DISCUSSION PAPER SERIES

IZA DP No. 3285

Job Market Signaling and Employer Learning

Carlos Alós-Ferrer
Julien Prat

January 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Job Market Signaling and Employer Learning

Carlos Alós-Ferrer
University of Konstanz

Julien Prat
University of Vienna
and IZA

Discussion Paper No. 3285
January 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3285
January 2008

ABSTRACT

Job Market Signaling and Employer Learning*

This paper extends the job market signaling model of Spence (1973) by allowing firms to learn the ability of their employees over time. Contrary to the model without employer learning, we find that the Intuitive Criterion does not always select a unique separating equilibrium. When the Intuitive Criterion bites and information is purely asymmetric, the separating level of education does not depend on the observability of workers' types. On the other hand, when workers are also uncertain about their productivity, the separating level of education is ambiguously related to the speed of employer learning.

JEL Classification: I20, C70, D82, D83

Keywords: signaling, job markets, education, employer learning, intuitive criterion

Corresponding author:

Julien Prat
Department of Economics
University of Vienna
Hohenstaufengasse 9
A-1010 Wien
Austria
E-mail: julien.prat@univie.ac.at

* We are very grateful to Egbert Dierker, Gerhard Orosel, Larry Samuelson, and Gerhard Sorger, as well as conference participants at the ESEM and SED 2007 meetings for helpful comments.

1 Introduction

The signaling model of Spence (1973) has provided a seminal impulse for ground-breaking research in game theory and econometrics. On the one hand, the multiplicity of equilibria in Spence's model has been a motivation for the vast literature on refinement concepts. On the other hand, econometricians have devised ingenious tests about unobservable characteristics in order to assess the importance of signaling for schooling. In spite of their common roots, these two strands of literature have followed diverging paths. The theoretical literature focuses on the signaling stage and generally ignores learning in the labor market.[1] Econometric research places more emphasis on the learning process but typically relies on *Informational equilibria*, rather than *Perfect Bayesian (Nash) equilibria*, to analyze the signaling game.[2] This paper proposes to reconcile these two views by embedding Spence's model into a dynamic framework with Bayesian learning on the side of firms.

As in Spence (1973), workers of different abilities can acquire education before entering the labor market. While workers know their productivity with certainty, firms ignore the actual ability of job applicants. This is why high-types have an incentive to acquire education in order to avoid being confused with low-types. They may prefer, however, to save on education costs by choosing to reveal their ability while being on-the-job. This alternative strategy is particularly attractive if there is a strong correlation between a worker's ability and his observable performance.

Spence's model does not take into account this countervailing incentive because it is based on the premise that all information is collected prior to the entry into the labor market. We assume instead that employers make use of observed employee performance to update their beliefs. Hence, workers' types are *gradually* revealed over time. Our model therefore proposes a synthesis of the signaling and job matching literatures. We find that the two dimensions interact with each other: the equilibria of the signaling game are affected by the learning process and the labor market outcomes are influenced by the signaling stage.

In order to make the model suitable for aggregation, we assume that workers' abilities are either high or low. Although this simplification is not crucial and can easily be relaxed in the original model,[3] it plays a key role in our enlarged set-up as it allows us to characterize the cross-sectional distribution of beliefs in closed-form. Our analysis bears similarities to Moscarini's (2005) matching model since we assume that workers' outputs are drawn from a Gaussian distribution and that firms learn in continuous time. The problem is complicated by the fact that uncertainty is not match-specific, as in Jovanovic (1979) and Moscarini (2005), but instead worker-specific. One contribution of the paper

[1] The paper by Gibbons an Katz (1991) is a notable exception.

[2] Informational equilibria are not Nash equilibria. See Riley (1979a) for details.

[3] Allowing for more than two types though, leads to well-known difficulties in the application of the Intuitive Criterion.

is to show that the matching model with asymmetric information remains tractable. An appropriate change of variable enables us to analytically characterize pooling equilibria and to derive new findings about the robustness and properties of separating equilibria.

We show that the Intuitive Criterion proposed by Cho and Kreps (1987) does not necessarily rule out all but one separating equilibrium. This stands in contrast to the model without employer learning where all pooling equilibria fail the Intuitive Criterion. The key difference between the two models is that, even when pooling is the equilibrium outcome, learning yields higher asset values for talented workers. The gap increases with the speed of learning as low and high-types become respectively less and more optimistic about future prospects. Hence, while the incentive for low-types to send misleading signals increases with the speed of employer learning, the incentive for high-types to signal their ability decreases.[4] When learning is efficient, high-types prefer working until their ability is recognized to paying the education costs. In other words, they find it optimal to pool with low-types.

We further analyze the model's implications for the relationship between the speed of employer learning and education. We find that the separating education level does not depend on the efficiency of the learning process. Given that education levels fully reveal types, signaling makes learning redundant. We also consider an extended model where workers are uncertain about their productivity. In this model, learning occurs also in separating equilibria. This extension illustrates that symmetric uncertainty is the crucial element linking education with the efficiency of the learning process.

As explained by Riley (1979b), if low-types are more easily detected, smaller levels of education should be sufficient to discourage them from trying to masquerade as high-types.[5] Several papers have tested this implication by assessing whether average years of schooling are lower in occupations where employers can better infer the ability of their employees. The current consensus is that inter-industry data do not substantiate such a correlation, a finding that is widely interpreted as evidence against the importance of job market signaling.[6]

We suggest instead a more ambivalent interpretation. In line with Riley's prediction, our model generates a positive correlation between the separating level of education and the noise hindering the learning process. Our results, however, also imply that higher returns to ability facilitate signal extraction. Since they raise the incentive to acquire education, signaling can be consistent with the data if learning is faster in some occupations, not because productivity is easier to observe, but because returns to talent are higher. This leads us to the conclusion that one would need precise information

[4]A recent working paper by Haberlmaz (2006) proposes a partial equilibrium model which also underlies the ambiguity of the relationship between the value of job market signaling and the speed of employer learning.

[5]See Kaymak (2006) for a signaling model with Bayesian learning that leads to a similar implication.

[6]See Lange and Topel (2006) for an overview of the empirical literature and Section 4 in this paper for more references.

about the two components of the signal/noise ratio in order to test conclusively whether or not job market signaling is empirically relevant.

The paper proceeds as follows. Section 2 characterizes the lifetime incomes of workers and the distribution of wages in pooling equilibria. It describes the signal extraction problem and obtains closed-form solutions for the workers' asset values. In Section 3, we analyze the conditions under which the Intuitive Criterion bites. Section 4 lays-out the extended model where workers are also uncertain about their productivity. It details the model's implications for the relationship between educational attainment and the speed of employer learning. Section 5 concludes. The proofs of the Propositions and Lemmata are relegated to the Appendix.

2 Pooling Equilibria

2.1 Signal extraction

The main ingredient of our model is information asymmetry. Workers differ in their innate *abilities*. They can be of different *types* which determine their productivity. While a worker knows his ability with certainty, his employer has to infer it. In this section, we concentrate on pooling equilibria, i.e. equilibria where all workers choose the same education level so that signalling carries no information. In Spence's (1973) static framework, the employer is left with no information at all because the game ends as the worker enters the labor market.

Contrary to Spence (1973), and in line with the recent literature on labor markets, we consider a dynamic framework where the output realizations for a given type of worker are not deterministic. Rather, they are random draws from a Gaussian distribution centered on the worker's average productivity. The variance of the shocks σ is common knowledge and independent of the worker's type. Thus the employer can update his belief on the worker's type from the observation of realized outputs. Obviously, the employer's ability to identify the mean of the output distribution is hindered by the variance of the realizations. Accordingly, we will hereafter refer to the inverse variance, $1/\sigma$, as signal precision.

We restrict our attention to the case where there are only two types of workers, $i = h$ (high) or $i = l$ (low). For simplicity, we also assume that the expected productivity of a given worker remains constant through time[7] and is equal to μ_h for the high-type and to $\mu_l < \mu_h$ for the low-type. The *cumulative* output X_t of a match of duration t with a worker of type $i = h, l$, follows a Brownian

[7] It would be reasonable to assume that workers accumulate general human capital. However, this would substantially complicate the aggregation procedure without adding substantial insights.

motion with drift

$$dX_t = \mu_i dt + \sigma dZ_t \ ,$$

where dZ_t is the increment of a standard Brownian motion. The cumulative output $\langle X_t \rangle$ is observed by both parties. The employer uses the filtration $\left\{\mathcal{F}_t^X\right\}$ generated by the output sample path to revise his belief about the worker's average productivity. Starting from a prior p_0 equal to the fraction of high ability workers in the population (identified with the *ex ante* probability that a randomly sampled worker is of the high-type), the employer applies Bayes' rule to update his belief $p_t \equiv \Pr\left(\mu = \mu_h | \, \mathcal{F}_t^X\right)$. His posterior is therefore given by

$$p\left(X_t, t | \, p_0\right) = \frac{p_0 g_h(X_t, t)}{p_0 g_h(X_t, t) + (1-p_0) g_l(X_t, t)} \ . \tag{1}$$

where $g_i(X_t, t) \equiv e^{-\frac{(X_t - \mu_i t)^2}{2\sigma^2 t}}$ is the rescaled[8] density for the worker of type i.

The analysis is simplified by the change of variable $P_t \equiv p_t/(1-p_t)$. P_t is the ratio of "good" to "bad" belief. Since p_t is defined over $]0,1[$, P_t takes values over the positive real line. It immediately follows from (1) that

$$P\left(X_t, t | \, P_0\right) = P_0 \frac{g_h(X_t, t)}{g_l(X_t, t)} = P_0 e^{\frac{s}{\sigma}\left(X_t - \frac{1}{2}(\mu_h + \mu_l) t\right)} \ . \tag{2}$$

where $s \equiv (\mu_h - \mu_l)/\sigma$ is the signal/noise ratio of output. The higher s, the more efficient is the learning process. By Ito's lemma, the law of motion of the posterior belief ratio is

$$\begin{aligned} dP\left(X_t, t | \, P_0\right) &= \frac{\partial P\left(X_t, t | \, P_0\right)}{\partial X_t} dX_t + \frac{\partial^2 P\left(X_t, t | \, P_0\right)}{\partial X_t^2} \frac{\sigma^2}{2} dt + \frac{\partial P\left(X_t, t | \, P_0\right)}{\partial t} dt \\ &= P\left(X_t, t | \, P_0\right) \left(\frac{s}{\sigma}\right) \left[dX_t - \mu_l dt\right] \ . \end{aligned} \tag{3}$$

Replacing in (3) the workers' beliefs on the law of motion of X_t, i.e. $dX_t = \mu_i dt + \sigma dZ_t$, yields the following stochastic differential equations:

(i) High ability worker: $dP_t = P_t s \left(s dt + dZ_t\right)$,

(ii) Low ability worker: $dP_t = P_t s dZ_t$.

Since the signal/noise ratio s is a positive constant, the posterior belief P_t increases in time for high-types and follows a martingale for low-types.[9] In both cases, an increase in σ lowers the variance of beliefs since larger idiosyncratic shocks hamper signal extraction.

[8] The factor $[\sigma\sqrt{2\pi t}]^{-1}$ is omitted because it simplifies in (1).

[9] It may be surprising that the belief ratio P_t does not drift downward when the worker is of the low-type. The technical reason is that the belief ratio P_t is a convex function of the posterior belief p_t. Reversing the change of variable shows that the belief p_t is a strict supermartingale when the worker's ability is low.

2.2 Asset values

After a worker enters the labor market, his cumulative output is publicly observed in continuous time. Employers compete *à la* Bertrand so that wages $w(p)$ are equal to the expected worker's productivity: $p\mu_h + (1-p)\mu_l$. Workers are risk neutral and discount the future at rate r. We further assume that workers leave the labor market at a constant rate δ.[10]

Although high and low ability workers earn the same wage for a given cumulative output, their asset values differ because high ability workers are more optimistic about future prospects. Moreover, workers' expectations also differ from those of their employers. Accordingly, three different asset values are associated to the same cumulative output. Using the laws of motion of beliefs given above, we derive the two Hamilton-Jacobi-Bellman (HJB henceforth) equations satisfied by the workers' asset values $W_i(P)$:

$$\begin{aligned} rW_h(P) &= w(P) + Ps^2 W_h'(P) + \frac{1}{2}(Ps)^2 W_h''(P) - \delta W_h(P) \ , \\ rW_l(P) &= w(P) + \frac{1}{2}(Ps)^2 W_l''(P) - \delta W_l(P) \ . \end{aligned}$$

These two ordinary differential equations have (somewhat cumbersome) closed-form solutions which are detailed in the following Proposition. Plots of these value functions for a particular numerical example and several values of σ are presented in Figure 1. It illustrates two things. First, for any given belief $p \in]0,1[$, high-types have higher expected lifetime incomes than low-types. Second, the gap increases as learning becomes more efficient.

Proposition 1. *The expected lifetime income of workers as a function of the belief ratio P are given by*

$$W_l(P) = \frac{2\sigma}{s\Delta}\left(P^{\alpha^-}\int_0^P \frac{1}{(1+x)\,x^{\alpha^-}}dx + P^{\alpha^+}\int_P^\infty \frac{1}{(1+x)\,x^{\alpha^+}}dx\right) + \frac{\mu_l}{r+\delta}$$

and

$$W_h(P) = \frac{2\sigma}{s\Delta}\left(P^{\gamma^-}\int_0^P \frac{1}{(1+x)\,x^{\gamma^-}}dx + P^{\gamma^+}\int_P^\infty \frac{1}{(1+x)\,x^{\gamma^+}}dx\right) + \frac{\mu_l}{r+\delta}$$

where $\alpha^+ = \frac{1}{2}(1+\Delta)$, $\alpha^- = \frac{1}{2}(1-\Delta)$, $\gamma^+ = \frac{1}{2}(-1+\Delta)$, $\gamma^- = \frac{1}{2}(-1-\Delta)$, *and* $\Delta = \sqrt{1+8\left(\frac{r+\delta}{s^2}\right)}$.

[10] Allowing δ to be an increasing function of labor market experience would improve the realism the model. Such a specification would greatly complicate the derivations by making the asset values non-stationary. We leave this task for future research.

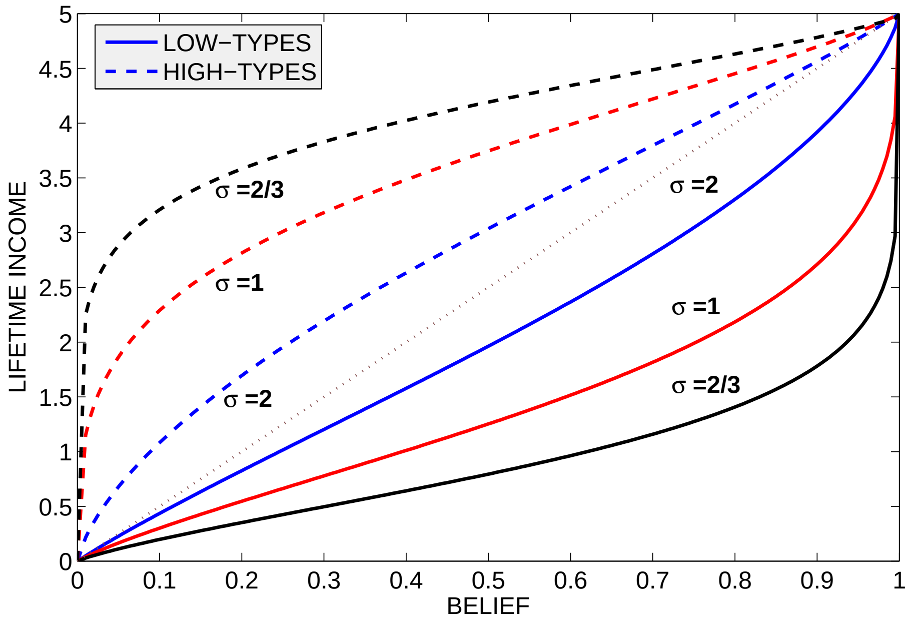

Figure 1: Workers value functions. Parameters: r $= \delta = 0.1$, $\mu_l = 0$, $\mu_h = 1$.

2.3 Wage distribution

Wages are a linear function of beliefs since $w(p) = p(\mu_h - \mu_l) + \mu_l$. Thus the wage distribution follows from the distribution of beliefs by a location transformation. In order to derive the ergodic distribution, we use the Kolmogorov forward equation to characterize the evolution of the cross-sectional density of beliefs $f(p)$. The expected law of motion for employers' beliefs is obtained replacing $dX_t = (p_t\mu_h + (1-p_t)\mu_l)\,dt + \sigma d\overline{Z}_t$, where $\overline{Z}_t$ is a standard Brownian motion with respect to the filtration $\left\{\mathcal{F}_t^X\right\}$, into (3). Using Ito's lemma to reverse the change of variable from P_t to p_t yields $dp_t = p_t\,(1-p_t)\,sd\overline{Z}_t$, so that the Kolmogorov forward equation reads

$$\frac{df\,(p)}{dt} = \frac{d^2}{dp^2}\left(\frac{1}{2}p^2\,(1-p)^2\,s^2 f\,(p)\right) - \delta f(p)\,. \tag{4}$$

The first term on the right-hand side deducts, for any given belief, the outflows from the inflows. The second term takes into account the fact that workers leave the labor market at rate δ. The ergodic density is derived imposing the stationarity condition: $df\,(p)\,/dt = 0$. As shown in the Appendix, the general solution involves solving for two constants of integration. They are pinned down by the normalization, $\int_0^1 f\,(p)\,dp = 1$, and continuity, $f\left(p_0^-\right) = f\left(p_0^+\right)$, requirements.

Proposition 2. *The ergodic distribution of beliefs is given by*

$$f(p) = \begin{cases} K_{0f} p^{\eta-2}(1-p)^{-1-\eta} & \text{if} \quad p \in (0, p_0) \\ \\ C_{1f} p^{\eta-2}(1-p)^{-1-\eta} & \text{if} \quad p \in [p_0, 1) \end{cases},$$

where $\eta = \frac{1}{2}\left(1 + \sqrt{1 + 8\left(\frac{\delta}{s^2}\right)}\right)$, $\xi = \frac{1}{2}\frac{s^2}{\delta}\sqrt{1 + 8\left(\frac{\delta}{s^2}\right)}$, $K_{0,f} = \frac{1}{\xi}\frac{1}{P_0^{\eta-1}(1+P_0)}$, *and* $C_{1,f} = \frac{1}{\xi}\frac{P_0^{\eta}}{(1+P_0)}$.

The ergodic distribution is piecewise and composed of two Beta functions. Depending on the values of the parameters, the wage distribution can be either U-shaped or hump-shaped. Empirically, the the second case seems to be more realistic. It occurs for sufficiently high values of the ratio δ/s^2. Then the right-tail of the distribution decreases at a slower rate than the Gaussian distribution from which underlying shocks are sampled. Hence, the model can replicate the heavy tail property of empirical wage distributions. As noticed by Moscarini (2005), another specific implication of the learning model is that the dispersion of beliefs is decreasing in the variance of shocks. This is of course different from models without signal extraction where the uncertainty of the economic environment naturally generates more inequality. To the contrary, when firms have to filter the noise from the observations, a higher degree of uncertainty reduces the rate at which information is acquired. This is why the inertia in belief revisions is stronger when there is more noise, which in turn implies that wage dispersion is decreasing in the degree of uncertainty.

3 Employer Learning and the Intuitive Criterion

We now consider the full model where workers signal their types by acquiring education before entering the labor market. Given that workers are also able to reveal their ability after the signaling stage, our framework extends the job market-signaling model of Spence (1973).[11] As in the previous section, nature assigns a productivity $\mu \in \{\mu_l, \mu_h\}$ to the worker (sender), with $\mu_h > \mu_l$. The worker then chooses an education level $e \in [0, +\infty)$. To isolate the effect of signaling, we assume that education does not increase the labor productivity of the worker. Thus its only use is to signal the worker's ability which is initially unobserved by the industry (receiver).

It is well known that one can find a plethora of Perfect Bayesian Equilibria (hereafter PBE) in Spence's (1973) model. As shown by Cho and Kreps (1987), the refinement concept known as the *Intuitive Criterion* rules out all but one separating equilibrium, thus conferring a predictive power to the model. The purpose of this section is to show that this does not hold true when signal extraction

[11] Spence's (1973) model corresponds to the limit case where the signal/noise ratio s is equal to 0.

also takes place on-the-job. More precisely, we prove that all pooling PBE also satisfy the Intuitive Criterion when the signal precision $1/\sigma$ exceeds a given threshold.

Let the cost function $c : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ specify the cost of acquiring education. That is, $c(e, \mu)$ is the cost that a worker with innate productivity μ has to pay in order to acquire education level e.[12] The cost function is assumed to satisfy $c_e(e, \mu) > 0$ and $c_{ee}(e, \mu) > 0$, thus it is strictly increasing and strictly convex in the level of education. This ensures the existence of an interior optimum. As commonly assumed in the literature, we also consider that total and marginal costs of education are strictly decreasing in the worker's ability, so that $c_\mu(e, \mu) < 0$ and $c_{e\mu}(e, \mu) < 0$. The last requirement ensures that low-types have steeper indifference curves than high-types.

The industry sets the wage $w \in [0, +\infty)$. In the classic signaling model, the payoffs of the worker are given by $w - c(e, \mu)$. As a proxy for either a competitive labor market or a finite number of firms engaged in Bertrand competition for the services of the worker,[13] the industry always offers a wage equal to the expected productivity. In a signalling PBE:

(a) The worker selects an education level which maximizes expected utility given the industry's offer;

(b) The industry offers a starting wage equal to the expected productivity of the worker given the industry's beliefs;

(c) The industry's beliefs are derived from Bayes' rule for any educational attainment e that is selected with a positive probability.

We depart from Spence's model as follows. Once the worker has accepted a wage and been matched to a firm, i.e. after the signalling game has been played out, production starts. The firm initial belief about the type of the worker is the equilibrium belief. Afterwards, the employer revises his prior as in Section 2. Hence, the worker payoff is not simply $w - c(e, \mu)$ but rather $W_i(p_0) - c(e, \mu)$, where $W_i(p_0)$ is the expected lifetime income of the worker ($i = l, h$) as a function of the firm prior p_0.

In separating equilibria, the ability of the worker is perfectly revealed by his education: depending on the education signal, the initial belief of the firm is either zero or one. As can be seen from the definition of p_t in equation (1), firms do not update their beliefs when $p_0 \in \{0, 1\}$. This implies that the asset values of high and low ability workers in separating equilibria are $W_h^{Sep} = \mu_h/(r + \delta)$ and $W_l^{Sep} = \mu_l/(r + \delta)$, respectively. Comparing these asset values with the ones derived in Section 2,

[12] Even though workers will only have productivity levels in $\{\mu_l, \mu_h\}$, it is convenient to consider the cost function to be defined on all potential productivities.

[13] If we explicitly model Bertrand competition among multiple firms, the equilibrium concept has to be refined to include the additional condition that, given an education level, all firms have the same beliefs about the worker.

it is possible to verify that, for some parameter configurations, all pooling equilibria do not fail the Intuitive Criterion. We now show this formally.

As shown by Lemma 1, sequential rationality is satisfied as long as the "participation constraint" for the low-type is fulfilled. This substantiates the claim made at the beginning of this section that the model admits a continuum of PBE.

Lemma 1. *A pooling PBE where both types educate at level e_p exists if and only if $e_p \in [0, \bar{e}]$, where $\bar{e}$ is given by*

$$W_l(P_0) - c(\bar{e}, \mu_l) = W_l^{Sep} - c(0, \mu_l) \ .$$

In order to refine the set of potential equilibria, we impose further restrictions on out-of-equilibrium beliefs. The Intuitive Criterion requires that, after receiving an out-of-equilibrium signal, firms place zero probability on the event that the sender is of type i whenever the signal is equilibrium-dominated for type i. In our set-up, a pooling equilibrium with education level e_p fails the Intuitive Criterion if and only if there exists an education level e^{Sep} such that:

$$W_h^{Sep} - c(e^{Sep}, \mu_l) < W_l(P_0) - c(e_p, \mu_l), \text{ equilibrium dominance for the low-type,} \tag{ED}$$

$$W_h^{Sep} - c(e^{Sep}, \mu_h) > W_h(P_0) - c(e_p, \mu_h), \text{ high-type's participation constraint.} \tag{PC}$$

The argument is as follows. The equilibrium dominance condition (ED) implies that, even in the best-case scenario where the worker could forever deceive his employer, deviating to e^{Sep} is not attractive to the low-type. The firm can therefore infer by forward induction that any worker with an off the equilibrium signal e^{Sep} is of the high-type. But the participation constraint (PC) implies in turn that credibly deviating to e^{Sep} is profitable for the high-type. Thus e_p is not stable[14] or, in other words, fails the Intuitive Criterion.

Combining conditions (ED) and (PC), it is straightforward to show analytically that all pooling equilibria fail the Intuitive Criterion in the model without learning. Let $e^*(e_p)$ denote the minimum education level that does not trigger a profitable deviation for low ability workers, so that (ED) holds with equality at $e^*(e_p)$. Analogously, let $e^{**}(e_p)$ be such that (PC) holds with equality.[15] These two thresholds always exist because $c(e, \mu)$ is continuous and strictly increasing in e. Conditions (ED) and

[14]We follow the terminology of Cho and Kreps (1987) by referring to equilibria that satisfy the Intuitive Criterion as stable equilibria.

[15]Note that, since $W_l(P_0) < W_h^{Sep}$ and $c_e(e, \mu) > 0$, condition (ED) implies that $e^*(e_p) > e_p$. Similarly, condition (PC) implies that $e^{**}(e_p) > e_p$.

(PC) above reduce to $e^*(e_p) < e^{Sep} < e^{**}(e_p)$. Now assume that e_p is stable, so that condition (PC) is not satisfied at $e^*(e_p)$. This can be true if and only if

$$W_h(P_0) - W_l(P_0) \geq c(e^*(e_p), \mu_l) - c(e_p, \mu_l) - [c(e^*(e_p), \mu_h)) - c(e_p, \mu_h)] \ . \tag{5}$$

Abusing notation, we refer to the case without learning as $\sigma = \infty$. When there is no learning, the lifetime incomes of workers in a pooling equilibrium do not depend on their abilities, i.e. $W_h(P_0|\sigma = \infty) = W_l(P_0|\sigma = \infty)$. Thus the left hand side of equation (5) becomes zero, while the right-hand side is strictly positive since $c_{\mu e}(e, \mu) < 0$. This contradiction shows that, in the basic signaling model, one can always find a credible and profitable deviation for the high-type.

When workers' abilities are also revealed on-the-job, the premise leading to a contradiction is not anymore true. With learning, high-ability workers have an higher asset values even in a pooling equilibrium.[16] Hence $W_h(P_0|\sigma) > W_l(P_0|\sigma)$ for all $\sigma < \infty$ and equation (5) can hold true for some parameter configurations. Obviously, the lower σ is, the bigger the difference between the workers' asset values. This suggests that a pooling equilibrium is more likely to be stable when signal extraction is efficient. Proposition 3 substantiates this intuition.

Proposition 3. *For any arbitrary pooling level of education e_p, there exists a threshold variance $\sigma^*(e_p)$ such that e_p fulfills the Intuitive Criterion for any $\sigma \leq \sigma^*(e_p)$.*

Figure 2 illustrates the mechanism behind Proposition 3. It displays the high and low-types' indifference curves in the limit case $\sigma \to \infty$ (abusing notation, we write simply $\sigma = \infty$) and when $\sigma = \sigma^*(e_p)$. The dotted curves correspond to the latter case, the undotted ones to the former case, that is the Spence model without learning. The level of education $e^*(e_p|\sigma)$ where condition (ED) holds with equality is given by the point where the low-type indifference curve crosses the horizontal line with intercept W_h^{Sep}. Similarly, the level of education $e^{**}(e_p|\sigma)$ where condition (PC) holds with equality is given by the point where the high-type indifference curve crosses the same horizontal line. As discussed above, the pooling equilibrium e_p fails the Intuitive Criterion if and only if $e^*(e_p) < e^{**}(e_p)$. Thus we can conclude that e_p is not stable when the low-type's indifference curve intersects the horizontal line with intercept W_h^{Sep} before the high-type's indifference curve.

First of all, consider the basic model without learning. At the pooling level of education e_p, the two types enjoy the same asset value $W(P_0|\sigma = \infty) \equiv W_h(P_0|\sigma = \infty) = W_l(P_0|\sigma = \infty)$. The *single-crossing property* therefore implies that $e^*(e_p|\sigma = \infty)$ lies to the left of $e^{**}(e_p|\sigma = \infty)$, as shown by the undotted indifference curves in Figure 2. This illustrates that any pooling equilibrium fails the

[16] See Figure 1 for an illustration of this property.

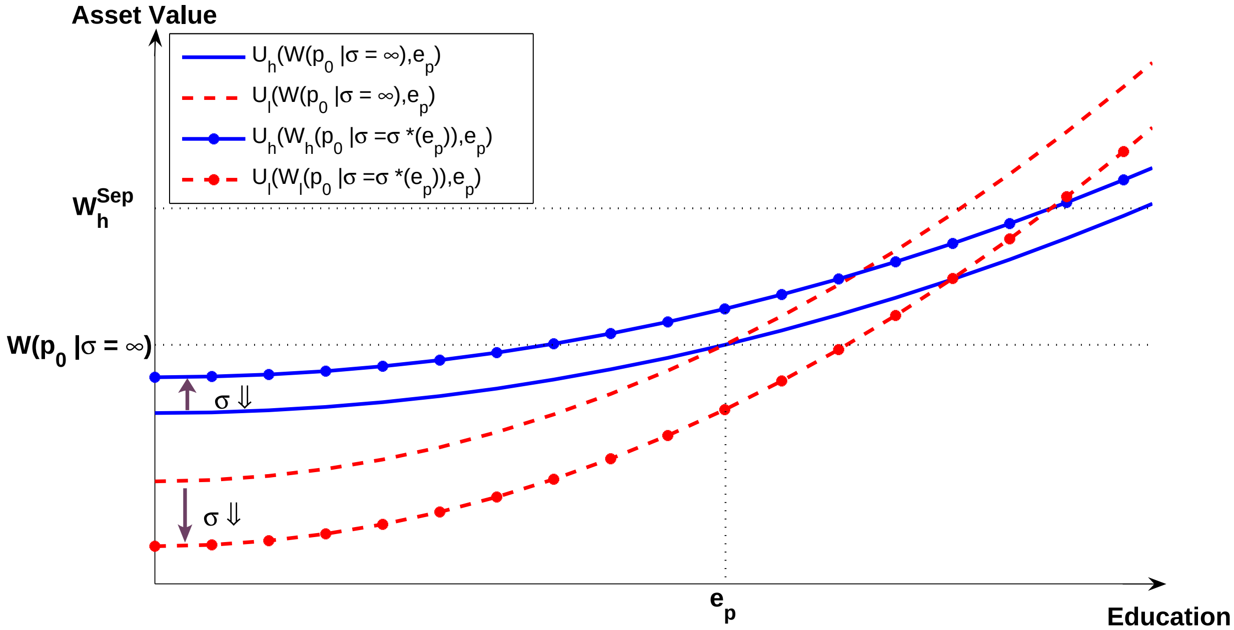

Figure 2: Workers' indifference curves.

Intuitive Criterion when the signal/noise ratio $s = 0$. Now, consider what happens when σ decreases. Since high-types are more quickly recognized, their asset value increases and so their indifference curve shifts up. Conversely, the indifference curve of low-types shifts down when σ decreases because they are more rapidly detected. These opposite adjustments obviously shrink the gap between $e^*(e_p|\sigma)$ and $e^{**}(e_p|\sigma)$. The threshold variance $\sigma^*(e_p)$ is identified by the point where this gap vanishes as the two indifference curves concurrently cross the horizontal line with intercept W_h^{Sep}. One can always find such a point for any given e_p since $\lim_{\sigma\to 0} W_h(P_0|\sigma) = W_h^{Sep}$. It is also clear from Figure 2 that $\sigma^*(e_p)$ is unique.

By checking that (ED), (PC) and the low-type's "participation constraint" given in Lemma 1 are simultaneously satisfied, one can establish whether or not e_p is a stable pooling PBE. This approach solely yields local results. We now show that an additional assumption on the cost function allows for a global characterization of the region where the Intuitive Criterion fails to select a unique separating equilibrium.

Proposition 4. *When the marginal cost of education $c_e(e, \mu)$ is weakly log-submodular, there exists a unique level of noise, which we denote σ_p, such that*

(a) For all $\sigma \in [\sigma_p, \infty)$, the Riley separating equilibrium is the only PBE satisfying the Intuitive

Criterion.[17]

(b) For all $\sigma < \sigma_p$, $\exists \varepsilon(\sigma) > 0$ such that the pooling equilibrium meets the Intuitive Criterion for any education values $e \in [0, \varepsilon(\sigma))$. As σ goes to zero, $\varepsilon(\sigma)$ converges to zero.

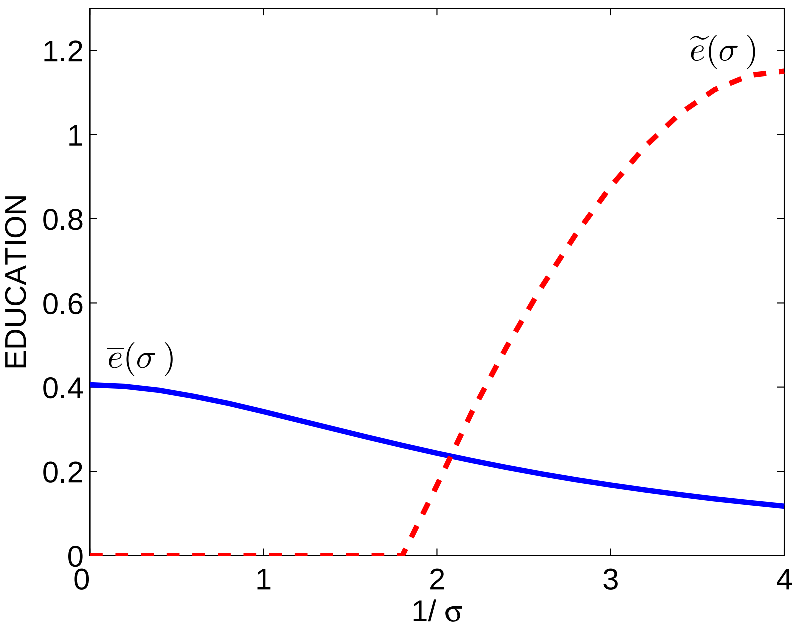

Figure 3: Graphical interpretation of Proposition 4. A pooling equilibrium with education level e exists if and only if $e \leq \overline{e}(\sigma)$. It satisfies the Intuitive Criterion if and only if $e \leq \widetilde{e}(\sigma)$. Parameters: r $= \delta = 0.1$, $\mu_l = 0.5$, $\mu_h = 1$, $p_0 = 0.5$ and $c(e, \mu) = \exp(e/\mu) - 1$.

The economics behind Proposition 4 is quite intuitive. When the signal/noise ratio is high, firms easily infer the actual type of their employees. Then the benefits derived from ex-ante signaling are not important. Symmetrically, when the signal/noise ratio is low, firms learn little from observing output realizations. This leaves fewer opportunities for high-types to reveal their ability while being on-the-job and so raises their incentives to send a signal. In the extreme case where signal precision goes to zero, all the relevant information is collected prior to the entry into the labor market. This situation corresponds to the signaling model of Spence (1973). Thus, it should not be surprising that we recover, as a limit case, the well known result according to which all PBE but one separating equilibrium fail the Intuitive Criterion when there is no learning.

[17] For separating equilibria, our model coincides with Spence (1973). Thus the Riley equilibrium is the only separating equilibrium fulfilling the Intuitive Criterion.

In order to derive Proposition 4, we have imposed slightly stronger restrictions on the cost of education: we assume that the log of the derivative of $c(e, \mu)$ with respect to e has weakly increasing differences in (e, μ) or, in other words, that $c_e(e, \mu)$ is weakly log-submodular. Although the weak log-submodularity of $c_e(e, \mu)$ might seem quite restrictive, it is actually satisfied by the distributions commonly used to illustrate the single crossing property, such as power and exponential functions. The requirement is sufficient but not necessary to ensure that the single crossing property is satisfied. It is more stringent since it implies that, not only total educational costs, but also marginal educational costs diverge. If that property is not satisfied, there are cases where an increase in the level of education restores the stability of a pooling equilibrium.[18] Hence, when $c_e(\mu, e)$ is not weakly log-submodular, equilibrium stability does not generally divide the (σ, e) space into two non-overlapping regions, as done in Figure 3 by the $\widetilde{e}(\sigma)$ locus.[19]

We finish this Section with a remark on equilibrium refinements. The Intuitive Criterion is of course not the only available refinement for signaling games. However, it is well known that more sophisticated criteria generally have a bite only for more than two types. Banks and Sobel (1987) introduce the concept of *Universal Divinity*, which is a strengthening of the Intuitive Criterion based on the elimination of pairs of types and signals (education levels) such that the set of receivers' answers (wages) making the type weakly prefer the signal to the equilibrium one is contained in the set of receivers' answers making *another* type strictly prefer the signal to the equilibrium one.[20] In our framework, the identification of such signals for the low-type boils down to condition (5). Thus, the pooling equilibria identified in Proposition 3 also satisfy Universal Divinity. Proposition 4 would also be unaffected by the application of Universal Divinity, because the Riley equilibrium always satisfies this latter refinement.

4 Educational Attainment and the Speed of Learning

While the previous section focused on pooling equilibria, we now turn our attention to separating equilibria and how they vary with the speed of employer learning. This connects our analysis to the long lasting debate about the empirical content of job market signaling. One drawback of signaling theory is that it shares most of its testable predictions with human capital theory. This makes it

[18]See Claim 2 in the Proof of Proposition 4.

[19]Notice that the $\widetilde{e}(\sigma)$ locus is vertical in the knife-edge case where the cost function is log-linear, e.g. $c(e, \mu) = e^2/\mu$. See the proof of Proposition 4 for a discussion of this point.

[20]Cho and Kreps (1987) refer to this as Condition D1. Actually, with more than two types, Universal Divinity is based on their Condition D2, where the pair is eliminated if the referred set is included in a union of sets across several other types. This makes no difference in our case.

particularly difficult to distinguish the two approaches using data from a single labor market.[21]

To circumvent this problem, Wolpin (1977) and Riley (1979b) have proposed to exploit differences in the speed of learning across occupations and industries. Riley notices that the expected payoff from masquerading as a high-type is larger when learning is difficult. This implies that workers employed in occupations with faster learning should have the following characteristics: lower average education and higher ability for a given level of education. In other words, high wage industries and occupations should employ a labor force with relatively less education. Unfortunately that prediction is not supported by the data. Murphy and Topel (1990) and more recently Lange and Topel (2006) document that workers with higher schooling tend to work in industries and occupations that reward more unobserved characteristics. They interpret this finding as evidence against the importance of signaling in labor markets.

In our model, however, the separating level of education does not depend on the efficiency of the learning process. Low-types choose the minimum education requirement (normalized to zero in our set-up), whereas high-types select the lowest education level which allows them to credibly signal their ability. Thus the separating investment e^{Sep} is such that

$$c\left(e^{Sep}, \mu_l\right) - c(0, \mu_l) = W_h^{Sep} - W_l^{Sep}.$$

The right-hand side is equal to $W_h^{Sep} - W_l^{Sep} = (\mu_h - \mu_l)/(r + \delta)$, which does not vary with σ. Since the cost function $c(e, \mu)$ is also independent of σ, Spence's signaling model does not generate any correlation between schooling and signal precision. This is because, when uncertainty is purely asymmetric, workers' types are perfectly revealed in separating equilibria. Firms put full weight on the belief that the ability of educated workers is high and do not revise their prior.

To ensure that signaling does not make learning redundant, we extend the model by allowing for uncertainty on the workers' side. We continue to assume that average productivity can take only one of two values, namely μ_h or μ_l. Let π_h and π_l denote the objective probabilities that a worker's productivity equals μ_h when his educational ability is respectively high or low.[22] As in the standard model, we assume that educational and labor market abilities are correlated, so that $\pi_h > \pi_l$. The case

[21] There exists, however, a strand of literature that directly measures the speed of employer learning. The identifying assumption is that econometricians have access to a correlate of workers' abilities that is not available to employers (generally Armed Forces Qualification Test scores of workers). Farber and Gibbons (1996) and Altonji and Pierret (2001) document that the impact of this correlate on wages increases with labor market experience. Although this result indicates that employers learn over time, it does not distinguish between symmetric and asymmetric learning. Hence, it does not test job market signaling. Nevertheless, as explained by Lange (2007), the estimated speed of employer learning can be used to place an upper-bound on the contribution of signaling to the gains of schooling.

[22] Hence, the distinction between low and high types now refers to their schooling abilities.

where $\pi_h = 1$ and $\pi_l = 0$ obviously corresponds to the model analyzed in the previous sections. For other parameter values, there is a positive probability that a talented student has a low productivity in the labor market. This is why firms revise their beliefs even in separating equilibria. Notice that, since workers ignore the move of nature, they also try to infer their actual productivity. The filtering problems faced by firms and their employees are similar except that they do not necessarily share the same prior: in pooling equilibria, workers have private information and thus more accurate priors than their employers. It follows from equation (2) that

$$P\left(X_t, t \middle| P_0\right) = \left(\frac{P_0}{P_0^i}\right) P^i\left(X_t, t \middle| P_0^i\right) \text{ , for } i = h, l \text{ ,} \tag{6}$$

where $P^i(\cdot)$ is the posterior and P_0^i is the prior of type i. Equation (6) shows that the initial difference in belief between an employer and his employee is preserved for any employment history. Reinserting the worker's belief about the law of motion of X_t into (3) and using (6) yields

$$dP_t = P_t s \left[s \left(\frac{P_t}{P_0/P_0^i + P_t} \right) dt + dZ_t \right] .$$

Using this law of motion and following the same steps than in the proof of Proposition 1, one can derive the following closed-form solution.

Proposition 5. *The expected lifetime income of a worker of type $i = h, l$, as a function of the current belief ratio P and of the firm prior P_0, is given by*

$$W_i\left(P \middle| P_0\right) = \frac{2\sigma}{s\Delta} \left(\frac{P^{\alpha^-}}{P_0/P_0^i + P} \int_0^P \frac{P_0/P_0^i + x}{(1+x)\, x^{\alpha^-}} dx + \frac{P^{\alpha^+}}{P_0/P_0^i + P} \int_P^\infty \frac{P_0/P_0^i + x}{(1+x)\, x^{\alpha^+}} dx \right) + \frac{\mu_l}{r+\delta} \text{ ,}$$

where Δ, α^- and α^+ are as in Proposition 1.

Distinguishing pooling and separating equilibria is a simple matter of defining initial beliefs. In pooling equilibria, firms form their prior based on their knowledge of the share, χ_h, of workers with low educational costs, so that $p_0 = \chi_h \pi_h + (1 - \chi_h)\pi_l$. In separating equilibria, a worker's private information is revealed. Then employers and employees share the same prior, that is $P_0 = P_0^l = \pi_l/(1-\pi_l)$ when the worker education $e = 0$, and $P_0 = P_0^h = \pi_h/(1-\pi_h)$ when the worker education is equal to the separating level e^{Sep}. A separating profile is an equilibrium when it is not profitable for low-types to send the high-types' signal

$$W_l\left(P_0^h \middle| P_0^h\right) - c(e^{Sep}, \mu_l) \leq W_l\left(P_0^l \middle| P_0^l\right) - c(0, \mu_l) \text{ .} \tag{7}$$

The key difference with the purely asymmetric information model is that low-types have a stronger incentive to deviate when σ increases. This property allows us to establish the following Proposition.

Proposition 6. *If $\pi_h < 1$, the separating level of education e^{Sep} is an increasing function of the variance σ of output realizations.*

Proposition 6 illustrates that Riley's prediction holds true in our extended set-up but also that it crucially hinges on the introduction of symmetric uncertainty. Thus, strictly speaking, it cannot be interpreted as a test of the basic Spence's model. Another important caveat is that the relationship between the speed of learning and schooling is unambiguously negative solely when it is driven by an increase in signal precision. Conversely, the effect on schooling of more informative signals is ambiguous.

To see this, consider an increase in the return to ability μ_h. It obviously raises the signal/noise ratio and so the speed of learning. For the same reasons than before, this lowers the future gains from masquerading as a high-type. On the other hand, for any given belief, the immediate gains from masquerading are higher. The overall effect of an increase in μ_h is therefore ambiguous. As shown in Figure 4, when signal precision is relatively low, the second effect dominates because low-types are able to deceive their employers for a longer period.

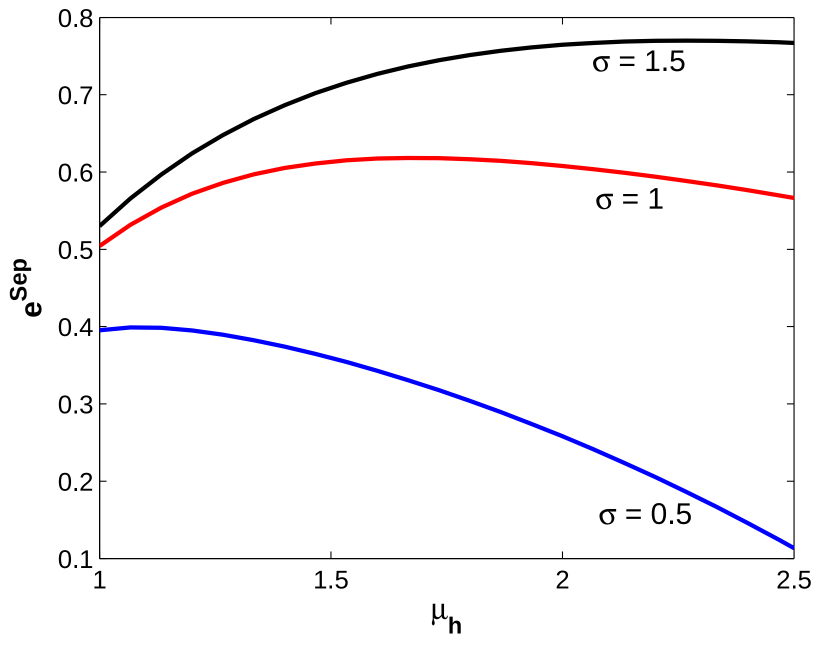

Figure 4: Separating level of education as a function of μ_h. Parameters: r $= \delta = 0.1$, $\mu_l = 0.5$, $\pi_l = 0.1$, $\pi_h = 0.9$, $\chi = 0.5$ and $c(e, \mu) = exp(e/\mu) - 1$.

Given that the speed of learning is increasing in μ_h, Figure 4 illustrates that its relationship with the separating level of education is ambiguous. This implies that predictions about inter-industry differentials are implicitly based on the premise of constant returns to ability. Consider for example the positive correlation between average schooling and industry wage differentials (Murphy and Topel, 1990). When signal precision is high, as in the lower curve of Figure 4, education is decreasing in returns to ability which obviously reinforces the effect described in Riley (1979b). On the other hand, when signal precision is low, as in the upper curve of Figure 4, the correlation goes in the opposite direction. Moreover, the relationship is concave because the negative learning effect grows stronger while the positive effect on wages remain constant.[23] Hence, the productivy/education gradient is higher in sectors where educated workers have more schooling, that is precisely the opposite to what happens when σ varies across industries. For brevity, we do not explicitly model how workers allocate themselves across sectors since reasonable mechanisms would imply that talented workers sort themselves into industries with higher returns to ability.[24] Accordingly, the average educational attainment in these industries will also be higher. Thus, for the parameter of the upper curve in Figure 4, returns to ability generate a positive correlation between average schooling and wage differentials, as documented in the data.

Similarly, it is easily seen that the relationship between schooling and residual wage dispersion is ambiguous. The aggregate wage distribution of the model with symmetric uncertainty is composed of two underlying density functions, that is one for each level of education. Both densities are given by the expression in Proposition 2 with $p_0 = \pi_l$ when $e = 0$, and $p_0 = \pi_h$ when $e = e^{Sep}$. As explained in Section 2.3, inequality is increasing in the speed of employer learning. One can therefore infer from Proposition 6 that schooling and residual wage dispersion should be negatively correlated across occupations as long as these occupations differ solely with respect to the observability of workers' types. Instead, if occupations have different returns to ability, the sign of the relationship can go either way.

We have shown in this section that returns to ability and signal precision may have opposite and potentially countervailing effects on the relationship between education and the speed of learning. This makes it difficult, if not impossible, to draw conclusive evidence from inter-industry data without disentangling the two components of the signal/noise ratio. Obviously, one could avoid this difficulty by maintaining the premise that returns to ability are constant across industries. Empirical evidence, however, points to the opposite. Murphy and Topel (1990) illustrate that sorting on abilities is

[23] More precisely, $\partial w(\pi_h)/\partial \mu_h = \partial(\pi_h(\mu_h - \mu_l) + \mu_l)/\partial \mu_h = \pi_h$.

[24] For example, it would be relatively straightforward to introduce a sorting mechanism similar to the one in Kaymak (2006) where workers choose their occupations on the basis of personal tastes. Assuming that the average worker is indifferent would imply that industries with higher returns to ability attract more high-types.

a primary determinant of observed occupational wage differentials. Moreover, recent findings by Gibbons et al. (2005) suggest that this sorting pattern is largely explained by higher returns to skills in high wage occupations.

5 Conclusion

We have postulated an analytically tractable model where education plays a role as a job-market signal and worker productivity is revealed on-the-job as the result of Bayesian updating on the part of the firms. The addition of this realistic element causes the failure of standard arguments, such as the unavoidable selection of the 'Riley' separating equilibrium via the Intuitive Criterion.

The separating equilibrium, though, still plays a key role provided learning is not too fast. We find that the separating level of education does not depend on the output variance when, as in Spence's (1973) model, workers know their innate ability with certainty. Introducing some uncertainty on the side of the worker generates a potentially negative correlation between the speed of learning and educational attainment. But the relationship can also be positive when returns to ability increase.

We see our framework as a stepping stone to the integration of asymmetric uncertainty into dynamic models of the labor market. Its tractability makes it amenable to several extensions. Introducing search frictions would bridge the gap with modern theories of the labor market and make it possible to characterize the interactions between signaling and unemployment. Another natural extension would combine worker with match-specific uncertainty. Actually, if π_h and π_l are interpreted as match-specific, Section 4 already contains the asset values for such a model. The only remaining challenge consists in characterizing the optimal separation of matches.

APPENDIX: Proofs

Proof of Proposition 1: The wage does not directly depend on the worker's type, but solely on the current belief P. It is equal to the expected output $\overline{\mu}(P) \equiv (\mu_h - \mu_l)\left(\frac{P}{1+P}\right) + \mu_l$.

For a low ability worker, $dP_t = P_t s dZ_t$ and thus the asset value solves the HJB equation

$$(r+\delta) W_l(P) - \frac{1}{2}(Ps)^2 W_l''(P) = (\mu_h - \mu_l)\left(\frac{P}{1+P}\right) + \mu_l \ ,$$

which is a second order non-homogenous ODE with non-constant coefficients. An Euler equation allow to obtain the solution to the associated homogenous problem,

$$W_l^H(P) = C_{1l} P^{\alpha^-} + C_{2l} P^{\alpha^+} \ ,$$

where α^- and α^+ are the negative and positive roots of the quadratic equation

$$\alpha\left(\alpha-1\right)\frac{s^2}{2}-r-\delta=0\,.$$

Thus $\alpha^- = \frac{1}{2}(1-\Delta)$ and $\alpha^+ = \frac{1}{2}(1+\Delta)$ with $\Delta = \frac{1}{s}\sqrt{s^2+8(r+\delta)}$. Notice that $\alpha^+ - \alpha^- = \Delta$ and $\alpha^+ + \alpha^- = 1$.

To solve for the non-homogenous equation we use the method of variations of parameters. The non-homogenous term is composed of a non-linear function of P plus a constant term. Thus we can assume that the particular solution is of the form

$$W_l^{NH}\left(P\right)=\left[y_1\left(P\right)P^{\alpha^-}+y_2\left(P\right)P^{\alpha^+}\right]+\frac{\mu_l}{r+\delta}\,.$$

Standard derivations yield the system of equations

$$\begin{pmatrix} P^{\alpha^-} & P^{\alpha^+} \\ \alpha^- P^{\alpha^- -1} & \alpha^+ P^{\alpha^+ -1} \end{pmatrix}\begin{pmatrix} y_1'\left(P\right) \\ y_2'\left(P\right) \end{pmatrix}=\begin{pmatrix} 0 \\ -\frac{2\sigma}{(1+P)Ps} \end{pmatrix}.$$

The Wronskian of the two linearly independent solutions is

$$P^{\alpha^-}\alpha^+P^{\alpha^+-1}-P^{\alpha^+}\alpha^-P^{\alpha^--1}=\alpha^+-\alpha^-=\Delta\,.$$

Therefore

$$y_1\left(P\right)=\frac{2\sigma}{s\Delta}\int\frac{1}{\left(1+x\right)x^{\alpha^-}}dx \quad \text{and} \quad y_2\left(P\right)=\frac{2\sigma}{s\Delta}\int\frac{1}{\left(1+x\right)x^{\alpha^+}}dx\,.$$

Thus the general form of the particular solution reads

$$W_l^{NH}\left(P\right)=\frac{2\sigma}{s\Delta}\left(P^{\alpha^-}\int\frac{1}{\left(1+x\right)x^{\alpha^-}}dx+P^{\alpha^+}\int\frac{1}{\left(1+x\right)x^{\alpha^+}}dx\right)+\frac{\mu_l}{r+\delta}\,. \tag{8}$$

The bounds of integration and constants C_{1l} and C_{2l} of the homogenous solution are pinned down by the boundary conditions

$$W_l\left(P\right)\xrightarrow[P\to 0]{}\frac{\mu_l}{r+\delta} \quad \text{and} \quad W_l\left(P\right)\xrightarrow[P\to\infty]{}\frac{\mu_h}{r+\delta}\,. \tag{9}$$

Let us first consider the homogenous solution. Since $P^{\alpha^-}\to\infty$ as $P\uparrow 0$, the first boundary condition can be satisfied if and only if C_{1l} equals zero. Similarly, since $P^{\alpha^+}\to\infty$ as $P\uparrow\infty$, the second boundary condition allows us to set C_{2l} equal to zero. All that remains is to determine the integration bounds in equation (8). Consider the following function

$$W_l\left(P\right)=\frac{2\sigma}{s\Delta}\left(P^{\alpha^-}\int_0^P\frac{1}{\left(1+x\right)x^{\alpha^-}}dx+P^{\alpha^+}\int_P^\infty\frac{1}{\left(1+x\right)x^{\alpha^+}}dx\right)+\frac{\mu_l}{r+\delta}\,. \tag{10}$$

Let us examine first the limit when $P \uparrow 0$. Given that $P^{\alpha^-} \to \infty$ and $\int_0^P [(1+x)\, x^{\alpha^-}]^{-1} dx \to 0$ as $P \uparrow 0$, we can apply l'Hôpital's rule to determine the limit. Straightforward calculations show that $P^{\alpha^-} \int_0^P [(1+x)\, x^{\alpha^-}]^{-1} dx \to -P/[(1+P)\alpha^-] \to 0$ as $P \uparrow 0$. A similar argument yields $P^{\alpha^+} \int_P^\infty [(1+x)\, x^{\alpha^+}]^{-1} dx \to P/[(1+P)\alpha^+] \to 0$ as $P \uparrow 0$.[25] Hence, (10) satisfies the first boundary condition in (9). Now, consider the limit when $P \uparrow \infty$. We can again use l'Hôpital's rule since $P^{\alpha^-} \to 0$ and $\int_0^P [(1+x)\, x^{\alpha^-}]^{-1} dx \to \infty$ as $P \uparrow \infty$, so that $P^{\alpha^-} \int_0^P [(1+x)\, x^{\alpha^-}]^{-1} dx \to -1/\alpha^-$ as $P \uparrow \infty$. Similarly, we obtain $P^{\alpha^+} \int_P^\infty [(1+x)\, x^{\alpha^+}]^{-1} dx \to 1/\alpha^+$ as $P \uparrow \infty$. Hence we have

$$\lim_{P \to \infty} W_l(P) = \frac{2\sigma}{s\Delta} \left(\frac{1}{-\alpha^-} + \frac{1}{\alpha^+} \right) + \frac{\mu_l}{r+\delta} = \frac{2\sigma}{s} \left(\frac{-1}{\alpha^- \alpha^+} \right) + \frac{\mu_l}{r+\delta} = \frac{\mu_h}{r+\delta} ,$$

where the last equality follows from $\alpha^- \alpha^+ = -2(r+\delta)/s^2$. Hence we have established that (10) also satisfies the second boundary condition in (9), which completes the derivation of $W_l(P)$.

The asset value of the high-type is derived similarly. For a high ability worker, $dP_t = P_t s\,(sdt + dZ_t)$ and thus the asset value solves

$$(r+\delta)\, W_h(P) - Ps^2 W_h'(P) - \frac{1}{2} (Ps)^2 W_h''(P) = (\mu_h - \mu_l) \left(\frac{P}{1+P} \right) + \mu_l .$$

The homogenous solution is

$$W_l^H(P) = C_{1h} P^{\gamma^-} + C_{2h} P^{\gamma^+} ,$$

where γ^- and γ^+ are the negative and positive roots of the quadratic equation

$$\gamma(\gamma+1) \frac{s^2}{2} - r - \delta = 0 ,$$

so that $\gamma^- = \frac{1}{2}(-1-\Delta)$ and $\gamma^+ = \frac{1}{2}(-1+\Delta)$. The non-homogenous solution is of the form

$$W_h^{NH}(P) = \left[z_1(P)\, P^{\gamma^-} + z_2(P)\, P^{\gamma^+} \right] + \frac{\mu_l}{r+\delta} ,$$

where the functions $z_1(P)$ and $z_2(P)$ satisfy

$$\begin{pmatrix} P^{\gamma^-} & P^{\gamma^+} \\ \gamma^- P^{\gamma^- - 1} & \gamma^+ P^{\gamma^+ - 1} \end{pmatrix} \begin{pmatrix} z_1'(P) \\ z_2'(P) \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{2\sigma}{(1+P)Ps} \end{pmatrix} .$$

Following the same steps as before yields the solution in Proposition 1. ■

Proof of Proposition 2: The dynamics of the transition density of beliefs is captured by the Kolmogorov forward equation given in equation (4). By definition, the ergodic density satisfies the stationarity condition $df(p)/dt = 0$. The general solution of $f(p)$ reads

$$f(p) = C_{j(p)f} p^{-1-\eta} (1-p)^{\eta-2} + K_{j(p)f} p^{\eta-2} (1-p)^{-1-\eta} ,$$

[25] Notice that $\int_P^\infty [(1+x)\, x^{\alpha^+}]^{-1} dx < \int_P^\infty x^{-\alpha^+ - 1} dx = P^{-\alpha^+}/\alpha^+$. Thus $\int_P^\infty [(1+x)\, x^{\alpha^+}]^{-1} dx$ is bounded for all $P > 0$ and the asset equation is well defined.

where $j(p) = 1_{\{p \geq p_0\}}$ and η is as given in the statement. The value of C_{0f} can be set to zero since $\int_0^1 f(p) = 1$, whereas $\int_p^{p_0} x^{-1-\eta} dx \to \infty$ as $p \uparrow 0$. Similarly $K_{1f} = 0$ since $\int_{p_0}^p (1-x)^{-1-\eta} dx \to \infty$ as $p \uparrow 1$. The values of the two remaining constants are pinned down by the following requirements: (i) $\int_0^1 f(p)\, dp = 1$, (ii) $\lim_{p \to p_0^-} f(p) = \lim_{p \to p_0^+} f(p)$. Condition (ii) yields

$$\left(p_0^{\eta-2}(1-p_0)^{-1-\eta}\right) K_{0f} = \left(p_0^{-1-\eta}(1-p_0)^{\eta-2}\right) C_{1f}$$

thus

$$C_{1f} = \left(\frac{p_0}{1-p_0}\right)^{2\eta-1} K_{0f} \ . \tag{11}$$

Condition (i) can be expressed analytically by integration of the ergodic density and use of the change of variable $P = p/(1-p)$ to obtain

$$\begin{aligned} \int_0^{p_0} f(p)\, dp &= K_{0f} \int_0^{p_0} p^{\eta-2}(1-p)^{-1-\eta} dp \\ &= K_{0f} \int_0^{P_0} P^{\eta-2}(1+P) dP = K_{0f} \left(\frac{P_0^{\eta-1}}{\eta-1} + \frac{P_0^{\eta}}{\eta} \right) \ , \end{aligned}$$

and

$$\int_{p_0}^1 f(p)\, dp = C_{1f} \int_{p_0}^1 p^{-1-\eta}(1-p)^{\eta-2} dp = C_{1f} \left(\frac{P_0^{-\eta}}{\eta} + \frac{P_0^{1-\eta}}{\eta-1} \right) \ .$$

Thus conditions (i) translates to

$$\left(\frac{P_0^{\eta-1}}{\eta-1} + \frac{P_0^{\eta}}{\eta} \right) K_{0f} + \left(\frac{P_0^{-\eta}}{\eta} + \frac{P_0^{1-\eta}}{\eta-1} \right) C_{1f} = 1 \ .$$

And using (11) yields

$$K_{0f} \left(P_0^{\eta-1} + P_0^{\eta} \right) \left(\frac{1}{\eta-1} + \frac{1}{\eta} \right) = 1.$$

The proof is completed observing that $\frac{1}{\eta-1} + \frac{1}{\eta} = \xi$. ■

Proof of Lemma 1: In a pooling equilibrium, no information is revealed by the education level e_p. Hence, firms' initial beliefs are equal to the proportion of high-ability worker in the population, that is p_0. Define

$$L(e) = W_l(P_0) - c(e, \mu_l) - W_l^{Sep} + c(0, \mu_l).$$

By definition, $\bar{e}$ is such that $L(\bar{e}) = 0$. In a pooling equilibrium, the low-type's "participation constraint" is satisfied when

$$W_l(P_0) - c(e_p, \mu_l) \geq W_l^{Sep} - c(0, \mu_l),$$

i.e. $L(e_p) \geq 0$. Actually, the condition $L(e_p) \geq 0$ is necessary and sufficient for the existence of a pooling equilibrium at level e_p. For, if $L(e_p) \geq 0$, we can specify weakly consistent beliefs, e.g.

$P(e_p) = P_0$ and $P(e) = 0$ for $e \neq e_p$. It follows that $W_l^{Sep} - c(0, \mu_l) > W_l^{Sep} - c(e, \mu_l)$ for all $e \neq 0, e_p$; and thus e_p is sequentially rational for the low-type. Since $W_h(P_0) > W_l(P_0)$, we have that $W_h(P_0) - c(e_p, \mu_l) > W_l^{Sep} - c(0, \mu_l)$. Thus the "participation constraint" for the high-type is also satisfied, which in turn implies sequential rationality for the high-type with the given equilibrium beliefs.

All that remains is to notice that, since $L'(e) = -c_e(e, \mu_l) < 0$, $L(e)$ is a strictly decreasing function, and hence $L(e_p) \geq 0$ if and only if $e_p \in [0, \bar{e}]$. ∎

The following pure-calculus property will be useful in the proof of Proposition 3.

Lemma 2. *Let $H :]0, +\infty[\to \mathbb{R}$ be a twice continuously differentiable real function and let $C :]0, +\infty[\to \mathbb{R}_+$ be a continuous real function such that*

(H1) $\lim_{x \to 0} H(x) = \lim_{x \to \infty} H(x) = 0$, *and*

(H2) whenever $H(x) \leq 0$, it follows that $H''(x) + C(x)H'(x) < 0$.

Then $H(x) > 0$ for all $x \in]0, +\infty[$ (in other words, condition (H2) can only be fulfilled vacuously).

Proof. By contradiction, let $H(x_1) \leq 0$ for some $x_1 \in]0, +\infty[$. Suppose first that $H'(x_1) > 0$. We start by establishing that $H'(x)$ is strictly decreasing in $]0, x_1[$. (H2) implies that $H''(x_1) < -C(x_1)H'(x_1) \leq 0$. By continuity of H'', if $H''(x)$ is not strictly negative in $]0, x_1[$, there is a "last inflexion point before x_1", i.e. $x_0 < x_1$ such that $H''(x_0) = 0$ but $H''(x) < 0$ for all $x \in]x_0, x_1[$. This implies that H' is strictly decreasing on $[x_0, x_1]$ and thus (since $H'(x_1) > 0$) strictly positive. This implies in turn that H is strictly increasing on $[x_0, x_1]$ and thus (since $H(x_1) \leq 0$) strictly negative on $[x_0, x_1[$. It therefore follows from (H2) that

$$\begin{aligned} 0 = H''(x_0) < -C(x_0)H'(x_0) &= -C(x_0)\left(H'(x_1) - \int_{x_0}^{x_1} H''(x)dx\right) \\ &= -C(x_0)H'(x_1) + C(x_0)\int_{x_0}^{x_1} H''(x)dx \leq 0 \end{aligned}$$

a contradiction which shows that $H''(x) < 0$ for all $x \in]0, x_1]$. But this implies that for all $\underline{x} \in]0, x_1[$

$$H(\underline{x}) = H(x_1) - \int_{\underline{x}}^{x_1} H'(x)dx \leq -\int_{\underline{x}}^{x_1} H'(x_1)dx = -(x_1 - \underline{x})H'(x_1) < 0\,,$$

Hence $\lim_{x \to 0} H(x) \leq -x_1 H'(x_1) < 0$ which contradicts (H1). This proves that there exists no $x_1 > 0$ such that $H(x_1) \leq 0$ and $H'(x_1) > 0$.

Suppose now that $H'(x_1) < 0$. Then there exists $\varepsilon > 0$ such that $H(x) < 0$ for all $x \in]x_1, x_1 + \varepsilon]$. If there exists some $x > x_1 + \varepsilon$ such that $H(x) = 0$, by continuity of H there exists a "first zero after x_0", i.e. x^* such that $H(x^*) = 0$ but $H(x) < 0$ for all $x \in]x_1, x^*[$. But then it follows that $H'(x_2) > 0$ for some $x_2 \in]x_1, x^*[$ (else H is decreasing on $]x_1, x^*[$ and thus $0 = H(x^*) \leq H(x_1 + \varepsilon) < 0$, a contradiction). That is, we have found $x_2 > 0$ such that $H(x_2) < 0$ and $H'(x_2) > 0$. This contradicts our previous finding. It follows that $H(x) \neq 0$ for all $x > x_1 + \varepsilon$. By continuity of H, this means that $H(x) < 0$ for all $x > x_1$. But then there must exist some $x_2 > x_1$ such that $H'(x_2) > 0$. For, if not, H is strictly decreasing on $[x_1 + \varepsilon, +\infty[$ and thus bounded above by $H(x_1 + \varepsilon) < 0$, a contradiction with the boundary condition $\lim_{x \to \infty} H(x) = 0$. Again we have found $x_2 > 0$ such that $H(x_2) < 0$ and $H'(x_2) > 0$, contradicting our previous finding.

We conclude that $H'(x_1) = 0$ for all $x_1 \in]0, +\infty[$ with $H(x_1) \leq 0$. If there exists any x_1 with $H(x_1) < 0$, since $\lim_{x \to 0} H(x) = 0$ by continuity there must exist $0 < x' < x_1$ with $H(x') < 0$ and $H'(x') < 0$, a contradiction. Thus $H(x) \geq 0$ for all x. But, if $H(x_1) = 0$ for some $x_1 \in]0, +\infty[$, we also have $H'(x_1) = 0$. By (H2) this implies that $H''(x_1) < 0$ so that $H'(x) < 0$ on an interval $]x_1, x_1 + \varepsilon'[$, a final contradiction. ■

Proof of Proposition 3: Define for $i = h, l$

$$I_i(e, e_p) = W_h^{Sep} - c(e, \mu_i) - W_i(P_0) + c(e_p, \mu_i) \; . \tag{12}$$

As explained in Section 3, the pooling equilibrium e_p fails the Intuitive Criterion if and only if there exists an education level e^{Sep} such that $I_l(e^{Sep}, e_p) < 0$ and $I_h(e^{Sep}, e_p) > 0$. Let $e^*(e_p)$ denote the minimum education level that does not trigger a profitable deviation for the low-type, so that $I_l(e^*(e_p), e_p) = 0$. Since $I_h(e, e_p)$ is strictly decreasing in e, the pooling equilibrium is stable when $I_h(e^*(e_p), e_p) < 0$. To show that one can always find an output variance σ below which this requirement is fulfilled, we need the following claim.

Claim 1: The asset values of high-ability and low-ability workers are respectively decreasing and increasing functions of σ for all $P \in (0, \infty)$.

The claim is most easily established reversing the change of variable from P_t to p_t to obtain[26]

(i) High-type's expectation: $dp_t = p_t\,(1 - p_t)\,s\,(s\,(1 - p_t)\,dt + dZ_t)$,

(ii) Low-type's expectation: $dp_t = p_t\,(1 - p_t)\,s\,(-sp_t dt + dZ_t)$.

By definition

$$\begin{aligned} W_i\,(p_t) &= E_{p_t}\left[\int_t^{+\infty} e^{-(r+\delta)(\tau - t)} w(p_\tau) d\tau \middle|\, \mu_i\right] \\ &= \int_t^{+\infty} e^{-(r+\delta)(\tau - t)} E_{p_t}\left[\, w(p_\tau)|\, \mu_i\right] d\tau \text{ , for all } p_t \in (0,1) \text{ and } i = h, l \;, \end{aligned} \tag{13}$$

[26]Notice that when p_t goes to zero or one, its stochastic component vanishes which provides us with the boundary conditions described in equation (9).

where the second equality follows from Fubini's theorem. When the worker is of the high-type, so that $\mu_i = \mu_h$, we know from condition (i) above that p_t has a positive deterministic trend: $p_t\,(1-p_t)^2\,s^2$. Since $w\,(p_t) = p_t(\mu_h - \mu_l) + \mu_l$ is a linear function of p_t, it follows that $E_{p_t}\left[\,w(p_\tau)|\,\mu_h\right] > w(p_t)$ for all $\tau > t$, and so $W_h\,(p_t) > w(p_t)/(r+\delta)$. Similarly, condition (ii) above shows that p_t has a negative deterministic trend when the worker is of the low-type, so that $W_l\,(p_t) < w(p_t)/(r+\delta)$.

We first establish the claim for low-types. Consider two different values of σ such that $\overline{\sigma} > \underline{\sigma}$. We define the function $H_l(\,P|\,\overline{\sigma}, \underline{\sigma}) = W_l\,(\,P|\,\sigma = \overline{\sigma}) - W_l\,(\,P|\,\sigma = \underline{\sigma})$. The HJB equation implies that for all $P \in (0, \infty)$

$$\begin{aligned} H_l''\,(P) &= \frac{2}{(Ps(\overline{\sigma}))^2}\left[(r+\delta)W_l\,(\,P|\,\sigma = \overline{\sigma}) - w(P)\right] - \frac{2}{(Ps(\underline{\sigma}))^2}\left[(r+\delta)W_l\,(\,P|\,\sigma = \underline{\sigma}) - w(P)\right] \\ &< \frac{2(r+\delta)}{(Ps(\overline{\sigma}))^2}\left[W_l\,(\,P|\,\sigma = \overline{\sigma}) - W_l\,(\,P|\,\sigma = \underline{\sigma})\right]\ , \end{aligned}$$

where the second equality follows from $W_l\,(P) < w(P)/(r+\delta)$ and $s(\overline{\sigma}) < s(\underline{\sigma})$. Thus $H_l(P)$ satisfies property (H2) in Lemma 2 for the trivial case where $C(P) = 0$. Since (H1) follows from the boundary conditions (9), it must be the case that $H_l(P) > 0$, that is $W_l\,(\,P|\,\sigma = \overline{\sigma}) > W_l\,(\,P|\,\sigma = \underline{\sigma})$, for all $P\ \in]0, +\infty[$. The property for high-types is established in the same fashion considering $H_h(P) = W_h\,(\,P|\,\sigma = \underline{\sigma}) - W_h\,(\,P|\,\sigma = \overline{\sigma})$ and applying Lemma 2 with $C(P) = Ps^2$.

We are now in a position to prove the Proposition. Recalling the definition of $I_h(e, e_p)$, we have

$$\lim_{\sigma\to 0} I_h(e^*(e_p), e_p) = \lim_{\sigma\to 0}(W_h^{Sep} - W_h(P_0)) + \lim_{\sigma\to 0}(c(e_p, \mu_h) - c(e^*(e_p), \mu_h))\ .$$

We first show that the second term on the right-hand side is negative. Given that $W_l(P_0|\sigma = \infty) < W_h^{Sep}$ and $c_e(e, \mu) > 0$, the relation $I_l(e^*(e_p), e_p) = 0$ requires that $e^*(e_p|\sigma = \infty) > e_p$. Claim 1 implies that $e^*(e_p|\sigma) > e^*(e_p|\sigma = \infty)$ for all $\sigma < \infty$, which in turn yields $\lim_{\sigma\to 0}(c(e_p, \mu_h) - c(e^*(e_p), \mu_h)) < 0$. But $\lim_{\sigma\to 0}(W_h^{Sep} - W_h(P_0)) = 0$. Hence, one can always find a sufficiently low output variance to ensure that $I_h(e^*(e_p), e_p) < 0$.

Finally define $\sigma^*(e_p)$ to be the unique variance such that $I_h(e^*(e_p), e_p|\sigma) = 0$ if $I_h(e^*(e_p), e_p|\sigma = \infty) > 0$, and $\sigma^*(e_p) = \infty$ otherwise. The proposition is established noticing that, since $I_h(e, e_p|\sigma)$ is decreasing in e and $e^*(e_p|\sigma)$ is decreasing in σ, $I_h(e^*(e_p), e_p|\sigma) < 0$ if and only if $\sigma < \sigma^*(e_p)$. ■

Proof of Proposition 4: For ease of notation, we denote $I_i(e) = I_i(e^*(e), e)$. Recall that a pooling equilibrium with education level e_p fulfills the Intuitive Criterion if and only if $I_h(e_p) \leq 0$. In order to characterize the region where this requirement is fulfilled, we prove the following claim.

Claim 2: When $c_e(e, \mu)$ is weakly log-submodular, if the pooling equilibrium fails the Intuitive Criterion for a given level of education, this is also true for any education level above it.

To prove this claim, we differentiate the equality defining $e^*(e_p)$ to obtain

$$\frac{de^*(e_p)}{de_p} = \frac{c_e(e_p, \mu_l)}{c_e(e^*(e_p), \mu_l)} \ ,$$

hence $de^*(e_p)/de_p \in (0,1)$. Differentiating $I_h(e)$ with respect to e yields

$$\begin{aligned} \frac{dI_h(e_p)}{de_p} &= -c_e(e^*(e_p), \mu_h)\left(\frac{de^*(e_p)}{de_p}\right) + c_e(e_p, \mu_h) \\ &= \frac{c_e(e_p, \mu_h)c_e(e^*(e_p), \mu_l) - c_e(e_p, \mu_l)c_e(e^*(e_p), \mu_h)}{c_e(e^*(e_p), \mu_l)} \ . \end{aligned}$$

When $c_e(e, \mu)$ is weakly log-submodular, the last expression is larger than or equal to zero. Thus $dI_h(e_p)/de_p \geq 0$ and so, if $I_h(e_p) > 0$ for a given e_p, this is also true or any education level above it, as stated in Claim 2.

We now prove the Proposition. Let $\widetilde{e}(\sigma)$ denote the lowest education level such that $I(\widetilde{e}(\sigma)) = 0$, if it exists. According to Claim 2, the region where pooling equilibria fulfill the Intuitive Criterion should be below $\widetilde{e}(\sigma)$. In the no-learning case $\sigma = \infty$, we have that $I(0) > 0$ and so all pooling equilibria fail the Intuitive Criterion. To characterize how $\widetilde{e}$ changes with σ, consider the derivative of $I_h(e_p|\sigma)$ with respect to σ. Claim 1 implies that

$$\frac{\partial I_h(e_p|s)}{\partial \sigma} = -\frac{\partial W_h(P_0|\sigma)}{\partial \sigma} - c_e\left(e^*(e_p), \mu_h\right)\frac{\partial e^*(e_p)}{\partial \sigma} > 0 \ .$$

Differentiating the equality $I\left(\widetilde{e}(\sigma)\right) = 0$ with respect to σ therefore yields

$$\frac{d\widetilde{e}(\sigma)}{d\sigma} = -\frac{\partial I_h(\widetilde{e}(\sigma))/\partial \sigma}{\partial I_h(\widetilde{e}(\sigma))/\partial e} < 0 \ .$$

Hence the function $\widetilde{e}(\sigma)$ is strictly decreasing.[27] In the limit

$$\lim_{\sigma \to 0} I_h(e|\sigma) = -c\left(e^*(e), \mu_h\right) + c(e, \mu_h) \ ,$$

since $W_h(P_0|\sigma)$ converges to W_h^{Sep} as σ goes to zero. It follows that $\lim_{\sigma \to 0} e^*\left(\widetilde{e}(\sigma)\right) = \widetilde{e}(\sigma)$. From the definition of $e^*(\cdot)$ and the convexity of the educational costs, this can be true if and only if $\lim_{\sigma \to 0} \widetilde{e}(\sigma) = \infty$.

All that remains is to recall that, as shown in Lemma 1, e_p is a PBE if and only if $e_p \leq \overline{e}$. But $\overline{e}$ is strictly increasing in σ because $\partial W_l(P_0|\sigma)/\partial \sigma > 0$. Furthermore, since $W_l(P_0|\sigma)$ converges to W_l^{Sep} as σ goes to 0, it must be the case that $\lim_{\sigma \to 0} \overline{e}(\sigma) = 0$. We can therefore conclude that $\overline{e}(\sigma)$ and $\widetilde{e}(\sigma)$ eventually intersect for sufficiently high values of σ. ■

[27] Notice that when the cost function is log-linear, e.g. $c(e, \mu) = e^2/\mu$, $dI_h(e)/de = 0$ for all e. Then $d\widetilde{e}(\sigma)/d\sigma$ becomes infinite, because the $\widetilde{e}(\sigma)$ locus is vertical.

Proof of Proposition 5: Since the proof is essentially similar to that of Proposition 1, the exposition can be brief. The HJB equation for type i is

$$(r+\delta) W_i(P) - \frac{(Ps)^2}{P_0/P_0^i + P} W_i'(P) - \frac{(Ps)^2}{2} W_i''(P) = (\mu_h - \mu_l)\left(\frac{P}{1+P}\right) + \mu_l .$$

Direct verification shows that the homogenous solution reads

$$W_i^H(P) = A_{1i}\left(\frac{P^{\alpha^-}}{P_0/P_0^i + P}\right) + A_{2i}\left(\frac{P^{\alpha^-}}{P_0/P_0^i + P}\right) ,$$

where α^- and α^+ are defined in Proposition 1. Hence the particular solution is of the form

$$W_i^{NH}(P) = \left[v_1(P)\frac{P^{\alpha^-}}{P_0/P_0^i + P} + v_2(P)\frac{P^{\alpha^+}}{P_0/P_0^i + P}\right] + \frac{\mu_l}{r+\delta} ,$$

Standard derivations yield the following system of equations

$$\begin{pmatrix} \frac{P^{\alpha^-}}{P_0/P_0^i+P} & \frac{P^{\alpha^+}}{P_0/P_0^i+P} \\ \frac{(\alpha^- -1)P^{\alpha^-}+(\alpha^-)P^{\alpha^- -1}}{P_0/P_0^i+P} & \frac{(\alpha^+ -1)P^{\alpha^+}+(\alpha^+)P^{\alpha^+ -1}}{P_0/P_0^i+P} \end{pmatrix} \begin{pmatrix} v_1'(P) \\ v_2'(P) \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{2(\mu_h-\mu_l)}{(1+P)Ps^2} \end{pmatrix} .$$

The particular solution reads

$$W_i^{NH}(P) = \frac{2(\mu_h-\mu_l)}{s^2(\alpha^+ - \alpha^-)}\left(\frac{P^{\alpha^-}}{P_0/P_0^i + P}\int \frac{P_0/P_0^i + x}{(1+x)\,x^{\alpha^-}}dx + \frac{P^{\alpha^+}}{P_0/P_0^i + P}\int \frac{P_0/P_0^i + x}{(1+x)\,x^{\alpha^+}}dx\right) + \frac{\mu_l}{r+\delta} ,$$

Finally, l'hospital's rule allows one to establish that the boundary conditions (9) are satisfied by the expression in Proposition 5. ■

Proof of Proposition 6: In the 'Riley' separating equilibrium, low ability workers choose the minimum level of education, i.e. zero. Given that $c(0,\mu_l)$ can be normalized to be zero without loss of generality, e^{Sep} such that

$$c\left(e^{Sep},\mu_l\right) = W_l\left(P_0^h \middle| P_0^h\right) - W_l\left(P_0^l \middle| P_0^l\right) .$$

First, notice that evaluating the expression in Proposition 5 yields $W_l\left(P_0^l \middle| P_0^l\right) = \mu_l/(r+\delta)$, which is independent of σ. This is because the posterior belief p follows a martingale when the worker's and firm's priors coincide. Since the wage is an affine function of the belief and workers are risk-neutral, their asset values do not depend on the volatility of beliefs.

Conversely, $W_l\left(P_0^h \middle| P_0^h\right)$ varies with σ. An argument similar to the one in Claim 1 of Proposition 3 shows that $W_l\left(P_0^h \middle| P_0^h\right)$ is indeed an increasing function of σ. First, we establish that $W_l\left(P \middle| P_0^h\right) < w(P)/(r+\delta)$. Reversing the change of variable from P_t to p_t yields

$$dp_t = s^2 p_t^2\left[\frac{1-p_t}{(P_0/P_0^i)(1-p_t)+p_t} - (1-p_t)\right]dt + (1-p_t)p_t s^2 dZ_t .$$

Thus the deterministic trend of p_t is negative when the worker is of the low-type, so that $P_0^i = P_0^l$, and the initial belief $P_0 = P_0^h$. As shown in Claim 1, this implies that $W_l\left(P|\, P_0^h\right) < w(P)/(r+\delta)$. The proof that $W_l\left(P|\, P_0^h\right)$ is a decreasing function of s can now easily be established as in Claim 1, i.e. by showing that the functions $H_l(P) = W_l\left(P|\, P_0^h, \sigma = \overline{\sigma}\right) - W_l\left(P|\, P_0^h, \sigma = \underline{\sigma}\right)$ and $C(P) = \dfrac{(Ps)^2}{P_0/P_0^l + P}$ satisfy properties (H1) and (H2) in Lemma 2. ■

References

[1] Altonji, Joseph and Charles Pierret. 2001. "Employer learning and statistical discrimination". *Quarterly Journal of Economics*, 116, No. 1: 313-350.

[2] Banks, Jeffrey S. and Joel Sobel. 1987. "Equilibrium Selection in Signalling Games". *Econometrica* 55, No. 3 (May): 647–662.

[3] Cho, In-Koo and David M. Kreps. 1987. "Signaling Games and Stable Equilibria". *Quarterly Journal of Economics* 102, No. 2 (May): 179–221.

[4] Farber, Henry S. and Robert Gibbons. 1996. "Learning and Wage Dynamics". *Quarterly Journal of Economics* 111, No. 4: 1007-47.

[6] Gibbons, Robert and Lawrence F. Katz. 1991. "Layoffs and Lemons". *Journal of Labor Economics* 9, No. 4: 351-80.

[6] Gibbons, Robert, Lawrence F. Katz., Thomas Lemieuxa and Daniel Parent. 2005. "Comparative Advantage, Learning, and Sectoral Wage Differences". *Journal of Labor Economics* 23, No. 4: 681-723.

[7] Habermalz, Steffen. 2006. "The Speed of Employer Learning and Job Market Signaling Revisited". *IZA Discussion Paper* , No. 2309.

[8] Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover". *Journal of Political Economy* 87, No. 5: 972-90.

[9] Kaymak, Baris. 2006. "Signaling, Self-Selection and Employer Learning". Mimeo, University of Rochester.

[11] Lange, Fabian. 2007. "The Speed of Employer Learning". *Journal of Labor Economics* 25, No. 1: 1-35.

[11] Lange, Fabian and Robert Topel. 2006. "The Social Value of Education and Human Capital" *Handbook of Educational Economics*, Vol. 1, eds. Eric Hanushek and Finis Welch.

[12] Moscarini, Giuseppe. 2005. "Job Matching and the Wage Distribution". *Econometrica* 73, No. 2 (March): 481–516.

[13] Murphy, Kevin and Robert Topel. 1990. "Efficiency wages reconsidered: theory and evidence", in: Y. Weiss and G. Fishelson, eds., *Advances in the Theory and Measurement of Unemployment*, Macmillan, London.

[14] Riley, John G. 1979a. "Informational Equilibrium". *Econometrica* 47, No. 2 (March): 331-60.

[15] Riley, John G. 1979b. "Testing the Educational Screening Hypothesis". *Journal of Political Economy* 87, No. 5: S227-S252.

[16] Spence, Michael. 1973. "Job Market Signaling". *Quarterly Journal of Economics* 87, No. 3 (August): 355-374.

[17] Wolpin, Kenneth. 1977. "Education and screening". *American Economic Review* 67, No. 5 (August): 949-958.