Wage-Training Contracts and Wage Dynamics
In a Job Search Model with General Human Capital\textsuperscript{1}

Chao Fu\textsuperscript{2}
University of Pennsylvania

First Draft: December, 2006
This Version: March, 2008

\textsuperscript{1}I thank Kenneth Burdett, Philipp Kircher and Guido Menzio for their invaluable guidance and support. I thank Roberto Pinheiro, Xi Weng and Randall Wright for insightful comments and encouragements. I also thank participants of the April 2007 Penn Search and Matching workshop for helpful comments and discussions. All remaining errors are mine.

\textsuperscript{2}Department of Economics, University of Pennsylvania, Room 160 McNeil, 3718 Locust Walk, Philadelphia, PA 19104, USA.
Email: chaofu@sas.upenn.edu
Abstract

This paper combines on-the-job search and human capital accumulation into the study of firm-provided general training and its implication for wage dynamics. Although firms are ex ante identical, some but not all of them offer general training. When training is provided, the cost is borne jointly by the firm and the worker. Moreover, firms with training are those offering higher-value jobs. Workers may take wage cuts in order to transit from a job without training to one with training. Although workers are ex ante identical, wage dispersion exists not only because workers of the same quality are paid differently across firms, but also because workers differ in their productivity ex post. The interaction between job search and general human capital accumulation yields a distribution of wages that features a density with a long and decreasing right tail, as observed in the data.

Keywords: On-the-job search, general human capital, wage-training contract, wage dynamics
JEL codes: J64, J24, J31
1 Introduction

Should a firm offer general training to its workers? According to the standard theory by Becker (1964) and Mincer (1974), the answer is no. General training increases a worker’s productivity in a range of employment opportunities, and therefore translates into higher earnings in a competitive labor market. Thus, it is the worker who should pay for general training. However, in reality, we do see some firms offer general training to their employees. Moreover, even if one takes into account the possibility that firms may internalize the cost of general training by reducing the worker’s wage, it is shown that employers pay at least part of the cost, see, for example, Acemoglu and Pischke (1999).

Why would a firm offer general training? It is well known in the training literature that some labor market imperfection must exist so that the mobility of workers is restricted and that employers earn rents on trained workers. However, a theory on firm-provided training is not satisfactory if it yields sizable training only when job-to-job transitions are rare, given the high frequency of such transitions observed in the data. The coexistence of firm-provided general training and frequent job-to-job transitions calls for a model that can accommodate both. In this paper, I propose a theory based on the interaction between on-the-job search and general human capital accumulation. Due to on-the-job search, even a firm with training may frequently lose its workers to other firms that offer more valuable jobs. However, worker mobility is still restricted by search friction and firms still have substantial incentive to provide general training. In addition to explaining why a firm would offer general training, my model also generates different training decisions by ex-ante identical firms, and predicts which firms would train workers and which would not.

Substantial ex-post heterogeneity also exists among workers although they are ex ante identical. Workers with the same quality could differ in the pay rates they earn due to search friction. And more importantly, workers are essentially different, in that there is a non-degenerate distribution of human capital levels among them. In this model, wages grow as workers become more experienced for two reasons: they get better jobs as they sample more, and workers themselves become better (more productive) over time. The structure of this model sets up a building block for the assessment of the relative importance of each of these two components.

To be specific, in this paper, I analyze an equilibrium in a labor market where homogenous firms post contracts that consist of a pay rate for each efficiency unit of labor and a specification of general training. Homogenous workers—both employed and unemployed—search for more valuable jobs, where the value of a job is the expected lifetime income for the worker if she accepts the job and behaves optimally afterwards. For a given job value, pay rate and the wage growth resulting from on-the-job training substitute each other. Therefore, firms offering more training can attract workers with a relatively lower pay rate than firms with less training.

\footnote{Fallick and Fleischman (2001), for instance, estimate that in the United States in 1999, on average 4 million workers changed employers from one month to the next (about 2.7 percent of employment)—more than twice the number who transited from employment to unemployment.}
In equilibrium, some firms offer general training and share the cost with their workers. The time-consuming nature of job search gives the firm monopsony power and compresses the worker’s pay rate below her marginal productivity. As a result, the firm and the worker share the total benefit from general training, hence they share the cost as well.

More importantly, although firms are ex ante homogenous, some but not all of them offer training in equilibrium. Moreover, firms with training are those offering higher-value jobs. The positive relationship between job value and training reflects the fact that high-value jobs tend to keep the worker longer, and the longer horizon of harvesting from general training motivates the firm to offer training. However, given that workers are more likely to quit low-value jobs very soon, it does not pay for low-value firms to offer training. In addition, given a higher value for each of her efficiency units, the worker has a stronger desire for more human capital and greater willingness to pay for training. A high-value firm utilizes this fact to deliver its promised job value in a cheaper way by offering training and a lower pay rate. By contrast, faced with a lower value per efficiency unit, the worker’s willingness to pay for training is weaker. Thus, it would be more costly for a lower value firm to deliver its promised value by offering training. Different wage-training offers from firms might lead to the case where a worker takes wage cut over transition from a job without training to one with training. These workers take wage cuts as a "tuition" for training in order to obtain higher wage growth in the future.

The plan of the rest of the paper is as follows: the next section reviews the relevant literature; section 3 lays out the basic framework; section 4 characterizes the optimal contract for the firm; section 5 discusses the model implication about wage dynamics; section 6 carries out the general equilibrium analysis, and a numerical simulation of the model; a brief discussion about efficiency of the market provision of training will be given in section 7; in section 8, I allow firms to choose the training intensity under a standard cost function and show that the main results still hold in this extension of the model; the final section concludes the paper.

2 Related Literature:

Technically, this paper builds on the on-the-job search literature a la Burdett and Mortensen (1998) (termed B-M hereafter). As stated in the introduction, job-to-job transition plays a key role in firm’s incentive to provide training. In B-M, job-to-job transitions are explicitly modeled and the strategic wage posting by firms directly affects job turnover. To capture the idea that workers value the long-run value of a job, not just the instantaneous wage payment, I extend B-M by allowing firms to post a training opportunity, as well as pay rate, to attract workers. Time-consuming job search serves as a self-enforcing source of job value dispersion: it gives firms monopsony power, which in turn motivates the worker’s search activity in hope of finding a higher-value job. Workers with the same human capital

---

2Since workers and firms are ex ante homogeneous and there is no match-specific value, the value of a job refers to the value for the worker, not the firm nor the match. This convention will be maintained throughout the paper when I mention high/low-value jobs or high/low-value firms.
level are paid differently across firms due to search frictions; and wage dispersion also exists within a firm due to human capital accumulation. In B-M, the density of wage distribution increases monotonically, which is in conflict with the data. By decomposing wage into pay rate and human capital level, this paper shows that the density of wage distribution has a declining right tail due to the declining density of human capital distribution.

Substantially, this paper contributes to the discussion of the firm’s willingness to offer general training. There have been studies on firm-sponsored training that are based on another source of market imperfection—private information. For example, Katz and Ziderman (1990) and Acemoglu and Pischke (1998) present models in which the firm may pay for general training because of its ex post informational monopsony power: the training firm possesses more information about the worker’s ability than outside firms. Such informational asymmetry between a training and a recruiting firm creates an adverse selection problem in the second-hand labor market, which, in turn, reduces the net benefits that a worker with general training can obtain by moving to another firm. As a result, the training firm makes greater profits from more skilled workers. In their model, either the worker has to quit her current job before she can look for another job; or, when poaching is allowed, the incumbent firm is allowed to make counteroffer to match the worker’s outside offer. This leads to a severe "winner’s curse" on the poaching firm and essentially shuts down job-to-job transitions, unless the worker experiences a negative shock on her current job. By contrast, job-to-job transitions in my model result from the worker’s desire and capability to move up along the job ladder. Another key element in their approach is the heterogeneity of workers: if workers are of the same ability, then the asymmetry of information, which provides the incentive for training, does not arise. Exploring search friction, my analysis shows that no heterogeneity nor any informational asymmetry is needed. Rather, search frictions alone are sufficient to generate the incentive for firms to provide training. Moreover, ex-ante homogeneous firms could have different choices with respect to training. Introducing heterogeneity would only enrich my current model without adding substantial insights.

In this paper, I assume that worker’s human capital grows as long as she stays on the training job, while it is usually assumed that training takes place only for an initial period as, for example, in Stevens (1994). To justify my assumption, it is important to realize that although the actual training program may last only for a short period of time, the accumulation of human capital is a continuous process on the job. Given the various problems a worker might face on the job, it is hard to imagine that a short training program would be able to teach her all. Rather, the role of the training program is to provide the worker with the ability-to-learn on the job, and human capital accumulation requires both the ability-to-learn and the learning environment provided by the firm. As long as the worker stays in the learning environment, she continues to improve her productivity after the actual training program ends. By contrast, on a non-training job, the worker cannot enhance her human capital beyond the level she previously obtained, because although the human capital accumulated can be carried across jobs, the learning environment cannot.

This paper also contributes to the literature on wage cuts over job-to-job transitions, which is ruled
out in standard search models such as B-M, but is not rare in the data, see for example, Postel-Vinay and Robin (2002). Workers may take wage cuts on transition in exchange for better job prospects in the future. In Postel-Vinay and Robin (2002), firms are heterogenous and respond to the offers received by their employees from poaching firms. In the Betrand competition between the incumbent firm and the poaching firm, the best offer the firm can make is its marginal productivity. If the marginal productivity of the poaching firm is high enough relative to the incumbent firm, the worker would be willing to accept a wage cut on transition in exchange of a greater potential best offer in the new firm. In my model, both firms and workers are homogenous, a worker may take wage cut as tuition for training, which leads to higher wage growth in the future. Although both papers model wage cuts as some investment for future wage growth, in Postel-Vinay and Robin (2002), the return to such investment realizes only when good luck strikes and the worker is poached by a highly productive firm. By contrast, in my model, the return to wage cut is guaranteed as long as the worker stays on the training job.

Relatively little work has been done combining on-the-job human capital accumulation and search to study wage dynamics. A recent paper by Bagger et al.(2006) allows for human capital accumulation in a job search framework with heterogenous agents. Workers accumulate human capital through learning-by-doing, which is an exogenous process, in that any job provides the same accumulation mechanism at zero cost. By contrast, in my paper, human capital accumulation is a costly process, and whether or not a job provides such opportunity results from the firm’s optimal wage-training contract strategy.

3 Basic Framework

Time is discrete and the economy is in steady state. There is a continuum of workers and firms, each of measure one. Both workers and firms are risk neutral.

Firms are homogenous in that, each period, any firm generates revenue $p$ for each efficiency unit it employs. The objective of a firm is to maximize steady-state flow profit.

Workers are also homogenous in that every worker is initially endowed with the same level of human capital (normalized to 1). Human capital is general and can be carried across all jobs and employment status without loss. For simplicity, I assume no depreciation of human capital, hence, the lowest human capital level of a worker is 1. The objective of a worker is to maximize expected life-time income.

While unemployed, a worker obtains a net benefit $b$ from unemployment for each efficiency unit which satisfies $p > b \geq 0$. Both unemployed and employed workers receive a new job offer with probability $\lambda$ each period. A job offer is composed of a pay rate for each efficiency unit ($\theta$) and a specification of training ($d$).\footnote{Given a linear production technology, pay rate per efficiency unit is equivalent to a piece rate wage.} In the basic model, I only consider the case where a firm offers either no training ($d = 0$), or training ($d = 1$) that increases the worker’s human capital linearly with net rate $g$, at cost $c$ for each unit of human capital. That is, should the firm offer training to a worker with human capital $h$, the current net profit is $(p - \theta - c)h$. In each period, a job is destroyed exogenously with probability
δ. All workers retire and leave the market for good with probability σ, each retired worker is replaced with a new unemployed worker with initial human capital level. Since there is worker retirement, for simplicity, I assume no discounting. Job destruction, job offer and retirement are mutually exclusive events, and (δ + λ + σ) < 1.

There are several basic assumptions on which the following analysis is based: first of all, since the pay rate is measured by efficiency unit, I assume a firm offers all new hires the same contract, regardless the worker’s former employment status and her human capital level, and the firm is committed to the contract. Secondly, I assume there is no recall should a worker quit or reject a job offer. Lastly I assume (1 − σ)(1 + g) < 1, which guarantees that the per efficiency unit value functions are bounded. 4

3.1 Worker Payoffs and Job Search Strategies

When a job offer arrives, each worker, unemployed or employed, takes as given the firm’s posted contract (θ, d) and computes her expected lifetime income conditional on accepting the job offer and using an optimal quit strategy in the future.

As long as she stays on a job with (θ, d = 0), the worker gets θ for each unit of her human capital and the level of her human capital stays constant, which corresponds to the first term in equation (1). The next period, with probability λ, the worker gets a new offer, upon which she chooses the best among staying with the current firm, transiting to the new firm and quitting into unemployment, which is the second term in (1). With probability δ, the worker is laid off, which is the third term in (1). If hit by the retirement shock, which occurs with probability σ, the worker leaves the market with a continuation value normalized, without loss of generality, to zero. And finally, if no shock of any sort occurs, the worker stays with the current firm as long as this is better than quitting into unemployment (the last term in (1)). In sum, the value of a job (θ, d = 0) for a worker with human capital h is:

\[
V(\theta, 0; h) = \theta h + \lambda E_{(\theta', d')} \max \{V(\theta, 0; h), V(\theta', d'; h), U(h)\} \\
+ \delta U(h) + (1 - \delta - \lambda - \sigma) \max \{V(\theta, 0; h), U(h)\}.
\]

4To see the sufficiency of this assumption for the boundedness of value function, consider a worker who works on a job with training and is paid her marginal productivity p for each of her efficiency units. The value of such a job will be the upper bound of her expected life-time value.

\[
\bar{V}(h) = ph + \delta U(h(1 + g)) + (1 - \sigma - \delta)\bar{V}(h(1 + g)) \\
< ph + (1 - \sigma)\bar{V}(h(1 + g)) \\
= ph + (1 - \sigma)(1 + g)V(h)
\]

where the second equality follows from the linearity of value function in h.

\[
\frac{\bar{V}(h)}{h} < \frac{p}{1 - (1 - \sigma)(1 + g)}
\]

Therefore, (1 − σ)(1 + g) < 1 guarantees that \(\frac{\bar{V}(h)}{h}\) is finite.
On a job with \((\theta, d = 1)\), the events that can happen to a worker are the same as when she is on a job with \((\theta, d = 0)\), except that her human capital grows at rate \((1 + g)\), hence, the value of a job \((\theta, d = 1)\) for a worker with human capital \(h\) is:

\[
V(\theta, 1; h) = \theta h + \lambda E_{(\theta', d')} \max\{V(\theta, 1; h(1 + g)), V(\theta', d'; h(1 + g)), U(h(1 + g))\}
+ \delta U(h(1 + g)) + (1 - \delta - \lambda - \sigma) \max\{V(\theta, 1; h(1 + g)), U(h(1 + g))\}.
\]

When unemployed, the worker gets \(b\) for each unit of her human capital and there is no change in the level of her human capital. With probability \(\lambda\), she gets a job offer, which will be accepted as long as its value is above unemployment value. Like employed workers, the unemployed also face the retirement risk. Hence the value of unemployment for a worker with human capital \(h\) is:

\[
U(h) = bh + \lambda E_{(\theta', d')} \max\{U(h), V(\theta', d'; h)\} + (1 - \lambda - \sigma)U(h).
\]

Due to the linearity of these value functions in \(h\), I can define the per efficiency unit value functions as \(v_u = U(h)/h\) and \(\tilde{v} = V(\theta, d; h)/h\).

When offered a job, the worker does not care about \(\theta\) and \(d\) separately; what matters to her is the value associated with this job offer, i.e. given the same expected \(\tilde{v}\), a worker is indifferent between a job with training and one without. Therefore, the important information for her decision is the distribution of \(\tilde{v}\) in the labor market. Denote as \(F(\tilde{v})\) the fraction of offers with per efficiency unit value no greater than \(\tilde{v}\). When making her job search decision, an employed worker compares the per-efficiency-unit value of her current job and that of the outside offer; and an unemployed worker compares the per-efficiency-unit value of the offer and that of unemployment \((v_u)\). For brevity, I will call these per-efficiency-unit values the values of a job and of unemployment, with the understanding that they are measured for each efficiency unit. These values can be expressed as follows:

\[
\tilde{v} = \theta + (1 + dg)\{\lambda \int_{\underline{v}}^{\tau} \max\{\tilde{v}, \tilde{v}', v_u\} dF(\tilde{v}') + \delta v_u + (1 - \delta - \lambda - \sigma) \max\{\tilde{v}, v_u\}\}, \text{ for } d = 0, 1. \tag{2}
\]

For unemployment:

\[
v_u = b + \lambda \int_{\underline{v}}^{\tau} \max\{v_u, \tilde{v}'\} dF(\tilde{v}') + (1 - \lambda - \sigma)v_u, \tag{3}
\]

where \(\tau\) and \(\underline{v}\) are the upper and lower bound for job values in equilibrium and will be specified later.

Optimal job search implies the following strategies for the worker:\(^5\)

1. when unemployed, the worker accepts a job offer if it has value \(\tilde{v} \geq v_u\);

2. when employed with contract \((\theta, d)\) which delivers \(\tilde{v}\), the worker quits if and only if a job offer is

\(^5\)I assume that in case of indifference, an unemployed worker accepts the job offer and an employed worker stays with the current employer.
received with value $\tilde{v}' > \tilde{v}$;  

3. if the job value falls below the value of unemployment and no better offer is received, the worker quits into unemployment.

Although I do allow the worker to quit her job into unemployment, it is obvious that this will never happen. Since the value is measured for each unit of human capital, and the per-efficiency-unit value of a job and that of unemployment are constant. As a result, if a job is acceptable at the first place, it is never optimal for a worker to quit into unemployment later.

Therefore, equation (2) can be simplified as:

$$\tilde{v} = \theta + (1 + dg)\lambda \int_{\tilde{v}}^{\tilde{v}'} \max\{\tilde{v}, \tilde{v}'\} dF(\tilde{v}') + \delta v_u + (1 - \delta - \lambda - \sigma)\tilde{v}, \text{ for } d = 0, 1. \quad (4)$$

As mentioned above, when offered a job with $(\theta, d)$, the worker will be able to compute the $\tilde{v}$ of this job as the basis for making her acceptance/rejection decision. Lemma 1 shows the other way, that is, given a specification of training, pay rate can be defined as a function of $\tilde{v}$, and I will denote these functions as $\theta_0(.)$ and $\theta_1(.)$.

**Lemma 1** Given $d$, $\theta_d(.)$ is strictly increasing in $\tilde{v}$.

**Proof.** See appendix. ■

Moreover, for any given $\tilde{v}$, a worker can always compute $\theta_0(\tilde{v})$ and its relationship with $\theta_1(\tilde{v})$, which is specified in the next lemma.

**Lemma 2** Given $\tilde{v}$, the worker demands $\theta_1(\tilde{v}) = \theta_0(\tilde{v}) - g(\tilde{v} - \theta_0(\tilde{v}))$ in order to be indifferent between a job that offers pay rate $\theta_0(\tilde{v})$ but no training and a job with pay rate $\theta_1(\tilde{v})$ and training, moreover, the gap between $\theta_0(\tilde{v})$ and $\theta_1(\tilde{v})$ is increasing in $\tilde{v}$.

**Proof.** See appendix. ■

The gap between $\theta_0(\tilde{v})$ and $\theta_1(\tilde{v})$, $g(\tilde{v} - \theta_0(\tilde{v}))$, represents the future gain for the worker from her increased human capital. Therefore, when offered a job with training, the worker is willing to pay the amount of the benefit she can get from her human capital accumulation on the job. The willingness to pay for training is greater when the worker is offered a higher $\tilde{v}$. Intuitively, the value per efficiency unit amounts to the "price" of the worker’s human capital. The worker makes investment in her human capital by accepting a lower pay rate, such compromise is worth making if she can sell her human capital at a higher price. Therefore, given a higher $\tilde{v}$, the worker has a stronger desire for more human capital and is willing to take a greater pay rate cut in order to get training.
3.2 Firm’s Problem:

Since workers only care about the value \( \tilde{v} \) of a job, a firm is free to choose any combination of \( \theta \) and \( d \) as long as it delivers the same \( \tilde{v} \) to the worker. Therefore, the competitiveness of a firm in the labor market depends only on the value \( \tilde{v} \) it promises, to the extent that the expected human capital level the firm can hire and the separation rate depend only on its promised \( \tilde{v} \) value. Let \( u \) be the steady state unemployment rate, \( E(h|u) \) be the average human capital level of unemployed workers, and let \( \Pr(\tilde{v}' < \tilde{v}, h) \) denote the measure of workers with human capital level \( h \) that are hired at jobs with value lower than \( \tilde{v} \), so that the joint distribution of \((\tilde{v}, h)\) among employed workers is \( \Pr(\tilde{v}' < \tilde{v}, h)/(1-u) \).

The expected human capital level that can be hired by a firm with \( \tilde{v} \) (denoted as \( l(\tilde{v}) \)) is:

\[
l(\tilde{v}) = \lambda[I(\tilde{v} \geq v_u)uE(h|u) + (1-u)\sum_h h \frac{\Pr(\tilde{v}' < \tilde{v}, h)}{1-u}],
\]

where \( I(\cdot) \) is an indicator function that equals 1 if the argument is true, and 0 otherwise. That is, with probability \( \lambda \) the firm meets a worker, it can attract an unemployed worker if the promised \( \tilde{v} \) is no less than \( v_u \), and the expected human capital level of this worker is \( E(h|u) \). Likewise, an employed worker would be attracted to the firm if she currently stays on a job with value less than \( \tilde{v} \), and the expected human capital level of this worker is \( \sum_h h \Pr(\tilde{v}' < \tilde{v}, h)/(1-u) \).

The worker leaves the firm when the job is destroyed, or when she retires, or if she receives an outside offer with value higher than \( \tilde{v} \) (recall that a worker never quits into unemployment), hence the separation rate for a firm offering \( \tilde{v} \) is:

\[
s(\tilde{v}) = \delta + \sigma + \lambda(1-F(\tilde{v})).
\]

The firm’s steady state flow profit equals the total efficiency units it hires, multiplied by the profit from each of these efficiency units. For a firm with job value \( \tilde{v} \), the profit is given as:

\[
\pi(\tilde{v}) = \max_{\theta,d}\left\{ l(\tilde{v}) \sum_{t=0}^{\infty} (1-s(\tilde{v}))^t (p - \theta - cd)(1+g)^{dt} \right\}
\]

\[
= \max_{\theta,d}\left\{ l(\tilde{v}) \frac{p - \theta - dc}{1-(1-s(\tilde{v}))(1+g)^d} \right\}
\]

\[
s.t. \tilde{v} \leq \theta + (1+dg)\left\{ \lambda \int_{\tilde{v}}^{\infty} \max\{\tilde{v}, \tilde{v}'\} dF(\tilde{v}') + \delta v_u + (1-\delta - \lambda - \sigma)\tilde{v} \right\}
\]

That is, if the firm does not provide training, it extracts \((p - \theta)\) from each efficiency unit it hires for as long as the worker stays in the firm. If the firm provides training, besides pay rate \( \theta \), it pays training cost \( c \) for each efficiency unit per period. In return, the efficiency units it hires grow at rate \((1+g)\), hence its profit also grows at the same rate. The constraint guarantees that the firm keeps its promise of delivering \( \tilde{v} \) per efficiency unit to the worker. Obviously, the promise keeping constraint always binds.
The firm’s problem can be decomposed into two steps: first, it chooses the value \( \hat{v} \) it will delivers to the worker; second, it chooses the most efficient combination of \( \theta \) and \( d \) to deliver the \( \hat{v} \) it has promised. As shown in Lemma 1, given \( d \), there is a unique pay rate \( \theta \) that, combined with \( d \), can deliver \( \hat{v} \). Therefore, the second stage problem boils down to the choice of \( d \). And the firm’s problem can be written as

\[
\pi = \max_{\hat{v}} \{ \pi(\hat{v}) \} \\
= \max_{\hat{v}} \max_d [\pi(d = 0; \hat{v}), \pi(d = 1; \hat{v})]
\]

where

\[
\pi(d = 0; \hat{v}) = \frac{(p - \theta_0(\hat{v}))}{s(\hat{v})} l(\hat{v})
\]

\[
\pi(d = 1; \hat{v}) = \frac{(p - \theta_1(\hat{v}) - c)}{1 - (1 - s(\hat{v}))(1 + g)} l(\hat{v}).
\]

The next section is devoted to the analysis of the firm’s optimal choice of wage-training contracts.

4 Optimal Wage-Training Contracts

Having promised \( \hat{v} \) to the worker, the firm is free to choose any pair of \((\theta, d)\) that maximizes its profit as long as the pair chosen can deliver \( \hat{v} \). Lemma 2 specifies the relationship between the pay rate under training and the pay rate under non training when they deliver the same \( \hat{v} \) for the worker. The next lemma characterizes the relationship between such pay rates that deliver the same profit for the firm.

**Lemma 3** Given the promised value \( \hat{v} \), with pay rate

\[
\theta_1'(\hat{v}) \equiv \theta_0(\hat{v}) - c + \frac{g(1 - s(\hat{v}))(p - \theta_0(\hat{v}))}{s(\hat{v})},
\]

a firm providing training earns equal profit as a firm offering wage \( \theta_0(\hat{v}) \) and no training.

**Proof.** If the firm chooses \( d = 0 \) and keeps its promise \( \hat{v} \), its profit is:

\[
\pi(d = 0; \hat{v}) = \frac{(p - \theta_0(\hat{v}))}{s(\hat{v})} l(\hat{v}).
\]

(7)

If, instead, the firm chooses \( d = 1 \), its profit is:

\[
\pi(\theta_1, d = 1; \hat{v}) = \frac{(p - \theta_1 - c)}{s(\hat{v})(1 + g) - g} l(\hat{v}).
\]

(8)
To find the $\theta_1^I(\hat{v})$ that earns the firm equal profit as it would get by posting $(\theta_0(\hat{v}), d = 0)$, I equalize the right hand sides of (7) and (8) to get:

$$\theta_1^I(\hat{v}) = \theta_0(\hat{v}) - c + \frac{g(1 - s(\hat{v}))(p - \theta_0(\hat{v}))}{s(\hat{v})}. \quad (9)$$

The last term on the right hand side of equation (9) represents the expected future gain for the firm from the increased human capital in production. Instead of fully internalizing the cost of training through cutting the pay rate by $c$, the firm is willing to bear the part of cost that is equal to the future benefit it can expect.

Since it is constrained to deliver $\hat{v}$ to the worker, the firm can decide on training provision by comparing the pay rate demanded by the worker, i.e. $\theta_1(\hat{v})$, and the pay rate necessary for equal profit, i.e. $\theta_1^I(\hat{v})$.

**Proposition 1 (Optimal wage-training contract)** The firm’s optimal training choice $d$, given the value $\hat{v}$ it has promised to its worker, is characterized by the following:

(i) if $c > B(\hat{v})$, firm with $\hat{v}$ chooses $d = 0$;

(ii) if $c < B(\hat{v})$, firm with $\hat{v}$ chooses $d = 1$;

(iii) if $c = B(\hat{v})$, firm with $\hat{v}$ chooses $d = 0$ or $d = 1$;

where

$$B(\hat{v}) \equiv g\left[\frac{(1 - s(\hat{v}))(p - \theta_0(\hat{v}))}{s(\hat{v})} + (\hat{v} - \theta_0(\hat{v}))\right]$$

is the worker-firm joint benefit from general training.

**Proof.** To keep its promise of $\hat{v}$, the firm has to give the worker $\theta_1(\hat{v})$ should it choose $d = 1$. If $\theta_1(\hat{v}) > \theta_1^I(\hat{v})$, when offered a job with $d = 1$, the worker demands more than the pay rate that keeps equal profit for the firm as it would get by offering no training, therefore, it is cheaper for the firm to deliver $\hat{v}$ with $(\theta_0(\hat{v}), d = 0)$ rather than with $(\theta_1(\hat{v}), d = 1)$. Similarly, if $\theta_1(\hat{v}) < \theta_1^I(\hat{v})$, it is cheaper for the firm to deliver $\hat{v}$ with $(\theta_1(\hat{v}), d = 1)$. When $\theta_1(\hat{v}) = \theta_1^I(\hat{v})$, the firm is indifferent between offering training and not offering it. The rest of the proof is obtained once I combine the expressions of $\theta_1(\hat{v})$ (from Lemma 2) and $\theta_1^I(\hat{v})$ (from equation (9)).

Proposition 1 says, given the promised $\hat{v}$ level, the choice of whether or not to offer training is based on the comparison between the cost of training and the worker-firm joint benefit from training. Moreover, this joint benefit is just the sum of the amounts that the two parties are willing to pay for the training. Search frictions enable the firm to pay a pay rate lower than the worker’s marginal
productivity, hence the firm and the worker share the rent from the accumulation of general human capital, as a result, they also share the cost.

The following lemma shows that the joint benefit is increasing in $\hat{v}$. The intuition behind this result is that, the worker is strictly better off with an increase in $\hat{v}$. As for firms, although a higher $\hat{v}$ may imply a higher pay rate, by promising a higher $\hat{v}$, the firm keeps the worker longer and hence extracting more from the worker; when the latter effect dominates the former effect, the firm’s benefit$^6$ is also increasing in $\hat{v}$.

**Lemma 4** The worker-firm joint benefit from increased human capital,

$$B(\hat{v}) \equiv g\left[\frac{(1 - s(\hat{v}))(p - \theta_0(\hat{v}))}{s(\hat{v})} + (\hat{v} - \theta_0(\hat{v}))\right],$$

is increasing in the promised $\hat{v}$ value.

**Proof.** See appendix. ■

Combining Lemma 4 with the fact that training cost $c$ is a fixed value, there will be three different situations with respect to the optimal training-provision decisions, depending on the value of $c$. Let $\tau^0$, $v_u^0$ denote the highest value and the lowest value offered in a market where no firm offers training, and let $v_u^1$ denote the lowest value offered in a market where all firms offer training.

Case 1: if

$$c \geq B(\tau^0) = g\left[\frac{(1 - \delta - \sigma)(p - \theta_0(\tau^0))}{\delta + \sigma} + (\tau^0 - \theta_0(\tau^0))\right] = g\left[\frac{(1 - \delta - \sigma)\bar{p} + \delta v_u^0}{\delta + \sigma}\right]$$

then, all firms will optimally choose not to offer training.$^7$

Case 2: if

$$c \leq B(v_u^1) = g\left[\frac{(1 - \delta - \sigma - \lambda)(p - \theta_0(v_u^1))}{\delta + \sigma + \lambda} + (v_u^1 - \theta_0(v_u^1))\right] = g\left[\frac{(1 - \delta - \sigma - \lambda)(p - b)}{\delta + \sigma + \lambda} + (v_u^1 - b)\right]$$

$^6$In equilibrium, the cost of training would offset the firm’s benefit from providing a higher $\hat{v}$, and therefore equalizes profits across firms with different $\hat{v}$’s.

$^7$The second equality follows from the fact that

$$\tau = \frac{\theta_0(\tau) + \delta v_u}{\delta + \sigma}.$$

The derivation of $\tau$ utilizes the fact that given the most generous offer, the worker will never quit.
then all firms will choose to offer training.\footnote{The second equality follows from the fact that \( \theta_0(v_u) = b \).}

Case 3: if \( B(v_u^1) < c < B(v_u^0) \), there exists\footnote{By Intermediate Value Theorem, the existence of the cutoff value \( \hat{v}^c \) follows from the continuity of \( B(\cdot) \) in \( \hat{v} \).} a cutoff value \( \hat{v}^c \) such that \( c = B(\hat{v}^c) \). Firms offering \( \hat{v} < \hat{v}^c \) will do so by offering \( \theta_0(\hat{v}) \) without training. Firms offering \( \hat{v} > \hat{v}^c \) will do so by offering \( \theta_1(\hat{v}) \) with training. Firms offering \( \hat{v}^c \) will do so by either offering \( \theta_0(\hat{v}^c) \) without training or offering \( \theta_1(\hat{v}^c) \) with training.

In case 1, consider a deviating firm when no other firms offer training. Lemma 4 shows that firms with higher \( \hat{v} \) is more likely to offer training, therefore, if the highest value firm finds the training cost too high to benefit from training, given the other firms’ decision, then, no firm will provide training. Case 2 describes the opposite situation: if the lowest value firm finds training profitable, then all firms will provide training. If neither of the two cases hold, and if parameter values are such that \( B(v_u^1) < B(v_u^0) \), there will be some \( c \) that induces the coexistence of training and non-training firms.\footnote{A numerical example will be given later to show that this case exists under reasonable parameter values.} In other words, in the case where the cost of training is not too low or too high, firms that offer higher \( \hat{v} \) also offer training. This results from the fact that by offering the worker a higher \( \hat{v} \), the firm is more likely to keep the worker for a longer time, given that the increase in human capital contributes to the firm’s total output, the firm is willing to make this investment. While a firm that offers low \( \hat{v} \) anticipates that it will lose the worker before it can extract enough from the worker to recover its share of training cost, therefore, it is better not to offer training. In addition, given a higher value for each of her efficiency units, the worker has a stronger desire for more human capital and greater willingness to pay for training. High-value firms utilize this fact to deliver their promised job values in a cheaper way by offering training and lower pay rate. By contrast, faced with a lower value per efficiency unit, the worker’s willingness to pay for training is weaker. Thus, it would be more costly for the lower value firm to deliver its promised value by offering training.

### 5 Wage Dynamics

What effects do the firms’ different wage-training offers have on the worker’s wage dynamics? To see this, let’s now assume that the cost of training is within the medium range, so that some but not all firms provide training.

First, on some jobs, there is no change in the worker’s wage over time because both the pay rate and the worker’s human capital level remain constant. While on other jobs, the worker’s wage grows smoothly over tenure, which results from the smooth growth of the worker’s human capital. Although in this model, workers are risk neutral, and the contract does not deliberately back-load the worker’s wage in order to reduce the worker’s incentive to quit, the observed wage still increases smoothly over tenure as a natural result of the growth of human capital.
The next two implications are illustrated in Figure 1. I assume toward the end of this section that parameter values are such that \( b < \theta_1(\tilde{v}^c) < \theta_0(\tilde{v}^c) < \theta_1(\overline{v}) \). I focus on this case since it is most interesting. Cases where the ranking of these pay rates is different can be analyzed in a similar way.

The second implication is that there might be wage cuts when the worker makes a voluntary job-to-job transition\(^{11}\). Despite wage cuts, the worker always moves to a better job. Wage cuts might happen when the worker transits from a job without training to one with training. Given the same \( \tilde{v} \) value, wage growth due to training and the pay rate substitute each other, therefore conditional on the same \( \tilde{v} \) value, pay rate on a job without training is always higher than that on a job with training. Although firms offering training also offer higher \( \tilde{v} \) values, wage cuts are still possible when the worker moves from a job without training to a job with training and both job values are close to the cutoff value \( \tilde{v}^c \).

The third finding is more surprising: it is possible for a worker to find a training job with higher pay rate than the pay rate on any non-training job. This happens if the value of the training job is sufficiently high and the training cost is relatively low, so that \( \theta_0(\tilde{v}^c) < \theta_1(\overline{v}) \). But given that it already bears the cost of training, why would a training firm be willing to accept an all-time lower profit ("smaller slice") per unit of human capital it employs? The answer lies in the fact that although the profit per unit of human capital is lower, the size of the human capital per employee in this firm is increasing over time. In other words, the firm and the worker shares a "growing pie" with training, while they share a "fixed-sized pie" without training. When the "growing pie" effect dominates the "smaller slice" effect, the firm would earn higher total profit out of a worker. The domination of the former effect follows from the fact that by offering a higher job value, the training firm keeps its worker longer, and this allows sufficient growth of the "pie". In addition, the training firm also "creates" more "pies" by attracting more new workers. As the argument gets further, the average pay rate on training jobs could be higher than that on non-training jobs. This finding yields a testable prediction: given the same level of human capital, between two unemployed workers, the one that finds a training job could, on average, earn more than the one that finds a non-training job, even at the beginning of tenure.

To be more specific, the following notations will be used:

Let \( \tilde{v}^d_0 \) be the value of a job without training such that \( \theta_0(\tilde{v}^d_0) = \theta_1(\tilde{v}^c) \), that is, \( \tilde{v}^d_0 \) is the lowest \( \tilde{v} \) level such that there is still a probability of wage cuts on transition. (Recall that \( \tilde{v}^c \) is the cutoff offer value above which a contract will provide training.) Let \( \tilde{v}^*_1(\tilde{v}_0) \) denote the value of a job with training such that the pay rate on this job is equal to the pay rate on the job without training that offers value \( \tilde{v}_0 \) i.e. \( \theta_0(\tilde{v}_0) = \theta_1(\tilde{v}^*_1(\tilde{v}_0)) \), let \( \tilde{v}^{**}_1(\tilde{v}_0) = \max\{\tilde{v}^*_1(\tilde{v}_0), \overline{v}\} \). Notice that the values \( \tilde{v}^d_0 \) and \( \tilde{v}^*_1(\tilde{v}_0) \) are well defined because of monotonicity of wage functions and the continuity of the job offer distribution \( F \), the latter will be confirmed in the next section.

**Proposition 2** For each worker who works on non-training jobs with value \( \tilde{v} \in [\tilde{v}^d_0, \tilde{v}^c] \), there is a

---

\(^{11}\)A cut in the observed wage over transition happens whenever there is a cut in the pay rate, since the worker’s human capital level remains the same at the time of job transition.
The chance of wage cuts on transition, and this probability is given by

\[ \Pr(\text{wage cuts on transition}) = \frac{\int_{\tilde{\nu}_0^c}^{\tilde{\nu}_1^*} dF(\tilde{\nu})}{\int_{\tilde{\nu}_0}^{\tilde{\nu}_1^*} dF(\tilde{\nu})} \quad \text{for } \tilde{\nu}_0 \in [\tilde{\nu}_0^l, \tilde{\nu}_0^c]. \]  

**Proof.** A worker hired at non-training job with value \( \tilde{\nu}_0 \in [\tilde{\nu}_0^l, \tilde{\nu}_0^c] \) will move to a new job if the new job is of value \( \tilde{\nu} \geq \tilde{\nu}_0 \), which happens with probability \( \lambda \int_{\tilde{\nu}_0}^{\tilde{\nu}_1} dF(\tilde{\nu}) \). Among these possible transitions, wage cuts happen when she transits to a training job with value between \( \tilde{\nu}_0^c \) and \( \tilde{\nu}_1^* \), which happens with probability \( \lambda \int_{\tilde{\nu}_0^c}^{\tilde{\nu}_1^*} dF(\tilde{\nu}) \). The ratio of these two probabilities gives the conditional probability of wage cuts given that a transition is observed. \( \blacksquare \)

At a glance, one might think the model suggests that workers with more human capital, typically older workers, are more likely to receive training, since training jobs are more valuable and older workers have sampled more jobs. This seems to be inconsistent with the observed career paths. However, as mentioned earlier, it is crucial to distinguish between an actual training program and the human capital accumulation on the job. A job with training would involve an actual training program at the beginning of the job but the worker continues to learn on the job in the learning environment provided by the firm. The typical career path for a worker, implied by the model, is: at the early stage of career, the worker takes low-value, short-tenure jobs without prospects; then she moves to jobs with training and relatively longer tenure, but she still changes jobs quite often and receives training programs on transitions; finally, she settles into a job where she stays for long and continuously improves herself. Therefore, instead of the older workers, it is younger workers who are more likely to be involved in training programs.
6 General Equilibrium Analysis

Continue the use of notation \( \Pr(\hat{v}' \leq \hat{v}, h) \) for the measure of workers with human capital \( h \) and employed at jobs with values no greater than \( \hat{v} \).

**Definition 1** A market equilibrium is:

1. a job offer distribution \( F \) of expected lifetime per-efficiency unit value \( \hat{v} \), such that:
   \[
   \pi^* = \pi(\hat{v}) = \max_d \left[ \pi(d = 0; \hat{v}), \pi(d = 1; \hat{v}) \right] \text{ for all } \hat{v} \in [v, \overline{v}]
   \]
   \[
   \pi^* \geq \pi(\hat{v}) \text{ otherwise.}
   \]

   i.e. any contract offered maximizes the firm’s profit and the maximized profit is equalized across optimizing firms;

2. a set of optimal wage-training contracts \((\theta_d(\hat{v}), d)\) that delivers \( \hat{v} \), according to equation (4);

3. optimal job search and quit strategies by the workers;

4. a steady state unemployment rate \( u \), a distribution of human capital among unemployed workers \( D_u(h) \), a joint distribution of job values and human capital among employed workers \( \frac{1}{1-u} \Pr(\hat{v}' \leq \hat{v}, h) \) that are consistent with steady state turnover, given \( F(.) \) and workers’ optimal strategy.

Given \( p > b \geq 0 \), in any equilibrium, firm’s profit will be strictly positive, and because all firms are ex-ante homogenous, they all earn equal profit in equilibrium.

**Lemma 5** In any equilibrium, firm’s profit is strictly positive.

**Proof.** By offering \( d = 0 \) and \( \theta = b \), a firm will be able to deliver the value \( v_u \) to an unemployed worker, who will accept the offer. Then the positive profit results directly from the assumption that \( p > b \).

Given this result, and that no worker will accept an offer lower than \( v_u \), in equilibrium, no firm will offer \( \hat{v} \) lower than \( v_u \). Moreover, the hiring rate and the separation rate for firm that offers the lowest \( \hat{v} \) are independent of the exact value of \( \hat{v} \) as long as it is no less than \( v_u \). Therefore, the lower bound for the equilibrium offer distribution \( \hat{v} \) will be \( v_u \). Moreover, given positive equilibrium profit, the maximum pay rate offered is less than \( p \), and therefore, the support for \( F(\hat{v}) \) is bounded above by some \( \overline{v} \), where

\[
\overline{v} < \frac{p + \delta(1 + g)}{(\delta + \sigma)(1 + g) - g}.
\]

\[(11)\]

**Lemma 6** Any equilibrium market distribution of job offers, represented by C.D.F. \( F(\hat{v}) \), is continuous, has a connected support, is bounded below by \( v_u \), and bounded from above by \( \overline{v} \) that satisfies (11).
Proof. See appendix. ■

The steady-state unemployment level is

\[ u = \frac{\delta + \sigma}{\delta + \lambda + \sigma}, \]

while the steady-state employment value distribution \( G(\hat{v}) \) can be derived from offer distribution \( F(\hat{v}) \) by equalizing the flow-in and flow-out of \( G(\hat{v}) \), which delivers the following:\(^{13}\)

\[ G(\hat{v}) = \frac{(\delta + \sigma) F(\hat{v})}{\delta + \sigma + \lambda(1 - F(\hat{v}))}. \]

Next, I derive the steady state human capital distribution among unemployed workers, \( D^u(\cdot) \), and the joint distribution of job values and human capital among employed workers \( \frac{1}{1 - u} \Pr(\hat{v}' \leq \hat{v}, h) \). The model is then closed as \( F(\hat{v}) \) must imply equal profit condition as specified in the definition of Market Equilibrium.

6.1 Human Capital Distribution

The next proposition characterizes the steady-state human capital distribution. The following notations will be used: \( F^c = F(\hat{v}^c) \) is the measure of the non-training sector; \( s^c = \sigma + \delta + \lambda(1 - F^c) \) is the separation rate for the non-training sector; \( D(h) \) is the steady state measure of all workers with human capital \( h \); and \( uD^u(h) \) is the steady state measure of unemployed workers with human capital \( h \).

**Proposition 3** In the steady state, the distribution of human capital is given by the following: for the lowest human capital level \( h=1 \):

\[ D(1) = \frac{\sigma(\sigma + \delta + \lambda)}{s^c(\sigma + \lambda) - \delta \lambda F^c} \]

\[ uD^u(1) = \frac{\sigma s^c}{s^c(\sigma + \lambda) - \delta \lambda F^c} \]  

For all \( n \geq 1 \),

\[ D[(1 + g)^n] = D(1) \frac{s^c\lambda(\sigma + \delta + \lambda)(1 - F^c)}{s^c(\sigma + \lambda) - \delta \lambda F^c} y^{n-1} \]

\[ uD^u[(1 + g)^n] = D(1) \frac{s^c\lambda(1 - F^c)}{s^c(\sigma + \lambda) - \delta \lambda F^c} y^{n-1} \]

where

\[ y = \frac{\delta \lambda(\sigma + \delta + \lambda)(1 - F^c)}{s^c(\sigma + \lambda) - \delta \lambda F^c} + 1 - \sigma - \delta. \]

---

\(^{12}\)See appendix for proof.  
\(^{13}\)See appendix for proof.
And for any \( h \notin \{ (1 + g)^n \}_{n=0}^\infty \)

\[ D(h) = 0. \]

Moreover, the mean human capital in the whole market in the steady state exists and is finite:

\[ E(h) = D(1) \{ 1 + \frac{s^c\lambda(\sigma + \delta + \lambda)(1 - F^c)(1 + g)}{s^c\sigma(\sigma + \delta + \lambda)(1 + g) - g[s^c(\sigma + \lambda) - \delta F^c]} \}, \]

and the mean human capital among unemployed workers is:

\[ E(h|u) = D^u(1) \{ 1 + \frac{\delta\lambda(\sigma + \delta + \lambda)(1 - F^c)(1 + g)}{s^c\sigma(\sigma + \delta + \lambda)(1 + g) - g[s^c(\sigma + \lambda) - \delta F^c]} \}. \]

**Proof.** See appendix.  

The measure of workers declines exponentially and converges to zero as the level of human capital increases. As long as the human capital growth rate is not too high relative to the retirement rate, the average human capital remains finite. Given the marginal distribution of pay rate and that of human capital, the following subsection characterizes the joint distribution of job value and human capital.

### 6.2 Joint Distribution of Job Values and Human Capital

**Proposition 4** The measure of workers with human capital \( h \) and employed at jobs with values no greater than \( v \) is given as:

**Case 1.** \( v < \hat{v}^c \)

\[ \Pr(\hat{v} \leq v, h = (1 + g)^n) = \frac{\lambda F^c(v)}{s(v)} uD^u[(1 + g)^n] \text{ for } n \geq 0, \]

where \( s(v) \) is the separation rate for firm that offers value \( v \), i.e.

\[ s(v) = \delta + \sigma + \lambda(1 - F(v)). \]

**Case 2.** \( v \geq \hat{v}^c \)

\[ \Pr(\hat{v} \leq v, h = 1) = \Pr(\hat{v} \leq \hat{v}^c, h = 1) = \frac{\lambda F^c}{s^c} uD^u(1); \]

for \( n \geq 1, \)

\[ \Pr(\hat{v} \leq v, h = (1 + g)^n) = \frac{\lambda F^c}{s^c} uD^u[(1 + g)^n] \]

\[ + \frac{\lambda(\sigma + \lambda + \delta)(F(v) - F^c)}{s^c} \sum_{m=1}^{n-1} (1 - s(v))^{m-1} uD^u[(1 + g)^{n-m}]. \]

**Proof.** See appendix.
Given the measure \( \Pr(\hat{v} \leq v, h = (1 + g)^n) \), one can easily obtain the joint distribution of job values and human capital among employed workers by dividing the measure \( \Pr(\hat{v} \leq v, h = (1 + g)^n) \) by the measure of all employees \((1 - u)\).

**Corollary 1** The distribution of job values \(v\) conditional on human capital level \(h\), \(\Pr(\hat{v} \leq v|h)\), is first order stochastically increasing in \(h\) for any \(v \geq \bar{v}^c\), and is invariant to \(h\) for \(v < \bar{v}^c\).

**Proof.** See appendix. ■

A worker with very high level of human capital must have been in the training sector for long. The longer she stays in the training sector, the higher her human capital level is, due to on-the-job training; and at the same time, the longer she stays in the training sector, the higher her job value is, due to on-the-job search. This constitutes the intuition behind the first part of Corollary 1. However, unemployed workers are the only inflow for non-training jobs, i.e. jobs with value lower than \(\bar{v}^c\). Moreover, layoff occurs with the same probability for workers regardless of their human capital level. Therefore, conditional on being employed in the non-training sector, the job value a worker obtains is not correlated with her human capital level.

### 6.3 Job Offer Distribution

I now turn to the equal profit condition to pin down job offer distribution. To calculate a firm’s profit, I first establish the average productivity of its employees.

**Claim 1** The average human capital level in a firm with value \(v\), \(l(v)\), is given by:

If \(v_u \leq v \leq \bar{v}^c\), then

\[
l(v) = \lambda \frac{\sigma + \lambda + \delta}{s(v)} uE(h|u). \tag{18}
\]

If \(\bar{v}^c < v \leq \bar{v}\), then

\[
l(v) = \lambda \frac{(\sigma + \lambda + \delta)(s^c(1 + g) - g)}{s^c[s(v)(1 + g) - g]} uE(h|u). \tag{19}
\]

**Proof.** See appendix. ■

Notice that, in the case where no firm provides training, \(\bar{v}^c \geq \bar{v}\), only (18) applies. While if all firms provide training, \(\bar{v}^c \leq v_u\), then only (19) applies. When we have both non-training firms and training firms, then, (18) applies to the former and (19) applies to the latter. It is easy to see that \(l(v)\) is increasing in \(v\). That is, higher value jobs can hire more efficiency units. Two forces drive this result: first, by offering a higher value \(v\), the firm attracts more workers, i.e. the hiring rate is higher. Secondly, among the workers that are hired at jobs with values lower than \(v\), the \(v\) firm can attract those hired at relatively higher values, and these workers are, on average, relatively more productive, as shown in Corollary 1. Therefore, the average human capital level hired by a firm increases with the value it promises. The next proposition characterizes the steady state job offer distribution.
**Proposition 5** In a market equilibrium, the steady state job offer distribution is given as follows:

if \( v_u \leq v \leq \bar{v}^c \)

\[
F(v) = \frac{\sigma + \lambda + \delta}{\lambda} \left[ 1 - \sqrt{\frac{p - \theta_0(v)}{p - b}} \right];
\]

(20)

if \( \bar{v}^c \leq v \leq \bar{v} \)

\[
F(v) = \frac{(\sigma + \lambda + \delta)(1 + g) - g}{\lambda(1 + g)} - \frac{s^c(1 + g) - g}{\lambda(1 + g)} \sqrt{\frac{p - \theta_1(v) - c}{p - \theta_1(\bar{v}^c) - c}}.
\]

(21)

**Proof.** See appendix. ■

When \( \bar{v}^c \geq \bar{v} \), no firm offers training, only (20) applies, and the distribution is the same as in B-M. If \( \bar{v}^c \leq v_u \), all firms provide training, only (21) applies. When there are both non-training firms and training firms, \( F(.) \) is specified separately for the two cases, however, as shown in the proof, the distribution is still continuous. The distribution given here involves endogenous variables \( \theta_1(\bar{v}^c) \), \( \theta_0(\bar{v}^c) \) and \( s^c \), but it can be expressed in primitives and is unique given parameter values, which implies the existence and uniqueness of the market equilibrium.

**Proposition 6** A market equilibrium exists and is unique.

**Proof.** See appendix. ■

Depending on parameter values, the market equilibrium could feature universal training, or no training at all, or training in some firms and no training in others, but for given parameter values, there exists a unique equilibrium. An analytical solution to the market equilibrium is long and tedious without being very informative, instead, a numerical solution will be given later.

Next, I lay out the distribution of observed wages, which depends on the joint distribution of pay rate and human capital.

### 6.4 Wage Distribution Among Employed Workers

Because conditional on training/non-training, pay rate is strictly increasing with job value, I can derive the joint and conditional distributions of pay rate and human capital from those of job value and human capital. Moreover, the properties of latter distributions also apply to the derived ones. In the next corollary, I focus on the most interesting case when parameter values are such that \( b < \theta_1(\bar{v}^c) < \theta_0(\bar{v}^c) < \theta_1(\bar{v}) \), cases where the ranking of these pay rates is different can be analyzed in a similar way.

**Corollary 2** The joint distribution of pay rate and human capital among employed workers is given by

Case 1. \( \theta \in [b, \theta_1(\bar{v}^c)) \)

\[
\frac{1}{1 - u} \Pr(\tilde{\theta} \leq \theta, h = (1 + g)^n) = \frac{1}{1 - u} \Pr(\tilde{\theta} \leq \theta_0^{-1}(\theta), h = (1 + g)^n)
\]

(22)
Case 2. \( \theta \in (\theta_0(\bar{v}^c), \theta_1(\bar{v})) \)

\[
\frac{1}{1-u} \Pr(\bar{\theta} \leq \theta, h = (1 + g)^n) = \frac{1}{1-u} \Pr(\bar{v} \leq \theta^{-1}(\theta), h = (1 + g)^n)
\]

Case 3. \( \theta \in [\theta_1(\bar{v}^c), \theta_0(\bar{v}^c)] \)

\[
\frac{1}{1-u} \Pr(\bar{\theta} \leq \theta, h = (1 + g)^n) = \frac{1}{1-u} \{ \Pr(\bar{v} \leq \theta_0^{-1}(\theta), h = (1 + g)^n) \\
+ \Pr(\bar{v} \leq \theta_0^{-1}(\theta), h = (1 + g)^n) - \Pr(\bar{v} \leq \bar{v}^c, h = (1 + g)^n) \}
\]

Moreover, the distribution of pay rate \( \theta \) conditional on human capital level \( h \), \( \Pr(\bar{\theta} \leq \theta|h) \), is first order stochastically increasing in \( h \) for any \( \theta \geq \theta_1(\bar{v}^c) \), and is invariant to \( h \) for \( \theta < \theta_1(\bar{v}^c) \).

**Proof.** Conditional on \( d \), \( \theta_d(v) \) function is strictly increasing in \( v \), hence \( \theta_d^{-1}(\theta) \) is well-defined for \( d = 0, 1 \). In case 1 and 2, there is only one type of firm in the market offering the pay rate \( \theta \), each pay rate corresponds to a unique job value. As a result, the \((\theta, h)\) distribution is the same as \((\bar{v}, h)\) distribution. In case 1, \( \theta \) is so low that only firms without training would offer such a pay rate. In case 2, \( \theta \in (\theta_0(\bar{v}^c), \theta_1(\bar{v})) \), \( \theta \) is offered only by firms with training. When \( \theta \in [\theta_1(\bar{v}^c), \theta_0(\bar{v}^c)] \), the same pay rate is offered by both types of firms. \( \Pr(\bar{\theta} \leq \theta, h) \), in this case, is composed of \( \Pr(\bar{v} \leq \theta_0^{-1}(\theta), h) \) from the non-training firms, and \( \Pr(\bar{v} \leq \theta_0^{-1}(\theta), h) - \Pr(\bar{v} \leq \bar{v}^c, h) \) from the training firms. Given the relationship between \((\theta, h)\) distribution and \((\bar{v}, h)\) distribution, one can see that \((\theta|h)\) distribution must preserve the first order stochastic dominance property of \((\bar{v}|h)\) when the pay rate \( \theta \) is paid by some training firms.

Thus, employed workers with more human capital typically earn higher pay rates. This is because these workers have been in the market for long, and hence have sampled more offers, which allows them to find jobs with higher pay rate. Given the joint distribution of pay rate and human capital, I now derive the distribution of the wages earned by employed workers, denoted as \( Q(w) \), where \( w = \theta h \). Let \( f(\theta, h) \) be the joint density of \((\theta, h)\) across employed workers.

\[
f(\theta, h) = \frac{\partial \Pr(\bar{\theta} \leq \theta, h)/(1-u)}{\partial \theta} \quad \text{for} \quad h \in \{(1+g)^n\}_{n=0}^{\infty} \quad \text{and} \quad \theta \in [\underline{\theta}, \bar{\theta}]
\]

where \( \underline{\theta} \) and \( \bar{\theta} \) are the lowest and highest pay rate paid to workers. Since \( w = \theta h \), by definition:

\[
Q(w) = \sum_{n=0}^{\infty} \int_{\underline{\theta}}^{\bar{\theta}} f(\theta, (1+g)^n) d\theta
\]
Differentiating with respect to $w$ yields the density of wages:

$$Q'(w) = \sum_{n=0}^{\infty} \frac{1}{(1+g)^n} f\left(\frac{w}{(1+g)^n}, (1+g)^n\right).$$

To derive the analytic solution for $Q$ is straightforward but tedious, however, two properties are immediate and insightful.

First consider the left tail of $Q'(w)$. If $w \in [\underline{\theta}, \underline{\theta}(1+g))$, the worker cannot have human capital higher than 1, otherwise, her wage must be at least as high as $\underline{\theta}(1+g)$. Therefore, $Q'(w) = f(w, 1)$, for $w \in [\underline{\theta}, \underline{\theta}(1+g))$. By differentiating the right hand side of equation (22) twice, one gets $Q''(w)$, which is positive. Therefore, the density of wages earned by employees is increasing when wage is sufficiently low.

Second, consider the right tail of $Q'(w)$. If $w$ becomes large, since pay rate is bounded above by $\bar{\theta}$, it must be that human capital level is large, i.e. $h \to \infty$ as $w \to \infty$. The conditional distribution of pay rate $\Pr(\tilde{v} \leq v|h)$ converges to $\Pr(\tilde{v} \leq v|\infty)$, moreover, since human capital distribution declines exponentially, the distribution of wages must decline at the same rate. Therefore, this model generate a wage distribution with a declining right tail, which is consistent with the data.
6.5 Simulation

This subsection carries out numerical simulation to further demonstrate the wage distribution derived from this model. Using a month as a time unit, I set \( \sigma = 1/480 \) such that an average worker is expected to work for 40 years, \( g = 0.00042 \) such that the annual growth of human capital is 0.5\%. Following Shimer (2005), I set the value of leisure net of search cost \( b = 0.4 \), job destruction rate \( \delta = 0.026 \), and job arrival rate \( \lambda = 0.45 \), which implies an unemployment rate of 5.87\%, consistent with OECD countries. Together with these parameter values, I set the cost of training \( c = 0.1862 \) such that training and non-training firms coexist with \( F(\bar{v}) = 0.2826 \), hence, there are more firms with training than firms without.

As shown in Figure 1, the pay rates are ranked as \( b < \theta_1(\bar{v}) < \theta_0(\bar{v}) < \theta_1(\bar{v}) \), in this case, only workers hired at medium pay rates may take wage cuts over job-to-job transition, while workers with very low or very high pay rates would never do so\(^{14}\). Moreover, the average pay rate on non-training jobs is 0.5524, while the average pay rate on training jobs is 0.7505. As a result, at a point in time, workers with the same human capital would, on average, earn more on a training job than they would on a non-training job. Figure 1 also shows that the pay rate is increasing linearly on non-training jobs as the value of the job goes up, however, the relationship is concave between the pay rate and the job value for training jobs. This reflects the idea that workers are more willing to pay for training if they are hired at high-value jobs, therefore, the increase in pay rates on training jobs slows down as the job value goes up.

Figure 2 shows the density of wage earned by employed workers. It features an increasing left tail and a decreasing long right tail, with an interior mode around the highest pay rate \( \theta_1(\bar{v}) \). The left tail increases as in B-M, but due to the exponentially declining distribution of human capital, the density of wages declines at the right tail. The interaction between search and human capital enables me to produce a more realistic wage distribution similar to what is observed in the data.

7 Efficiency of Training-Provision

Section 4 characterizes the optimal wage-training contracts offered by firms in the market, in particular, it is shown that the fraction of firms that offer training is endogenously determined. Next, I will compare this result with the social planner’s problem to see whether market provision of training is efficient.

Consider a social planner who also faces the search friction, and has to decide the fraction \( \alpha \) of jobs with training. In assuming that the social planner is also subject to search friction, I concentrate on the efficiency of training-provision. Let \( A \) be the social value of a unit of human capital given that it is used for production this period; and \( E \) be the social value of a unit of human capital given it is not

\(^{14}\)In particular, \( b = 0.4, \theta_1(\bar{v}) = 0.4908, \theta_0(\bar{v}) = 0.6767 \) and \( \theta_1(\bar{v}) = 0.8118 \).
used for production this period. The social planner’s problem is:

\[
A = \max_{0 \leq \alpha \leq 1} \left\{ (1 - \alpha)[p + \delta E + (1 - \delta - \sigma)A] + \alpha[p - c + \delta(1 + g)E + (1 - \delta - \sigma)(1 + g)A] \right\}
\]

where \( E = b + [\lambda A + (1 - \lambda - \sigma)E] \) \hspace{1cm} (24)

A job without training produces \( p \) today, and next period the job might be destroyed upon which the value becomes \( E \). Since \( p > b \) by assumption, the social planner would never choose to put the worker into unemployment, hence if no shock occurs, with probability \( (1 - \delta - \sigma) \), the value of one unit of human capital stays at \( A \). For a job with training, the cost of training \( c \) has to be deducted from the current output, but the continuation value is increased by a factor of \( (1 + g) \) as a result of growth in human capital. In case of a job destruction, the social value of a unit of human capital becomes \( E \): an unemployed unit of human capital produces \( b \) today, tomorrow if the worker gets matched with a firm, the value becomes \( A \), otherwise it stays at \( E \) provided the worker is still in the market. A unit of human capital, while unemployed, is valuable not only to the worker, but also it is valuable to the potential firms that might be matched with the worker in the future, hence the social value of an unemployed unit of human capital is greater than its value to the worker, i.e. \( E > v_u \).

Notice that the objective function is linear in \( \alpha \), which indicates a corner solution. Define

\[
A_0 = \frac{p + \delta E}{\delta + \sigma}, \quad A_1 = \frac{p - c + \delta(1 + g)E}{\delta + \sigma(1 + g) - g}.
\]

Case 1: if \( A_0 > A_1 \), then \( \alpha=0, \ A = A_0 \);
Case 2: if \( A_0 < A_1 \), then \( \alpha=0, \ A = A_1 \);
Case 3: if \( A_0 = A_1 \), then \( \alpha \in [0, 1], \ A = A_0 = A_1 \).

In terms of training cost, the social planner follows a cutoff cost strategy, where the cutoff cost is:

\[
c^* = g \left( \frac{1 - \delta - \sigma)p + \delta E}{\delta + \sigma} > g \left( \frac{1 - \delta - \sigma)p + \delta v_u}{\delta + \sigma} = B(\tau) \right.
\]

where the inequality follows from the fact that \( E > v_u \).

To be socially optimal, all firms should provide training if \( c < c^* \), no training should be provided if \( c > c^* \), and when \( c = c^* \), there will be no difference in social welfare whether there is training or not. Comparing this result with the market equilibrium result, one can see the efficiency of the market provision of training. Proposition 6 summarizes this finding.

**Proposition 7** In general, market-provided training is inefficiently low. Only when the cost of training goes above \( c^* \) defined in equation (25), will the market’s result coincide with the social planner’s choice,
which is not to provide training.

The inefficiency of training provision in the market equilibrium results from the externality of general training. When a firm chooses between offering and not offering training, it only considers the fact that increasing human capital will increase its own production and will provide a cheaper way to keep its promise of $\hat{v}$ to the worker (hence implicitly the firm takes $v_u$ into account). However, the firm does not consider that human capital, being general, can also contribute to the production of other firms should the worker leave this firm either through a direct job-to-job transition or through an indirect job transition via unemployment. Therefore, the social benefit from human capital accumulation is larger than the firm’s benefit, which leads to the inefficient market outcome. This inefficiency result falls into the general result that externality of training exists whenever there is uncertainty in labor turnover and imperfect competition that compresses wages below marginal product, see for example, Stevens (1994).

8 Extension: Endogenous Growth Rate

Up to now, I have focused on the case where firms can only choose between not providing training and providing training that increases human capital at a given rate $(1 + g)$. In this section, I relax this assumption and allow firms to choose the growth rate of its employed human capital. It is trivial that all firms will choose the highest possible growth rate if the cost does not increase with growth rate. Moreover, if the cost is linear in growth rate, one will go back to the dichotomy situation, where a firm will choose either zero training or the highest training level. That is because, with a linear cost function, growth rate $g$ does not enter the first order condition, the optimal choice of $g$ is always at one of the corners. To study a more interesting case, I make the following assumptions about the cost function:

Assumption: The per-efficiency-unit cost of training that increases human capital at rate $(1 + g)$ is represented by cost function $C(g)$, which satisfies 1) $C(0) = 0$, 2) $C'(.) > 0$, 3) $C''(.) > 0$ and 4) $\lim_{g \to \overline{g}} C'(g) = \infty$, where $\overline{g}$ is such that $(1 - \sigma)(1 + \overline{g}) = 1$.

Assumptions 1) to 3) define a standard cost function: Assumption 1) says the cost of no training is zero, assumptions 2) and 3) specify the cost function as an increasing, convex function of growth rate. Assumption 4) guarantees that no firm will choose a growth rate that is so high that the worker’s value functions might become unbounded.

Endogenizing the choice of $g$ has no effect on the firm’s optimal choice of $\hat{v}$: given $\hat{v}$, the competitiveness of the firm in the labor market is independent of the specific content of its contract. Therefore, I focus on the optimal wage-training contract problem for a firm that has already promised $\hat{v}$:

$$
\pi(\hat{v}) = \max_{g, \theta} \frac{p - \theta - C(g)}{s(\hat{v})(1 + g) - g} l(\hat{v})
$$

s.t.$$ \begin{align*}
\theta & \geq \theta_0(\hat{v}) - g(\hat{v} - \theta_0(\hat{v})) \\
g & \geq 0
\end{align*}
$$
where the first constraint is the promise keeping constraint, i.e. the right hand side of the constraint is the pay rate that the worker demands in order to be indifferent between a job without growth and one with growth rate \((1 + g)\). Since \(l(\tilde{v})\) is constant given \(\tilde{v}\), and since the promise keeping constraint is always binding, the maximization problem is equivalent to

\[
\max_{g \geq 0} \frac{p - \theta_0(\tilde{v}) + g(\tilde{v} - \theta_0(\tilde{v})) - C(g)}{s(\tilde{v})(1 + g) - g}
\]

**Proposition 8** When the choice of \(g\) is in the interior, \(\partial g / \partial \tilde{v} > 0\), that is, firms that offer higher \(\tilde{v}\) also offer higher growth rate.

**Proof.** See appendix. ■

Now, I will show the conditions under which the solution \(g\) is in the interior.

From the proof of Proposition 5, first order condition implies:

\[
\begin{aligned}
L(g; \tilde{v}) = s(\tilde{v})\tilde{v} + (1 - s(\tilde{v}))[p - C(g)] - \theta_0(\tilde{v}) - C'(g)[s(\tilde{v})(1 + g) - g]
\end{aligned}
\]

\[
\begin{aligned}
&\leq 0 \\
&= 0 \text{ if } g > 0
\end{aligned}
\]

Case (1) If \(L(g; v_u^1) = 0\), with \(g > 0\), then all firms will choose \(g > 0\) and \(g\) increases with \(\tilde{v}\); where \(v_u^1\) is the lowest job value offered in the market with all firms providing training.

Case (2) if \(L(0; \pi^0) \leq 0\), then no firm will offer training, where \(\pi^0\) is the highest job value in the market with no firm providing training.

Case (3) if neither of the above is true, then there will be a cutoff level \(\tilde{v}^c\), such that, firms that offer \(\tilde{v} \in [v_u, \tilde{v}^c)\) will not offer training, firms that offer \(\tilde{v} \in (\tilde{v}^c, \pi]\) will offer training and the growth rate will increase with \(\tilde{v}\); firms that offer \(\tilde{v} = \tilde{v}^c\) is indifferent between offering and not offering training.

In sum, when the firms are allowed to choose human capital growth rate under a convex cost function, the optimal growth rate \(g\) is non-decreasing in \(\tilde{v}\), and strictly increasing in \(\tilde{v}\) when \(g > 0\). This is consistent with the result from the basic model where the firm’s choice is restricted to be binary. Hence, endogenizing growth rate does not change the main findings from the basic model.

### 9 Conclusion

In labor economics, most of the studies on wage dynamics have been categorized into either search models or human capital models. This paper explores the potential of combining these two areas. Although only a tentative step toward a more unified approach, it is able to generate interesting new insights. In particular, the model studies the incentive for a firm to provide general training to its workers. It explains the coexistence of frequent job-to-job transitions and firm-sponsored general training.

Search frictions enable the firm to share the rent from the accumulation of general human capital. As a result, the firm and the worker share the cost of training. Although firms are ex ante homogenous, some but not all of them offer training to their employees. Moreover, firms that offer training are also firms that offer jobs valued higher by workers. As a result, workers may take wage cuts in order to transit...
from a non-training job to a training job. Compared to the socially optimal level, the market-provided training is inefficiently low because of externality of general human capital.

By decomposing wage into pay rate and human capital level, the model yields a wage distribution with a declining right tail, as is seen in the data. Assuming that workers are ex ante homogenous, the analysis focuses on the dispersion of human capital simply due to on-the-job training. Although a lot of studies have been done on wage dispersion, an important question that remains unanswered is the following: how much of the observed wage dispersion is due to pure pay rate dispersion and how much is due to human capital dispersion? This paper provides some building blocks for future studies on this question.

References


Appendix
A1. Proof for Lemma 1

Proof. Case (1): \( d = 0 \)

\[
\hat{\nu} = \theta_0 + \lambda \int_{\hat{\nu}}^{\nu} \hat{\nu}' dF(\hat{\nu}') + \lambda F(\hat{\nu})\hat{\nu} + \delta \nu_u + (1 - \delta - \lambda - \sigma)\hat{\nu}
\]

\[
= \theta_0 + (1 - \delta - \lambda - \sigma)\hat{\nu} + \lambda \nu + \delta \nu_u - \lambda \int_{\hat{\nu}}^{\nu} F(\hat{\nu}')d\hat{\nu}',
\]

where the second equality follows from integration by part. Rearrange terms to get

\[
\theta_0 = (\delta + \lambda + \sigma)\hat{\nu} - \lambda \nu - \delta \nu_u + \lambda \int_{\hat{\nu}}^{\nu} F(\hat{\nu}')d\hat{\nu}'.
\]

Take the derivative with respect to \( \hat{\nu} \)

\[
\frac{d\theta_0}{d\hat{\nu}} = (\delta + \lambda + \sigma) - \lambda F(\hat{\nu})
\]

\[
= s(\hat{\nu}) > 0,
\]

27
i.e. \( \theta_0 \) is strictly monotone in \( \hat{v} \), hence the mapping from \( \hat{v} \) to \( \theta_0 \) is one-to-one. Therefore, \( \theta_0(\hat{v}) \) is a well defined function.

Case(2) \( d = 1 \): similar arguments will yield the result that \( \theta_1(\hat{v}) \) is a well-defined function.

\[ \text{A2 Proof for Lemma 2} \]
(The relationship between \( \theta_1(\hat{v}) \) and \( \theta_0(\hat{v}) \), from the worker’s perspective)

**Proof.** For a job with \((\theta_0, d = 0)\)

\[
\hat{v} = \theta_0(\hat{v}) + \lambda \int_{\Sigma} \max\{\hat{v}, \hat{v}'\} dF(\hat{v}') + \delta v_u + (1 - \delta - \lambda - \sigma)\hat{v} \\
= \theta_0(\hat{v}) + \hat{v} - \theta_0(\hat{v}).
\]

For a job with \((\theta_1, d = 1)\)

\[
\hat{v} = \theta_1 + (1 + g)\{\lambda \int_{\Sigma} \max[\hat{v}, \hat{v}'] dF(\hat{v}') + \delta v_u + (1 - \delta - \lambda - \sigma)\hat{v}\} \\
= \theta_1 + (1 + g) (\hat{v} - \theta_0(\hat{v})).
\]

Equating these two expressions to solve for \( \theta_1 \), one gets:

\[
\theta_1(\hat{v}) = \theta_0(\hat{v}) - g(\hat{v} - \theta_0(\hat{v})).
\]

Now, take the derivative of the wage gap \( \text{Gap}(\hat{v}) = \theta_0(\hat{v}) - \theta_1(\hat{v}) = g(\hat{v} - \theta_0(\hat{v})) \) with respect to \( \hat{v} \)

\[
\text{Gap}'(\hat{v}) = g(1 - \theta_0'(\hat{v})) = g[1 - s(\hat{v})] > 0.
\]

\[ \text{A3. Proof for Lemma 4} \]
(The joint benefit from general training is increasing in \( \hat{v} \))

**Proof.** Rearrange terms

\[
B(\hat{v}) = \frac{g}{s(\hat{v})} \{(1 - s(\hat{v}))(p - \theta_0(\hat{v})) + s(\hat{v})(\hat{v} - \theta_0(\hat{v}))\} \\
= \frac{g}{s(\hat{v})} \{s(\hat{v})\hat{v} - \theta_0(\hat{v}) + (1 - s(\hat{v}))p\} \\
= g\hat{v} + \frac{g(1 - s(\hat{v}))p - \theta_0(\hat{v})}{s(\hat{v})}
\]

28
Take derivative

\[ B'(\bar{v}) = g + \frac{g}{[s(\bar{v})]^2} \{ (-ps'(\bar{v}) - \theta'_0(\bar{v}))s(\bar{v}) - [(1 - s(\bar{v}))p - \theta_0(\bar{v})]s'(\bar{v}) \} \]

\[ = g + \frac{g}{[s(\bar{v})]^2} \{ -(p - \theta_0(\bar{v}))s'(\bar{v}) - s(\bar{v})\theta'_0(\bar{v}) \} \]

\[ = \frac{-g}{[s(\bar{v})]^2} [p - \theta_0(\bar{v})]s'(\bar{v}) \]

\[ > 0 \]

where I use the following facts: \( \theta'_0(\bar{v}) = s(\bar{v}), s'(\bar{v}) = -\lambda F'(\bar{v}) < 0 \) and \( p - \theta_0(\bar{v}) > 0 \). Hence, the joint benefit is strictly increasing in \( \bar{v} \).  

A4. Proof for Lemma 6  
(Property of F distribution)

**Proof.** Step 1: The support of F is bounded below \( v_u \), and has upper bound \( \bar{v} \) that satisfies

\[ \bar{v} < \frac{p + \delta(1 + g)v_u}{(\delta + \sigma)(1 + g) - g} \]  

(26)

The reason for lower bound to be \( v_u \) has been shown in the main body of the paper. Here, I will just establish that the upper bound \( \bar{v} \) must satisfy (26). Consider a worker who is on a job with training and is paid her productivity for each of her efficiency unit. Her life-time value per-efficiency unit is given by

\[ \bar{\bar{v}} = p + (1 - \delta - \sigma)(1 + g)\bar{v} + \delta(1 + g)v_u \]

\[ \bar{\bar{v}} = \frac{p + \delta(1 + g)v_u}{(\delta + \sigma)(1 + g) - g} \]

Due to strict positive equilibrium profit, the upper bound of wage must be strictly less than \( p \), therefore:

\[ \bar{v} < \bar{\bar{v}} \]

Step 2: \( F(\bar{v}) \) has no mass point.

Suppose not, there exists \( \bar{v}_0 \in [v_u, \bar{v}] \), such that \( F(\bar{v}_0) > F^{-}(\bar{v}_0) \), this would imply that \( \Pr(\bar{v}' \leq \bar{v}_0, h) \geq \)
Pr(\(\hat{v}' < \hat{v}_0, h\)) for all \(h\), and with strict inequality for some \(h\).

\[
\lim_{\epsilon \to 0^+} \pi(d, \hat{v}_0 + \epsilon) = \lim_{\epsilon \to 0^+} \frac{p - \theta_d(\hat{v}_0 + \epsilon) - dc}{1 - (1 - \delta - \sigma - \lambda(1 - F(\hat{v}_0 + \epsilon))(1 + dg))} \lambda[uE(h|u) + \sum_h h \Pr(\hat{v}' < \hat{v}_0 + \epsilon, h)]
\]

Step 3: \(F(\hat{v})\) has connected support.
Suppose not. There is a gap between \((\hat{v}', \hat{v}'')\) in the support of \(F(\hat{v})\), where \(\hat{v}'\) and \(\hat{v}''\) are in the support. By strict monotonicity of wage function, for any \(\hat{v} \in (\hat{v}', \hat{v}'')\), \(\theta_d(\hat{v}') < \theta_d(\hat{v}'')\) for \(d = 0, 1\). Moreover, since \(F(\hat{v}') = F(\hat{v}'')\), and hence, \(\Pr(\hat{v} < \hat{v}', h) = \Pr(\hat{v} < \hat{v}'', h)\) for all \(h\), then:

\[
\pi(d, \hat{v}') = \frac{p - \theta_d(\hat{v}') - dc}{1 - (1 - \delta - \sigma - \lambda(1 - F(\hat{v}')))(1 + dg)} \lambda[uE(h|u) + \sum_h h \Pr(\hat{v} < \hat{v}', h)]
\]

\[
> \frac{p - \theta_d(\hat{v}'') - dc}{1 - (1 - \delta - \sigma - \lambda(1 - F(\hat{v}'')))(1 + dg)} \lambda[uE(h|u) + \sum_h h \Pr(\hat{v} < \hat{v}'', h)]
\]

\[
= \pi(d, \hat{v}''),
\]

a contradiction. ■

A5. Derivation of stationary unemployment level and \(G(\hat{v})\) distribution

**Proof.** The flow-in of unemployment comes only from exogenous job destruction, due to the fact that workers never quit into unemployment in a stationary environment. The flow-out of unemployment consists of workers that get a job offer, which will always be accepted in equilibrium:

\[
u \lambda = (1 - u) \delta \\
u = \frac{\delta + \sigma}{\lambda + \delta + \sigma}.
\]
The flow-in of workers hired at value no greater than \( \hat{v} \) comes only from the unemployed, while workers flow out of \( G(\hat{v}) \) because of job destruction, retirement, or the arrival of a job offer from \((1 - F(\hat{v}))\)

\[
\begin{align*}
\lambda u F(\hat{v}) &= (1 - u)G(\hat{v})[\delta + \sigma + \lambda(1 - F(\hat{v}))], \\
G(\hat{v}) &= \frac{\lambda u F(\hat{v})}{(1 - u)[\delta + \sigma + \lambda(1 - F(\hat{v}))]}, \\
&= \frac{(\delta + \sigma)F(\hat{v})}{[\delta + \sigma + \lambda(1 - F(\hat{v}))]}.
\end{align*}
\]

\( \blacksquare \)

A6 Proof for proposition 3
(Human Capital Distribution)

**Proof.** Denote \( D^i(h) \) as the steady-state distribution of workers with human capital \( h \) within the sector \( i \), where \( i = u \) (unemployed), \( 0 \) (employed without training), \( 1 \) (employed with training). Let \((h, i)\) represents the status of a worker who has human capital \( h \) and is in sector \( i \). Denote \((1 - u)G^c\) as the measure of workers in the nontraining sector, i.e. \((1 - u)G^c = \sum_h \Pr(v \leq \hat{v}^c, h)\), and denote \((1 - u)(1 - G^c)\) as the measure of workers in the training sector. Since time is discrete and a worker with \((h, 1)\) at the beginning of a period becomes \((h(1 + g), 1)\) at the end of the period, if she is still employed in the training sector. Without loss of generality, I will characterize the end-of-period human capital distribution, beginning-of-period distribution can also be derived in a similar way.

Unemployment sector: when \( h = 1 \), because I am characterizing end-of-period distribution, the human capital level of workers in the training sector is at least \((1 + g)\), the inflow of \((1, u)\) is composed only of workers who are laid off from the non-training sector with human capital \( 1 \) and the new entrants, while the outflow consists of workers that either retire or find a job. Equating outflow with inflow, I get

\[
(\lambda + \sigma)uD^u(1) = \sigma + \delta(1 - u)G^cD^0(1).
\]

(27)

For \( h \in \{(1 + g)^n\}_{n=1}^{\infty} \), the inflow of \((h, u)\) consists of workers that are laid off from either employment sector with human capital \( h \), while the outflow is the same as before, hence I have:

\[
(\lambda + \sigma)uD^u(h) = (1 - u)G^cD^0(h)\delta + (1 - u)(1 - G^c)D^1(h)\delta.
\]

(28)

Employment sector without training: for all \( h \), workers with \((h, 0)\) leave this group if they find a job in the \( d = 1 \) sector, or if they leave the market or if they are laid off, hence separation probability is \( s^c = \sigma + \delta + \lambda(1 - F^c) \). Since workers in sector \( d = 1 \) will never go directly down to sector \( d = 0 \) (recall training job is more valuable than non training job), only unemployed workers will join this group if
they find a job in this sector.

\[ s^c(1 - u)g^cD^0(h) = \lambda F^c uD^u(h) \]  \hspace{1cm} (29)

Employment sector with training:

\[ (1 - u)(1 - g^c)D^1(1) = 0. \]

For \( h \in \{(1 + g)^n\}_{n=1}^{\infty} \), workers in sector \( d = 1 \) with \( h \) will leave this group for sure regardless of whether they stay or leave this sector, (if they stay, their human capital becomes \( h(1 + g) \)). Those who were in \( d = 1 \) with \( \frac{h}{1 + g} \) moves into \((h, 1)\) group as long as they stay in the training sector. Workers who were unemployed or employed in non-training sector with human capital \( \frac{h}{1 + g} \) will join this \((h, 1)\) group if they find a job in the training sector.

\[ (1 - u)(1 - g^c)D^1(h) = (1 - u)(1 - g^c)D^1\left(\frac{h}{1 + g}\right)(1 - \sigma - \delta) + uD^u\left(\frac{h}{1 + g}\right)\lambda(1 - F^c) + (1 - u)g^cD^0\left(\frac{h}{1 + g}\right)\lambda(1 - F^c). \]  \hspace{1cm} (30)

In the whole economy:

\[ D(1) = uD^u(1) + (1 - u)g^cD^0(1), \]  \hspace{1cm} (31)

and for \( h \in \{(1 + g)^n\}_{n=1}^{\infty} \),

\[ D(h) = uD^u(h) + (1 - u)g^cD^0(h) + (1 - u)(1 - g^c)D^1(h). \]  \hspace{1cm} (32)

The relationships between the measure of workers with human capital \( h \) in the unemployment sector, in the non-training sector and in the training sector, i.e. \( uD^u(h), (1 - u)g^cD^0(h) \) and \( (1 - u)(1 - g^c)D^1(h) \), would be useful in deriving the results. These relationships can be shown to be as follows: for \( h \in \{(1 + g)^n\}_{n=1}^{\infty} \),

\[ uD^u(h) = \frac{\delta s^c}{s^c(\lambda + \sigma) - \lambda \delta F^c} (1 - u)(1 - g^c)D^1(h), \]

\[ (1 - u)g^cD^0(h) = \frac{\lambda \delta F^c}{s^c(\lambda + \sigma) - \lambda \delta F^c} (1 - u)(1 - g^c)D^1(h). \]

And for \( h = 1 \)

\[ uD^u(1) = \frac{\sigma s^c}{s^c(\lambda + \sigma) - \lambda \delta F^c}; \]

\[ (1 - u)g^cD^0(1) = \frac{\lambda F^c}{s^c}uD^u(1). \]
Solving the equations (27) to (32) gives us the distribution as specified in the proposition. One can check that this is indeed a distribution because $\forall h \in \{(1 + g)^n\}_{n=0}^\infty$, $D(h) \in (0, 1)$ and $\sum_{n=0}^\infty D[(1 + g)^n] = 1$. In particular, $\lim_{n \to \infty} D[(1 + g)^n] = 0$ since $y \in (0, 1)$.

The mean of human capital is

$$E(h) = \sum_{n=0}^\infty (1 + g)^n D[(1 + g)^n]$$

$$= \sum_{n=1}^\infty (1 + g)^n \frac{s^c \lambda (\sigma + \delta + \lambda)(1 - F^c)}{s^c(\sigma + \lambda) - \delta \lambda F^c} y^{n-1} D(1) + D(1)$$

$$= D(1) \{1 + (1 + g) \frac{s^c \lambda (\sigma + \delta + \lambda)(1 - F^c)}{s^c(\sigma + \lambda) - \delta \lambda F^c} \sum_{n=1}^\infty [y(1 + g)]^{n-1}\}$$

The assumption that $(1 + g)(1 - \sigma) < 1$ guarantees $y(1 + g) \in (0, 1)$, and therefore the expectation is finite. Using the relationship between $uD(\cdot)$ and $D(\cdot)$, one can get the expression of the average human capital among unemployed workers.\(^\text{15}\)

**A7 Proof for Proposition 4**

$$(Pr(\hat{\nu} \leq v, h))$$

**Proof.** Case 1. $v < \hat{\nu}^c$: In steady state, the inflow for $Pr(\hat{\nu} \leq v, [(1 + g)^n])$ comes only from the unemployed who have human capital $(1 + g)^n$ and find a job with value lower than $v$. i.e. $\lambda F(v) uD^n[(1 + g)^n]$. Workers of this group flow out due to layoff, retirement or finding a better job, i.e. $Pr(\hat{\nu} \leq v, h) s(v)$. Equalizing inflow with outflow, and utilizing the relationship between $uD^n(h)$ and $D(h)$ gives the result.

Case 2 $v \geq \hat{\nu}^c$: $Pr(\hat{\nu} \leq v, [(1 + g)^n]) = Pr(\hat{\nu} \leq \hat{\nu}^c, [(1 + g)^n]) + Pr(\hat{\nu}^c \leq \hat{\nu} \leq v, [(1 + g)^n])$. Notice that the first term is the measure of workers with human capital $(1 + g)^n$ in the non-training sector, i.e. $(1 - u) G^c D^0[(1 + g)^n]$. The inflow for $Pr(\hat{\nu}^c \leq \hat{\nu} \leq v, [(1 + g)^n])$ comes from workers, unemployed or employed at lower value jobs, who have human capital $(1 + g)^{n-1}$ last period and find a job with $\hat{\nu} \in [\hat{\nu}^c, v]$. Moreover, as long as they still stay in jobs within this range, the workers who had human capital $(1 + g)^{n-1}$ last period would also join this inflow. The outflow is the whole $Pr(\hat{\nu}^c \leq \hat{\nu} \leq v, (1 + g)^n)$, because workers with $(\hat{\nu}^c \leq \hat{\nu} \leq v, (1 + g)^n)$ would either retire, or get laid off, or get a job better than

\(^{15}\)More detailed proof is available from the author on request.
or if they stay in \((\tilde{v}^c \leq \tilde{v} \leq v)\), they would have human capital \((1 + g)^{n+1}\). Therefore,

\[
Pr(\tilde{v}^c \leq \tilde{v} \leq v, [(1 + g)^n]) = \lambda(F(v) - F^c)\{uD^u[(1 + g)^{n-1}] + (1 - u)G^cD^0[(1 + g)^{n-1}]\} + (1 - s(v)) \Pr(\tilde{v} \leq \tilde{v} \leq v, [(1 + g)^{n-1}] \\
= \lambda(F(v) - F^c) \sum_{m=1}^{n} (1 - s(v))^{m-1}\{uD^u[(1 + g)^{n-m}] + (1 - u)G^cD^0[(1 + g)^{n-m}]\} \\
= \frac{\lambda(F(v) - F^c)(\sigma + \lambda + \delta)}{s^c} \sum_{m=1}^{n-1} (1 - s(v))^{m-1}uD^u[(1 + g)^{n-m}],
\]

where the last equality follows from the relationship between \(uD^u(h)\) and \((1 - u)G^cD^0(h)\).

For \(n = 0\), since workers in the training sector have human capital at least as high as \((1 + g)\) at the end of any period, I have

\[
Pr(\tilde{v} \leq v, h = 1) = \Pr(\tilde{v} \leq \tilde{v}^c, h = 1) = \frac{\lambda F^c}{s^c}uD^u(1).
\]

\[\Box\]

A8 Proof for corollary 1

(Conditional distribution of \(\tilde{v}|h\))

**Proof. Part I.** I first derive the conditional distribution of \(v|h\). The conditional distribution of \(v|h\) is the measure of workers with \(h\) and employed with job values no greater than \(v\), divided by the measure of employed workers with human capital \(h\), and the latter is the measure of workers with \(h\) minus the measure of unemployed workers with \(h\).

\[
Pr(\tilde{v} \leq v|h = (1 + g)^n) = \frac{Pr(\tilde{v} \leq v, h = (1 + g)^n)}{[D((1 + g)^n) - uD^u((1 + g)^n)]}
\]

Case 1. \(v < \tilde{v}^c\):

\[
Pr(\tilde{v} \leq v|h = 1) = \frac{\lambda F(v)}{s(v)} \frac{uD^u(1)}{D(1) - uD^u(1)} = \frac{F(v)s^c}{s(v)F^c}
\]

for \(n \geq 1\)

\[
Pr(\tilde{v} \leq v|h = (1 + g)^n) = \frac{\lambda F(v)}{s(v)} \frac{uD^u((1 + g)^n)}{[D((1 + g)^n) - uD^u((1 + g)^n)]} = \frac{\delta}{s(v) (\sigma + \lambda)}
\]
where the second equality follows from the relationship between \( uD^u() \) and \( D() \). Notice that in this case, the conditional distribution is invariant to \( h \).

**Case 2.** \( v \geq \hat{v}^c \)

\[
Pr(\hat{v} \leq v | h = 1) = \frac{Pr(\hat{v} \leq \hat{v}^c, [(1 + g)^n])}{[D((1 + g)^n) - uD^u((1 + g)^n)]} = 1
\]

If an employed worker has human capital \( 1 \), she must be employed in the non-training sector, hence \( \hat{v} \leq \hat{v}^c \leq v \) for sure.

For \( n \geq 1 \), I have shown earlier that

\[
Pr(\hat{v} \leq v, h = (1 + g)^n) = \frac{\lambda F^c}{s^c} uD^u((1 + g)^n)
\]

\[
+ \lambda(F(v) - F^c) \sum_{m=1}^{n} (1 - s(v))^{m-1} \frac{\sigma + \lambda + \delta}{s^c} uD^u((1 + g)^{n-m}).
\]

Using the expression for \( uD^u((1 + g)^n) \), for \( n - m = 0 \),

\[
(\sigma + \lambda + \delta) uD^u(1) = D(1)
\]

and for \( n - m \geq 1 \),

\[
(\sigma + \lambda + \delta) uD^u((1 + g)^{n-m}) = \frac{\lambda D(1) \delta(\sigma + \lambda + \delta)(1 - F^c)}{s^c(\lambda + \sigma) - \lambda \delta F^c} y^{n-m-1},
\]

where

\[
y = 1 - \sigma - \delta + \frac{\delta(\lambda + \sigma + \delta)(1 - F^c)}{s^c(\lambda + \sigma) - \lambda \delta F^c}.
\]

Notice that \( y > 1 - \sigma - \delta > 1 - s(v) \). Plug the expressions for \( \frac{\sigma + \lambda + \delta}{s^c} uD^u((1 + g)^{n-m}) \) into (33), I have

\[
Pr(\hat{v} \leq v, h = (1 + g)^n) = \frac{\lambda F^c}{s^c} uD^u((1 + g)^n)
\]

\[
+ \lambda(F(v) - F^c) D(1) \left\{ \frac{\lambda D(1) \delta(\sigma + \lambda + \delta)(1 - F^c)}{s^c(\lambda + \sigma) - \lambda \delta F^c} \sum_{m=1}^{n-1} \left( \frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\}.
\]

Notice that for \( n \geq 1 \)

\[
[D((1 + g)^n) - uD^u((1 + g)^n)] = \frac{\lambda D(1)s^c(\lambda + \sigma)(1 - F^c)}{s^c(\lambda + \sigma) - \lambda \delta F^c} y^{n-1},
\]
hence,

\[
\Pr(\hat{v} \leq v|h = (1 + g)^n) = \frac{\lambda F^c \delta}{s^c(\sigma + \lambda)} + \frac{\lambda \delta (\lambda + \delta + \sigma)(F(v) - F^c)}{s^c(\lambda + \sigma)(y + s(v) - 1)} + \frac{(F(v) - F^c)[s^c(\lambda + \sigma) - \lambda \delta F^c]}{(1 - F^c)s^c(\lambda + \sigma)} \left(1 - s(v)\right)^{n-1} - \frac{\lambda \delta (\lambda + \delta + \sigma)}{y - (1 - s(v))} \right].
\]

Part II. Show first order stochastic dominance: From (36), one can see that the conditional probability is decreasing in \(n\) for any \(v < \bar{v}\) if the term in the curly bracket is positive, since \(F(v) > F^c\) and \(\frac{1 - s(v)}{y} < 1\).

I now turn to the term in the curly bracket

\[
\frac{[s^c(\lambda + \sigma) - \lambda \delta F^c]}{1 - F^c} - \frac{\lambda \delta (\lambda + \delta + \sigma)}{y - (1 - s(v))} = \frac{[s^c(\lambda + \sigma) - \lambda \delta F^c]}{(1 - F^c) [y - (1 - s(v))]},
\]

where the denominator > 0, using the definition of \(y\), one can show that the numerator is

\[
[(1 - \sigma - \delta) - (1 - s(v))] [s^c(\lambda + \sigma) - \lambda \delta F^c] > 0
\]

where the inequality follows from the fact that \((1 - \sigma - \delta) > 1 - s(v)\).

As a result, \(\Pr(\hat{v} \leq v|h = (1 + g)^n)\) is first order stochastically increasing in human capital level \(h\) for any \(v \geq \hat{v}\).
\textbf{Proof.} Recall the definition of \( l(v) \): \( l(v) = \lambda[I(v \geq v_u)uE(h|u) + \sum_{h} h \Pr(\bar{v}' \prec v, h)] \), the proof follows from the definition of \( \Pr(\bar{v}' \prec v, (1 + g)^n) \) given in Proposition 4. If \( v_u \leq v \leq \bar{v} \),

\[
\frac{l(v)}{\lambda} = uE(h|u) + \sum_{n=0}^{\infty} (1 + g)^n \Pr(\bar{v}' \prec v, (1 + g)^n) \\
= uE(h|u) + \lambda F(v) \frac{\lambda}{s(v)} \sum_{n=0}^{\infty} (1 + g)^n uD^u[(1 + g)^n] \\
= (1 + \frac{\lambda F(v)}{s(v)}) uE(h|u) \\
= \frac{\sigma + \lambda + \delta}{s(v)} uE(h|u)
\]

If \( \bar{v} < v \leq \bar{v} \),

\[
\frac{l(v)}{\lambda} = uE(h|u) + \sum_{n=0}^{\infty} (1 + g)^n \Pr(\bar{v}' \prec v, (1 + g)^n) \\
= uE(h|u) + \frac{\lambda F^c}{s^c} uE(h|u) \\
+ \lambda(F(v) - F^c)D(1) \sum_{n=0}^{\infty} (1 + g)^n \left\{ \frac{\lambda \delta (\lambda + \sigma + \delta)(1 - F^c)y^{n-2}}{s^c(\lambda + \sigma) - \lambda \delta F^c} \sum_{m=1}^{n-1} \left( \frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\}.
\]

where, the last equality uses the result from equation (35). Define \( X \) as the constant term \( \frac{\lambda \delta (\lambda + \sigma + \delta)(1 - F^c)}{[\lambda + \sigma]s^c - \delta \lambda F^c} \), compute the summation:

\[
\sum_{n=0}^{\infty} (1 + g)^n \left\{ \frac{Xy^{n-2}}{y} \sum_{m=1}^{n-1} \left( \frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\} \\
= \sum_{n=0}^{\infty} (1 + g)^n \left\{ \frac{Xy^{n-2}}{y} \sum_{m=1}^{n-1} \left( \frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\} \\
= \frac{(1 + g)X}{y(1 - s(v)\frac{1}{y})} \sum_{n=0}^{\infty} (1 + g)^{n-1} y^{-1} \left[ 1 - \frac{1 - s(v)}{y} \right]^{n-1} + (1 + g) \sum_{n=0}^{\infty} \left[ (1 + g)(1 - s(v)) \right]^{n-1} \\
= \frac{(1 + g)X}{y - 1 + s(v)\frac{1}{y}} \left[ 1 - \frac{1}{1 - (1 + g)} \right] + \frac{1 + g}{1 - (1 + g)(1 - s(v))} \\
= \frac{(1 + g)}{[(1 + g)s(v) - g] \left[ 1 - (1 + g)y \right] + 1}
\]
Plug in the definition of $y$ and $X$, using the relationship between $D(1)$ and $uD^u(1)$, I have

$$D(1) \sum_{n=0}^{\infty} (1 + g)^n \{ Xy^{n-2} \sum_{m=1}^{n-1} \left( \frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \}$$

$$= \frac{(\delta + \sigma + \lambda)(1 + g)}{[(1 + g)s(v) - g]s^e} uD^u(1) *$$

$$\{ \lambda \delta(\lambda + \sigma + \delta)(1 - F^c)(1 + g) \}
\frac{1}{(1 + g)\sigma s^e(\lambda + \sigma + \delta) - g[s^e(\sigma + \lambda) - \lambda \delta F^c] + 1}$$

$$= \frac{(\delta + \sigma + \lambda)(1 + g)}{[(1 + g)s(v) - g]s^e} uE(h|u)$$

Going back to $\frac{l(\widehat{v})}{\lambda}$,

$$\frac{l(\widehat{v})}{\lambda} = \frac{\lambda + \sigma + \delta}{s^e} uE(h|u) + \frac{\lambda(F(v) - F^c)(\delta + \sigma + \lambda)(1 + g)}{[(1 + g)s(v) - g]s^e} uE(h|u)$$

$$= \frac{(\delta + \sigma + \lambda)(1 + g)}{[(1 + g)s(v) - g]s^e} uE(h|u).$$

\[ \Box \]

A10. Proof for Proposition 5

(Job offer distribution)

**Proof.** Using the definitions of the firm’s profit and the average quality of workers it can hire, there are the following two cases.

If $v_u \leq v \leq \widehat{v}^c$, $d = 0$ and

$$\pi(d = 0; v) = \frac{(p - \theta_0(v))\lambda(\sigma + \lambda + \delta)}{s(v)^2} uE(h|u).$$

Equal profit condition $\pi(d = 0; v) = \pi(d = 0; v_u)$ implies:

$$\frac{(p - \theta_0(v))(\sigma + \lambda + \delta)}{s(v)^2} = \frac{(p - b)}{\sigma + \lambda + \delta}, \quad (37)$$

where I use the fact that $\theta_0(v_u) = b$ and $s(v_u) = \sigma + \delta + \lambda$. The result follows immediately from the fact that $s(v) = \sigma + \delta + \lambda(1 - F(v))$.

If $\widehat{v}^c < v \leq \overline{v}$, $d = 1$ and

$$\pi(d = 1; v) = \frac{(p - \theta_1(v) - c)}{s(v)(1 + g)} \frac{\lambda(\sigma + \lambda + \delta)}{s^e} \frac{s^e(1 + g) - g}{s^e[s(v)(1 + g) - g]} uE(h|u).$$

38
Equal profit condition \( \pi(d = 1; v) = \pi(d = 1; \tilde{v}^e) \) implies
\[
\frac{(p - \theta_1(v) - c)}{s(v)(1 + g) - g} = \frac{(p - \theta_1(\tilde{v}^e) - c)}{s^e(1 + g) - g} \tag{38}
\]
and the result, again, follows from the relationship between \( s(v) \) and \( F(v) \).
Using the relationship between \( \theta_1(\tilde{v}^e), \theta_0(\tilde{v}^e) \) and \( c \), one can prove the continuity of \( F(.) \) by showing \( F(\tilde{v}^e) \) in (20) is the same as \( F(\tilde{v}^e) \) in (21).

A11. Proof for Proposition 6
(Existence and Uniqueness of Market Equilibrium)

In the following, I lay out the logic in deriving the job offer distribution in primitives, from which I establish the existence and uniqueness of the market equilibrium.

1) From the worker’s Bellman equations, I get the following relationship between \( b_v \) and \( \theta \); for non-training job, \( d\tilde{v}_0/d\theta = 1/s(v) \), where \( s(v) \) is the separation probability. For training jobs, \( d\tilde{v}_1/d\theta = 1/[(1 + g)s(v) - g] \).

2) From the equal profit condition, i.e. every firm should get the same profit as the firm that posts \( v_u \), and from the fact that \( \theta_0(v_u) = b \), I can get the following distribution: for \( v < \tilde{v}^e \),
\[
s(v) = (\sigma + \delta + \lambda) \sqrt{\frac{p - \theta_0(v)}{p - b}} \tag{39}
\]
for \( v \geq \tilde{v}^e \)
\[
(1 + g)s(v) - g = (\sigma + \delta + \lambda) \sqrt{\frac{p - \theta_1(v) - c}{p - b}} \sqrt{\frac{(1 + g)s^e - g}{s^e}} \tag{40}
\]
where \( s^e = \sigma + \delta + \lambda(1 - F(\tilde{v}^e)) \), separation rate for the firm that is indifferent between training and non-training.

3) Plug (39) and (40) into part 1, I get the following relationship between \( \tilde{v} \) and \( \theta \), where \( M_0 \) and \( M_1 \) are constants.
\[
\begin{align*}
\tilde{v}_0(\theta) &= M_0 - \frac{2\sqrt{(p - b)(p - \theta)}}{\sigma + \delta + \lambda} \\
\tilde{v}_1(\theta) &= M_1 - \frac{2\sqrt{(p - b)(p - \theta - c)}}{\sigma + \delta + \lambda} \sqrt{\frac{s^e}{(1 + g)s^e - g}}
\end{align*} \tag{41}
\]
At \( \tilde{v}^e \), the right hand side of these two equations should be equal, therefore, I have
\[
\begin{align*}
M_0(\sigma + \delta + \lambda) - 2\sqrt{(p - b)(p - \theta_0^e)} &= M_1(\sigma + \delta + \lambda) - 2\sqrt{(p - b)(p - \theta_1^e - c)} \sqrt{\frac{s^e}{(1 + g)s^e - g}} \tag{42}
\end{align*}
\]
4). Evaluate (39) and (40) at \( \hat{v}^c \), I get the pay rate at \( \hat{v}^c \)

\[
\theta_0^c = p - \left( \frac{s^c}{\sigma + \delta + \lambda} \right)^2 (p - b) \tag{43}
\]

\[
\theta_1^c = p - c - \frac{s^c[(1 + g)s^c - g]}{(\sigma + \delta + \lambda)^2} (p - b)
\]

Plug these into (42), I get the relationship that \( M_0 = M_1 \equiv M \).

5). Using the definition of \( \hat{v}^c \), i.e. at \( \hat{v}^c \), the worker-firm joint benefit is equal to \( c \), and using (43), I get the following

\[
c = g \left[ M_0 - p - \frac{s^c}{(\sigma + \delta + \lambda)^2} (p - b) \right] \tag{44}
\]

6). The relationship between \( \bar{v} \) and \( v_u \), from the Bellman equation, is the following

\[
\bar{v} = \theta_1(\bar{v}) + \delta(1 + g)v_u \quad \frac{(\sigma + \delta)(1 + g) - g}{(\sigma + \delta)(1 + g) - g}
\]

where \( v_u \) can be derived from (41) as

\[
v_u = M_0 - \frac{2(p - b)}{(\sigma + \delta + \lambda)}
\]

Moreover, from (41), I also have

\[
\bar{v} = M_1 - \frac{2\sqrt{(p - b)(p - \theta_1(\bar{v}) - c)}}{\sigma + \delta + \lambda} \sqrt{\frac{s^c}{(1 + g)s^c - g}}
\]

where \( \theta_1(\bar{v}) \) can be derived from (40)

\[
\theta_1(\bar{v}) = p - c - \left( \frac{(1 + g)(\delta + \sigma) - g}{\sigma + \delta + \lambda} \right)^2 \frac{s^c}{(1 + g)s^c - g} (p - b).
\]

Therefore, I get the following relationship

\[
M_1 - \frac{2s^c[(1 + g)(\delta + \sigma) - g]}{[(1 + g)s^c - g][(\sigma + \delta + \lambda)^2]} (p - b) = \frac{p - c - (\frac{(1 + g)(\delta + \sigma) - g}{\sigma + \delta + \lambda})^2 \frac{s^c}{(1 + g)s^c - g} (p - b)}{(\sigma + \delta)(1 + g) - g} \left[ M_0 - \frac{2(p - b)}{(\sigma + \delta + \lambda)} \right] \tag{45}
\]

7). With \( M_0 = M_1 \), (44) and (45) are two equations in two unknowns \( s^c \) and \( M \). Plugging (44) into (45), I obtain one equation in one unknown: \( s^c \). Solving this equation for \( s^c \), I can then back out the whole distribution of \( \hat{v} \). Notice that for coexistence of training and non training firms, it is required that \( s^c \in (\sigma + \delta, \sigma + \delta + \lambda) \), which in turn puts restrictions on parameter values such as training cost. For parameter values that satisfy this requirement, it can be shown that the solution \( s^c \)
exists and is unique. For parameter values that do not satisfy this requirement, the market equilibrium features either universal training or no training at all (but not both), and the equilibrium can be solved similarly. Nonetheless, for given parameter values, distribution of $\tilde{\nu}$ exists and is unique, which implies the existence and uniqueness of the market equilibrium.

8) Given the distribution of $\tilde{\nu}$, I obtain the $D(h)$ and $Pr(\tilde{\nu} \leq \nu, h)$ that are characterized in Proposition 3 and 4.

A12 Proof for Proposition 8
(Firms that offer higher $\tilde{\nu}$ also offer higher growth rate)

**Proof.** FOC with respect to $g$:

$$\begin{align*}
\frac{[\tilde{\nu} - \theta_0(\tilde{\nu}) - C'(g)][1 - (1 - s(\tilde{\nu}))(1 + g)] + [p - \theta_0(\tilde{\nu}) + g(\tilde{\nu} - \theta_0(\tilde{\nu})) - C(g)](1 - s(\tilde{\nu}))}{[s(\tilde{\nu})(1 + g) - g]^2} &\leq 0 \\
= 0 & \text{ if } g > 0
\end{align*}$$

This is equivalent to

$$\begin{align*}
[\tilde{\nu} - \theta_0(\tilde{\nu}) - C'(g)][1 - (1 - s(\tilde{\nu}))(1 + g)] + [p - \theta_0(\tilde{\nu}) + g(\tilde{\nu} - \theta_0(\tilde{\nu})) - C(g)](1 - s(\tilde{\nu})) \leq 0 & \text{ if } g > 0
\end{align*}$$

Rearranging and collecting terms, one gets

$$\begin{align*}
s(\tilde{\nu})\tilde{\nu} + (1 - s(\tilde{\nu}))[p - C(g)] - \theta_0(\tilde{\nu}) - C'(g)[s(\tilde{\nu})(1 + g) - g] \leq 0 & \text{ if } g > 0
\end{align*}$$

For now, I assume interior solution, (the conditions for existence of interior solution will be discussed below) and determine the relationship between optimal growth rate and the promised $\tilde{\nu}$ value. Define

$$L(g; \tilde{\nu}) = s(\tilde{\nu})\tilde{\nu} + (1 - s(\tilde{\nu}))[p - C(g)] - \theta_0(\tilde{\nu}) - C'(g)[s(\tilde{\nu})(1 + g) - g]$$

Take partial derivative:

$$\frac{\partial L}{\partial g} = -C'(g)(1 - s(\tilde{\nu})) + C'(g)(1 - s(\tilde{\nu})) - C''(g)[s(\tilde{\nu})(1 + g) - g]$$

$$= -C''(g)[s(\tilde{\nu})(1 + g) - g]$$

$$< 0 \quad \text{by convexity of } C(.)$$
\[
\frac{\partial L}{\partial \tilde{v}} = s(\tilde{v}) + s'(\tilde{v})\tilde{v} - s'(\tilde{v})(p - C(g)) - \theta_0'(\tilde{v}) - C'(g)s'(\tilde{v})(1 + g)
\]
\[
= s(\tilde{v}) + s'(\tilde{v})[\tilde{v} - p + C(g) - C'(g)(1 + g)] - s(\tilde{v})
\]
\[
= s'(\tilde{v})[\tilde{v} - p + C(g) - C'(g)(1 + g)],
\]
where I use the fact that \( \theta_0'(\tilde{v}) = s(\tilde{v}) \). I need to sign the term \([\tilde{v} - p + C(g) - C'(g)(1 + g)]\), which is of the opposite sign as \( \partial L / \partial \tilde{v} \).

From \( s(\tilde{v})\tilde{v} + (1 - s(\tilde{v}))[p - C(g)] - \theta_0(\tilde{v}) - C'(g)[s(\tilde{v})(1 + g) - g] = 0 \) one can get
\[
[\tilde{v} - p + C(g) - C'(g)(1 + g)] = \frac{\tilde{v} - \theta_0(\tilde{v}) - C'(g)}{(1 - s(\tilde{v}))}
\]
From \( [\tilde{v} - \theta_0(\tilde{v}) - C'(g)][1 - (1 - s(\tilde{v}))(1 + g)] + [p - \theta_0(\tilde{v}) + g(\tilde{v} - \theta_0(\tilde{v})) - C(g)](1 - s(\tilde{v})) = 0 \) one gets:
\[
\frac{\tilde{v} - \theta_0(\tilde{v}) - C'(g)}{(1 - s(\tilde{v}))} = -\frac{[p - \theta_0(\tilde{v}) + g(\tilde{v} - \theta_0(\tilde{v})) - C(g)]}{[1 - (1 - s(\tilde{v}))(1 + g)]}
\]
\[
= -\frac{[p - \theta_0(\tilde{v}) - C'(g)] + g(\tilde{v} - \theta_0(\tilde{v}))}{[1 - (1 - s(\tilde{v}))(1 + g)]}
\]
\[
< 0
\]
where the inequality follow from the fact that the future value of a job offer is positive, and that profit is positive in equilibrium. Therefore I have \( \partial L / \partial \tilde{v} > 0 \). As a result, \( \frac{\partial g}{\partial \tilde{v}} = -\frac{\partial L}{\partial \tilde{v}} / \frac{\partial L}{\partial g} > 0 \). ■