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

Seung-Gyu Sim†
University of Tokyo

January, 2017

Abstract

This paper analyzes the coexistence of on-the-job (general) training and on-the-job search in a frictional labor market where firms post wage-training-promotion contracts to preemptively back-load compensation. The back-loaded compensation discourages trained workers' efficient job-to-job transition, as if they jointly accumulated relationship-specific capital, which induces over-intensified training among more productive firms. The quantitative analysis predicts that the market equilibrium, relative to the efficiency benchmark, gets more skilled workers (training inefficiency) and less output (allocation inefficiency). Aggregate efficiency loss, being offset by future employer externality in the previous literature, is moderate. Interestingly, it can be even improved, as search friction is mitigated.

Keywords: On-the-job Training, On-the-job Search, Wage-Training Contracts
JEL Classification: J24, J31, J64

*I am deeply indebted to Anton Braun, Yongsung Chang, Chao Fu, Junichi Fujimoto, Hidehiko Ichimura, Selahattin Imrohoroglu, Boyan Jovanovic, Michihiro Kandori, John Kennan, Young Sik Kim, Chul-In Lee, Rasmus Lentz, Espen Moen, Serene Tan, Chris Taber, and Yasuyuki Sawada for their insightful guidance and warm encouragement. Especially, I appreciate the hospitality of John Kennan, Rasmus Lentz, and Chris Taber at University of Wisconsin at Madison, a substantial revision having been done during my stay in Madison. I also thank all seminar participants at the 49th Annual Conference of the Canadian Economics Association, the Fourth SOLE/EALE World Conference, the 9th Biennial Conference of Hong Kong Economic Association, the 11th World Congress of The Econometric Society, the CIG End of Year Macroeconomics Conference, Spring 2015 Midwest Macro Meeting, City University of Hong Kong, Hitotsubashi University, Peking University HSBC Business School, Seoul National University, University of Tokyo, University of Wisconsin at Madison, and Yokohama National University. 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.

†Present Address: University of Wisconsin-Madison, 1180 Observatory Drive #7313, Madison, WI 53706. (seunggyusim@wisc.edu, +1-608-571-8883)

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 general training 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 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). She, consistent with Sanders and Taber (2012), predicts general job training to be under-provided in the market equilibrium relative to the constrained social planner's problem.

This paper extends the wage-posting framework proposed by Burdett and Mortensen (1998) and differentiates workers according to skill level. Unlike the ex-post wage-bargaining framework, the wage-posting framework allows the training firm to exploit the first mover-advantage through a contractual arrangement in advance.[5]

[1] For 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.

[2] Lynch 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.

[3] Fallick 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.

[4] An 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. Later, Lentz and Roys (2016) also demonstrate that the market equilibrium may provide an efficient level of general training. Those papers do not account, however, for the possibility of over-investment.

[5] It shares the same motivation with Moen and Rosen (2004). But this paper generalizes the environment by introducing firms' productivity differentials and allowing unskilled workers' job-to-job transition and shows that the market equilibrium may result in over-intensified general training.

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. Training firms exploit the training opportunity to back-load compensation as much as possible and to extract more surplus in earlier periods, especially when unskilled workers engage in on-the-job search behavior.[6] The back-loading compensation scheme suggests that those internally trained workers be paid higher skilled wages than those workers recruited as skilled. As a result, they may turn down offers from slightly more productive poaching firms and stay longer at the firm from which they got training, as if they accumulated *job-specific* or *relationship-specific* human capital through *general* training.

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 market equilibrium is subject to two sources of inefficiency, 'allocation inefficiency' and 'training inefficiency.' The back-loaded compensation discourages efficient job-to-job transition by the workers working at the firms from which they were trained, which deteriorates allocation efficiency. While the social planner considers only productivity improvement in determining training intensity, training firms in the market equilibrium exploit training for the purpose of back-loading compensation as well as improving productivity. Although the free rider problem causes training firms to make under-investment on general training, the additional purpose due to unskilled workers' on-the-job search behavior incentivizes training firms to make over-investment relative to the social planner at the same time. Depending on the relative impact of the former and latter, some firms do provide over-intensified general training, but the others less intensified general training in the market equilibrium. According to the quantitative analysis, the market equilibrium outcome, relative to the constrained social planner's outcome, retains a larger steady state mass of skilled workers (due to training inefficiency), but, by discouraging socially efficient job-to-job transitions, produces fewer units of the actual and net outputs (due to allocation inefficiency). The efficiency loss is, however, not so large as in other search models considering only the free rider problem. Furthermore, it is shown that both types of inefficiency in the model can be improved, as the offer arrival rate is accelerated, which, in contrast to Acemoglu (1997) as well as public fear, enables the decentralized equilibrium in a frictional labor market to move toward the Beckerian outcome, as search friction is mitigated. This result is consistent with those recent findings in Lentz and Roys (2016) and Huegerich

[6]The underlying motivation for the back-loading wage scheme coincides with that of the wage-tenure contract analyzed by Burdett and Coles (2003), Stevens (2004), and Shi (2009). 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 skill-dependent wage schedules rather than tenure-dependent schedules.

and Sim (2016).[7]

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 numerical analysis in Section 4. Section 5 concludes.

2 Model

2.1 Environment

This paper incorporates on-the-job training into the wage-posting framework proposed by Burdett and Mortensen (1998) 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.[8] "Skilled" and "unskilled" are denoted throughout the paper by subscript $i \in \{s, u\}$. Profit-maximizing firms are heterogeneous in terms of productivity, $H(p)$ denoting the proportion of those firms with productivity no greater than p. Assume that $H(\cdot)$ is continuously differentiable with support $(\underline{p}, \overline{p})$. 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. It is assumed that $\underline{p} + s < \overline{p}$. In addition to $(1 - \varepsilon)$-measure of those profit-maximizing firms, there exist ε-measure of noise firms such that they randomly draw and post wages from $[0, \overline{p} + s]$. 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 11), the only requirement being that the noise firms' value offer distribution be continuously differentiable and strictly increasing.[9] From the next subsection, it is assumed that the steady state equilibrium has a sufficiently small ε.

Unemployed Workers An unemployed worker collects unemployment benefit b and finds an employment opportunity at rate λ^0. 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 $F_{i\varepsilon} : \mathbb{R}_+ \to [0, 1]$ be the steady state distribution function of the values offered to i-type workers by both normal and noise firms. The

[7] Lentz and Roys (2016) study human capital accumulation over workers' careers using renegotiation proof employments contracts. They also show that increased labor market friction reduces training in equilibrium. Huegerich and Sim (2016) also show that depending on social norm, firms may provide more intensified (informal) training, as search friction mitigated. They analyze informal training without commitment by assuming that recruiting firms can post only wage payments, *i.e.* the piece rate sharing rule, without specifying training opportunity.

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

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

lifetime value of an unemployed worker is described by

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

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.

Employed Workers An employed worker receives a wage, finds another employment opportunity at rate λ^1, and is laid off at rate δ in every moment. 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

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

Given training intensity x, chosen by the firm, the unskilled employee acquires skills at rate μx, and upon acquiring skills is awarded a new labor contract with a higher value.[10] 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, $w_s^t(p)$, into the value equation (2). We will see in the later that $E_s^t(p) > E_s(p)$ for each $p \in (\underline{p}, \overline{p})$.

The unskilled employed worker receives on-the-job training and unskilled wage w_u. Let $E_u(p)$ be the value of unskilled employment offered by the firm with productivity p. Given $(w_u(p), x(p), E_s^t(p))$, the expected value of unskilled employment is given by

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

The value of unskilled 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.

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

[10] Owan (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.

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

$$rJ_s(p) = p + s - w_s(p) - [\rho + \delta + \lambda^1(1 - F_{s\varepsilon}(E_s(p)))]J_s(p). \tag{4}$$

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 $\gamma > 1$. The value of an unskilled match producing p at every instant is described by

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

subject to constraints (3) and (4). Isolating $w_u(p)$ in equation (3), plugging the expression for $w_u(p)$ into (5), and taking the derivative with respect to E_s^t yields the first order condition $(dF_{s\varepsilon}(E_s^t)/dE_s)J_s(E_s^t, p) \geq 0$, where the strict equality holds only when $J_s(E_s^t, p) = 0$, as long as $F_{s\varepsilon}(\cdot)$ is strictly increasing. The ε-measure of noise firms is introduced to avoid the case in which $F_{s\varepsilon}(\cdot)$ reaches one quickly.[11]

Lemma 1 *The worker internally promoted by the firm with productivity p gets*

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

It implies that the firm leaves the entire surplus for the worker after promotion.

Equation (6) implies that the training firm makes no claim after training, leaving the entire surplus for the trained worker. It is because the training firm gets nothing once the trainee quits for a better paying job, which makes the former discounts future more severely than the latter. The former wants to exploit the match surplus as early as possible. Apparently, no recruiting firms commit the entire surplus to potential employees in the market, which implies that $w_s(p) < p + s = w_s^t(p)$, $E_s(p) < E_s(p) + J_s(p) < E_s^t(p)$, and $J_s(E_s, p) > 0 = J_s(E_s^t, p)$. The training firm's back-loading strategy effectively deters potential job offers from slightly more productive poaching firms, as if the training firm and worker jointly accumulated job-specific or relationship-specific asset.

Given productivity $p \in (\underline{p}, \overline{p})$, the first order condition with respect to x yields

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

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 $p \in (\underline{p}, \overline{p})$, the interior solution is denoted by $x(p)$. Equation (7) implies that the training firm that commits to paying the worker the entire surplus

[11] Suppose there are no noise firms. The most productive firm posts $w_s(\overline{p})(< \overline{p} + s)$ to recruit a skilled worker in a steady state equilibrium. Then, some training firms can prevent their workers' job turnover after training by setting $w^t(\cdot) = w_s(\overline{p})$, which creates a mass point at $E_s(\overline{p})$ in the value offer distribution. Then, the most productive firm, by deviating from the original contract and posting slightly more, can attract those workers at the mass. In order to avoid this situation, we introduce ε-measure of noise firms.

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),[12] however, over-intensified training can be offered in this setting relative to their efficient outcome because of the discrepancy in the job turnover decisions by trained and promoted workers and unskilled workers, that is, $E_s^t(p) > E_s(p) + J_s(p)$ and $E_u(p) < E_u(p) + J_u(p)$. This will be examined in the next section. 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 $p \in (\underline{p}, \overline{p})$.

Recruiting Firms Let $\{G_{u\varepsilon}(E_u), G_{s\varepsilon}(E_s)\}$ denote the proportions of unskilled and skilled workers that receive lifetime values no greater than $\{E_u, E_s\}$, respectively, and $\{u_{u\varepsilon}, u_{s\varepsilon}\}$ the total mass of unskilled and skilled unemployed workers, respectively. Each profit-maximizing recruiting firm with productivity p posts the skill-dependent values (E_u, E_s) to maximize the expected value, $[\lambda^0 u_{s\varepsilon} + \lambda^1 G_{s\varepsilon}(E_s)]J_s(E_s, p) + [\lambda^0 u_{u\varepsilon} + \lambda^1 G_{u\varepsilon}(E_u)]J_u(E_u, p)$. Taking first order conditions with respect to E_s and E_u yields

$$\lambda^0 u_{i\varepsilon} + \lambda^1 G_{i\varepsilon}(E_i) = \Big[\frac{\lambda^0 u_{i\varepsilon} + \lambda^1 G_{i\varepsilon}(E_i)}{r + \rho + \delta + \lambda^1(1 - F_{i\varepsilon}(E_i))}\frac{dF_{i\varepsilon}}{dE_i} + \frac{dG_{i\varepsilon}}{dE_i}\Big]\lambda^1 J_i(E_i, p), \tag{8}$$

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. Finally, aggregating labor contracts posted by $(1 - \varepsilon)$-measure of profit-maximizing normal firms together with those by ε-measure of noise firms restores the value offer distributions $\{F_{u\varepsilon}, F_{s\varepsilon}\}$.

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 $\{F_{i\varepsilon}, G_{i\varepsilon}, u_{i\varepsilon}\}_{i\in\{u,s\}}$ that jointly satisfy the following conditions.*

(i) Given $\{F_{i\varepsilon}\}_{i\in\{u,s\}}$, 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 $\{F_{i\varepsilon}, E_i, U_i\}_{i\in\{u,s\}}$, 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 $\{G_{i\varepsilon}, u_{i\varepsilon}\}_{i\in\{u,s\}}$ and $\{J_i\}_{i\in\{u,s\}}$, each recruiting firm posts a contract which satisfies the first order condition in (8). The noise firms randomly choose how much they post.

(iv) $\{F_{i\varepsilon}, G_{i\varepsilon}, u_{i\varepsilon}\}_{i\in\{u,s\}}$ are stationary and consistent with workers' job turnover decisions, firms' provision of training and vacancy creation.

[12] Moen and Rosen (2004) did not pay attention to the possibility of job-to-job transition by unskilled workers and job separation shocks to skilled workers.

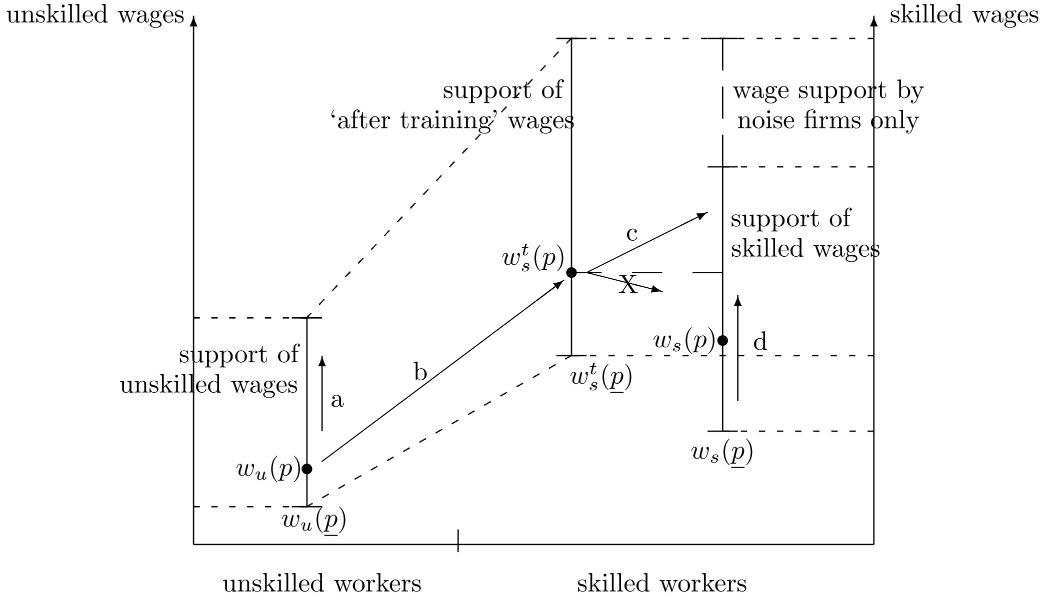


Figure 1: Equilibrium Support of Wages

[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, internally trained skilled workers, and recruited skilled workers, respectively. It depicts two wage supports for skilled workers, one for recruited, the other for promoted, skilled workers. Overall, the latter is located higher than the former. 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 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. The key message of [Figure 1] is that the promoted worker at the firm with productivity p will not accept skilled wage offers between $w_s(p)$ and $w_s^t(p)$, even though those are offered by more productive 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 more severely than their workers so that they trade the surplus before and after training. The strategic (skill-dependent) back-loading scheme[13] leads to general training being exploited as a means of enlarging the expected duration of the match as if it were job-specific training, further reinforcing the investment in general training. Remark that it takes place because the unskilled

[13]Note that while the previous literature posits tenure-dependent back-loading scheme, the current paper assumes skill-dependent back-loading wages.

workers also engage in on-the-job search behavior.

The dashed line on top of the wage support for skilled workers represents the interval that should be covered by those ε-measure of noise firms. Without those noise firms, the wage support for internally promoted workers may have a mass point at the upper bound of the skilled wage support, which incentivizes more productive firms to deviate from the mass point. The only requirement to support the equilibrium of our interest is that the offer distribution should be strictly increasing over the dashed line. Since the measure of noise firms itself does not matter, ε is sent to zero and dropped from the subscript of $\{F_{i\varepsilon}, G_{i\varepsilon}, u_{i\varepsilon}\}_{i=u,s}$ in what follows.

2.2 Characterization of Steady State Equilibrium

It is natural to think that the least productive normal 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 2 *Suppose that (U_u, U_s) are given. The optimal strategy by the least productive firm then implies*

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

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

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

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

where $\hat{F}_s(w(p)) = F_s(E_s(p))$ for any $p \in (\underline{p}, \overline{p})$. The above strategy by the least productive firm with $\underline{p}$ results in

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

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

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

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

The unemployed worker finds a job at rate λ^0; the 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 and skilled unemployment yields

$$(\lambda^0+\rho)u_u = \delta G_u(E_u(\overline{p})) + \rho \quad \text{and} \quad (\lambda^0+\rho)u_s = \delta G_s(E_s(\overline{p})), \tag{17}$$

respectively. Given that $E_u(p)$ is uniquely defined and strictly increasing in p,[14] the unskilled, unemployed worker finds a job with productivity no greater than p at rate $\lambda^0 F_u(E_u(p))$. 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

$$\lambda^0 F_u(E_u(p))u_u = (\rho+\delta+\lambda^1(1-F_u(E_u(p))))G_u(E_u(p)) + \int_{\underline{p}}^{p} \mu x(p')g_u(E_u(p'))dp', \tag{18}$$

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

$$[\rho+\delta+\lambda^1(1-F_s(E_s))]G_s^r(E_s) = \lambda^0 F_s(E_s)u_s + \lambda^1 \int_{E_s^t(\underline{p})}^{E_s} [F_s(E_s)-F_s(z)]dG_s^t(z). \tag{19}$$

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

$$\int_{\underline{p}}^{p} \mu x(p')g_u(E_u(p'))dp' = (\rho+\delta)G_s^t(E_s^t(p)) + \lambda^1 \int_{\underline{E}_s}^{E_s^t(p)} (1-F_s(E_s^t(p)))dG_s^t(z). \tag{20}$$

Taking the derivatives of (18), (19), and (20) 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 3 *Solving the unskilled workers' optimal job turnover decision and firms' optimal training decision yields*

$$u_u = \rho\Big[\lambda^0+\rho+\delta I_u^{-1}(\overline{p})\int_{\underline{p}}^{\overline{p}} \frac{I_u(p')\lambda^0(dF_u(E_u(p'))/dp)}{\rho+\delta+\lambda^1(1-F_u(E_u(p')))+\mu x(p')}dp'\Big]^{-1} \text{ and} \tag{21}$$

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

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

The implied flows of skilled workers are thus described by

$$u_s = \frac{\rho+\delta}{\rho+\delta+\lambda^0} - u_u, \tag{24}$$

[14] Figure 3 shows that both $E_u(\cdot)$, $E_s(\cdot)$, and $E_t(\cdot)$ are strictly increasing in $p \in (\underline{p}, \overline{p})$.

$$G_s^t(E_s^t(p)) = \int_{\underline{p}}^{p} \frac{\mu x(p')(dG_u(E_u(p'))/dp)}{\rho+\delta+\lambda^1(1-F_s(E_s^t(p')))}dp', \quad \text{and} \tag{25}$$

$$G_s^r(E_s(p)) = I_s^{-1}(p) \int_{\underline{p}}^{p} \frac{I_s(p')(\lambda^0 u_s + \lambda^1 G_s^t(E_s(p')))(dF_s(E_s(p'))/dp')}{\rho+\delta+\lambda^1(1-F_s(E_s(p')))}dp', \quad \text{where} \tag{26}$$

$$I_s(p) := \exp\Big[\int_{\underline{p}}^{p} \frac{-\lambda^1(dF_s(E_s(p'))/dp')}{\rho+\delta+\lambda^1(1-F_s(E_s(p')))}dp'\Big]. \tag{27}$$

Plugging the relevant expressions in Lemma 3 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 restores $(u_u, u_s, G_u(\cdot), G_s^t(\cdot), G_s^r(\cdot))$ and 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 guesses. This section ends with the characterization of training intensity.

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

$$c'(x(p))(r+\rho+\delta)/\mu + x(p)c'(x(p)) - c(x(p)) = s + \delta(U_s - U_u) \tag{28}$$
$$+\lambda^1 \int_{E_s^t(p)}^{\overline{E}_s} [z - E_s^t(p)]dF_s(z) - \lambda^1 \int_{E_u(p)}^{\overline{E}_s} [z - E_u(p) - J_u(p)]dF_u(z).$$

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

$$\int_{E_s^t(p)}^{\overline{E}_s} [z - E_s^t(p)]dF_s(z) - \int_{E_u(p)}^{\overline{E}_u} [z - E_u(p) - J_u(p)]dF_u(z) > 0. \tag{29}$$

It also suggests that training intensity rise with the future employer externality before training, and fall with the externality after training.[15]

The left-hand side in equation (28) rises convexly from the origin with $x(p)$. The right-hand side is independent of $x(p)$. Equation (28) implies that the training decision made by each firm is affected by workers' value differentials after and before completing training as well as firms; value differentials, 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 training, $\int_{E_s^t(p)}^{\overline{E}_s}[z - E_s^t(p)]dF_s(z) - \int_{E_u(p)}^{\overline{E}_u}[z - E_u(p) - J_u(p)]dF_u(z)$, 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. In other words, the expected loss from the future employer externality after training discourages firms' provision of general training, whereas the expected loss from the same externality but before completing training encourages provision of general training. In contrast to

[15]The notion of the future employer externality is borrowed from Lentz and Roys (2016).

Acemoglu (1997) and Moen and Rosen (2004) which concentrate on the future employer externality after training without considering the possibility of job-to-job transition by unskilled workers, this result shows that the externality caused by the existence of subsequent employers may encourage or discourage firms' provision of general training.

Furthermore, that the optimal back-loading scheme permits trained workers only jointly efficient job turnover keeps the expected gains positive, which partially internalizes the future employer externality after training. But, 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 (29) 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.[16] 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 $p \in (\underline{p}, \overline{p})$. Denote by $S_s^*(p)$ ($S_u^*(p)$) the present value of the output flow of a currently employed, skilled (unskilled) worker with productivity $p \in (\underline{p}, \overline{p})$.

$$\begin{aligned}
(r+\rho)S_s^*(p) &= p + s + \delta(U_s^* - S_s^*(p)) + \lambda^1 \int_p^{\overline{p}} [S_s^*(p') - S_s^*(p)]dH(p'), && (30)\\
(r+\rho)S_u^*(p) &= \max_{x^*(p)} \; p - c(x^*(p)) + \delta(U_u^* - S_u^*(p)) + \mu x^*(p)(S_s^*(p) - S_u^*(p)) \\
&\quad + \lambda^1 \int_p^{\overline{p}} [S_u^*(p') - S_u^*(p)]dH(p'), \quad \text{and} && (31)\\
(r+\rho)U_i^* &= b + \lambda^0 \int_{\underline{p}}^{\overline{p}} (S_i^*(p') - U_i^*)dH(p'), \quad \text{for} \quad i \in \{u, s\}. && (32)
\end{aligned}$$

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

$$c'(x^*(p)) = \mu(S_s^*(p) - S_u^*(p)). \qquad (33)$$

[16] 'Constrained' implies that the social planner takes the matching process by employed and unemployed workers as given. It also requires that the training decision by the planner at every decision node should be socially efficient. This requirement removes the possibility that the social planner manipulates the training decisions at different matches with different productivity levels to encourage job-to-job transition, which enables us to focus on the efficiency loss associated with on-the-job training free from inefficiency due to the lack of coordination.

Proposition 2 *The constrained social planner chooses a training intensity such that for each $p \in (\underline{p}, \overline{p})$,*

$$c'(x^*(p))(r+\rho+\delta)/\mu + x^*(p)c'(x^*(p)) - c(x^*(p)) = s + \delta(U_s^* - U_u^*) \qquad (34)$$

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,[17] 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 (34) and reinforces training intensity at all productivity levels.

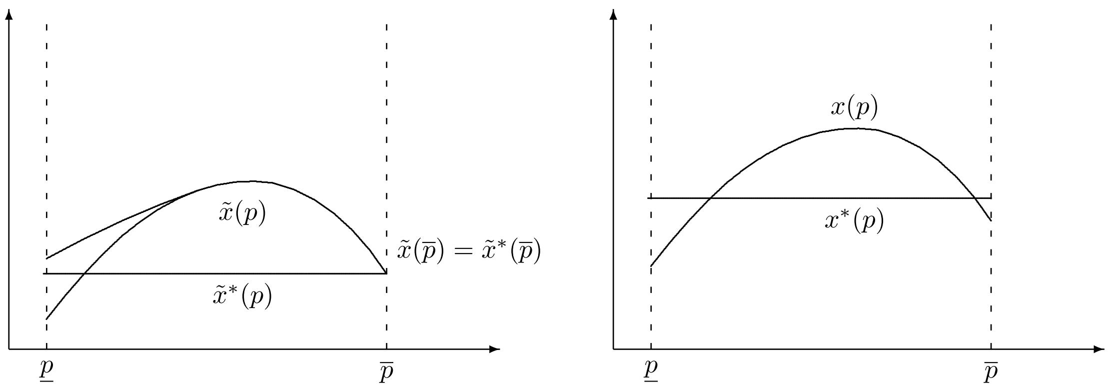
Comparing $x(p)$ in (28) and $x^*(p)$ in (34) results in interesting breakdowns. The difference in training intensity between the market equilibrium and the planner's problem reflects the difference in unemployment value differentials $(U_s - U_u)$ and $(U_s^* - U_u^*)$ and the difference in the expected joint gains from the worker's job-finding. The back-loading compensation scheme, in particular, by manipulating the price of the 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).[18] 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 $(\tilde{\ })$. $(U_s - U_u)$ and $(U_s^* - U_u^*)$ are dropped in equations (28) and (34), enabling a clear comparison by removing the general equilibrium feedback effect of shifting the offer distributions.

Proposition 3 *Consider the alternative benchmark setting with $\delta = 0$ and $\phi < 1$.*
(i) $\tilde{x}(\overline{p}) = \tilde{x}^*(\overline{p}) = \tilde{x}^*(p)$.
(ii) $\tilde{x}(p) > \tilde{x}(\overline{p}) = \tilde{x}^*(\overline{p}) = \tilde{x}^*(p)$ *for each $p \in (\underline{p}, \overline{p})$ if and only if the worker's and*

[17]This result relies on the restriction that the 'constrained' planner should choose the socially efficient training intensity at each decision node. As mentioned before, this restriction, by preventing the planner from manipulating the overall training intensities to facilitate the efficient job-to-job transition, enables us to abstract from efficiency loss due to the lack of coordination.

[18]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.



(a) Without Separation Shocks ($\delta = 0$) (b) With Separation Shocks ($\delta > 0$)

Figure 2: Training Intensity

firm's joint net gains from job turnover after training is greater than their joint gains before training, that is,

$$\int_{E_s^t(p)}^{\overline{E}_s} [z - E_s^t(p)] dF_s(z) - \int_{E_u(p)}^{\overline{E}_u} [z - E_u(p) - J_u(p)] dF_u(z) > 0. \tag{35}$$

In particular, there exists $\tilde{p}' \in (\underline{p}, \overline{p})$ *such that* $\tilde{x}(p) > \tilde{x}^*(p)$ *at every* $p \in (\tilde{p}', \overline{p})$.

Proposition 3 reveals that some firms, particularly firms paying higher wages, provide over-intensified training on the market equilibrium compared to the constrained planner's problem. Because workers at the most productive firm have no gains from on-the-job search behavior, equations (28) and (34) are identical, hence, $\tilde{x}(\overline{p}) = \tilde{x}^*(\overline{p})$ in the alternative setting without separation shocks, (no distortion at the top). As shown in Proposition 1, the most productive firm does not necessarily provide the most intensified training under condition (35). Especially, the back-loaded compensation allows only jointly efficient job-to-job transitions by trained and promoted workers,[19] which makes the first term non-negative at any $p \in (\underline{p}, \overline{p})$. On the other hands, the second term in (35) has a negative value (or a sufficiently small positive value) at the firms with sufficiently high productivity p, because more productive firms lose larger surplus upon workers' quit. As shown in Panel (a) of [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. Suppose that s is so large that $\underline{p} + s > \overline{p}$. In this extreme case, almost no trained and promoted workers can find another better paying firm.[20] While the first term will be zero, the second term in equation (35) has a negative value at $p = \underline{p}$,[21] which may result in over-intensified

[19]Note that $E_s^t(p) = E_s(p) + J_s(p) > E_s(p)$ for any $p \in (\underline{p}, \overline{p})$.

[20]They still can find a better paying 'noisy' firm. But at the limit with $\varepsilon \to 0$, the possibility converges to zero as well.

[21]An extremely large s also increases the value of $J_s(\cdot)$ and $J_u(\cdot)$. Thus, firms' loss is also larger upon

training by the least productive firm. On the other hands, if s is moderate, the least productive firm may provide less intensified training in the market equilibrium compared to the planner's outcome. Indeed, the quantitative analysis in the next section indicates less intensified training by the least productive firm.

Proposition 4 *Suppose that* $\delta > 0$ *and* $\phi = 1$.
(i) $x(\overline{p}) < x^*(\overline{p})$ *if and only if* $U_s - U_u < U_s^* - U_u^*$.
(ii) $x(p) < x^*(p)$ *for each* $p \in (\underline{p}, \overline{p})$, *if and only if*

$$\delta(U_s - U_u) + \lambda^1 \int_{E_s^t(p)}^{\overline{E}_s} [z - E_s^t(p)] dF_s(z) - \lambda^1 \int_{E_u(p)}^{\overline{E}_u} [z - E_u(p) - J_u(p)] dF_u(z) < \delta(U_s^* - U_u^*). \tag{36}$$

Proposition 4 circles back to the general environment with the exogenous separation shock. The proof of Proposition 4 is dropped provided that it is straightforward from Proposition 3. Panel (b) in [Figure 2] summarizes Proposition 4. The existence of employment-unemployment-employment flow aggravates the free-rider problem and dampens firms' incentives to provide training. 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, Proposition 4 provides a necessary and sufficient condition.

Propositions 3 and 4 examine, through a point-wise comparison at each productivity level $p \in (\underline{p}, \overline{p})$, whether individual firms provide over- or under-intensified general training. The result indicates that firms, by exploiting the training intensity as a means of back-loading compensation, may provide over-intensified training (training inefficiency) and effectively deter socially efficient job-to-job transition by trained workers (allocation inefficiency). Interestingly, as the offer arrival rate to employed searchers λ^1 is accelerated, those firms reduce their training intensities and their employees switch to more productive firms more quickly. Unlike the point-wise comparison, the overall efficiency associated with the market equilibrium can be measured based on the steady state measures. Once training intensity is obtained at each $p \in (\underline{p}, \overline{p})$, the steady state measures of skilled and unskilled, employed and unemployed workers should be adjusted. Given those steady state measures, the offer distributions are obtained. Then, the aggregate training intensity per unskilled worker not only in the social planner's problem but also in the decentralized market equilibrium can be restored by solving a fixed point problem in a functional space. Since this requires numerical approach, the next section conducts numerical analysis with a focus on the aggregate market equilibrium.

4 Quantitative Analysis: Aggregation

This section, which calibrates the model and illustrates its implications for social efficiency, shows that (i) the back-loading scheme causes efficiency loss due to over-intensified training and partial deterrence of (socially) efficient job turnover, (ii) the

workers' quit, when s is extremely large.

Table 1: Parameter Values

$[\underline{p}, \overline{p}]$	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

aggregate efficiency loss is moderate in a sharp contrast to the substantial loss predicted by the search model considering only the future employer externality after training, and (*iii*) as job turnover rates by employed searchers are accelerated, social efficiency, measured by net output, is improved.

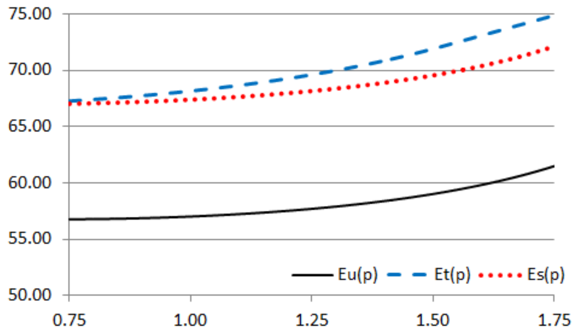
4.1 Specification and Parameterization

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

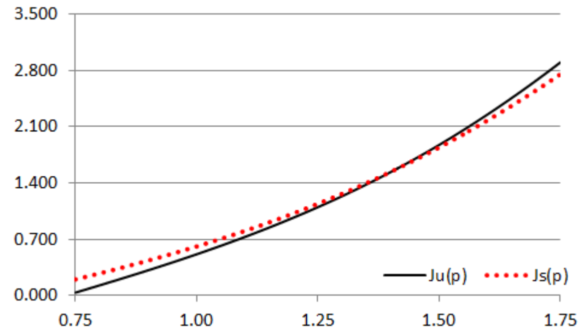
$$H(p) = \frac{1 - (\underline{p}/p)^\eta}{1 - (\underline{p}/\overline{p})^\eta}, \quad \text{where} \quad \eta > 0. \tag{37}$$

Productivity support is normalized to be the unit length interval, $[\underline{p}, \overline{p}] = [0.75, 1.75]$, and evenly discretized with 1,001 grid points such that $p_{j+1} - p_j = 10^{-3}$ for each $j = 1, 2, \cdots, 1000$. 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 which assumes a bounded Pareto distribution sets $\eta = 1.0$ following his finding. 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.

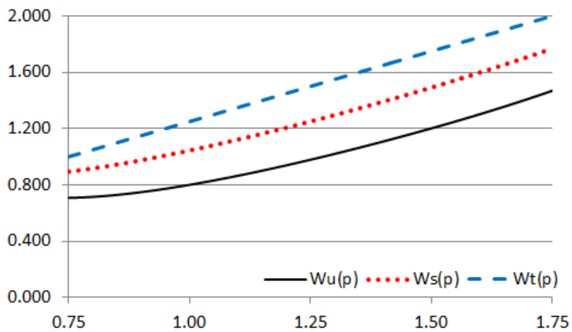
Setting the quarterly interest rate at 0.012, which roughly targets an annual interest rate of 0.048 and quarterly retirement rate at 0.008, implies that the 76 percent of workers who enter the labor market around age 20 retire before age 65. Setting the job finding rate for unemployed workers at 1.35, 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.



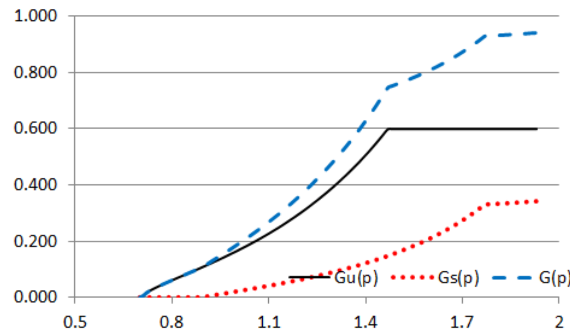
(a) Workers' Values



(b) Firms' Values



(c) Wages



(d) The Value-earning Distribution

Figure 3: The Baseline Simulation Result

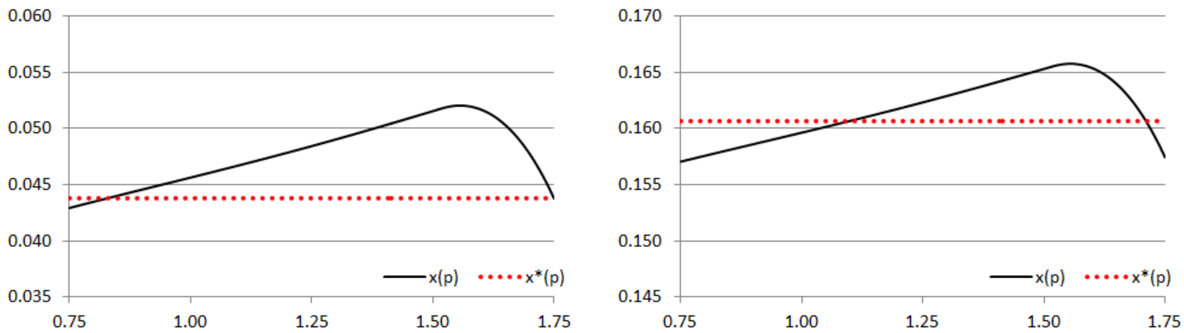
The horizontal axes of Panel (a), (b) and (c) represent firm productivity, and the vertical axes workers' values, firms' values, and wages, respectively. Panel (d) presents the cumulative wage earning distribution.

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 $0.072 (= 0.064 + 0.008)$. 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,



(a) In the Alternative Setting (b) In the Baseline Simulation

Figure 4: Training Intensity

The loci of training intensity in each panel are depicted based on the parameter values presented in Table 1. The horizontal axis in each panel represents firm productivity, and the vertical axis training intensity. The dotted line represents the training intensity in the planner's problem and the solid line the training intensity in the market equilibrium. Note that each panel has a different range on the vertical axis.

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

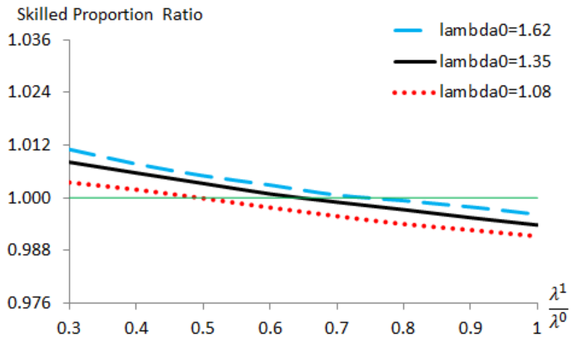
Panel (a) shows all lifetime values to increase monotonically with firm productivity, reflecting the high value that a productive firm attaches to worker recruitment. Panel (b) shows firm values to also increase monotonically with productivity, but relatively less productive firms to realize higher values 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

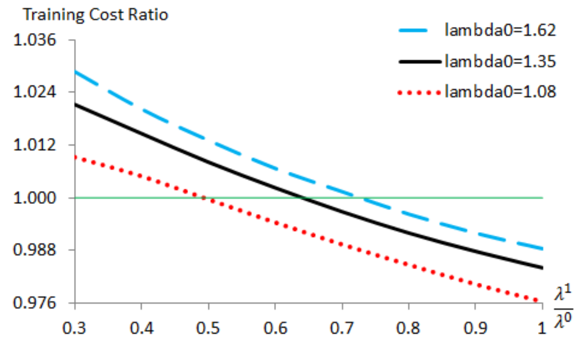
$$\frac{dx(p)}{dp} \quad = \quad \frac{\mu\lambda^1}{c''(x(p))}\Big[\frac{[F_s(E_s^t(p)) - F_u(E_u(p))]\frac{dE^t}{dp} - J_u(p)H'(p)}{r+\rho+\delta+\lambda^1(1-F_u(E_u(p)))+\mu x(p)}\Big]. \tag{38}$$

The first term in the square bracket in (38), $[F_s(E_s^t(p)) - F_u(E_u(p))]\frac{dE^t}{dp}$, 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 of potential job-to-job transition. 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 $p \in (\overline{p} - s, \overline{p}]$. 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 planer's decision.

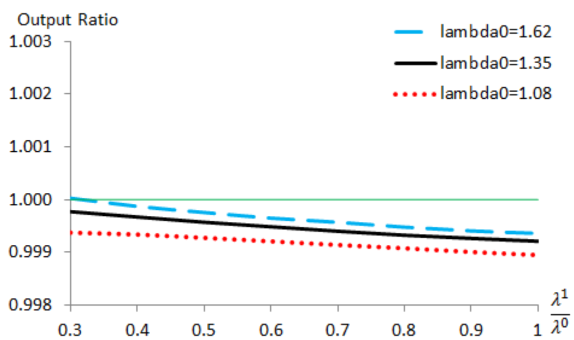
[Figure 5] further shows the (relative) impact of mitigating search frictions on the market equilibrium outcome. The horizontal axis represents the relative search efficiency, λ^1/λ^0, the vertical axis the ratio of the market equilibrium outcome to the planner's outcome, in all four panels. The dotted, 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. 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 in reality. When λ^0 is sufficiently large but λ^1/λ^0 small, Panel (b) consistently indicates a higher training cost in the market equilibrium (training inefficiency). 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 (allocation inefficiency). 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).



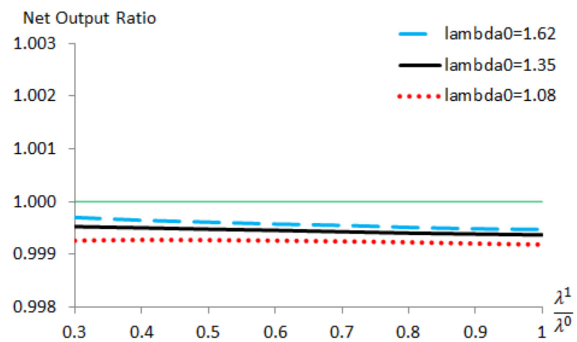
(a) The Ratio of the Skilled Proportions



(b) The Ratio of the Aggregate Training Cost



(c) The Ratio of the Aggregate Output



(d) The Ratio of the Aggregate Net Output

Figure 5: 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.’

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.

A Mathematical Appendix

Proof of Lemma 1 Rewriting equation (3) yields

$$w_u(p) = (r+\rho+\delta+\mu x(p)+\lambda^1)E_u(p) - \mu x(p)E_s^t(p) - \delta U_u - \lambda^1 \int \max[z, E_u(p)]dF_{u\varepsilon}(z). \tag{A1}$$

Plugging (A1) into (5) and taking the derivative with respect to E_s^t yields

$$\frac{dF_{s\varepsilon}(E_s^t)}{dE_s} J_s(E_s^t, p) \geq 0. \tag{A2}$$

Since $dF_{s\varepsilon}/dE_s > 0$, the strict equality in (A2) holds only when

$$J_s(p) = \frac{p + s - w_s(p)}{r+\rho+\delta+\lambda^1(1 - F_{s\varepsilon}(E_s(p)))} = 0. \tag{A3}$$

It implies that $w^t(p) = p + s$. Then, plugging $w^t(p) = p + s$ into (2) yields $E_s^t(p)$.

Proof of Lemma 2 Reordering and rewriting (1) yields that

$$\int (z - U_i)dF_i(z) = \frac{1}{\lambda^0}[(r+\rho)U_i - b] = \int (z - E_i(\underline{p}))dF_i(z). \tag{A4}$$

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 (A4) 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

$$\int z dF_u = \frac{1}{\lambda^0}[(r+\rho+\lambda^0)U_u - b], \tag{A5}$$

combining (3), (5), and (A5) yields

$$E_u(\underline{p}) + J_u(\underline{p}) = \frac{\underline{p} + \delta U_u + \frac{\lambda^1}{\lambda^0}[(r+\rho+\lambda^0)U_u - b] - c(x(\underline{p})) + \mu x(\underline{p})E^t(\underline{p})}{r+\rho+\delta+\lambda^1+\mu x(\underline{p})}. \tag{A6}$$

Plugging (A6) into (7) and reordering yields (12). Equations (13)-(16) immediately follow from (9)-(12).

Proof of Lemma 3 Taking the derivative of (18), (19), and (20) with respect to p and applying Leibniz rule results in

$$\begin{aligned}
\frac{dG_u(E_u(p))}{dE_u}\frac{dE_u(p)}{dp} &= \frac{\lambda^0 u_u + \lambda^1 G_u(E_u(p))}{\rho+\delta+\lambda^1(1-F_u(E_u(p)))+\mu x(p)}\frac{dF_u(E_u(p))}{dE_u}\frac{dE_u(p)}{dp}, && \text{(A7)}\\
\frac{dG_s^r(E_s)}{dE_s}\frac{dE_s}{dp} &= \frac{\lambda^0 u_s + \lambda^1 G_s^t(E_s) + \lambda^1 G_s^r(E_s)}{\rho+\delta+\lambda^1(1-F_s(E_s))}\frac{dF_s(E_s)}{dE_s}\frac{dE_s}{dp}, \quad \text{and} && \text{(A8)}\\
\frac{dG_s^t(E_s^t(p))}{dE_s^t}\frac{dE_s^t(p)}{dp} &= \frac{\mu x(p)}{\rho+\delta+\lambda^1(1-F_s(E_s^t(p)))}\frac{dG_u(E_u(p))}{dE_u}\frac{dE_u(p)}{dp}. && \text{(A9)}
\end{aligned}$$

Solving the differential equation (A7) yields equations (22) and (23) in Lemma 3. Plugging (22) into (17) and isolating u_u yields equation (21). Similarly, solving the system of differential equations (A8) and (A9) yields (25) and (27) in Lemma 3. Combining those with (17) yields (24).

Proof of Proposition 1 Plugging (3), (5), and (6) into (7) and reordering yields (28). Then, the last two terms in (28) constitutes condition (29).

Proof of Proposition 2 Plugging (30), (31), and (32) into (33) and reordering results in (34). Note that for any $p \in [\underline{p}, \overline{p}]$,

$$\int_p^{\overline{p}} \Big[(S_s^*(p') - S_s^*(p)) - (S_u^*(p') - S_u^*(p))\Big] dH(p') = 0. \tag{A10}$$

From (34), it is trivially true that $dx^*(p)/dp = 0$ and $dx^*(p)/d\lambda^1 = 0$. Since

$$U_s^* - U_u^* = \frac{\lambda^0}{r+\rho+\lambda^0}\int_{\underline{p}}^{\overline{p}} \Big[S_s^*(p') - S_u^*(p')\Big] dH(p'), \tag{A11}$$

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

Proof of Proposition 3 *(i)* In the alternative setting without separation shocks, the right hand side of (28) and (34) are same at $p = \overline{p}$ by construction.
(ii) Condition (35) is trivially obtained by comparing (28) and (34). In particular, the first term in (35) is always non-negative, whereas the second term should be negative around $\overline{p}$. More specifically, there exists $\epsilon > 0$ such that the second term has a negative value for any $p \in (\overline{p} - \varepsilon, \overline{p})$ because $J_u(p)$ is strictly positive, the value of $\int_{E_u(p)}^{\overline{E}_u}[z - E_u(p)]dF_u(z)$ is zero at $p = \overline{p}$ and the whole second term is continuous. Consequently, there exists $\varepsilon > 0$ such that $\tilde{x}(p) > \tilde{x}^*(p)$ for any $p \in (\overline{p} - \varepsilon, \overline{p})$.

References

Acemoglu, D. (1997): "Training and innovation in an imperfect labour market," *The Review of Economic Studies*, 64, 445–464. 2, 3, 12, 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. 16

Becker, G. (1964): *Human capital: A theoretical and empirical analysis, with special reference to education*, University of Chicago Press. 2

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. 21

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, 3

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. 16

Huegerich, T. and S.-G. Sim (2016): "On-the-Job Training without Commitments: Investment or Compensation?" Manuscript, November 2016. 3, 4

Lentz, R. and N. Roys (2016): "Training and Search On the Job," Manuscript, June 2016. 2, 3, 4, 11

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, 7, 12, 13

Owan, H. (2004): "Promotion, Turnover, Earnings, and Firm-Sponsored Training," *Journal of Labor Economics*, 22, 955–978. 5

Quercioli, 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. 16

Sim, S.-G. (2013): "Wage Dynamics with Private Learning-by-doing and On-the-job Search," Manuscript, September 2013. 2, 16

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