Training and Search On the Job*

Rasmus Lentz†　　Nicolas Roys‡

December 20, 2023

Abstract

The paper studies human capital accumulation over workers' careers in an on-the-job search setting with heterogeneous firms. In renegotiation-proof employment contracts, more productive firms provide more training. General and specific training both induce higher wages within jobs and with future employers, even conditional on the future employer type.

Because matches do not internalize the specific capital loss from employer changes, specific human capital can be over-accumulated, more so in low type firms. The analysis also establishes that general training can be efficient regardless of the level of labor market frictions.

We calibrate the model to the US economy using Compustat and NLSY79. While validating the Acemoglu and Pischke (1999) mechanisms, the analysis nevertheless arrives at the opposite conclusion: increased labor market friction reduces training in equilibrium.

Keywords: Wage contracts, human capital, training, wage dispersion, frictional labor markets, optimal contract design, firm heterogeneity, sorting.

JEL codes: D21, D43, D83, E24, J24, J31, J33, J41, J62, J63, J64

---

*We are grateful for helpful comments from Daron Acemoglu, Joseph Altonji, Melvyn Coles, John Kennan, Fabian Lange, Theodore Papageorgiou, Pascual Restrepo, and Chris Taber. Joseph Han provided highly capable research assistance. Rasmus Lentz acknowledges financial support from the National Science Foundation, grant No. SES-1459897.

†University of Wisconsin-Madison; LMDG; CAP; E-mail: rlentz@wisc.edu

‡Royal Holloway, University of London; E-mail: nicolas.roys@rhul.ac.uk

1 Introduction

Human capital is well recognized as a primary determinant of earnings and inequality. To improve our understanding of wage dynamics over workers' careers, we propose a framework of on-the-job training in a frictional labor market with firm heterogeneity and long-term contracts. The framework casts light on a number of significant questions such as the following: How are the gains and costs of training shared between firms and workers? Are labor market frictions detrimental to training? Are training levels socially efficient? Do training levels differ across firms? The analysis contributes in part to the theoretical understanding of active on the job training in frictional labor markets. It also calibrates the model to the US economy to quantify the relative magnitude of the different effects in the model on wage and training outcomes.

The frictional environment includes on-the-job search, firm heterogeneity and direct competition between employers for a worker's services. In the United States, worker reallocation between firms most commonly happens without an intervening unemployment spell.[1] Christensen et al. (2005) show that job-to-job transitions are motivated by the worker's search for the higher wages that come with more productive jobs. Match surplus heterogeneity is a fundamental motivation for labor market churning. We show that it has strong implications for training and wage determination. Following the analyses in Burdett and Mortensen (1998), Postel-Vinay and Robin (2002), Bagger et al. (2014), and Bagger and Lentz (2019) we maintain the assumption that match surplus heterogeneity is systematically related to firm type heterogeneity.

In our setup, firms design employment contracts to maximize profits subject to a given utility promise to the worker. Training results in an increased valuation of the worker across all potential employers. The occasional arrival of outside offers partially delivers the ex post training valuation increase to the worker through higher wages, either within the job if the current employer can retain the worker or in a new match if it cannot.[2] Thus, the provision of training implies a promise of increased future utility to the worker. As emphasized in Becker (1964), optimal training provision given these limited commitment employment contracts involves the exchange of future higher utility for the worker with present lower wages. The

[1] From Rogerson and Shimer (2011), the Current Population Survey implies an annual employment to employment rate of about 0.31 for the period 2000-05. The corresponding employment to unemployment hazard is 0.24. Finally, the unemployed job finding rate is 4.3 at an annual frequency. These numbers are based on a competing hazards interpretation of the monthly transition probabilities. As documented in Fujita and Moscarini (2013), the numbers over-estimate reallocation to new employers through the unemployment channel because a large part of the outflow from unemployment consists of recalls.

[2] The employment contracts are required to be renegotiation-proof which in combination with limited commitment implies that they respond to outside competition in a manner that is isomorphic to the offer-matching process in Postel-Vinay and Robin (2002).

firm could be constrained in its ability to make this exchange. For instance, a minimum wage implies a zero cost of adjusting the wage down to the minimum and an infinite adjustment cost further down. We choose a generalization of this cost through the introduction of risk-averse workers and the implied imperfect intertemporal substitution of wages.[3] The degree of this limitation is governed by the risk aversion parameter, which determines the magnitude of the backloading cost.

We generalize the Acemoglu and Pischke (1999) argument that increased friction can result in more training because it reduces the upward competitive pressure on wages after training. We confirm their argument as a partial equilibrium result in our setup. However, frictions impact training through two other channels in the model: Worker bargaining position and firm heterogeneity. These channels are sufficiently potent to alter the aggregate training response to changes in frictions. Specifically, increased friction implies an equilibrium shift in the match composition toward greater mismatch and a worsening of worker bargaining positions which are both associated with less training in our calibrated equilibrium. Indeed, we find that aggregate training declines given a counterfactual increase in frictions.

It is important to note that this result is not merely a case of adding independent effects to an existing model that happen to run counter to the original model. The bargaining position and mismatch effects naturally interact with the cost of backloading. To wit, the classical solution is to sell the job to the worker. In our setting, if the worker has full bargaining power the same result obtains. Bargaining power is in most frictional models an endogenous outcome and in our setting, increased friction results in a worsening of the worker's bargaining position.

In our analysis, inefficiencies in training arise from what we term the future employer externality: It is possible that the current match does not internalize training's impact on future employers' profits.[4] In our setting, the future employer externality arises when the worker's human capital is valued differently by the future employer relative to the current employer. This in turn depends on the shape of match production function. In the case where the production function is supermodular in the worker's human capital and the firm's productivity, the future employer associated with a job-to-job move values it more by virtue of being more productive than the current employer. In this case, the future employer externality dictates an underinvestment. If the match production function is submodular,

[3]The modeling choice has the advantage that the underprovision of training associated with this version of the hold-up problem does not have to rely on an assumption of some form of internal dysfunction between the worker and the firm. Contracts are jointly efficient. Consequently, the results are invariant to whether it is the firm or the worker that pays training costs.

[4]This happens to be the key source of underinvestment in Acemoglu (1997).

the reverse is true and there is an overinvestment. If the production function is modular, all employers value the worker's human capital by the same and the future employer externality is not present.

Our analysis has a special case where training is efficient regardless of frictions: If workers are risk neutral and the production function is modular, both the cost of backloading and the future employer externality channels are shut down, and firms choose the socially optimal training level. The result is striking particularly in light of Acemoglu (1997) who argues that underinvestment is a general feature of frictional labor markets. Our analysis demonstrates that the results in Acemoglu (1997) are not due to frictions per se, but rather the choice of the frictional wage determination mechanism. In Acemoglu (1997) wages are set by bargaining where the worker's outside option is unemployment. Necessarily in this setup, some gains from training flow to the future employer unless the worker has all the bargaining power.[5] In our setup, when two firms compete for a worker's services, the winning firm matches the losing firm's willingness to pay. If training has increased the old firm willingness to pay by the same as it did for the new firm, then the returns to the training investment are fully priced and there is no future employer externality. This is the case where the production function is modular in firm productivity and human capital. There is in this case an important link between our analysis and that of Moen and Rosén (2004) where training is studied in a directed search setup and efficient training also arises. Our results highlight that efficiency does not require the ability to direct search.[6]

We allow for firm heterogeneity to analyze variation in training across firms and the possible impact of mismatch through frictions on training. We emphasize the result that more productive firms provide more training. This result can come through both the cost of backloading channel as well as the future employer externality. At the steady state, more productive firms provide on average higher utility promise contracts that are less constrained by the cost of backloading. In the risk-neutral case where backloading is costless, training is increasing (decreasing) in the firm type if and only if the production function is supermodular (submodular). In this setting, we make the argument that increased frictions result in greater mismatch which is associated with less human capital accumulation, all else equal. This turns

[5]This is also the case in the Online Appendix of Sanders and Taber (2012), where this issue is also discussed.

[6]Our efficiency analysis is also related to frictionless matching models with endogenous investment and in particular Cole et al. (2001) and Nöldeke and Samuelson (2015). A static and frictionless version of our setup is a special case of Corollary 2 in the latter paper. They show the competitive equilibrium is efficient. They however do not allow for dynamics, long-term contracts nor frictions. Relatedly, Felli and Harris (1996) develop a dynamic model allowing for firm-specific human capital with two firms and one worker, establishing that the equilibrium is efficient. An important difference is that the worker's employer choice is the sole endogenous decision influencing training. In contrast, our model allows training intensity to be entirely endogenous.

out to be the dominant impact from a hypothetical increase in labor market frictions for our calibrated economy.

This paper is related to large literatures on labour market with search frictions, human capital accumulation and long-term contracts. A full survey of these literatures is beyond our scope but we briefly name some key papers. Bagger et al. (2014) estimate a model of search frictions that considers the accumulation of general human capital while individuals are employed. Taber and Vejlin (2020) expand upon this by incorporating non-pecuniary aspects of jobs and heterogeneity in pre-market skills. Lise and Postel-Vinay (2020) allow for multi-dimensional skills that accumulate at varying rates depending on the job an individual holds. Balke and Lamadon (2022) focus on long-term contract and the transmission employer- and worker-specific productivity shocks to earnings and employment. There is no human capital accumulation however. Our paper extends this literature by endogenizing human capital accumulation, enabling us to analyze questions of efficiency, variations in training provisions across heterogenous firms and the impact of frictions.

Engbom (2022) is particularly related to our work as it also incorporates search frictions and human capital accumulation. His analysis explores how labor market frictions affect life-cycle wage growth across different countries. His empirical analysis supports one of our key points, indicating that countries with more fluid labor markets experience reduced life-cycle wage growth. However, there are important differences between our study and Engbom's. First, we consider firm-specific skills which provide novel insights into life-cycle wages growth, job-to-job transitions and training efficiency. Second, heterogeneity in training intensity across firms arises endogenously in our setup while Engbom (2022) assumes that it is a feature of the technology for human capital accumulation.

An outline of the paper follows. Section 2 describes the model and the optimal contract. In Section 3, we derive some analytical results in the special case of risk neutrality and discuss efficiency. Section 4 presents a quantitative analysis of the model properties and shows that frictions are detrimental to training. Section 5 concludes.

2 Model

Time is continuous and both firms and workers discount time at rate ρ. There is a unit measure of workers who can be either employed or unemployed. Matches between workers and firms are formed through a frictional search process. They produce an output that generates a revenue stream. And, the firm can invest in worker's general human capital.

Workers are risk averse and hand-to-mouth. They consume whatever income they have at a given instant which delivers utility, $u(w)$, where w is the wage and they do not have access to savings. Denote by h a worker's general human capital level, which we also refer to as the worker's skill. Let m be the match specific capital level. Match specific capital affects only the productivity of the current match, whereas general human capital applies to all matches. While we refer to the accumulation of match specific capital as training, it accommodates a broader interpretation of investments in the surplus of the current match, including intangibles like goodwill.[7] Skill has two support points: skilled ($h = 1$) and unskilled ($h = 0$). And, for match specific human capital, $m \in \{0, 1\}$. Human capital does not depreciate. It is straightforward to relax these assumptions, but we maintain them throughout the paper for the sake of exposition and to ease the numerical solution of the model.

All workers are born into the labor market unskilled and unemployed. Workers die at rate d. Human capital does not depreciate. An unemployed worker with human capital h receives benefits b_h. Define $r = \rho + d$ as the discount rate including the death risk. Deaths are offset by births that maintain the worker population at a steady normalized measure of one.

All employed workers are laid off at an exogenous layoff rate δ. Unemployed and employed workers meet employment opportunities at rates λ_u and λ_e, respectively. A vacancy is characterized by its productivity index $p \in [0, 1]$. The distribution of productivity over vacancies is given by the CDF $\Phi(\cdot)$ with density ϕ. Search is random and each vacancy is equally represented, so Φ can also be referred to as the productivity offer distribution.

Let $f_{hm}(p)$ be the output of a match between a productivity p firm and a skill h worker with match specific capital m. It is strictly increasing in h, m, and p.

The firm's training decision is modeled as a choice that controls the stochastic process of the worker's human capital evolution. The firm picks the Poisson arrival rate η for which the unskilled worker becomes skilled, and the Poisson rate μ by which a $m = 0$ match transitions to $m = 1$. All firms have access to the same training technology, which is reflected in the monetary training cost $c_h(\eta)$ and specific training cost $c_m(\mu)$. Both functions are increasing and convex.

An employment contract specifies an employment history conditional path of wages and training, $(w, \eta, \mu)(\sigma)$, where σ is the history of the contract. The history of the contract includes the worker's meetings with outside vacancies, including the productivity type. Outside firms can attempt to poach the worker by matching the utility value of the worker's contract. To simplify, it is assumed that an outside firm can perfectly observe the worker's contract. There is limited commitment: At any time, both the worker and the firm can at no cost leave the relationship in favor of their respective outside options. Furthermore,

[7] Many of the paper's conclusions regarding the incentives to accumulate match specific capital apply to a setup where the parties are learning about the productivity of the match.

employment contracts must be renegotiation-proof.

2.1 Recursive formulation of employment contract design problem

For a given match, the firm designs an employment contract to maximize its profits subject to the utility promise it has made to the worker. The firm's instantaneous profit flow is $f_{hm}(p) - w - c_h(\eta) - c_m(\mu)$. Denote by $\Pi_{hm}(V,p)$ the expected net present value of the optimally designed future profit stream to the firm from a match with a skill level (h,m) worker and a current utility promise of V. Following Spear and Srivastava (1987), Thomas and Worrall (1988), and Sannikov (2008), we write the firm's contract design problem taking the worker's utility promise V as the state variable.

For a worker without general nor specific skills $(h,m) = (0,0)$, we have

$$\begin{aligned}(r+\delta)\Pi_{00}(V,p) \quad = \quad & \max_{\left(w,\dot{V},\eta,\mu,H,M,\Omega(\cdot)\right)\in\Gamma_{00}(V,p)} \Bigg\{ f_{00}(p) - w - c_h(\eta) - c_m(\mu) \\ & +\Pi'_{00}(V,p)\dot{V} + \lambda_e \int_0^1 \alpha\left(\Omega\left(p'\right),p'\right)\left[\Pi_{00}\left(\Omega\left(p'\right),p\right) - \Pi_{00}(V,p)\right]d\Phi\left(p'\right) \\ & +\eta\left[\Pi_{10}(H,p) - \Pi_{00}(V,p)\right] + \mu\left[\Pi_{01}(M,p) - \Pi_{00}(V,p)\right]\Bigg\}, \end{aligned} \qquad (1)$$

where $\Pi'_{00}(V,p) = d\Pi_{00}(V,p)/dV$. The contract elements are chosen from the set of feasible contract choices, $\Gamma_{00}(V,p)$, which will be described below. The effective discount rate on the match comes to include the death risk of the worker and the exogenous match destruction rate δ. In the absence of other events, the contract design dictates a change in the worker's utility promise over time of $\dot{V}$, which has a profit flow impact of $\Pi'_{00}(V,p)\dot{V}$. At Poisson rate η the worker receives a positive skill shock, at which point the contract is specified to continue with utility promise H. The net profit value to the firm from this event is $\Pi_{10}(H,p) - \Pi_{00}(V,p)$. The same intuition applies to the term $\mu\left[\Pi_{01}(M,p) - \Pi_{00}(V,p)\right]$ with specific human capital. At rate λ_e the worker meets an outside vacancy with productivity p' distributed according to CDF Φ. Given the vacancy p' meeting, the contract specifies a continuation utility promise of $\Omega(p')$. The outside firm observes $\Omega(p')$. If the outside firm offers the worker a contract with utility value greater than $\Omega(p')$, the worker moves. Otherwise, the worker stays with the current firm. Let the indicator function $\alpha(\Omega(p'),p') = 1$ reflects the worker's decision to stay given the outside type-p' firm's optimal response to $\Omega(p')$. And, $\alpha(\Omega(p'),p') = 0$ if the worker moves. The profit functions are simpler for workers with either general or specific skills and are therefore provided in the Appendix A.

Stated in utility terms, denote a type-p firm's willingness to pay for a skill level (h,m)

worker by $\bar{V}_{hm}(p)$. It is defined as the utility promise such that the firm's net present value of future profits from the match is exactly zero,

$$\Pi_{hm}\left(\bar{V}_{hm}(p), p\right) = 0. \tag{2}$$

Lemma 1 characterizes the resolution of the competition between firms.

Lemma 1. *Consider an (h, m, p) match with current utility promise V and an outside firm with productivity p'. The indicator function α reflecting the worker's decision to stay satisfies*

$$\alpha\left(\Omega(p'), p'\right) = \begin{cases} 1 & \text{if } \Omega(p') \geq \bar{V}_{hm}(p') \\ 0 & \text{if } \Omega(p') < \bar{V}_{hm}(p'). \end{cases}$$

The optimal renegotiation-proof choice of Ω is given by

$$\Omega(p') = \begin{cases} \bar{V}_{hm}(p) & \text{if } \bar{V}_{h0}(p') > \bar{V}_{hm}(p) \\ \bar{V}_{h0}(p') & \text{if } V \leq \bar{V}_{h0}(p') \leq \bar{V}_{hm}(p) \\ V & \text{if } \bar{V}_{h0}(p') < V. \end{cases}$$

If the worker's contract offers less than the outside firm's willingness to pay, the worker leaves in favor of a matching offer from the outside firm. If the outside firm has a greater willingness to pay than the current firm $\bar{V}_{h0}(p') \geq \bar{V}_{hm}(p)$, the contract's optimal renegotiation-proof response is to promise continuation utility equal to the current firm's willingness to pay $\bar{V}_{hm}(p)$. By giving the highest possible renegotiation-proof utility promise, the current firm does the best it can to allow the worker to extract rents from the future match. This increases the worker's current valuation of the contract, which the firm can translate into profits through lower current wages. If the firm could bluff, it would want to push the continuation utility promise all the way to the outside firm's willingness to pay. However, this is ruled out by requiring the contract to be renegotiation-proof.

When the outside firm's willingness to pay is greater than the worker's current contract utility promise, but less than the current employer's willingness to pay $\bar{V}_{hm}(p) > \bar{V}_{h0}(p') > V$, the optimal response is to offer exactly the outside firm's willingness to pay $\bar{V}_{h0}(p')$. If any less is offered, the worker would leave which is Pareto dominated. If any more is offered, the firm is giving up profits unnecessarily. If the outside firm's willingness to pay is less than $\bar{V}_{h0}(p') < V$, it is optimal to keep the contract at V. The latter follows from the worker's risk aversion.

Using the discussion above together with integration by parts, Equation (1) simplifies to

$$
\begin{aligned}
(r+\delta)\,\Pi_{00}(V,p) = \max_{\left(w,\dot{V},\eta,\mu,H,M\right)\in\Gamma_{00}(V,p)} & \Bigg\{ f_{00}(p) - w - c_h(\eta) - c_m(\mu) + \Pi'_{00}(V,p)\,\dot{V} \\
& + \lambda_e \int_V^{\bar{V}_{00}(p)} \Pi'_{00}(V',p)\,\hat{F}_{00}(V')\,dV' \\
& + \eta\left[\Pi_{10}(H,p) - \Pi_{00}(V,p)\right] + \mu\left[\Pi_{01}(M,p) - \Pi_{00}(V,p)\right] \Bigg\},
\end{aligned} \tag{3}
$$

where $F_{hm}(V) = \Phi\left(\bar{V}_{hm}^{-1}(V)\right)$ is the offer distribution in terms of willingness to pay and $\hat{F}_{hm}(V) = 1 - F_{hm}(V)$.

If $h = 0$ and $m = 0$, the utility promise constraint on the contract can be written as

$$
\begin{aligned}
(r+\delta+\eta+\mu)\,V &= u(w) + \eta H + \mu M + \delta U_0 + \dot{V} + \\
&\quad + \lambda_e \left[\int_V^{\bar{V}_{00}(p)} (V' - V)\,dF_{00}(V') + \hat{F}_{00}\left(\bar{V}_{00}(p)\right)\left(\bar{V}_{00}(p) - V\right) \right] \\
&= u(w) + \eta H + \mu M + \delta U_0 + \dot{V} + \lambda_e \int_V^{\bar{V}_{00}(p)} \hat{F}_{00}(V')\,dV',
\end{aligned} \tag{4}
$$

where the second equality follows from integration by parts. The worker receives utility flow $u(w)$. At Poisson rate δ, the match is destroyed, which has flow contribution $\delta(U_0 - V)$. At Poisson rate η, the worker receives a skill shock, which has flow contribution $\eta(H - V)$. Similarly, at rate μ match specific capital increases and the contract continues with utility promise M. The contract is designed to have a time change in the utility value of $\dot{V}$. Furthermore, at rate λ_e the worker meets an outside vacancy with willingness to pay V' distributed according to $F_{00}(V)$. If $V' \in \left[V, \bar{V}_{00}(p)\right]$, the worker stays with the current firm at continuation utility V'. If $V' > \bar{V}_{00}(p)$, the worker leaves to work at the outside firm at a utility promise of $\bar{V}_{00}(p)$. If either $h = 1$ or $m = 1$, the utility expressions are simpler and presented in Appendix A.

Finally, the contract must respect the participation constraints. The contract cannot give the worker utility value less than unemployment. Likewise, the contract cannot give the firm a negative profit value. The set of feasible contract choices for $h = 0$ and $m = 0$ is given by

$$\Gamma_{00}(V,p)=\Bigg\{\Big(w,\dot{V},\eta,\mu,H,M\Big)\ \Big|$$
$$u(w)+\eta H+\mu M+\delta U_h+\dot{V}+\lambda_e\int_V^{\bar{V}_{00}(p)}\hat{F}_{00}(V')\,dV'=(r+\delta+\eta+\mu)V$$
$$U_0\le M\le\bar{V}_{01}(p)$$
$$U_0\le H\le\bar{V}_{10}(p)\Bigg\}. \tag{5}$$

The expressions for the sets Γ_{10},Γ_{01} and $,\Gamma_{11}$ are presented in Appendix A.

Finally, the value of unemployment U_h is given by,

$$rU_h=u(b_h).$$

The worker meets firms at rate λ_u. However, regardless of the type of the meeting, the associated employment contract delivers utility promise U_h. Indeed, when a firm meets an unemployed worker, it makes her a take-it-or-leave-it lifetime utility offer, which she accept as long as it exceeds the value of unemployment.

2.2 Optimal contract design

To keep notation simple, the policy functions $H_m(V,p)$ and $\eta_m(V,p)$ refer to the optimal general training-related policies in the contract for an $h=0$ type worker with match specific capital m who is employed with a type-p firm and a utility promise of V. Analgously, $M_h(V,p)$ and $\mu_h(V,p)$ refer to the specific training policies.Where otherwise not obvious, the worker's skill state will be explicitly referenced in the notation. In the following, we discuss the properties of the optimal employment contract with a given productivity firm.

We assume $\Pi_{hm}(V,p)$ is strictly concave in V. Although the strict concavity of the utility function might suggest this property is intuitive, proving it is challenging and we do not relitigate the point in this paper.[8] It is a verified characteristic in all our numerical solutions.

The following lemma characterizes the optimal contract's utility promise and wage dynamics over the duration of the relationship. We characterize the training choices in section

[8]Lentz (2014) provides proof in a similar setting without endogenous human capital accumulation. Variations over this theme come up in settings like Phelan and Townsend (1991); Hopenhayn and Nicolini (1997); Lentz and Tranæs (2005); Lentz (2009); Lise (2013)

2.2.1.

Lemma 2. *The optimal contract is for any $p \in [0,1]$ and $V \in \left[U_h, \bar{V}_{hm}(p)\right]$ characterized by*

1. *The optimal employment contract is flat:*

$$\dot{V}_{hm}(V,p) = 0. \tag{6}$$

2. *Wages are increasing in the utility promise, $\partial w_{hm}(V,p)/\partial V > 0$.*

3. *For both general and specific human capital increases, the utility promise gains are less than full rent extraction by the worker:*

$$V < H_m(V,p) < \bar{V}_{1m}(p) \text{ and } V \leq M_h(V,p) < \bar{V}_{h1}(p) \text{ with strict inequality if } p < 1.$$

4. *Wages are smooth across human capital increases unless the participation constraint is binding: $w_{h0}(V,p) = w_{h1}(M_h(V,p),p)$. If $U_{1m} < H_m(V,p)$, $w_{0m}(V,p) = w_{1m}(H_m(V,p),p)$.* Otherwise, $w_{0m}(V,p) < w_{1m}(H_m(V,p),p)$.

There are no incentives to front- or backload because there is no moral hazard or particular joint inefficiencies in the match such as in Burdett and Coles (2003) and Lentz (2014). Thus, for any level of risk aversion, in the absence of arrivals of outside offers or changes in human capital, the contract's wage profile is flat in tenure. In the limit case where the worker is risk neutral, the flat contract remains optimal, but there is now a multitude of optimal paths. The analysis uses the flat contract in the limit case where the worker is risk neutral.

Unconditionally, the employment contract involves expected changes in the utility promise over job duration through two channels. The contract matches the willingness to pay of firms with which the worker meets, which in isolation implies an expected increasing utility promise path in duration. For a given utility promise V, the expected growth rate in the utility promise within the job due to on-the-job search is $\lambda_e \int_V^{\bar{V}_{hm}(p)} (V' - V)\, dF_{h0}(V') \geq 0$. Since wages are increasing in the utility promise, on-the-job search in isolation implies an increasing wage path in tenure. This is a simple replication of the offer-matching process in Postel-Vinay and Robin (2002).

The second channel ties value promise increases to increases in human capital. Lemma 2 states that the worker receives utility promise gains when human capital increases, whether it is specific or general. It is worthwhile to contrast this result with the traditional argument in a frictionless setting. Here, competition between firms deliver all rents from general human capital increases to the worker, and does not impose itself whatsoever on how rents from specific human capital gains are shared. In a setting where smooth wage paths are preferred, match specific human capital gains would be enjoyed by the firm, only. Thus, in the non-frictional setting, there is stark difference between how rents associated with human capital gains are allocated depending on the specificity of human capital. A major implication of frictions in our setting is to muddle this difference: Whether human capital gains are specific or general, they result in a utility promise increase to the worker, but this increase falls short of full rent extraction to the worker.

Why? The argument works through the following main mechanism: An increase in either general or specific capital is for a given utility promise associated with an increase in the worker's expected net utility gains from outside meetings. If the firm wanted to keep the utility promise constant across an increase in human capital, it would have to lower the worker's current wage at the point of the increase. This is suboptimal given the concave utility function. More broadly, Lemma 2 proves that the concave utility function induces a preference for a constant wage path such that wages are kept constant across human capital increases, all else equal. But since the expected gains to the worker from future employer meetings are now greater, the utility promise must jump when human capital increases. The only caveat to this is that the contract must satisfy the worker's participation constraint which may force the firm to increase wages associated with a general human capital increase.

In Figure 1 we illustrate the increased value from future employer meetings associated with increases in human capital. The figures are stylized in that one would not generally expect $\bar{V}_{hm}(p)$ to be linear. The exact source and magnitude of the increased value of future meetings depend on whether the capital gain is general or specific. Consider an $(h,m)=(0,0)$ worker with a current utility promise of V who is employed with a firm that has willingness to pay $\bar{V}_{00}(p)$. The worker's expected utility growth rate from on-the-job search is,

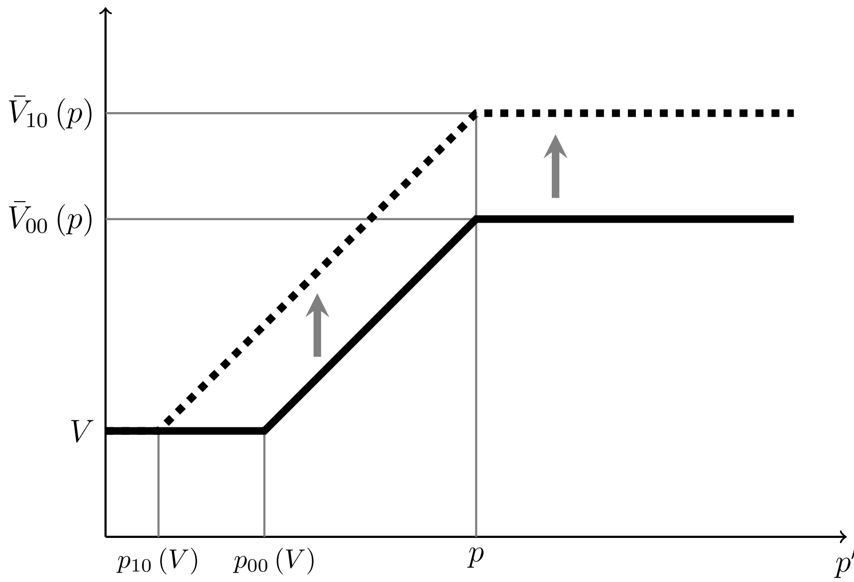
$$\lambda_e (1-\Phi(p)) \left[\bar{V}_{00}(p) - V\right] + \lambda_e \int_{p_{00}(V)}^{p} \left(\bar{V}_{00}(p') - V\right) d\Phi(p'),$$

where the first term reflects meetings that result in job-to-job moves to a new employer and the second reflects increased worker rent extraction due to matching of offers from inferior outside employers. The solid line in subpanels (a) and (b) shows the exact value to the worker from a meeting with an outside p' firm. We define $p_{hm}(V)$ as the productivity type who has willingness to pay equal to V, $\bar{V}_{hm}(p_{hm}(V)) = V$.

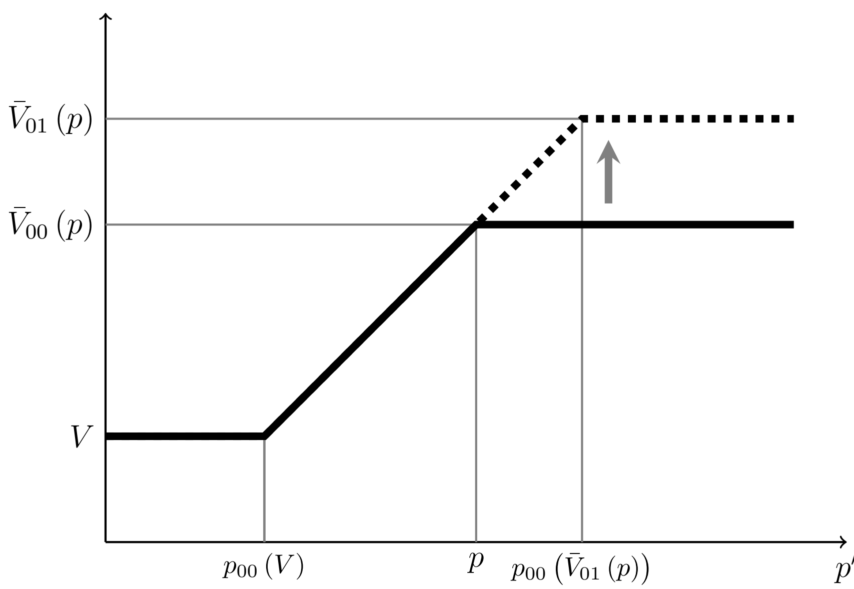
Suppose the worker becomes generally-skilled. Holding the current utility promise fixed at V, the increased competitive pressure on the match is reflected in the now greater expected

Figure 1: Expected Gains from On-the-Job Search

(a) A general human capital increase



(b) A specific human capital increase



utility growth rate from on-the-job search,

$$\lambda_e (1 - \Phi(p)) \left[\bar{V}_{10}(p) - V\right] + \lambda_e \int_{p_{10}(V)}^{p} \left(\bar{V}_{10}(p') - V\right) d\Phi(p').$$

This is illustrated with the dashed line in subpanel (a). All outside firms are now willing to pay more for the worker, $\bar{V}_{10}(p') > \bar{V}_{00}(p')$. In addition, the support of firm types that can impose competitive pressure on the match expands downward from $p_{00}(V)$ to the lower firm type $p_{10}(V)$. Should the worker move, she will move with a higher utility promise, $\bar{V}_{10}(p)$.

Consider alternatively an increase in specific capital from $m = 0$ to $m = 1$. Holding the utility promise fixed at V, the competitive pressure on the match increases to,

$$\lambda_e \left(1 - \Phi\left(p_{00}\left(\bar{V}_{01}(p)\right)\right)\right) \left[\bar{V}_{01}(p) - V\right] + \lambda_e \int_{p_{00}(V)}^{p_{00}\left(\bar{V}_{01}(p)\right)} \left(\bar{V}_{00}(p') - V\right) d\Phi(p').$$

This is illustrated with the dashed line subpanel (b). In this case, the incumbent firm's willingness to pay for the worker increases to $\bar{V}_{01}(p) > V_{00}(p)$. Outside firms do not change their willingness to pay for the worker, but the upper bound on the set of firms that will force up the utility promise within the job increases to $p_{00}\left(\bar{V}_{01}(p)\right) > p$. Furthermore, when the worker moves to a better firm, she moves with a greater utility promise of $\bar{V}_{01}(p)$. This is an effect from the frictional framework: The increased specific capital has increased rent extraction from future employers.

It is worth noting that in contrast to piece-rate contracts such as in Bagger et al. (2014), the wage does not immediately respond to a skill increase. The worker's utility value jumps in anticipation of future wage gains. However, as shown in Lentz (2014), in a setting with hidden search, the growth rate in wages would immediately jump upon a human capital increase and while the wage is constant across the skill jump, it immediately begins to grow faster, even in the absence of outside meetings. This begins to soften the contrast between the piece-rate contract and our paper.

2.2.1 Training rates are increasing in the utility promise. The cost of backloading.

In this section we emphasize that competition between firms generally associate training with backloading of wages. The greater the backloading cost, the greater the discouragement of investment. This is in close kinship with the argument in Acemoglu and Pischke (1999) that when backloading is costly (in their constrained regime case, impossible), increased competition discourages investment because it increases backloading pressure. Indeed, our model contains and validates their argument as a partial equilibrium effect (we return in greater detail in section 4.3). In this section we characterize the optimal contract's training choices and make the point that backloading costs are greater the lower the utility promise and therefore training is increasing in the utility promise.

The first-order condition for the training rate is given by

$$[\Pi_{1m}(H_m(V,p),p) - \Pi_{0m}(V,p)] - \Pi'_{0m}(V,p)[H_m(V,p) - V] = c'_h(\eta_m(V,p)) \tag{7}$$

$$[\Pi_{h1}(M_h(V,p),p) - \Pi_{h0}(V,p)] - \Pi'_{h0}(V,p)[M_h(V,p) - V] = c'_m(\mu_h(V,p)), \tag{8}$$

where the skill change conditional utility promise satisfy Equation (18) and (19).

The first-order condition on training state that the marginal cost of training must equal the marginal profit gain from the increase in either general or specific skills. The first bracketed term on the right-hand side of the first-order conditions (7) and (8) is the direct jump in profits due to the skill increase. The second term reflects the profit value of the change in the worker's utility promise, where by equation (15), $\Pi'_{hm}(V,p) = -1/u'(w_{hm}(V,p))$ is the profit impact of a one-unit increase in the utility promise.

Increases in the skill change conditional utility promise, $H_m(V,p)$ and $M_h(V,p)$, reduce the direct profit gains from training, but the loss is compensated by the worker's greater expected utility gains, which are translated into current profits through reduced wages today. In a risk-neutral setting, these two effects exactly offset each other and the training decisions are unaffected by the particular choice of H_m and M_h. Thus, Becker's (1964) insight that even though a perfectly competitive environment dictates that the firm has to deliver all of the match surplus to the skilled worker, $H_m(V,p) = \bar{V}_{1m}(p)$, the training choice remains privately efficient since the firm is perfectly compensated via lower wages during the training period. In the incomplete contracts setting combined with the narrative that the firm provides and pays the cost of training, this is sometimes referred to as the worker "holding-up" the firm when the market forces, $H_m(V,p) = \bar{V}_{1m}(p)$ and the firm cannot extract ex post rents from its investment in the worker. Becker (1964) can then be understood as an insight that it is not the hold-up problem per se that discourages investment. It is the cost of responding to the hold-up by, in this case, backloading wages.

In our analysis, the cost of backloading is tied to risk aversion. When the worker is risk averse, future utility promises can no longer be translated into profits one to one through a lowering of current wages. Backloading is expensive. Therefore, training comes to depend on both the current utility promise V and the contract's optimal choice of $H_m(V,p)$ and $M_h(V,p)$. It follows by differentiation of the first order conditions (7) and (8) as well as concavity of the profit function that,

$$\frac{\partial \mu_h(V,p)}{\partial V} = \frac{-\Pi''_{h0}(V,p)\left[M_h(V,p)-V\right]}{c''_\mu\left(\mu_h(V,p)\right)} \geq 0$$
$$\frac{\partial \eta_m(V,p)}{\partial V} = \frac{-\Pi''_{0m}(V,p)\left[H_m(V,p)-V\right]}{c''_\eta\left(\eta_m(V,p)\right)} \geq 0. \tag{9}$$

By Lemma 2, both specific and general training are increasing in the utility promise given concavity of the profit function. We discuss variation in training across firm types in detail in Section 4.1.

Increased human capital (both general and specific) implies increased competitive pressure on the match and consequently greater rents to the worker. The firm will want to reduce current wages to capture the ex post rents flowing to the worker. However, risk aversion imposes a cost on this mechanism. The concavity of the profit function is a reflection that the this cost is more severe for low utility promises because there is already greater backloading built into the contract through the lower utility promise.

2.2.2 Job-to-job mobility, tenure and wages

Both Altonji and Shakotko (1987) and Topel (1991) emphasize that tenure effects in wages may be associated with a selection effect on the type of future firms which complicates the distinction between experience and tenure effects in their analyses. Our analysis exhibits exactly this effect in the case of specific human capital accumulation. As specific capital increases, the firm type threshold such that the worker is indifferent between moving to it and staying with the current firm goes up. Thus, conditional on moving, the expected firm type of the new firm increases as specific capital goes up, and consequently increased specific capital will have a positive wage impact beyond the current match through this selection effect.

In addition, our analysis contains another important channel through which specific training will result in higher wages with future firms: Even though specific capital is not portable between firms, bargaining position carries over. Specific training increases the willingness to pay of the worker's current employer, which means that conditional on moving, the worker will do so with a greater utility promise with the new firm. Hence, even conditional on the type of the future employer, specific training in the current firm raises wages with future employers - this despite the fact that the willingness to pay of the future employer is unchanged.

Thus, specific training raises wages with future employers and within the current match. General training does as well.

2.3 Steady state

Denote by e_{hm} the mass of employment of general skill h workers in jobs with match specific capital, m. Let u_h be the mass of unemployed general skill h workers. Normalize the population at unity, $1 = \sum_h \left(u_h + \sum_m e_{hm}\right)$. Furthermore, denote by $G_{hm}(V,p)$ the cumulative distribution of match states for type (h,m) matches, where by definition $G_{hm}\left(\bar{V}_{hm}(1),1\right) = 1$. The steady state conditions on the employment and unemployment stocks follow the simple logic that the flow into the stock must equal the flow out.

The steady state condition on $e_{00}G_{00}(V,p)$ is given by,

$$
\begin{aligned}
&\lambda_u u_0 \Phi(p) + \lambda e_{01} \int_0^{\bar{p}_{01}(V)} \int_{U_0}^{\bar{V}_{01}(p')} \left[F_0\left(\bar{V}_{00}(p)\right) - F_0\left(\bar{V}_{01}(p')\right)\right] g_{01}(V',p')\, dV'dp' = \\
&e_{00}\left\{ \int_0^{\bar{p}_{00}(V)} \int_U^{\bar{V}_{00}(p')} \left[d + \delta + \eta_0(V',p') + \mu_0(V',p') + \lambda \hat{F}_0\left(\bar{V}_{00}(p)\right)\right] g_{00}(V',p')\, dV'dp' + \right. \\
&\qquad \left. \int_{\bar{p}_{00}(V)}^{p} \int_U^{V} \left[d + \delta + \eta_0(V',p') + \mu_0(V',p') + \lambda \hat{F}_0(V)\right] g_{00}(V',p')\, dV'dp' \right\}.
\end{aligned}
$$

The first term on the left hand side is the flow into the $e_{00}G_{00}(V,p)$ pool from unemployment. The second term is the flow in from the pool of matches with high match specific capital where the worker nevertheless receives a better offer and consequently moves into low match specific capital. The integral is over types of matches with high match specific capital. The outer integral is over firms that have willingness to pay less than V. Any firm with a willingness to pay more than V may be beat, but the worker would move into the e_{00} pool with a utility promise greater than V. The inner integral is then all the possible utility promises that workers may have in these firms. The term $\left[F_0\left(\bar{V}_{00}(p)\right) - F_0\left(\bar{V}_{01}(p')\right)\right]$ is the probability that a worker in a type p' firm will receive an offer that is better than her current firm's willingness to pay, but is from a type firm less than p. If that happens, the worker moves into the $e_{00}G_{00}(V,p)$ pool. The terms on the right hand side are standard: The worker leaves the pool upon death, unemployment, general and specific skill acquisition, and if the worker receives an outside offer that takes her out of the pool. The latter can happen in two ways: If a worker is currently employed with a firm that has willingness to pay less than V then an outside offer must be from a firm better than p to make her leave the pool. If she is with a firm with willingness to pay greater than V, then it is sufficient that the outside offer be better than V.

The steady state conditions on $e_{01}G_{01}(V,p)$, $e_{10}G_{10}(V,p)$, and $e_{11}G_{11}(V,p)$ follow the same type of argument and are given in Appendix D.

3 The risk-neutral case

In the special case of risk neutrality, the cost of backloading is eliminated from the analysis. Training inefficiencies are in this case due only to the future employer externality. We obtain two new analytical results in this case. First, we find that variation in training across firms are only driven by technological properties, and in particular complementarities in the production functions. Second, we find that specific training tends to be inefficiently high in low-type firms.

3.1 Training decisions and future employer externalities

Assume $u'' = 0$ and without loss of generality transform the utility function so that $u'(w) = 1$. The profit function takes the form $\Pi_h(V, p) = \bar{V}_h(p) - V$. By equations (7) and (8), the first-order condition for the optimal contract's training rate reduces to,

$$
\begin{aligned}
c_h'(\eta_m(V, p)) &= \bar{V}_{1m}(p) - \bar{V}_{0m}(p) \\
c_m'(\mu_h(V, p)) &= \bar{V}_{h1}(p) - \bar{V}_{h0}(p).
\end{aligned}
$$

It is immediately seen that the training rates do not depend on the particular utility promise in the contract. The risk-neutral case eliminates the cost of backloading from the analysis and, in particular, the variation of the severity of the problem as a function of the utility promise.

The firm's willingness to pay solves

$$
\begin{aligned}
(r + \delta)\bar{V}_{hm}(p) = f_{hm}(p) + \delta U_h + (1 - h)\left[\eta_m(p)\left[\bar{V}_{1m}(p) - \bar{V}_{hm}(p)\right] - c_h(\eta_m(p))\right] \\
+ (1 - m)\left[\mu_h(p)\left[\bar{V}_{h1}(p) - \bar{V}_{hm}(p)\right] - c_m(\mu_h(p))\right], \qquad (10)
\end{aligned}
$$

where the dependency of the training rates on V has been eliminated. By differentiation it follows that,

$$
\begin{aligned}
\eta_1'(p) &= \frac{f_{11}'(p) - f_{01}'(p)}{[r + \delta + \eta_1(p)]\, c_h''(\eta_1(p))} \\
\mu_1'(p) &= \frac{f_{11}'(p) - f_{10}'(p)}{[r + \delta + \mu_1(p)]\, c_m''(\mu_1(p))}.
\end{aligned}
$$

The expressions for $\eta_0'(p)$ and $\mu_0'(p)$ account for possible complementarities between general and specific training both direct and through firm productivity,

$$c_h''(\eta_0(p))\,\eta_0'(p) = \frac{f_{10}'(p) - f_{00}'(p) + \mu_1(p)\,c_m''(\mu_1(p))\,\mu_1'(p) - \mu_0(p)\,c_m''(\mu_0(p))\,\mu_0'(p)}{r + \delta + \eta_0(p)}$$
$$c_m''(\mu_0(p))\,\mu_0'(p) = \frac{f_{01}'(p) - f_{00}'(p) + \eta_1(p)\,c_h''(\eta_1(p))\,\eta_1'(p) - \eta_0(p)\,c_h''(\eta_0(p))\,\eta_0'(p)}{r + \delta + \mu_0(p)}.$$

In the case of a modular production function one immediately obtains that, $\eta_m'(p) = \mu_h'(p) = 0$. Thus, in the risk neutral case, if the production function does not have complementarities between firm productivity and training, then training is constant across firm types. Furthermore, training is increasing (decreasing) in firm type if and only if the production function is supermodular (submodular).

Competitive pressure varies across firms, but whatever the share of ex post gains to training it delivers to the worker, the firm can translate it into profits through lower wages at the time of training without any efficiency loss. Specifically, notice that the meeting rates λ_u and λ_e do not affect $\bar{V}_{hm}(p)$ and therefore do not impact the training levels.

These results highlight the importance of the wage determination process as to whether the firm's position in the firm hierarchy directly affects training through the implied job-to-job transition rate. In our setting it does not. However, if the worker's gains associated with a move to another firm fall short of the old firm's losses, one would expect that general training be decreasing in the degree of competitive pressure on the match, since it now raises the effective discount rate on the returns to human capital investments. Furthermore, specific investment is stimulated since it is a way to reduce the match surplus destruction associated with job-to-job transitions. In the Supplemental Appendix, Sanders and Taber (2012) discuss such a case in an environment where wages are statically bargained based on an outside worker option of unemployment and the current firm cannot provide side payments to avoid the destruction of match surplus. Fu (2011) presents an analysis with a super-modular production function and a piecewise wage posting environment where matches are not necessarily fully compensated for their destruction when workers reallocate.

We turn to analyzing the efficiency of the decentralized equilibrium. For clarity, we analyze general human capital and specific human capital separately.

3.2 Efficiency

In this section we discuss social efficiency with risk neutral workers. We do so in a simplified setting that focuses on the human capital accumulation decisions in isolation from the standard externalities in the random search model. Thus, consider a planner version of the problem where the population of firms is fixed and the hiring intensity of any given firm is normalized at unity. The constant returns to scale firm production technology has the implication that a firm's agreement to match with a given worker today does not impact the value of a future worker meeting.

Assume a standard constant returns to scale matching function $m = m(s, v)$ where s and v are the aggregate measures of search and vacancies, respectively. A unit of search meets a vacancy at rate $\lambda = m/s$. For the sake of simplicity assume unemployed and employed workers search with equal intensity. Assume unmatched productivity is $b_h = f_{h0}(0)$, with the implication that all jobs have a higher social value than unemployment.

With the fixed population sizes and given job destruction technology, δ, the measures s and v are outside of the planner's control. By implication, and also outside of the planner's control, any given worker meets a production technology p at rate $\lambda\phi(p)$. Consider a social planner problem of maximizing the net present value of the future stream of output net of training costs for a given initial worker population characterized by match distribution G_0,

$$\mathcal{V}(G_0) = \max_{\left\{\eta_{hm,t}(p),\mu_{hm,t}(p),a_{hm,t}(p,p')\right\}} \int_0^\infty e^{-\rho t} \sum_{h,m} \int_0^1 \left(f_{hm}(p) - c_h(\eta_{hm}(p)) - c_m(\mu_{hm}(p))\right) dG_{hm,t}(p)dt, \tag{11}$$

where the optimal choices are done for the $t \geq 0$ time path and for $(p, p') > 0$ and $(h, m) \in \{0, 1\}^2$. The acceptance decision $a_{hm,t}(p, p') \in \{0, 1\}$ states that a worker in state (h, m, p) accepts an offer from a type p' firm. The match distribution is for the sake of brevity defined so that $1 = \sum_{h,m} \int_0^1 G_{hm,t}(p)dp$. Specifically, $G_{h,t}(0)$ includes unemployment and is a mass point. Furthermore, the notation adopts the convention $G_{h,t}(p) = G_{h0,t}(p) + G_{h1,t}(p)$. The law of motion for the population is,

$$\begin{aligned}
\dot{G}_{0,t}(0) &= d + \delta G_{0,t}(1) - G_{0,t}(0)\left[\delta + d + \lambda \int_0^1 a_0(0,p)d\Phi(p)\right] \\
\dot{G}_{1,t}(0) &= \delta G_{1,t}(1) - G_{1,t}(0)\left[\delta + d + \lambda \int_0^1 a_1(0,p)d\Phi(p)\right] \\
\dot{G}_{0,t}(p) &= d + \lambda\left[\int_p^1 \int_0^p a_0(p'',p')d\Phi(p')dG_{0,t}(p'') - \int_0^p \int_p^1 a_0(p'',p')d\Phi(p')dG_{0,t}(p'')\right] \\
&\quad - dG_{0,t}(p) - \int_0^p \eta_0(p')dG_{0,t}(p'),\ \forall p > 0 \\
\dot{G}_{1,t}(p) &= \lambda\left[\int_p^1 \int_0^p a_1(p'',p')d\Phi(p')dG_{1,t}(p'') - \int_0^p \int_p^1 a_1(p'',p')d\Phi(p')dG_{1,t}(p'')\right] \\
&\quad - dG_{1,t}(p) + \int_0^p \eta_0(p')dG_{0,t}(p'),\ \forall p > 0.
\end{aligned}$$

The planner faces time independent technologies: $\lambda\Phi(\cdot)$, δ, d, $f_h(\cdot)$ and $c(\cdot)$. Given that none of these technologies depend on the aggregate state of the economy either, the planner problem can be restated as a collection of independent problems,

$$\mathcal{V}(G_0) = \sum_{h,m} \int_0^1 \mathcal{V}_{hm}(p)\, dG_{hm,0}(p),$$

where $\mathcal{V}_{hm}(p)$ states the maximized net present value of future production net of training cost for a state (h, m, p) worker who experiences the death shock as a rebirth into unskilled unemployment. $\mathcal{V}_{hm}(p)$ also coincides with the co-state variable associated with the law of motion for $g_{hm}(p)$ in the Hamiltonian for the planner problem in (11). In the following efficiency analysis we discuss specific and general human capital investments in isolation.

3.2.1 General human capital

In this section, ignore the specific human capital investment. Consider the planner's problem associated with an unskilled worker who is currently matched to a production technology p. Denote the net present value of production net of training costs associated with such a worker by,

$$\begin{aligned}
(r+\delta)\,\mathcal{V}_0(p) =& \max_{\eta}\Big[f_0(p) - c_h(\eta) + (\delta+d)\,\mathcal{U}_0 + \eta\,(\mathcal{V}_1(p) - \mathcal{V}_0(p)) + \\
&\lambda \int_p^1 [\mathcal{V}_0(p') - \mathcal{V}_0(p)]\, d\Phi(p')\Big], \\
=& f_0(p) + \delta\mathcal{U}_0 + \mathcal{M}(p) + \lambda \int_p^1 \frac{[f_0'(p') + \mathcal{M}'(p')]\,\hat{\Phi}(p')}{r+\delta+\lambda\hat{\Phi}(p')} dp',
\end{aligned}$$

where the value contribution of a skilled worker is

$$\begin{aligned}
(r+\delta)\,\mathcal{V}_1(p) =& f_1(p) + \delta\mathcal{U}_1 + d\mathcal{U}_0 + \lambda \int_p^1 [\mathcal{V}_1(p') - \mathcal{V}_1(p)]\, d\Phi(p) \\
=& f_1(p) + \delta\mathcal{U}_1 + d\mathcal{U}_0 + \lambda \int_p^1 \frac{f_1'(p')\,\hat{\Phi}(p')}{r+\delta+\lambda\hat{\Phi}(p')} dp',
\end{aligned}$$

and the value of the investment option is

$$\mathcal{M}(p) = \max_{\eta}\left[-c_h(\eta) + \eta\,(\mathcal{V}_1(p) - \mathcal{V}_0(p))\right].$$

The socially optimal investment choice solves

$$c_h' \left(\eta^{sp} (p)\right) = \mathcal{V}_1 (p) - \mathcal{V}_0 (p) .$$

Some algebra yields

$$\begin{aligned}(r+\delta) \left[\mathcal{V}_1 (p) - \mathcal{V}_0 (p)\right] = f_1 (p) - f_0 (p) + \frac{\delta}{r} \left[f_1 (0) - f_0 (0)\right] - \mathcal{M} (p) \\ + \lambda \int_p^1 \frac{\left[f_1' (p') - f_0' (p')\right] \hat{\Phi} (p')}{r+\delta+\eta^{sp} (p') + \lambda \hat{\Phi} (p')} dp' \\ + \frac{\delta}{r} \lambda \int_0^1 \frac{\left[f_1' (p') - f_0' (p')\right] \hat{\Phi} (p')}{r+\delta+\eta^{sp} (p') + \lambda \hat{\Phi} (p')} dp'. \end{aligned} \tag{12}$$

With this, the analysis can immediately establish a significant efficiency result in Lemma 3.

Lemma 3. *In the risk-neutral case, general training is efficient if the production function is modular. If the production function is supermodular (submodular), training is too low (high).*

Proof. It follows from Equation 10 that

$$(r+\delta) \left[\bar{V}_1 (p) - \bar{V}_0 (p)\right] = f_1 (p) - f_0 (p) + \frac{\delta}{r} \left[f_1 (0) - f_0 (0)\right] - M (p) , \tag{13}$$

where the value of the training option is given by

$$M (p) = \max_{\eta} \left[-c_h (\eta) + \eta \left[\bar{V}_1 (p) - \bar{V}_0 (p)\right]\right] .$$

The proof then follows from a simple comparison of equations (13) and (12). When the production function is modular, $f_1' (p) - f_0' (p) = 0$ for all $p \in [0, 1]$. Hence, the integral terms in equation (12) fall away. In this case, it is immediate that $V_1 (p) - V_0 (p) = \mathcal{V}_1 (p) - \mathcal{V}_0 (p)$, and hence the decentralized training choice coincides with that of the social planner. When the production function is supermodular, the integral terms are positive. Thus, the social planner has greater returns to training than the decentralized contract. Therefore, $\eta (p) < \eta^{sp} (p)$. When the production function is submodular, the integral terms are negative, and the opposite holds; there is too much investment. □

We discuss the intuition for the efficiency results in terms of the future employer externality in the following section.

Future employer externality To poach a worker a firm must promise the worker a utility value equal to the old firm's willingness to pay. Hence, if the worker is swayed to move, the old match is fully compensated for its destruction. Given this wage determination, for there to be a future employer externality, it has to be that skill is more or less valuable with a future employer than in the current match. When the production function is modular, increased skill adds the same to all matches regardless of firm productivity. Lemma 3 shows that training levels are efficient in this case. If the production function is supermodular, future employers value increased skill by more, and there is underinvestment in training due to a positive future employer externality. Furthermore, the underinvestment problem is stronger for low-productivity employers, and so we find in this case that more productive employers provide more training. There is a subtle distinction of these results relative to Moen and Rosén (2004): In their competitive search setting, efficiency obtains also in the supermodular and submodular settings.

The efficiency result uses the assumption that unemployment benefits are $b_h = f_h(0)$. In the modular production function case, this ensures that the gains from increased skill carry through unemployment so as to match the social planner valuations. The assumption ensures that the time spent in unemployment is associated with an income flow that increases with skill as it does in production. While the time spent in unemployment must reflect the increased skill, the more important issue in our view is the future employer externality coming out of unemployment. The assumption implies that an employer that hires a worker out of unemployment must promise utility as if it were matching the willingness to pay of the lowest firm type. This means that the worker gets to carry the return from skill out of unemployment. If this is not the case, the effective discount rate on human capital investments in the decentralized case will include the layoff rate into unemployment, which differs from that of the planner.

It is not a special feature that a worker who is hired from unemployment would have the willingness to pay of another firm as an outside option. This is, for example, the case in Burdett and Judd (1983) where workers compare multiple offers within a period. The same idea applies to environments where job offers can be held for some period of time. While not widely representative, this is for example the case for the junior market for academic economists. It is also a feature of the models on recalls and search capital.[9] The assumption that $b_h = f_h(0)$ is a simple way of ensuring that gains from skill carry through the unemployed state as would be implied in the above-mentioned models.

Interestingly, even in the sub- or super-modular production function case, notice that the meeting rate λ does not affect $\bar{V}_{hm}(p)$ and therefore do not impact the decentralized training level. In other words, even in the presence of a future employer externality, the poaching risk does not affect training levels. How can this be? The degree of friction in the market does not

[9]See Fujita and Moscarini (2013) and Carrillo-Tudela and Smith (2014).

enter the training decision in this model because a match is always perfectly compensated for its destruction when the worker moves to another firm. Hence, if training increases the value of a match, a poaching firm will deliver the value increase to the match upon the worker's departure to the new firm. Thus, a change in the strength of frictions will not impact the training levels in the risk-neutral case. In the case of a supermodular production function, this is a source of disagreement between the planner and the decentralized economy: Human capital is in this case more valuable with a more productive future employer. Not only does the decentralized solution imply too little investment at any $p < 1$, the shortfall is increasing in the meeting rate as the probability that the worker will be with a more productive employer in the future increases. By the same logic, if the production function is submodular, the decentralized solution overinvests relative to the planner and again the difference between planner and decentralized investment rate for given p is increasing in λ.

If the worker's gains associated with a move to another firm fall short of the old firm's losses, one would expect training to be decreasing in the rate at which the worker meets outside vacancies, since it now raises the effective discount rate on the returns to human capital investments. The wage determination mechanism can possibly modify this effect since less friction may be associated with an increased rent extraction from any future employer and thereby a reduction in the future employer externality. In Acemoglu (1997), there is no on-the-job search and wages are set statically through bargaining where the worker's outside option is unemployment. Here, the strength of frictions is controlled by the arrival rate of offers out of unemployment. In this case, a faster meeting rate and thereby more competitive pressure, results in greater rent extraction by workers since wages are set by bargaining with unemployment as the outside option. Consequently, less friction implies a reduction in the future employer externality, and training increases. There is no ambiguity in this case because the rate at which workers are separated from their old employers is held constant in this argument. In a model such as this with on- the-job search, reduced frictions would discourage training through the increased discount rate on returns to training, modifying the unambiguous result in Acemoglu (1997).

Commitment Absent risk aversion, training inefficiencies are in the model purely due to the future employer externality. If the current match can extract all rents from future employer meetings, it will internalize the value of the destruction of match-specific capital in case the worker moves. We propose an instrument that achieves this outcome. The current firm can issue the following obligation: If the worker moves, the firm will pay the holder of the obligation the difference between the outside firm's willingness to pay and its own willingness to pay; that is $B = \bar{V}(p') - \bar{V}(p)$ where $p' > p$ is the type of the outside firm and

p is the type of the current firm. In a competitive market, the firm can sell this obligation at flow rate $\lambda \int_{\bar{V}(p)}^{\bar{V}(1)} \left[V - \bar{V}(p)\right] dF(V)$. With the obligation, the firm's willingness to pay for the worker comes to equal that of the outside firm.[10] Thus, the obligation allows for efficient separation and the current match extracts all the rents from future employers.

Subject to the obligation, the value of the current contract to the unskilled worker is

$$(r + \delta + \eta) V = w + \eta H + \delta U_0 + \lambda \int_V^{\bar{V}_0(1)} \hat{F}_0(V') \, dV'.$$

The firm is maximizing the profit expression,

$$\begin{aligned}(r + \delta) \Pi_0(V, p) = & f_0(p) - w - c(\eta) + \eta \left[\Pi_1(H, p) - \Pi_0(V, p)\right] - \\ & \lambda \int_V^{\bar{V}_1(p)} \hat{F}_0(V') \, dV',\end{aligned}$$

where the expected liability payment from the obligation resulting the worker quitting is perfectly offset by the revenue flow from the sale of the obligation. Furthermore, linearity of the profit function simplifies the profit loss integral from outside offers.

Insert the utility promise expression into the firm's profits to obtain

$$(r + \delta) \bar{V}_0(p) = f_0(p) - c(\eta) + \delta U_0 + \eta \left[\bar{V}_1(p) - \bar{V}_0(p)\right] + \lambda \int_p^1 \bar{V}_0(p') \, d\Phi(p'),$$

where the optimal training rate solves

$$c'(\eta) = \bar{V}_1(p) - \bar{V}_0(p).$$

The firm's willingness to pay for a skilled worker satisfies

$$(r + \delta) \bar{V}_1(p) = f_1(p) + \delta U_1 + \lambda \int_p^1 \bar{V}_1(p') \, d\Phi(p').$$

Notice that the expressions for $\bar{V}_h(p)$ perfectly match the social planner values, $\mathcal{V}_h(p)$, for $h = 0, 1$. Hence, the privately optimal training intensity coincides with that of the social planner.

[10] This implies that with the obligation a type-p firm will be setting a continuation value conditional on a higher-type outside firm meeting of $V^o\left(\bar{V}'\right) = \bar{V}' > \bar{V}$, which would involve a violation of the firm's participation constraint should the worker decide to stay with the firm. Thus, the obligation needs to state that in case the worker ends up staying with the current firm and it subsequently lays off the worker due to a violation of the participation constraint, then the firm must honor the payment, B, to the holder of the obligation in this case as well.

In our analysis, variations on the style of obligation as that above can undo the limitations to commitment that are implied by the renegotiation proofness restriction. We rule out the existence of markets for such instruments. Nevertheless, the mechanisms of the obligation instrument above are instructive: Efficiency is obtained by adoption of side payments not within the match, but rather with a third party so as to ensure a credible bargaining position with a possible future employer of the worker.[11]

3.2.2 Specific human capital

For specific training, the environment has an intriguing inefficiency: In the modular production function case, the decentralized economy provides the same level of specific training everywhere on the ladder. However, the social planner solution for specific training is increasing in firm type: The planner discounts match specific capital in low productivity firms at a greater rate because workers are more likely to reallocate to better firms. Therefore, the social planner invests more in specific training further up the ladder. The inefficiency in the decentralized economy is a result of future employers perfectly compensating the old match for its destruction, which includes the value of the match specific capital. Thus, there is a private return to match specific capital investment that is not present in the social returns. It implies that there tends to be too much specific training in low type firms. The following formalizes the argument.

In this section, ignore general human capital investments. Consider a modular production function with $f_m(p) = f(p) + m$. As before the social planner problem can be divided into a series of problems of maximizing the net present value of match output net of training of each given worker in each given state. In particular, consider the problem of maximizing the value of a worker in a low match specific capital match,

$$\begin{aligned}(r+\delta)\mathcal{V}_0(p) = & \max_{\mu}\left[f(p) - c_m(\mu) + \delta\mathcal{U} + \mu\left(\mathcal{V}_1(p) - \mathcal{V}_0(p)\right) + \lambda\int_p^1\left[\mathcal{V}_0(p') - \mathcal{V}_0(p)\right]d\Phi(p)\right] \\ = & f(p) + \delta\mathcal{U} + \mathcal{M}(p) + \lambda\int_p^1\frac{\left[f'(p') + \mathcal{M}'(p')\right]\hat{\Phi}(p')}{r+\delta+\lambda\hat{\Phi}(p')}dp',\end{aligned}$$

where the value of a high match specific capital match is,

$$\begin{aligned}(r+\delta)\mathcal{V}_1(p) &= f(p) + m + \delta\mathcal{U} + \lambda\int_{\tilde{p}(p)}^1\left[\mathcal{V}_0(p') - \mathcal{V}_0(\tilde{p}(p))\right]d\Phi(p) \\ &= f(p) + m + \delta\mathcal{U} + \lambda\int_{\tilde{p}(p)}^1\frac{\left[f'(p') + \mathcal{M}'(p')\right]\hat{\Phi}(p')}{r+\delta+\lambda\hat{\Phi}(p')}dp'.\end{aligned}$$

[11] Recently, Shi (2023) uses the insight that noncompete contracts work exactly the same way.

The threshold $\tilde{p}(p)$ is defined by $\mathcal{V}_1(p) = \mathcal{V}_0(\tilde{p}(p))$. Since $\mathcal{V}_0(p) < \mathcal{V}_1(p)$ and the value is increasing in p, it must be that $\tilde{p}(p) > p$. The loss of firm specific capital that is associated with switching firms must be compensated by a sufficiently large gain in firm type. The value of the investment option is,

$$\mathcal{M}(p) = \max_{\mu} \left[-c_m(\mu) + \mu\left(\mathcal{V}_1(p) - \mathcal{V}_0(p)\right)\right].$$

And the socially optimal specific investment choice solves,

$$c'_m(\mu(p)) = \mathcal{V}_1(p) - \mathcal{V}_0(p).$$

Some algebra yields,

$$\mathcal{V}_1(p) - \mathcal{V}_0(p) = \max_{\mu} \frac{m + c_m(\mu) - \int_p^{\tilde{p}(p)} \frac{\left[f'(p') + \mu\left[\mathcal{V}'_1(p') - \mathcal{V}'_0(p')\right]\right]\hat{\Phi}(p')}{r+\delta+\lambda\hat{\Phi}(p')} dp'}{r + \delta + \mu}. \tag{14}$$

Differentiation and the envelope theorem leads to,

$$\mathcal{V}'_1(p) - \mathcal{V}'_0(p) = \frac{f'(p)}{r+\delta}\left[\frac{\hat{\Phi}(p)}{r+\delta+\lambda\hat{\Phi}(p)} - \frac{\hat{\Phi}(\tilde{p}(p))}{r+\delta+\lambda\hat{\Phi}(\tilde{p}(p))}\right].$$

By $\tilde{p}(p) > p$ it follows that $\mathcal{V}'_1(p) - \mathcal{V}'_0(p) > 0$. Therefore, the social planner's choice of specific investment is increasing in p, $\mu'(p) = \left[\mathcal{V}'_1(p) - \mathcal{V}'_0(p)\right]/c''_m(\mu(p)) > 0$.

The inefficiency in specific human capital training in low-type firms arises because a future employer fully compensates the destruction of the match if the worker moves. Therefore, if the current match can extract all rents from future employer meetings, it will internalize the value of the destruction of match specific capital in case the worker moves. Therefore, and as in Section 3.2.1, an obligation issued by the current firm could restore efficiency.

4 Quantitative Analysis

We return to the full model with two aims. First, we analyze the impact of firm heterogeneity on endogenous training provision. Second, we reexamine the link between training and frictions.

4.1 Model parameterization and calibration

The model is calibrated to the U.S. economy. We use the following functional forms:

$$\begin{aligned}
u(w) &= \log(w) \\
c_h(\eta) &= \frac{\left(c_0^h \eta\right)^{1+c_1}}{1+c_1^h} \\
c_m(\mu) &= \frac{\left(c_0^m \mu\right)^{1+c_1}}{1+c_1^m} \\
f_{ij}(p) &= h_i + m_j + p, \qquad (i,j) \in \{0,1\}^2
\end{aligned}$$

Firm productivity is Pareto truncated below and above. The vector of parameters (ρ, θ, d) is set a priori using estimates from the literature. The death rate reflects an average working life of 40 years, $d = 0.025$. The discount rate is set to a 5% annual rate, $\rho = 0.05$. We choose a log-utility (with a coefficient of risk aversion of 1). Workers are hand-to-mouth and are therefore more risk-averse than workers with access to savings.

The remaining parameters are chosen to fit salient features of the U.S. labor market. To discipline the model's accumulation processes, we reproduce the age-earning profile. The model's distribution of firm productivity is calibrated to fit the firm-average wage distribution. In addition, the job destruction rate $\delta = 0.24$ is set to match the U.S. monthly layoff rate of 2%, and the calibration of $\lambda_0 = 4.3$ is set the fit the U.S. job-finding rate out of unemployment. We also match the U.S. annual job-to-job transition rate of 0.31.[12] Finally, we impose that both types of human capital are provided at the same intensity on average.

The age-earnings profile is obtained through the 1979–2020 survey years of the National Longitudinal Survey of Youth, 1979 (NLSY79). It is a representative sample of US households that was administered yearly from 1979-1994 by the Bureau of Labor Statistics, and once every two years since. We measure wages as the hourly pay rate at the time of the interview and deflate wages using the Personal Consumption Expenditures index (PCE) and trimmed for values below 3 and above 200. The sample is restricted to individuals' wage observations after they left school and never returned. Potential experience is defined as age minus the age of entry in the labor market. We restrict the sample to non-negative potential experience and above 40. Our final sample contains 12,655 individuals and 159,806 individual-year observations. The average number of observations per individuals is 16.63.

The firm wage distribution is obtained from Compustat, which provides annual accounting data on publicly listed US firms. We focus on the year 2014 as it is the year with the

[12] This corresponds to the employment to employment hazard rate reported by Rogerson and Shimer (2011).

Table 1: Model parameters

		I. Fixed
h_0	0.00	Unskilled productivity
d	0.03	Death rate
ρ	0.05	Discount factor
$\underline{p}$	1.00	Pareto distribution: lower bound
		II. Calibrated
δ	0.24	Job destruction rate
λ_u	4.30	Job offer rate: unemployed
λ_e	1.71	Job offer rate: employed
c_0^h	38.54	Training costs: constant (general)
c_0^m	6.18	Training costs: constant (specific)
$c_1^h = c_1^m$	0.81	Training costs: variable
$h_1 = m_1$	0.68	Skilled productivity
α	0.29	Pareto distribution: slope
$\overline{p}$	17.02	Pareto distribution: upper bound

largest number of observations and we drop firms with fewer than 10 employees. Our final sample contains 1,997 firms. Compustat has the advantage to be a publicly accessible dataset extensively utilized across various contexts. A significant limitation is its exclusion of a large number of smaller firms. We calculate the average wage per firms using the number of employees and the total wage bill.[13]

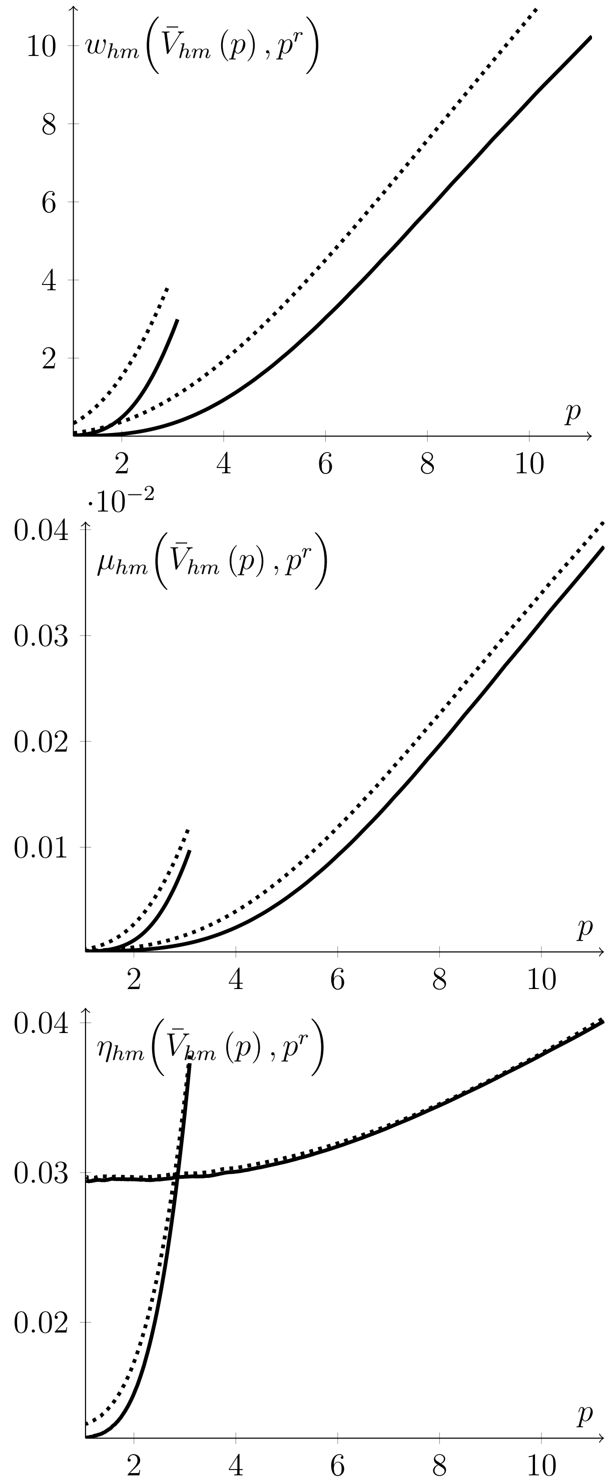
Wages are regressed on a full set of dummies for potential experience and individual fixed effects. We ask our model to match the dummies for potential experience. The first year of wages is excluded from the calibration due to the model's sharp assumption that all workers start their careers with a lifetime utility equal to the value of unemployment. The model parameters are chosen to reproduce the dispersion in average wage per worker, weighting each observation by the number of employees. The calibrated parameter values are reported in Table 1.

4.2 Calibrated contracts and steady state

Figure 2 shows the employment contracts for the 50th and 90th percentile firm productivity types as a function of the utility promise in the contract. The figure expresses the utility promise in terms of the willingness to pay of a firm with a given productivity p. This is done to facilitate comparison across contracts. Within a contract, the wage is increasing in the utility promise. Holding the utility promise constant, wages are decreasing in the firm type.

[13]The wage bill measure "represents salaries, wages, pension costs, profit sharing and incentive compensation, payroll taxes and other employee benefits."

Figure 2: Employment contracts by firm type



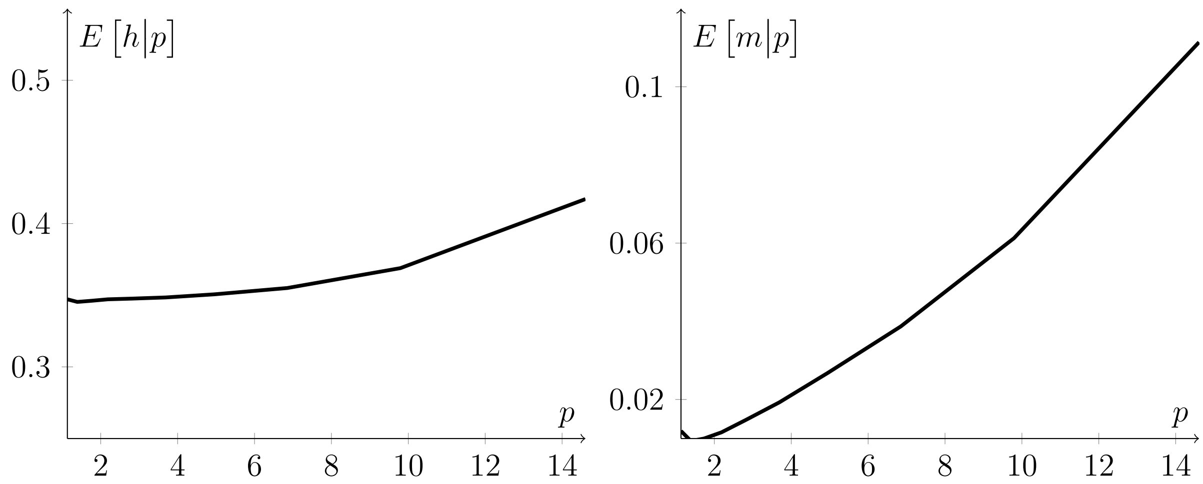
Note. Firm type conditional contracts drawn for $r = 0.5$ and $r = 0.9$. Top panel: Solid lines for $m = 0$ and dotted lines for $m = 1$. Middle panel: Solid lines for $h = 0$ and dotted lines for $h = 1$. Bottom panel: Solid lines for $(h, m) = (0, 0)$ and dotted lines for $(h, m) = (1, 1)$.

The latter is a well-known feature of the outside offer-matching feature of the wage mechanism, also seen in Postel-Vinay and Robin (2002). For a given utility promise, an increase in firm type implies greater expected gains from the on-the-job search process, which the firm can translate into higher profits through lower current wages. Production function complementarities between human capital and firm productivity can introduce a compensating differential between wages and training, but such considerations are not relevant given the modular production function specification in the current calibration. However, even in this case, whether higher-type firms on average pay higher wages depends on the composition of utility promises across their workers. We explore this in the next section. As shown in the previous section, both general and specific training increase in the utility promise within a contract. For a given firm type, a lower utility promise implies a steeper expected future wage path, which increases the cost of backloading. Therefore, training is lower for lower utility promises.

The differences in competitive pressure across the two types of training show up in the figures as well. Specific training within the 90^{th} percentile firm is almost constant in the utility promise whereas general training is considerably more sensitive to the utility promise. The competitive pressure on future utility promises associated with specific training is determined primarily by the firm's position in the firm hierarchy: As the match becomes more productive due to an increase in m, competitive pressure on the worker's future utility promises is only affected in the event that the worker meets a more productive firm than the current firm. The wage is lowered up front to reflect the expected utility promise gains associated with training. The only reason the current utility promise does play a role in the provision of specific training is because the surplus loss associated with lowering the worker's wage is proportional to the worker's marginal utility, which is decreasing in the utility promise. The wage is lowered up front to reflect the expected utility promise gains associated with training. The increased competitive pressure associated with increased general human capital is on the other hand primarily determined by the current utility promise, V. A meeting with any productivity firm greater than $p_{1m}(V)$ is associated with an increased utility promise pressure due to the increase in h. Thus, for a lower V there is a larger mass of outside firms that can exert pressure on the match. In combination with the greater marginal utility of wages associated with the lower utility promise, V, the surplus loss of reducing the worker's wages up front in expectation of the future utility promise gains from general training is more sensitive to V.

Finally, risk aversion is a separate source of positive complementarity between general and specific training. If a worker's skill increases, her utility promise increases and her wages come to increase faster. The lower marginal utility of wages reduces the surplus loss

Figure 3: Average human capital level by firm type.



associated with the backloading of wages due to one type of training. This effect is related to the strategic complementarity results in Balmaceda (2005) and Kessler and Lülfesmann (2006) where the existence of non-contractable specific training can counteract the hold-up problem in particular wage bargaining settings.

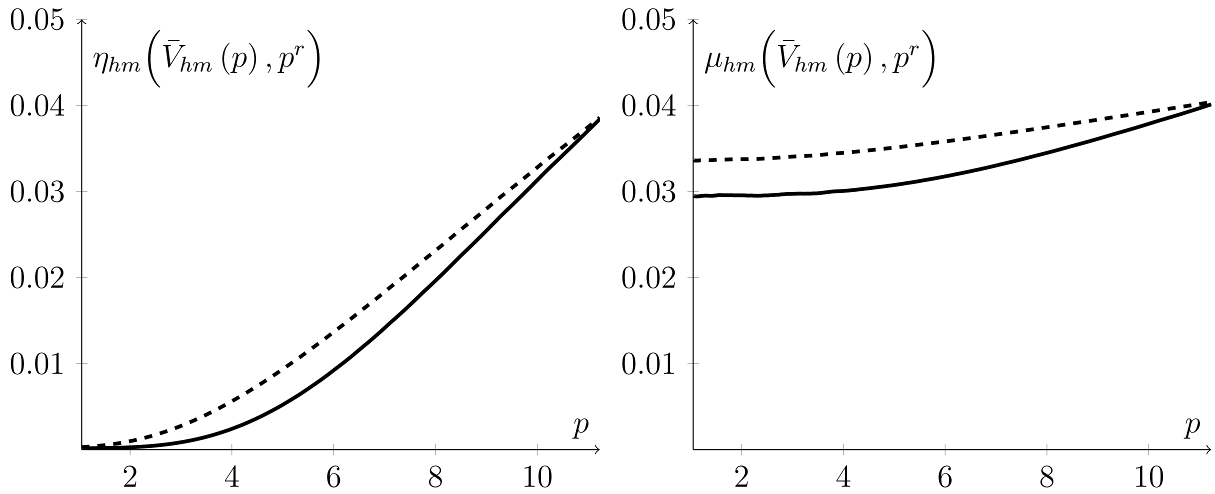
Figure 3 shows the average levels of human capital by firm type in the steady state. As can be seen, the model implies significant positive sorting between firm productivity and worker skill. The labor force of higher-ranked firms is more skilled and has higher match specific capital. Therefore, more productive firms yield higher output due to better technology and the presence of more skilled workers. This is not a result of positive assortative matching as there are no complementarities in production in the calibration. It is a reflection of the state dependence in the model that fortunate employment draws with more productive firms contribute to a better contract value and also to a faster development of both general and specific skills.

4.3 Training and frictions

Acemoglu and Pischke (1999) emphasize that increased labor market friction allows firms to provide more general training when it is costly to resolve the hold-up problem by making the worker pay for training up front through lower wages. Wasmer (2006) adds to the argument that increased labor market friction will increase specific training in a setup where matches invest in specific capital to reduce the risk of job destruction.

Figure 4 demonstrates the Acemoglu and Pischke (1999) mechanism within a given firm's contract. It shows the training choices for a firm at the 90th percentile productivity when the contact rate is low ($\lambda_e = 0.88$) and when the contact rate is high ($\lambda_e = 1.76$). The horizontal

Figure 4: Firm type $\Phi(p) = 0.9$ employment contract for $\lambda_e = 1.76$ and $\lambda_e = 0.88$



Note. Solid line drawn for $\lambda_e = 1.76$ and dashed line drawn for $\lambda_e = 0.88$.

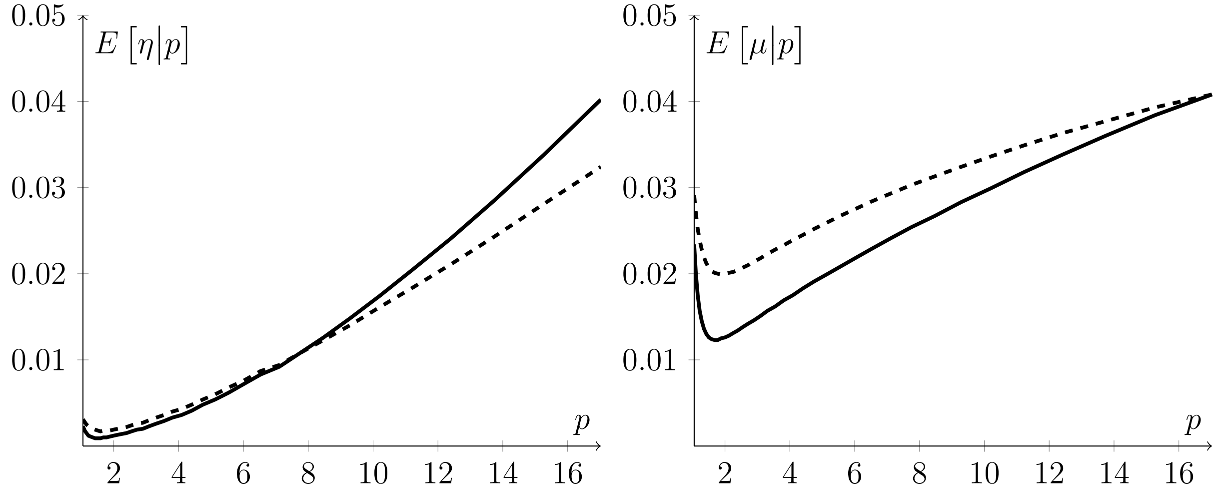
axis is the utility promise support of the contract represented by the willingness to pay of a type-p firm. The figure shows that holding the firm's utility promise fixed, an increase in the contact rate is associated with a decrease in training. For the given competitive position as represented by the utility promise, the greater contact rate implies steeper future wages, a lower current wage, and therefore a more severe cost of backloading. This is precisely the Acemoglu and Pischke (1999) argument that holds in our environment and it applies not only to general training but also to specific training.

But the steady-state utility promise composition within a firm's labor force is not constant in changes in the contact rate. Specifically, a higher contact rate implies a right shift of utility promises resulting from greater competitive pressure between firms.

Figure 5 shows the average training levels by firm type in steady state for $\lambda_e = 0.88$ and $\lambda_e = 1.76$. As shown, this effect by itself substantially modifies the Acemoglu and Pischke (1999) mechanism. High productivity firms and middle productivity firms offer more general training when the contact rate goes up. At the bottom of the distribution, training is reduced. For specific training, the Acemoglu and Pischke (1999) argument dominates throughout except at the very top where training is invariant to frictions.

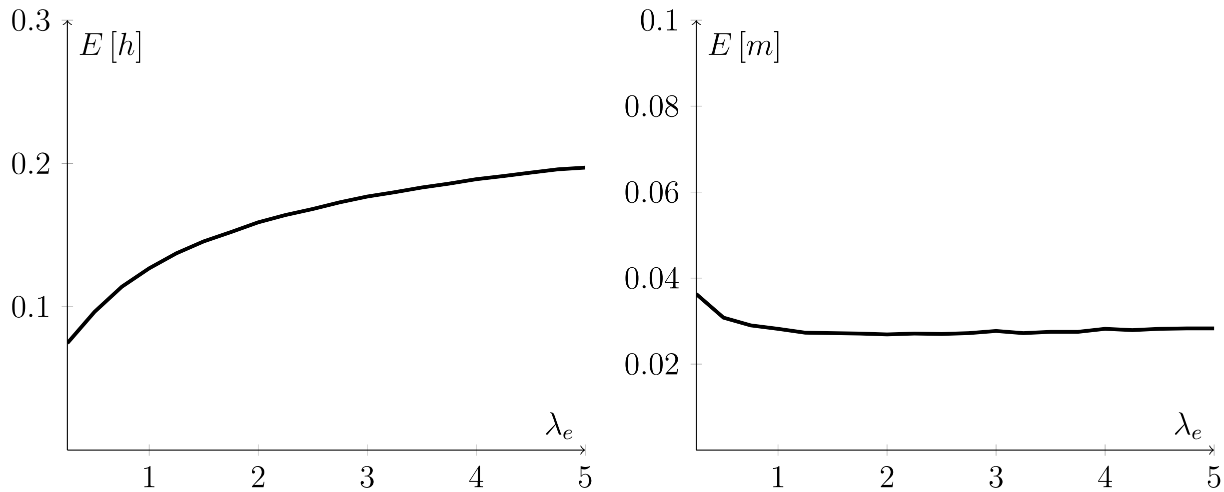
In addition to the composition of utility promises, overall training and accumulation of skills in the economy also depend on the match distribution, which is also affected by changes in frictions. As the contact rate increases, mismatch declines since workers are matching with better firms. General training is decreasing in mismatch which will tend to increase general training. The overinvestment in specific training at the lower end will be alleviated. Figure 6 shows the average human capital levels in the steady-state economy for different levels

Figure 5: Average steady state firm type conditional training levels for $\lambda_e = 0.88$ and $\lambda_e = 1.76$.



Note. Solid line drawn for $\lambda_e = 1.76$ and dashed line drawn for $\lambda_e = 0.88$.

Figure 6: Steady state share of skilled workers by λ_e.



of contact rates. As can be seen, generally skills are robustly increasing in the contact rate, which is opposite to the intuition developed in Acemoglu and Pischke (1999). The analysis in this paper embodies the central mechanism in their paper, but it is dominated by composition effects from search on the job and the presence of firm heterogeneity that is a natural consequence of a frictional labor market environment.

For the given calibration, match specific capital is stable in the contact rate. Specific training is increasing in firm type and eventually the improved match distribution will result in more training as mismatch declines. However, for lower contact rates, the lower training levels within firm type for given utility promises dominate and result in less specific training.

5 Concluding remarks

We have put forth a framework for the study of wage dynamics that allows for search frictions, firm heterogeneity, and human capital accumulation. In contrast to passive learning processes, we model the active investment in general human capital and match specific capital in response to the magnitude of the returns. The intensity of labor market competition is a primary factor in the determination of the returns to training, and we perform the analysis in a frictional setting where heterogeneous firms naturally coexist and workers can move directly between firms through a standard on-the-job search process. Optimally designed employment contracts set wages and training rates conditional on the history of the match.

We find that training varies by firm type. In isolation, the moral hazard problem associated with training implies that more productive firms train more. A supermodular production function in human capital and firm productivity will amplify this relationship. Thus, aggregate human capital accumulation ultimately depends on the equilibrium match distribution of worker over firm types, and we show that it is of first-order importance in the model calibrated to the US economy. The classic Acemoglu and Pischke (1999) result that increased labor market friction alleviates the hold up problem in training and therefore results in more training is overturned through dominating equilibrium effects. Increased labor market friction results in worse matches and reduced bargaining positions, both of which imply reduced training.

In terms of the classic decomposition of labor market outcomes into luck and skill, the current analysis demonstrates that variation in skill is at least in part a result of variation in luck. The calibrated economy displays substantial sorting despite the absence of complementarities in production and the absence of assortative matching. The positive relationship between worker skill and firm productivity is a result of the faster accumulation of skill. Consequently, more productive firms tend to have more skilled workers as well as higher match specific capital.

The presence of firm heterogeneity also allowed us to point to an important feature of wage dynamics and specific training: The presence of more productive firms than the current firm implies that the match value increase associated with specific training can be contested in the market by these more productive firms. Therefore, specific training is associated with both increasing wages within the job as well as increased wages with future employers. Consequently, the distinction between tenure and experience effects in wage dynamics is not by itself sufficient to evaluate the importance of specific relative to general training. An avenue for future research is to utilize worker reallocation patterns to help with the separate identification of the two processes: Specific training reduces reallocation whereas general training has no impact on mobility in the model.

A Additional utility and profit expressions

We reported the profit expression and utility promise constraint for a worker with $(h,m)=(0,0)$ in the main text. The expressions for the other cases are simpler and reported below. We start by presenting the profit expressions.

$$(r+\delta)\,\Pi_{11}(V,p)= \max_{w\in\Gamma_{11}(V,p)}\left\{f_{11}(p)-w+\lambda_e\int_V^{\bar{V}_{11}(p)}\Pi_{11}'(V',p)\,\hat{F}_{10}(V')\,dV'\right\}$$

$$\begin{aligned}(r+\delta)\,\Pi_{01}(V,p)= &\max_{\{w,\eta,H\}\in\Gamma_{01}(V,p)}\left\{f_{01}(p)-w-c_h(\eta)+\lambda_e\int_V^{\bar{V}_{01}(p)}\Pi_{01}'(V',p)\,\hat{F}_{00}(V')\,dV'\right.\\ &\left.+\,\eta\left[\Pi_{11}(H,p)-\Pi_{01}(V,p)\right]\right\}\end{aligned}$$

$$\begin{aligned}(r+\delta)\,\Pi_{10}(V,p)= &\max_{\{w,\mu,M\}\in\Gamma_{10}(V,p)}\left\{f_{10}(p)-w-c_m(\mu)+\lambda_e\int_V^{\bar{V}_{10}(p)}\Pi_{10}'(V',p)\,\hat{F}_{10}(V')\,dV'\right.\\ &\left.+\,\mu\left[\Pi_{11}(M,p)-\Pi_{10}(V,p)\right]\right\}\end{aligned}$$

We now turn to the utility promise constraints.

$$\begin{aligned}(r+\delta)\,V_{11} &= u(w)+\delta U_1+\lambda_e\int_V^{\bar{V}_{11}(p)}\hat{F}_{10}(V')\,dV'\\ (r+\delta+\eta)\,V_{01} &= u(w)+\eta H+\delta U_0+\lambda_e\int_V^{\bar{V}_{01}(p)}\hat{F}_{00}(V')\,dV'\\ (r+\delta+\mu)\,V_{10} &= u(w)+\mu M+\delta U_1+\lambda_e\int_V^{\bar{V}_{10}(p)}\hat{F}_{10}(V')\,dV'\end{aligned}$$

Finally, we present the set of feasible contract choices:

$$\begin{aligned}
\Gamma_{11}(V,p) &= \left\{ w \,\middle|\, u(w) + \delta U_h + \lambda_e \int_V^{\bar{V}_{11}(p)} \hat{F}_{10}(V')\,dV' = (r+\delta)V \right\} \\
\Gamma_{01}(V,p) &= \Bigg\{ (w,\eta,H) \,\Bigg|\, u(w) + \eta H + \delta U_0 + \lambda_e \int_V^{\bar{V}_{01}(p)} \hat{F}_{00}(V')\,dV' = (r+\delta+\eta)V, \\
&\qquad U_0 \le H \le \bar{V}_{10}(p) \Bigg\} \\
\Gamma_{10}(V,p) &= \Bigg\{ (w,\mu,M) \,\Bigg|\, u(w) + \mu M + \delta U_1 + \lambda_e \int_V^{\bar{V}_{10}(p)} \hat{F}_{10}(V')\,dV' = (r+\delta+\mu)V, \\
&\qquad U_1 \le M \le \bar{V}_{01}(p) \Bigg\}
\end{aligned}$$

B Proof of Lemma 1

Consider the case, $\bar{V}_{h0}(p') \in \left[V, \bar{V}_{hm}(p)\right]$. If $\Omega_{hm}(p') < \bar{V}_{h0}(p')$, the outside firm would match $\Omega_{hm}(p') + \epsilon$ where $\epsilon > 0$ is arbitrarily small and the worker moves the new firm. This is not renegotiation proof as the incumbent firm and the worker would agree to renegotiate the existing contract which will make both worker and incumbent firm weakly better off. If $\Omega_{hm}(p') > \bar{V}_{h0}(p')$, the incumbent firm could win the worker services with a lower utility promise. Thus, the outcome in this case if that worker stays with current firm $\alpha(\Omega(p'), p') = 1$ and $\Omega_{hm}(p') = \bar{V}_{h0}(p')$.

Now, consider the case where $\bar{V}_{h0}(p') > \bar{V}_{hm}(p)$. If $\Omega_{hm}(p') < \bar{V}_{hm}(p)$ and $\alpha(\Omega(p'), p') = 0$, the worker moves to the outside firm with a utility promise of $\Omega_{hm}(p')$. This is not renegotiation proof. Both worker and incumbent firm would agree to a change in the contract so that $\bar{V}_{h0}(p') > \Omega_{hm}(p') \ge \bar{V}_{hm}(p)$. This strictly improves the worker's position and leaves the incumbent firm no worse off. Yet if the incumbent firm offers $\Omega_{hm}(p') > \bar{V}_{hm}(p)$,the outside firm counter the offer with a promise of $\Omega_{hm}(p') + \epsilon$ where $\epsilon > 0$ is arbitrarily small. The worker will accept the outside firm's offer because if he accepts the incumbent firm offer, he knows it will be renegotiated down to a utility promise no greater than $\bar{V}_{hm}(p)$. Therefore, the optimal renegotiation proof contract must be such that $\Omega_{hm}(p') = \bar{V}_{hm}(p)$ and $\alpha(\Omega(p'), p') = 0$.

Finally consider the case, $\bar{V}_{h0}(p') < V$. In this case, limited commitment does not impose a constraint on the optimal design of the utility promise path of the contract which is flat due to the concavity of worker utility. Subgame perfection refines to $\Omega_{hm}(p') = V$ and $\alpha(\Omega(p'), p') = 1$.

C Proof of Lemma 2

Let the Lagrange multiplier on the promise-keeping constraint be $\gamma_{hm}(V,p)$, where $\gamma_{hm}(V,p) > 0$ is a sufficient condition for the recursive formulation of the contracting problem to be valid. Furthermore, denote by $\varphi_m(V,p)$ the Lagrange multiplier on the worker's participation constraint, $U_1 \leq H$. It is verified that the other constraints are not binding for the optimal contract. Since unemployment benefits depend on general human capital, the worker participation constraint might be binding when the worker becomes generally skilled. Since an increase in match specific capital involves an increase in joint match value and neither the worker's or firm's outside options are affected, the participation constraints will not bind in the case when the worker becomes specifically skilled.

In the absence of minimum wages or other constraints on the wage design, the slope of the profit function satisfies

$$\Pi'_{hm}(V,p) = -(r+\delta)\,\gamma_{hm}(V,p) = -\frac{1}{u'(w_{hm}(V,p))} < 0, \tag{15}$$

which follows from the first-order conditions on the choices of w and $\dot{V}$. That is, the profit function is strictly decreasing in the utility promise. In addition, wages $w_{hm}(V,p)$ are strictly increasing in the utility promise given the concavity of the profit function.

By the derivative of the Lagrangian $\partial\mathcal{L}/\partial V = \Pi'(V)$ and the envelope theorem, one obtains

$$\Pi'_{hm}(V,p) + (r+\delta)\,\gamma_{hm}(V,p) = \frac{\Pi''_{hm}(V,p)}{r+\delta+\lambda\hat{F}_h(V)+\eta_m(V,p)+\mu_h(V,p)}\dot{V}_{hm}(V,p)\,. \tag{16}$$

Together with equation (15), it therefore must be that in the absence of outside offers and skill increases, the optimal employment contract is flat:

$$\dot{V}_{hm}(V,p) = 0. \tag{17}$$

The human capital change conditional utility promises satisfy the first order equations,

$$\Pi'_{1m}(H_m(V,p),p) - \Pi'_{0m}(V,p) = \frac{-(r+\delta)\,\varphi_m(V,p)}{\eta_m(V,p)} \tag{18}$$

$$\Pi'_{h1}(M_h(V,p),p) - \Pi'_{h0}(V,p) = 0. \tag{19}$$

If the worker participation constraint is not binding, $\varphi_{hm}(V,p) = 0$, the wage profile is flat over human capital jumps,

$$\Pi'_{0m}(V,p) = \Pi'_{1m}(H_m(V,p),p)$$
$$\Pi'_{h0}(V,p) = \Pi'_{h1}(M_h(V,p),p),$$

which implies,

$$w_{0m}(V,p) = w_{1m}(H_m(V,p),p)$$
$$w_{h0}(V,p) = w_{h1}(M_h(V,p),p).$$

If the worker's participation constraint is binding following a skill increase, wages jump up because wages are increasing in the utility promise. The participation constraint forces the firm to offer a greater utility promise than the one that makes wages smooth across the skill jump. The binding participation implies, $\Pi'_{0m}(V,p) > \Pi'_{1m}(H_m(V,p),p)$ and therefore $w_{0m}(V,p) < w_{1m}(H_m(V,p),p)$.

Now, consider the claim that $V < M_h(V,p) < \bar{V}_{h1}(p)$ for $p < 1$. Proof is by contradiction. Suppose first that $M_h(V,p) = \bar{V}_{h1}(p)$. For the sake of simplicity, take the case where $h = 1$. Trivially, it must be that $w_{11}\left(\bar{V}_{11}(p),p\right) = f_{11}(p)$ since there is no possibility of future wage gains within the contract. It must then be that $w_{10}(V,p) \leq w_{10}\left(\bar{V}_{10}(p),p\right) \leq f_{10}(p) - c_m\left(\mu\left(\bar{V}_{10}(p),p\right)\right) < f_{11}(p)$. This is because at $M_1(V,p) = \bar{V}_{11}(p)$ the firm hands over all gains to specific training to the worker. Hence, $\Pi_{10}\left(\bar{V}_{10}(p),p\right) = 0$ implies that wages w_{10} cannot exceed production less training costs. Thus, $M_h(V,p) = \bar{V}_{h1}(p)$ implies that $w_{10}(V,p) < w_{11}(M_h(V,p),p)$, violating (19). Suppose instead by contradiction that $M_1(V,p) \leq V$. By the utility promise constraint we have that,

$$\begin{aligned}
(r+\delta)M_1(V,p) &= u(w_{11}(M_1(V,p),p)) + \delta U_1 + \lambda_e \int_{M_1(V,p)}^{\bar{V}_{11}(p)} \hat{F}_1(V')\,dV' \\
&= u(w_{10}(V,p)) + \delta U_1 + \lambda_e \int_{M_1(V,p)}^{\bar{V}_{11}(p)} \hat{F}_1(V')\,dV' \\
&> u(w_{10}(V,p)) + \mu(V,p)[M_1(V,p) - V] + \delta U_1 + \lambda_e \int_{V}^{\bar{V}_{10}(p)} \hat{F}_1(V')\,dV' \\
&= (r+\delta)V.
\end{aligned}$$

The second equality follows from (19). The inequality follows directly from the presumption that $M_1(V,p) \leq V$ and that $\bar{V}_{11}(p) > \bar{V}_{10}(p)$. Therefore $M_1(V,p) \leq V$ is contradicted. The basic intuition is that since wages are smooth across the human capital change, a utility promise $M_1(V,p) \leq V$ implies greater future utility promise growth than prior to the human capital increase. At an unchanged current wage level, a greater future utility promise growth is inconsistent with a reduction in the utility promise. Hence it must be that $V < M_h(V,p) < \bar{V}_{h1}(p)$. The $p = 1$ case is the exception. In this case $M_1(V,1) = V$. The reason being that $\bar{V}_{10}(1)$ is the upper bound on the support of $F_1(V)$. The fact that the firm's willingness to pay increases from $\bar{V}_{10}(1)$ to $\bar{V}_{11}(1)$ does not result in an increase in the worker's expected utility promise growth rate for any given utility promise, because there are no outside firms to challenge the increase.

Arguments for $h = 0$ as well as the skill increase conditional utility promise $V < H_m(V,p) < \bar{V}_{1m}(p)$ go along the same lines.

D Steady state conditions

Assuming that unemployed workers do not turn down any meetings, the steady state conditions on the employment and unemployment stocks are,

$$(d+\lambda_u)\,u_0 = d + \delta\,(e_{00}+e_{01}) \tag{20}$$

$$(d+\lambda_u)\,u_1 = \delta\,(e_{10}+e_{11}) \tag{21}$$

$$(d+\delta+\bar{\eta}_0+\bar{\mu}_0)\,e_{00} = \lambda_u u_0 + e_{01}\int_0^1\int_{U_0}^{\bar{V}_{01}(p')} \lambda\hat{F}_0\left(\bar{V}_{01}(p')\right) g_{01}(V',p')\,dV'dp' \tag{22}$$

$$(d+\delta+\bar{\mu}_1)\,e_{10} = \lambda_u u_1 + \bar{\eta}_0 e_{00} + e_{11}\int_0^1\int_{U_1}^{\bar{V}_{11}(p')} \lambda\hat{F}_1\left(\bar{V}_{11}(p')\right) g_{11}(V',p')\,dV'dp' \tag{23}$$

$$\bar{\mu}_0 e_{00} = \left(d+\delta+\bar{\eta}_1+\int_0^1\int_{U_0}^{\bar{V}_{01}(p')} \lambda\hat{F}_0\left(\bar{V}_{01}(p')\right) g_{01}(V',p')\,dV'dp'\right) e_{01} \tag{24}$$

$$\bar{\eta}_1 e_{01} + \bar{\mu}_1 e_{10} = \left(d+\delta+\int_0^1\int_{U_1}^{\bar{V}_{11}(p')} \lambda\hat{F}_1\left(\bar{V}_{11}(p')\right) g_{11}(V',p')\,dV'dp'\right) e_{11}, \tag{25}$$

where $\bar{\mu}_h = \int_0^1\int_{U_h}^{\bar{V}_{h0}(p')} \mu_h(V',p')\,dG_{h0}(V,p)$ and $\bar{\eta}_m = \int_0^1\int_{U_0}^{\bar{V}_{0m}(p')} \eta_m(V',p')\,dG_{0m}(V,p)$.

The steady state conditions on $e_{01}G_{01}(V,p)$, $e_{10}G_{10}(V,p)$, and $e_{11}G_{11}(V,p)$ are respectively,

$$\begin{aligned} e_{00}\int_0^p\int_{U_0}^{\bar{V}_{00}(p')} 1\left[M_0(V',p')\le V\right]\mu_0(V',p')\,g_{00}(V',p')\,dV'dp' = \\ e_{01}\int_0^p\int_{U_0}^{\bar{V}_{01}(p)} \left[m+\delta+\eta_1(V',p')+\lambda\hat{F}_0\left(\bar{V}_{01}(p')\right)\right] g_{01}(V',p')\,dV'dp'. \end{aligned}$$

The steady state condition on $e_{10}G_{10}(V,p)$ is,

$$\lambda_u u_1 \Phi(p) + \lambda e_{11} \int_0^{p_{11}(V)} \int_{U_1}^{\bar{V}_{11}(p')} \left[F_1\left(\bar{V}_{10}(p)\right) - F_1\left(\bar{V}_{11}(p')\right)\right] g_{11}(V',p')\, dV'dp'$$
$$+ e_{00} \int_0^p \int_{U_0}^{\bar{V}_{00}(p')} 1\left[H_0(V',p') \leq V\right] \eta_0(V',p')\, g_{00}(V',p')\, dV'dp' =$$
$$e_{10}\Bigg\{ \int_0^{p_{10}(V)} \int_U^{\bar{V}_{10}(p')} \left[d + \delta + \mu_1(V',p') + \lambda \hat{F}_1\left(\bar{V}_{10}(p)\right)\right] g_{10}(V',p')\, dV'dp' +$$
$$\int_{\bar{p}_{10}(V)}^{p} \int_U^V \left[d + \delta + \mu_1(V',p') + \lambda \hat{F}_1(V)\right] g_{10}(V',p')\, dV'dp' \Bigg\}.$$

And finally, the steady state condition on $e_{11}G_{11}(V,p)$ is,

$$e_{10} \int_0^p \int_{U_0}^{\bar{V}_{10}(p')} 1\left[M_1(V',p') \leq V\right] \mu_1(V',p')\, g_{10}(V',p')\, dV'dp' =$$
$$e_{11} \int_0^p \int_{U_0}^{\bar{V}_{11}(p)} \left[d + \delta + \lambda \hat{F}_1\left(\bar{V}_{11}(p')\right)\right] g_{11}(V',p')\, dV'dp'.$$

E Numerical Solution

Firm productivity is discretized and each element of the grid $\{p_j\}_{j=1,\dots,N_p}$ has equal probability $\frac{1}{N_p}$. All integrals are numerically approximated with Gauss-Legendre quadrature. We use linear interpolation to approximate policy functions off the grid. We solve for the optimal contracts using the following iterative algorithm. Use initial guesses for the functions F_1 and F_0.

1. Solve for $\bar{V}_{11}^j$ and the corresponding w using (4), (3) and (2).
2. For any $V_{11} \in \left[U_1, \bar{V}_{11}^j\right]$, solve for the corresponding (w, Π) using (4), (3) and (15).
3. Solve for $\bar{V}_{10}^j$ and the corresponding (w, μ, Π) using (4), (3), (2) and (8).
4. Using the previous step, update F_1. Return to 1. until convergence.
5. For any $V_{10} \in \left[U_1, \bar{V}_{10}^j\right]$, solve for the corresponding (w, η, Π) using (4), (3), (15) and (8).
6. Solve for $\bar{V}_{01}^j$ and the corresponding (w, η) using (4), (3), (2) and (7).
7. For any $V_{01} \in \left[U_0, \bar{V}_{01}^j\right]$, solve for the corresponding (w, μ, Π) using (4), (3), (15) and (7). If the participation constraint is not satisfied, we set $H = U_1$.

8. Solve for $\bar{V}_{00}^{j}$ and the corresponding (w, η, μ) using (4), (3), (2), (7) and (8). If the participation constraint is not satisfied, we set $H = U_1$.

9. Using the previous step, update F_0. Return to 5. until convergence.

10. For any $V_{00} \in \left[U_0, \bar{V}_{00}^{j}\right]$, solve for the corresponding (w, η, μ, Π) using (4), (3), (15), (7) and (8).

In step 5 and above, we check whether the participation constraint $H \geq U_1$ is binding.

References

Acemoglu, D. (1997). Training and innovation in an imperfect labour market. *Review of Economic Studies*, 64:445–464.

Acemoglu, D. and Pischke, J.-S. (1999). The structure of wages and investment in general training. *Journal of Political Economy*, 107(3):539–572.

Altonji, J. G. and Shakotko, R. A. (1987). Do wages rise with job seniority? *Review of Economic Studies*, 54(3):437–459.

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.

Bagger, J. and Lentz, R. (2019). An Empirical Model of Wage Dispersion with Sorting. *The Review of Economic Studies*, 86(1):153–190.

Balke, N. and Lamadon, T. (2022). Productivity shocks, long-term contracts, and earnings dynamics. *American Economic Review*, 112(7):2139–77.

Balmaceda, F. (2005). Firm-sponsored general training. *Journal of Labor Economics*, 23(1):115–133.

Becker, G. S. (1964). *Human Capital*. University of Chicago Press, Chicago.

Burdett, K. and Coles, M. G. (2003). Equilibrium wage-tenure contracts. *Econometrica*, 71(5):1377–1404.

Burdett, K. and Judd, K. L. (1983). Equilibrium price dispersion. *Econometrica*, 51(4):955–69.

Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, 39(2):257–273.

Carrillo-Tudela, C. and Smith, E. (2014). Search capital. *Working Paper*.

Christensen, B. J., Lentz, R., Mortensen, D. T., Neumann, G., and Werwatz, A. (2005). On the job search and the wage distribution. *Journal of Labor Economics*, 23(1):31–58.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (2001). Efficient non-contractible investments in large economies. *Journal of Economic Theory*, 101(2):333–373.

Engbom, N. (2022). Labor market fluidity and human capital accumulation. Technical report, National Bureau of Economic Research.

Felli, L. and Harris, C. (1996). Learning, wage dynamics, and firm-specific human capital. *Journal of Political Economy*, 104(4):838–868.

Fu, C. (2011). Training, search and wage dispersion. *Review of Economic Dynamics*, 14:650–666.

Fujita, S. and Moscarini, G. (2013). Recall and unemployment. *Working Paper*.

Hopenhayn, H. A. and Nicolini, J. P. (1997). Optimal unemployment insurance. *Journal of Political Economy*, 105(21):412–38.

Jolivet, G., Postel-Vinay, F., and Robin, J.-M. (2006). The empirical content of the job search model: Labor mobility and wage distributions in europe and the US. *European Economic Review*, 50(4):877–907.

Kessler, A. S. and Lülfesmann, C. (2006). The theory of human capital revisited: On the interaction of general and specific investments. *The Economic Journal*, 116:903–923.

Lentz, R. (2009). Optimal unemployment insurance in an estimated job search model with savings. *Review of Economic Dynamics*, 12(1):37–57.

Lentz, R. (2014). Optimal employment contracts with hidden search. *NBER Working Papers*, (19988).

Lentz, R. and Tranæs, T. (2005). Job search and savings: Wealth effects and duration dependence. *Journal of Labor Economics*, 23(3):467–90.

Lise, J. (2013). On-the-job search and precautionary savings. *Review of Economic Studies*, 80(3):1086–1113.

Lise, J. and Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, 110(8):2328–2376.

Moen, E. R. and Rosén, Å. (2004). Does poaching distort training? *Review of Economic Studies*, 71:1143–1162.

Nöldeke, G. and Samuelson, L. (2015). Investment and competitive matching. *Econometrica*, 83(3):835–896.

Phelan, C. and Townsend, R. M. (1991). Computing multi-period, information-constrained optima. *Review of Economic Studies*, 58(5):853–81.

Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350.

Rogerson, R. and Shimer, R. (2011). Search in macroeconomic models of the labor market. In *Handbook of Labor Economics, Volume 4a*. Elsevier Science B.V.

Sanders, C. and Taber, C. (2012). Life-cycle wage growth and heterogenous human capital. *Annual Review of Economics*, 41(1):399–425.

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *Review of Economic Studies*, 75:957–984.

Shi, L. (2023). Optimal regulation of noncompete contracts. *Econometrica*, 91(2):425–463.

Spear, S. E. and Srivastava, S. (1987). On repeated moral hazard with discounting. *Review of Economic Dynamics*, pages 599–617.

Taber, C. and Vejlin, R. (2020). Estimation of a roy/search/compensating differential model of the labor market. *Econometrica*, 88(3):1031–1069.

Thomas, J. and Worrall, T. (1988). Self-enforcing wage contracts. *Review of Economic Studies*, pages 541–554.

Topel, R. H. (1991). Specific capital, mobility and wages: Wages rise with job seniority. *Journal of Political Economy*, 99(1):145–176.

Wasmer, E. (2006). Specific skills, search frictions, and firing costs. *American Economic Review*, 96(3):811–31.