Can On-the-Job Training Bring Back Wage Growth in the UK?*

Robert Jacek Wlodarski †

Advised by Nezih Guner †

Abstract: While government-subsidised training programmes have been widely studied by applied microeconomists, less attention is paid to their macroeconomic effects. In this paper, I study the distributional consequences of a subsidised training programme in a frictional search and matching model with heterogeneous workers and firms, featuring industry dynamics and skill accumulation through on-the-job learning and training. I capture the salient characteristics of the 2017 Apprenticeship Levy policy in the United Kingdom in a framework where the government taxes the largest firms and subsidises on-the-job training costs. A calibration to the UK economy between 2010 and 2016 matches the key earnings, training, inequality, and firm-size distribution moments. I find that the subsidy increases training incidence along extensive and intensive margins, even at smaller firms. The average wage and earnings growth are both boosted but mainly through the positive impact on the largest firms. Examining whether expanding the subsidy to all employers leads to better economic outcomes, I show that it still increases the average wage but worsens the industry's performance.

Keywords: Earnings Inequality ⋄ Labour Market Frictions ⋄ Human Capital ⋄ On-the-Job Trainning ⋄ Productivity ⋄ Firm Size ⋄ Life-Cycle Earnings Profile ⋄ Training Subsidy ⋄ Productivity Puzzle

JEL Classification: E23 ⋄ E24 ⋄ H32 ⋄ J08 ⋄ J24 ⋄ J64

---

*I am grateful to Nezih Guner for outstanding supervision and assistance. I thank Alessandro Ruggieri for help with computation and modelling. I also thank Josep Pijoan-Mas and Fabio Canova for their useful comments and Cemfi's master thesis workshop participants for their invaluable feedback. I acknowledge that my studies, including this project, have been funded by the Cemfi Fellowship. I appreciate Maria Dinu and Anna Chesa's help with proofreading this work.

†Centro de Estudios Monetarios y Financieros (Cemfi)

1 Introduction

Across Europe, the output per worker and wages have barely recovered since the Great Financial Crisis. The stagnation has been especially persistent across the entire earnings distribution and at all education levels in the United Kingdom (UK), as shown by Chiappori et al. (2020). At the same time, the UK has witnessed an across-the-board decline in human capital investment, notably when it comes to on-the-job training (De Lyon and Dhingra, 2020). This lends credibility to the conjecture that lack of investment in skills is one of the sources of the country's "productivity puzzle."

Several general and UK-specific macroeconomic explanations of training underprovision have been proposed. Flinn et al. (2017) argue that firms do not internalise the resulting aggregate human capital gains of on-the-job training. In Lentz and Roys (2024), firms also do not account for future matches' profits resulting from skill investments. Fu (2011) stipulates that on-the-job search discourages firms from providing training. Further, citing economic uncertainty in Britain, De Lyon and Dhingra (2020) point out a steady decline in the provision of on-the-job training, especially in the wake of the crisis and the surprising vote to leave the EU in 2016.

This paper studies the labour market consequences of subsidised training programmes, using the 2017 UK Apprenticeship Levy as a compelling case study in this context. To do so, I expand the framework in Guner and Ruggieri (2021) and leverage it as a quantitative laboratory to analyse how a subsidised training programme impacts wage distribution and life-cycle wage growth. I also apply the model to analyse the impact on the firm size distribution and training decisions. I find that the training subsidy, which targets larger firms, can increase wage growth and average earnings, but expanding the policy to all firms reduces their average size and revenue.

I develop a macroeconomic model with frictional labour markets that captures the salient features of the UK economy. Workers differ by their ex-ante exogenous entry human capital that can be grown via on-the-job learning and training. They search for jobs in a frictional environment, either matching with potential employers or staying non-employed. Firms are heterogeneous along the productivity and training cost dimensions. Workers at the more productive firms enjoy higher earnings and growth. The industry and labour dynamics allow for generating the essential characteristics for an analysis of the subsidy's heterogeneous influence on workers and firms. Calibrated by using the pre-policy data, the model generates the salient worker and firm attributes of the UK economy. In particular, it replicates the size and training incidence distributions on the firms' side of the economy. On the workers' end, it successfully simulates earnings inequality and the wage-size effect outlined by Elsby and Michaels (2013).

I model the policy as a tax on the largest businesses which is spent on providing training subsidies to the same pool of firms. I find that more on-the-job training is provided along extensive and intensive margins, especially by the large subsidised firms and for the workers around the 25th percentile of the human capital distribution. The average firm size and before-tax revenues increase by 0.25 workers and 1.8 per cent, respectively. Yet, this is driven entirely by the improved outcomes at the biggest firms. These changes translate into higher average pay but mostly at said large firms. On the workers' side of the economy, the training subsidy leads to a boost in training incidence along both margins, a 0.8 per cent higher mean wage, and lower earnings inequality. In light of that evidence, I ask whether expanding the subsidy to all employers in the economy leads to better firm results and labour outcomes in the UK. I show that the expanded subsidy mainly benefits smaller and medium-sized businesses. In doing so, it changes the composition of the industry, lowering the average firm size by an average of 2 workers. Despite improving the average wage by almost 1 per cent, this leads to a worse economic performance of firms and concentrates the subsidy expansion's benefits on older workers at the expense of deteriorating prospects for the entrants.

I contribute to three strands of economic literature. First, I build on the burgeoning body of works linking a frictional labour market environment with human capital accumulation and training. The closest study to mine is that of Guner and Ruggieri (2021). They leverage a framework with heterogeneous workers and firms featuring search frictions, on-the-job learning and training, and firm-size distribution to explore the cross-country differences in output and inequality. The strength of their methodology lies in generating the observed facts on training decisions, life-cycle outcomes, and earnings inequality for multiple countries. I expand their modelling strategy by accounting for on-the-job search and using it to evaluate the chosen training subsidy policy. My paper also corresponds with the findings of Flinn et al. (2017), who develop a specification with frictional labour markets and match-specific and general human capital types of investment. They use it to analyse the impact of minimum wage, another policy, on the training environment. My article's novelty is adapting a similar environment to assess the training subsidy's influence on wage distribution in the UK. Other related works are that of Gregory (2020) and Arellano-Bover and Saltiel (2024), both of whom investigate the roles of on-the-job learning and training in driving life-cycle inequality. Similarly, I demonstrate the potential of the policy to reduce earnings inequality in the UK.[1]

Second, I add to the literature researching the causes and consequences of the UK "productivity puzzle." Turrell et al. (2021) and Patterson et al. (2016) run search and matching models to answer if the Great Financial Crisis led to occupational mismatch and the productivity slump in the UK. The former paper uses a novel weighted vacancy dataset of 15 million job adverts to show that there is little evidence to support the conjecture. Like them, I apply a frictional labour markets approach to answer a question related to the UK's economic adversities. In particular, I seek to leverage it to assess if expanding human capital investment can spur wage growth. My article is also related to the agenda of the *IFS Deaton Review of Inequalities* that assembles applied research of labour conditions in the UK (e.g., see Giupponi and Machin (2022)). My innovation lies in developing a general equilibrium

[1] Another notable related work is that of Fu (2011) who investigates wage dispersion in a frictional labour environment with training. More loosely related papers include Lentz and Roys (2024) who study training provision in the context of long-term labour contracts and Bagger et al. (2014) who apply a similar framework to contrast the contribution of job search and human capital accumulation to earnings inequality.

environment to evaluate the steady-state outcomes of a training policy.[2]

Third, I contribute to the labour field that links human capital and market outcomes of individuals. The closest works to mine are that of Bartel (1995) and Dearden et al. (2006). Both use the UK *Labour Force Survey* (LFS) to identify the productivity and wage returns of on-the-job training.[3] My model is estimated with the same dataset to study the training environment in the UK. The novelty of my article is using a general equilibrium framework to evaluate the impact of a training incentive policy. In other pieces, Acemoglu and Pischke (1998, 1999b) argue that when the market is tilted in favour of lower-skilled workers, in the presence of informational asymmetry or skill heterogeneity, firms have an incentive to provide on-the-job training.[4] My input lies in studying the potential of training subsidy policies to correct training underprovision.

The article is structured as follows. Section 2 introduces the 2017 Apprenticeship Levy. Section 3 lays out the modelling approach, focusing on the labour market environment, the problems of workers and firms, and the determinants of training, matching, vacancy posting, and entry policy functions. Section 4 presents the estimation strategy. In Section 5, I show the simulated impact of the training subsidy policy. Section 6 documents the potential subsidy extension to all firms. Section 7 concludes.

2 The 2017 Apprenticeship Levy

Introduced in 2016 and implemented in 2017, the Apprenticeship Levy constitutes a relevant case study for assessing the macroeconomic effects of training programmes. The policy distinguishes two categories of firms based on size and payroll. Those with over 50 employees or £3m in annual payroll are considered large. They face a 0.5 per cent extra payroll tax, referred to as the apprenticeship levy. Its proceeds can fund 90 per cent of apprenticeship costs for a worker while the government tops up the remaining 10 per cent (ESFA, 2017a). Small firms do not pay the levy, but the only benefit they receive is a 90 per cent subsidy for the apprentices between 16 and 19.[5] Further details can be found in Appendix A.

The policy has three key characteristics from the point of view of economic theory. First, it

[2] Other UK research and policy institutions also inquire into the human capital investment in the UK "productivity puzzle's" context. Notable examples include the *Institute for Government* (Pope et al., 2022), the LSE's *Centre for Vocational Education Research* (Britton et al., 2020), the *Social Mobility Commission* Battiston et al. (2020), and the *Resolution Foundation* (Resolution Foundation, 2023).

[3] Their findings also echo newer quasi-experimental and RCT-based evidence from Portugal (Almeida and Carneiro, 2009; Martins, 2021), Belgium (Konings and Vanormelingen, 2015), Italy (Conti, 2005), the developing countries (Attanasio et al., 2011; Almeida and Faria, 2014; Bandiera et al., 2023) and more. For more, see the meta-analysis of Card et al. (2018).

[4] For further review of similar works, see Acemoglu and Pischke (1999a).

[5] This effectively means the small firm subsidy is aimed at school leavers who do not pursue university degrees. Modelling the small-firm aspect of the policy would require endogenising the higher education decision, which is beyond the scope of the model.

targets only the larger and, hence, more productive firms by providing a clear size cut-off.[6] Second, the tax can be redeemed if spent on an apprenticeship. This potentially corrects the inefficiency resulting from firms' underprovision of training outlined by Flinn et al. (2017). Third, given that not all firms use their levy to cover the costs of apprenticeships, the policy induces a cross-subsidy from the low to high training cost firms. The scarce empirical evidence corroborates the story. Battiston et al. (2020) present an exploratory analysis of the consequences of the introduction of the apprenticeship levy. They document a larger increase in apprenticeship take-up among those from better-off areas than their counterparts from the most deprived regions. This trend is shown to have been driven by the growth in the apprenticeship provision in large firms. Conversely, the incidence rate fell among small and medium firms, which are already more likely to pay their apprentices the minimum wage.

3 Modelling Labour Markets and Training

In this section, I develop a model that is suitable for studying the consequences of training programmes on the labour market and captures these three characteristics. The framework features a closed economy with three types of agents: workers, firms, and the government. There is a unit measure of infinitely lived workers who face a constant probability of death in each period. They differ in their initial level of human capital, representing general skill or ability. In each period, they are either employed or non-employed. While employed, a worker can increase her skills via on-the-job learning and training. She also searches for alternative jobs, and if she prefers the new job, she moves. Conversely, a non-employed worker seeks a job and faces the risk of skill depreciation. Searching for jobs, workers are subject to labour market frictions. On the other side of the economy, there are firms which are heterogeneous along two dimensions: productivity and training costs. These qualities determine the size and skill distribution of their workforces. In each period, firms decide on whether to enter the market and open new vacancies. Finally, the government taxes firms' revenues and subsidises on-the-job training while balancing the budget.

3.1 Labour Market Environment

First, consider workers. They maximise the present discounted value of their lifetime linear utility:

$$U = \sum_{t=0}^{\infty} \left(\frac{1-\delta_w}{1+r} \right)^t c_t, \tag{3.1}$$

where c_t is period t's consumption, δ_w represents the probability of dying or retiring, and $\frac{1}{1+r}$ acts as the discount rate. Similar to the frameworks in Guner and Ruggieri (2021) and Flinn et al. (2017),

[6] The notion that larger firms are more productive is present in multiple strands of macroeconomic literature. Bento and Restuccia (2017) highlight that increasing an average boosts aggregate productivity. Hopenhayn (2014) provides a review of similar works. On top of that, Brown and Medoff (1989) and Elsby and Michaels (2013) document the wage-size effects in the labour markets.

workers are heterogeneous in their ex-ante human capital, which is denoted by $a \in \mathscr{A} \subset \mathbb{R}_+$ and distributed according to density $\psi_a(\cdot)$. While employed, they engage in on-the-job learning and training, each of which raises their skill level to $a + \Delta_a \in \mathscr{A}$ at rates p^e and p^t, respectively. Moving to non-employment or between jobs, workers retain their human capital.[7] While remaining non-employed, they face a Δ_a skill depreciation taking place at rate p_d.

The industry consists of an endogenously determined measure of firms. They are heterogeneous along two dimensions. First, ahead of entering the market, each firm draws their constant productivity, represented by $z \in \mathscr{Z} \subset \mathbb{R}_+$ and with support $\psi_z(\cdot)$. Second, they face different training costs, $\xi \in \mathscr{E} \in \mathbb{R}_+$, distributed according to $\psi_\xi(\cdot)$. The output of firm (z, ξ) is given by:

$$y(z,\xi,l,\psi) = \int_0^l g(z,a_i)\,\psi(a_i|z,\xi,l)\,\mathrm{d}i, \tag{3.2}$$

where $g(z,a_i) = z a_i$ is the value added produced by a match of firm z and worker a_i. $\psi(a_i|z,\xi,l)$ is the measure of workers with ability a_i at firm (z,ξ) with l employees. It follows that the value added is linear in l:

$$y(z,\xi,l,\psi) = z\overline{a}\big(z,\xi,l,\psi\big)\,l, \tag{3.3}$$

where

$$\overline{a}\big(z,\xi,l,\psi\big) \equiv \int_{a\in\mathscr{A}} a\psi(a|z,\xi,l)\,\mathrm{d}a. \tag{3.4}$$

This implies that each firm seeks to hire as many workers as possible, a process which is only constrained by the entry fee, convex hiring costs, and matching frictions.[8] Given that the linearity result effectively implies that firms treat workers as independent production units, hiring and training decisions, as well as wage bargaining, take place at the match level. Finally, firms face two types of exogenous shocks. They see their matches randomly terminated with probability δ_s and exit the market at rate δ_f.

Finally, labour market frictions ensure the presence of involuntary non-employment and non-perfect assortative matching. Building on Diamond (1982), Mortensen (1982), and Pissarides (1985), firms post vacancies at a non-zero cost to match with workers. Both employed and non-employed workers search for jobs at no cost. There are N non-employed workers, E employed workers, and V

[7] While Flinn et al. (2017) splits human capital into general ability and match-specific skills, I opt to model only the general ability.

[8] These two phenomena lead to a non-degenerate distribution of firm size. An alternative method of achieving differing firm sizes is to impose a production function with a decreasing marginal product of labour, akin to the framework in Elsby and Michaels (2013). This approach, however, depends on the intra-firm bargaining protocol of Stole and Zwiebel (1996) which breaks down with human capital heterogeneity.

vacancies in each period.

I impose a constant returns to scale matching function from Den Haan et al. (2000):[9]

$$m(S,V)=\frac{SV}{(S^{\eta}+V^{\eta})^{\frac{1}{\eta}}}, \tag{3.5a}$$

where η is the parameter determining the efficiency of matching. S represents the total measure of searchers:

$$S\equiv U+\gamma E, \tag{3.5b}$$

where γ governs the relative effort of search of employed workers (Flinn et al., 2017). If one sets $\gamma=0$, the model collapses to that with no on-the-job search in Guner and Ruggieri (2021).

Defining $\theta\equiv\frac{V}{S}$ as the labour market tightness, it follows that the contact rates for firms, non-employed workers, and employed workers are:

$$\phi_f=(1+\theta^{\eta})^{-\frac{1}{\eta}}, \tag{3.6a}$$

$$\phi_w^u=\left(1+\theta^{-\eta}\right)^{-\frac{1}{\eta}}, \tag{3.6b}$$

and

$$\phi_w^e=\gamma\left(1+\theta^{-\eta}\right)^{-\frac{1}{\eta}}=\gamma\phi_w^u. \tag{3.6c}$$

Once matched, worker a and firm (z,ξ) bargain over wage $w(z,\xi,a)$. If the match does not materialise or is not accepted by either party, the worker receives b, the flow payoff from non-employment.

3.2 Workers

Consider a non-employed worker with skill a. At the beginning of the period, her value function is:

$$\begin{aligned} J^u(a)=&\left(1-\phi_w^u\right)\left[p^d J^{u,h}(a-\Delta_a)+\left(1-p^d\right)J^{u,h}(a)\right]+\\ &\phi_w^u\iint\limits_{(z,\xi)\in\mathscr{Z}\times\mathscr{E}}\left\{J^{u,h}(a)+\mathbb{I}^h(z,\xi,a)\left[J^{e,h}(z,\xi,a)-J^{u,h}(a)\right]\right\}\psi_v(z,\xi)\,\mathrm{d}\xi\,\mathrm{d}z. \end{aligned} \tag{3.7}$$

The first component represents the expected value of remaining non-employed at the end of the period. Note that the worker accounts for the probability of her ability deteriorating by Δ_a in $p^d J^{u,h}(a-\Delta_a)$. This reduces the value of non-employment by adding the new "downside" of waiting - the possibility of a drop in one's human capital. The expression's second component denotes the expected

[9] Alternatively, one could use the Cobb-Douglas functional form: $m(U,V)=m_0U^{\alpha}V^{1-\alpha}$. It, however, features a major drawback by allowing for contact rates outside the $[0,1]$ interval (Petrongolo and Pissarides, 2001).

value of matching with firms from the distribution of vacancies, $\psi_v(z,\xi)$. If the match is formed, $\mathbb{I}^h(z,\xi,a)=1$, then the worker receives the end-of-period value from being employed at firm (z,ξ), denoted by $J^{e,h}(z,\xi,a)$. As it is formalised later, the match is formed when the end-of-period value of employment exceeds that of non-employment. In the case of remaining non-employed, her end-of-period value is:

$$J^{u,h}(a)=b+\frac{1-\delta_w}{1+r}J^u(a), \tag{3.8}$$

where $1-\delta_w$ stands for the probability of staying in the labour market (by not retiring or dying).

An employed worker accounts for three labour flows: staying in her current job, being poached by another firm, and moving to non-employment. The value of employment for worker a at firm (z,ξ) at the beginning of the period is:

$$\begin{aligned} J^e(z,\xi,a)=\phi_w^e\mathbb{I}^h(z,\xi,a)\iint\limits_{(z,\xi)\in\mathscr{Z}\times\mathscr{E}} \Big\langle J^{e,h}(z,\xi,a)+\mathbb{I}^h\big(z',\xi',a\big)\max\big\{J^{e,h}\big(z',\xi',a\big)- \\ J^{e,h}(z,\xi,a),0\big\}\Big\rangle\psi_v(z,\xi)\,\mathrm{d}\xi\,\mathrm{d}z+\big(1-\phi_w^e\big)\mathbb{I}^h(z,\xi,a)\,J^{e,h}(z,\xi,a)+\big[1-\mathbb{I}^h(z,\xi,a)\big]\,J^{u,h}(a). \end{aligned} \tag{3.9}$$

The first element denotes the expected value of receiving an alternative job offer from firm (z',ξ'). For this to happen, the current match must continue, $\mathbb{I}^h(z,\xi,a)=1$, and the worker must receive a new offer, which takes place at rate ϕ_w. New offers arrive from the distribution of vacancies, $\psi_v(\cdot,\cdot)$. The worker transitions to the competing employer only if she is accepted, $\mathbb{I}^h(z',\xi',a)=1$, and the offer is appealing enough, $J^{e,h}(z',\xi',a)\geq J^{e,h}(z,\xi,a)$. The formula's second component refers to the expected value of not receiving an alternative offer and remaining in the current firm. The last item accounts for the probability of becoming non-employed.

The end-of-period value function of working at firm (z,ξ) is:

$$\begin{aligned} J^{e,h}(z,\xi,a)=w(z,\xi,a)+\left(\frac{1-\delta_w}{1+r}\right)\big[\delta_f+\big(1-\delta_f\big)\delta_s\big]\,J^{u,h}(a)+ \\ \left(\frac{1-\delta_w}{1+r}\right)\big\{1-\big[\delta_f+\big(1-\delta_f\big)\delta_s\big]\big\}\big\{J^e(z,\xi,a)+p^h(z,\xi,a)[J^e(z,\xi,a+\Delta_a)-J^e(z,\xi,a)]\big\}. \end{aligned} \tag{3.10}$$

The first element references the wage she earns in that period. The second block accounts for the exogenous “routes” to non-employment: the match termination and firm exit. Conditional on staying in the job, the worker also accounts for the possibility of a skill upgrade, which is given by:

$$p^h(z,\xi,a)=p^e+\mathbb{I}^t(z,\xi,a)\,p^t. \tag{3.11}$$

The presence of on-the-job training, $\mathbb{I}^t(z,\xi,a)=1$, increases the value of remaining employed. This implies that lower-paying jobs can be preferred if they entail more training and, consequently, human capital accumulation in the future.

3.3 Firms

Consider firm (z,ξ) matched with worker a. The value of the match at the beginning of the period is:

$$V(z,\xi,a)=\mathbb{I}^h(z,\xi,a)\big[1-\phi_w^e\epsilon(z,\xi,a)\big]V^h(z,\xi,a), \tag{3.12}$$

which represents the end-of-period value conditioned on not losing the match. This can happen via endogenous termination, $\mathbb{I}^t(z,\xi,a)=0$, or seeing the worker poached. $\epsilon(z,\xi,a)$ shows the proportion of workers with skill a leaving to the firm's competitors:

$$\epsilon(z,\xi,a)\equiv\iint\limits_{(z',\xi')\in\mathscr{Z}\times\mathscr{E}}\mathbb{I}^h(z,\xi,a)\times\mathbb{I}\left\{J^{e,h}\left(z',\xi',a\right)\geq J^{e,h}(z,\xi,a)\right\}\psi_v(z,\xi)\,\mathrm{d}\xi\,\mathrm{d}z. \tag{3.13}$$

Assuming no non-labour production costs, the gross revenue of a match between firm (z,ξ) and worker a, the revenue is given by: $r(z,a)=za$. This yields the end-of-the-period match value of:

$$\begin{aligned}&V^h(z,\xi,a)=[1-\tau(z,\xi,a)]\,r(z,a)-w(z,\xi,a)+\\&\left(\frac{1-\delta}{1+r}\right)\left\{-\mathbb{I}^t(z,\xi,a)\xi[1-\lambda(z,\xi,a)]+V(z,\xi,a)+p^h(z,\xi,a)[V(z,\xi,a+\Delta_a)-V(z,\xi,a)]\right\},\end{aligned} \tag{3.14}$$

where $\mathbb{I}^t(z,\xi,a)$ is the firm's training decision.[10] $[1-\tau(z,\xi,a)]\,r(z,a)-w(z,\xi,a)$ represents the firm's after-tax flow profit. $\tau(z,\xi,a)$ is the tax levied on the firm's revenue. The second component consists of the training cost, subject to proportional subsidy $\lambda(z,\xi,a)$, and the potential benefit of seeing worker a increase their skill within the period. The value is discounted by:

$$\delta\equiv\delta_w+(1-\delta_w)\delta_s+(1-\delta_w)(1-\delta_s)\delta_f, \tag{3.15}$$

which is the aggregate probability of the match being exogenously terminated.

This means that the firm's decision to post a vacancy depends on both its productivity and training costs, z and ξ. Assume that all firms face the same convex cost function of advertising a vacancy:

[10] Unlike in Flinn et al. (2017), I model training incidence as a binary event. The motivation is twofold. First, doing so makes the framework more tractable and reduces the dimension of the vector of parameters that needs to be estimated. Second, I use the UK *Labour Force Survey*. The dataset contains information on whether a worker receives on-the-job training in a given quarter, but the training length is often not included (Office For National Statistics, 2023). For example, 6,737 private sector workers report non-negative earnings and do not have their employment length value missing from the dataset between 2010 and 2016. In 1,141 cases, the training length variable is blank. Out of the remaining values, 2,809 respondents report *Does Not Apply* and 16 refuse to answer. Even where it is present, it belongs to the category of variables likely suffering from the "recall bias." For the discussion of the "recall bias" in the LFS, see Postel-Vinay and Sepahsalari (2023).

$$c(v) = \frac{v^{\lambda_1}}{\lambda_1}, \tag{3.16}$$

where v is the number of vacancies and $\lambda_1 > 1$ is the convexity parameter. Firm (z, ξ) maximises the total value of new hires, subject to the cost function:

$$\begin{aligned} \pi(z,\xi) = \max_{v(z,\xi)\geq 0} \Bigg\langle -c\left[v(z,\xi)\right] + \phi_f \Bigg[\frac{U}{S} \int\limits_{a\in\mathscr{A}} \mathbb{I}^h(z,\xi,a) V^h(z,\xi,a) \psi_a^u(a) \mathrm{d}a + \\ \frac{\gamma E}{S} \iiint\limits_{(z',\xi',a)\in\mathscr{Z}\times\mathscr{E}\times\mathscr{A}} \mathbb{I}^h(z,\xi,a) \mathbb{I}\left\{J^{e,h}(z,\xi,a) > J^{e,h}\left(z',\xi',a\right)\right\} V^h(z,\xi,a) \psi\left(z',\xi',a\right) \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}a \Bigg] \Bigg\rangle. \end{aligned} \tag{3.17}$$

Unlike in the optimisation problem without on-the-job search in Guner and Ruggieri (2021), the firm now differentiates between those hired from the stock of non-employed workers and those poached from other firms, distributed according to $\psi_a^u(a)$ and $\iint \psi(z',\xi',a) \mathrm{d}z\, \mathrm{d}\xi$, respectively. The first-order condition leads to the following vacancy-posting policy function:

$$\begin{aligned} v(z,\xi) = \Bigg\langle \phi_f \Bigg[\frac{U}{S} \int\limits_{a\in\mathscr{A}} \mathbb{I}^h(z,\xi,a) V^h(z,\xi,a) \psi_a^u(a) \mathrm{d}a + \frac{\gamma E}{S} \iiint\limits_{(z',\xi',a)\in\mathscr{Z}\times\mathscr{E}\times\mathscr{A}} \mathbb{I}^h(z,\xi,a) \\ \mathbb{I}\left\{J^{e,h}(z,\xi,a) > J^{e,h}\left(z',\xi',a\right)\right\} V^h(z,\xi,a) \psi\left(z',\xi',a\right) \mathrm{d}z\, \mathrm{d}\xi\, \mathrm{d}a \Bigg] \Bigg\rangle^{\frac{1}{\lambda_1 - 1}}. \end{aligned} \tag{3.18}$$

Note that the presence of $\mathbb{I}\left\{J^{e,h}(z,\xi,a) > J^{e,h}(z',\xi',a)\right\}$ implies that the high-productivity and low-cost firms are more likely to poach employees from their competition. As a result, workers in the model move up the job ladder towards firms with higher z and lower ξ.

Finally, I introduce the entry cost that, along with search frictions and vacancy-posting costs, disciplines the firm size distribution in the equilibrium. I assume that there is a constant measure of entrants, N_e, who draw productivities and training costs from two independent distributions: $\psi_z(\cdot)$ and $\psi_\xi(\cdot)$. Each of the entrants then decides on whether to access the market, which happens when they can recover the entry cost, c_e:

$$\Pi(z,\xi) \geq c_e, \tag{3.19a}$$

where

$$\Pi(z,\xi) \equiv \sum_{t=0}^{\infty} \left(\frac{1-\delta_f}{1+r}\right)^t \pi(z,\xi) = \left(\frac{1+r}{r+\delta_f}\right) \pi(z,\xi) \tag{3.19b}$$

and $\pi(\cdot,\cdot)$ is given by Equation (3.17). Note that the search frictions and non-zero option value of search imply that there exists threshold $(z^\star, \xi^\star)$ that determines the entry policy function's contour in $\mathscr{Z} \times \mathscr{E} \subset \mathbb{R}^2_+$.[11]

3.4 Hiring, Wage Bargaining, and Training

Let $S(z,\xi,a)$ and $S^h(z,\xi,a)$ represent the beginning and end-of-period match surplus value functions. Each represents the joint match value to the worker-firm pair, net of the value of non-employment. Built from the end-of-period value functions for workers and firms, the end-of-period surplus accounts for the exogenous processes affecting the match. Both functions are expressed as:

$$S(z,\xi,a) = J^e(z,\xi,a) + V(z,\xi,a) - J^u(a) = M(z,\xi,a) - J^u(a) \tag{3.20a}$$

and

$$S^h(z,\xi,a) = J^{e,h}(z,\xi,a) + V^h(z,\xi,a) - J^{u,h}(a) = M^h(z,\xi,a) - J^{u,h}(a). \tag{3.20b}$$

$M(\cdot,\cdot,\cdot)$, the joint match value, is the sum of the value of employment and the match value for the firm. Appendix B.1 outlines the full relationship between the beginning and end-of-period joint match values and surplus functions.

Upon meeting, workers and firms determine whether to form a match. A worker wishes to accept or continue the match with firm (z,ξ) when $J^{e,h}(z,\xi,a) \geq J^{u,h}(a)$. Firm (z,ξ) keeps the match when $V^h(z,\xi,a) \geq 0$. Combining both conditions indicates that the match's formation (or continuation) depends on the surplus:

$$\mathbb{I}^h(z,\xi,a) = \begin{cases} 1 & S^h(z,\xi,a) \geq 0 \\ 0 & S^h(z,\xi,a) < 0. \end{cases} \tag{3.21}$$

Wages are determined in Nash bargaining between employers and employees. Worker a and firm (z,ξ) negotiate to maximise the Nash product:

$$\max_{w(z,\xi,a)} \left\{ \left[J^{e,h}(z,\xi,a) - J^{u,h}(a) \right]^{\beta} V^h(z,\xi,a)^{1-\beta} \right\}, \tag{3.22}$$

where $\beta \in (0,1)$ represents the workers' bargaining power. The first-order condition yields the following end-of-the-period value of being employed:

$$J^{e,h}(z,\xi,a) = J^{u,h}(a) + \beta S^h(z,\xi,a). \tag{3.23}$$

A similar process takes place for the training decision. Each worker-firm pair maximises the

[11] Firm (z,ξ) enters the market if $z \geq z^\star$ and $\xi \leq \xi^\star$.

joint match value:

$$\mathbb{I}^t(z,\xi,a) = \arg\left\langle \max_{\mathbb{I}^t(z,\xi,a)\in\{0,1\}} \mathbb{I}^t(z,\xi,a)\left\{p^t[M(z,\xi,a+\Delta_a)-M(z,\xi,a)]-\xi[1-\lambda(z,\xi,a)]\right\}\right\rangle. \tag{3.24}$$

This means that on-the-job training takes place when:

$$\mathbb{I}^t(z,\xi,a) = \begin{cases} 1 & p^t[M(z,\xi,a+\Delta_a)-M(z,\xi,a)] \geq \xi[1-\lambda(z,\xi,a)] \\ 0 & p^t[M(z,\xi,a+\Delta_a)-M(z,\xi,a)] < \xi[1-\lambda(z,\xi,a)]. \end{cases} \tag{3.25}$$

Like in Guner and Ruggieri (2021) and Flinn et al. (2017), this constitutes the source of aggregate inefficiency in the economy. That is, firms only internalise the positive impact of training on the human capital distribution within the match itself.[12]

3.5 Government

The government runs a training subsidy programme. To mimic the UK's training policy, I set a productivity-dependent revenue tax: $\tau(z) = \tau\mathbb{I}\{z \geq \hat{z}\}$, where $\tau \in (0,1)$.[13] $\hat{z}$ is the productivity threshold chosen so that the tax affects only the firms of 50 and more employees. The government uses the tax proceeds to fund proportional training subsidies for the same firms: $\lambda(z) = \lambda\mathbb{I}\{z \geq \hat{z}\}$, where $\lambda \in (0,1)$. λ is chosen so that the budget in every period remains balanced:

$$\iiint_{(z',\xi',a)\in\mathscr{Z}\times\mathscr{E}\times\mathscr{A}} \tau(z)r(z,a)\psi(z,\xi,a)\,\mathrm{d}z\,\mathrm{d}\xi\,\mathrm{d}a = \iiint_{(z',\xi',a)\in\mathscr{Z}\times\mathscr{E}\times\mathscr{A}} \lambda(z)\xi\mathbb{I}^t(z,\xi,a)\psi(z,\xi,a)\,\mathrm{d}z\,\mathrm{d}\xi\,\mathrm{d}a. \tag{3.26}$$

Given the presence of $\mathbb{I}^t(z,\xi,a)$, some worker-firm pairs always pay the tax, but do not use the subsidy. Combined with the balanced budget condition and the proportional nature of the subsidy, this means that the firms with higher training costs that choose to provide on-the-job training to their workers are net beneficiaries of the government's intervention.

In Appendices B.2 and B.3, I define the recursive competitive equilibrium and outline the algorithm used to solve this model, respectively.

[12] By the same token, Lentz and Roys (2024) stipulate that firms do not internalise the future matches' profits in their decision-making. Fu (2011) highlights that firms under-provide training in the presence of on-the-job search. Finally, Martins (2021) provides a similar argument based on the study of the ESF grant applications in Portugal. The author highlights that the relatively low estimated cost of training hints at the human capital investment underprovision.

[13] The algorithm's solution depends on $w(z,\xi,a)$ being "cancelled out" from Equation (B.4). This is why, I impose the tax on revenue and choose the tax level to correspond to the same level as targeting 0.5 per cent of payroll at the firm.

4 Bringing the Model to UK Data

I parameterise the model to correspond to the UK economy between 2010 and 2016, before the introduction of the Apprenticeship Levy. The model period is a quarter. All data sets used in the calibration are outlined in Appendix C.1. The benchmark framework matches the targeted earnings and training moments for workers well. It also does an excellent job of matching firm size and training distributions. The ability of the model to match both worker and firm side moments is essential for an analysis of the subsidy's heterogeneous influence on workers and firms in the UK.

4.1 Model Parameters

I directly set four parameters based on their data counterparts or other works. First, the interest rate is directly calibrated to match a 4 per cent annual return rate $r = 0.0033$. Second, the worker exit rate is calibrated to correspond to the average tenure in the labour force of 40 years, $\delta_w = 0.0099$. Third, the firm exit rate mirrors the annual firm exit rate of 10.5 per cent. Finally, the elasticity of the matching function is $\eta = 0.5416$ based on the GMM estimation in Guner and Ruggieri (2021).[14]

The remaining parameters are estimated using the method of simulated moments. To do so, I assume that the initial human capital and firm productivity are drawn from two log-normal distributions:

$$\log a \sim \mathcal{N}(0, \sigma_a) \tag{4.1a}$$

and

$$\log z \sim \mathcal{N}(0, \sigma_z) \tag{4.1b}$$

which requires two variance terms, σ_a and σ_z, to be estimated. Further, I impose the uniform distribution on the training cost:

$$\xi \sim \mathcal{U}\left(\underline{\xi}, \overline{\xi}\right), \quad \text{where } 0 < \underline{\xi} < \overline{\xi}. \tag{4.2}$$

Given these functional assumptions, there are 14 parameters to be estimated:

$$\boldsymbol{\vartheta} \equiv \left[c_e, \sigma_z, \underline{\xi}, \overline{\xi}, \lambda_1, N_e, \beta, \sigma_a, p^d, p^e, p^t, b, \delta_s, \gamma\right]^T. \tag{4.3}$$

[14] Turrell et al. (2021) estimate the Cobb-Douglas matching function's elasticity parameter, $\alpha \equiv \frac{\partial m(S,V)}{\partial V} \frac{V}{m(S,V)}$, for different levels of industry aggregation with IV and OLS using the ONS LFS-based data and the daily vacancies data from *Reed*, a popular job-posting website. Their estimates range from 0.420 to 0.526 for the monthly model. Plugging Guner and Ruggieri (2021)'s estimate of η to Equations (3.6a), (3.6b), and (3.6c) and re-computing the same values for Turrell et al. (2021)'s Cobb-Douglas functional form and estimates yields nearly identical arrival rates as functions of the labour market tightness, θ. Patterson et al. (2016) find similar results using the OLS and the *JobCentrePlus* vacancy data.

Table 4.1: Estimated Parameters $\hat{\boldsymbol{\vartheta}}$.

Meaning	Parameter	Value
Non-Employment Payoff	b	21.1977
Entry Cost	c_e	38.3073
Lower Bound of Training Cost	$\underline{\xi}$	2.7971
Higher Bound of Training Cost	$\overline{\xi}$	26.4656
Dispersion of Initial Human Capital	σ_a	1.2109
Dispersion of Firm Productivity	σ_z	0.99595
Experience Jump (Rate)	p^e	0.23855
Training Jump (Rate)	p^t	0.04888
Depreciation Jump (Rate)	p^d	0.43017
Bargaining Power of Workers	β	0.40769
Hiring Cost, Convexity	λ_2	2.3474
Exogeneous Match Separation (Rate)	δ_s	0.0032421
Search Efficiency of Employed Workers	γ	0.56152
Measure of Potential Entrants	N_e	0.024846

Notes: The parameters are estimated using the method of simulated moments, by minimising Equation (4.4) with the use of the genetic algorithm.

Letting $\boldsymbol{M}_K$ denote the K observed moments based on the 2010-2016 UK data and $\tilde{\boldsymbol{M}}(\boldsymbol{\vartheta})$ be the model-originating counterparts, the vector of parameter values is given by:

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}\in\boldsymbol{\Theta}} \left\{ \left[\boldsymbol{M}_K - \tilde{\boldsymbol{M}}(\boldsymbol{\vartheta})\right]^T \boldsymbol{W_K} \left[\boldsymbol{M}_K - \tilde{\boldsymbol{M}}(\boldsymbol{\vartheta})\right] \right\}, \tag{4.4}$$

where $\boldsymbol{W_K}$ is a positive semi-definite weighting matrix. Following Lise et al. (2016) and Guner and Ruggieri (2021), I pick the identity matrix. $\boldsymbol{\Theta}$ is a 14-dimensional space representing the permissible values of parameters. Table 4.1 presents the estimated parameters.

4.1.1 Worker Parameters

Consider the parameters governing the non-employed worker's position in the labour market. The flow payoff from non-employment stands at around $\hat{b} = 21.2$, which is equivalent to 23.5 per cent of the average earnings. The probability of skill depreciation while non-employed is estimated to be 43 per cent. Jarosch (2023) applies a similar process for Germany. He finds that the probability of monthly skill depreciation stands at 24 per cent, which is equivalent to 55 per cent at the quarterly rate.

When it comes to the results concerning the job search and wage negotiation, the relative search effort of the employed workers of $\hat{\gamma} = 0.56$ lies closely to what Flinn et al. (2017) find. The estimated worker's bargaining weight of $\hat{\beta} = 0.41$ for the model with on-the-job search is similar to that of Flinn (2006), Flinn et al. (2017), and Elsby and Michaels (2013).[15]

[15] The notable exception is the work of Gregory (2020) whose estimate of workers' bargaining power stands at 0.66. This, however, is likely the result of a very small discount rate chosen to avoid producing negative wages.

Finally, the procedure estimates the parameters relevant to the skill distribution in the economy. The estimate of the variance of the logged initial human capital implies that 84 per cent of values of $\log a$ lie in $[-2.5, 2.5]$. The probability of increasing one's human capital via on-the-job learning stands at 24 per cent, over 5 times the rate of the increase of a via on-the-job training equalling 4.3 per cent.

4.1.2 Firm Parameters

I estimate three parameters governing a firm's entry into the industry and vacancy opening. The cost of entry stands at $\hat{c}_e = 38.3$, which corresponds to approximately 24 per cent of the average quarterly firm revenues. Discipling the firm size distribution, the hiring convexity of $\hat{\lambda}_2 = 2.35$ is relatively high and in line with the findings of Merz and Yashiv (2007). The resulting measure of entrants, $\hat{N}_e$, is 0.025.

The estimates of the minimum and maximum training costs, $\underline{\xi}$ and $\overline{\xi}$, point out a large dispersion in ξ across the industry. These figures can be interpreted in two ways. That is, they represent the monetary training costs or the opportunity cost of allocating work time to on-the-job training. Such a large range is in line with the empirical evidence in Almeida and Carneiro (2009) and Martins (2021), who also find a broad dispersion in the training cost across firms. Further, my results are consistent with the findings of Flinn et al. (2017), who showcase that around 11 per cent of average work time is spent on training. Figure C.2 indicates that the above costs translate to between 1 and 12 per cent of the average firm revenue, making the findings for most of the firms in line with Flinn et al. (2017).[16]

Left are two estimates governing the firms' performance and exogenous match termination. The dispersion of firm productivity, σ_z, is smaller than that of the initial human capital. It translates into the coefficient of variation of labour productivity at 0.64, which is lower than that of the firm-level productivity estimated by Office For National Statistics (2024). The exogenous separation rate, δ_s, disciplines the average job duration. Given that the average tenure is now also impacted by job-to-job transitions (governed by γ), the estimate is smaller than that of Guner and Ruggieri (2021). It stands at around 0.3 per cent at the quarterly frequency.

4.2 Model Fit

I find that the model fits the UK data relatively well. When it comes to the worker side of the economy, the crucial earnings-related moments are presented in Figures 4.1a and 4.1b, with further estimated results shown in Table C.3. On the firm's end, I focus on the well-matched size and training moments that discipline the economy's reaction to the training subsidy policy, shown in Figures 4.1c, 4.1d, and 4.1e. The remaining firm statistics can be found in Table C.4. I also explore the sensitivity of the fit

[16] It is worth pointing out that the model without on-the-job search generates a larger dispersion of the training costs relative to the average revenue compared to its counterpart without on-the-job search. This is driven by two factors. First, the average revenue is larger in the model with on-the-job search. That is, firms can poach more productive workers from their competitors. This means that the firms in the right tail of the productivity distribution can attract even more productive workers. Second, the lower bound of training costs, $\underline{\xi}$, is bigger in the model featuring on-the-job search.

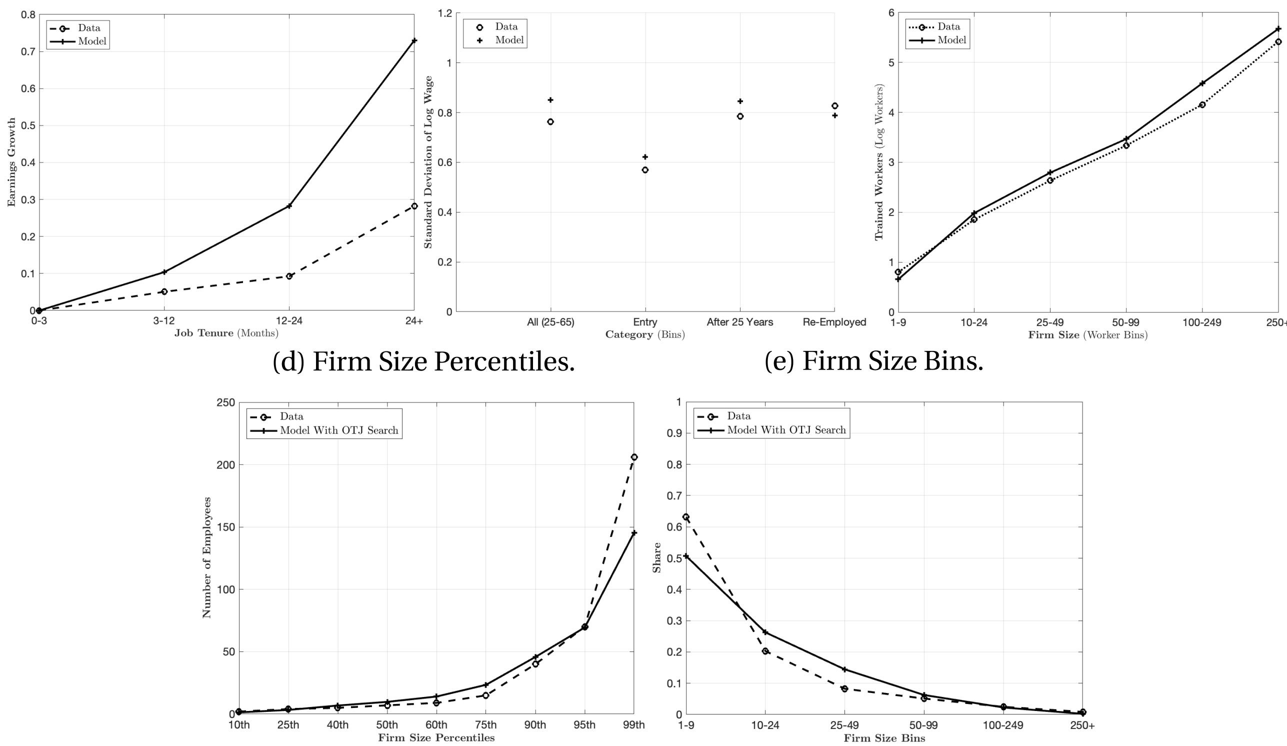
to the values of parameters. This is done by estimating the objective function's Jacobian, which is presented in Table C.5.

4.2.1 Worker Statistics

Figure 4.1: Model Fit.

(a) Job Tenure-Earnings Profiles. (b) Experience Inequality Profiles. (c) Trained Workers by Size.

(d) Firm Size Percentiles. (e) Firm Size Bins.



Notes: Panel 4.1a represents the average earnings for workers across different job tenures. Panel 4.1b demonstrates the fit of the standard deviation of log wage for different categories of workers. Panel 4.1c shows the average number of employees receiving training across different firm-size bins. Panel 4.1d presents the average number of workers across different firm-size percentiles. Panel 4.1e highlights the share of firms in different firm-size bins.

The model successfully matches the aggregate labour market moments that are key for studying a training subsidy programme. The flow payoff of non-employment parameter, b, generates the benefits-wage ratio which is in line with its data counterpart. Similarly, the average tenure statistic from the model is just a quarter short of its data counterpart. This is enough to generate tenure-dependent wage growth in the economy. Like in Flinn et al. (2017), the relative search efficiency of employed workers (γ) determines the job-to-job transition rate, which nearly perfectly corresponds to its data counterpart. Yet, the perfect match of the job-to-job transitions combined with a relatively low estimate of the measure of potential entrants, N_e (disciplined by the firm-size distribution), leads to an underestimation of the employment rate.

The framework fits the earnings-related statistics well for the purpose of analysing the impacts of the training subsidy. The model fits the tenure return relatively well for those who work up to a year, but it overshoots for the longer periods spent on the same job, as evidenced in Figure 4.1a. That is, a worker with a quarter of on-the-job experience is expected to earn 5 per cent more than an entrant, compared to the 3 per cent result in the data. At the same time, an employee with more than 2 years of tenure commands over a 60 per cent higher wage than an entrant. In the LFS data, the difference stands at around 30 per cent. The discrepancy is larger than in the framework of Guner and Ruggieri (2021), who do not allow for on-the-job search, due to the relatively big estimate of the training jump rate, p^t. The parameter is identified by the training statistics and must be kept large enough to compensate firms for the possibility of their workers being poached. Further, as shown in Table C.3, the experience return statistics are relatively well-matched. The only exception for the average entry wage which remains suppressed to contain the overestimated tenure returns for those who do not change jobs. As a result, the model satisfactorily mirrors the data for the inequality statistics, shown in Figure 4.1b. These are especially important as I analyse the inequality consequences of the training subsidy in Section 5.2. Finally, I ensure the model matches an appropriate training return moment. The statistic I pick is the return on training from regressing the log wage on the past period's training. I find a 25 per cent return on past training in the framework, compared to the 20 per cent value in the LFS data.

4.2.2 Firm Statistics

The model does a remarkable job of targeting the size and training-related characteristics of the UK firms. By estimating the average firm size to be 19.52 employees, my framework successfully matches the same moment in the firm survey data, where it stands at 19.55, as seen in Table C.4. The statistic is almost solely identified by the entry cost parameter, c_e, which has a more limited impact on other results, as shown in the Jacobian of the objective function in Table C.5. Another important moment I target is the share of training firms, which represents the human capital investment's extensive margin. In my model, 63 per cent of firms provide on-the-job training to their workers. In the data, this stands at 65. Along the intensive margin, I find that 44 per cent of employees receive training, compared to the 51 per cent in the LFS. The findings along both margins depend on the training cost bounds, $\underline{\xi}$ and $\overline{\xi}$.

In what is significant for analysing the heterogeneous impact of the training subsidy, my model succeeds at imitating the distributional statistics of the UK business environment. The framework closely tracks the training patterns by firm size, as shown in Figure 4.1c. These statistics are sensitive to the training cost bounds ($\underline{\xi}$ and $\overline{\xi}$) and training jump rate (p^t). Unlike in the version with no on-the-job search, a relatively high p^t serves the role of counteracting the poaching-risk reduction in the value function of a match, given by Equation (3.12). Further, Figures 4.1d and 4.1e highlight the close fit of the size distribution of firms. This is identified with the productivity dispersion (σ_z), hiring cost convexity (λ_2), and the training jump rate (p^t) parameters. Lowering p^t impacts the vacancy posting rule in Equation (3.18) for smaller firms and shifts more mass of the distribution to the right. Finally, the model's close match of both distributions, combined with the linearity of the production function, allows for generating the wage-size effects outlined by Elsby and Michaels (2013) and

Brown and Medoff (1989), which usually remains a challenge in the frictional search and matching environments.

5 Results: Training Incentive Subsidy

In this section, I argue that the training subsidy increases the proportion of workers trained and wages via three mechanisms.

1. *Human Capital Effect.* The policy leads to an aggregate increase in training provision. This translates into a more skilled workforce and higher wages.

2. *Job Value Effect.* Expanded on-the-job training incidence raises the value of working at the subsidy-affected firms.[17] This results in shifting firm decisions and the wage bargaining environment.

3. *Distribution Effect.* Impacted firms adjust their vacancy posting decisions. The share of productive and better-paying employers consequently increases.

I leverage the model to identify the impact of these three causal links in driving the policy's outcome. To do so, I use the elements of different estimated equilibria: one without the government and one with the training subsidy. I term them baseline and subsidy equilibria, respectively.[18] First, I isolate the job value effect, by simulating the worker and firm outcomes based on the baseline distributions of vacancies and workers. The policy and value functions are taken from the subsidy equilibrium. Second, I isolate the job value and distribution effects together. I do so by recomputing the first step with the vacancies distribution from the subsidy equilibrium and the human capital distribution of workers from the baseline model.

Comparing these two sets of results with their counterparts from the baseline and subsidy equilibria, I assess the role of each mechanism separately. I contrast the baseline and first sets of results to gauge the impact of the job value effect. Setting the first and second sets of results against each other allows for disentangling the role of the distribution effect. Finally, I compare the firm and worker outcomes from the second set of results with those of the subsidy equilibrium to analyse the human capital mechanism's influence.

[17] In Figure D.4a, D.4b, and D.4c in Appendix D.2, I review the change in the job value functions with and without the policy, finding that for most of the levels of human capital, the job value function increases in the policy's presence. In the case of non-employed workers, Figure D.5a highlights that the option value of search increases by at most around 1 per cent for those with mid-high ability levels. Conversely, the circumstances are exactly the opposite for the employed workers. For them, the option value of search (on average) increases most for the least and more skilled employees, as shown in Figure D.5b. For the former, it happens as the subsidy means they become more "employable." For the latter, this results from a higher training incidence and a larger proportion of high-productivity firms that hire them.

[18] In the model's quantitative environment, I need four model-generated elements to simulate workers' and firms' outcomes: (1) the distribution of vacancies, (2) the distribution of workers, (3) the policy functions, and (4) the value functions.

5.1 Impact of Subsidised Training on Firms

Analysing the policy's impact on firms, I find that the three mechanisms are mainly driven by the changing vacancy and training provision decisions. As a result, more on-the-job training is provided along extensive and intensive margins, especially by the larger firms. While this means that the average firm size and revenue increase, this is almost entirely driven by the positive impact of the job value effect on the largest firms. They also see an improvement in the average wage paid. At the same time, smaller employers see the job value mechanism push down their revenues and wages. Finally, I show that the human capital and distribution effects are secondary, usually only restraining the strength of the job value causal link.

5.1.1 Mechanisms' Origins: Firm Decisions and Training Provision

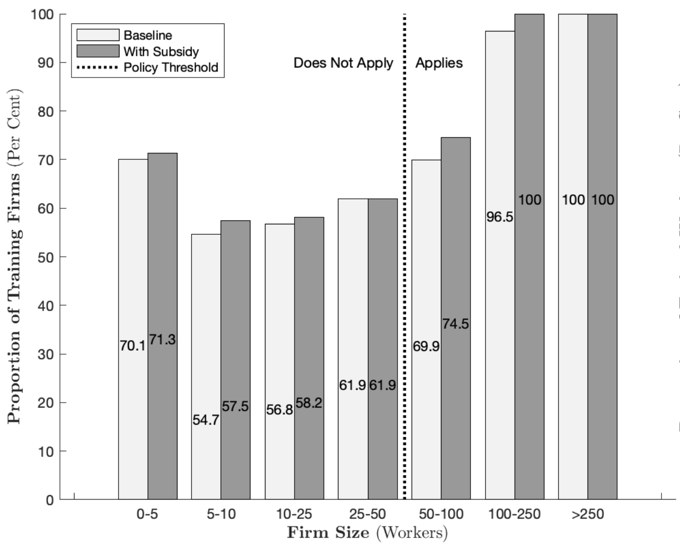
The policy does not affect the match acceptance and entry decisions, but it induces a change in the training and vacancy posting policy functions. First, Figures D.1a, D.1b, and D.1c map the evolution of the match acceptance policy functions at different categories of firms. There is only a marginal adjustment in the acceptance decisions of firms with the government subsidy programme compared to the baseline. This is in line with the argument in Fu (2011): firms have the incentive to hire lower-ability workers and train them due to the presence of search frictions in the economy even without the subsidy. Figure D.3a also demonstrates that the policy does not have an impact on the firms' entry decisions. On the other hand, the vacancy posting reacts visibly to the training programme. Employers with high training costs and lower productivity hire fewer workers. Those more productive post more vacancies, as seen in Figure D.3b. Meanwhile, the training provision decision undergoes a noticeable adjustment, especially for the high training cost employers, as shown in Figures D.2a, D.2b, and D.2c.

Namely, the subsidy induces firms to provide more training along the extensive and intensive margins. More firms provide at least one quarter of training in each size category, as shown in Figure 5.1a. Especially among those in the 50-100 workers bin, the policy leads to an almost 5 percentage points increase in the proportion of training employers. All of the largest firms already train their workforces without the policy. This is in line with the argument that bigger firms constitute better learning environments for young employees (Arellano-Bover, 2024). By the same token, more workers get trained in each size bin of firms, as pictured in Figure 5.1b. In the group of the taxed firms, 1 to 6 percentage points more workers are trained. In each category of firms not subject to the levy, more employees receive on-the-job training, albeit the difference ranges from 0.2 to only 0.9 percentage points. My findings of a robust impact of the training subsidy are conceptually in line with the empirical evidence of Martins (2021) who illustrates that the training grants increase training participation at Portuguese firms.[19]

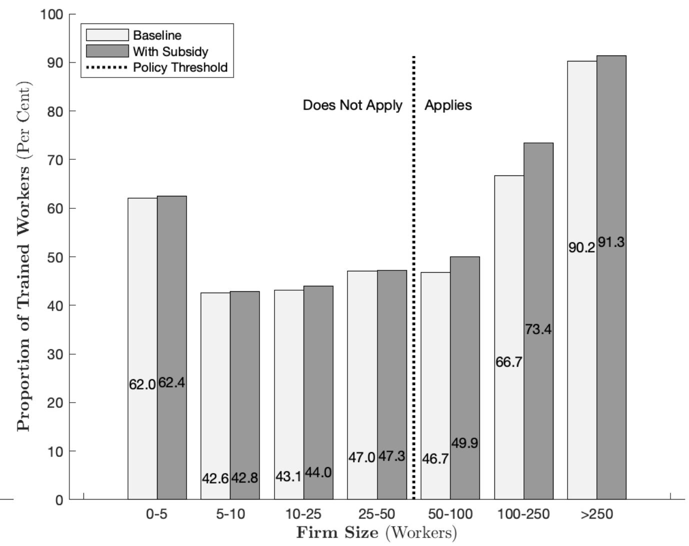
[19] For more, see Card et al. (2018) for an extensive summary of active labour market programmes in both developing and developed countries.

Figure 5.1: Training Incidence by Firm Size: Extensive and Intensive Margins.

(a) Extensive Margin.



(b) Intensive Margin.



Notes: Figure 5.1a features the proportion of firms training their workers in each category. For example, the value of 100 means that all firms in that category provide at least a quarter of training to at least one worker (the extensive margin). Figure 5.1b shows the proportion of workers receiving training in each category of firms (the intensive margin).

The three mechanisms originate from these changes. The shifting training choices change employment and match value functions for workers and firms, respectively. This gives rise to the job value mechanism. Adjusted vacancy posting decisions translate into firm distribution tilted more heavily towards productive firms, sparking the distribution effect. That even the firms not affected by the subsidy train more frequently is a testament to the strength of these two mechanisms. Finally, with firms providing more training along the extensive and intensive margins, in turn, the human capital distribution in the economy shifts up.

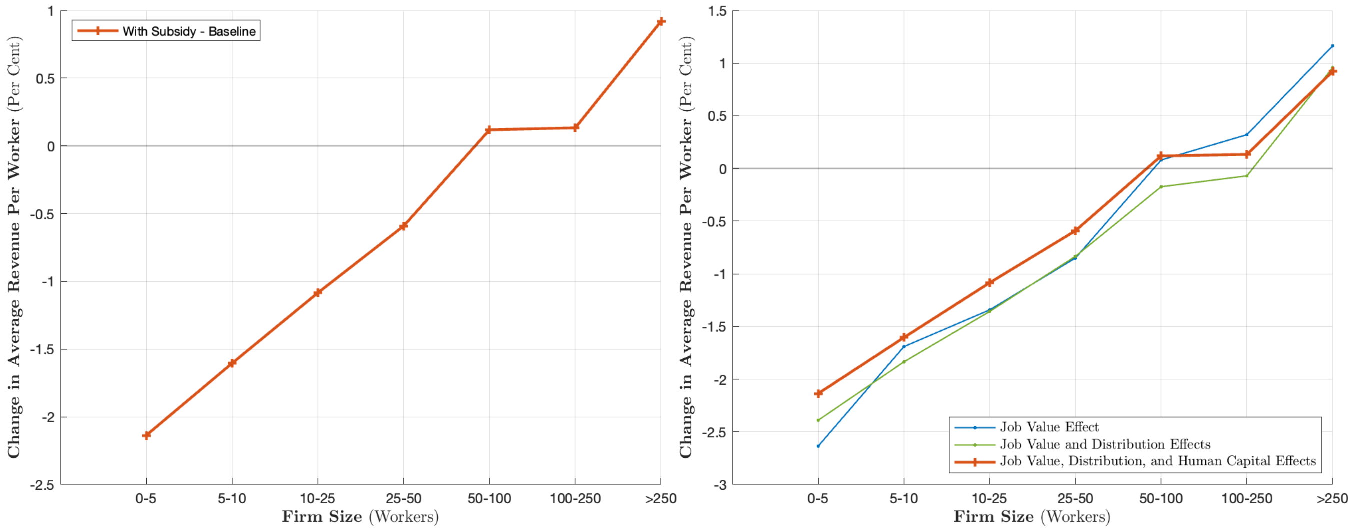
5.1.2 Mechanisms' Influence: Economic Outcomes of Firms

As a result, the training subsidy leads to changes in the key firm outcomes, especially revenues and wages. There is a 1 per cent increase in the average size but the overall size dispersion does not adjust much, as shown in Table D.1. Firms improve their average productivity and revenues, with the latter jumping by around 1.5 per cent. Figure 5.2a highlights that this is mostly driven by the firms with more than 50 employees as they are the only ones to see their average per-worker before-tax revenue increase. Meanwhile, the policy's impact on small firms is negative, translating into even larger inequality among employers. The same process is reflected by the joint growth of the P90-50 and P50-10 ratios of firm productivity in Table D.1.[20] Finally, the average wage rises only for the largest category of firms, as shown in Figure 5.3a. This implies that the training subsidy deepens the wage-size effect

[20]Note that, with and without the subsidy, the dispersion of revenue is moderately larger among the smaller firms, which echoes the argument about small firms being less productive in the TFP literature (e.g., see Bento and Restuccia (2017)).

Figure 5.2: Difference between Per-Worker Before-Tax Revenues by Firm Size With Subsidy and Baseline.

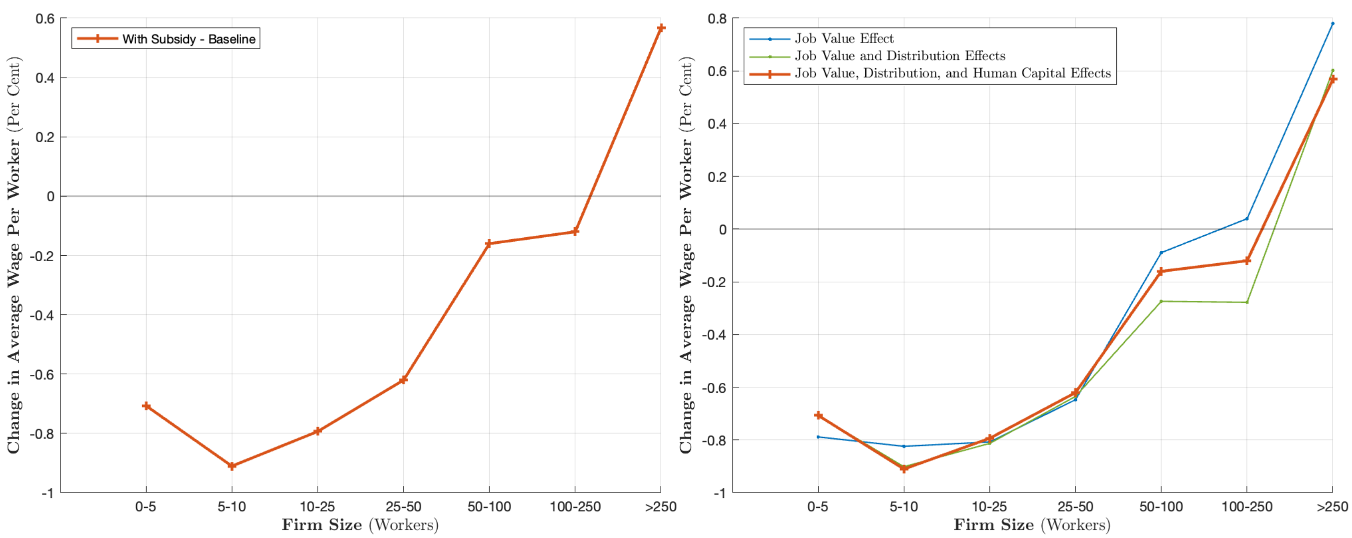
(a) Difference. (b) Mechanisms Decomposition.



Notes: Figure 5.2a shows the per cent change in the per-worker before-tax revenue of firms between the environment with the subsidy and the baseline. For example, 2 indicates that imposing the government training programme is followed by a 2 per cent higher before-tax revenue per worker than in the baseline. Figure 5.2b shows the decomposition of the change into the three causal links. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy. The blue line plot indicates the adjustment when only the job value effect is active. The green line highlights the change when the job value and distribution effects are operational. As a result, comparing the blue and green lines allows for assessing the distribution effect alone. Finally, the orange line shows the influence of all three mechanisms, which is equivalent to the policy's total impact.

Figure 5.3: Difference between Average Wage by Firm Size With Subsidy and Baseline.

(a) Difference. (b) Mechanisms Decomposition.



Notes: Figure 5.3a shows the per cent change in the average wage of firms between the environment with the subsidy and the baseline. For example, 0.5 indicates that imposing the government training programme is followed by a 0.5 per cent higher average wage than in the baseline. Figure 5.3b shows the decomposition of the change into the three mechanisms. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy. The blue line plot indicates the change when only the job value effect is active. The green line highlights the change when the job value and distribution effects are operational. As a result, comparing the blue and green lines allows for assessing the distribution effect alone. Finally, the orange line shows the impact of all three mechanisms, which is equivalent to the policy's total impact.

from Elsby and Michaels (2013) and Brown and Medoff (1989).

Decomposing the three mechanisms' roles, I show that the job value effect is the main culprit behind the policy's heterogeneous impact on firms. Figures 5.2b and 5.3b map the contribution of each mechanism to the average before-tax per-worker revenue and wage changes. Consider the policy's impact on firms with more than 100 employees. Working at large firms becomes more valuable which allows them to poach productive workers from their slightly smaller counterparts. This explains why the job value effect has a positive impact on both the wages and revenues of such businesses. Further, as there are more skilled workers, bigger firms attract better talent. Still, since the largest of them already target the best candidates, the human capital mechanism exerts a secondary influence on their outcomes. These two effects are countered by the distribution mechanism. That is, with the industry's composition shifting in favour of large firms, the top workers are more likely to receive competing offers that they are willing to accept. The story is different for smaller firms. The job value effect now harms revenues and wages as they are more likely to see their productive workers poached. The decomposition also suggests that their gain from the upskilled labour force is minor - it is the large firms that benefit from the human capital mechanism.

5.2 Impact of Subsidised Training on Workers

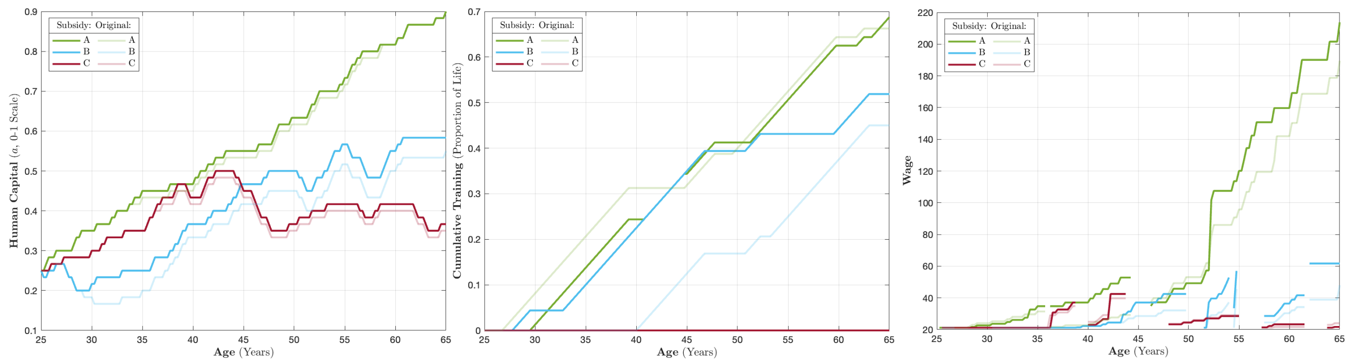
Given the evidence that the policy positively impacts the largest firms, mainly through the job value effect, I move to the workers' side of the economy to study how the subsidy's presence affects different types of agents. I begin with a pedagogical example of how workers along different labour market paths are influenced by the mechanisms analysed. This allows me to set the stage for showcasing how these causal links lead to greater training incidence, a higher mean wage, and lower earnings dispersion.

5.2.1 Stylised Example: Policy and Heterogeneous Sources of Wage Growth

In Figure 5.4, I plot the labour market paths of three ex-ante identical model-simulated workers to highlight the salient features of the policy. Figures 5.4a, 5.4b, and 5.4c show the evolution of their skill, cumulative training incidence, and wage levels. The darker lines represent the environment with the training subsidy while the tinted counterparts are derived from the government-free benchmark. I pick three workers who start with the same low level of human capital and face heterogeneous histories to point out how the subsidy impacts different pockets of the labour market. Successful in the benchmark, worker A (in green) does not receive extra training, but benefits from the job value and distribution effects of the policy. Worker B (in blue), less fortunate in the baseline, sees the subsidy correct for the initial human capital drop. This improves her labour outcomes via the human capital effect. Finally, worker C (in maroon) faces the same problems in both scenarios: the government programme cannot rectify skill depreciation caused by a prolonged period of non-employment in her 40s.

Figure 5.4: Simulated Labour Market Paths of Three Ex-Ante Identical Workers.

(a) Human Capital Paths. (b) Training Paths (Cumulative). (c) Wage Paths.



Notes: All figures show the labour market paths of three *ex-ante* identical workers between 25 anf 65. Each of them is subject to different shocks and meets different firms. The dark lines represent the paths in the subsidy's presence. The tinted lines show the baseline model outcome. All workers are subject to the same shocks and meet the same firms in both economies, but the decisions are endogenous and subject to different economic environments. Figures 5.4a, 5.4b, and 5.4c show the paths of the human capital, cumulative training provision, and wages of the chosen workers.

The policy's strong impact on worker A serves as an example of the subsidy's potential positive impact on workers. In the baseline, she matches with productive firms from the beginning of her career. After two years of increasing her human capital only through on-the-job learning, she begins to receive regular training. This continues throughout her working life, allowing her to reach one of the highest possible skill levels by age 65. As a result, her wage raises from a little over 20 per cent of the economy's average pay at 25 years old to more than 200 per cent at 65. Now, notice how her circumstances evolve in the subsidy-impacted economy. She successfully matches with a more productive firm with lower training costs earlier on. Once employed, she improves her skills solely through on-the-job learning until she starts obtaining firm-sponsored training in her late 20s. Thanks to this initial hike, she improves her abilities in her 30s more than in the benchmark. By her early 50s, this makes enough difference for her wages to grow and remain noticeably higher.

The government training subsidy's influence on worker B highlights the effectiveness and limitations of the training subsidy policy. In the baseline economy, she faces a non-employment streak in her late 20s. This translates to a period of skill depreciation that sets her earnings back for over 10 years. Further, matching with a high training cost firm implies that she does not receive on-the-job training until her 40s. Only then does she catch up and her wage picks up, reaching almost 50 per cent of the average pay by the age of 65. Her labour market trajectory changes in the presence of the training subsidy. The policy corrects the unemployment spell, as she receives more training earlier in her career. This lifts her human capital for the rest of her life. Her wage takes up from her early 40s, culminating at almost 70 per cent of the average earnings in the economy at 65. Note that the growth starts taking place around the age of 43 when she matches with a more productive employer than in the baseline.

Worker C's trajectory in both scenarios indicates that the policy is not a panacea for all the risks of the frictional labour markets. Take the baseline economy. By matching with a medium productivity and high training cost firm early on in her career, she fails to secure on-the-job training. She then loses her accumulated ability in a non-employment spell in her mid-40s. Against that setting, she is not trained later in life, even when she works for highly productive employers. The human capital stagnation translates to permanently low wages that never pick up from around 20 per cent of the economy's average. The story is similar in the environment with the government training subsidy. The only difference to her path is that she temporarily matches with a high-productivity firm in her late 30s. Still, given the employer's high training costs, she only upskills via on-the-job learning. Like in the baseline economy, she does not avoid the non-employment spell and is set back until her 60s and her wage remains stagnant.

Note that each of the workers is differently impacted by the three mechanisms. Worker A matches with a more productive firm earlier on, which points out the joint impact of the job value and distribution mechanisms on her labour outcomes. That is, these effects boost the surplus value of her match with a more productive firm above zero, helping her secure an attractive job offer she does not get in the benchmark scenario. At the same time, worker B sees the subsidy improve her outcomes around the age of 40. Around then, she has a higher skill level than in the baseline and manages to match with a more productive employer. This highlights that the human capital mechanism works mostly together with the remaining two effects. Namely, the job value and distribution effects allow worker B to leverage her higher ability for a better job offer. Finally, worker C sees little change with the training subsidy policy. Even with the job value and distribution channels, she does not avoid the detrimental non-employment spell.

5.2.2 Workers: Training Incidence, Lifecycle Labour Outcomes, and Inequality

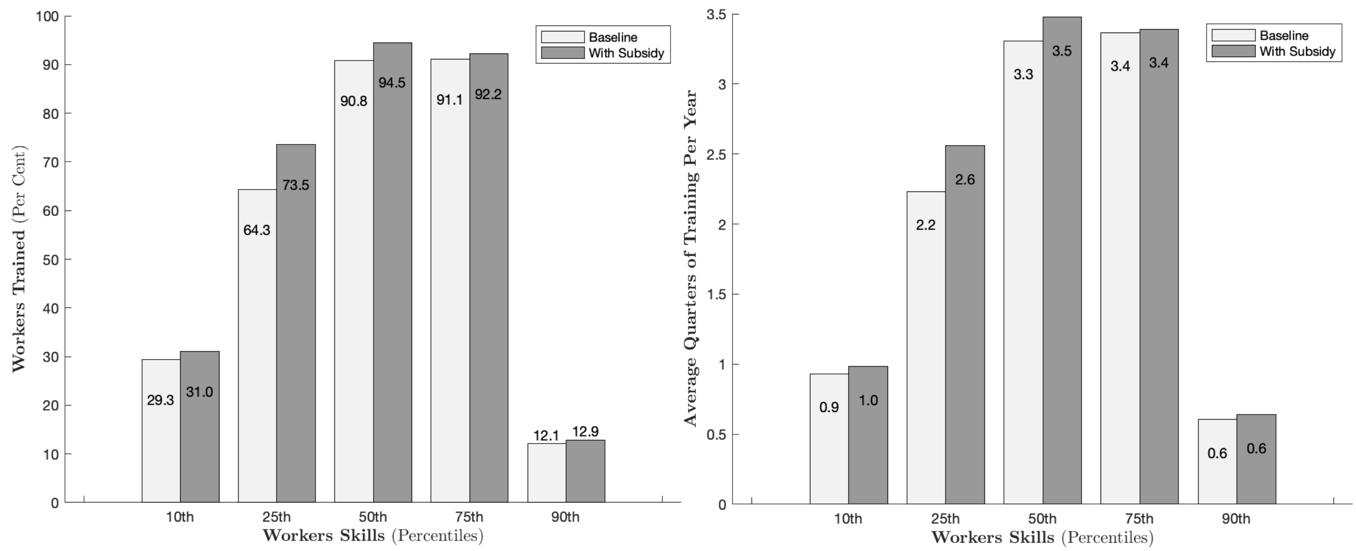
I find that the training subsidy translates into a higher training incidence, more robust wage growth, increased average wage, and lower earnings inequality. Figures 5.5a and 5.5b show how training incidence evolves along extensive and intensive margins with the subsidy for different human capital bins. All but the most skilled workers receive more training along both margins in an environment with the training incentive policy. In both cases, those around the 25th percentile of the ability distribution see the biggest increases in training incidence. They obtain training in 9 percentage points more cases and spend an average of 0.4 quarters more being trained. Conversely, the most skilled workers are at a disadvantage. They see their training rates barely move along both margins. Like in Fu (2011)'s argument, firms are wary of investing in such workers as they are most likely to leave. Then, Table 5.1 shows the changing wage structure of the economy. While the average entry wage drops as a result of the policy, the subsidy pushes up pay growth across the lifecycle. All in all, the average wage is boosted by stronger growth and depressed by the lower entry wage.[21] This induces a change in earnings inequality. The second and third lines of Table 5.2 indicate the reduction in both the P90-50 and P50-10 ratios. Further, Figure D.7 shows that the fall of the variance in log wages

[21]The same can be inferred from Figures D.6a and D.6b. That is, larger average earnings are induced by the higher pay of the more experienced workers (which itself is a result of more robust growth rates earlier in life).

Figure 5.5: Training Incidence by Human Capital Levels: Extensive and Intensive Margins.

(a) Extensive Margin.

(b) Intensive Margin.



Notes: Figure 5.5a features the proportion of workers receiving any training by human capital bin. For example, 65 means that 65 per cent of workers receive at least one quarter of training (the extensive margin). Figure 5.5b shows the average number of quarters of training received within a year for each human capital category (the intensive margin).

Table 5.1: Wage Statistics: Policy's Impact and Mechanisms Decomposition.

Worker Variable	Baseline	Job Value Effect	Job Value and Distribution Effects	Job Value, Distribution, and Human Capital Effects
Average Entry Wage (Proportion of Baseline)	1.000	1.015	0.949	0.943
Average Wage (Proportion of Baseline)	1.000	1.056	1.003	1.008
Wage Growth (Entry - 5 Yrs.)	0.129	0.141	0.137	0.150
Wage Growth (Entry - 10 Yrs.)	0.341	0.365	0.361	0.354
Wage Growth (Entry - 25 Yrs.)	0.840	0.884	0.900	0.885
Employment Rate	0.608	0.596	0.608	0.618

Notes: To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy.

is achieved primarily in the latter half of one's life. This suggests that the decrease in inequality is achieved primarily through the compression of wages among older workers.

Finally, I disentangle the impact of the three mechanisms on labour outcomes, showing that the job value effect plays a central role in increasing average wage and lowering earnings inequality. The falling entry wage can be explained by the joint impact of the distribution and human capital effects. That is, in an environment with more productive firms and skilled workers, labour market entrants are at a disadvantage while negotiating with their prospective employers. Meanwhile, those in the 25th percentile of the skill distribution benefit from the changes. Like worker A in her early life in the stylised example in Section 5.2.1, they are trained more and increase their human capital.[22] Their outside options also become more attractive in an economy where more training takes

[22] This is in line with the remedial and career-advancing interpretations for on-the-job training in Bartel (1995).

Table 5.2: Inequality Statistics: Policy's Impact and Mechanisms Decomposition.

Inequality Statistics	Baseline	Job Value Effect	Job Value and Distribution Effects	Job Value, Distribution, and Human Capital Effects
Gini	0.421	0.412	0.418	0.418
P90-50	2.804	2.844	2.747	2.768
P50-10	4.472	4.537	4.489	4.451
Mean-Median	1.274	1.252	1.249	1.261
Var. of Log Wage	0.823	0.806	0.811	0.813

Notes: To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy.

place.[23] These two changes explain the joint roles of job value and human capital effects in driving up the wage growth for those in their first 5 years in the labour market, as shown in Table 5.1. The story, however, becomes different once we consider older and more skilled employees. The human capital effect lowers their wages as there are more skilled workers competing with them for jobs at the most productive firms. This explains the negative role of this mechanism for those over the age of 45 highlighted by Figure D.6b. The combination of these patterns translates to a fall in inequality statistics. Note that it is the job value that is the main mechanism through which those around the 25th percentile of the human capital distribution benefit more than other agents and wage growth is pushed across the life cycle. This gives the reason of why Table 5.2 points out to it as the main driver of reducing earnings inequality (e.g., in the cases of the Gini coefficient and the variance of log wages).

6 Subsidy Extension

I next ask whether expanding the subsidy to all employers in the economy leads to an improvement in firm economic results and labour outcomes in the UK. In answering the question, I find that such a change benefits small and medium firms and changes the composition of the industry. While still improving the average wage, this leads to worse economic performance of firms and shifts the expanded policy's advantages to older workers at the cost of worse prospects for the entrants.

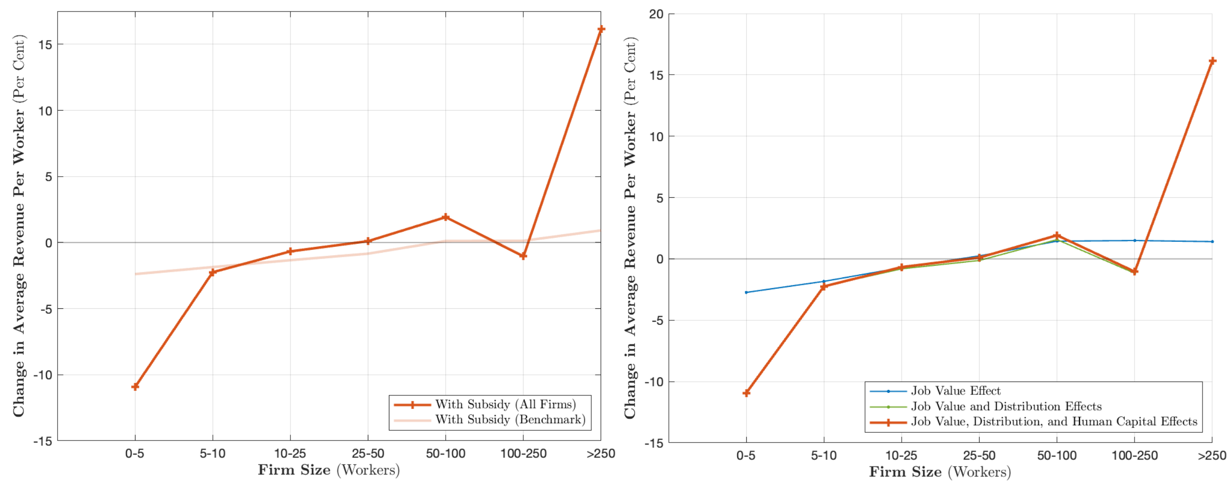
6.1 Impact on Firms

The subsidy's expansion to the entire industry positively impacts training incidence but fails to improve economic outcomes. Small and medium businesses increase their training provision by over 10 percentage points along the extensive margin, as shown in Figure D.8a. Those with 50 to 100 workers see much higher growth than with the benchmark subsidy in section 5.1 (to 81 instead of 75 per

[23] In an environment with more productive firms, less skilled young workers see a higher growth of the option value of search while non-employed than while employed. For example, compare the changes in the option values of search for the non-employed and employed workers in Figures D.5a and D.5b. Take a worker from the bottom 25th per cent of the human capital grid. Her search value increases by 0.5 per cent when she is non-employed. At the same time, it almost does not change (on average) when employed. With the distribution mechanism causing more firms to be productive, the search value is even more dampened when she is employed - she is less likely to be poached while already working at a good firm. This lowers the surplus divided by the employer and employee in Equation (3.23).

Figure 6.1: Difference between Average Per Worker Before-Tax Revenue by Firm Size With Expanded Subsidy and Baseline.

(a) Expanded and Benchmark Subsidies. (b) Mechanisms Decomposition.



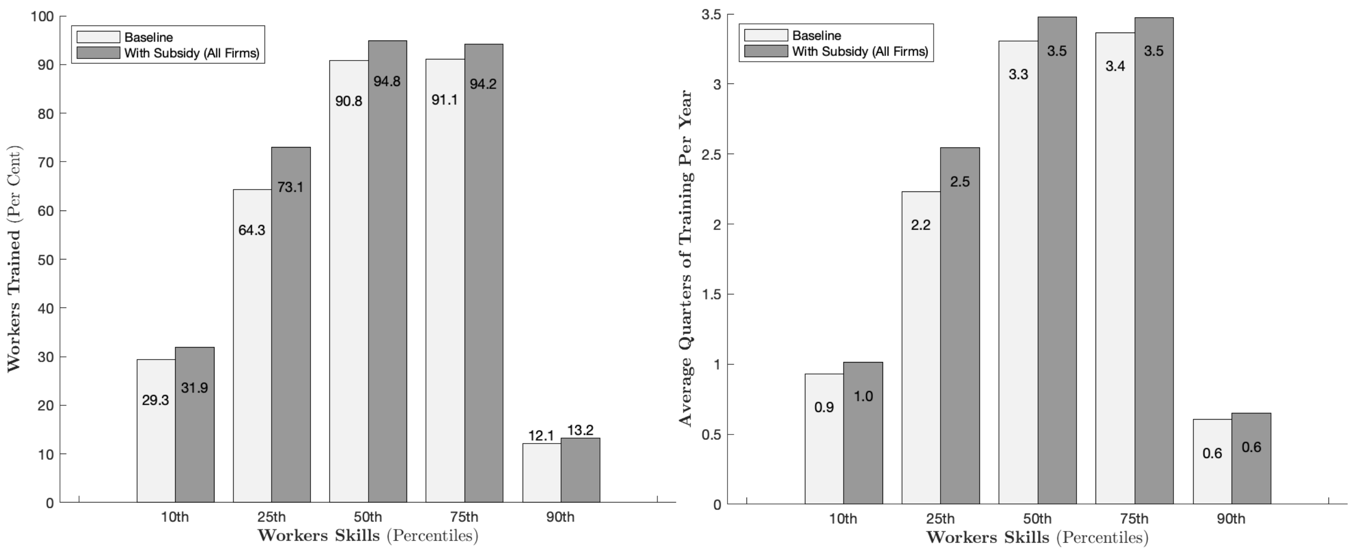
Notes: Figure 6.1a compares the average before-tax per worker revenues by firm size for the baseline, benchmark subsidy from Section 5, and the expanded training subsidy (to all firms). In both cases with the training incentive policy, the subsidy level is determined via the balanced budget condition in Equation (3.26). Figure 6.1b shows how the decomposition of the change in the average before-tax per worker revenue (relative to the baseline) into the three mechanisms for the expanded subsidy. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy. The blue line plot indicates the change when only the job value effect is active. The green line highlights the change when the job value and distribution effects are operational. As a result, comparing the blue and green lines allows for assessing the distribution effect alone. Finally, the orange line shows the impact of all three mechanisms, equivalent to the policy's total impact.

cent). The pattern is similar along the intensive margin. Figure D.8b highlights a 2 to 10 percentage points increase in the proportion of employees obtaining on-the-job training, especially at medium firms. The subsidy's expansion induces the rearrangement in the pool of firms, which is tilted towards smaller sizes. Against this backdrop, the average size falls by almost 2 workers, as shown in Table D.2. Further, Figure 6.1a maps out a relative fall in the average per-worker revenue at the smaller firms. These two patterns contribute to a large, over 10 per cent, fall in the average before-tax revenue. The dynamics of earnings vary across the firm size distribution and closely track the pattern in revenues, as plotted in Figure D.9b. Pay falls at the smallest firms and follows a heterogeneous pattern at their larger counterparts.

Decomposing the contribution of the three mechanisms to the policy's impact allows for understanding how it disrupts the large firms' performance. First, the policy is advantageous to small and less productive businesses. With lower revenues, they pay little in levy in absolute terms but receive the advantageous proportional training subsidy. This especially helps businesses with 25 to 100 workers provide more training along extensive and intensive margins and benefit from the job value effect, as indicated by shown in Figure 6.1b. Further, this change implies a drop in the match value for the bigger high-cost firms, suppressing the optimal number of vacancies governed by Equation (3.18). In leading to a fall in the proportion of the largest employers, the distribution effect leaves only the most productive ones in the highest category, which explains the mechanism's contribution to the significant revenue boost for them. Given that the policy's expansion causes the selection of only the

Figure 6.2: Training Incidence by Human Capital Levels: All Firms Eligible.

(a) Extensive Margin. (b) Intensive Margin.



Notes: Figure 5.5a features the proportion of workers receiving any training by human capital bin. For example, 65 means that 65 per cent of workers receive at least one quarter of training (the extensive margin). Figure 5.5b shows the average number of quarters of training received within a year for each human capital category (the intensive margin).

Table 6.1: Wage Statistics: Expanded Impolicy's Impact and Mechanisms Decomposition.

Worker Variable	Baseline	Job Value Effect	Job Value and Distribution Effects	Job Value, Distribution, and Human Capital Effects
Average Entry Wage (Proportion of Baseline)	1.000	0.962	0.950	0.920
Average Wage (Proportion of Baseline)	1.000	1.009	1.002	1.011
Wage Growth (Entry - 5 Yrs.)	0.129	0.144	0.134	0.107
Wage Growth (Entry - 10 Yrs.)	0.341	0.358	0.346	0.336
Wage Growth (Entry - 25 Yrs.)	0.840	0.868	0.898	0.889
Employment Rate	0.608	0.613	0.604	0.614

Notes: Table 6.1 shows wage results of extending the policy to all firms. The subsidy level is determined via the balanced budget condition in Equation (3.26). The panel shows the decomposition of the change into the three causal links. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. Using the environment above, I add the distribution effect with the entry and vacancy posting policies taken from the subsidy-present economy.

best establishments into the top category, their employees are compensated accordingly.[24] Finally, these two changes are especially detrimental to firms with 100 to 250 employees. They now face the combined impact of the distribution and job value effects as they can see their workers poached by both smaller and larger counterparts at much higher rates. All in all, the advantageous position of medium firms and the relative reduction in the proportion of their largest peers contribute not only to the fall in aggregate revenue but also to the hike in the P50-10 productivity ratio in the economy, as seen in Figure 6.1b.

6.2 Impact on Workers

On the workers' side of the economy, I find that training provision and average wages are boosted similarly to the benchmark policy, but earnings dynamics and inequality follow a different pattern

[24]This implies that the subsidy deepens the "superstar" firms phenomenon of Autor et al. (2020).

than before. The training incidence increases by between 1 and 9 percentage points along the extensive margin, as shown in Figure 6.2a. Highest for those in the 25th percentile of the skill distribution, on-the-job training provision grows more than in the benchmark subsidy. Still, Figure 6.2b highlights the results are nearly identical to those from the benchmark training subsidy for the changes along the intensive margin. In Table 6.1, I find that expanding the subsidy to all firms leads to similar average wage growth as in Section 5.2. The difference lies in how this average earning result is achieved. The expanded subsidy suppresses the average entry wage even more than before. Unlike with the benchmark subsidy, average wage growth falls between the entry and 10th year in the labour market. Here, what drives the wage growth is the boost by the 25th year of working. This indicates that the expanded subsidy benefits older workers more than their younger counterparts. Like in the previous case, one observes a drop in earnings inequality. This time, however, this is achieved primarily via the reduction in the P50-10 ratio, as shown in Table D.3

In decomposing the role of the three mechanisms in driving these changes, I evidence a pattern consistent with the expanded subsidy's impact on the industry. The fall in the proportion of the largest and most productive firms caused by the job effect harms those entering the labour market. That is, new entrants agree to lower pay as there are fewer alternatives at large businesses providing on-the-job training. Table 6.1 indeed highlights a noticeably larger drop in the average entry wage compared to the benchmark subsidy scenario from Section 5.2. For those already in the labour market, the evolution in the pool of firms adds to the fall in wage growth in the first 10 years of experience. As a result, the distribution mechanism suppresses the opportunities to boost earnings by changing employers. The same effect, on the other hand, pushes up the wage growth for the older workers. With more human capital on average, older workers in the middle of the ability grid can choose from a larger number of medium firms that traditionally hire them.

7 Concluding Remarks and Policy Implications

A burgeoning body of macroeconomic literature has studied human capital accumulation via on-the-job learning and training, as well as its impact on wage dynamics and inequality. I build on such works to analyse the impact of the Apprenticeship Levy, a subsidised training programme in the UK. In doing so, I find that the subsidy has a heterogeneous impact on firms and a positive influence on workers. Training incidence is found to increase along both extensive and intensive margins. The average before-tax revenue and average firm size increase, but these changes are driven by the positive outcomes at large firms only. At the same time, workers experience higher wage growth and an increase in the average pay. Most of the measures of earnings inequality drop. In light of that evidence, I ask whether expanding the policy to all employers necessarily leads to better economic outcomes. I show that the extended subsidy benefits small and medium businesses and changes the firm-size distribution in the economy, lowering the average business size. While the average wage increases as well, I highlight that the gains are now concentrated among older workers.

My analysis has relevant implications for policymakers. Low levels of human capital investment are a subject of the policy discussion concerning the UK's "productivity puzzle" (e.g., see Resolution Foundation (2023)). The model's results present a tractable set of trade-offs of the current version of the Apprenticeship Levy and outcomes for expanding the policy to smaller firms. The framework can also be used to explore increasing the tax or targeting a smaller set of firms. The mechanism decomposition allows for a further understanding of what drives the training subsidy and could be applied for better policy design. In what constitutes a further item in my research agenda, the modelling framework can be further expanded to account for match-specific and general human capital types, akin to the work of Flinn et al. (2017).

Bibliography

Acemoglu, D. and Pischke, J. (1999a). Beyond Becker: Training in Imperfect Labour Markets. *The Economic Journal*, 109(453):112–142.

Acemoglu, D. and Pischke, J. (1999b). The Structure of Wages and Investment in General Training. *Journal of Political Economy*, 107(3):539–572.

Acemoglu, D. and Pischke, J.-S. (1998). Why Do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics*, 113(1):79–119.

Almeida, R. and Carneiro, P. (2009). The Return to Firm Investments in Human Capital. *Labour Economics*, 16(1):97–106.

Almeida, R. K. and Faria, M. (2014). The Wage Returns to On-the-Job Training: Evidence from Matched Employer-Employee Data. *IZA Journal of Labor & Development*, 3(1):19.

Arellano-Bover, J. (2024). Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size. *Journal of Labor Economics*, Forthcoming.

Arellano-Bover, J. and Saltiel, F. (2024). Differences in On-the-Job Learning across Firms. *Journal of Labor Economics*, Forthcoming.

Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics*, 3(3):188–220.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Bagger, J., Fontaine, F., Postel-Vinay, F., and Robin, J.-M. (2014). Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics. *American Economic Review*, 104(6):1551–1596.

Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2023). The Search for Good Jobs: Evidence from a Six-year Field Experiment in Uganda. Technical Report w31570, National Bureau of Economic Research, Cambridge, MA.

Bartel, A. P. (1995). Training, Wage Growth, and Job Performance: Evidence from a Company Database. *Journal of Labor Economics*, 13(3):401–425.

Battiston, A., Williams, R., and Conlon, G. (2020). Apprenticeships and Social Mobility: Fulfilling Potential. Technical report, Social Mobility Commission.

Bento, P. and Restuccia, D. (2017). Misallocation, Establishment Size, and Productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

Britton, J., Espinoza, H., McNally, S., Speckesser, S., Tahir, I., and Vignoles, A. (2020). Post-18 Education - Who Is Taking the Different Routes and How Much Do They Earn? Technical Report 013, Centre for Vocational Education Research.

Brown, C. and Medoff, J. (1989). The Employer Size-Wage Effect. *Journal of Political Economy*, 97(5):1027–1059.

Card, D., Kluve, J., and Weber, A. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3):894–931.

Chiappori, P., Costa-Dias, M., Crossman, S., and Meghir, C. (2020). Changes in Assortative Matching and Inequality in Income: Evidence for the UK. *Fiscal Studies*, 41(1):39–63.

Conti, G. (2005). Training, productivity and wages in Italy. *Labour Economics*, 12(4):557–576.

De Lyon, J. and Dhingra, S. (2020). Firm investments in Skills and Capital in the UK Services Sector. OECD Economics Department Working Papers 1632. Series: OECD Economics Department Working Papers Volume: 1632.

Dearden, L., Reed, H., and Van Reenen, J. (2006). The Impact of Training on Productivity and Wages: Evidence from British Panel Data*. *Oxford Bulletin of Economics and Statistics*, 68(4):397–421.

Den Haan, W. J., Ramey, G., and Watson, J. (2000). Job Destruction and Propagation of Shocks. *American Economic Review*, 90(3):482–498.

Diamond, P. A. (1982). Wage Determination and Efficiency in Search Equilibrium. *The Review of Economic Studies*, 49(2):217.

Elsby, M. W. L. and Michaels, R. (2013). Marginal Jobs, Heterogeneous Firms, and Unemployment Flows. *American Economic Journal: Macroeconomics*, 5(1):1–48.

ESFA (2017a). Apprenticeship Funding: Rules and Guidance for Employers May 2017 to July 2018. *Education and Skills Funding Agency*.

ESFA (2017b). Apprenticeship Technical Funding Guide for Starts from May 2017. *Education and Skills Funding Agency*.

ESFA (2019). Apprenticeship Technical Funding Guide April 2019 to July 2020. *Education and Skills Funding Agency*.

ESFA (2020). Apprenticeship Technical Funding Guide from August 2020. *Education and Skills Funding Agency*.

Flinn, C., Gemici, A., and Laufer, S. (2017). Search, Matching and Training. *Review of Economic Dynamics*, 25:260–297.

Flinn, C. J. (2006). Minimum Wage Effects on Labor Market Outcomes under Search, Matching, and Endogenous Contact Rates. *Econometrica*, 74(4):1013–1062.

Fu, C. (2011). Training, Search and Wage Dispersion. *Review of Economic Dynamics*, 14(4):650–666.

Giupponi, G. and Machin, S. (2022). Labour Market Inequality. Technical report, Institute for Fiscal Studies.

Gregory, V. (2020). Firms as Learning Environments: Implications for Earnings Dynamics and Job Search. Technical report.

Guner, N. and Ruggieri, A. (2021). Misallocation and Inequality. Technical report, University of Nottingham, Centre for Finance, Credit and Macroeconomics (CFCM). Issue: 2021/01.

Hopenhayn, H. A. (2014). Firms, Misallocation, and Aggregate Productivity: A Review. *Annual Review of Economics*, 6(1):735–770.

Jarosch, G. (2023). Searching for Job Security and the Consequences of Job Loss. *Econometrica*, 91(3):903–942.

Konings, J. and Vanormelingen, S. (2015). The Impact of Training on Productivity and Wages: Firm-Level Evidence. *Review of Economics and Statistics*, 97(2):485–497.

Lentz, R. and Roys, N. (2024). Training and Search On the Job. *Review of Economic Dynamics*, 53:123–146.

Lise, J., Meghir, C., and Robin, J.-M. (2016). Matching, sorting and Wages. *Review of Economic Dynamics*, 19:63–87.

Martins, P. S. (2021). Employee Training and Firm Performance: Evidence from ESF Grant Applications. *Labour Economics*, 72:102056.

Merz, M. and Yashiv, E. (2007). Labor and the Market Value of the Firm. *American Economic Review*, 97(4):1419–1431.

Mortensen, D. (1982). The Matching Process as a Noncooperative Bargaining Game. In *The Economics of Information and Uncertainty*, pages 233–258. National Bureau of Economic Research, Inc.

Office For National Statistics (2023). Labour Force Survey. Version Number: 7th Release.

Office For National Statistics (2024). Annual Business Survey Quality and Methodology Information. Version Number: 7th Release.

Patterson, C., Åđahin, A., Topa, G., and Violante, G. L. (2016). Working Hard in the Wrong Place: A Mismatch-Based Explanation to the UK Productivity Puzzle. *European Economic Review*, 84:42–56.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the Black Box: A Survey of the Matching Function. *Journal of Economic Literature*, 39(2):390–431.

Pissarides, C. (1985). Short-run Equilibrium Dynamics of Unemployment Vacancies, and Real Wages. *American Economic Review*, 75(4):676–90.

Pope, T., Shearer, E., and Hourston, P. (2022). Levelling Up and Skills Policy: How Qualifications and Training Can Help Boost Regional Productivity. Technical report, Institute for Government.

Postel-Vinay, F. and Sepahsalari, A. (2023). Labour Mobility and Earnings in the UK, 1992-2017. *The Economic Journal*, 133(656):3071–3098.

Resolution Foundation (2023). Ending Stagnation A New Economic Strategy for Britain. Version Number: 7th Release.

Stole, L. A. and Zwiebel, J. (1996). Organizational Design and Technology Choice under Intrafirm Bargaining. *American Economic Review*, 86(1):195–222. Publisher: American Economic Association.

Turrell, A., Speigner, B., Copple, D., Djumalieva, J., and Thurgood, J. (2021). Is the UK's Productivity Puzzle Mostly Driven by Occupational Mismatch? An Analysis Using Big Data on Job Vacancies. *Labour Economics*, 71:102013.

UK Commission For Employment And Skills and Department For Education (2024). UK Commission's Employer Skills Survey; ESS; Investment in TrainingEmployer Skills Survey, 2011-2019: Secure Access.

A Policy Appendix

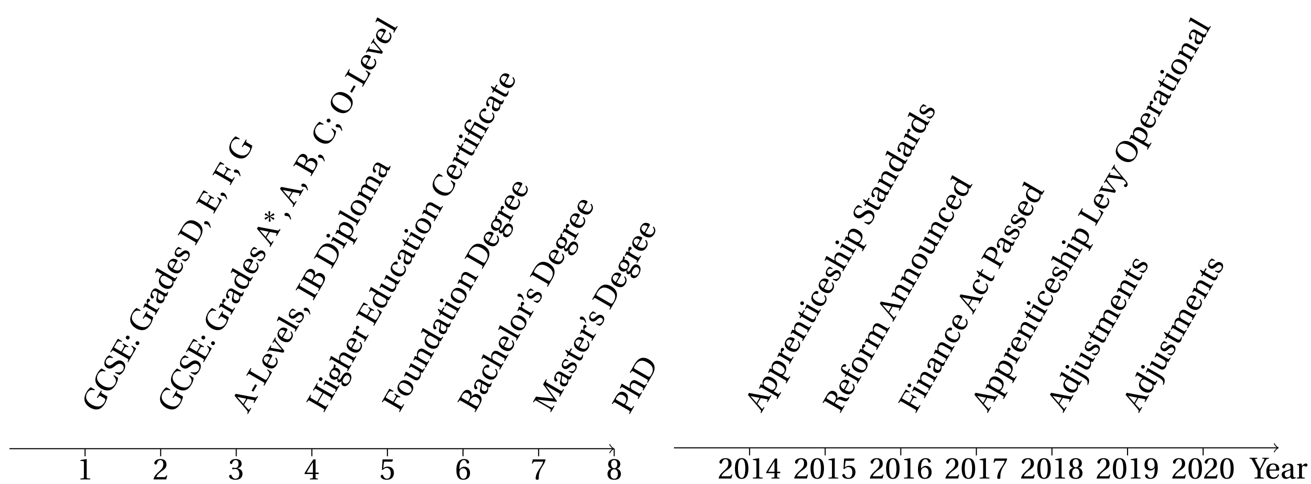
Introduced by the Conservative government in 2016 and enforced since 2017, the Apprenticeship Levy is a modest payroll tax, of 0.5 per cent, on large firms. Such employers can use its proceeds to fund apprenticeships for their workforce within 24 months of the payment, with only a minor co-investment required from the firm. Small businesses do not pay the tax but could receive full funding to provide apprenticeships to workers between 19 and 24.

A.1 Apprenticeships in the UK

The Apprenticeship Levy is an adequate case study for assessing the macroeconomic effects of training programmes across varying levels of human capital. Apprenticeships in the UK entail a combination of receiving practical training and working. They could be used as a substitute for higher education, especially between the 4th and 6th levels of educational attainment in the UK (see figure A.1a). Apprenticeships also operate as complements to those with a higher education background who work and aim to improve their skills. For example, a professional pursuing a part-time master's degree or a CFA certificate is also classified as an apprentice. Seeking to incentivise firms to provide more training to their workers, the UK government introduced the Apprenticeship Levy, a measure streamlining the apprenticeship system in England in 2015. The reform was passed as a part of the *Finance Act* in 2016, beginning to operate from 2017 (Battiston et al., 2020). It was subsequently amended, as shown in figure A.1b.

Figure A.1: The Apprenticeship Levy Reform Timeline and Education Levels in the UK.

(a) Education Outcome Levels in the UK. (b) Apprenticeship Levy: Reform Timeline.



Notes: This is based on Britton et al. (2020).

Notes: The timeline is based on Battiston et al. (2020).

A.2 The Original 2016 Reform

The original reform imposes varying incentives and taxes depending on the firm size, keeping the standard of an apprenticeship uniform across them. Small firms are defined as those with fewer than 50 employees (ESFA, 2017a) and less than £3m in annual payroll.[25] If these conditions are not met, the firm belongs to the larger size category.

The reform harmonises the apprenticeship rules and provides incentives to hire young workers and those from more deprived areas. An apprentice must work at least 30 hours per week, with a minimum of 20 per cent of which is to be spent on the training programme. The minimum duration of the scheme is one year, with the potential for a shorter timeframe under special circumstances (e.g., prior achievements and certificates). An apprentice must receive at least the minimum wage pay depending on their status.[26] Apprenticeships are divided into 15 (30 from 2018) categories of the maximum cost bands, ranging from £1,500 at the lowest to £27,000 at the highest one. Further, modest additional top-ups are provided for the apprentices from the most deprived areas, determined by the *Index of Multiple Deprivation*. The firm receives an extra £600, £300, and £100 of funding towards the training if an apprentice comes from the 10, 20, and 30 per cent most deprived areas, respectively (ESFA, 2017b). The government also provides an apprenticeship grant for employers (AGE) to incentivise businesses to hire those between 16 and 24. To be eligible for the AGE grant of £1500 (for up to 5 employees), the employer must hire an apprentice between 16 and 24 and be classified as a small firm (ESFA, 2017a). The regulation allows for flexibility for the apprentices moving between employers.

The reform imposed new rules on large firms. The levy is collected through the PAYE taxation system every month and deposited in the employer account (with the government). 0.5 per cent of the total payroll of the English workforce is paid. The government tops it with an extra 10 per cent of the final sum, with the final employer account balance determined as:

$$\text{Final Balance} = \text{Monthly Levy} \times \frac{\text{English Workforce}}{\text{Total UK Workforce}} \times 1.1. \tag{A.1}$$

The balance can be used to fund apprenticeships within 24 months of the payment (ESFA, 2017a). Once a firm hires an apprentice, they can use their account balance to pay 90 per cent of the apprenticeship cost within the maximum cost bands. The remaining 10 per cent is to be paid by the employer through the co-investment. The government transfers the training provider 80 per cent of the levy contribution from the employer account in equal monthly payments. The remaining 20 per cent is cashed out once the apprenticeship has been completed (ESFA, 2017b). The final payment is subject to the completion element that needs to be sufficiently documented in the so-called *Individ-*

[25] While the guidance does not directly mention the £3m condition (ESFA, 2017a,b), it is a result of the annual levy allowance imposed by the *Finance Act* of 2006. The regulation stipulates the £15,000 annual levy allowance for the levy payment (per firm). Consider a firm with 50 employees and exactly £3 in annual payroll. Its levy liability stands at exactly £15,000. Taking the allowance under consideration, this means the firm does not pay the levy. This implies that an employer with less or equal to £3m in payroll for up to 50 workers pays no apprenticeship levy.

[26] For example, the minimum wage for an apprentice below the age of 19 or in their first year of the training programme was £3.40 in 2016. For the workers above 21, the minimum pay stood at £6.95 then.

ual Learning Record (ILR). The government fully covers the training costs of English and mathematics up to the 2nd level.

The reform imposes no extra costs on the small firms while providing them with incentives to hire young workers. No levy is collected for firms with an English workforce below or equal to 50 people and with less than £3m in annual payroll. If such a small firm hires an apprentice, the government covers the full cost of the apprenticeship, up to the maximum cost band if they are between 16 and 19. If the worker is between 19 and 24, the government covers the full cost subject to extra conditions (ESFA, 2017a). No funding is provided for the apprentices above 24. The eligibility is based on the firm's size at the beginning of the apprenticeship.

A.3 Changes Since 2016

The government has introduced several minor adjustments since the reform was passed. First, the new co-investment required was announced in 2019, with the large firms required to contribute 5 per cent of the apprenticeship cost instead of 10 per cent as before. Second, employers have been able to transfer 25 per cent of their levy account to other firms in their ownership network since 2019 (as opposed to the previous limit of 10 per cent). Third, the government introduced minor legal and reporting requirements. The HM Revenues and Customs filing policies were adjusted in 2019. The reporting rules were streamlined in 2019 as well (ESFA, 2019). In 2020, the apprenticeship standards were tightened, largely to abide by the 2014 framework (ESFA, 2020).

B Model Appendix

B.1 The Surplus and Joint Match Value Functions

I define the match values at the beginning and end of the period:

$$M(z,\xi,a) = J^e(z,\xi,a) + V(z,\xi,a); \text{ and} \tag{B.1a}$$

$$M^h(z,\xi,a) = J^{e,h}(z,\xi,a) + V^h(z,\xi,a), \tag{B.1b}$$

which are the same as in Guner and Ruggieri (2021).

Define the surplus functions at the beginning and end of the period that satisfy:

$$M(z,\xi,a) = J^e(z,\xi,a) + V(z,\xi,a) = S(z,\xi,a) + J^u(a) \implies \tag{B.2a}$$

$$S(z,\xi,a) = J^e(z,\xi,a) + V(z,\xi,a) - J^u(a); \text{ and} \tag{B.2b}$$

$$M^h(z,\xi,a) = J^{e,h}(z,\xi,a) + V^h(z,\xi,a) = S^h(z,\xi,a) + J^{u,h}(a) \implies \tag{B.2c}$$

$$S^h(z,\xi,a) = J^{e,h}(z,\xi,a) + V^h(z,\xi,a) - J^{u,h}(a). \tag{B.2d}$$

Note that this definition of the period-beginning surplus breaks the relationship from Guner and Ruggieri (2021): $S(z,\xi,a) = \mathbb{I}^h(z,\xi,a) S^h(z,\xi,a)$ (due to the presence of the on-the-job search).

These definitions allow me to find the recursive relationship between $M(z,\xi,a)$ and $M^h(z,\xi,a)$. Start by expanding the match value:

$$\begin{aligned} M^h(z,\xi,a) &= w(z,\xi,a) + \frac{(1-\delta_w)}{1+r}\left[\delta_f + \left(1-\delta_f\right)\delta_s\right] J^{u,h}(a) + \\ &+ \frac{(1-\delta_w)}{1+r}\left\{1 - \left[\delta_f + \left(1-\delta_f\right)\delta_s\right]\right\}\left\{p^h(z,\xi,a) J^e(z,\xi,a+\Delta_a) + \left[1-p^h(z,\xi,a)\right] J^e(z,\xi,a)\right\} + \\ &+ r(z,a) - w(z,\xi,a) + \frac{1-\delta}{1+r}\left\{-\mathbb{I}^t(z,\xi,a)\xi + p^h(z,\xi,a) V(z,\xi,a+\Delta_a)\right. \\ &\left. + \left[1-p^h(z,\xi,a)\right] V(z,\xi,a)\right\}. \end{aligned} \tag{B.3}$$

After re-arranging, it follows that:

$$\begin{aligned} M^h(z,\xi,a) &= \frac{(1-\delta_w)}{1+r}\left[\delta_f + \left(1-\delta_f\right)\delta_s\right] J^{u,h}(a) + r(z,a) + \\ &+ \frac{1-\delta}{1+r}\left\{-\mathbb{I}^t(z,\xi,a)\xi + p^h(z,\xi,a) M(z,\xi,a+\Delta a) + \left[1-p^h(z,\xi,a)\right] M(z,\xi,a)\right\}. \end{aligned} \tag{B.4}$$

B.2 Equilibrium

Definition B.1 (Stationary Recursive Equilibrium) *A stationary recursive competitive equilibrium consists of:*

- *workers' value functions for unemployment and employment, given by equations* (3.7) *and* (3.9)*;*
- *firms' value functions for active jobs, given by equation* (3.12)*;*
- *the policy functions for job creation, training, entry, and vacancy posting, given by equations* (3.21), (3.25), (3.19a), *and* (3.18)*;*
- *the wage schedule, given by equation* (3.23)*;*
- *the job contact probabilities for non-employed and employed workers and firms, given by equations* (3.6b), (3.6c), *and* (3.6a)*;*
- *the unemployment rate, U;*
- *the distribution of employed and non-employed workers across states, $\psi(z,\xi,a)$ and $\psi^u_a(a)$;*
- *the distribution of open vacancies and firms across states, $\psi_v(z,\xi)$ and $\psi(z,\xi)$; and*
- *the tax and subsidy schedules for a given productivity threshold $\hat{z}$, τ and λ,*

such that:

1. ***optimality**: the value functions for workers and firms, given by equations* (3.7), (3.9), *and* (3.12), *attain their maxima;*

2. ***bargaining**: the wage schedule solves problem* (3.23)*;*

3. ***training**: the training decision solves problem* (3.24)*;*

4. ***market clearing**: the goods and labour markets are cleared;*

5. ***measure of entrants**: for all Borel sets $\mathscr{Z} \times \mathscr{E} \subset \mathbb{R}^2_+$ it must be that:*

$$N(\mathscr{Z} \times \mathscr{E}) = N_e \iint_{(z,\xi)\in\mathscr{Z}\times\mathscr{E}} \mathbb{I}^e(z,\xi)\psi(z,\xi)\, dz\, d\xi, \tag{B.5}$$

   where $\mathbb{I}^e(z,\xi)$ represents the solution to problem (3.19a)*;*

6. ***measure of incumbent**: for all Borel sets $\mathscr{Z} \times \mathscr{E} \subset \mathbb{R}^2_+$ it must be that:*

$$N^\star(\mathscr{Z} \times \mathscr{E}) = \frac{N(\mathscr{Z} \times \mathscr{E})}{\delta_f}; \tag{B.6}$$

7. ***aggregate consistency**: the workers' and vacancies' distributions replicate themselves through the workers' and firms' policy functions; and*

8. ***balanced budget condition**: the tax and subsidy schedules and the productivity threshold, τ, λ, and $\hat{z}$, are set such that the government budget, given by equation* (3.26), *remains balances.*

B.3 Algorithm

B.3.1 Equilibrium Algorithm Without Taxes

The following is an adaptation of the algorithm in Guner and Ruggieri (2021). I discretise the space into equally spaced grids:

- 50 for firm productivity (a);
- 20 for firm-specific training costs (ξ); and
- 60 for worker human capital.

1. Guess the labour market tightness, θ_0. Compute the associated contact rates:

(B.7a)
$$\phi_w^{e,0} = \gamma\left[1+(\theta_0)^{-\eta}\right]^{-\frac{1}{\eta}},$$

(B.7b)
$$\phi_w^{u,0} = \left[1+(\theta_0)^{-\eta}\right]^{-\frac{1}{\eta}},$$

and

(B.7c)
$$\phi_f^{0} = \left[1+(\theta_0)^{\eta}\right]^{-\frac{1}{\eta}}.$$

2. Guess the distribution of vacancies, $\psi_v^0(z,\xi)$.

   (a) Solve for the end-of-period surplus value function, $S^h(z,\xi,a)$ based on the guesses of $\phi_w^{e,0}$, $\phi_w^{u,0}$, ϕ_f^0, and $\psi_v^0(z,\xi)$.

   (b) Compute the match formation and training decision functions, $\mathbb{I}^h(z,\xi,a)$ and $\mathbb{I}^t(z,\xi,a)$ given the surplus value function $S^h(z,\xi,a)$.

   (c) Use the above results to simulate a large panel of workers, which can be used to construct:

   - the distributions of employed and non-employed workers ($\psi_u(a)$ and $\psi_w(z,\xi,a)$),
   - the aggregate measures of employed and non-employed workers (E and U), and
   - the aggregate measure of searchers (S).

   (d) Use the above results to compute the vacancy decision function, $v(z,\xi)$.

   (e) Use the above to calculate the entry decision, $\mathbb{I}^e(z,\xi)$.

   (f) Use the above to compute the new guess of the distribution of vacancies, $\psi_v^1(z,\xi)$.

   (g) Check for convergence: $\psi_v^0(z,\xi)$ and $\psi_v^1(z,\xi)$ are close enough, save $\psi_v^\star(z,\xi)$. Otherwise, set $\psi_v^0(z,\xi) = \psi_v^1(z,\xi)$ and return to setp 2.

   (h) Iterate till convergence.

3. Compute the measure of entrants:

(B.8)
$$N = N_e \iint\limits_{(z,\xi)\in\mathscr{Z}\times\mathscr{E}} \mathbb{I}^e(z,\xi)\psi_z(z)\psi_\xi(\xi)\,\mathrm{d}z\,\mathrm{d}\xi,$$

and use the stationarity to compute the total number of incumbent firms:

$$N^{\star}\delta_f = N. \tag{B.9}$$

4. Compute the aggregate measure of the vacancies posted:

$$v = N \iint_{(z,\xi)\in\mathscr{Z}\times\mathscr{E}} \mathbb{I}^e(z,\xi)\psi_z(z)\psi_\xi(\xi)\,\mathrm{d}z\,\mathrm{d}\xi. \tag{B.10}$$

5. Combine the above results with the matching function to obtain a new guess for the contact rates of non-employed workers, $\phi_w^{u,1}$ (other contact rates can be computed from it).

6. Iterate till convergence.

C Estimation Appendix

C.1 Data

I collect the worker-side based on the *Labour Force Survey* (LFS) (Office For National Statistics, 2023). The data for January 2010 - June 2016 come from 12 separate waves of the survey as indicated in Table C.1. The employer data come from the 2011 *Employer Skills Survey* (UK Commission For Employment And Skills and Department For Education, 2024). Table C.2 lays out sources of all the key moments used in the calibration procedure.

Table C.1: Summary of the *Labour Force Surveys* Waves Collected.

Dates (From)	Dates (To)	Serial Number
April-June Datasets		
2010	2011	7026
2011	2012	7157
2012	2013	7379
2013	2014	7722
2014	2015	7790
2015	2016	8042
January-March Datasets		
2010	2011	6796
2011	2012	7035
2012	2013	7279
2013	2014	7504
2014	2015	7729
2015	2016	7988

Table C.2: MSS-Calibration Targets: Sources.

Target Moment	Expression	Source
Firm-Level Employment.		
Average (Log) Size	$\mathbb{E}\,l,\, \mathbb{E}\big(\log l\big)$	Firm Survey (2011)
St. Deviation Log Size	$\sigma_{\log l} = \sqrt{\mathbb{V}\big(\log l\big)}$	Firm Survey (2011)
Firm-Size Distribution & Percentiles		
Firm-Size Bins	$\mathbb{P}(Q_{i-1} \leq l \leq Q_i),\ i \in \{1,\ldots,6\}, Q_0 = 0$	Firm Survey (2011)
Firm-Size Percentiles	$Q_i = \frac{i}{7},\ i \in \{1,\ldots,6\}$	Firm Survey (2011)
Extra Firm-Size Percentiles	$Q_i = \frac{i}{7},\ i \in \{1,2\}$	Firm Survey (2011)
Firm-Size Training Shares		
Firms with 1-49 Employees	$\mathbb{E}\left(\frac{\text{\# Training Firms}}{\text{\# Firms}}\middle\vert\, l \leq 49\right)$	Eurostat CVTS
Firms with 50-249 Employees	$\mathbb{E}\left(\frac{\text{\# Training Firms}}{\text{\# Firms}}\middle\vert\, 20 \leq l \leq 250\right)$	Eurostat CVTS
Firms with 250+ Employees	$\mathbb{E}\left(\frac{\text{\# Training Firms}}{\text{\# Firms}}\middle\vert\, 250 \leq l\right)$	Eurostat CVTS
Wage Profiles (Age $\in [25, 65]$ Years)		
Relative Wage at Entry	$\mathbb{E}\log w_{0\text{ Years}} - \mathbb{E}\log w$	LFS
Relative Wage After 20 Years	$\mathbb{E}\log w_{20\text{ Years}} - \mathbb{E}\log w$	LFS
Relative Wage at Re-Employment	$\mathbb{E}\log w_{\text{Re-Employment}} - \mathbb{E}\log w$	LFS
Dispersion of Wage at Entry	$\sqrt{\mathbb{V}\big(\log w_{0\text{ Years}} - \log w\big)}$	LFS
Dispersion of Wage After 25 Years	$\sqrt{\mathbb{V}\big(\log w_{25\text{ Years}} - \log w\big)}$	LFS
Dispersion of Wage at Re-Employment	$\sqrt{\mathbb{V}\big(\log w_{\text{Re-Employment}} - \log w\big)}$	LFS
Aggregate Worker Statistics		
Average Job Duration (Years)	$\frac{\mathbb{E}\,l}{4}$	LFS
Average Tenure (Quarters)	$\mathbb{E}\,l$	LFS
Average Non-Employment	N	LFS
Home Production Share	b	OECD
Worker-Level Training Return (Regressions)		
Training Share	$\mathbb{E}\,\mathbb{I}^t(z,\xi,a)$	LFS
Training Return (β)	$\log w_{i,t} = \beta\,\mathbb{I}^t_{i,t-1} + \varepsilon_{i,t}$	LFS, from Guner and Ruggieri (2021)
Worker-Level Tenure Return (Age $\in [25, 65]$ Years)		
Relative Wage: 3-12 Quarters	$\mathbb{E}\log w_{[3,12]} - \mathbb{E}\log w_{[0,3]}$	LFS
Relative Wage: 12-24 Quarters	$\mathbb{E}\log w_{[12,24]} - \mathbb{E}\log w_{[0,3]}$	LFS
Relative Wage: 12-24 Quarters	$\mathbb{E}\log w_{>24} - \mathbb{E}\log w_{[0,3]}$	LFS

Note: Firm Survey (2011) refers to the 2011 *Employer Skills Survey* (UK Commission For Employment And Skills and Department For Education, 2024). The CVTS refers to Eurostat's *Continuing Vocational Training Survey* and I obtain these moments from Guner and Ruggieri (2021) (as I do not have access to that survey). The LFS stands for the *Labour Force Survey* (Office For National Statistics, 2023).

C.2 Estimation Figures

Table C.3: Worker Moments and Their Fit.

Moment	Data	Model
Non-Employment	0.23	0.39
Avg. Tenure, Quarters	8.33	7.27
Benefits-Wage Ratio	0.35	0.37
Avg. Log Wage at Entry	-0.24	-0.56
Avg. Log Wage after 20 Years	0.07	0.12
Avg. Re-Emp Log Wage	-0.23	-0.28
Diff. Return from Tenure: 4-12 m. - 0-3m	0.05	0.03
Diff. Return from Tenure: 12-24 m. - 0-3m	0.09	0.21
Diff. Return from Tenure: 24+ m. - 0-3m	0.28	0.66
Std. Log Wage at Entry	0.57	0.62
Std. Log Wage after 25y.	0.78	0.84
Std. Log Ee-Emp Wage	0.83	0.79
Training Premium	0.20	0.25
Job-to-Job Transitions, Rate	0.02	0.02

Figure C.1: Estimation General Fit.

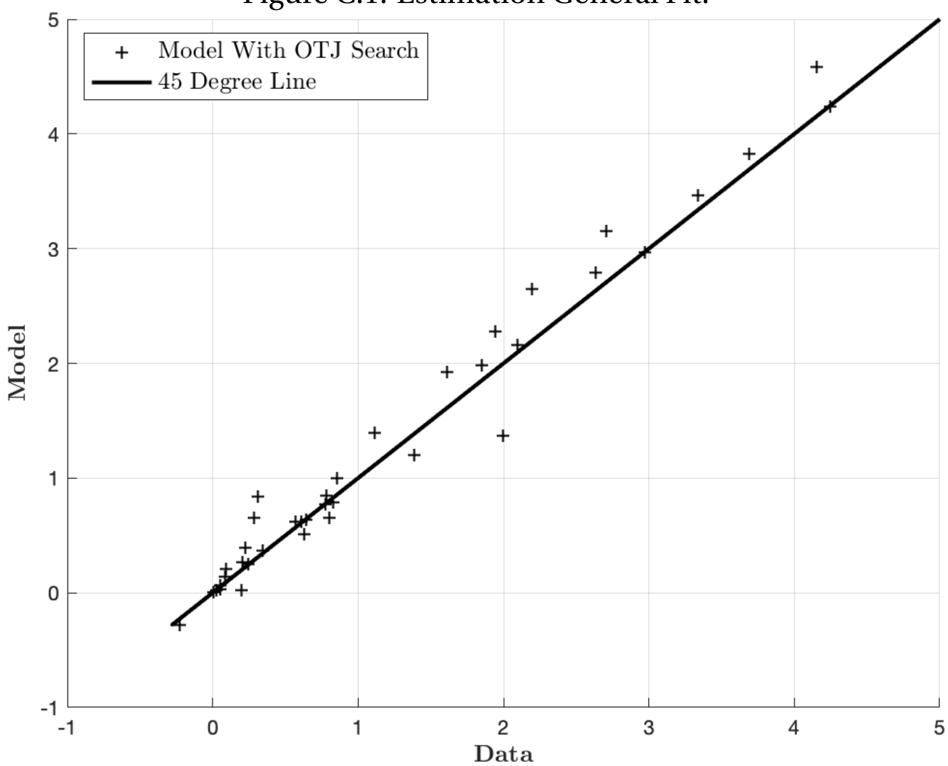
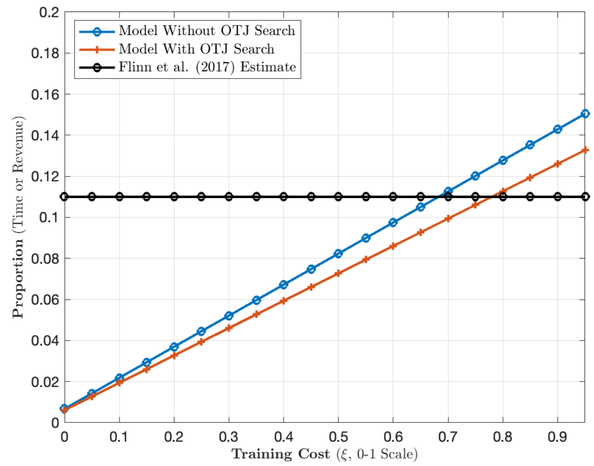


Table C.4: Firm Moments and Their Fit.

Moment	Data	Model
$\mathbb{E}(\log l)$, Annual	2.10	2.16
St. Dev$(\log l)$, Annual	1.11	1.39
$\mathbb{E}\, l$, Annual	19.55	19.52
#Firms, 1-9 Emp	0.63	0.51
#Firms, 10-24 Emp	0.20	0.26
#Firms, 25-49 Emp	0.08	0.14
#Firms, 50-99 Emp	0.05	0.06
#Firms, 100-249 Emp	0.02	0.02
#Firms, 250+ Emp	0.01	0.00
Firm Size Pct, 10th	2.00	1.37
Firm Size Pct, 25th	4.00	3.32
Firm Size Pct, 50th	7.00	9.80
Firm Size Pct, 75th	15.00	23.44
Firm Size Pct, 90th	40.00	45.75
Firm Size Pct, 95th	70.00	69.42
Firm Size Pct, 99th	206.00	145.47
Firm Size Pct, 40th	5.00	6.86
Firm Size Pct, 60th	9.00	14.07
Training Firms, Share	0.65	0.63
Training Firms, Share 1-49	0.61	0.62
Training Firms, Share 50-249	0.78	0.77
Training Firms, Share 250+	0.86	1.00
Trained Workers within Firms, Share	0.44	0.51

C.3 Parameters Analysis

Figure C.2: Comparing Training Cost Proportions with Flinn et al. (2017).



Notes: The model without OTJ corresponds to the specification outlined in section 3 with the employed workers' search effort parameter set to 0, $\gamma = 0$ and to the framework in Guner and Ruggieri (2021). The black line is based on Flinn et al. (2017). It represents the average proportion of time spent on training by the workers (without distinguishing across the training cost). The blue and red lines represent the training cost divided by the average revenue (across all firms). It could also be represented as the proportion of time lost to the training (assuming that the revenue lost is directly proportional to the time allocated to training instead of work).

Table C.5: Jacobian of the Loss Function With Respect to the Parameters.

Parameter	Derivative of Loss Function
b	45.1056
ϕ_w	73.8382
c_e	1.0027
$\underline{\xi}$	79.0727
$\overline{\xi}$	104.6139
σ_a	103.9595
σ_z	36.2843
p^e	36.4769
p^t	97.0036
p^d	105.2084
β	71.2438
λ_2	182.0765
δ_s	109.5094
γ	69.6234

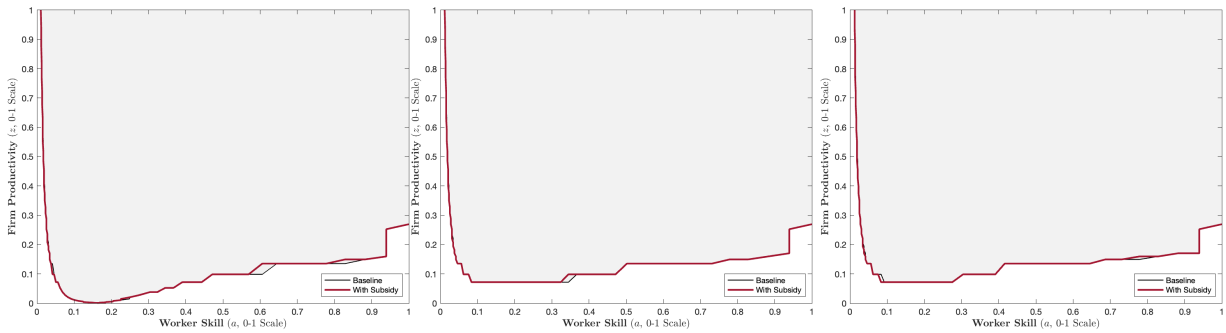
Notes: The Jacobian is estimated by monitoring the objective function's reaction to small changes in one of the parameters. Then, the outcome is used to compute the first derivative from its definition: $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$.

D Results Appendix

D.1 Policy Functions

Figure D.1: Countours of Matching Policy Function: $\mathbb{I}^h(z, \xi, a)$.

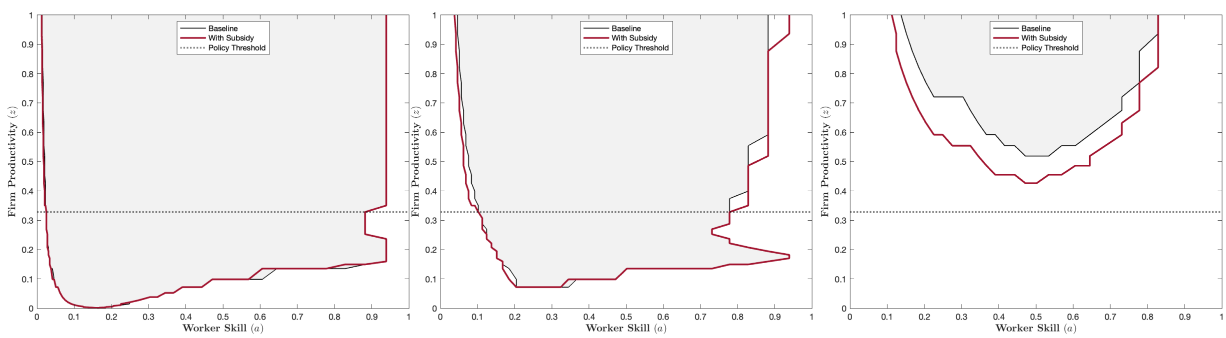
(a) Low Training Costs: ξ_2. (b) Medium Training Costs: ξ_8. (c) High Training Costs: ξ_{18}.



Notes: The grey areas represent the region of accepted matches without the training subsidy. The maroon contours indicate how the boundary changes following the policy's introduction. Figures D.1a, D.1b, and D.1c represent the matching functions at low, medium, and high training costs, respectively. The plots are based on equation (3.21).

Figure D.2: Countours of Training Policy Function: $\mathbb{I}^t(z, \xi, a)$.

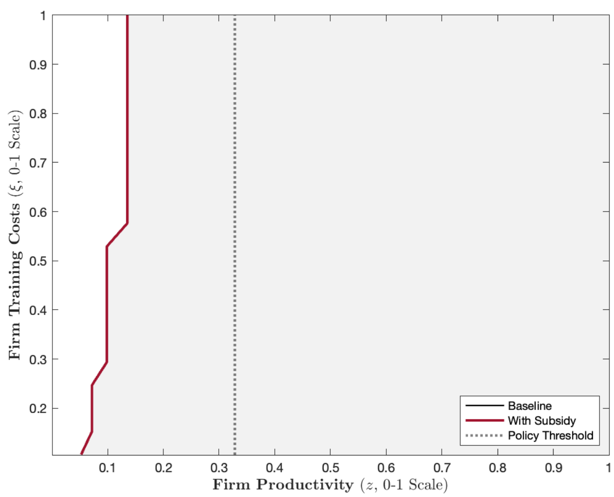
(a) Low Training Costs: ξ_2. (b) Medium Training Costs: ξ_8. (c) High Training Costs: ξ_{18}.



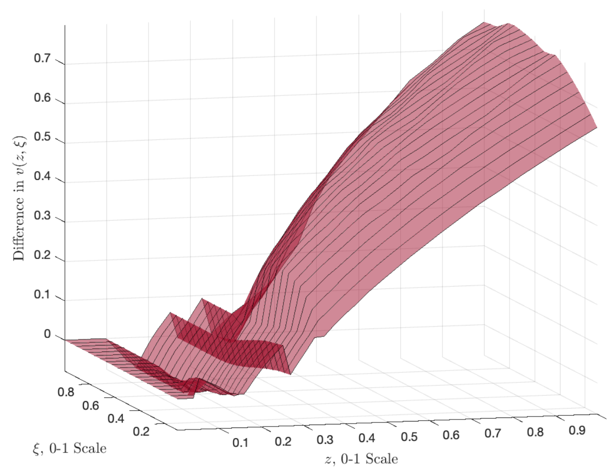
Notes: The grey areas represent the region of positive training decisions (conditional on accepting the match) without the training subsidy. The maroon contours indicate how the boundary changes following the policy's introduction. Figures D.2a, D.2b, and D.2c represent the training policy functions at low, medium, and high training costs, respectively. The plots are based on equation (3.25).

Figure D.3: Entry and Vacancy Posting Policy Functions Summary.

(a) Entry Policy Function: $v(z,\xi)$.



(b) Difference in Vacancy Posting With and Without the Subsidy.

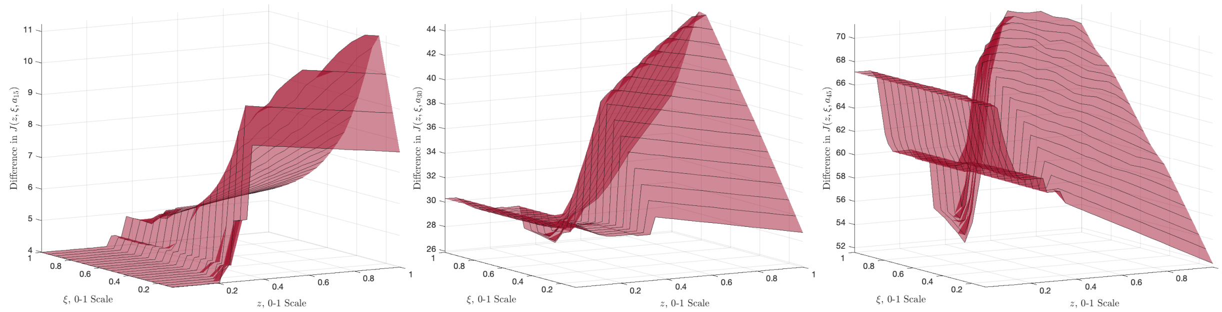


Notes: Figure D.3a displays the contour of the firm entry policy function. The grey areas represent the region of positive training decisions (conditional on accepting the match) without the training subsidy. The maroon contours indicate how the boundary changes in the policy's presence. Figure D.3b presents the differences in the vacancy posting policy function between its values with and without the policy.

D.2 Value Functions

Figure D.4: Differences between With and Without-Levy Value Functions of Employment.

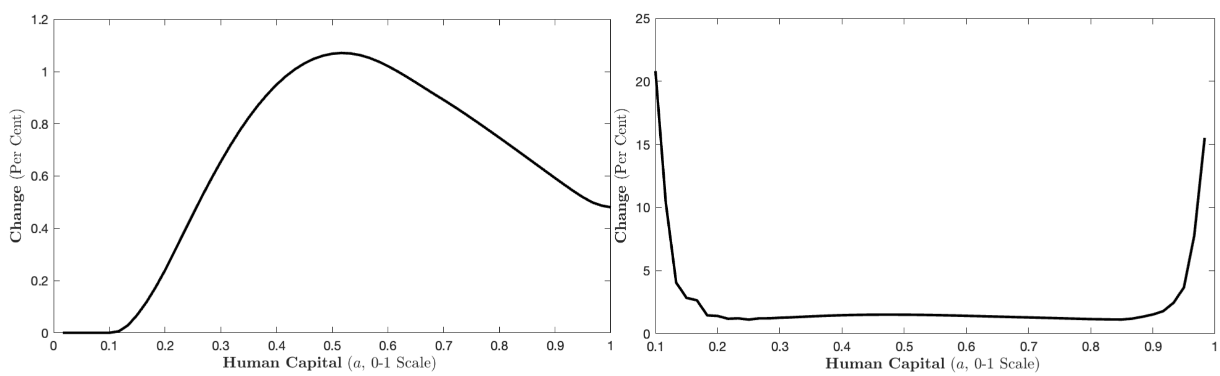
(a) Low Skill. (b) Medium Skill. (c) High Skill.



Notes: The figures present surfaces of the differences in the job value functions between their values with and without the policy: $\Delta^\tau J^e(z,\xi,a) \equiv J^e(z,\xi,a \mid \tau = 0.05) - J^e(z,\xi,a \mid \tau = 0)$, based on equation (3.9). Positive values along the vertical axis indicate that the value of employment increases following the subsidy's introduction.

Figure D.5: The Difference between the With and Without-Levy Option Values of Search.

(a) Non-Employed Workers. (b) Employed Workers (Expected Value).



Notes: The left figure shows the change between the option values of search for the non-employed workers between the scenarios with and without the subsidy, based on equation (3.7). The right figure presents the difference between the mean values of the option value of search for employed workers between the subsidy and baseline scenarios, based on equation (3.9). Positive values indicate that, on average, the expected value of an alternative job offer is larger with the subsidy. Missing values for $a \in [0, 0.1) \cup (0.95, 1]$ indicate that in both cases the expected values of alternative job offers are zero. The reason is twofold. The least skilled workers are not accepted by any employer in either scenario. The most skilled workers accept only the jobs at the most productive firms, implying that they account for no alternative job offers.

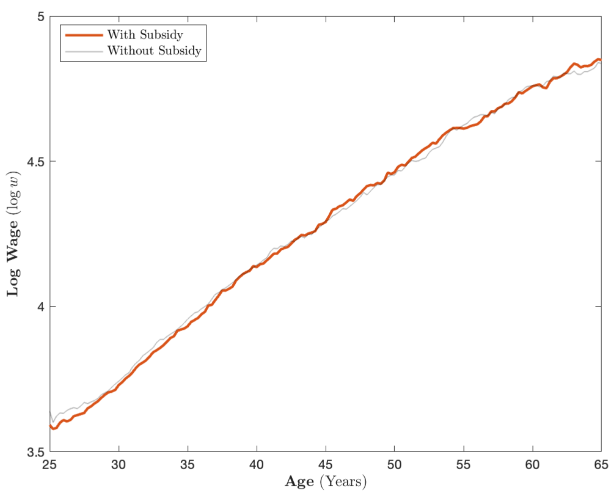
D.3 Firm and Workers Outcomes

Table D.1: Firm Aggregate Statistics.

Firm Statistic	Baseline	With Subsidy
Avg. Size	19.99	20.26
St. Dev. Size	30.52	30.91
Skewness Size	4.75	4.73
Kurtosis Size	40.69	40.58
Avg. Pre-Tax Revenue	6755.30	6875.98
Avg. Post-Tax Revenue	6755.30	6858.79
St. Dev. Revenue	19957.17	20319.90
P90-50 Revenue	1.95	1.99
P50-10 Revenue	2.25	2.29
Labour Productivity CV	0.64	0.65

Figure D.6: Lifecycle Wages.

(a) Comparing Environments With and Without the Subsidy.



(b) Mechanisms Decomposition (Differences from Baseline).

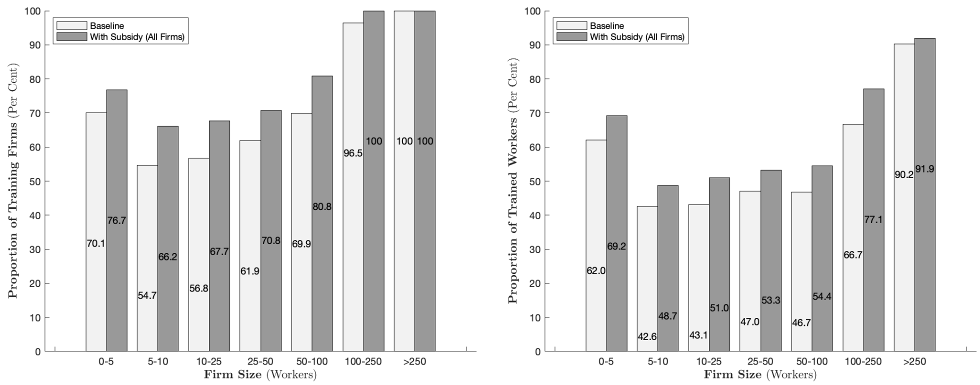
Notes: Figure D.6a shows the log wages with and without the government training subsidy programme. Figure D.6b shows the decomposition of the change into the three mechanisms. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy. The blue line plot indicates the adjustment when only the job value effect is active. The green line highlights the change when the job value and distribution effects are operational. As a result, comparing the blue and green lines allows for assessing the distribution effect alone. Finally, the orange line shows the impact of all three mechanisms, which is equivalent to the policy's total impact.

Figure D.7: Variance of Log Wages With and Without the Subsidy.

D.4 Policy Extensions

Figure D.8: Extensive Margin Expansion: All Firms Eligible - Training Incidence by Firm Size

(a) Training Extensive Margin. (b) Training Intensive Margin.



Notes: Figure D.8a features the proportion of training their workers in each category. For example, a value of 100 means that all firms in that category provide at least a quarter of training to at least one worker (the extensive margin). Figure D.8b shows the proportion of workers receiving training in each category of firms (the intensive margin).

Table D.2: Extensive Margin Expansion: All Firms Eligible.

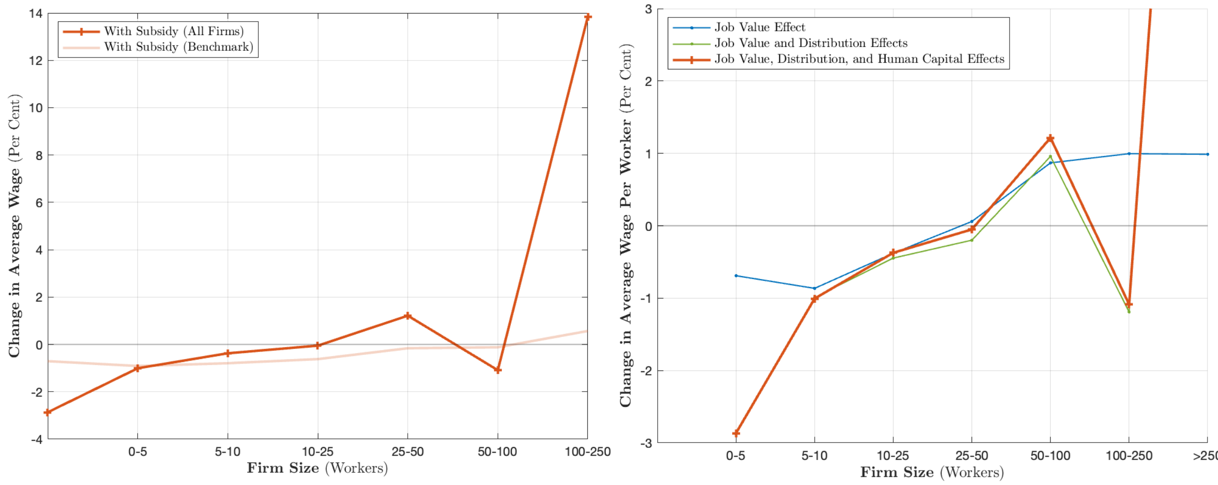
Firm Statistic	Baseline	With Subsidy
Avg. Size	19.99	17.58
St. Dev. Size	30.52	27.22
Skewness Size	4.75	5.54
Kurtosis Size	40.69	63.18
Avg. Pre-Tax Revenue	6755.30	5623.77
Avg. Post-Tax Revenue	6755.30	5609.71
St. Dev. Revenue	19957.17	17906.45
P90-50 Revenue	1.95	1.99
P50-10 Revenue	2.25	3.58
Labour Productivity CV	0.64	0.67

Notes: Table D.2 shows firm statistics for the environment with the policy extended to all firms.

Figure D.9: Difference between Average Wage With Expanded Subsidy and Baseline.

(a) Expanded and Benchmark Subsidies.

(b) Mechanisms Decomposition.



Notes: Figure D.9a contrasts the average wage per firm category for three scenarios: the extended subsidy (to all firms), the benchmark subsidy, and the baseline. In both cases with the training incentive policy, the subsidy level is determined via the balanced budget condition in equation (3.26). Figure D.9b shows the decomposition of the change (relative to the baseline) in average wage into the three mechanisms. To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy. The blue line plot indicates the change when only the job value effect is active. The green line highlights the change when the job value and distribution effects are operational. As a result, comparing the blue and green lines allows for assessing the distribution effect alone. Finally, the orange line shows the impact of all three mechanisms, equivalent to the policy's total impact.

Table D.3: Inequality Statistics of the Extensive Margin Expansion: All Firms Eligible.

Inequality Statistics	Baseline	Job Value Effect	Job Value and Distribution Effects	Job Value, Distribution, and Human Capital Effects
Gini	0.421	0.419	0.419	0.417
P90-50	2.804	2.747	2.803	2.865
P50-10	4.472	4.493	4.472	4.574
Mean-Median	1.274	1.251	1.272	1.250
Var. of Log Wage	0.823	0.814	0.816	0.823

Notes: The subsidy level is determined via the balanced budget condition in equation (3.26). To isolate the job value effect, I compute the outcomes based on the baseline distribution of firms and the policy functions of training, vacancy posting, and entry. The remaining policies and value functions are taken from the training incentive environment. I add the distribution effect by using the environment above with the entry and vacancy posting policies taken from the subsidy-present economy.