

# Job Matching within and across Firms, International Economic Review, 2015

Wenzhi Wang \*

August 30, 2024

Pastorino (2015)

## 1. Introduction

Over their careers, workers typically advance through a hierarchy of jobs within firms and turn over across firms, experiencing both wage increases and decreases. These two types of job mobility within and across firms have been largely analyzed as distinct phenomena by two virtually separate literatures on so-called *internal* and *external* labor markets. In this article, I propose a simple unified framework that integrates both dimensions of job dynamics in the presence of uncertainty and learning about workers' ability. I characterize equilibrium assignment and wages for varying degrees of generality of ability across jobs and firms. I also explore how differences in the speed of learning about ability across jobs and firms affect salient features of the dynamics of jobs and wages. The implications of the model are consistent with well-documented features of careers, cast new light on common findings, and explain patterns of careers previously thought difficult to reconcile with the presence of uncertainty and learning.

This unified framework addresses a shortcoming of the two literatures mentioned above: Models of job mobility within firms are mostly silent about worker turnover, whereas models of turnover across firms commonly ignore job mobility in firms. By relaxing the assumption that ability is either fully general or fully specific to jobs or firms, the framework proposed can simultaneously explain findings on careers in firms as well as on turnover across firms and occupations. This framework also suggests a new endogenous mechanism of compensating wage differentials for the *differential informativeness* of jobs that leads wages to deviate from a worker's (expected) output in an otherwise frictionless environment. *This mechanism gives rise to discrete wage increases at promotion in a firm, and discrete wage changes upon turnover across firms, for workers with any level of experience.* Since the ability of experienced workers may

---

\*This note is written in my MPhil period at the University of Oxford.

be known precisely, this feature of wage dynamics has long been considered hard to explain simply based on job matching and learning about ability.

*Notes: I don't really understand how this mechanism leads to discrete increase in wages.*

Formally, I consider a competitive overlapping generations economy in which a worker's ability (either high or low) is initially unknown to all agents, including that worker. Firms operate technologies that are described by a collection of jobs, which produce output and information. These technologies display complementarity between ability and jobs so that jobs can be ordered by the additional output produced by a worker of high ability relative to a worker of low ability.

When a worker is employed, a two-sided heterogeneous trade in labor services and information takes place. The worker provides labor services to a firm, the perceived quality of which is summarized by the common belief about his ability. The firm provides information to the worker that is indexed by the informativeness of the assigned job. Hence, workers select into jobs, in part, because of the information acquisition possibilities that employment offers. As a result, workers' career paths reflect a gradual process of accumulation of information about their suitability for the jobs in the economy, which leads workers to optimally advance through jobs both within and across firms.

I study first a simple *general ability* model in which ability is general across firms and jobs, and jobs are equally informative about ability. In this case, equilibrium assignment follows a *hierarchy rule*: Success leads to promotion to higher-ranked jobs in a firm and possibly turnover to other firms, at which workers are paid higher wages and high-ability workers have greater comparative advantage. Similarly, failure leads to demotion to lower-ranked jobs and possibly turnover to other firms, at which workers are paid lower wages and low-ability workers have greater comparative advantage.

This model captures several observed features of careers in firms. For instance, for individuals continually employed at a firm, wages increase at promotion, real wage decreases are common but demotions are rare, and promotions and wage increases are serially correlated. In addition, when firms differ in their technologies, workers turn over across firms, and their turnover is associated with either wage increases (in the case of successful workers) or wage decreases (in the case of unsuccessful workers). This version of the model also generates a cross-sectional distribution of wages that is positively skewed, as in the data, and more skewed as the chance of failure for young workers increases. Intuitively, when success is difficult to achieve, high wages reward the relatively small proportion of individuals whose ability is perceived to be high.

In this general ability model, job and wage changes within a firm are mainly generated by the assumption that ability is general across a firm's jobs, whereas turnover and the associated dynamics of wages follow from the assumption that ability is general across firms. Some evidence challenges this latter assumption.

Motivated by this evidence, I consider a *firm-specific* version of the model in which a worker's ability is specific to firms but general across a firm's jobs. A novel implication of this model is that equilibrium job assignment has now a two-part form. First, for each firm, a modified Gittins index is computed, and a worker is assigned to the firm with the highest value of such index. Second, given the assigned firm based on this index, the assigned job within the firm follows a within-firm hierarchy rule. Suc-

cess in a given firm increases the index of that firm only, whereas failure lowers it. Hence, continual success leads workers to move up the hierarchy of jobs of a firm but remain continuously employed at a same firm. Continual failure, instead, leads to demotion within a given firm until the modified Gittins index of the firm drops below that of the next-best firm, in which case a worker turns over to the next-best firm. Thus, by retaining the assumption that ability is general across a firm's jobs, the model still generates the key features of careers in firms reproduced by the general ability model. By allowing ability to be specific to firms, the model also gives rise to a declining hazard of separation with tenure in a firm and positive returns to tenure.

More recent empirical evidence suggests that ability may be specific to occupations instead of firms. Based on this evidence, I also analyze an *occupation-specific* version of the model, in which a worker's ability is specific to occupations, interpreted as collections of jobs with similar skill requirements, and each firm has a collection of jobs of possibly different occupations. In this case, equilibrium assignment follows a modified Gittins index rule for occupational choice and a hierarchy rule for job assignment within each occupation. This version of the model gives rise to a hazard of occupational switching that eventually decreases with tenure in an occupation and experience in the labor market as well as returns to tenure in occupation. This model also matches the observation that the relatively low and high wage earners in an occupation have the highest probability of leaving it.

Although the framework proposed reproduces salient features of careers by allowing for varying degrees of generality of ability across jobs and firms, one important aspect of the data is known to be difficult to reconcile with learning models of careers. In firms, wage increases at promotion tend to rise with the level of a job in a firm's hierarchy of positions, and so they are also larger for workers with more tenure and labor market experience, typically assigned to higher levels. Models that assume jobs are equally informative about ability are partly at odds with this feature of the data. The reason is that after a number of periods of observation of a worker's output, a worker's ability should be fairly well known. Then additional output observations on workers with greater tenure and labor market experience should have a small impact on the market expectation about their ability and, thus, on their wages, leading to progressively smaller wage increases at promotion.

To address this limitation, I consider a *differential learning* version of the model in which jobs differ in informativeness, that is, in the information about ability that job performance conveys. I show that the resulting different speed of learning across jobs gives rise to compensating wage differentials across jobs for their different informativeness. This mechanism can naturally account for the pattern of wage increases at promotion just discussed. Intuitively, when higher-level jobs in a firm's hierarchy are less informative about ability than lower-level ones, each promotion compensates a worker for the loss in information associated with the new assignment. As a result, wage premia in the form of discontinuous increases in wages accrue upon promotion from a *more* informative job to a *less* informative one to compensate a worker for the lost information. In this case, equilibrium assignment is no longer solely dictated by the static comparative advantage of workers across jobs. Rather, the trade-off between output and information implied by job assignment leads workers to be assigned to jobs at which they have a static comparative disadvantage but to be dynamically compensated for the associated current loss in wages with the greater information that these jobs provide as well as with wage premia at promotion.

## 2. The Economy

### 2.1. Setup

Consider a market in which  $N$  firms compete for workers in each period of an infinite (or finite) horizon, with dates  $t \geq 1$ . All firms and workers share a common discount factor,  $\delta \in [0, 1)$ . In the economy, one good is produced and consumed, the price of which is normalized to one. A worker's talent or productive ability at each firm  $f$  is described by the skill type  $s \in \{a, b\}$ , which is assumed to be unobserved by firms and workers – including the given worker. I refer to a worker of type  $a$  as a high-ability worker and a worker of type  $b$  as a low-ability worker; later, I consider the case in which ability is described by a vector of skill types. Each worker inelastically supplies one unit of labor each period. At  $t = 1$ , the information available about a worker's skill type is summarized by the prior probability  $p_1$  that the worker is of high ability. I assume that there is a continuum of workers with measure one. I elaborate on the structure of the population in the economy in the next section.

Each firm is endowed with a production set that consists of  $K_f$  tasks or jobs. A worker's output in a period is stochastic, and its distribution depends on the worker's ability, the employing firm, and the assigned job. The output produced in period  $t$ , when the worker is employed at job  $k$  of firm  $f$ ,  $1 \leq k \leq K_f$ , is denoted by  $y_{fk} \in \{y_{fHk}, y_{fLk}\}$ , with  $y_{fHk} > y_{fLk} > 0$ . The probability of high output at job  $k$  of firm  $f$  is  $\alpha_{fk}$  for a worker of ability  $a$  and  $\beta_{fk}$  for a worker of ability  $b$ , with  $0 < \beta_{fk} < \alpha_{fk} < 1$ . Thus, a job is a four-tuple  $(\theta_{fk}, y_{fHk}, y_{fLk})$ , where  $\theta_{fk} = (\alpha_{fk}, \beta_{fk})$  records the *informativeness* of the job. Depending on how these job parameters vary across jobs and firms, the productivity of a worker of a given ability correspondingly varies. I assume that a worker's output, or performance, is symmetrically observed by all firms and workers, and denote by  $p_t$  the prior probability that firms and workers assign to a worker being of high ability at the beginning of period  $t$  or, in short, the *prior* at the beginning of period  $t$ . Since  $\alpha_{fk} > \beta_{fk}$ , observing high output,  $y_{fHk}$ , in any period at the job of any firm increases  $p_t$ .

At the beginning of period  $t \geq 1$ , the one-period expected output of a worker assigned to job  $k$  of firm  $f$ , conditional on his ability being high and low, respectively, is

$$\bar{y}_f(a, k) = \alpha_{fk} y_{fHk} + (1 - \alpha_{fk}) y_{fLk}, \quad \bar{y}_f(b, k) = \beta_{fk} y_{fHk} + (1 - \beta_{fk}) y_{fLk}. \quad (1)$$

Given  $p_t$ , the one-period expected output  $y_f(p_t, k)$  of a worker at job  $k$  of firm  $f$  in  $t \geq 1$ , averaged over the two skill types of the worker, is affine in the prior  $p_t$  and given by

$$y_f(p_t, k) = p_t \bar{y}_f(a, k) + (1 - p_t) \bar{y}_f(b, k) = c_{fk} + d_{fk} p_t, \quad (2)$$

with  $c_{fk} = \bar{y}_f(b, k)$  and  $d_{fk} = \bar{y}_f(a, k) - \bar{y}_f(b, k)$ . Note that  $f$  and  $k$  in  $y_f(p_t, k)$  simply index  $(\theta_{fk}, y_{fHk}, y_{fLk})$  so that  $y_f(p_t, k)$  is shorthand for  $y(p_t, \theta_{fk}, y_{fHk}, y_{fLk})$ .

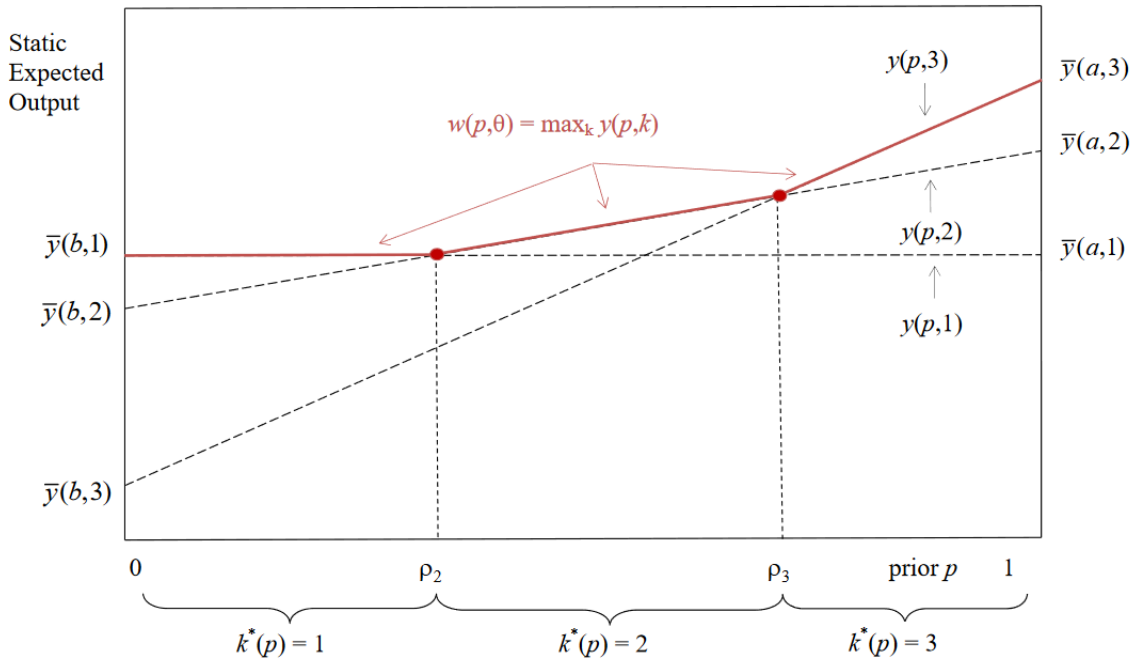
The assumptions that  $y_{fHk} > y_{fLk}$  and  $\alpha_{fk} > \beta_{fk}$  imply that a high-ability worker has an *absolute advantage* at all jobs. Hence, expected output,  $y_f(p_t, k)$ , increases with  $p_t$ . This pattern of absolute advantage implies that in equilibrium, wages increase after high output but decrease after low output

in a job, as often observed. I also assume *complementarity* between ability and jobs in that

$$\bar{y}_f(a, k+1) \geq \bar{y}_f(a, k), \quad \bar{y}_f(b, k) \geq \bar{y}_f(b, k+1), \quad k = 1, \dots, K_f - 1. \quad (3)$$

Figure 1 illustrates this assumption for three jobs. *High-ability workers (ability  $a$ ) have the highest expected output in job 3, the next highest in job 2, and the lowest in job 1, whereas low-ability workers (ability  $b$ ) display the opposite pattern.* By assumption (3), high-ability workers have a *comparative advantage* at higher-index jobs. To interpret this assumption, imagine ranking jobs by their complementarity with ability, which reflects, for instance, the complexity of their tasks, from least (job 1) to most complementary (job  $K_f$ ). Then, higher-indexed jobs are those involving more complex tasks whose complementarity with ability is more pronounced. This pattern of comparative advantage leads successful individuals to advance over time to higher-indexed jobs of a firm's hierarchy, as consistent with the data.

Figure 1: Benchmark assignment rule and wages with equally informative jobs



For intuition, consider an economy with two jobs: job 1, a mailroom job, and job 2, an executive job. Assumption (3) captures the idea that expected output in the mailroom job is relatively insensitive to ability: A low-ability worker produces only slightly less output than a high-ability worker. On the contrary, expected output in the executive job is very sensitive to ability: A low-ability worker produces much less output as an executive compared with a high-ability worker. So, a high-ability worker has a comparative advantage as an executive as well as an absolute advantage in both jobs.

## 2.2. Information

Priors about ability are updated using Bayes' rule: If a worker with prior  $p_t$  is assigned to job  $k$  to firm  $f$ , then the updated prior  $p_{t+1}$  after a success,  $y_{fHk}$ , and a failure,  $y_{fLk}$ , is

$$P_H(p_t, \theta_{fk}) = \frac{\alpha_{fk} p_t}{\alpha_{fk} p_t + \beta_{fk} (1 - p_t)}, \quad P_L(p_t, \theta_{fk}) = \frac{(1 - \alpha_{fk}) p_t}{(1 - \alpha_{fk}) p_t + (1 - \beta_{fk}) (1 - p_t)}. \quad (4)$$

Job  $k$  of firm  $f$  is said to be *more informative* than job  $k'$  of firm  $f'$  if

$$P_H(p_t, \theta_{fk}) \geq P_H(p_t, \theta_{f'k'}), \quad P_L(p_t, \theta_{fk}) \leq P_L(p_t, \theta_{f'k'}), \quad (5)$$

that is, the posterior reached after output is realized at job  $k$  is a mean-preserving spread of the posterior reached after output is realized at job  $k'$ . I model the distribution of output at each job as possibly asymmetric across skill types in that  $\beta_{fk}$  need not equal  $1 - \alpha_{fk}$ . With symmetric and identically distributed signals at all jobs, that is, with  $\beta_{fk} = \beta = 1 - \alpha = 1 - \alpha_{fk}$ , it follows that  $P_H(P_L(p_t, \theta_{fk}), \theta_{fk}) = P_L(P_H(p_t, \theta_{fk}), \theta_{fk}) = p_t$ , so one failure exactly offsets one success. In this case, if a worker is promoted from job  $k$  to  $k + 1$  after a success, then that worker is necessarily demoted to job  $k$  after a failure in job  $k + 1$ . Asymmetric signals, instead, allow for more flexible patterns of promotion and demotion. For example, consider an extreme case in which  $\alpha = 1$  and  $\beta$  is close to 1, so that high-ability workers never fail and low-ability workers rarely fail. In this case, a success is not very informative about ability, so promotions are gradual, occurring only after sufficiently many successes, since no single success perfectly reveals ability. Demotions, instead, are rare, since the probability of failure for both worker types is small. This combination of gradual, frequent promotions and rare demotions is commonly observed in the data.

The discrete Bernoulli structure assumed here also allows for a flexible time profile of the variance of posterior beliefs about a worker's ability and thus job assignment wages.<sup>1</sup> In practice, individual performance evaluations, normally interpreted as proxies for output signals, are recorded based on coarse classification scales. It is easy to show that if performance evaluations are discrete, then assuming discrete signals entails no loss of generality. Hence, modelling output signals as discrete is both conceptually and empirically appealing.

## 3. A Competitive Equilibrium

Each worker alive in period  $t$  faces a constant probability of dying of  $1 - \gamma$  with  $\gamma \in (0, 1)$ . At time  $t$  the probability of being alive at  $t + s$  is thus  $\gamma^s$ . In each period, a measure  $1 - \gamma$  of workers are born with an initial distribution of priors about ability denoted by  $G(p_1)$ . Using the law of large numbers,

<sup>1</sup>In contrast, a worker's output is often modelled as normally distributed in the literature. Coupled with normally distributed priors, this assumptions implies that variance of posterior beliefs about a worker's ability declines deterministically with the number of signals observed, that is, with tenure or with tenure and experience. Hence, the learning component of these models predicts that the longitudinal variance of wages for a single worker should also decline deterministically with tenure or experience, which is not universally observed.



the measure of any given cohort of workers declines deterministically with time at rate  $1 - \gamma$ . Using  $n_{ts}$  to denote the measure in period  $s$  of a cohort born in period  $t$ , it follows that  $n_{ts} = (1 - \gamma) \gamma^{s-t}$ . The measure of the total population of workers in period  $s$  is  $n_s = \sum_{t=-\infty}^s n_{ts}$ , which is constant and equal to 1.

I assume there exist perfect annuity markets, in which firms make annuity payments to workers when annuity payments to workers when they are alive and inherit their wealth when they die, and complete insurance markets against the idiosyncratic risk of success or failure in a job.

At birth, workers differ in their initial prior  $p_1$  that they are of high ability. Hence, workers' consumption paths are indexed by  $p_1$ . If  $\tilde{\delta}$  denotes a worker's discount factor, ignoring the possibility of death, then a worker's effective discount factor is  $\delta = \tilde{\delta}\gamma$ . A worker born at time  $t$  with initial prior  $p_1$  has utility given by  $(1 - \delta) \sum_{s=t}^{\infty} \delta^{s-t} u(c_{ts}(p_1))$ , where  $c_{ts}(p_1)$  is the consumption of a worker of cohort  $t$  in period  $s$  and the period utility function  $u(\cdot)$  is concave.

Consider a stationary equilibrium for this economy. In such an equilibrium, the intertemporal price of goods is constant in calendar time. In particular, the price of goods at time  $s$  in units of goods at time  $t$  is  $1/R^{s-t}$  with  $R = 1/\delta$ . Since there is no aggregate uncertainty in the economy, the combination of annuity markets and insurance markets lets consumers perfectly insure consumption risk. Hence, a worker's budget constraint in a stationary equilibrium is  $\sum_{s=t}^{\infty} \delta^{s-t} c_{ts} = W(p_1)$ , where  $W(p_1)$  is the expected present discounted value of the worker's wages using the market interest rate  $R$  in units of the consumption good at date  $t$ . A worker's problem then conveniently separates into two parts: an *intertemporal consumption* problem and a *job choice* problem. The optimal path of consumption satisfies  $\delta^{s-t} u'(c_{ts}(p_1)) / u'(c_{tt}(p_1)) = 1/R^{s-t}$ , which implies full consumption smoothing in that  $c_{ts}(p_1) = c_{tt}(p_1)$  for all  $s$  and  $t$ . A worker's job choice problem consists of choosing a sequence of jobs so as to maximize the expected present discounted value of wages  $W(p_1)$ , with wages discounted at the market rate  $R = 1/\delta$ .

Consider how wages are determined. When a worker supplies labor to a firm in a given job, a two-sided heterogeneous trade in labor services and information takes place. Heterogeneity in labor service arises from workers differing in their unobserved skill types and is summarized by the prior  $p$ , which may be thought of as a worker's *information capital* at date  $t$ . Thus, a worker with prior  $p'$  supplies a labor input that is different from that of a worker with prior  $p''$ . Heterogeneity in information arises from each job supplying a possibly different increment to this information capital. Specifically, in the economy there are  $\bar{K} = \sum_f K_f$  jobs and thus at most  $\bar{K}$  levels of informativeness. Let  $I$  with  $I \leq \bar{K}$  denote the number of distinct informativeness levels of the  $\bar{K}$  jobs, and let  $\Theta = \{\theta_1, \dots, \theta_I\}$  denote the set of informativeness levels. The commodities exchanged in the labor market, labor input and information, are indexed by  $(p, \theta)$ . Think of there being a market for each such pair: In market  $(p, \theta)$ , the wage in a stationary equilibrium is  $w(p, \theta)$ . If a worker is employed in market  $(p, \theta)$ , then the worker receives the wage  $w(p, \theta)$  today, and the worker's information capital is (stochastically) augmented according to (4) with  $\theta_{fk} = \theta$ . The belief updating rules in (4) can be interpreted as production functions of information that a firm supplies to a worker by hiring him.

Given the wage function  $w(p, \theta)$ , a worker with prior  $p$  must then choose in which of the  $I$  *information*

markets indexed by  $\theta$  to supply one unit of labor services of type  $p$ , by solving

$$V(p) = \max_{\tilde{d}(\theta)} \sum_{\theta \in \Theta} \tilde{d}(\theta) [(1 - \delta) w(p, \theta) + \delta EV(p' | p, \theta)], \quad (6)$$

where  $EV(p' | p, \theta) = r(p, \theta)V(P_H(p, \theta)) + [1 - r(p, \theta)]V(P_L(p, \theta))$  is the worker's expected continuation value, given the prior  $p$  and the informativeness level  $\theta = (\alpha, \beta)$ , and  $r(p, \theta) = \alpha p + \beta(1 - p)$  is the probability of high output at a job with informativeness  $\theta$ . Note that  $\tilde{d}(\theta) = 1$  at some  $\theta$  and  $\tilde{d}(\theta) = 0$  at all others. Then, the optimal decision rule can be written as  $d(p, \theta)$ , where  $d(p, \theta) = 1$  when the worker supplies one unit of labor in the market with informativeness level  $\theta$  and  $d(p, \theta) = 0$  otherwise.

By (6), when choosing a job, a worker with prior  $p$  does not simply select the information market with the highest current wage,  $w(p, \theta)$ . Rather, in choosing between information markets  $\theta$  and  $\theta'$ , a worker with prior  $p$  balances the difference in current wage,  $w(p, \theta) - w(p, \theta')$ , against the value of the corresponding difference in information,  $EV(p' | p, \theta) - EV(p' | p, \theta')$ . Thus, a worker may choose a market in which current wages are lower than in another because in the chosen market, more information can be acquired, so the worker can achieve a higher expected value of future wages.

Consider the problem of the representative firm  $f$  that takes the wage function  $w(p, \theta)$  as given. Since the firm rents labor services each period, there are no state variables for the firm, and the firm's problem simply reduces to a sequence of static problems. In turn, each static problem reduces to a job-by-job problem for workers of each level of information capital  $p$ ,

$$\max_{l_{fk}} [y(p, \theta_{fk}, y_{fHk}, y_{fLk}) - w(p, \theta_{fk})] l_{fk}, \quad (7)$$

where  $w(p, \theta)$  is the wage that firm  $f$  must pay at any job  $k$  with informativeness level  $\theta_{fk} = \theta$  in its production set. The solution  $l_{fk}(p, \theta_{fk})$  denotes the labor demand of firm  $f$  for workers with priors  $p$  at job  $k$ . Notice that since the firm hires a measure of workers and uncertainty about a worker's production is purely idiosyncratic, the law of large numbers implies no uncertainty at the firm level.

In a stationary equilibrium, the initial distribution  $G(p_1)$  of priors over newborn workers leads to a time-invariant distribution over all workers,  $H(p)$ . To define this invariant distribution, note that given any possibly noninvariant distribution  $H_t(p)$ , the decision rules of workers about labor supply imply a transition law from  $H_t(p)$  to  $H_{t+1}(p)$  of the form  $H_{t+1}(\cdot) = \Gamma(H_t(\cdot), d_t)$ , where  $d_t$  records the acceptance decisions of all workers in  $t$ . In a stationary equilibrium, this distribution reproduces itself in that  $H(p) = \Gamma(H(p), d)$  at each  $p$ . The labor market clearing condition, given a distribution of priors  $H(p)$  across all workers, can be written at each  $\tilde{p} \in [0, 1]$  and for each  $\theta$  as

$$\int_0^{\tilde{p}} d(p, \theta) dH(p) = \int_0^{\tilde{p}} \sum_{\{f, k | \theta_{fk} = \theta\}} l_{fk}(p, \theta_{fk}) dp. \quad (8)$$



## References

Pastorino, Elena (2015) "Job Matching within and across Firms," *International Economic Review*, 56 (2), 647–671, [10.1111/iere.12117](#).