

On-the-job Training and On-the-job Search: Wage-Training Contracts in a Frictional Labor Market*

Seung-Gyu (Andrew) Sim[†]

December 22, 2014

Abstract

This paper analyzes the coexistence of on-the-job training and on-the-job search in a frictional labor market in which both skilled and unskilled workers engage in on-the-job search behavior and firms post skill-dependent wage-training contracts to preemptively back-load compensation and extract more surplus in earlier periods. The paper demonstrates that because back-loaded compensation after training effectively enlarges the expected duration, and joint value, of the match as if it were job-specific training, general training is over-intensified overall. Comparative static analysis predicts aggregate training intensity to be reinforced, but net output to improve toward the Beckerian outcome, as search friction is mitigated.

Keywords: On-the-job Training, On-the-job Search, Wage-Training Contracts

JEL Classification: J24, J31, J64

*This is a revised version of the third chapter of my Ph.D dissertation. I am deeply indebted to John Kennan, Rasmus Lentz, and Chris Taber for their insightful guidance and invaluable support. I also thank Junichi Fujimoto, Tim Huegerich, Hidehiko Ichimura, Michihiro Kandori, Young Sik Kim, Chul-In Lee, Dan Sasaki, Yasuyuki Sawada, Yongs Shin, Serene Tan, Kenichi Ueda, and all seminar participants at Hitotsubashi University, Seoul National University, University of Tokyo, and Yokohama National University for their helpful comments and warm encouragement. I also acknowledge financial support by the Grants-in-Aid for Scientific Research (Kakenhi No. 26780170) from the Japan Society for the Promotion of Science. All remaining errors are mine.

[†]Address: University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, 113-0033. Phone: +81-3-5841-5620. Email: sgsim@e.u-tokyo.ac.jp

1 Introduction

Most economists believe human capital accumulation on the job to be the major contributor to both individual wages and broader economic growth.¹ Rapid technological progress and knowledge spillover dictate that workers continuously upgrade their skills post schooling, and firms pay the cost of on-the-job training to encourage workers to catch up and keep up with new technologies.² Absent commitment by workers to stay with their firms,³ it is unclear why, and how much, firms will pay for general training. This paper investigates how firm productivity, wage payments, and investment in human capital interact in response to workers' on-the-job search behavior. The focus is on efficiency analysis of the market equilibrium and comparative static analysis associated with the mitigation of search friction.

Becker (1964) documents that under perfect competition firms provide the socially efficient level of general training to retain workers, such training paid for by workers through lower wages during training. This result is shown by Acemoglu and Pischke (1999), however, to be subject to a wide range of deviation from perfect information and competition in the labor market. Firms provide inefficiently low levels of training and share the cost of training in a labor market with high job turnover, according to Acemoglu (1997), because training firms cannot raise a claim on the expected benefit of next employers (the free-rider problem). Sanders and Taber (2012) acknowledge the widely accepted notion that potential job turnover causes under-investment in general, and over-investment in job-specific, training, especially in a wage bargaining environment.⁴ Fu (2011) investigates on-the-job training and wage dispersion in a frictional labor market by incorporating general training and the piece-rate sharing rule into the framework proposed by Burdett and Mortensen (1998) and Burdett et al. (2011), who, consistent with Sanders and Taber (2012), shows general job training to be under-provided in the market equilibrium relative to the constrained social planner's problem.

This paper extends the Burdett and Mortensen (1998) framework by differentiating workers according to skill level. Workers enter the labor market as unskilled and acquire general skills through firm-sponsored training on the job. Both employed and unemployed workers search for better employment opportunities. Deviating from Fu (2011) with respect to the piece-rate sharing rule, this paper posits that recruiting firms, being heterogeneous in terms of productivity, post skill-dependent lifetime values that reflect their optimal recruiting strategies. Committed values are delivered to employed workers according to operating firms' optimal training and retention strategies. When

¹For example, Altonji and Shakotko (1987) argue that general human capital accumulation accounts for the lion's share of wage growth, and Topel and Wald (1992) report that the wage of a typical male worker in the U.S. labor market doubles over the course of a 40-year career.

²Lynch and Black (1998) find that more than half of (but not all) U.S. firms provide and pay for general training, for example, in computer skills and teamwork.

³Fallick and Fleischman (2004) find 60 percent of vacancies in the U.S. labor market to be taken by employed, and only 40 percent by unemployed, searchers, Topel and Wald (1992) find that a typical worker holds seven jobs during the first 10 years of employment, and Sim (2013) reports the average job duration of white male high school graduates in the United States to be slightly more than two years.

⁴An interesting exception is Moen and Rosen (2004), who argue that an efficient level of training can be achieved in a frictional labor market if employers and employees are able to coordinate efficiently, as by means of long-term contracts on wages, training intensity, and search intensity. The authors do not account, however, for the possibility of over-investment.

unskilled as well as skilled workers engage in on-the-job search behavior, training firms should, because they lose the entire rent when trainees leave for better paying jobs, care about job turnover even before skills are acquired. This consideration leads training firms to commit to paying trainees the entire surplus after training, choose the training intensity that maximizes the sum of the workers' and firms' values, and extract surplus by paying lower wages during training. In other words, firms exploit the training opportunity to back-load compensation as much as possible in order to extract more surplus in earlier periods. The underlying motivation for the back-loading wage scheme coincides with that of the wage-tenure contract proposed by [Burdett and Coles \(2003\)](#), [Stevens \(2004\)](#), and [Shi \(2009\)](#).⁵

It is useful to compare the market equilibrium with the problem of the *constrained* social planner, who, taking labor market friction as given, maximizes the present value of the expected net output flow throughout the life of a newly born worker. The *constrained* social planner, not caring from which firm a worker receives training, considers only productivity improvement in determining training intensity. That training firms exploit training for purposes of back-loading compensation as well as improving productivity effectively increases the expected joint value and duration of matches post training. Relative to the social planner, training firms thus provide over-intensified general training as if it were job-specific training, which possibility has thus far not been properly considered, the literature having paid little attention to job turnover among unskilled workers or training firms' preemptive responses thereto. The the back-loading scheme, although it dilutes the problem of free riders raised by [Acemoglu \(1997\)](#), suggests other sources of inefficiency, namely, 'training inefficiency' and 'allocation inefficiency.' According to the quantitative analysis based on the calibrated model, the market equilibrium outcome, relative to the constrained social planner's outcome, retains a larger steady state mass of skilled workers (due to the former inefficiency), but produces fewer units of the actual and net outputs (due to the latter inefficiency).

One possible take on [Acemoglu \(1997\)](#) is that, although he delivers a plausible insight, as search friction is mitigated the market outcome converges to almost "no training" in the opposite direction to the prediction of [Becker \(1964\)](#). The current paper's contracting solution predicts that accelerating the job turnover rate among employed workers both enables them to find more efficient matches quickly and discourages firms from providing general training, thereby improving, but not as much as in the constrained planner's problem, the market equilibrium's inefficiency in job turnover and training. It also predicts that firms will provide more intensified training as the job finding rate of unemployed workers is accelerated because the value differential between skilled and unskilled workers becomes larger. When offer arrival rates to employed and unemployed searchers are accelerated in the same proportion, aggregate training intensity is reinforced further from, but net output increased towards, the planner's outcome owing to the enhanced skill supply and efficiency improvement in job turnover behavior in spite of higher training cost. The calibrated model demonstrates that as search friction is mitigated through labor market reforms or the arrival

⁵[Stevens \(2004\)](#) shows firms, in the presence of on-the-job search, to *take* in early periods, and *leave for the workers* in later periods, the entire surplus. [Burdett and Coles \(2003\)](#), assuming risk aversion, describe a unique back-loading wage schedule such that different firms choose different starting wages and all firms follow the wage schedule. [Shi \(2009\)](#) develops a directed search model with the wage-tenure contract and proposes 'block recursivity.' The current paper departs from the foregoing studies in proposing that training firms improve the probability of retention by ceding the entire surplus to workers post training.

of an efficient search technology, aggregate outcome in the planner’s problem moves towards the Beckerian outcome and the outcome of the market equilibrium follows the planner’s outcome.

The rest of the paper proceeds as follows. Section 2 develops the theoretical model and characterizes the steady state equilibrium. The efficiency analysis is presented in Section 3, the quantitative assessment in Section 4. Section 5 concludes.

2 Model

2.1 Environment

This paper incorporates on-the-job training into the [Burdett and Mortensen \(1998\)](#) framework to analyze how firm productivity, wage payment, and provision of job training interact in response to workers’ on-the-job search behavior. It defines unit measures of risk neutral workers and firms, both of which discount the future at rate r . A newly born worker enters the labor market as an unskilled unemployed worker. The unskilled worker acquires general skills through on-the-job training and becomes a skilled worker.⁶ “Skilled” and “unskilled” are denoted throughout the paper by the subscript $i \in \{s, u\}$. Firms are heterogeneous in terms of productivity, $H(p)$ denoting the proportion of firms with productivity no greater than p and $H(\cdot)$ continuously differentiable with support $[\underline{p}, \bar{p}]$. Each firm has one vacant job at every instant. A recruiting firm posts and commits skill-dependent lifetime values, and an operating firm delivers the committed value through its wage-training schedule. The operating job with productivity p occupied by an unskilled worker accrues revenue p , by a skilled worker, revenue $p + s$, at every instant. Assume that $\underline{p} + s < \bar{p}$.

To ensure the existence of an equilibrium, assume that there are ε -measure of noise firms such that they randomly draw and post skilled wages from the cumulative distribution $\hat{F}_N : [\underline{p} + s, \bar{p} + s] \rightarrow [0, 1]$. The overhead hat ($\hat{\cdot}$) indicates a distribution of wages, not values. It is assumed that the steady state equilibrium has a sufficiently small $\varepsilon (> 0)$. Noise firms are introduced as a means of removing a mass point in the distribution of skilled wage earnings (a detailed explanation of their role is provided in footnote 10), the only requirement being that the noise firms’ wage offer distribution be continuously differentiable and strictly increasing.⁷

Workers An unemployed worker collects unemployment benefit b and finds an employment opportunity at the Poisson arrival rate λ^0 . An employed worker receives a wage, finds another employment opportunity at rate λ^1 , and is laid off at rate δ . All workers retire (or die) at rate ρ , retirees being replaced by newly born workers who enter the labor market as unskilled and unemployed. Denote by U_i and E_i the lifetime values of an i -type unemployed and employed worker, respectively, and let $F_i : \mathbb{R}_+ \rightarrow [0, 1]$ be the steady state distribution function of the values offered to i -type workers. The latter’s support, denoted by $[\underline{E}_i, \bar{E}_i]$, as well as the distribution function F_i , will be endogenously determined later. The lifetime value of an unemployed worker

⁶[Quercioli \(2005\)](#) uses the [Burdett and Mortensen \(1998\)](#) framework to analyze the case in which firms provide training opportunities for firm-specific skills.

⁷The concept of ε -measure of noise firms is borrowed from [Galenianos and Kircher \(2009\)](#).

is described by

$$rU_i = b - \rho U_i + \lambda^0 \int \max\{z - U_i, 0\} dF_i(z), \text{ for each } i \in \{u, s\}. \quad (1)$$

The left-hand side of asset equation (1) represents the opportunity cost of holding asset “ i -type unemployment,” the right-hand side, the dividend flow from the asset, potential losses from retirement shock, and gains from job-finding.

Given that recruiting firms with different productivity post different values, the value of skilled employment offered by a particular recruiting firm with productivity p is denoted by $E_s(p)$. The committed value is delivered through skilled wages, potential losses from exogenous retirement and separation shocks, and potential gains from job-to-job transition, described by

$$rE_s(p) = w_s - \rho E_s(p) + \delta(U_s - E_s(p)) + \lambda^1 \int \max\{z - E_s(p), 0\} dF_s(z). \quad (2)$$

The one-to-one relationship between $E_s(p)$ and w_s , described in equation (2), allows us to denote by $w_s(p)$ the skilled wage flow paid by the firm with productivity p .

The unskilled employed worker receives on-the-job training and unskilled wage w_u . Given training intensity x , chosen by the firm, the worker acquires skills at rate μx , and upon acquiring skills is awarded a new labor contract with a higher value.⁸ In the steady state equilibrium, skilled workers who remain at the firms at which they were trained and promoted are termed “promoted (skilled) workers,” workers hired as skilled who acquired the skills elsewhere “recruited (skilled) workers.” The value of promotion, denoted by E_s^t , is obtained by plugging skilled wages after promotion, w_s^t , into the value equation (2). Let $E_u(p)$ be the value of unskilled employment offered by the firm with productivity p . Given (w_u, x, E_s^t) , the expected value of unskilled employment is given by

$$\begin{aligned} rE_u(p) = & w_u + x\mu(E_s^t - E_u(p)) - \rho E_u(p) + \delta(U_u - E_u(p)) \\ & + \lambda^1 \int \max[z - E_u(p), 0] dF_u(z). \end{aligned} \quad (3)$$

The value of unskilled employment, like that of skilled employment, is delivered through wage payments during training, expected gains from skill acquisition and subsequent promotion, potential losses from exogenous retirement and separation shocks, and potential gains from job-to-job transition. Equation (3) is used as the promise-keeping constraint in the training firm’s optimization.

Operating Firms Let $J_s(E_s(p), p)$ and $J_u(E_u(p), p)$ be the values of the skilled and unskilled matches having productivity p and committed values $E_s(p)$ and $E_u(p)$, respectively. These can be rewritten as $J_s(p)$ and $J_u(p)$ interchangeably. Given the one-to-one relationship between $E_s(p)$ and $w_s(p)$ in equation (2), the asset value of the skilled match producing $p + s$ at every instant is given by

$$rJ_s(p) = p + s - w_s(p) - [\rho + \delta + \lambda^1(1 - F_s(E_s(p)))]J_s(p). \quad (4)$$

⁸Owan (2004), to shed light on the promotion strategy and hierarchical structure of firms, treats strategic promotion and training provision separately. This paper, the organizational structure of firms being beyond its scope, assumes that an unskilled worker who acquires skills will immediately be promoted to a skilled position.

The firm having an operating unskilled match chooses a wage-training schedule to deliver the committed value. The cost of training associated with training intensity x is given by $c(x) = x^\gamma$, where $\gamma > 1$. Note that $c(\cdot)$ is (infinitely) continuously differentiable with $c(0) = 0$, $c'(0) = 0$, $c' > 0$, and $c'' > 0$. The value of an unskilled match producing p at every instant is described by

$$rJ_u(p) = \max_{w_u, x, E_s^t} p - w_u - c(x) - [\rho + \delta + \lambda^1(1 - F_u(E_u(p)))]J_u(p) + x\mu(J_s(p) - J_u(p)) \quad (5)$$

subject to constraints (3) and (4). Plugging (3) into (5) and taking the derivative with respect to E_s^t yields the first order condition $(dF_s(E_s^t)/dE_s)J_s(E_s^t, p) \geq 0$, where the strict equality holds only when $J_s(E_s^t, p) = 0$, as long as $F_s(\cdot)$ is strictly increasing.⁹ The ε -measure of noise firms is introduced to avoid the case in which dF_s/dE_s reaches zero quickly.¹⁰ Equation (4) implies that $J_s(E_s, p) = 0$ if and only if $w_s = p + s$. Let $w_s^t(p)$ be the wage payment after promotion by the training firm having productivity p . After promotion, the firm thus offers

$$E_s^t(p) = \frac{w_s^t(p) + \delta U_s + \lambda^1 \int_{E_s^t(p)} z dF_s(z)}{r + \delta + \rho + \lambda^1(1 - F_s(E_s^t(p)))}, \quad \text{where } w_s^t(p) = p + s > w_s(p). \quad (6)$$

Equation (6) implies that the training firm makes no claim after training, leaving the entire surplus for the trained worker. Apparently, no recruiting firms commit the entire surplus to potential employees. The training firm's back-loading strategy effectively deters potential job offers from slightly better poaching firms, as if the training firm and worker jointly accumulated job-specific human capital.

Given productivity $p \in [p, \bar{p}]$, the first order condition with respect to x yields

$$c'(x) = \mu(E_s^t(p) - E_u(p) - J_u(p)). \quad (7)$$

The left-hand side represents the marginal cost, the right-hand side the marginal benefit, of providing training. As $c(x)$ is convexly increasing with the Inada condition and the right-hand side independent of x , a global maximum is uniquely obtained in the interior. For each $p \in [p, \bar{p}]$, the interior solution is denoted by $x(p)$. Once the value of promotion $E_s^t(p)$ and training intensity $x(p)$ are determined, the promise keeping constraint (3) determines wage payment $w_u(p)$ at each productivity level p . Equation (7) implies that the training firm that commits to paying the worker the entire surplus after promotion chooses the training intensity that seemingly maximizes the summation of the worker's and firm's net gains from training, as Moen and Rosen (2004) suggest. Unlike Moen and Rosen (2004), however, the intensity that maximizes the summation of the worker's and firm's gains is not jointly efficient due to job turnover by unskilled workers. This will be examined in the next section.

⁹Note that the optimal decision on E_s^t satisfies the second order sufficient condition as well.

¹⁰Suppose there are no noise firms. For any $p \in [p, \bar{p}]$, the firm with productivity p never posts a skilled wage more than or equal to $(p + s)$, $w_s(\bar{p}) < \bar{p} + s$. If $w_s(\bar{p}) < \bar{p} + s$, the first order condition with respect to E_s^t implies that all training firms, regardless of productivity, choose $w_s^t(\cdot) = w(\bar{p})$ to retain their trained workers. If $w_s(\bar{p}) \geq \bar{p} + s$, there exists $\hat{p} \in [p, \bar{p}]$ such that $w_s^t(\hat{p}) = w_s(\bar{p})$ due to the continuity argument. Firms with productivity $p \in (\hat{p}, \bar{p}]$ set $w_s^t(\cdot) = w_s(\bar{p})$. There being, in both cases, a mass point at $w_s(\bar{p})$ in the wage earning distribution, no equilibrium exists.

Recruiting Firms Let $\{G_u(E_u), G_s(E_s)\}$ denote the proportions of unskilled and skilled workers that receive lifetime values no greater than $\{E_u, E_s\}$, respectively, and $\{u_u, u_s\}$ the total mass of unskilled and skilled unemployed workers, respectively. The total measure of workers being fixed at unity in the steady state equilibrium, a recruiting firm with productivity p posts the skill-dependent values (E_u, E_s) to maximize the expected value, $[\lambda^0 u_s + \lambda^1 G_s(E_s)]J_s(E_s, p) + [\lambda^0 u_u + \lambda^1 G_u(E_u)]J_u(E_u, p)$. Taking first order conditions with respect to E_s and E_u yields

$$\lambda^0 u_i + \lambda^1 G_i(E_i) = \left[\frac{\lambda^0 u_i + \lambda^1 G_i(E_i)}{r + \rho + \delta + \lambda^1(1 - F_i(E_i))} \frac{dF_i}{dE_i} + \frac{dG_i}{dE_i} \right] \lambda^1 J_i(E_i, p), \quad (8)$$

for each $i \in \{u, s\}$. Equation (8) implicitly determines the optimal pair of $(E_u(p), E_s(p))$ posted by the recruiting firm with productivity p , and equations (2) and (3) show how the committed values are delivered. [Figure 3] reveals both $E_s(p)$ and $E_u(p)$ to be strictly increasing in p under reasonable parameter values.

Let $\hat{F}_s : [\underline{w}_s, \bar{w}_s] \rightarrow [0, 1]$ be the cumulative distribution function of skilled wages offered by profit-maximizing normal firms, and $\hat{F}_{s\varepsilon} : [\underline{p} + s, \bar{p} + s] \rightarrow [0, 1]$ be the cumulative distribution function of the skilled wages offered by both noise firms and profit-maximizing normal firms, obtained, given $\varepsilon (> 0)$, by $\hat{F}_{s\varepsilon}(w) = (1 - \varepsilon)\hat{F}_s(w) + \varepsilon\hat{F}_N(w)$. As $\varepsilon \rightarrow 0$, $\hat{F}_{s\varepsilon}(w(p))$ converges to $\hat{F}_s(w(p)) = F_s(E_s(p))$. Because the noise firms randomly draw skilled wages from the wage offer distribution $\hat{F}_N : [\underline{p} + s, \bar{p} + s] \rightarrow [0, 1]$, the supremum value offered, \bar{E}_s , is strictly higher than $E_s(\bar{p})$, the *actual* supremum value posted by the profit-maximizing normal firms. As ε goes to zero, $F_s(E_s(\bar{p}))$ converges to one and $F_s(\bar{E}_s) - F_s(E_s(\bar{p}))$ to zero.

Equilibrium Configuration *Given firms' productivity distribution $H(p)$, a steady state equilibrium with on-the-job training and on-the-job search consists of value equations $\{U_i, E_i, J_i\}_{i \in \{u, s\}}$, compensation packages $\{(w_u(\cdot), x(\cdot), E_s^t(\cdot)), (w_s(\cdot))\}$, and steady state measures $\{F_i, G_i, u_i\}_{i \in \{u, s\}}$ that jointly satisfy the following conditions.*

- (i) *Given $\{F_i\}_{i \in \{u, s\}}$, workers make optimal job turnover decisions, which determines $\{E_i, U_i\}_{i \in \{u, s\}}$ from (1), (2), and (3) together with the firms' training decisions.*
- (ii) *Given $H(p)$ and $\{F_i, E_i, U_i\}_{i \in \{u, s\}}$, each operating firm with a skilled worker delivers $E_s(p)$, which determines (4). The operating firm with an unskilled worker optimally chooses $(w_u(p), x(p), E_s^t(p))$ following (3), (6) and (7), which determines (5).*
- (iii) *Given $\{G_i, u_i\}_{i \in \{u, s\}}$ and $\{J_i\}_{i \in \{u, s\}}$, each recruiting firm posts a contract which satisfies the first order condition in (8).*
- (iv) *$\{F_i, G_i, u_i\}_{i \in \{u, s\}}$ are stationary and consistent with workers' job turnover decisions and firms' provision of training.*

[Figure 1] illustrates the dynamic worker flow in the steady state equilibrium. The left, middle, and right vertical lines represent the equilibrium wage supports for unskilled workers, promoted workers, and recruited workers, respectively. Newly born workers start their careers as unskilled and, once employed, are afforded opportunities for training and continue to search for better paying jobs. Arrow "a" in [Figure 1] represents unskilled workers' job-to-job transition, arrow "b," promotion after acquiring skills. An unskilled worker receiving unskilled wage $w_u(p)$ is promoted to a skilled

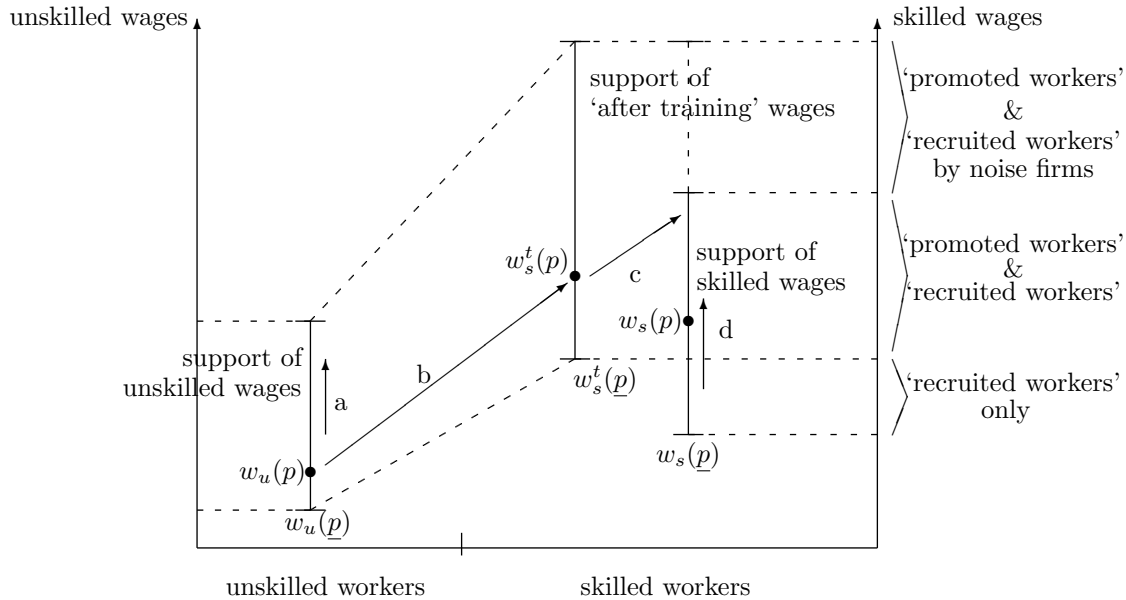


Figure 1: Equilibrium Support of Wages

position with wage $w_s^t(p)$, which is strictly larger than the skilled wage received by skilled workers recruited by the same firm $w_s(p)$. “Promoted workers” leave their jobs voluntarily or involuntarily and become “recruited workers” or “unemployed workers,” respectively. Arrow “c” represents “promoted workers’,” arrow “d” “recruited workers’,” job-to-job transition. [Figure 1] depicts two wage supports for skilled workers, one exclusively for recruited, the other for both recruited and promoted, skilled workers. The dotted line on top of the support of skilled wages indicates that they are, in the interval, offered by noise firms.

Training firms give their workers the entire surplus after training for the same reason firms offer wage-tenure contracts in [Burdett and Coles \(2003\)](#), [Stevens \(2004\)](#), and [Shi \(2009\)](#). Because unskilled workers engage in on-the-job search behavior during training, training firms discount the future much more severely than their workers so that they trade the surplus before and after training. Training firms take the surplus in earlier periods and leave the full surplus for workers after training. Because firms pay much higher skilled wages to internally trained and promoted workers than to recruited workers, the former are less likely to leave for slightly more productive poaching firms. This strategic back-loading scheme leads to general training being exploited as a means of enlarging the expected duration, and thus joint value of, the match as if it were job-specific training, further reinforcing the investment in general training.

2.2 Characterization of Steady State Equilibrium

It is natural to think that the least productive firm having \underline{p} , being able to attract unemployed but not employed searchers, posts for each i -type worker the lifetime value equivalent to U_i , which determines $E_i(\underline{p}) = \underline{E}_i$, the infimum value offered to i -type workers.

Lemma 1 *Suppose that (U_u, U_s) are given. The optimal strategy by the least productive firm then implies*

$$w_s(\underline{p}) = (r + \rho)U_s - \frac{\lambda^1}{\lambda^0} \left[(r + \rho)U_s - b \right], \quad (9)$$

$$E_s^t(\underline{p}) = \frac{1}{r + \delta + \rho} \left[\underline{p} + s + \delta U_s + \lambda^1 \int_{\underline{p}+s}^{\bar{p}} \frac{(1 - \hat{F}_s(w'))dw'}{r + \delta + \rho + \lambda^1(1 - \hat{F}_s(w'))} \right], \quad (10)$$

$$w_u(\underline{p}) = (r + \rho)U_u - \frac{\lambda^1}{\lambda^0} \left[(r + \rho)U_u - b \right] - \mu x(\underline{p})(E_s^t(\underline{p}) - U_u), \quad \text{and} \quad (11)$$

$$\begin{aligned} & \frac{1}{\mu} (r + \rho + \delta + \lambda^1) c'(x(\underline{p})) + x(\underline{p}) c'(x(\underline{p})) - c(x(\underline{p})) \\ & = (r + \rho + \delta + \lambda^1) E_s^t(\underline{p}) - [\underline{p} + \delta U_u + \frac{\lambda^1}{\lambda^0} ((r + \rho + \lambda^0)U_u - b)]. \end{aligned} \quad (12)$$

The above strategy by the least productive firm with \underline{p} results in

$$E_s(\underline{p}) = \frac{w_s(\underline{p}) + \delta U_s + \lambda^1 ((r + \rho)U_s - b) / \lambda^0}{r + \rho + \delta + \mu x(\underline{p})}, \quad (13)$$

$$E_u(\underline{p}) = \frac{w_u(\underline{p}) + \delta U_u + \lambda^1 ((r + \rho)U_u - b) / \lambda^0 + \mu x(\underline{p}) E_s^t(\underline{p})}{r + \rho + \delta + \mu x(\underline{p})}, \quad (14)$$

$$J_s(\underline{p}) = \frac{\underline{p} + s - w_s(\underline{p})}{r + \rho + \delta + \lambda^1}, \quad \text{and} \quad (15)$$

$$J_u(\underline{p}) = \frac{\underline{p} - c(x(\underline{p})) - w_u(\underline{p})}{r + \rho + \delta + \lambda^1 + \mu x(\underline{p})}. \quad (16)$$

Given that $U_s > U_u$, Lemma 1 implies that as long as unemployed workers are more efficient than employed workers in job search, the infimum of skilled wages, $\lambda^0 \geq \lambda^1$, is strictly greater than that of unskilled wages, $w_s(\underline{p}) > w_u(\underline{p})$. It is trivial that $F_u(E_u(\underline{p})) = F_s(E_s(\underline{p})) = G_u(E_u(\underline{p})) = G_s(E_s(\underline{p})) = 0$, in which case the steady state equilibrium is characterized by the system of differential equations (1)-(8) and (A4)-(A6) in Appendix A together with initial condition (9)-(16) in Lemma 1. The differential equations (A4)-(A6) are derived from the following argument.

The unskilled, unemployed worker finds a job at rate λ^0 ; the unskilled, employed worker is laid off at rate δ . A retiree is replaced with a newly born, unskilled, unemployed worker. Equating the outflow from, and inflow to, the steady state unskilled unemployment yields

$$(\lambda^0 + \rho)u_u = \delta G_u(E_u(\bar{p})) + \rho \phi. \quad (17)$$

The steady state unemployment rate being given by $(\rho + \delta) / (\rho + \delta + \lambda^0)$, the proportion of skilled unemployed workers is given by

$$u_s = (\rho + \delta) / (\rho + \delta + \lambda^0) - u_u. \quad (18)$$

Given that $E_u(\underline{p})$ is uniquely defined and strictly increasing in \underline{p} , the unskilled, unemployed worker finds a job with productivity no greater than \underline{p} at rate $\lambda^0 F_u(E_u(\underline{p}))$. The

unskilled worker working at a job with productivity less than p switches to a higher valued job at rate $\lambda^1(1 - F_u(E_u(p)))$, acquires skills at rate $\mu x(p)$, is laid off at rate δ , and retires at rate ρ . The steady state measure $G_u(E_u(p))$ is characterized by

$$\lambda^0 F_u(E_u(p))u_u = (\rho + \delta + \lambda^1(1 - F_u(E_u(p))))G_u(E_u(p)) + \int_{\underline{p}}^p \mu x(p')g_u(E_u(p'))dp', \quad (19)$$

where $g_u(E_u(p)) = dG_u(E_u(p))/dp$. There are two types of skilled employed workers, recruited and promoted workers. Denote by $G_s^r(E_s(p))$ ($G_s^t(E_s^t(p))$) the mass of recruited (promoted) workers whose values are less than $E_s(p)$ ($E_s^t(p)$). Then, $G_s(E_s) = G_s^r(E_s) + G_s^t(E_s)$. Equating the inflow to and outflow from the steady state measure of recruited workers receiving less than value E_s yields

$$[\rho + \delta + \lambda^1(1 - F_s(E_s))]G_s^r(E_s) = \lambda^0 F_s(E_s)u_s + \lambda^1 \int_{E_s^t(p)}^{E_s} [F_s(E_s) - F_s(z)]dG_s^t(z). \quad (20)$$

The left-hand side implies that recruited workers experience retirement, separation, and job-to-job transition; the right-hand side represents the flow of skilled workers newly recruited at values no greater than E_s . Promoted workers leave their training firms due to retirement, separation, and job-to-job transition. Equating the inflow and outflow yields

$$\int_{\underline{p}}^p \mu x(p')g_u(E_u(p'))dp' = (\rho + \delta)G_s^t(E_s^t(p)) + \lambda^1 \int_{E_s}^{E_s^t(p)} (1 - F_s(E_s^t(p)))dG_s^t(z). \quad (21)$$

Taking the derivatives of (19), (20), and (21) with respect to p yields a system of differential equations with initial conditions $G_u(E_u(\underline{p})) = 0$, $G_s^r(E_s(\underline{p})) = 0$, and $G_s^t(E_s^t(\underline{p})) = 0$.

Lemma 2 *Solving the unskilled workers' optimal job turnover decision and firms' optimal training decision yields*

$$G_u(E_u(p)) = I_u^{-1}(p) \int_{\underline{p}}^p \frac{I_u(p')\lambda^0(dF_u(E_u(p'))/dp)u_u}{\rho + \delta + \lambda^1(1 - F_u(E_u(p'))) + \mu x(p')} dp', \quad \text{where} \quad (22)$$

$$I_u(p) := \exp \left[\int_{\underline{p}}^p \frac{-\lambda^1(dF_u(E_u(p'))/dp')}{\rho + \delta + \lambda^1(1 - F_u(E_u(p'))) + \mu x(p')} dp' \right]. \quad (23)$$

The implied flows of skilled workers are thus described by

$$G_s^t(E_s^t(p)) = \int_{\underline{p}}^p \frac{\mu x(p')(dG_u(E_u(p'))/dp)}{\rho + \delta + \lambda^1(1 - F_s(E_s^t(p')))} dp' \quad \text{and} \quad (24)$$

$$G_s^r(E_s(p)) = I_s^{-1}(p) \int_{\underline{p}}^p \frac{I_s(p')(\lambda^0 u_s + \lambda^1 G_s^t(E_s(p')))(dF_s(E_s(p'))/dp')}{\rho + \delta + \lambda^1(1 - F_s(E_s(p')))} dp', \quad \text{where} \quad (25)$$

$$I_s(p) := \exp \left[\int_{\underline{p}}^p \frac{-\lambda^1(dF_s(E_s(p'))/dp')}{\rho + \delta + \lambda^1(1 - F_s(E_s(p')))} dp' \right]. \quad (26)$$

Plugging the relevant expressions in Lemma 2 into equation (8), solving for the skill-dependent values offered by individual recruiting firms, and integrating those

values yields (F_s, F_u) . Taking (F_s, F_u) as given, solving for the optimal strategies by economic agents, and aggregating the worker flow constitutes a fixed point argument. Given the complexity of the overall system, in section 4 the analytical proof of the existence and uniqueness of the result is replaced with numerical experiments with different initial values. This section ends with the characterization of training intensity.

Proposition 1 *The training firm with productivity p chooses a training intensity such that*

$$c'(x(p))(r + \rho + \delta)/\mu + x(p)c'(x(p)) - c(x(p)) = s + \delta(U_s - U_u) \quad (27)$$

$$+ \lambda^1 \int_{E_s^t(p)}^{\bar{E}_s} [z - E_s^t(p)] dF_s(z) - \lambda^1 \int_{E_u(p)}^{\bar{E}_u} [z - E_u(p) - J_u(p)] dF_u(z).$$

Moreover, $x(\bar{p}) < x(p)$ for any $p \in [p, \bar{p})$ if and only if the summation of the workers' and firms' gains from job turnover after training is greater than the summation of their gains before training, that is,

$$\int_{E_s^t(p)}^{\bar{E}_s} [z - E_s^t(p)] dF_s(z) - \int_{E_u(p)}^{\bar{E}_u} [z - E_u(p) - J_u(p)] dF_u(z) > 0. \quad (28)$$

The left-hand side in equation (27) rises convexly from the origin with $x(p)$. The right-hand side is independent of $x(p)$. Equation (27) implies that the training decision made by each firm is affected by the sum of value differentials before and after training, which consists of the value differentials of skilled and unskilled unemployment, $(U_s - U_u)$, and differentials of joint gains from job-to-job transition after and before completing training, $\int_{E_s^t(p)}^{\bar{E}_s} [z - E_s^t(p)] dF_s(z) - \int_{E_u(p)}^{\bar{E}_u} [z - E_u(p) - J_u(p)] dF_u(z)$, as well as productivity improvement through training, s . Interestingly, training intensity rises with the joint expected gains from job-to-job transition after, and falls with the joint gains from job-to-job transition during, training. That the optimal back-loading scheme permits trained workers only jointly efficient job turnover keeps the expected gains positive. Moreover, the sum of expected gains from job turnover by unskilled workers is more likely to be negative due to the training firm's loss, especially when the firm's values are sufficiently high. The former being always positive under the optimal back-loading scheme, the latter can be negative due to the firm's loss, especially when p is sufficiently high or λ^1 is sufficiently small. This implies that condition (28) is not so restrictive, and also explains why the current study, which considers job turnover by unskilled workers, unlike previous studies that considered only job turnover after training, finds the possibility of over-investment in general training.

3 Efficiency Analysis

Consider the problem of the *constrained* social planner who maximizes the present value of the expected output flow throughout the life of a newly born worker. Variables associated with the planner's problem are designated by an asterisk. A typical worker produces b when unemployed, $p - c(x^*(p))$ when employed as unskilled, and $p + s$ when employed as skilled, depending on productivity $p \in [p, \bar{p}]$. Denote by $S_s^*(p)$ ($S_u^*(p)$)

the present value of the output flow of a currently employed, skilled (unskilled) worker with productivity $p \in [\underline{p}, \bar{p}]$.

$$(r + \rho)S_s^*(p) = p + s + \delta(U_s^* - S_s^*(p)) + \lambda^1 \int_p^{\bar{p}} [S_s^*(p') - S_s^*(p)] dH(p'), \quad (29)$$

$$(r + \rho)S_u^*(p) = \max_{x^*(p)} p - c(x^*(p)) + \delta(U_u^* - S_u^*(p)) + \mu x^*(p)(S_s^*(p) - S_u^*(p)) \\ + \lambda^1 \int_p^{\bar{p}} [S_u^*(p') - S_u^*(p)] dH(p'), \quad \text{and} \quad (30)$$

$$(r + \rho)U_i^* = b + \lambda^0 \int_p^{\bar{p}} (S_i^*(p') - U_i^*) dH(p'), \quad \text{for } i \in \{u, s\}. \quad (31)$$

The socially efficient training intensity, $x^*(p)$, is determined by

$$c'(x^*(p)) = \mu(S_s^*(p) - S_u^*(p)). \quad (32)$$

Once training intensity is obtained at each $p \in [\underline{p}, \bar{p}]$, the steady state measures of $\{u_u, u_s, G_u(\cdot), G_s(\cdot)\}$ should be adjusted as follow:

$$\frac{dG_u^*(p)}{dp} = \frac{\lambda^0 u_u^* + \lambda^1 G_u^*(p)}{\rho + \delta + \lambda^1(1 - H(p)) + \mu x^*(p)} \frac{dH(p)}{dp}, \quad \text{and} \quad (33)$$

$$\frac{dG_s^*(p)}{dp} = \frac{\lambda^0 u_s^* + \lambda^1 G_s^*(p) + \mu x^*(p)(dG_u^*(p)/dp)}{\rho + \delta + \lambda^1(1 - H(p))} \frac{dH(p)}{dp} \quad (34)$$

where $G_u^*(\underline{p}) = G_s^*(\underline{p}) = 0$, $u_u^* = \rho\phi(\rho + \delta + \mu x^*(\bar{p}))[(\rho + \delta + \mu x^*(\bar{p}))(\rho + \lambda^0) - \delta\lambda^0]^{-1}$ and $u_s^* = (\rho + \delta)/(\rho + \delta + \lambda^0) - u_u^*$.

Proposition 2 *The constrained social planner chooses a training intensity such that for each $p \in [\underline{p}, \bar{p}]$,*

$$c'(x^*(p))(r + \rho + \delta)/\mu + x^*(p)c'(x^*(p)) - c(x^*(p)) = s + \delta(U_s^* - U_u^*) \quad (35)$$

In particular, $dx^/dp = 0$, $dx^*/d\lambda^1 = 0$, and $dx^*/d\lambda^0 > 0$.*

Proposition 2 implies training intensity in the social planner's problem to be flat regardless of productivity, and affected by acceleration of the job-finding rate of unemployed, but not by acceleration of the job turnover rate of employed, searchers. Intuitively, there being no heterogeneity in training technology across productivity levels, the training intensity in the social planner's problem is affected by neither productivity differentials nor job turnover. If the job-finding rate of unemployed searchers is accelerated, the value of skilled unemployment rises further than that of unskilled unemployment because the opportunity cost of being unemployed is larger in the former than in the latter case. This raises the marginal benefit of training in equation (35) and reinforces training intensity at all productivity levels.

Comparing $x(p)$ in (27) and $x^*(p)$ in (35) results in interesting breakdowns. The difference in training intensity between the market equilibrium and the planner's problem reflects the difference in unemployment value differentials ($U_s - U_u$) and ($U_s^* - U_u^*$) and the difference in the expected joint gains from job-finding. The back-loading compensation scheme, in particular, by imposing a 'proper price of the

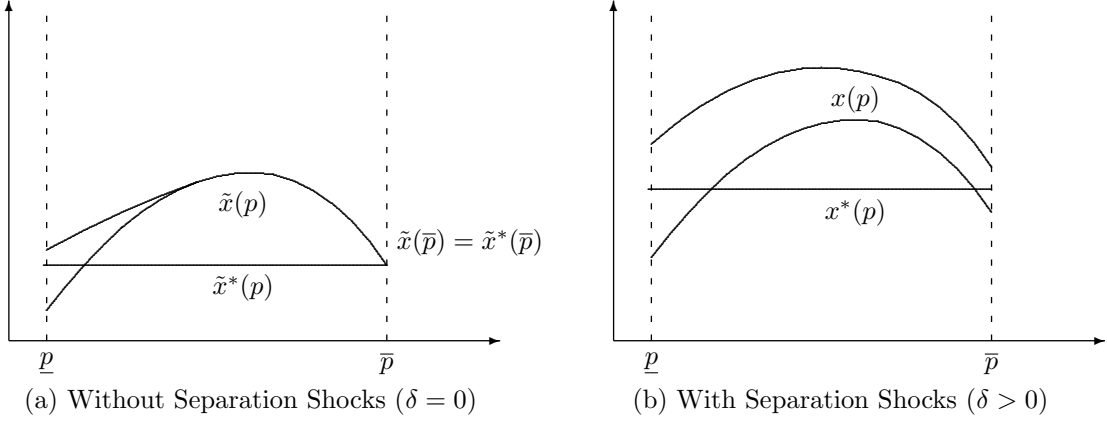


Figure 2: Training Intensity

skill,' enables training firms to internalize the positive externality for subsequent poaching firms. That training firms are not able to raise a claim when their trained workers switch to other firms through unemployment, however, still discourages firms from providing general training (the free-rider problem as in [Acemoglu \(1997\)](#)) if $(U_s - U_u) < (U_s^* - U_u^*)$. To isolate the effect of the training firms' strategic back-loading scheme from the free rider problem raised in the skilled employment-unemployment-employment flow, consider the case without the separation shock, $\delta = 0$, as in [Moen and Rosen \(2004\)](#).¹¹ To keep a positive mass of skilled unemployed workers, it is assumed that a newly born worker enters the labor market as unskilled with probability $\phi \in (0, 1]$ and as skilled with probability $(1 - \phi)$. Variables associated with the benchmark setting (hereafter alternative benchmark) are designated by an overhead tilde ($\tilde{\cdot}$). $(U_s - U_u)$ and $(U_s^* - U_u^*)$ are dropped in equations (27) and (35), enabling a clear comparison by removing the general equilibrium effect of sifting the offer distributions.

Proposition 3 Consider the alternative benchmark setting with $\delta = 0$ and $\phi < 1$.

(i) $\tilde{x}(\bar{p}) = \tilde{x}^*(\bar{p})$.

(ii) $\tilde{x}(p) > \tilde{x}(\bar{p}) = \tilde{x}^*(\bar{p}) = \tilde{x}^*(p)$ for each $p \in [\underline{p}, \bar{p})$ if and only if the sum of the worker's and firm's gains from job turnover after training is greater than the summation of their gains before training, that is,

$$\int_{E_s^t(p)}^{\bar{E}_s} [z - E_s^t(p)] dF_s(z) - \int_{E_u(p)}^{\bar{E}_u} [z - E_u(p) - J_u(p)] dF_u(z) > 0. \quad (36)$$

¹¹[Moen and Rosen \(2004\)](#), to show that the efficient level of training can be provided through coordination between the trainee and training firm, assume only retirement, and not job separation, shock, with the result that all unemployed workers in their model are newly born workers. They further assume that unskilled employees are unable to search for other jobs. These are not innocuous assumptions in deriving their main argument.

Proposition 3 reveals that the joint surplus maximizing training intensity in the market equilibrium outcome can be over-intensified. Because workers at the most productive firm have no gains from on-the-job search behavior, (27) is the same as (35), hence, $\tilde{x}(\bar{p}) = \tilde{x}^*(\bar{p})$ in the alternative benchmark setting. As shown in Proposition 1, the most productive firm provides the least training under condition (36). Training intensity in the market equilibrium is thus over-intensified relative to the socially efficient level under condition (36). If the second term in (36) has a negative value or a sufficiently small positive value, firms provide over-intensified training. As shown in Panel (a) in [Figure 2], which summarizes Proposition 3, more productive firms provide over-intensified general training relative to the social planner. Whether less productive firms, including the least productive firm, provide over- or under-intensified training, however, remains ambiguous.

Corollary 1 *Suppose that $\delta > 0$.*

- (i) $x(\bar{p}) > x^*(\bar{p})$ if and only if $U_s - U_u > U_s^* - U_u^*$.
- (ii) $x(p) > x^*(p)$ for each $p \in [\underline{p}, \bar{p}]$, if and only if

$$\delta(U_s - U_u) + \lambda^1 \int_{E_s^t(p)}^{\bar{E}_s} [z - E_s^t(p)] dF_s(z) - \lambda^1 \int_{E_u(p)}^{\bar{E}_u} [z - E_u(p) - J_u(p)] dF_u(z) > \delta(U_s^* - U_u^*). \quad (37)$$

Corollary 1 circles back to the general environment with the exogenous separation shock. The proof of Corollary 1 is dropped provided that it is straightforward from Proposition 3. Panel (b) in [Figure 2] summarizes Corollary 1. When $\delta > 0$, it is unclear whether the most and least productive firms provide over- or under-intensified training relative to the social planner's training intensity. Because it is affected by the distribution assumption and other parameter values, Corollary 1 provides a sufficient condition.

Proposition 3 and Corollary 1 examine, through a point-wise comparison at each productivity level $p \in [\underline{p}, \bar{p}]$, whether individual firms provide over- or under-intensified general training. To determine social efficiency in aggregation requires a careful choice of the distribution assumption and parameter values. The next section performs a quantitative assessment with a focus on the aggregate market equilibrium.

4 Quantitative Analysis

This section, which calibrates the model and illustrates its implications for social efficiency, shows that (i) the most productive firm provides less intensified training, (ii) aggregate training intensity is over-intensified in the market equilibrium relative to the planner's problem under reasonable choices of the distribution function and parameter values, (iii) the back-loading scheme causes efficiency loss due to over-intensified training and partial deterrence of (socially) efficient job turnover, and (iv) as offer arrival rates to both employed and unemployed searchers are accelerated in the same proportion, social efficiency, measured by net output, is improved.

Table 1: Parameter Values

$[p, \bar{p}]$	the productivity support	$[0.75, 1.75]$
s	productivity improvement through training	0.25
η	the shape parameter of $H(p)$	1.0
μ	human capital accumulation rate	0.03
γ	cost function parameter	2.0
r	interest rate	0.012
ρ	retirement rate	0.008
δ	separation rate	0.064
λ^0	job finding rate by unemployed workers	1.35
λ^1	job finding rate by employed workers	0.45

4.1 Specification and Parameterization

The simulation experiments in this paper proceed with $H(p) : [p, \bar{p}] \rightarrow [0, 1]$, defined as

$$H(p) = \frac{1 - (p/p)^{\eta}}{1 - (p/\bar{p})^{\eta}}, \quad \text{where } \eta > 0. \quad (38)$$

Productivity support is normalized to be the unit length interval, $[p, \bar{p}] = [0.75, 1.75]$, and evenly discretized with 1,001 grid points such that $p_{j+1} - p_j = 10^{-3}$ for each $j = 1, 2, \dots, 1000$. Using data on the entire population of tax-paying firms in the United States, [Axtell \(2001\)](#) shows the size distribution of U.S. firms to be characterized by the Pareto distribution with shape parameter between 0.99 and 1.1. This paper assumes $\eta = 1.0$ following his finding, which assumes a bounded Pareto distribution. Further experiments indicate that a Pareto distribution with lower η results in more over-intensified training. [Hornstein et al. \(2011\)](#) suggest that the mean-min wage ratio, *Mm*-ratio, should lie between 1.7 and 1.9. To accommodate this empirical finding, this paper sets $s = 0.25$, which results in an *Mm*-ratio of 1.77 in cooperation with other parameter values. Given the firm's training cost $c(x)$, the worker acquires skills at rate μx . Without any good reference on firm training cost, this paper proceeds with $\gamma = 2.0$ and $\mu = 0.03$, and checks robustness.

Setting the quarterly interest rate at 0.012, which roughly targets an annual interest rate of 0.048 and quarterly retirement rate at 0.008, implies that the 76 percent of workers who enter the labor market around age 20 retire before age 65. Setting the job finding rate for unemployed workers at 1.35, as in [Shimer \(2005\)](#), implies that the average U.S. unemployment spell is about 10 weeks. The steady state unemployment rate $(\rho + \delta)/(\rho + \delta + \lambda^0)$ is usually targeted around 5 or 6 percent, which fixes the separation rate δ at 0.064. [Sim \(2013\)](#) documents average job duration in the U.S. labor market to be slightly more than two years among white male high school graduates. This paper reconciles his suggestion by choosing $\lambda^1 = 0.45$. This, together with other choices, results in average job duration of eight quarters, which is consistent with the targeted value.

In addition to the baseline simulation, this section analyzes the alternative benchmark setting with $\delta = 0$. To maintain the unemployment rate, ρ is set to be 0.072(=

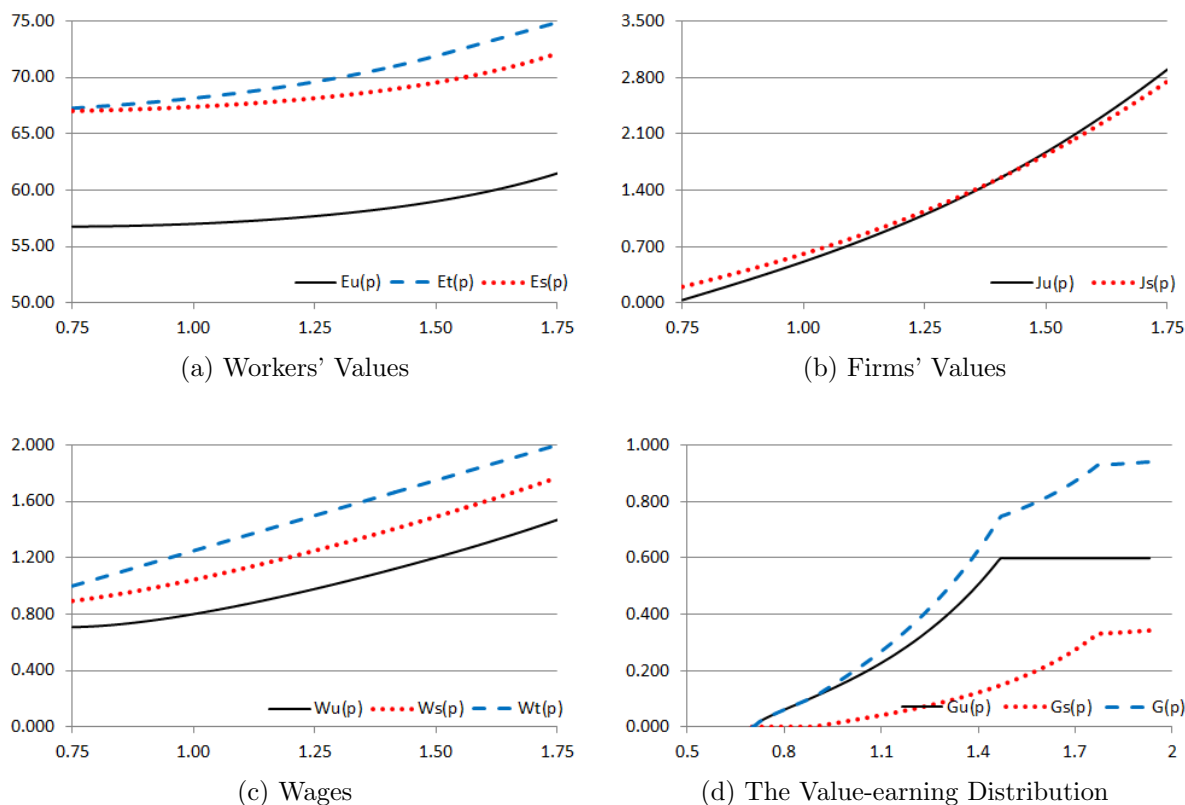


Figure 3: The Baseline Simulation Result

The horizontal axes of Panel (a), (b) and (c) represent firm productivity, and the vertical axes workers' values, firms' values, and wages, respectively. Panel (d) presents the cumulative wage earning distribution.

0.064 + 0.008). It is necessary, when δ is set to be zero, to assume a certain fraction of newly born workers to be skilled in order to avoid a zero mass of the skilled unemployed in the general equilibrium environment. Panel (a) of [Figure 4] reports the implied training intensity under $\phi = 0.4$. But note that the training intensity in the market equilibrium is robustly over-intensified regardless of the choice of ϕ .

4.2 Baseline Simulation

[Figure 3] summarizes the result of the numerical experiments based on the parameter values chosen in the previous section. In [Figure 3], the solid lines represent, in Panels (a), (b), (c), and (d), respectively, workers' values of unskilled employment, firms' values of unskilled jobs, unskilled wages, and the proportion of unskilled workers receiving a certain level of wage or less. The dotted lines represent, in Panels (a), (b), (c), and (d), respectively, workers' values of skilled recruitment, firms' values of jobs with recruited workers, skilled wages received by recruited workers, and proportion of skilled workers receiving a certain level of wage or less. The dashed line represents, in Panels (a), (c), and (d), respectively, the value of, and wage payment to, the promoted worker and proportion of employed workers receiving a certain level of wage or less.

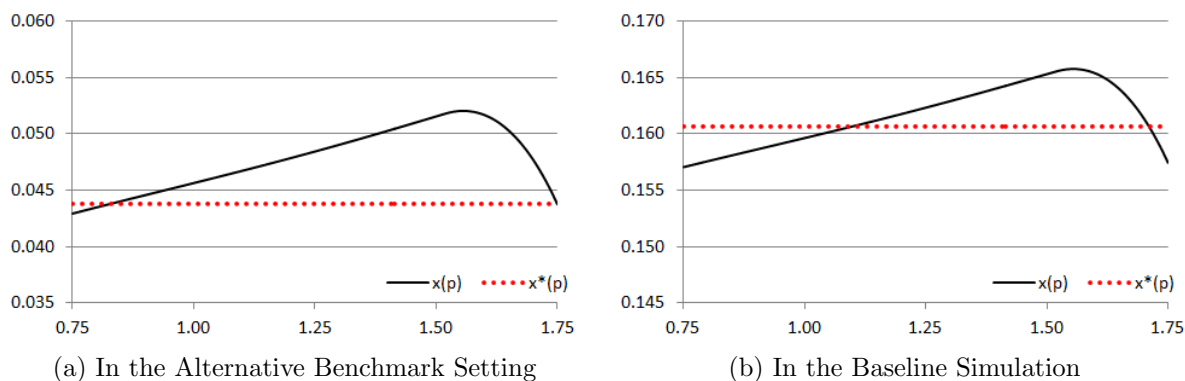


Figure 4: Training Intensity

The horizontal axis in each panel represents firm productivity, and the vertical axis training intensity. Note that each panel has a different range on the vertical axis.

	unskilled workers	skilled workers	training cost	total output	Net output
ME	0.6001	0.3493	0.0159	1.4455	1.4297
PP	0.6027	0.3467	0.0156	1.4458	1.4303
ME/PP	0.9958	1.0074	1.0192	0.9998	0.9995

ME: the market equilibrium outcome PP: the planner's solution ME/PP: the ratio of ME to PP

Table 2: The Outcome of the Baseline Simulation

The value of a firm with a promoted worker, always being zero, is dropped in Panel (b).

Panel (a) shows all lifetime values to increase monotonically with firm productivity, reflecting the high value a productive firm attaches to worker recruitment. Panel (b) shows firm values to also increase monotonically with productivity, but relatively less productive firms to realize higher values from skilled, and more productive firms from unskilled, matches, implying that productive firms more effectively trade wage payments before and after training through the back-loading wage scheme. This is also captured in Panel (c). Although wage payment after internal promotion increases monotonically with firm productivity, the unskilled wage is monotonically increasing depending on the parameter values. That the gap between wages before and after training apparently becomes larger as productivity rises implies that firms with high productivity effectively suppress the unskilled wage by committing to higher skilled wages. Comparing the dotted and dashed lines in Panel (c) reveals the distinct intervals of skilled wages, as in [Figure 1] and also reflected in Panel (d). There being four distinct intervals overlapped with each other, there exist three kink points in the wage support. Shifting the dashed line by the proportion of unemployed workers pushes it up to one at the right end.

In [Figure 4], the solid and dotted lines represent the training intensity in the market equilibrium and in the planner's problem, respectively. Panel (a) in [Figure 4] shows the training intensity of the market equilibrium in the alternative benchmark setting, the

solid line mostly above the dotted line, to exhibit a hump-shaped relationship between productivity and training intensity. This implies that the condition in Proposition 3 is satisfied at all productivity levels except for a short interval at the bottom. The training intensity is characterized by

$$\frac{dx(p)}{dp} = \frac{\mu\lambda^1}{c''(x(p))} \left[\frac{[F_s(E_s^t(p)) - F_u(E_u(p))] \frac{dE^t}{dp} - J_u(p)H'(p)}{r + \rho + \delta + \lambda^1(1 - F_u(E_u(p))) + \mu x(p)} \right]. \quad (39)$$

The first term in the square bracket in (39), $[F_s(E_s^t(p)) - F_u(E_u(p))] \frac{dE^t}{dp}$, is interpreted as an increase in the encouraging effect that leads productive firms to train their workers more intensively in response to the enlarged retention probability and, hence, increase the joint value after training. The second term, $-J_u(p)H'(p)$, captures a decrease in the back-loading pressure that leads productive firms to provide training less intensively consequent to mitigation of the threat posed by potential poaching firms. As p increases, the improvement in the retention probability, $F_s(E_s^t(p)) - F_u(E_u(p))$, goes to zero and $dx(p)/dp < 0$ for any $p \in (\bar{p} - s, \bar{p}]$. This implies that optimal training intensity declines with firm productivity among highly productive firms. But because $F_s(E_s^t(p)) - F_u(E_u(p))$ has a large value among relatively less productive firms (under the Pareto distribution), optimal training intensity increases with productivity p among less productive firms. As predicted by Proposition 2, the dotted line is a flat straight line regardless of productivity, that is, $dx^*(p)/dp = 0$. Panel (b) shows that when $\delta > 0$ and the skilled employment-unemployment-employment flow exists, the most productive firm in the market equilibrium offers under-intensified training relative to the planner's decision.

[Table 2] presents clear evidence of over-investment in general training in the steady state market equilibrium. The first and second columns show there to be more skilled (unskilled) workers in the market equilibrium (planner's problem) as a result of the 'over-intensified training' shown in the third column. Total employment is fixed at $0.9494 (\approx \lambda^0 / (\rho + \delta + \lambda^0))$ in both cases. Although the mass of skilled workers is larger in the market equilibrium than in the planner's solution, total output in the fourth column is smaller in the former than in the latter, indicating that some (socially) efficient job turnover is deterred in the market equilibrium, and lowering the average productivity of skilled matches. Because the negative effect of the higher aggregate cost and partial deterrence of efficient job turnover dominates the positive effect of enhanced skill supply, the market equilibrium results in smaller net output.

4.3 Comparative Static Analysis

[Figure 5] shows how training firms adjust training intensity in response to variation in the magnitude of search friction. The baseline parameterization dictates $\lambda^1 \leq \lambda^0$, which indicates that constraints on search behavior, such as time, render job search less efficient among employed than among unemployed workers. Panel (a) shows the extent to which training firms must adjust training intensity if employed workers are to effectively overcome such hurdles. Training intensity in the market equilibrium is represented by solid, in the planner's problem by dotted, lines. The thin lines follow from the baseline simulation; the thick lines represent training intensity in each setting with 20 percent reinforcement in the offer arrival rate to employed workers. If the job turnover rate of employed workers is accelerated, training firms train workers

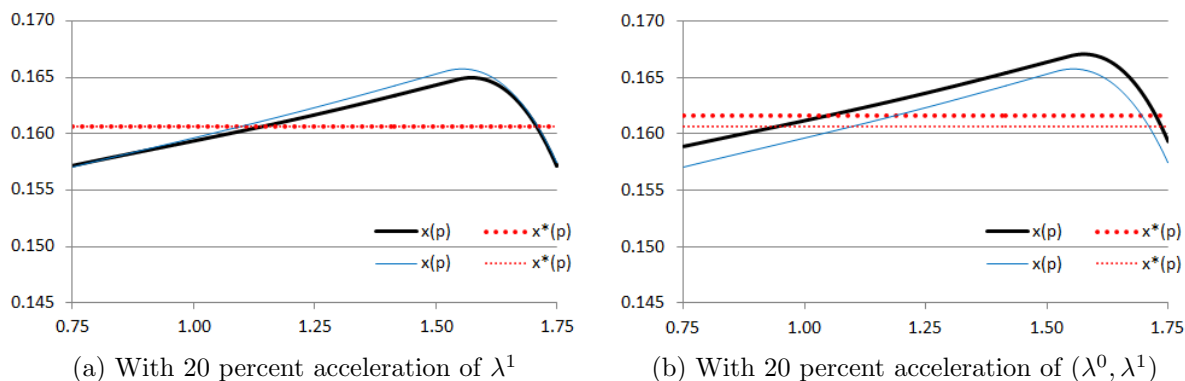


Figure 5: Comparative Statics

The horizontal axis in each panel represents firm productivity, and the vertical axis training intensity.

	unskilled workers	skilled workers	training cost	total output	Net output
With 20 percent acceleration of λ^1					
ME	0.6007	0.3487	0.0158	1.4672	1.4514
PP	0.6027	0.3467	0.0156	1.4677	1.4521
ME/PP	0.9968	1.0056	1.0144	0.9997	0.9995
With 20 percent acceleration of (λ^0, λ^1)					
ME	0.6013	0.3561	0.0162	1.4808	1.4647
PP	0.6048	0.3527	0.0158	1.4809	1.4651
ME/PP	0.9944	1.0096	1.0251	1.0000	0.9997

ME: the market equilibrium outcome PP: the planner's solution ME/PP: the ratio of ME to PP

Table 3: With the Acceleration of the Offer Arrival Rates

less intensively. Intuitively, as the job turnover rate is accelerated expected job duration becomes shorter, which, together with other general equilibrium effects, reduces training intensity in the market equilibrium.

Panel (b) shows how training firms adjust training intensity in response to the improvement of search efficiency among unemployed as well as employed workers. As before, the thin lines represent the baseline simulation, the thick lines training intensity with 20 percent reinforcement. Interestingly, the results reported in Panel (b) are opposite those reported in Panel (a), both thick lines lying above the thin lines. In other words, training firms provide more intensive training as search efficiency improves, whether for employed or unemployed workers. In particular, the accelerated offer arrival rate to unemployed searchers increases their reservation value, forcing recruiting firms to raise the value of their offers to potential employees, and, hence, the values of skilled and unskilled unemployment. But because the opportunity cost of unskilled unemployment is not as great as that of skilled unemployment, the value of the latter

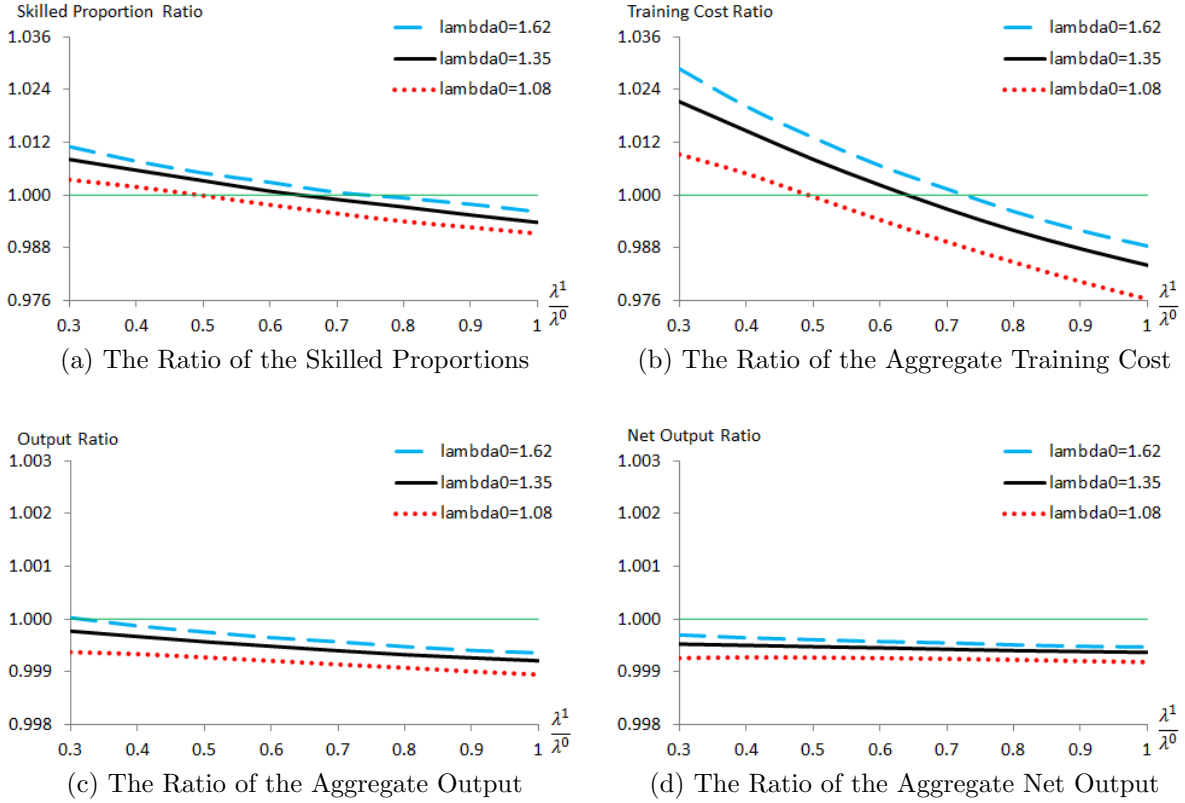


Figure 6: Comparative Statics

The horizontal axis in each panel represents the ratio of λ^1/λ^0 . The different curves in each panel represent the values associated with different intensities of λ^0 . As a result, unilateral improvement in only λ_1 generates ‘movement along the curve’ and proportional improvement in λ_0 and λ_1 ‘shift-up.’

improves further with λ^0 , encouraging training firms to provide more intensive training.

[Table 3] shows the impact of the changes on the market equilibrium outcome. When only λ^1 increases, the market equilibrium gets a smaller fraction of skilled workers but generates more output than before because the skilled workers are more likely to work at more productive firms. Aggregate training intensity decreases, but output increases towards the planner’s outcome. When λ^0 and λ^1 increase together, training intensity, the fraction and average productivity of skilled workers, and outputs all increase and net output improves toward the planner’s outcome.

[Figure 6] further compares the market equilibrium with the planner’s problem. The horizontal axis represents the relative search efficiency, λ^1/λ^0 , the vertical axis the ratio of the market equilibrium outcome to the planner’s outcome, in all four panels. The dotted, solid, and dashed lines represent the market equilibrium outcomes associated with $\lambda^0 = 1.08, 1.35$, and 1.62 , respectively. The thin horizontal line in the middle represents the benchmark case in which the market equilibrium outcome coincides with the planner’s outcome. In Panel (a), the skilled proportion ratio rises with λ^0 , but falls with λ^1/λ^0 , confirming the implication of [Figure 5]. When λ^1/λ^0 is small (large), the market equilibrium outcome possesses a larger (smaller) mass of skilled workers than

the planner’s outcome, which indicates ‘over-investment’ in general training. When λ^0 is sufficiently large but λ^1/λ^0 small, Panel (b) consistently indicates a higher training cost (over-investment) in the market equilibrium. Comparing Panels (a) and (c) raises an interesting point. When λ^1/λ^0 is small, the market equilibrium outcome keeps a larger mass of skilled workers than the planner’s outcome (in Panel (a)), but the former seldom produces more than the latter (in Panel (c)). Total employment being the same in both cases, the efficiency loss caused by deterring efficient job turnover can be inferred to be substantial. The contracting solution’s deterrence of efficient job turnover translates into an aggregate output loss in the market equilibrium relative to the planner’s outcome. But as λ^0 and λ^1 increase in the same proportion maintaining λ^1/λ^0 constant, the aggregate output loss shrinks as the measure of skilled workers increases and the distribution is also improved. Because aggregate training cost declines with λ^1/λ^0 , the net output ratios in Panel (d) are flatter than those in Panel (c).

Overall, improvement in λ_1 generates ‘movement along the curve.’ Starting from the baseline point $(\lambda^0, \lambda^1/\lambda^0) = (1.35, 0.333)$, the ratios of skilled proportion, aggregate training cost, aggregate output, and net output decline along the solid line in each panel. The accelerated job turnover rate reduces both training intensity and the net outputs of the market equilibrium. Improving λ_0 and λ_1 in the same proportion generates a vertical ‘shift-up.’ Finally, as search friction is mitigated, the net output in the planner’s outcome is expected to move toward the Bekerian outcome and the net output in the market equilibrium to catch up with the planner’s net output.

5 Conclusion

This paper develops a job search model with on-the-job training and on-the-job search to analyze how productivity, wage payment, and training intensity interact in response to potential job turnover. It demonstrates that when unskilled, as well as skilled, workers engage on-the-job search behavior and firms are allowed to post skill-dependent wage-training schedule, firms exploit training as a means of back-loading, as much as possible, compensation to extract more surplus in earlier periods. Hence, the back-loaded compensation after training effectively deters potential offers from slightly more productive firms, which encourage firms to make over-investment on general training, as if it were job-specific training. The implied training intensities chosen by heterogeneous firms increase with productivity among relatively less productive, and decrease among more productive, firms.

The market equilibrium suffers from two sources of inefficiency, ‘training inefficiency’ and ‘allocation inefficiency.’ Overall, the market equilibrium provides over-intensified general training relative to the planner’s problem because the latter considers only productivity improvement in determining training intensity, whereas the former exploits general training for purposes of intertemporal substitution as well as productivity improvement. Back-loaded compensation after training effectively deters job offers from slightly more productive firms, rendering general training inefficiently over-intensified (inefficiency in training), as if it were job-specific training. Deterring socially efficient job turnover distorts the distribution of skilled workers (inefficiency in job turnover), which degrades aggregate match quality in the market equilibrium.

The quantitative assessment based on the calibrated model demonstrates that (i) if only the job turnover rate of employed searchers is accelerated, the ratios of training

intensity and net output of the market equilibrium outcome to the planner's outcome fall together, and (ii) if offer arrival rates to employed and unemployed searchers are reinforced proportionally, both ratios rise together. This predicts that as search friction is mitigated due to labor market reform or the arrival of an efficient search technology, aggregate training intensity may be reinforced, but net outputs improve toward the planner's outcome.

A Mathematical Appendix

Proof of Lemma 1 Reordering and rewriting (1) yields that

$$\int (z - U_i) dF_i(z) = \frac{1}{\lambda^0} [(r + \rho)U_i - b] = \int (z - E_i(\underline{p})) dF_i(z). \quad (\text{A1})$$

The last equality follows from $E_u(\underline{p}) = U_u$ and $E_s(\underline{p}) = U_s$. Note that the least productive firm with \underline{p} should offer $(E_u(\underline{p}), E_s(\underline{p})) = (U_u, U_s)$ on any equilibrium. Plugging (A1) into (2), replacing $E_s(\underline{p})$ with U_s , and reordering yields (9). The detailed derivation of (10) is presented in Christensen et al. (2005). (11) is obtained via the same procedure. Given

$$\int z dF_u = \frac{1}{\lambda^0} [(r + \rho + \lambda^0)U_u - b], \quad (\text{A2})$$

combining (3), (5), and (A2) yields

$$E_u(\underline{p}) + J_u(\underline{p}) = \frac{\underline{p} + \delta U_u + \frac{\lambda^1}{\lambda^0} [(r + \rho + \lambda^0)U_u - b] - c(x(\underline{p})) + \mu x(\underline{p}) E^t(\underline{p})}{r + \rho + \delta + \lambda^1 + \mu x(\underline{p})}. \quad (\text{A3})$$

Plugging (A3) into (7) and reordering yields (12). Equations (13)-(16) immediately follow from (9)-(12).

Proof of Lemma 2 Taking the derivative of (19), (20), and (21) with respect to p and applying Leibniz rule results in

$$\frac{dG_u(E_u(p))}{dE_u} \frac{dE_u(p)}{dp} = \frac{\lambda^0 u_u + \lambda^1 G_u(E_u(p))}{\rho + \delta + \lambda^1 (1 - F_u(E_u(p))) + \mu x(p)} \frac{dF_u(E_u(p))}{dE_u} \frac{dE_u(p)}{dp}, \quad (\text{A4})$$

$$\frac{dG_s^r(E_s)}{dE_s} \frac{dE_s}{dp} = \frac{\lambda^0 u_s + \lambda^1 G_s^t(E_s) + \lambda^1 G_s^r(E_s)}{\rho + \delta + \lambda^1 (1 - F_s(E_s))} \frac{dF_s(E_s)}{dE_s} \frac{dE_s}{dp}, \quad \text{and} \quad (\text{A5})$$

$$\frac{dG_s^t(E_s^t(p))}{dE_s^t} \frac{dE_s^t(p)}{dp} = \frac{\mu x(p)}{\rho + \delta + \lambda^1 (1 - F_s(E_s^t(p)))} \frac{dG_u(E_u(p))}{dE_u} \frac{dE_u(p)}{dp}. \quad (\text{A6})$$

Solving the system of differential equations (A4)-(A6) yields Lemma 2.

Proof of Proposition 1 Plugging (3), (5), and (6) into (7) and reordering yields (27). Then, the last two terms in (27) constitutes condition (28).

Proof of Proposition 2 Plugging (29), (30), and (31) into (32) and reordering results in (35). Note that for any $p \in [\underline{p}, \bar{p}]$,

$$\int_{\underline{p}}^{\bar{p}} [(S_s^*(p') - S_s^*(p)) - (S_u^*(p') - S_u^*(p))] dH(p') = 0. \quad (\text{A7})$$

From (35), it is trivially true that $dx^*(p)/dp = 0$ and $dx^*(p)/d\lambda^1 = 0$. Since

$$U_s^* - U_u^* = \frac{\lambda^0}{r + \rho + \lambda^0} \int_{\underline{p}}^{\bar{p}} [S_s^*(p') - S_u^*(p')] dH(p'), \quad (\text{A8})$$

it is obtained that $dx^*(p)/d\lambda^0 > 0$.

Proof of Proposition 3 (i) In the alternative benchmark setting, the right hand side of (27) and (35) are same at $p = \bar{p}$ by construction.

(iii) It's trivially obtained by comparing (27) and (35).

References

- ACEMOGLU, D. (1997): "Training and innovation in an imperfect labour market," *The Review of Economic Studies*, 64, 445–464. 2, 3, 13
- ACEMOGLU, D. AND J. PISCHKE (1999): "Beyond Becker: training in imperfect labour markets," *Economic Journal*, 109, 112–142. 2
- ALTONJI, J. G. AND R. A. SHAKOTKO (1987): "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, 54, 437–59. 2
- AXTELL, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820. 15
- BECKER, G. (1964): *Human capital: A theoretical and empirical analysis, with special reference to education*, University of Chicago Press. 2, 3
- BURDETT, K., C. CARRILLO-TUDELA, AND M. G. COLES (2011): "Human Capital Accumulation And Labor Market Equilibrium," *International Economic Review*, 52, 657–677. 2
- BURDETT, K. AND M. COLES (2003): "Equilibrium Wage-Tenure Contracts," *Econometrica*, 71, pp. 1377–1404. 3, 8
- BURDETT, K. AND D. T. MORTENSEN (1998): "Wage Differentials, Equilibrium Size, and Unemployment," *International Economic Review*, 39, 257–273. 2, 4
- CHRISTENSEN, B. J., R. LENTZ, D. T. MORTENSEN, G. R. NEUMANN, AND A. WERWATZ (2005): "On-the-Job Search and the Wage Distribution," *Journal of Labor Economics*, 23, 31–58. 22
- FALLICK, B. AND C. FLEISCHMAN (2004): "Employer-to-Employer Flow in the U.S. Labor Market: The Complete Picture of Gross Worker Flows," Manuscript, May 2004. 2
- FU, C. (2011): "Training, Search and Wage Dispersion," *Review of Economic Dynamics*, 14, 650–666. 2
- GALENIANOS, M. AND P. KIRCHER (2009): "Directed search with multiple job applications," *Journal of Economic Theory*, 144, 445–471. 4

- HORNSTEIN, A., P. KRUSELL, AND G. L. VIOLANTE (2011): “Frictional Wage Dispersion in Search Models: A Quantitative Assessment,” *American Economic Review*, 101, 2873–98. [15](#)
- LYNCH, L. M. AND S. E. BLACK (1998): “Beyond the incidence of employer-provided training,” *Industrial and Labor Relations Review*, 52, 64–81. [2](#)
- MOEN, E. R. AND A. ROSEN (2004): “Does Poaching Distort Training?” *Review of Economic Studies*, 71, 1143–1162. [2](#), [6](#), [13](#)
- OWAN, H. (2004): “Promotion, Turnover, Earnings, and Firm-Sponsored Training,” *Journal of Labor Economics*, 22, 955–978. [5](#)
- QUERCIOLO, E. (2005): “Training, Turnover, And Search,” *International Economic Review*, 46, 133–143. [4](#)
- SANDERS, C. AND C. TABER (2012): “Life-Cycle Wage Growth and Heterogeneous Human Capital,” *Annual Review of Economics*, 4, 399–425. [2](#)
- SHI, S. (2009): “Directed Search for Equilibrium Wage-Tenure Contracts,” *Econometrica*, 77, 561–584. [3](#), [8](#)
- SHIMER, R. (2005): “The Cyclical Behavior of Equilibrium Unemployment and Vacancies,” *American Economic Review*, 95(1), 25. [15](#)
- SIM, S.-G. (2013): “Wage Dynamics with Private Learning-by-doing and On-the-job Search,” Manuscript, September 2013. [2](#), [15](#)
- STEVENS, M. (2004): “Wage-Tenure Contracts in a Frictional Labor Market: Firms’ Strategies for Recruitment and Retention,” *Review of Economic Studies*, 71, 535–551. [3](#), [8](#)
- TOPEL, R. AND M. WALD (1992): “Job Mobility and the Careers of Young Men,” *Quarterly Journal of Economics*, 107, 439–479. [2](#)