

Training, Search and Wage Dispersion*

Chao Fu[†]
University of Wisconsin-Madison

First Draft: December 2006
This Version: October 2010

*I thank Kenneth Burdett, Philipp Kircher and Guido Menzio for their invaluable advice and support. I thank Hanming Fang, John Kennan, Rasmus Lentz, Iouri Manovskii, Roberto Pinheiro, Xi Weng and Randall Wright for their comments and encouragement. I thank the editor and two anonymous referees for insightful suggestions. I also thank participants at the April 2007 Penn Search and Matching workshop and participants at the January 2008 5th SED Conference for helpful comments and discussions. All errors are mine.

[†]Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706, USA.
Email: cfu@ssc.wisc.edu

Abstract

This paper combines on-the-job search and human capital theory to study the coexistence of firm-funded general training and frequent job turnovers. Although ex ante identical, firms differ in their training decisions. The model generates correlations between various firm characteristics that are consistent with the data. Wage dispersion exists among ex ante identical workers because workers of the same productivity are paid differently across firms, and because workers differ in their productivity ex post. Endogenous training breaks the perfect correlation between work experience and human capital, which yields new insights on wage dispersion and wage dynamics.

Keywords: On-the-job search, on-the-job training, general human capital, wage dispersion, wage persistence, wage growth

JEL codes: J64, J24, J31

1 Introduction

Modern technologies and work organizations require continuous upgrading of worker skills, which is best accomplished via on-the-job training. Some on-the-job training improves skills specific to the firm providing it, but some improves the worker's general skills. For example, Lynch and Black (1998) find that over 50% (but not all) of U.S. firms provide and pay for general training such as computer skills training and teamwork training. By improving trainees' earning capability across jobs, general training creates persistent income dispersion among ex ante identical workers with different training experiences. Given that these different training experiences are largely due to firms' different training decisions, the latter demands further investigation.

In a perfectly competitive market, firms will not pay for general training (Becker (1962)). With market friction, a firm may pay for general training because workers' job mobility is restricted and the firm can earn rents on its trained workers. In other words, restrictions on worker mobility are key to firms' training decisions. However, job-to-job transitions are a significant phenomenon in real life. 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 develop such a model by integrating human capital theory and job search theory. The model generates cross-firm comparisons that are consistent with the data. It also yields new insights on wage dispersion and wage dynamics.

To address job-to-job transitions, I draw on Burdett and Mortensen (1998) (BM), who explicitly model job turnovers with search frictions: with a given probability, unemployed and employed workers receive wage offers from firms with vacancies. To incorporate firms' training decisions, I extend BM by allowing the firm to post a training opportunity as well as a pay rate per unit of human capital.¹ The combination of these two elements determines the value of a job offer for the worker. Although firms are ex ante identical, there is a non-degenerate distribution of job values offered in equilibrium. Firms differ in their training and pay rate decisions. Only some but not all firms provide training. When training is provided, the firm and the worker share the cost and benefit of training. Firms with training make offers that yield higher values to workers. By offering higher values to its workers, training firms are more likely to keep their workers for longer, which justifies their provision

¹Training refers to general training in this paper.

of training. Consistent with the data, the model predicts a positive correlation between firm size and the likelihood of general training, and a positive correlation between wage growth rate and average tenure within a firm. It also provides a new way to explain the positive correlation between within-firm wage dispersion and within-firm mean wage.

At the worker level, a worker's wage grows over time because she climbs up the job ladder via on-the-job search, and because she becomes more productive via on-the-job training. Although workers are ex ante identical, at any point in time, wage dispersion exists because 1) identical workers are paid differently, 2) workers differ in their productivity ex post, and 3) there is a positive correlation between pay rate earned and human capital level. The positive correlation results from the fact that more experienced workers find better-paying jobs and they also accumulate more human capital. By decomposing the wage into human capital and pay rate per unit of human capital, the model yields a distribution of wages with a long declining right tail, as observed in the data. Due to different training levels across firms, workers with the same years of work experience may differ in their actual human capital due to different training experience. Moreover, regardless of the worker's current employment status, her entire work history (not only her years of experience) matters for her future wage profile. This leads to important implications for wage dispersion and wage persistence.

The rest of the paper is organized as follows: the next section reviews the literature. Section 3 lays out the model. Section 4 analyzes the market equilibrium. In section 5, I endogenize the growth rate of human capital by allowing firms to choose the intensity of training and show that the main results still hold. Section 6 summarizes model predictions and relates them to the empirical literature. The final section concludes the paper. The appendix contains some proofs. More detailed proofs are contained in the online appendix.

2 Related Literature

This paper is related to various studies that have focused on different aspects of the following issues: training and job turnovers, wage dynamics and wage dispersion.

In Rosen (1972), there is an implicit market for training opportunities that is dual to the market for jobs. The worker pays for training by "buying" the job from the firm. To explain the existence of firm-funded general training, Acemoglu and Pischke (1998) draw on worker heterogeneity and information friction. The superior knowledge about its workers' ability

encourages the incumbent firm to fund general training.² In Moen and Rosen (2004), some firms provide general training because they have the comparative advantage in doing so.

There are a few papers that incorporate general human capital into similar job turnover frameworks as in my paper.³ Rubinstein and Weiss (2007) study human capital accumulation and on-the-job search without modeling the equilibrium. Bagger, Fontaine, Postel-Vinay and Robin (2006) incorporate learning-by-doing with individual productivity shocks within the framework developed in Postel-Vinay and Robin (2002). They focus on estimating wage patterns over the life cycle for individual workers.

Independently, Burdett, Carrillo-Tudela and Coles (2009) also study both wage dynamics and wage dispersion, but with exogenous learning-by-doing. Their analysis yields a standard Mincer wage equation with worker fixed effects and firm fixed effects. For various specifications of worker heterogeneity, their simulation results show the relative importance of each effect in wage dispersion. My paper abstracts from worker heterogeneity and learning-by-doing and, instead, models firms' endogenous training decisions for the following considerations.⁴ First, wage growth differs significantly across firms and it is systematically correlated with other firm characteristics. This cannot be easily explained by exogenous learning-by-doing. Second, learning-by-doing models reduce the worker's working history into a single summary statistic, i.e., years of working experience. Relaxation of this assumption leads to new insights on wage dispersion and wage persistence.

This model also contributes to the discussion about wage cuts over voluntary job-to-job transitions. Workers may take wage cuts on transition as an investment for better job prospects in the future. For example, in the offer matching framework developed by Postel-Vinay and Robin (2002), the return to such investment is realized when good luck strikes and the worker is poached by a highly productive firm. In my model, a worker takes a wage cut in return for more training and hence higher wage growth.

3 The Model

In this section, I will analyze the basic model in which firms' training decisions are binary. Later in the paper, I will extend the model and allow firms to choose their own training

²Katz and Ziderman (1990) use similar arguments to explain firm-provided general training.

³Quercioli (2005) considers firms' decisions on specific training in a BM framework.

⁴I also abstract from wage-tenure contracts studied in Burdett and Coles (2003) and Stevens (2004).

intensities.

3.1 Basic Framework

Time is discrete. There is a continuum of risk-neutral workers and firms, each of measure one. On entering the market, each worker is endowed with one unit of human capital.⁵ Human capital is general, and for simplicity, I assume it does not decay. All workers retire and leave the market for good with probability σ per period. Each retired worker is replaced with a new unemployed worker, so that the economy is in steady state.

Firms are homogeneous in that any firm generates revenue p from each unit of human capital it employs. Each firm posts a job offer (θ, d) , where θ is the pay rate per unit of human capital and d specifies the provision of training: $d = 1$ if training is provided, $d = 0$ otherwise. If the firm provides training ($d = 1$), a per period cost c must be paid for each unit of human capital it employs. If a worker has been exposed to training for t periods during her life, her human capital is $h = (1 + g)^t$. When a worker with human capital h is employed at a job with (θ, d) , the current wage she receives is θh , and $(p - \theta - dc)h$ is the firm's per period profit from the worker. With probability δ per period any given job is destroyed. With probability λ per period, a job offer arrives for the worker, regardless of her employment status. Job destruction, job offer and retirement are mutually exclusive events, and $(\delta + \lambda + \sigma) < 1$. For an unemployed worker with h , she obtains bh each period, where $p > b > 0$. Hence, b can be viewed as the productivity of human capital in home production. Since there is worker retirement, for simplicity, I assume no discounting. Since pay rate is measured by efficiency unit, I assume that a firm offers all new employees the same binding contract. I also assume no recall should a worker quit or reject a job offer.

3.2 Worker Problem

Let $V(\theta, d; h)$ denote the expected lifetime income of a worker who has h units of human capital and is employed at a firm that offers (θ, d) . Clearly, a worker who accepts any job offer will never freely quit employment into unemployment. Therefore, if the job does not

⁵In this paper, I use efficiency unit and unit of human capital interchangeably.

offer training, i.e., $d = 0$, the job value is:

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

As long as she stays on a job with $(\theta, d = 0)$, 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 (1). The next period, with probability λ , the worker gets a new offer, upon which she chooses between staying with the current firm or moving to the new firm, 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 (the last term in (1)). If a job offers training, i.e., $d = 1$, the job value is:⁶

$$\begin{aligned} V(\theta, 1; h) &= \theta h + \lambda E_{(\theta', d')} \max\{V(\theta, 1; h(1 + g)), V(\theta', d'; h(1 + g))\} \\ &\quad + \delta U(h(1 + g)) + (1 - \delta - \lambda - \sigma)V(\theta, 1; h(1 + g)). \end{aligned}$$

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 define the per-efficiency-unit value functions by $v_u = U(h)/h$ and $v = V(\theta, d; h)/h$. Denote by $F(v)$ the fraction of offers with per-efficiency-unit value no greater than v . When making her search decision, an employed worker compares the per-efficiency-unit value of her current job with that of the outside offer; an unemployed worker compares the per-efficiency-unit value of the offer with that of unemployment (v_u). For brevity, I will call these per-efficiency-unit values the values of a job and of unemployment. For those employed at a firm offering (θ, d) that yields v , it follows

$$v = \theta + (1 + dg)\left\{\lambda \int_{\underline{v}}^{\bar{v}} \max\{v, v'\} dF(v') + \delta v_u + (1 - \delta - \lambda - \sigma)v\right\}, \text{ for } d = 0, 1. \quad (2)$$

For the unemployed,

⁶The events that can happen to the worker are the same as when she is on a job with no training, except that her human capital grows at rate $(1 + g)$.

$$v_u = b + \lambda \int_{\underline{v}}^{\bar{v}} \max\{v_u, v'\} dF(v') + (1 - \lambda - \sigma)v_u, \quad (3)$$

where \bar{v} and \underline{v} are the upper and lower bounds for job values in equilibrium and will be specified later. It is assumed that $(1 - \sigma)(1 + g) < 1$, which guarantees boundedness of the value functions.

In case of indifference, I assume that an unemployed worker accepts the job offer but an employed worker stays with the current employer. Given these harmless tie-breaking restrictions, optimal job search implies the following strategies for the worker:

1. When unemployed, the worker accepts a job offer if it has value $v \geq v_u$;
2. When employed with contract (θ, d) that delivers v , the worker quits if and only if a job offer is received with value $v' > v$.

Given training/no training, i.e., $d = 0/1$, it is simple to show that there is a unique pay rate that can yield the job value v . Define such pay rate by $\theta_d(v)$, for $d = 0, 1$.

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

For any given v , a worker can always compute $\theta_0(v)$ and its relationship with $\theta_1(v)$.

Lemma 2 *Given v , the worker demands $\theta_1(v) = \theta_0(v) - g(v - \theta_0(v))$ in order to be indifferent between a job that offers pay rate $\theta_0(v)$ but no training and a job with pay rate $\theta_1(v)$ and training. The gap between $\theta_0(v)$ and $\theta_1(v)$ is increasing in v .⁷*

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. As such, wage cuts may occur over voluntary job-to-job transitions, as seen in the data.⁸ Moreover, if the job is of higher v , the worker is offered more for each unit of her human capital. In that case, accumulating human capital is more rewarding and so the pay rate on a job with training would be even lower for the worker to remain indifferent to a job that does not.

⁷The proofs for Lemma 1 and Lemma 2 are in the online appendix.

⁸For example, Connolly and Gottschalk (2008).

3.3 Firm Problem

Let u be the steady-state unemployment rate and $E(h|u)$ be the average human capital level of unemployed workers. Let $\Pr(v' < v, h)$ denote the measure of workers with human capital level h that are employed at jobs with value lower than v . So the joint distribution of (v, h) among employed workers is $\Pr(v' < v, h)/(1 - u)$. The expected human capital level that can be employed by a firm with v (denoted by $l(v)$) is:

$$l(v) = \lambda [I(v \geq v_u)uE(h|u) + (1 - u) \sum_h h \frac{\Pr(v' < v, h)}{1 - u}], \quad (4)$$

where $I(\cdot)$ is an indicator function that equals 1 if the argument is true, and 0 otherwise. With probability λ the firm meets a worker, it can attract an unemployed worker if the promised 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 works at a job with value less than v , and the expected human capital level of this worker is $\sum_h h \Pr(v' < v, h)/(1 - u)$.

The worker leaves the firm when the job is destroyed, when she retires, or if she receives an outside offer with value higher than v . Hence the separation rate for a firm offering v is:

$$s(v) = \delta + \sigma + \lambda(1 - F(v)). \quad (5)$$

For a firm with job value v , the steady-state flow profit is given by:

$$\begin{aligned} \pi(v) &= \max_{\theta, d} \left\{ l(v) \sum_{t=0}^{\infty} (1 - s(v))^t (p - \theta - dc)(1 + g)^{dt} \right\} \\ &= \max_{\theta, d} \left\{ l(v) \frac{p - \theta - dc}{1 - (1 - s(v))(1 + g)^d} \right\} \\ \text{s.t. } v &= \theta + (1 + dg) \left\{ \lambda \int_v^{\bar{v}} \max\{v, v'\} dF(v') + \delta v_u + (1 - \delta - \lambda - \sigma)v \right\} \end{aligned}$$

If the firm does not provide training, it extracts $(p - \theta)$ from each efficiency unit it employs for as long as the worker stays at the firm. If the firm provides training, besides pay rate θ , it pays cost c for each efficiency unit per period. In return, the efficiency units it employs 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 v per efficiency unit to the worker.

The firm's problem can be decomposed into two steps: first, it chooses the value v it will

deliver to the worker; second, it chooses the most efficient combination of θ and d to deliver v . With the pay rate function $\theta_d(\cdot)$, the second-step problem boils down to the choice of d , and the firm's problem can be written as

$$\begin{aligned}\pi &= \max_v \{\pi(v)\} \\ &= \max_v \{\max_d \{\pi(d=0;v), \pi(d=1;v)\}\},\end{aligned}$$

where

$$\pi(d=0;v) = \frac{(p - \theta_0(v))}{s(v)} l(v), \quad (6)$$

$$\pi(d=1;v) = \frac{(p - \theta_1(v) - c)}{1 - (1 - s(v))(1 + g)} l(v). \quad (7)$$

3.3.1 Optimal Pay Rate-Training Contracts

Lemma 3 *Given the promised value v , with pay rate*

$$\theta_1^f(v) = \theta_0(v) - c + \frac{g(1 - s(v))(p - \theta_0(v))}{s(v)}, \quad (8)$$

a firm providing training earns the same profit as a firm offering wage $\theta_0(v)$ and no training.

Proof. It follows from equalizing the right-hand sides of (6) and (7) and solving for θ_1 . ■

The last term on the right-hand side of (8) represents the expected future gain for the firm from the increased human capital. Instead of fully internalizing the cost of training through cutting the pay rate by c , the firm is willing to bear the part of the cost that is equal to its expected benefit. Since it is constrained to deliver v to the worker, the firm can decide on training provision by comparing the pay rate demanded by the worker, i.e., $\theta_1(v)$, and the pay rate necessary for equal profit, i.e., $\theta_1^f(v)$.

Proposition 1 *Given the value v it has promised to its worker, the firm's optimal training choice d is characterized by the following:*

(i)

If $c > B(v)$, firm with v chooses $d = 0$.

(ii)

If $c < B(v)$, firm with v chooses $d = 1$.

(iii)

If $c = B(v)$, firm with v chooses $d = 0$ or $d = 1$;

where

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

is the worker-firm joint benefit from training.

Proof. To keep its promise of v , the firm has to give the worker $\theta_1(v)$ should it choose $d = 1$. If $\theta_1(v) > \theta_1^f(v)$, when offered a job with $d = 1$, the worker demands a higher pay rate than the pay rate that maintains the same profit the firm gets when it offers no training. Therefore, it is cheaper for the firm to deliver v with $(\theta_0(v), d = 0)$ rather than with $(\theta_1(v), d = 1)$. Similarly, if $\theta_1(v) < \theta_1^f(v)$, it is cheaper for the firm to deliver v with $(\theta_1(v), d = 1)$. When $\theta_1(v) = \theta_1^f(v)$, the firm is indifferent between offering training and not offering it. The rest of the proof is obtained once I combine the expressions of $\theta_1(v)$ (from Lemma 1) and $\theta_1^f(v)$ (from (8)). ■

Given the promised v , the choice of whether to offer training is based on the comparison between the cost of training and the worker-firm joint benefit from training. This joint benefit equals the sum of the amounts that the two parties are willing to pay for 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, and consequently, they also share the cost.

Lemma 4 *The worker-firm joint benefit from training, $B(v)$, is increasing in v .⁹*

At a higher v , both the firm and the worker are more willing to pay for training. For the firm, offering the worker a higher v keeps the worker for a longer time and extracts more from her, which justifies its investment in training. Given a higher value per unit of her human capital, the worker is rewarded more for her human capital and hence she will have greater willingness to pay for training. Therefore, $B(v)$ is increasing in v .

Combining Lemma 4 with the fact that training cost c is a fixed value, there will be three different cases with respect to the optimal training decisions, depending on the value of c . Let \bar{v}^0, v_u^0 denote the highest value and the lowest value offered in a market where no firm

⁹See Appendix A for the proof.

offers training, and let v_u^1 denote the lowest value offered in a market where all firms offer training.

Case 1: If

$$\begin{aligned} c &> B(\bar{v}^0) \\ &= g \frac{(1 - \delta - \sigma)p + \delta v_u^0}{\delta + \sigma}, \end{aligned}$$

all firms will optimally choose not to offer training.¹⁰

Case 2: If

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

all firms will choose to offer training.¹¹

Case 3: If $B(v_u^1) < c \leq B(\bar{v}^0)$, there exists a cutoff value v^c such that $c = B(v^c)$.¹² Firms offering $v < v^c$ will do so by offering $\theta_0(v)$ without training. Firms offering $v > v^c$ will do so by offering $\theta_1(v)$ with training. Firms offering v^c will do so by either offering $\theta_0(v^c)$ without training or offering $\theta_1(v^c)$ with training.

Lemma 4 shows that firms with higher v are more likely to offer training. Therefore, if the highest value firm finds the training cost too high to benefit from it, then no firm will provide training, which is case 1. 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 holds and if parameter values are such that $B(v_u^1) < c \leq B(\bar{v}^0)$, the economy will be divided into a training sector and a non-training sector. In the former, firms provide training and higher job values. In the latter, firms do not provide training and jobs are of lower values.¹³

¹⁰The equality follows from the fact that

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

The derivation of \bar{v} uses the fact that given the most generous offer, the worker will never quit.

¹¹The equality follows from the fact that $\theta_0(v_u) = b$.

¹²By the Intermediate Value Theorem, the existence of the cutoff value v^c follows from the continuity of $B(\cdot)$ in v .

¹³In line with previous studies, market provision of general training is inefficiently low due to the externality of general training. A formal proof is in the online appendix.

4 Market Equilibrium

Definition 1 *A market equilibrium is:*

1. *a job offer distribution F of the expected lifetime per-efficiency-unit value v , such that:*

$$\begin{aligned}\pi^* &= \pi(v) = \max_d \{\pi(d=0; v), \pi(d=1; v)\} \text{ for all } v \in [\underline{v}, \bar{v}] \\ \pi^* &\geq \pi(v) \text{ otherwise.}\end{aligned}$$

That is, any contract offered maximizes the firm's profit and the maximized profit is equalized across optimizing firms;

2. *an optimal pay rate-training contract $(\theta_d(v), d)$ that delivers v , according to (2), for every $v \in [\underline{v}, \bar{v}]$;*
3. *workers' optimal job search and quit strategies;*
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 $\frac{1}{1-u} \Pr(v' \leq v, h)$ that are consistent with the steady-state turnover, given $F(\cdot)$ and workers' optimal strategies.*

A more detailed analysis of market equilibrium can be found in Appendix B. There I first derive the steady-state human capital distribution among unemployed workers. I show in Appendix B1 that beyond the basic human capital level ($h = 1$), 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, i.e., $(1 + g)(1 - \sigma) < 1$, the average human capital remains finite. In Appendices B2 and B3, I derive the joint distribution of job values and human capital among employed workers and show that there is a positive correlation between job value and worker productivity.

4.1 Expected Human Capital Employed By Firm v

Denote $F^c \equiv F(v^c)$ as the measure of the non-training sector, and $s^c = \sigma + \delta + \lambda(1 - F^c)$ as the separation rate for the non-training sector.

Claim 1 *The expected human capital level in a firm with value v , $l(v)$, is given by:*¹⁴

If $v_u \leq v \leq v^c$, then

$$l(v) = \lambda \frac{\sigma + \lambda + \delta}{s(v)} u E(h|u). \quad (9)$$

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

$$l(v) = \lambda \frac{(\sigma + \lambda + \delta) [s^c(1 + g) - g]}{s^c[s(v)(1 + g) - g]} u E(h|u). \quad (10)$$

It is easy to see that $l(v)$ is increasing in v : higher value jobs can employ more human capital.¹⁵ Two forces drive this result: first, by offering a higher value v , the firm attracts more workers and keeps its workers longer, i.e., its hiring rate is higher and its separation rate is lower. Second, a firm with job value v attracts workers employed at jobs with values lower than v , and their average productivity is higher when v is higher. Therefore, the model predicts that conditional on training/no training, firms that pay more are likely to be larger, have lower turnover rate and more productive workers. Moreover, since higher v firms are more likely to provide training, the likelihood of training is positively correlated with the firm's size, the average tenure and the productivity of its workers.

4.2 Job Offer Distribution

From the standard arguments as in BM, the following lemma must hold.

Lemma 5 *Any equilibrium market distribution of job offers, $F(v)$, is continuous, has a connected support, is bounded below by v_u , and bounded from above by \bar{v} , and $\bar{v} < \frac{p + \delta(1+g)v_u}{1 - (1-\delta-\sigma)(1+g)}$.*

The following propositions are shown to hold in Appendix B.

Proposition 2 *In a market equilibrium, the steady-state job offer distribution is given as follows:*

if $v_u \leq v \leq v^c$

$$F(v) = \frac{\sigma + \lambda + \delta}{\lambda} \left[1 - \sqrt{\frac{p - \theta_0(v)}{p - b}} \right]; \quad (11)$$

¹⁴See Appendix B for the proof.

¹⁵In the case where no firm provides training ($v^c \geq \bar{v}$), only (9) applies. If all firms provide training ($v^c \leq v_u$), then only (10) applies. When we have both non-training firms and training firms, then (9) applies to the former and (10) applies to the latter.

if $v^c \leq v \leq \bar{v}$

$$F(v) = \frac{(\sigma + \lambda + \delta)(1 + g) - g}{\lambda(1 + g)} - \frac{s^c(1 + g) - g}{\lambda(1 + g)} \sqrt{\frac{p - \theta_1(v) - c}{p - \theta_1(v^c) - c}}. \quad (12)$$

When $v^c \geq \bar{v}$, no firm offers training, only (11) applies, and the distribution is the same as in BM. If $v^c \leq v_u$, all firms provide training, and only (12) applies. When there are both non-training firms and training firms, $F(\cdot)$ is specified separately for the two types of firms. However, as shown in the proof, the distribution is still continuous. The distribution given here involves endogenous variables $\theta_1(v^c)$, $\theta_0(v^c)$ and s^c , but it can be expressed in primitives and is unique given parameter values.

Proposition 3 *A market equilibrium exists and is unique.*

Depending on parameter values, the market equilibrium could feature universal training provision, no training at all, or training in only some firms. But for given parameter values, there exists a unique equilibrium.¹⁶ Although all firms earn equal profits in the equilibrium, if parameters are such that firms differ in their training decisions, only those located at the higher end of the $F(v)$ distribution, i.e., those with $v \geq v^c$, will offer training.¹⁷ Within the training/non-training category, firms with higher v offer higher pay rates. When the worker makes a job-to-job transition, she will always move to a job with higher v . Therefore, she will never move from a training job to a non-training job. If she moves to a job with the same training opportunity, the pay rate on the new job must be higher. If she moves from a non-training job to a training job, although she is better off, she might experience a wage cut on transition.^{18,19} Given the optimal strategies of the worker and the firm, I show in Appendix B that the shape of the distribution of human capital among workers features an exponential tail, which guarantees that the expected human capital in the population is finite.

¹⁶Details on how to express the market equilibrium in primitives are in the online appendix.

¹⁷For example, when training cost is neither too high nor too low.

¹⁸A formal discussion of wage cuts over job-to-job transition is available in the online appendix.

¹⁹As shown later, in the general case where firms choose training intensities, firms with higher v will offer more training; and workers will move to jobs with more training than their current jobs.

4.3 Wage Distribution Among Employed Workers

In the online appendix, I derive the joint distribution of the two components of wage: pay rate and human capital. From the joint distribution, I have the following finding:

Proposition 4 *The distribution of pay rate θ conditional on human capital level h , $\Pr(\theta' \leq \theta|h)$, is first-order stochastically increasing in h for any $\theta \geq \theta_1(v^c)$, and is invariant to h for $\theta < \theta_1(v^c)$.²⁰*

The longer a worker stays in the training sector, the higher her human capital level is, due to on-the-job training, and the higher her pay rate is, due to on-the-job search. Therefore, we see a positive correlation between pay rate and human capital for workers employed at a pay rate higher than or equal to $\theta_1(v^c)$, the lowest pay rate in the training sector. However, workers with pay rates lower than $\theta_1(v^c)$ must be in the non-training sector. Other than the newly born workers, the only potential inflow to the non-training sector is unemployed workers, whose jobs have been destroyed with equal probability regardless of their human capital levels. Moreover, all unemployed workers use the same reservation value. Therefore, the pay rate a worker obtains is not correlated with her human capital level, as long as her pay rate is lower than $\theta_1(v^c)$.

Given the distribution of pay rate and human capital, we can now study the distribution of wages. Let $Q(w)$ be the distribution of the wages earned by employed workers, where wage $w = \theta h$. Let $f(\theta, h)$ be the joint density of (θ, h) across employed workers, and we have

$$Q(w) = \sum_{n=0}^{\infty} \int_{\underline{\theta}}^{w/(1+g)^n} f(\theta, (1+g)^n) d\theta.$$

Differentiating with respect to w yields the density of wages:

$$Q'(w) = \sum_{n=0}^{\infty} \frac{1}{(1+g)^n} f\left(\frac{w}{(1+g)^n}, (1+g)^n\right). \quad (13)$$

Two properties of $Q'(w)$ 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, for $w \in [\underline{\theta}, \underline{\theta}(1+g))$, $Q'(w) = f(w, 1)$, i.e., the marginal distribution of the pay rate holding human capital constant at 1. Since the

²⁰See Appendix B for the proof.

marginal distribution of the pay rate is similar to the pay rate distribution in BM, it can be shown that $Q''(w)$ is positive in this region. Therefore, the density of wages earned by employees is increasing when the wage is sufficiently low. Second, consider the right tail of $Q'(w)$. If w becomes large, since the pay rate is bounded above by $\bar{\theta}$, it must be that the human capital level is large, i.e., $h \rightarrow \infty$ as $w \rightarrow \infty$. The conditional distribution of the pay rate $\Pr(\theta' \leq \theta|h)$ converges to $\Pr(\theta' \leq \theta|\infty)$. Moreover, since human capital distribution declines exponentially as shown in the appendix, the distribution of wages must decline at the same rate. Therefore, the wage distribution in this model exhibits a density with an interior mode and a long decreasing right tail.^{21,22}

5 Extension: Endogenous Growth Rate

In the basic model, I assume that the firm's choice of training is binary. In this section, I relax this assumption and allow the firm to choose its training intensity or, equivalently, the growth rate of its employed human capital. This extension will improve the model's capability to capture patterns found in the data.

Assumption: The per-efficiency-unit cost of training that increases human capital at rate $(1 + g)$ is represented by cost function $C(g)$. It satisfies (1) $C(0) = 0$, (2) $C'(\cdot) > 0$, (3) $C''(\cdot) > 0$ and (4) $\lim_{g \rightarrow \bar{g}} C'(g) = \infty$, where \bar{g} is such that $(1 - \sigma)(1 + \bar{g}) = 1$.²³

Endogenizing the choice of g has no effect on the firm's optimal choice of v : given 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 pay rate-training contract problem for a firm that has already promised v :

$$\begin{aligned} \pi(v) &= \max_{g, \theta} \frac{p - \theta - C(g)}{1 - (1 - s(v))(1 + g)} l(v) \\ \text{s.t. } \theta &\geq \theta_0(v) - g(v - \theta_0(v)) \\ g &\geq 0. \end{aligned}$$

²¹A simulation of the wage distribution is available from the author upon request.

²²The model by Burdett, Carrillo-Tudela and Coles (2009) also generates a wage density with a similar shape.

²³Assumptions (1) to (3) define a standard increasing convex cost function. Assumption (4) guarantees that no firm will choose a growth rate that is so high that the worker's value functions might become unbounded.

The first constraint is the promise-keeping constraint: 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 $l(v)$ is constant given v , and the promise-keeping constraint is always binding, the maximization problem is equivalent to

$$\max_{g \geq 0} \frac{p - \theta_0(v) + g(v - \theta_0(v)) - C(g)}{1 - (1 - s(v))(1 + g)}.$$

Proposition 5 *When the choice of g is in the interior, $\partial g / \partial v > 0$, that is, firms that offer higher v also offer a higher growth rate.²⁴*

Now, I show the conditions under which the solution g is in the interior. From the proof of Proposition 5, the first-order condition implies:

$$L(g; v) = s(v)v + (1 - s(v))[p - C(g)] - \theta_0(v) - C'(g)[s(v)(1 + g) - g] \begin{cases} \leq 0 \\ = 0 \text{ if } g > 0 \end{cases}.$$

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

Case (2) If $L(0; \bar{v}^0) \leq 0$, then no firm will offer training, where \bar{v}^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 v^c , such that firms that offer $v \in [v_u, v^c)$ will not offer training, firms that offer $v \in (v^c, \bar{v}]$ will offer training, and the growth rate will increase with v ; firms that offer $v = v^c$ are indifferent between offering and not offering training.

In sum, when the firms are allowed to choose the human capital growth rate under a convex cost function, the optimal growth rate g is non-decreasing in v , and strictly increasing in v when $g > 0$. This is consistent with the result from the basic model in which the firm's choice is restricted to be binary.

6 Summary of Model Predictions

In this section, I will summarize some of the important predictions from my model and compare them with the predictions from other models in the literature. The first subsection focuses on predictions about wages at the worker level. Endogenous training allows for essen-

²⁴See Appendix A for the proof.

tial differences among ex ante identical workers who have the same years of work experience. This yields new insights on wage dispersion and wage persistence. The second subsection focuses on predictions about wages at the firm level. My model predicts a positive correlation between wage growth and tenure. It also offers a new explanation for the systematic difference in within-firm wage dispersion across firms. The last subsection summarizes the predictions on training.

6.1 Wage Dispersion Across Workers and Wage Persistence

Since wage $w = \theta h$, the variance of log wages among ex ante identical workers can be written as

$$\text{var}(\ln(w)) = \text{var}(\ln(\theta)) + \text{var}(\ln(h)) + 2\text{cov}(\ln(\theta), \ln(h)). \quad (14)$$

In my model, wage dispersion across workers exists because 1) identical workers are paid differently, 2) workers differ in their productivity ex post, and 3) to magnify wage dispersion, there is a positive correlation between pay rate earned and human capital level. Burdett, Carrillo-Tudela and Coles (2009) also derive a decomposition of wage dispersion similar to that in (14). The crucial difference between their model and my model is the correlation between years of work experience and human capital. With exogenous learning-by-doing, the correlation is perfect: after controlling for experience and other worker characteristics, wages differ only because some workers are luckier and get better draws of pay rates than others (reason 1)).²⁵ In my model, due to different training levels across firms, workers with the same years of work experience may have different human capital, and all three reasons for wage dispersion still exist among observationally equivalent workers.

In the standard on-the-job search model as BM, observed wages out of unemployment are independent of the workers' wage histories: a lucky draw enables the worker to enjoy higher wages only until she is laid off. Conditional on unemployment, history does not matter for a worker's future wage profile. In my model, wage persistence exists even after job destruction because higher wage earners are, on average, more productive. A worker's work history, not only her years of work experience, should always matter for her wage profile. In particular, past training should have persistent effects on a worker's wage. This prediction is supported by the data. For example, using information on the entire work histories

²⁵As a result, the residual wage distribution will have an increasing and convex density as in BM.

(including unemployment spells) of young people, Lynch (1992) finds that past training has a significantly positive effect on a worker's wage at the current job, after controlling for years of experience, tenure and other worker characteristics.

6.2 Wage Growth, Tenure and Within-Firm Wage Dispersion

Because firms differ in their training decisions, this model predicts that wage growth rates differ across firms. Firms with higher wage growth (more training) are firms that provide higher job values to workers, and their workers will have longer tenure. Interestingly, this prediction differs from the predictions of some other models. For example, in Burdett, Carrillo-Tudela and Coles (2009), wage grows because of learning-by-doing and wage growth rates are the same across firms. The prediction from my (extended) model can be supported by the data. For example, Bronars and Famulari (1997) find that there is significant dispersion in wage growth across employers, conditional on worker characteristics, and that employers with higher average current tenure tend to have faster wage growth.

Empirical studies have found that within-firm wage dispersion is higher in firms that offer better jobs. For example, Postel-Vinay and Robin (2002) find a positive correlation between within-firm wage variance and within-firm mean wage. In their model, this is because more productive firms have greater monopsony power and offer lower wages to unemployed workers than less productive firms do. At the same time, they poach the employees of the less productive firms by offering them a better future value. My model provides a different explanation of this empirical fact. First, given the same training intensity, a firm offering a higher pay rate will keep its worker longer. This leads to a greater difference in tenure, hence a greater difference in human capital levels and wages, between senior workers and junior workers. Second, to magnify this effect, better jobs offer higher growth rates. Holding the difference in tenure fixed, the difference in human capital levels between a senior worker and a junior worker is greater when the growth rate is higher.

6.3 Training

The existence of firm-funded general training suggests the existence of frictions in the labor market that restrict worker mobility. For example, Acemoglu and Pischke (1998) provide evidence of firm-funded general training and use information asymmetry to explain it. But

information asymmetry essentially shuts down job-to-job transitions. However, Fallick and Fleischman (2001), for instance, estimate that in the United States in 1999, on average four million workers changed employers from one month to the next (about 2.7 percent of employment)—more than twice the number who transitioned from employment to unemployment. My model draws on search frictions and predicts that firm-funded general training and frequent job-to-job transitions coexist.

As another contribution, my model generates endogenous correlations between training and other firm characteristics. Acemoglu and Pischke (1998) focus on why firms train their workers but are silent about the characteristics of training firms. In Moen and Rosen (2004), training firms are *ex ante* better at training. In my model, firms with training provide jobs that are of higher values to the worker; therefore, training firms typically have lower turnover rates, are larger and have more productive employees. Such correlations are found to be true by previous empirical studies (see Bishop (1996) for a comprehensive literature review).

My model also predicts that when training is provided, the firm and the worker share both the cost and the benefit of training. Bishop (1996) reviews eight different types of evidence and confirms that cost sharing exists. In his review, Mincer (1989) summarizes that the estimated effects in terms of earnings received with an additional year of training range from 4.4% to 11%. Barron et al. (1997) and Bishop (1991) find that employers claim training is valuable with other firms, but their measures of productivity growth associated with training exceed wage growth by a factor of ten. These findings suggest that both the worker and the firm benefit from training. However, in line with the literature on firm training (Stevens (1994)), the provision of general training is lower than the socially optimal level because of the externality of training.²⁶

7 Conclusion

This paper develops a model in which *ex ante* identical firms decide, in addition to their pay rates, whether or not to provide costly training that improves their workers' general skills. Combining on-the-job search and endogenous training, the model explains the coexistence of firm-sponsored general training and frequent job-to-job transitions. It helps to explain the correlations between various firm characteristics found in the data. It also yields new

²⁶A formal discussion of the efficiency of training provision can be found in the online appendix.

insights on wage dynamics and wage dispersion.

Search frictions enable the firm to share the rent from its workers' general human capital accumulated via training. Therefore, the firm and the worker share the cost of training. Although firms are ex ante homogeneous, they differ in their training decisions. Firms with more training make offers that yield higher values to workers. Therefore, they are likely to be larger and have lower turnover and more productive workers. Wage growth differs across firms, and it is positively correlated with the within-firm average tenure. Endogenous training also helps to explain the positive correlation between within-firm wage dispersion and within-firm mean wage.

Although workers are ex ante identical, wage dispersion exists because 1) identical workers are paid differently, 2) workers differ in their productivity ex post, and 3) there is a positive correlation between pay rate earned and human capital level. By decomposing the wage into human capital and pay rate per unit of human capital, the model yields a distribution of wages with a long declining right tail, as observed in the data. Due to endogenous training, observationally equivalent workers may differ in their productivity. A worker's entire work history, not only her years of work experience, matters for her future wage profile.

There are several interesting extensions to the current model. The most straightforward extension is to introduce worker heterogeneity. When workers have different productivity ex ante, worker fixed effects have to be considered in the decomposition of wage dispersion. One can also introduce firm heterogeneity either through productivity or through training cost. For example, firms with lower training cost would provide more training, which implies higher wage growth. Extensions along these lines will be useful in empirical applications.

A more difficult extension is to allow more flexible pay rate structures.²⁷ One way is to allow firms to backload wages, as in Burdett and Coles (2003) and Stevens (2004). In this case, a contract would consist of a wage-tenure profile along with training intensity. In firms with general training, workers with longer tenure are, on average, more productive, and hence more valuable for the firm. Training firms, therefore, will have an even greater incentive to backload wages and reward loyalty than non-training firms. Within a training firm, junior workers earn less than senior workers not only because they are less productive, but also because their human capital is "priced down" (they pay more for their training).

²⁷I thank an anonymous referee for pointing this out.

References

- [1] Acemoglu, Daron. and Jorn-Steffen Pischke. 1998. "Why do firms train? Theory and evidence," *Quarterly Journal of Economics*, Vol. 113, 79-119.
- [2] Bagger, Jesper, Francois Fontaine, Fabien Postel-Vinay and Jean-Marc Robin. 2006. "A Feasible Equilibrium Search Model of Individual Wage Dynamics with Experience Accumulation," working paper, University of Bristol.
- [3] Barron, John, Mark Berger, and Dan Black. 1997. *On-the-job Training*. Kalamazoo: W. E. Upjohn Institute for Employment Research.
- [4] Becker, Gary. S. 1962. "Investment in human capital: a theoretical analysis," *Journal of Political Economy*, Vol. 70, 9-49.
- [5] Bishop, John H. 1991. "On-the-job training of new hires," in (David Stern and Joseph M. M. Ritzen, eds.) *Market Failure in Training? New Economic Analysis and Evidence on Training of Adult Employees*. Berlin: Springer Verlag, 61-98.
- [6] Bishop, John H. 1996. "What We Know About Employer-Provided Training: A Review of Literature," CAHRS Working Paper 96-109.
- [7] Bronars, Stephen G. and Melissa Famulari. 1997. "Wage, Tenure, and Wage Growth Variation within and across Establishments," *Journal of Labor Economics*, Vol. 15, No. 2, 285-317
- [8] Burdett, Kenneth, Carlos Carrillo-Tudela and Melvyn Coles. 2009. "Human Capital Accumulation and Labor Market Equilibrium," Discussion Paper No. 4215, IZA, Bonn.
- [9] Burdett, Kenneth and Melvyn Coles. 2003. "Equilibrium Wage-Tenure Contracts," *Econometrica*, Vol. 71, No. 5, 1377-1404
- [10] Burdett, Kenneth and Dale Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, Vol. 39, No. 2, 257-273
- [11] Connolly, Helen and Peter Gottschalk. 2008. "Wage Cuts as Investment in Future Wage Growth," *LABOUR: Review of Labour Economics and Industrial Relations* 22, 1-22.

- [12] Fallick, Bruce C., and Charles A. Fleischman. 2001. "The Importance of Employer-to-Employer Flows in the U.S. Labor Market," Finance and Economics Discussion Series, no. 2001-18, Bd. Governors, Fed. Reserve System, Washington, DC.
- [13] Katz, Eliakim and Adrian Ziderman. 1990. "Investment in General Training: The Role of Information and Labour Mobility," *Economic Journal*, Vol. 100, 1147-58.
- [14] Lynch, Lisa M. 1992. "Private-Sector Training and the Earnings of Young Workers," *American Economic Review*, Vol. 82, No. 1, 299-312.
- [15] Lynch, Lisa M. and Sandra Black. 1998. "Beyond the Incidence of Employer-Provided Training," *Industrial and Labor Relations Review*, Vol. 52, No. 1, 64-81.
- [16] Mincer, Jacob. 1989. "Human Capital and the Labor Market: A Review of Current Research," *Educational Researcher*, May 1989, 27-34.
- [17] Moen, Espen R. and Asa Rosen. 2004. "Does Poaching Distort Training?" *Review of Economic Studies*, Vol. 71, No. 4, 1143-1162.
- [18] Postel-Vinay, Fabien and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, Vol. 70, No. 6, 2295-2350.
- [19] Quercioli, Elena. 2005. "Training, Turnover and Search," *International Economic Review*, Vol. 46, No. 1, 133-143.
- [20] Rosen, Sherwin. 1972. "Learning and Experience in the Labor Market," *Journal of Human Resources*, Vol. 7, No. 3, 326-342
- [21] Rubinstein, Yona and Yoram Weiss. 2007. "Post Schooling Wage Growth: Investment, Search and Learning," in Eric A. Hanushek and Finis Welch. eds., *Handbook of Economics of Education*, Vol. 1, Amsterdam: Elsevier Science B.V.
- [22] Stevens, Margaret. 1994. "A Theoretical Model of On-the-Job Training with Imperfect Competition," *Oxford Economic Papers*, Vol. 46, 537-562.
- [23] Stevens, Margaret. 2004. "Wage-Tenure Contracts in a Frictional Labour Market: Firms Strategies for Recruitment and Retention," *Review of Economic Studies*, Vol. 71, No.2, 535-551.

Appendix A

A1. Proof for Lemma 4

Proof. Rearrange terms

$$B(v) = gv + g \frac{(1-s(v))p - \theta_0(v)}{s(v)}.$$

$$\begin{aligned} B'(v) &= g + \frac{g}{[s(v)]^2} \{(-ps'(v) - \theta'_0(v))s(v) - [(1-s(v))p - \theta_0(v)]s'(v)\} \\ &= \frac{-g}{[s(v)]^2} [p - \theta_0(v)]s'(v) > 0, \end{aligned}$$

because $\theta'_0(v) = s(v)$, $s'(v) = -\lambda F'(v) < 0$ and $p - \theta_0(v) > 0$. ■

A2 Proof for Proposition 5

Proof. FOC with respect to g :

$$\frac{[v - \theta_0(v) - C'(g)][1 - (1-s(v))(1+g)] + [p - \theta_0(v) + g(v - \theta_0(v)) - C(g)](1-s(v))}{[s(v)(1+g) - g]^2} \begin{cases} \leq 0 \\ = 0 \text{ if } g > 0 \end{cases}.$$

This is equivalent to

$$s(v)v + (1-s(v))[p - C(g)] - \theta_0(v) - C'(g)[s(v)(1+g) - g] \begin{cases} \leq 0 \\ = 0 \text{ if } g > 0 \end{cases}, \quad (15)$$

Assume interior solution, define

$$L(g; v) = s(v)v + (1-s(v))[p - C(g)] - \theta_0(v) - C'(g)[s(v)(1+g) - g]. \quad (16)$$

$$\frac{\partial L}{\partial g} = -C''(g)[s(v)(1+g) - g] < 0 \quad \text{by convexity of } C(\cdot);$$

$$\frac{\partial L}{\partial v} = s'(v)[v - p + C(g) - C'(g)(1+g)].$$

From $s(v)v + (1 - s(v))[p - C(g)] - \theta_0(v) - C'(g)[s(v)(1 + g) - g] = 0$,

$$[v - p + C(g) - C'(g)(1 + g)] = \frac{v - \theta_0(v) - C'(g)}{(1 - s(v))}.$$

From $[v - \theta_0(v) - C'(g)][1 - (1 - s(v))(1 + g)] + [p - \theta_0(v) + g(v - \theta_0(v)) - C(g)](1 - s(v)) = 0$,

$$\frac{v - \theta_0(v) - C'(g)}{(1 - s(v))} = -\frac{[p - \theta_0(v) - C(g)] + g(v - \theta_0(v))}{[1 - (1 - s(v))(1 + g)]} < 0,$$

since the future value of a job offer is positive, and that profit is positive in equilibrium.

Therefore, $\frac{\partial L}{\partial v} > 0$, and $\frac{\partial g}{\partial v} > 0$. ■

Appendix B: Market Equilibrium Analysis

B1. Human Capital Distribution

As in standard on-the-job search model, the steady-state unemployment level $u = \frac{\delta + \sigma}{\delta + \lambda + \sigma}$. The steady-state employment value distribution $G(v)$ is given by $G(v) = \frac{(\delta + \sigma)F(v)}{\delta + \sigma + \lambda(1 - F(v))}$. Notations: $F^c = F(v^c)$ 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 B1. *In the steady state, the distribution of human capital is given by the following: For the lowest human capital level $h = 1$,*

$$D(1) = \frac{\sigma(\sigma + \delta + \lambda)}{s^c(\sigma + \lambda) - \delta\lambda F^c}, \quad (17)$$

$$uD^u(1) = \frac{\sigma s^c}{s^c(\sigma + \lambda) - \delta\lambda F^c}. \quad (18)$$

For all $n \geq 1$,

$$D[(1 + g)^n] = D(1) \frac{s^c \lambda (\sigma + \delta + \lambda) (1 - F^c)}{s^c (\sigma + \lambda) - \delta \lambda F^c} y^{n-1}, \quad (19)$$

$$uD^u[(1 + g)^n] = D(1) \frac{s^c \lambda \delta (1 - F^c)}{s^c (\sigma + \lambda) - \delta \lambda F^c} y^{n-1}, \quad (20)$$

where

$$y = \frac{\delta\lambda(\sigma + \delta + \lambda)(1 - F^c)}{s^c(\sigma + \lambda) - \delta\lambda F^c} + 1 - \sigma - \delta. \quad (21)$$

And for any $h \notin \{(1 + g)^n\}_{n=0}^\infty$, $D(h) = 0$. The mean human capital in the whole market in the steady state exists and is finite:

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

and the mean human capital among unemployed workers is:

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

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 non-training sector, i.e., $(1 - u)G^c = \sum_h \Pr(v \leq v^c, h)$, 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,

$$(\lambda + \sigma)uD^u(1) = \sigma + \delta(1 - u)G^cD^0(1). \quad (22)$$

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,

$$(\lambda + \sigma)uD^u(h) = (1 - u)G^cD^0(h)\delta + (1 - u)(1 - G^c)D^1(h)\delta. \quad (23)$$

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 $d = 0$, only unemployed workers will join this group if they find a job in this sector:

$$s^c(1 - u)G^cD^0(h) = \lambda F^c u D^u(h). \quad (24)$$

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.

$$\begin{aligned} (1 - u)(1 - G^c)D^1(h) &= (1 - u)(1 - G^c)D^1\left(\frac{h}{1 + g}\right)(1 - \sigma - \delta) \\ &\quad + uD^u\left(\frac{h}{1 + g}\right)\lambda(1 - F^c) + (1 - u)G^cD^0\left(\frac{h}{1 + g}\right)\lambda(1 - F^c). \end{aligned}$$

In the whole economy:

$$D(1) = uD^u(1) + (1 - u)G^cD^0(1), \quad (25)$$

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

$$D(h) = uD^u(h) + (1 - u)G^cD^0(h) + (1 - u)(1 - G^c)D^1(h). \quad (26)$$

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 are as follows: for $h \in \{(1 + g)^n\}_{n=1}^\infty$,

$$\begin{aligned} uD^u(h) &= \frac{\delta s^c}{s^c(\lambda + \sigma) - \lambda\delta F^c}(1 - u)(1 - G^c)D^1(h), \\ (1 - u)G^cD^0(h) &= \frac{\lambda\delta F^c}{s^c(\lambda + \sigma) - \lambda\delta F^c}(1 - u)(1 - G^c)D^1(h). \end{aligned}$$

$$\begin{aligned}
uD^u(1) &= \frac{\sigma s^c}{s^c(\lambda + \sigma) - \lambda \delta F^c}, \\
(1 - u)G^c D^0(1) &= \frac{\lambda F^c}{s^c} uD^u(1).
\end{aligned}$$

Solving the equations (22) to (26) gives us the distribution as specified in the proposition. 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, $\lim_{n \rightarrow \infty} D[(1 + g)^n] = 0$ because $y \in (0, 1)$.

The mean of human capital is

$$\begin{aligned}
E(h) &= \sum_{n=0}^\infty (1 + g)^n D[(1 + g)^n] \\
&= D(1) \left\{ 1 + (1 + g) \frac{s^c \lambda (\sigma + \delta + \lambda) (1 - F^c)}{s^c (\sigma + \lambda) - \delta \lambda F^c} \sum_{n=1}^\infty [y(1 + g)]^{n-1} \right\}.
\end{aligned}$$

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 $uD^u(\cdot)$ and $D(\cdot)$, one can get the expression of the average human capital among unemployed workers.²⁸ ■

B2. Joint Distribution of Job Values and Human Capital

Proposition B2. *The measure of workers with human capital h who are employed at jobs with values no greater than v is given by:*

Case 1. $v < v^c$

$$\Pr(v' \leq v, h = (1 + g)^n) = \frac{\lambda F(v)}{s(v)} uD^u[(1 + g)^n] \text{ for } n \geq 0, \quad (27)$$

where $s(v) = \delta + \sigma + \lambda(1 - F(v))$ is the separation rate for firm v .

Case 2. $v \geq v^c$

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

²⁸More detailed proof is available from the author on request.

for $n \geq 1$,

$$\begin{aligned} \Pr(v' \leq v, h = (1+g)^n) &= \frac{\lambda F^c}{s^c} u D^u[(1+g)^n] \\ &+ \frac{\lambda(\sigma + \lambda + \delta)(F(v) - F^c)}{s^c} \sum_{m=1}^n (1 - s(v))^{m-1} u D^u[(1+g)^{n-m}]. \end{aligned}$$

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

Case 2 $v \geq v^c$: $\Pr(v' \leq v, [(1+g)^n]) = \Pr(v' \leq v^c, [(1+g)^n]) + \Pr(v^c \leq v' \leq v, [(1+g)^n])$. 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 $\Pr(v^c \leq v' \leq v, [(1+g)^n])$ 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 $v' \in [v^c, v]$. 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 $\Pr(v^c \leq v' \leq v, (1+g)^n)$, because workers with $(v^c \leq v' \leq v, (1+g)^n)$ would either retire, or get laid off, or get a job better than v , or if they stay in $(v^c \leq v' \leq v)$, they would have human capital $(1+g)^{n+1}$. Therefore,

$$\begin{aligned} &\Pr(v^c \leq v' \leq v, (1+g)^n) \\ &= \lambda(F(v) - F^c) \{u D^u[(1+g)^{n-1}] + (1-u) G^c D^0[(1+g)^{n-1}]\} + (1-s(v)) \Pr(v^c \leq v' \leq v, (1+g)^{n-1}) \\ &= \frac{\lambda(F(v) - F^c)(\sigma + \lambda + \delta)}{s^c} \sum_{m=1}^n (1 - s(v))^{m-1} u D^u[(1+g)^{n-m}], \end{aligned}$$

where the last equality follows from the relationship between $u D^u(h)$ and $(1-u) G^c D^0(h)$.

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,

$$\Pr(v' \leq v, h = 1) = \frac{\lambda F^c}{s^c} u D^u(1).$$

■

The joint distribution of job values and human capital among employed workers is $\Pr(v' \leq v, h = (1+g)^n) / (1-u)$.

B3. Proof for Proposition 4

Corollary B1 *The distribution of job values v conditional on human capital level h , $\Pr(v' \leq v|h)$, is first order stochastically increasing in h for any $v \geq v^c$, and is invariant to h for $v < v^c$.*

Proof. Part I. 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 :

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

Case 1. $v < v^c$:

$$\begin{aligned} \Pr(v' \leq v|h = 1) &= \frac{\lambda F(v)}{s(v)} \frac{uD^u(1)}{D(1) - uD^u(1)} \\ &= \frac{F(v)s^c}{s(v)F^c}; \end{aligned}$$

for $n \geq 1$

$$\begin{aligned} \Pr(v' \leq v|h = (1+g)^n) &= \frac{\lambda F(v)}{s(v)} \frac{uD^u((1+g)^n)}{[D((1+g)^n) - uD^u((1+g)^n)]} \\ &= \frac{\lambda F(v)}{s(v)} \frac{\delta}{(\sigma + \lambda)}, \end{aligned}$$

where the second equality follows from the relationship between $uD^u()$ and $D()$. In this case, the conditional distribution is invariant to h .

Case 2. $v \geq v^c$

$$\Pr(v' \leq v|h = 1) = 1.$$

If an employed worker has human capital 1, she must be employed in the non-training sector, hence $v' \leq v^c \leq v$ for sure.

For $n \geq 1$, I have shown earlier that

$$\begin{aligned} \Pr(v' \leq v, h = (1+g)^n) &= \Pr(v' \leq v^c, (1+g)^n) + \Pr(v^c \leq v' \leq v, [(1+g)^n]) \\ &= \frac{\lambda F^c}{s^c} uD^u[(1+g)^n] \\ &\quad + \lambda(F(v) - F^c) \sum_{m=1}^n (1-s(v))^{m-1} \frac{(\sigma + \lambda + \delta)}{s^c} uD^u[(1+g)^{n-m}]. \end{aligned}$$

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

$$\frac{(\sigma + \lambda + \delta)}{s^c} uD^u(1) = D(1).$$

For $n - m \geq 1$,

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

where

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

Notice that $y > 1 - \sigma - \delta > 1 - s(v)$.

$$\begin{aligned} \Pr(v' \leq v, h = (1+g)^n) & \tag{28} \\ &= \frac{\lambda F^c}{s^c} uD^u[(1+g)^n] \\ &\quad + \lambda(F(v) - F^c) D(1) \left\{ \frac{\lambda \delta (\lambda + \sigma + \delta) (1 - F^c) y^{n-2}}{s^c (\lambda + \sigma) - \lambda \delta F^c} \sum_{m=1}^{n-1} \left(\frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\}. \end{aligned}$$

For $n \geq 1$

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

After some algebraic manipulation,

$$\begin{aligned}
\Pr(v' \leq v|h = (1+g)^n) & & (29) \\
&= \frac{\Pr(v' \leq v^c, (1+g)^n) + \Pr(v^c \leq v' \leq v, (1+g)^n)}{[D((1+g)^n) - uD^u((1+g)^n)]} \\
&= \frac{\lambda F^c \delta}{s^c(\sigma + \lambda)} \\
&\quad + \frac{\lambda \delta (\lambda + \delta + \sigma)(F(v) - F^c)}{s^c(\lambda + \sigma)(y + s(v) - 1)} \\
&\quad + \frac{(F(v) - F^c)}{s^c(\lambda + \sigma)} \left(\frac{1 - s(v)}{y}\right)^{n-1} \left\{ \frac{[s^c(\lambda + \sigma) - \lambda \delta F^c]}{1 - F^c} - \frac{\lambda \delta (\lambda + \delta + \sigma)}{y - (1 - s(v))} \right\}.
\end{aligned}$$

Part II. From (29), 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 $\frac{1-s(v)}{y} < 1$. The term in the curly bracket is equal to

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

The denominator is positive, and the numerator is equal to

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

As a result, $\Pr(v' \leq v|h = (1+g)^n)$ is first order stochastically increasing in h for any $v \geq v^c$.

■

Given the relationship between (θ, h) distribution and (v, h) distribution, one can see that $(\theta|h)$ distribution must preserve the first order stochastic dominance property of $(v|h)$ when the pay rate θ is paid by some training firms.²⁹

B4. Proof for Claim 1

Proof. If $v_u \leq v \leq v^c$,

$$\begin{aligned}
\frac{l(v)}{\lambda} &= uE(h|u) + \sum_{n=0}^{\infty} (1+g)^n \Pr(v' < v, (1+g)^n) \\
&= \frac{\sigma + \lambda + \delta}{s(v)} uE(h|u)
\end{aligned}$$

²⁹See online appendix for more details.

If $v^c < v \leq \bar{v}$,

$$\begin{aligned}
\frac{l(v)}{\lambda} &= uE(h|u) + \sum_{n=0}^{\infty} (1+g)^n \Pr(v' < v, (1+g)^n) \\
&= uE(h|u) + \frac{\lambda F^c}{s^c} uE(h|u) + \lambda(F(v) - F^c)D(1) * \\
&\quad \sum_{n=0}^{\infty} (1+g)^n \left\{ \frac{\lambda \delta (\lambda + \sigma + \delta) (1 - F^c) y^{n-2}}{s^c (\lambda + \sigma) - \lambda \delta F^c} \sum_{m=1}^{n-1} \left(\frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\},
\end{aligned}$$

from (28). Define X as the constant term $\lambda \delta (\lambda + \sigma + \delta) (1 - F^c) / [(\lambda + \sigma) s^c - \delta \lambda F^c]$,

$$\begin{aligned}
&\sum_{n=0}^{\infty} (1+g)^n \left\{ X y^{n-2} \sum_{m=1}^{n-1} \left(\frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\} \\
&= \frac{(1+g)}{[(1+g)s(v) - g]} \left[\frac{(1+g)X}{[1 - (1+g)y]} + 1 \right].
\end{aligned}$$

Plug in the definition of y and X , using the relationship between $D(1)$ and $uD^u(1)$,

$$\begin{aligned}
&D(1) \sum_{n=0}^{\infty} (1+g)^n \left\{ X y^{n-2} \sum_{m=1}^{n-1} \left(\frac{1 - s(v)}{y} \right)^{m-1} + (1 - s(v))^{n-1} \right\} \\
&= \frac{(\delta + \sigma + \lambda)(1+g)}{[(1+g)s(v) - g]s^c} uE(h|u).
\end{aligned}$$

$$\begin{aligned}
\frac{l(v)}{\lambda} &= \frac{\lambda + \sigma + \delta}{s^c} uE(h|u) + \frac{\lambda(F(v) - F^c)(\delta + \sigma + \lambda)(1+g)}{[(1+g)s(v) - g]s^c} uE(h|u) \\
&= \frac{(\delta + \sigma + \lambda)[(1+g)s^c - g]}{[(1+g)s(v) - g]s^c} uE(h|u).
\end{aligned}$$

■

B5. Proof for Proposition 2

Proof. If $v_u \leq v \leq v^c$, $d = 0$ 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:

$$\frac{(p - \theta_0(v))(\sigma + \lambda + \delta)}{s(v)^2} = \frac{(p - b)}{\sigma + \lambda + \delta}, \quad (30)$$

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 $v^c < v \leq \bar{v}$, $d = 1$ and

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

Equal profit condition $\pi(d = 1; v) = \pi(d = 1; v^c)$ implies

$$\frac{(p - \theta_1(v) - c)}{[s(v)(1 + g) - g]^2} = \frac{(p - \theta_1(v^c) - c)}{[s^c(1 + g) - g]^2} \quad (31)$$

and the result follows from the relationship between $s(v)$ and $F(v)$. Using the relationship between $\theta_1(v^c)$, $\theta_0(v^c)$ and c , one can prove the continuity of $F(\cdot)$ by showing $F(v^c)$ in (11) is the same as $F(v^c)$ in (12). ■