage

	Dependent variable: $ln(w_{ijkft})$				
	(1)	(2)	(3)	(4)	(5)
Skill level	All	Low	Int	High	All
$\tilde{R}_{ft}$	0.015***	0.019***	0.010***	0.000	0.000
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)
$\times$ Int-skill					0.011***
					(0.001)
$\times$ Low-skill					0.018***
					(0.002)
Age	0.031***	0.020***	0.028***	0.039***	0.023***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.014***	0.015***	0.009***	0.003***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.019***	-0.014***	-0.009*	-0.005	-0.011***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.003)
Gender	0.052***	0.077***	0.057***	0.088***	0.069***
	(0.004)	(0.005)	-0.007	(0.007)	(0.004)
Full-Time	0.154***	0.106***	0.092***	0.030**	0.094***
	(0.008)	(0.009)	(0.024)	(0.012)	(0.007)
Int-skill dummy					-0.327***
					(0.010)
Low-skill dummy					-0.547***
					(0.011)
Initiall Wage	0.592***	0.328***	0.479***	0.536***	0.429***
	(0.010)	(0.011)	(0.010)	(0.010)	(0.009)
Geo-year	✓	✓	✓	✓	✓
R^2	0.616	0.446	0.469	0.534	0.699
Observations	612,871	386,319	107,546	118,403	612,871

Notes: The dependent variable is log of wage which is defined in Appendix A.2.2. $\tilde{R}_{ft} = ln(1 + R_{ft})$ where R_{ft} is total R&D expenditures of firm f during year t divided by employment. Other covariates definitions are given in Table A7. All columns include year-labor market 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.

Table B3: Relationship between wages and R&D firm dummy - control for initial wage

	Dependent variable: $ln(w_{ijkft})$				
	(1)	(2)	(3)	(4)	(5)
Skill level	All	Low	Int	High	All
R&D firm	0.066***	0.064***	0.058***	-0.003	-0.014
	(0.011)	(0.011)	(0.010)	(0.010)	(0.011)
× Int-skill					0.075***
					(0.013)
× Low-skill					0.075***
					(0.015)
Age	0.031***	0.020***	0.028***	0.039***	0.024***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.015***	0.016***	0.010***	0.003***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.022***	-0.018***	-0.011***	-0.005	-0.015***
	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
Gender	0.054***	0.080***	0.058***	0.088***	0.071***
	(0.004)	(0.005)	(0.007)	(0.007)	(0.004)
Full-Time	0.160***	0.111***	0.094***	0.031**	0.101***
	(0.008)	(0.009)	(0.024)	(0.012)	(0.008)
Int-skill dummy					-0.328***
					(0.011)
Low-skill dummy					-0.539***
					(0.011)
Initial Wage	0.601***	0.339***	0.485***	0.536***	0.435***
	(0.010)	(0.011)	(0.010)	(0.010)	(0.009)
Geo-year	✓	✓	✓	✓	✓
R^2	0.612	0.435	0.467	0.534	0.695
Observations	612,871	386,319	107,546	118,403	612,871

Notes: This Table replicates Table B2 but replaces $\tilde{R}_{ft}$ by a binary variable equal to 1 if the firm reports any R&D expenditures in the past.

Table B4: Robustness to alternative fixed effects

	Dependent variable: $ln(w_{ijkft})$				
	(1)	(2)	(3)	(4)	(5)
$\tilde{R}_{ft}$	0.000	0.004*	0.003***	-0.006***	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
$\times$ Int. skill	0.012***	0.007***	0.002***	0.004***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\times$ Low-skill	0.019***	0.010***	0.006***	0.012***	0.004***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
Age	0.023***	0.021***		0.021***	
	(0.001)	(0.001)		(0.001)	
Age Squared	-0.000***	-0.000***	-0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.012***	0.012***	0.008***	0.010***	0.008***
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure Squared	-0.000***	-0.000***	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.011***	-0.004	-0.006***	-0.023***	-0.001
	(0.003)	(0.003)	(0.002)	(0.006)	(0.004)
Gender	0.069***	0.073***		0.067***	
	(0.004)	(0.003)		(0.003)	
Full-Time	0.094***	0.056***	-0.001	0.064***	-0.025***
	(0.007)	(0.006)	(0.005)	(0.006)	(0.005)
Int-skill dummy	-0.327***	-0.332***	-0.077***	-0.266***	-0.061***
	(0.01)	(0.011)	(0.004)	(0.007)	(0.004)
Low-skill dummy	-0.547***	-0.417***	-0.162***	-0.480***	-0.136***
	(0.01)	(0.013)	(0.006)	(0.01)	(0.006)
Initial Wage	0.429***	0.391***		0.369***	
	(0.008)	(0.009)		(0.008)	
Fixed Effects					
Geo-Year	✓				
Geo-Year-Occupation		✓			
Individual			✓		✓
Firm				✓	✓
Year	✓	✓	✓	✓	✓
R^2	0.699	0.726	0.888	0.742	0.896
Observations	612,871	602,056	634,542	612,505	633,899

Notes: The dependent variable is log of wage which is defined in Appendix A.2.2. $\tilde{R}_{ft} = ln(1 + R_{ft})$ where R_{ft} is total R&D expenditures of firm f during year t divided by employment. Other covariates definitions are given in Table A7. Column 1 replicates column 3 in Table 2 and includes year and individual effects. Column 2 includes firm and year effects. Column 3 includes worker, firm and year effects. Heteroskedasticity robust standard errors clustered at the firm level (columns 1, 2 and 3) or individual levels (columns 4 and 5) are reported in parenthesis. ***, ** and * respectively indicate 0.01, 0.05 and 0.1 levels of significance.

Table B5: R&D and hourly wages low-skilled occupations - Robustness

	Dependent variable: $ln(w_{ijkft})$					
	(1)	(2)	(3)	(4)	(5)	(6)
Tenure:	< 5 years	< 5 years	< 5 years	< 5 years	< 5 years	< 5 years
R&D	0.044***	0.035***	-0.008**	0.026**	0.099***	0.010***
	(0.014)	(0.011)	(0.004)	(0.012)	(0.031)	(0.002)
×Med-lambda	0.048***	0.005	-0.001	0.051***	0.026	0.006***
	(0.017)	(0.013)	(0.006)	(0.013)	(0.046)	(0.002)
×High-lambda	0.046*	-0.048**	0.012	0.057**	0.027	-0.008**
	(0.028)	(0.022)	(0.008)	(0.023)	(0.05)	(0.004)
× Tenure	0.001	0.003	0.001	-0.001	-0.009	0.001*
	(0.002)	(0.002)	(0.001)	(0.002)	(0.007)	(0.000)
×Tenure ×Med-lambda	-0.001	-0.001	0.000	0.001	0.003	-0.001
	(0.004)	(0.004)	(0.002)	(0.004)	(0.012)	(0.001)
×Tenure ×High-lambda	0.010*	0.011**	0.004*	0.009*	0.014	0.002**
	(0.005)	(0.005)	(0.003)	(0.005)	(0.010)	(0.001)
Age	0.021***	0.019***	0.016***	0.019***	0.046***	0.019***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.015***	0.021***	0.018***	0.011***	0.902***	0.020***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.021)	(0.002)
× Med-lambda	0.000	-0.001**	-0.001***	0.001	-0.129***	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)
× High-lambda	0.003	0.005**	0.003***	0.001	0.004	0.004**
	(0.002)	(0.002)	(0.001)	(0.002)	(0.005)	(0.002)
Tenure Squared	0.001	-0.008***	0.001	0.001	-0.009	-0.007***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.006)	(0.002)
Med-lambda dummy	0.061***	0.045***	0.027***	0.063***	0.115***	0.041***
	(0.008)	(0.008)	(0.003)	(0.007)	(0.019)	(0.008)
High-lambda dummy	0.098***	0.104***	0.123***	0.095***	0.188***	0.098***
	(0.001)	(0.013)	(0.005)	(0.010)	(0.020)	(0.013)
Firm Size	-0.011***	-0.003	-0.025***	-0.014***	-0.018***	0.000
	(0.004)	(0.005)	(0.002)	(0.003)	(0.005)	(0.005)
Gender	0.063***	0.052***	0.047***	0.035***	0.150***	0.053***
	(0.004)	(0.003)	(0.001)	(0.004)	(0.007)	(0.003)
Full-Time	0.109***	0.071***	0.047***	0.110***	0.817***	0.068***
	(0.009)	(0.006)	(0.001)	(0.007)	(0.015)	(0.006)
Initial Wage	0.224***	0.173***	0.150***	0.201***	0.244***	0.171***
	(0.008)	(0.006)	(0.001)	(0.009)	(0.01)	(0.006)
Fixed Effects						
Geo-Year	✓		✓	✓	✓	✓
Geo-Year-Occupation		✓				
Firm			✓			
R^2	0.325	0.410	0.490	0.337	0.526	0.414
Observations	220,214	212,821	219,588	220,139	217,178	212,821

Notes: Column 1 of this Table is the same as column 1 in Table 5. Column 2 adds 2 digit occupation fixed effects while column 3 adds firm fixed effects to column 1. Column 4 and 5 change the measure of wage in the left-hand side by considering respectively the logarithm of total fixed pay per hours (i.e. when incentive and overtime pays are excluded) and the logarithm of total gross annual revenue. Column 6 uses the logarithm of R&D expenditure over employment as a measure of R&D.

Table B6: Robustness to using different measures of wages

	Dependent variable: $ln(w_{ijkft})$			
Income	Total hourly pay	Fixed hourly pay	Total pay (inc. incentive)	Fixed pay
	(1)	(2)	(3)	(4)
$\tilde{R}_{ft}$	0.001	-0.001	0.000	0.002
	(0.002)	(0.002)	(0.001)	(0.002)
× Int. skill	0.012***	0.009***	0.009***	0.009***
	(0.001)	(0.003)	(0.002)	(0.002)
× Low-skill	0.014***	0.023***	0.018***	0.021***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.021***	0.036***	0.040***	0.036***
	(0.001)	(0.001)	(0.001)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.011***	0.058***	0.014***	0.056***
	(0.000)	(0.002)	(0.001)	(0.002)
Tenure Squared	-0.000***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.014***	-0.015***	-0.013***	-0.013***
	(0.002)	(0.003)	(0.003)	(0.003)
Gender	0.037***	0.139***	0.126***	0.139***
	(0.004)	(0.006)	(0.004)	(0.006)
Full-Time	0.099***	0.790***	0.849***	0.784***
	(0.006)	(0.014)	(0.015)	(0.014)
Int-skill dummy	-0.356***	-0.294***	-0.281***	-0.284***
	(0.009)	(0.011)	(0.009)	(0.010)
Low-skill dummy	-0.564***	-0.537***	-0.494***	-0.504***
	(0.012)	(0.013)	(0.012)	(0.012)
Initial Wage	0.410***	0.457***	0.407***	0.417***
	(0.009)	(0.010)	(0.008)	(0.008)
Fixed Effects				
Geo-Year	✓	✓	✓	✓
R^2	0.709	0.627	0.742	0.615
Observations	612,661	608,851	612,871	608,544

Notes: This table is similar to the last column of Table 2 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 multiplicative travel to work area 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 B7: Testing different functions of R&D

	Dependent variable: $ln(w_{ijkft})$			
R&D function	$log(1+\frac{x}{l})$ (1)	$log(1+x)$ (2)	$\frac{x}{l}$ (3)	$H\left(\frac{x}{l}\right)$ (4)
$\tilde{R}_{ft}$	0.000	-0.002**	0.000*	-0.001
	(0.002)	(0.001)	(0.000)	(0.002)
× Int. skill	0.012***	0.006***	0.000**	0.011***
	(0.001)	(0.001)	(0.000)	(0.001)
× Low-skill	0.019***	0.022***	0.018***	0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Age	0.023***	0.023***	0.023***	0.023***
	(0.001)	(0.001)	(0.001)	(0.001)
Age Squared	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Tenure	0.012***	0.012***	0.012***	0.012***
	(0.001)	(0.001)	(0.001)	(0.001)
Tenure Squared	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Firm Size	-0.011***	-0.011***	-0.011***	-0.011***
	(0.003)	(0.003)	(0.003)	(0.003)
Gender	0.069***	0.069***	0.070***	0.069***
	(0.004)	(0.004)	(0.003)	(0.004)
Full-Time	0.094***	0.093***	0.094***	0.094***
	(0.007)	(0.007)	(0.008)	(0.007)
Int-skill dummy	-0.327***	-0.335***	-0.300***	-0.328***
	(0.01)	(0.01)	(0.009)	(0.01)
Low-skill dummy	-0.547***	-0.555***	-0.540***	-0.548***
	(0.011)	(0.011)	(0.01)	(0.011)
Initial Wage	0.429***	0.429***	0.430***	0.429***
	(0.009)	(0.008)	(0.009)	(0.008)
Fixed Effects				
Geo-Year	✓	✓	✓	✓
R^2	0.699	0.699	0.698	0.699
Observations	612,871	612,871	612,871	612,871

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 2 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 multiplicative travel to work area 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.

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.3 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:[24]

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

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:

$$4C(T - q_L) = \frac{z+1}{z+2}(Q_L - 1)(1 - \bar{\lambda}^{z+2}) + (1 - \bar{\lambda}^{z+1}) - \bar{\lambda}^z(1 - \bar{\lambda})(z + 1)\left(\bar{\lambda}(Q_L - 1) + 1\right) \quad (6)$$

[24]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 (6) yields:

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

whereas in Firm B the outsourcing equation (6) 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 (6) 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.4 Extension to multiple workers in the same task

We now consider the more general case with $n \geq 1$ workers in low-skilled occupations and $m \geq 1$ workers in high-skilled occupations. To determine the equilibrium wages resulting from *ex-post* negotiation, we rely on Stole and Zwiebel (1996). In their framework, if the n^{th} worker in a low-skilled occupation refuses the wage offer w_n^q, then the remaining $n-1$ worker in a low-skilled occupation 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 workers in high- and low-skilled occupations. 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:

$$
\begin{aligned}
w_{n,m}^q(Q,q,\lambda_z) = &\ \lambda_z\frac{q-q_L}{n(n+1)}\sum_{i=0}^{n}\ iQ^m q^{i-1} + \frac{1-\lambda_z}{2}(q-q_L) \\
w_{n,m}^Q(Q,\overline{q},\lambda_z) = &\ \lambda_z\frac{Q-Q_L}{m(m+1)}\sum_{i=0}^{m} iq^n Q^{i-1} + \frac{1-\lambda_z}{2}(Q-Q_L),
\end{aligned}
\tag{7}
$$

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

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