Training, Search and Wage Dispersion
Technical Appendix

Chao Fu*
University of Wisconsin-Madison

October, 2010

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, pay rate-training contract, wage dispersion, wage dynamics

JEL codes: J64, J24, J31

Appendix A

A1. Proof for Lemma 1

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

$$\begin{aligned} v &= \theta_0 + \lambda \int_v^{\overline{v}} v' dF(v') + \lambda F(v) v + \delta v_u + (1 - \delta - \lambda - \sigma) v \\ &= \theta_0 + (1 - \delta - \lambda - \sigma) v + \lambda \overline{v} + \delta v_u - \lambda \int_v^{\overline{v}} F(v') dv', \end{aligned}$$

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

$$\theta_0 = (\delta + \lambda + \sigma) v - \lambda \overline{v} - \delta v_u + \lambda \int_v^{\overline{v}} F(v') dv'.$$

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

Take the derivative with respect to v

$$\begin{aligned}\frac{d\theta_0}{dv} &= (\delta+\lambda+\sigma)-\lambda F(v)\\ &= s(v)\\ &> 0,\end{aligned}$$

i.e., θ_0 is strictly monotone in v, hence the mapping from v to θ_0 is one-to-one. Therefore, $\theta_0(v)$ is a well defined function.
Case(2) $d = 1$: similar arguments will yield the result that $\theta_1(v)$ is a well-defined function. ■

A2 Proof for Lemma 2 (The relationship between $\theta_1(v)$ and $\theta_0(v)$, from the worker's perspective)

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

$$\begin{aligned}v &= \theta_0(v)+\lambda\int_{\underline{v}}^{\overline{v}}\max\{v,v'\}dF(v')+\delta v_u+(1-\delta-\lambda-\sigma)v\\ &= \theta_0(v)+v-\theta_0(v).\end{aligned}$$

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

$$\begin{aligned}v &= \theta_1+(1+g)\{\lambda\int_{\underline{v}}^{\overline{v}}\max[v,v']dF(v')+\delta v_u+(1-\delta-\lambda-\sigma)v\}\\ &= \theta_1+(1+g)(v-\theta_0(v)).\end{aligned}$$

Equating these two expressions to solve for θ_1, one gets:

$$\theta_1(v) = \theta_0(v) - g(v - \theta_0(v)).$$

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

$$\begin{aligned}Gap'(v) &= g(1-\theta_0'(v))\\ &= g[1-s(v)]\\ &> 0.\end{aligned}$$

■

A3. Proof for Lemma 4 (The joint benefit from training is increasing in v)

Proof. Rearrange terms

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

Take derivative

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

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

A4. Wage cuts over job-to-job transition

The following notations will be used: Let v_0^l be the value of a job without training such that $\theta_0(v_0^l) = \theta_1(v^c)$, that is, v_0^l is the lowest v level such that there is still a probability of wage cuts on transition. Let $v_1^*(v_0)$ 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 v_0 i.e., $\theta_0(v_0) = \theta_1(v_1^*(v_0))$, and let $v_1^{**}(v_0) = \max\{v_1^*(v_0), \overline{v}\}$.[1]

Proposition A1. *For each worker who works on non-training jobs with value $v \in [v_0^l, v^c]$, there is a chance of wage cuts on transition, and this probability is given by*

$$
\Pr(\text{wage cuts on transition}) = \frac{\int_{v^c}^{v_1^{**}(v_0)} dF(v')}{\int_{v_0}^{\overline{v}} dF(v')} \qquad \text{for } v_0 \in [v_0^l, v^c]. \tag{1}
$$

Proof. A worker hired at a non-training job with value $v_0 \in [v_0^l, v^c]$ will move to a new job if the new job is of value $v' \geq v_0$, which happens with probability $\lambda \int_{v_0}^{\overline{v}} dF(v')$. Among these possible transitions, wage cuts happen when she transits to training job with value between v^c and $v^{**}(v_0)$, which happens with probability $\lambda \int_{v^c}^{v^{**}(v_0)} dF(v')$. The ratio of these two probabilities gives the conditional probability of wage cuts given that a transition is observed. ■

A5 Proof for Proposition 5 (Firms that offer higher v also offer higher growth rate)

Proof. FOC with respect to g:

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

[1] The values v_0^l and $v_1^*(v_0)$ are well defined because of monotonicity of wage functions and the continuity of the job offer distribution F; the latter will be confirmed in Appendix B.

This is equivalent to

$$[v-\theta_0(v)-C'(g)][1-(1-s(v))(1+g)]+[p-\theta_0(v)+g(v-\theta_0(v))-C(g)](1-s(v)) \begin{cases} \leq 0 \\ = 0 \text{ if } g > 0 \end{cases} . \tag{3}$$

Rearranging and collecting terms, one gets

$$s(v)v+(1-s(v))[p-C(g)]-\theta_0(v)-C'(g)[s(v)(1+g)-g] \begin{cases} \leq 0 \\ = 0 \text{ if } g > 0 \end{cases} , \tag{4}$$

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 v value. Define

$$L(g;v) = s(v)v + (1-s(v))[p-C(g)] - \theta_0(v) - C'(g)[s(v)(1+g)-g]. \tag{5}$$

Take partial derivative:

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

$$\begin{aligned} \frac{\partial L}{\partial v} &= s(v) + s'(v)v - s'(v)(p - C(g)) - \theta_0'(v) - C'(g)s'(v)(1+g) \\ &= s(v) + s'(v)[v - p + C(g) - C'(g)(1+g)] - s(v) \\ &= s'(v)[v - p + C(g) - C'(g)(1+g)], \end{aligned}$$

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

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

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, $\frac{\partial L}{\partial v} > 0$, and $\frac{\partial g}{\partial v} = -\frac{\partial L}{\partial v} / \frac{\partial L}{\partial g} > 0$. ■

Appendix B: Market Equilibrium Analysis

B1. The steady-state unemployment level is

$$u = \frac{\delta + \sigma}{\delta + \lambda + \sigma}.$$

The steady-state employment value distribution $G(v)$ can be derived from offer distribution $F(v)$ by equalizing the flow-in and flow-out of $G(v)$, which delivers the following:

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

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:

$$\begin{aligned} u\lambda &= (1-u)\delta \\ u &= \frac{\delta + \sigma}{\lambda + \delta + \sigma}. \end{aligned}$$

The flow-in of workers hired at value no greater than v comes only from the unemployed, while workers flow out of $G(v)$ because of job destruction, retirement, or the arrival of a job offer from $(1 - F(v))$. Therefore,

$$\begin{aligned} \lambda u F(v) &= (1-u)G(v)[\delta + \sigma + \lambda(1 - F(v))] \\ G(v) &= \frac{\lambda u F(v)}{(1-u)[\delta + \sigma + \lambda(1 - F(v))]} \\ &= \frac{(\delta + \sigma)F(v)}{[\delta + \sigma + \lambda(1 - F(v))]}. \end{aligned}$$

■

B2. Proof for Lemma 5 (Property of F distribution)

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

$$\overline{v} < \frac{p + \delta(1+g)v_u}{(\delta + \sigma)(1+g) - g}. \tag{6}$$

Given $p > b > 0$, in any equilibrium, firm's profit will be strictly positive: 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. Since no worker will accept an offer lower than v_u, in equilibrium, no firm will offer v lower than v_u. Moreover, the hiring rate and the separation rate for the firm that offers the lowest $\underline{v}$ are independent of the exact value of $\underline{v}$ as long as $\underline{v}$ is no less than v_u. Therefore, the lower bound for the equilibrium offer distribution $\underline{v}$ will be v_u.

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

$$\begin{aligned}\overline{\overline{v}} &= p + (1-\delta-\sigma)(1+g)\overline{\overline{v}} + \delta(1+g)v_u \\ \overline{\overline{v}} &= \frac{p+\delta(1+g)v_u}{(\delta+\sigma)(1+g)-g}.\end{aligned}$$

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

$$\overline{v} < \overline{\overline{v}}.$$

Step 2: $F(v)$ has no mass point.
Suppose not, there exists $v_0 \in [v_u, \overline{v}]$, such that $F(v_0) > F^-(v_0)$, this would imply that $\Pr(v' \leq v_0, h) \geq \Pr(v' < v_0, h)$ for all h, and with strict inequality for some h.

$$\begin{aligned}&\lim_{\epsilon\to 0^+} \pi(d, v_0+\epsilon) \\ =& \lim_{\epsilon\to 0^+} \frac{p-\theta_d(v_0+\epsilon)-dc}{1-(1-\delta-\sigma-\lambda(1-F(v_0+\epsilon))(1+dg)}\lambda[uE(h|u)+\sum_h h\Pr(v'<v_0+\epsilon,h)] \\ =& \frac{p-\lim_{\epsilon\to 0^+}\theta_d(v_0+\epsilon)-dc}{1-(1-\delta-\sigma-\lambda(1-\lim_{\epsilon\to 0^+}F(v_0+\epsilon))(1+dg)}\lambda[uE(h|u)+\sum_h h\lim_{\epsilon\to 0^+}\Pr(v'<v_0+\epsilon,h)] \\ =& \frac{p-\theta_d(v_0)-dc}{1-(1-\delta-\sigma-\lambda(1-F(v_0))(1+dg)}\lambda[uE(h|u)+\sum_h h\Pr(v'\leq v_0,h)] \\ >& \frac{p-\theta_d(v_0)-dc}{1-(1-\delta-\sigma-\lambda(1-F(v_0))(1+dg)}\lambda[uE(h|u)+\sum_h h\Pr(v'<v_0,h)] \\ =& \pi(d,v_0),\end{aligned}$$

where the third equality follows from the continuity of wage function $\theta_0(v)$, and the inequality follow from the fact that $\Pr(v' \leq v_0, h) \geq \Pr(v' < v_0, h)$ for all h, with strict inequality for some h. This implies that firms that offer v_0 cannot be maximizing: a contradiction.
Step 3: $F(v)$ has connected support.
Suppose not. There is a gap between (v', v'') in the support of $F(v)$, where v' and v'' are in the support. By strict monotonicity of wage function, for any $v \in (v', v'')$, $\theta_d(v') < \theta_d(v'')$ for $d = 0, 1$. Moreover, since $F(v') = F(v'')$, and hence, $\Pr(v < v', h) = \Pr(v < v'', h)$ for all h, then:

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

a contradiction. ■

B3. Human Capital Distribution

The following notations will be used: $F^c = F(v^c)$ 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* $h = 1$,

$$D(1) = \frac{\sigma(\sigma+\delta+\lambda)}{s^c(\sigma+\lambda)-\delta\lambda F^c}, \tag{7}$$

$$uD^u(1) = \frac{\sigma s^c}{s^c(\sigma+\lambda)-\delta\lambda F^c}. \tag{8}$$

For all $n \geq 1$,

$$D[(1+g)^n] = D(1)\frac{s^c\lambda(\sigma+\delta+\lambda)(1-F^c)}{s^c(\sigma+\lambda)-\delta\lambda F^c}y^{n-1}, \tag{9}$$

$$uD^u[(1+g)^n] = D(1)\frac{s^c\lambda\delta(1-F^c)}{s^c(\sigma+\lambda)-\delta\lambda F^c}y^{n-1}, \tag{10}$$

where

$$y = \frac{\delta\lambda(\sigma+\delta+\lambda)(1-F^c)}{s^c(\sigma+\lambda)-\delta\lambda F^c}+1-\sigma-\delta. \tag{11}$$

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:

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

and the mean human capital among unemployed workers is:

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

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, I get

$$(\lambda+\sigma)uD^u(1)=\sigma+\delta(1-u)G^cD^0(1). \tag{12}$$

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:

$$(\lambda+\sigma)uD^u(h)=(1-u)G^cD^0(h)\delta+(1-u)(1-G^c)D^1(h)\delta. \tag{13}$$

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:

$$s^c(1-u)G^cD^0(h)=\lambda F^c uD^u(h). \tag{14}$$

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.

$$\begin{aligned}(1-u)(1-G^c)D^1(h) \quad = \quad & (1-u)(1-G^c)D^1(\frac{h}{1+g})(1-\sigma-\delta) \\ & +uD^u(\frac{h}{1+g})\lambda(1-F^c) \\ & +(1-u)G^cD^0(\frac{h}{1+g})\lambda(1-F^c).\end{aligned} \tag{15}$$

In the whole economy:

$$D(1) = uD^u(1) + (1-u)G^c D^0(1), \tag{16}$$

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

$$D(h) = uD^u(h) + (1-u)G^c D^0(h) + (1-u)(1-G^c)D^1(h). \tag{17}$$

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., $uD^u(h)$, $(1-u)G^c D^0(h)$ 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 $h \in \{(1+g)^n\}_{n=1}^{\infty}$,

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

And for $h = 1$

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

Solving the equations (12) to (17) 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, $\lim_{n\to\infty} D[(1+g)^n] = 0$ because $y \in (0,1)$.
The mean of human capital is

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

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 $uD^u(.)$ and $D(\cdot)$, one can get the expression of the average human capital among unemployed workers.[2] ■

[2] More detailed proof is available from the author on request.

B4. 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. $v < v^c$

$$\Pr(v' \leq v, h = (1+g)^n) = \frac{\lambda F(v)}{s(v)} uD^u[(1+g)^n] \text{ for } n \geq 0, \tag{18}$$

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. $v \geq v^c$

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

for $n \geq 1$,

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

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

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

■

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

B5. The cutoff value v^c divides the economy into training and non-training sectors. A worker with very high level of human capital must have been in the training sector for a long time. The longer she stays in the training sector, the higher her human capital level is, due to on-the-job training; and the higher her job value is, due to on-the-job search. However, unemployed workers are the only inflow for the non-training sector, and layoff occurs with the same probability for workers regardless of their human capital levels. Therefore, conditional on being employed in the non-training sector, the job value a worker obtains is not correlated with her human capital level. This idea is formalized in the following.

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. 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:

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

Case 1. $v < v^c$:

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

for $n \geq 1$

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

where the second equality follows from the relationship between $uD^u()$and $D()$. Notice that in this case, the conditional distribution is invariant to h.
Case 2. $v \geq v^c$

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

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

$$\begin{aligned}\Pr(v' \leq v, h = (1+g)^n) = & \Pr(v' \leq v^c, (1+g)^n) + \Pr(v^c \leq v' \leq v, [(1+g)^n]) \quad (19) \\ = & \frac{\lambda F^c}{s^c} uD^u[(1+g)^n] \\ & + \lambda(F(v) - F^c) \sum_{m=1}^{n} (1 - s(v))^{m-1} \frac{(\sigma + \lambda + \delta)}{s^c} uD^u[(1+g)^{n-m}].\end{aligned}$$

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 $y > 1 - \sigma - \delta > 1 - s(v)$. Plug the expressions for $\frac{(\sigma+\lambda+\delta)}{s^c} uD^u[(1+g)^{n-m}]$ into (19), I have

$$\begin{aligned}\Pr(v' \leq & \; v, h = (1+g)^n) \quad (20) \\ = & \; \frac{\lambda F^c}{s^c} uD^u[(1+g)^n] \\ & + \lambda(F(v) - F^c)D(1)\{\frac{\lambda\delta(\lambda + \sigma + \delta)(1 - F^c)y^{n-2}}{s^c(\lambda + \sigma) - \lambda\delta F^c} \sum_{m=1}^{n-1} (\frac{1 - s(v)}{y})^{m-1} + (1 - s(v))^{n-1}\}.\end{aligned}$$

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,

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

After some algebraic manipulation,

$$\begin{aligned}
\Pr(v' &\leq v|h=(1+g)^n) \\
&= \frac{\lambda F^c \delta}{s^c(\sigma+\lambda)} \\
&\quad + \frac{\lambda\delta(\lambda+\delta+\sigma)(F(v)-F^c)}{s^c(\lambda+\sigma)(y+s(v)-1)} \\
&\quad + \frac{(F(v)-F^c)}{s^c(\lambda+\sigma)} (\frac{1-s(v)}{y})^{n-1} \{ \frac{[s^c(\lambda+\sigma)-\lambda\delta F^c]}{1-F^c} - \frac{\lambda\delta(\lambda+\delta+\sigma)}{y-(1-s(v))} \}.
\end{aligned} \tag{21}$$

Part II. Show first order stochastic dominance: From (21), one can see that the conditional probability is decreasing in n for any $v < \overline{v}$ 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

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

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

B6. Joint distribution of pay rate and human capital

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

Corollary B2 *The joint distribution of pay rate and human capital among employed workers is given by*
Case 1. $\theta \in [b, \theta_1(v^c))$

$$\frac{1}{1-u}\Pr(\theta' \leq \theta, h = (1+g)^n) = \frac{1}{1-u}\Pr(v' \leq \theta_0^{-1}(\theta), h = (1+g)^n) \qquad (22)$$

Case 2. $\theta \in (\theta_0(v^c), \theta_1(\overline{v})]$

$$\frac{1}{1-u}\Pr(\theta' \leq \theta, h = (1+g)^n) = \frac{1}{1-u}\Pr(v' \leq \theta_1^{-1}(\theta), h = (1+g)^n)$$

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

$$\begin{aligned}\frac{1}{1-u}\Pr(\theta' \leq \theta, h = (1+g)^n) = \quad & \frac{1}{1-u}\{\Pr(v' \leq \theta_0^{-1}(\theta), h = (1+g)^n) \\ + \quad & \Pr(v^c \leq v' \leq \theta_1^{-1}(\theta), h = (1+g)^n)\}\end{aligned}$$

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 $\theta < \theta_1(v^c)$.
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 cases 1 and 2, there is only one type of firms in the market offering the pay rate θ, and each pay rate corresponds to a unique job value. As a result, the (θ, h) distribution is the same as (v, h) distribution. In case 1, θ is so low that only firms without training would offer such a pay rate. In case 2, $\theta \in (\theta_0(v^c), \theta_1(\overline{v}))$, θ is offered only by firms with training. When $\theta \in [\theta_1(v^c), \theta_0(v^c)]$, the same pay rate is offered by both types of firms. In this case, $\Pr(\theta' \leq \theta, h)$ is composed of $\Pr(v' \leq \theta_0^{-1}(\theta), h)$ from non-training firms, and $\Pr(v^c \leq v' \leq \theta_1^{-1}(\theta), h)$ from training firms. 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. ■

B7. Proof for Claim 1 (Average Human Capital Level Hired by A Firm With Value v)

Proof. Recall the definition of $l(v) : l(v) = \lambda[I(v \geq v_u)uE(h|u) + \sum_h h\Pr(v' < v, h)]$, the proof follows from the definition of $\Pr(v' < v, (1+g)^n)$ given in Proposition 4.

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

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

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

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

where the last equality uses the result from (20). Define X as the constant term $\lambda\delta(\lambda+\sigma+\delta)(1-F^c)/[(\lambda+\sigma)s^c - \delta\lambda F^c]$, compute the summation:

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

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

$uD^u(1)$, I have

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

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

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

■

B8. Proof for Proposition 2 (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 $v_u \le v \le v^c$, $d=0$ 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:

$$\frac{(p-\theta_0(v))(\sigma+\lambda+\delta)}{s(v)^2}=\frac{(p-b)}{\sigma+\lambda+\delta}, \tag{23}$$

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 $v^c < v \le \overline{v}$, $d=1$ 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

$$\frac{(p-\theta_1(v)-c)}{[s(v)(1+g)-g]^2}=\frac{(p-\theta_1(v^c)-c)}{[s^c(1+g)-g]^2} \tag{24}$$

and the result, again, 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 (**??**) is the same as $F(v^c)$ in (**??**). ■

B9. Proof for Proposition 3 (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 v and θ: for non-training job, $dv_0/d\theta = 1/s(v)$, where $s(v)$ is the separation probability. For training jobs, $dv_1/d\theta = 1/[(1+g)s(v) - g]$.

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 $v < v^c$,

$$s(v) = (\sigma + \delta + \lambda)\sqrt{\frac{p - \theta_0(v)}{p - b}}. \tag{25}$$

For $v \geq v^c$,

$$(1+g)s(v) - g = (\sigma + \delta + \lambda)\sqrt{\frac{p - \theta_1(v) - c}{p - b}}\sqrt{\frac{(1+g)s^c - g}{s^c}}, \tag{26}$$

where $s^c = \sigma + \delta + \lambda(1 - F(v^c))$, separation rate for the firm that is indifferent between training and non-training.

3) Plug (25) and (26) into part 1, I get the following relationship between v and θ, where M_0 and M_1 are constants:

$$\begin{aligned} v_0(\theta) &= M_0 - \frac{2\sqrt{(p-b)(p-\theta)}}{\sigma + \delta + \lambda}, \\ v_1(\theta) &= M_1 - \frac{2\sqrt{(p-b)(p-\theta-c)}}{\sigma + \delta + \lambda}\sqrt{\frac{s^c}{(1+g)s^c - g}}. \end{aligned} \tag{27}$$

At v^c, the right-hand side of these two equations should be equal, therefore, I have

$$\begin{aligned} & M_0(\sigma + \delta + \lambda) - 2\sqrt{(p-b)(p-\theta_0^c)} \\ = \; & M_1(\sigma + \delta + \lambda) - 2\sqrt{(p-b)(p-\theta_1^c - c)}\sqrt{\frac{s^c}{(1+g)s^c - g}}. \end{aligned} \tag{28}$$

4) Evaluate (25) and (26) at v^c, I get the pay rate at v^c,

$$\begin{aligned} \theta_0^c &= p - (\frac{s^c}{\sigma + \delta + \lambda})^2(p - b), \\ \theta_1^c &= p - c - \frac{s^c[(1+g)s^c - g]}{(\sigma + \delta + \lambda)^2}(p - b). \end{aligned} \tag{29}$$

Plug these into (28), I get the relationship that $M_0 = M_1 \equiv M$.

5) Using the definition of v^c, i.e., at v^c,the worker-firm joint benefit is equal to c, and using (29), I get the following:

$$c = g[M_0 - p - \frac{s^c}{(\sigma+\delta+\lambda)^2}(p-b)]. \tag{30}$$

6) The relationship between $\overline{v}$ and v_u, from the Bellman equation, is the following:

$$\overline{v} = \frac{\theta_1(\overline{v}) + \delta(1+g)v_u}{(\sigma+\delta)(1+g)-g},$$

where v_u can be derived from (27) as

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

Moreover, from (27), I also have

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

where $\theta_1(\overline{v})$ can be derived from (26)

$$\theta_1(\overline{v}) = p - c - (\frac{(1+g)(\delta+\sigma)-g}{\sigma+\delta+\lambda})^2\frac{s^c}{(1+g)s^c-g}(p-b).$$

Therefore, I get the following relationship

$$\begin{aligned} & M_1 - \frac{2s^c[(1+g)(\delta+\sigma)-g]}{[(1+g)s^c-g](\sigma+\delta+\lambda)^2}(p-b) \\ = & \frac{p-c-(\frac{(1+g)(\delta+\sigma)-g}{\sigma+\delta+\lambda})^2\frac{s^c(p-b)}{(1+g)s^c-g}+\delta(1+g)[M_0-\frac{2(p-b)}{(\sigma+\delta+\lambda)}]}{(\sigma+\delta)(1+g)-g}. \end{aligned} \tag{31}$$

7) With $M_0 = M_1$, (30) and (31) are two equations in two unknowns s^c and M. Plugging (30) into (31), I obtain one equation in one unknown: s^c. Solving this equation for s^c, I can then back out the whole distribution of 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 v exists and is unique, which implies the existence and uniqueness of the market equilibrium.

8) Given the distribution of v, I obtain the $D(h)$ and $\Pr(v' \leq v, h)$ that are characterized in Propositions B1 and B2.

Appendix C: Efficiency of Training Provision

Consider a social planner who also faces search frictions and has to decide the fraction (α) of jobs with training. In assuming that the social planner is also subject to search frictions, I concentrate on the efficiency of training provision. Let A be the social value of a unit of human capital if it is currently employed, and E be the value if it is currently unemployed. A unit of human capital, while unemployed, is valuable not only to the worker, but also 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., $E > v_u$. The social planner's problem is:[3]

$$\begin{aligned} A = \max_{0\leq\alpha\leq 1} \quad & \{(1-\alpha)[p+\delta E+(1-\delta-\sigma)A] \\ & +\alpha[p-c+\delta(1+g)E+(1-\delta-\sigma)(1+g)A]\}, \qquad (32) \\ \textit{where} \quad E \quad = \quad & b+\lambda A+(1-\lambda-\sigma)E. \end{aligned}$$

The objective function is linear in α, which indicates a corner solution. Define

$$\begin{aligned} A_0 &= \frac{p+\delta E}{(\delta+\sigma)} \\ A_1 &= \frac{p-c+\delta(1+g)E}{(\delta+\sigma)(1+g)-g}. \end{aligned}$$

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 $\alpha \in [0,1]$, $A = A_0 = A_1$.

The social planner follows a cutoff cost strategy, where the cutoff cost c^* equalizes A_0 and A_1:

$$c^* = g\frac{(1-\delta-\sigma)p+\delta E}{\delta+\sigma} > g\frac{(1-\delta-\sigma)p+\delta v_u}{\delta+\sigma} = B(\overline{v}). \qquad (33)$$

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.

Proposition 1 *In general, market-provided training is inefficiently low. Only when the cost of training goes above c^* defined in (33) will the market's result coincide with the social planner's choice, which is not to provide training.*

[3] 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. As a result, 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 the 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.

The inefficiency of training provision in the market equilibrium results from the externality of general training. When a firm chooses between training and no training, it considers only the fact that increasing human capital will increase its own production and will provide a cheaper way to keep its promise of 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. 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 is in line with previous studies, for example, Stevens (1994).