The Advantage of Hiring the Disadvantaged

Jennifer Brown[*] Justin Tumlinson[†] & Felix Várdy[‡ §¶]

September 2024

SHORT ABSTRACT

In this extended abstract we derive conditions when an employer, considering two candidates with equal contemporaneous qualifications, should hire the one with a more lackluster early career performance or disadvantage. (1) When workers' learning rate increases in ability, and signals of human capital are precise, then a steeper trajectory indicates greater future productivity—hiring the poor early performer is optimal. (2) Introducing sufficient noise in (late) achievement flips the result—more strong signals indicate higher ability.

Preliminary empirical tests support these predictions: The unconditional probability of becoming CEO increases 0.02% (on a baseline of 14% among Compustat Capital IQ listed workers) for every 10 positions higher an individual's undergraduate *alma mater* is in the QS World University Rankings. However, conditional on having a top-5 MBA (according to the QS Global MBA Rankings), the probability of becoming CEO increases 0.022% for every 5 positions lower an individual's undergraduate *alma mater* is ranked. The effect magnitude monotonically increases from 0.00% as the definition of "top-X" MBA is made more precise.

JEL Classification: D81, D84, J24, J71

[*] *University of Utah*; Email: jen.brown@eccles.utah.edu
[†] *University of Exeter*; Email: j.tumlinson@exeter.ac.uk
[‡] *International Monetary Fund*; Email: fvardy@gmail.com
[§] Authors are listed alphabetically.
[¶] The authors wish to thank seminar audiences at *Harvard Business School* and *University of Exeter*.

1 Introduction

This extended abstract theoretically derives and empirically tests conditions under which an employer, considering two candidates with equal contemporaneous qualifications, should hire the one with the more lackluster early career performance or early disadvantage.[1]

We model the development of ability (or acquired human capital) according to Brownian motion with an initial position and drift, corresponding to workers' privilege (initial human capital) and innate talent (or learning rate), respectively.[2] Neither privilege, talent nor ability are directly observable. Only noisy signals of ability, taken at various points in time, like educational pedigree, are visible to potential employers.

First, we consider the perspective of an employer who wishes to select the most *talented* employee. If signals of human capital are precise (and the correlation between privilege and talent is not too positive), then a steeper trajectory indicates greater talent, and hiring the worker with poorer early performance is optimal. Introducing sufficient noise in (late) achievement measurement flips the result. In that case, more strong signals indicate higher ability. We derive necessary and sufficient condition for choosing the candidate with the lower early signal—i.e., when it is advantageous to hire the disadvantaged.

Preliminary empirical tests support these theoretical predictions: The unconditional probability of becoming CEO increases 0.02% (on a baseline of 14% in our rarefied sample of Compustat Capital IQ listed workers) for every 10 positions *higher* an individual's undergraduate *alma mater* is in the QS World University Rankings. However, conditional on having a top-5 MBA (according to the QS Global MBA Rankings), the probability of becoming CEO increases 0.43% for every position *lower* an individual's undergraduate *alma mater* is ranked. This relation increases monotonically from 0.00% when "top-X" MBA is defined as "any" to "top-50" to "top-10" to "top-5" (i.e. making the MBA signal more precise).

Next, we consider the perspective of an employer who wishes to select the employee with the

[1] This "extended abstract" will evolve into a full paper: (1) The basic theoretical results are complete; however, we are improving their exposition through graphical examples and more thorough supporting economic intuition. (2) Our empirical results included here are very preliminary. We have just obtained access to geocoded sample of the *National Longitudinal Study of Youth*, which will allow us to worker career/compensation evolution as a function of educational signals in a representative sample of the US population. (3) We will place our contribution in the literature better.

[2] Privilege and talent are assumed distributed according to a bivariate normal with a correlation coefficient.

highest *ability* at the moment of the late signal. We show that at the time of the late signal, this is always the worker with the highest early signal. Finally, we consider an extension of the model where the employer cares about about the cumulative ability of the worker from the late signal until his employment ends (i.e., he separates or the contract ends). We show that whenever an employer seeking the most *talented* worker would choose the one with the lower early signal, then there exists a combined discount-separation factor, such that for all lower factors, the employer prefers the worker with the lower signal.

The implication is that occupations with longer expected tenure favor talented disadvantaged individuals, because they have longer for their talents to develop and show through in productivity. Empirical verification of these results is ongoing.

1.1 Related Literature

Becker (1964) and Spence (1973) respectively laid the foundations of human capital and job market signalling upon which our work builds. Jovanovic (1979) introduced the concept that employers learn workers' true productivity by observing them over time. Although our implementation differs, Cunha and Heckman (2007) introduced a model of human capital development recognizing the importance of potential correlations between early household conditions and learning in later life. We are confident many others have added to the basic question of "How do employers choose employees for their talent and acquired ability, when these are unobservable?" but our search for their work is just getting started. Generally, the literature is rich with (1) models in which employers infer worker ability from a collection of noisy signals, and (2) models in which workers' human capital develops over time. In a few cases the heterogeneity of starting conditions matters. In this paper, we add the insights that (i) the noisy signals' rate of change over a career informs whether the signals are driven more by privilege or talent, and (ii) that the information contained in that rate change depends on the signal precision.

2 Theory

2.1 Theoretical Model

Let talent be given by random variable Θ with mean $\mu_\Theta \in \mathbb{R}$, variance $\sigma^2_\Theta > 0$, and realization $\theta \in \mathbb{R}$. Initial human capital, K_0, which we refer to as "privilege," is also random, with mean $\mu_0 \in \mathbb{R}$, variance $\sigma^2_0 > 0$, and realization $k_0 \in \mathbb{R}$. Talent and privilege are distributed jointly Normal, $N\left[\left(\mu_\Theta, \sigma^2_\Theta\right), \left(\mu_0, \sigma^2_0\right), \rho\right]$, with correlation coefficient $\rho \in [-1, 1]$. Conditional on realizations θ and k_0, human capital, K_t, develops over time according to a Brownian motion with initial position k_0, drift θ, and diffusion parameter $\sigma \geq 0$. That is, at time $t \in [0, \infty)$,

$$K_t = k_0 + \theta t + \sigma W_t \ ,$$

where W_t is a standard Wiener process. The unconditional mean and variance of K_t are denoted by μ_t and σ^2_t, while conditional means and variances are written as $\mu_{t|..}$ and $\sigma^2_{t|..}$, where $|..$ denotes the conditioning event.

Talent, privilege, and human capital are unobservable. However, at time t, an employer observes the realization s_t of a noisy signal $S_t = K_t + \varepsilon_t$, where noise ε_t is distributed $N\left[0, \sigma^2_{\varepsilon_t}\right]$ and independent. The (unconditional) mean and variance of signal S_t are $\hat{\mu}_t$ and $\hat{\sigma}^2_t$. The PDFs of random variables Θ, K_t and S_t are $\zeta[\theta]$, $\psi_t[k_t]$ and $\hat{\psi}_t[s_t]$, respectively, while $\psi[k_t, k_{t'}]$, $\hat{\psi}[s_t, s_{t'}]$, refer to the joint densities of human capital, and the signals thereof, at different times t and t'. Finally $\Phi[\cdot]$ and $\phi[\cdot]$ denote the CDF and PDF of the standard Normal distribution.

An employer observes a signal of a job candidate's human capital at two different points in time, t_1 and t_2, where $t_1 < t_2$. For readability, we slightly abuse notation and, in subscripting, drop the 't'. For example, $k_i \equiv k_{t_i}$ and $\mu_{i|\theta} \equiv \mu_{K_i|\theta}$, while $\hat{\rho}_{12|\theta}$ denotes the correlation between $S_1|\theta$ and $S_2|\theta$.

Observe that

$$\begin{aligned}
\mu_{1|\theta} &= \mu_0 + \rho\sigma_0 \frac{\theta - \mu_\Theta}{\sigma_\Theta} + t_1\theta \\
\sigma^2_{1|\theta} &= \sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_1 \\
\mu_{2|\theta} &= \mu_0 + \rho\sigma_0 \frac{\theta - \mu_\Theta}{\sigma_\Theta} + t_2\theta \\
\sigma^2_{2|\theta} &= \sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_2
\end{aligned}$$

From the standard formula for the correlation between sequential realizations of a Brownian motion, it follows that

$$\rho_{12|\theta} = \sqrt{\frac{\sigma^2_{1|\theta}}{\sigma^2_{2|\theta}}} = \sqrt{\frac{\sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_1}{\sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_2}}\,.$$

The "signalized" values are

$$\begin{aligned}
\hat{\sigma}^2_{1|\theta} &= \sigma^2_{1|\theta} + \sigma^2_{\varepsilon_1} = \sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_1 + \sigma^2_{\varepsilon_1} \\
\hat{\sigma}^2_{2|\theta} &= \sigma^2_{2|\theta} + \sigma^2_{\varepsilon_2} = \sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_2 + \sigma^2_{\varepsilon_2} \\
\hat{\rho}_{12|\theta} &= \rho_{12|\theta}\frac{\sigma_{1|\theta}\sigma_{2|\theta}}{\hat{\sigma}_{1|\theta}\hat{\sigma}_{2|\theta}} = \frac{\sigma_{1|\theta}}{\hat{\sigma}_{1|\theta}\hat{\sigma}_{2|\theta}}
\end{aligned}$$

Notice that conditional variances and correlations are the same for all $\theta \in \mathbb{R}$. This holds for human capital, K_t, as well as for signals, S_t.

Let $(X,Y) \sim N\left[\left(\mu_X, \sigma_X^2\right), \left(\mu_Y, \sigma_Y^2\right), \rho\right]$, $\mu_X, \mu_Y \in \mathbb{R}, \sigma_X^2, \sigma_Y^2 > 0$. Define $\hat{X} \equiv X + \varepsilon_X$, $\hat{Y} \equiv Y + \varepsilon_Y$, where the noise terms $\varepsilon_X \sim N\left[0, \sigma^2_{\varepsilon_X}\right]$ and $\varepsilon_Y \sim N\left[0, \sigma^2_{\varepsilon_Y}\right]$, $\sigma^2_{\varepsilon_X}, \sigma^2_{\varepsilon_Y} > 0$, are independent. The following technical lemma plays an important role in the analysis. It is proved in the Appendix.

Lemma 1. *$\left(\hat{X}, \hat{Y}\right)$ is distributed $N\left[\left(\mu_X, \hat{\sigma}_X^2\right), \left(\mu_Y, \hat{\sigma}_Y^2\right), \hat{\rho}\right]$, where $\hat{\sigma}_X^2 \equiv \sigma_X^2 + \sigma^2_{\varepsilon_X}$, $\hat{\sigma}_Y^2 \equiv \sigma_Y^2 + \sigma^2_{\varepsilon_Y}$, and $\hat{\rho} \equiv \rho\frac{\sigma_X\sigma_Y}{\hat{\sigma}_X\hat{\sigma}_Y}$.*

3 Selecting for Θ

Claim 2. *An employer selecting for greater talent, Θ, prefers the candidate with the higher first signal if the correlation between privilege and talent is high, the second signal is noisy, or the time*

between observations is short.

Specifically, he chooses the candidate with the higher first signal iff

$$\sigma_{\varepsilon_2}^2 > \frac{\sigma_0^2\left(1-\rho^2\right)-\rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_1}\left(t_2-t_1\right) \ .$$

The claim follows from the following proposition.

Proposition 1. *Let* $s_1' > s_1$.

$$\left(\Theta|s_2,s_1'\right) \overset{(\prec)}{\succ}_{FOSD} \left(\Theta|s_2,s_1\right) \iff \sigma_{\varepsilon_2}^2 \overset{(<)}{>} \frac{\sigma_0^2\left(1-\rho^2\right)-\rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_1}\left(t_2-t_1\right)$$

Proof of Proposition 1:

Conditional on signal realizations $(s_1,s_2)\in\mathbb{R}^2$, the odds ratio, R, that innate talent is θ' rather than θ is

$$R\left[s_1,s_2\right] \equiv \frac{\zeta\left[\theta'|\, s_1,s_2\right]}{\zeta\left[\theta|\, s_1,s_2\right]} \ .$$

Applying Bayes' Rule,

$$R\left[s_1,s_2\right] = \frac{\hat{\psi}\left[s_1,s_2|\,\theta'\right]}{\hat{\psi}\left[s_1,s_2|\,\theta\right]}\frac{\zeta\left[\theta'\right]}{\zeta\left[\theta\right]} \ .$$

By Lemma 1, the joint density of the signals conditional on θ is

$$\hat{\psi}\left[s_1,s_2|\,\theta\right] = \frac{e^{V\left[s_1,s_2;\left(\mu_{1|\theta},\hat{\sigma}_{1|\theta}^2\right),\left(\mu_{2|\theta},\hat{\sigma}_{2|\theta}^2\right),\hat{\rho}_{12|\theta}\right]}}{2\pi\hat{\sigma}_{1|\theta}^2\hat{\sigma}_{2|\theta}^2\sqrt{1-\hat{\rho}_{12|\theta}^2}} \tag{1}$$

where the function $V\left[\cdot\right]$ is given by

$$V\left[x,y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\hat{\sigma}_Y^2\right),\hat{\rho}\right] \equiv -\frac{\left(\frac{x-\mu_X}{\hat{\sigma}_X}\right)^2-2\hat{\rho}\frac{x-\mu_X}{\hat{\sigma}_X}\frac{y-\mu_Y}{\hat{\sigma}_Y}+\left(\frac{y-\mu_Y}{\hat{\sigma}_Y}\right)^2}{2\left(1-\hat{\rho}^2\right)}$$

Lemma 3. *Let* $\theta' > \theta$.

$$\frac{\partial \mathrm{R}\left[s_1,s_2\right]}{\partial s_1} > 0 \iff \sigma_{\varepsilon_2}^2 > \frac{\sigma_0^2\left(1-\rho^2\right)-\rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_1}\left(t_2-t_1\right)$$

Proof. Differentiating R with respect to s_1,

$$\frac{\partial R\left[s_1,s_2\right]}{\partial s_1}=\left(\Delta R_{\theta'}-\Delta R_{\theta}\right)R\left[s_1,s_2\right] \tag{2}$$

where

$$\Delta R_{\theta}\equiv\frac{\partial\hat{\psi}\left[s_1,s_2|\,\theta\right]/\partial s_1}{\hat{\psi}\left[s_1,s_2|\,\theta\right]}$$

Substituting $\hat{\psi}\left[s_1,s_2|\,\theta\right]$ from (1),

$$\begin{aligned}
\Delta R_{\theta} &= \frac{2\pi\hat{\sigma}_{1|\theta}^2,\hat{\sigma}_{2|\theta}^2\sqrt{1-\hat{\rho}_{12|\theta}^2}}{e^{V\left[s_1,s_2;\left(\mu_{1|\theta},\hat{\sigma}_{1|\theta}^2\right),\left(\mu_{2|\theta},\hat{\sigma}_{2|\theta}^2\right),\hat{\rho}_{12|\theta}\right]}}\frac{\partial}{\partial s_1}\left[\frac{e^{V\left[s_1,s_2;\left(\mu_{1|\theta},\hat{\sigma}_{1|\theta}^2\right),\left(\mu_{2|\theta},\hat{\sigma}_{2|\theta}^2\right),\hat{\rho}_{12|\theta}\right]}}{2\pi\hat{\sigma}_{1|\theta}^2,\hat{\sigma}_{2|\theta}^2\sqrt{1-\hat{\rho}_{12|\theta}^2}}\right]\\
&= \frac{\partial}{\partial s_1}V\left[s_1,s_2;\left(\mu_{1|\theta},\hat{\sigma}_1^2\right),\left(\mu_{2|\theta},\hat{\sigma}_2^2\right),\hat{\rho}_{12|\theta}\right]\\
&= \frac{\partial}{\partial s_1}\left[-\frac{\left(\frac{s_1-\mu_{1|\theta}}{\hat{\sigma}_{1|\theta}}\right)^2-2\hat{\rho}_{12|\theta}\frac{s_1-\mu_{1|\theta}}{\hat{\sigma}_{1|\theta}}\frac{s_2-\mu_{2|\theta}}{\hat{\sigma}_{2|\theta}}+\left(\frac{s_2-\mu_{2|\theta}}{\hat{\sigma}_{1|\theta}}\right)^2}{2\left(1-\hat{\rho}_{12|\theta}^2\right)}\right]\\
&= -\frac{1}{2\left(1-\hat{\rho}_{12|\theta}^2\right)}\frac{2}{\hat{\sigma}_{1|\theta}}\left(\frac{s_1-\mu_{1|\theta}}{\hat{\sigma}_{1|\theta}}-\hat{\rho}_{12|\theta}\frac{s_2-\mu_{2|\theta}}{\hat{\sigma}_{2|\theta}}\right)\\
&= -\frac{1}{2\left(1-\hat{\rho}_{12|\theta}^2\right)}\frac{2}{\hat{\sigma}_{1|\theta}}\left\{\frac{s_1-\left(\mu_0+\rho\sigma_0\frac{\theta-\mu_{\Theta}}{\sigma_{\Theta}}+t_1\theta\right)}{\hat{\sigma}_{1|\theta}}-\hat{\rho}_{12|\theta}\frac{s_2-\left(\mu_0+\rho\sigma_0\frac{\theta-\mu_{\Theta}}{\sigma_{\Theta}}+t_2\theta\right)}{\hat{\sigma}_{2|\theta}}\right\}
\end{aligned}$$

So, from (2), $\partial R\left[s_1,s_2\right]/\partial s_1$ takes the sign of

$$\begin{aligned}
\Delta R_{\theta'}-\Delta R_{\theta} &= \frac{\theta'-\theta}{\left(1-\hat{\rho}_{12|\theta}^2\right)\hat{\sigma}_{1|\theta}}\left(\frac{\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_1}{\hat{\sigma}_{1|\theta}}-\hat{\rho}_{12|\theta}\frac{\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_2}{\hat{\sigma}_{2|\theta}}\right)\\
&= \frac{\theta'-\theta}{\left(1-\hat{\rho}_{12|\theta}^2\right)\hat{\sigma}_{1|\theta}}\left(\frac{\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_1}{\hat{\sigma}_{1|\theta}}-\frac{\sigma_{1|\theta}}{\hat{\sigma}_{1|\theta}\hat{\sigma}_{2|\theta}}\frac{\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_2}{\hat{\sigma}_{2|\theta}}\right)
\end{aligned}$$

$$= \frac{\theta'-\theta}{\left(1-\hat{\rho}_{12|\theta}^2\right)\hat{\sigma}_{1|\theta}^2\hat{\sigma}_{2|\theta}}\left\{\left(\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_1\right)\hat{\sigma}_{2|\theta}-\sigma_{1|\theta}\left(\rho\frac{\sigma_0}{\sigma_{\Theta}}+t_2\right)\right\}$$

Since $\theta' > \theta$,

$$\begin{aligned}
\frac{\partial R\left[s_1, s_2\right]}{\partial s_1} > 0 \quad &\iff \quad \left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_1\right)\hat{\sigma}_{2|\theta} > \sigma_{1|\theta}\left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_2\right) \\
&\iff \quad \left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_1\right)\left(\sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_2 + \sigma_{\varepsilon_2}^2\right) > \left(\sigma_0^2\left(1-\rho^2\right) + \sigma^2 t_1\right)\left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_2\right) \\
&\iff \quad \sigma_0^2\left(1-\rho^2\right) + \rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2 t_2 + \left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_1\right)\sigma_{\varepsilon_2}^2 > t_2\sigma_0^2\left(1-\rho^2\right) + \rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2 t_1 \\
&\iff \quad \sigma_{\varepsilon_2}^2 > \frac{\sigma_0^2\left(1-\rho^2\right) - \rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta} + t_1}\left(t_2 - t_1\right)
\end{aligned}$$

□

From Lemma 3 we know that for $\theta' > \theta$ and $s_1' > s_1$,

$$\frac{\zeta\left[\theta'|s_1', s_2\right]}{\zeta\left[\theta|s_1', s_2\right]} \overset{(<)}{>} \frac{\zeta\left[\theta'|s_1, s_2\right]}{\zeta\left[\theta|s_1, s_2\right]} \iff \sigma_{\varepsilon_2}^2 \overset{(<)}{>} \frac{\sigma_0^2\left(1-\rho^2\right) - \rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta} + t_1}\left(t_2 - t_1\right)$$

Rearranging the LHS of the equivalence yields

$$\frac{\zeta\left[\theta'|s_1', s_2\right]}{\zeta\left[\theta'|s_1, s_2\right]} \overset{(<)}{>} \frac{\zeta\left[\theta|s_1', s_2\right]}{\zeta\left[\theta|s_1, s_2\right]} \iff \sigma_{\varepsilon_2}^2 \overset{(<)}{>} \frac{\sigma_0^2\left(1-\rho^2\right) - \rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta} + t_1}\left(t_2 - t_1\right)$$

Hence, MLRP holds for Θ with respect to s_1, with the "direction" (i.e., whether the ratio is increasing or decreasing in θ) turning on the inequality on the RHS. Since MLRP implies FOSD, Θ can be ranked as claimed.

This completes the proof of Proposition 1.

The employer's inference problem contains a tension between the second signal's noise and human capital development. Lower noise ($\sigma_{\varepsilon_2} \downarrow$), higher dispersion of privilege ($\sigma_0 \uparrow$) and greater time between signals ($t_2 - t_1 \uparrow$), all make it more likely that the wedge between the two candidates early signals ($s_1' > s_1$) is driven by the greater talent of the lower signaled individual closing the human capital gap, whereas the opposite conditions make human capital development less important relative to cutting through the noise (in the late signal) with an additional (early) signal.

4 Selecting for K_2

Claim 4. *An employer selecting for greater K_2 always prefers the candidate with the higher first signal.*

The claim follows from the following proposition.

Proposition 2.

$$s_1' \overset{(<)}{>} s_1 \iff \left(K_2|s_2, s_1'\right) \overset{(\prec)}{\succ}_{FOSD} \left(K_2|s_2, s_1\right)$$

Proof of Proposition 2:

Conditional on signal realizations s_1, s_2, the odds ratio, Q, that ability at time t_2 is k_2' versus k_2 is

$$Q\left[s_1, s_2\right] \equiv \frac{\psi_2\left[k_2'|\, s_1, s_2\right]}{\psi_2\left[k_2|\, s_1, s_2\right]} = \frac{\hat{\psi}\left[s_1, s_2|\, k_2'\right]}{\hat{\psi}\left[s_1, s_2|\, k_2\right]} \frac{\psi_2\left[k_2'\right]}{\psi_2\left[k_2\right]} = \frac{\hat{\psi}_1\left[s_1|\, k_2'\right]}{\hat{\psi}_1\left[s_1|\, k_2\right]} \frac{\hat{\psi}_2\left[s_2|\, k_2'\right]}{\hat{\psi}_2\left[s_2|\, k_2\right]} \frac{\psi_2\left[k_2'\right]}{\psi_2\left[k_2\right]}$$

where the first equality follows from Bayes' Rule and the second because s_1 and s_2 are independent conditional on k_2.

Lemma 5. *Let $k_2' > k_2$.*

$$\frac{\partial \mathrm{Q}\left[s_1, s_2\right]}{\partial s_1} > 0 \ .$$

Proof. Differentiating Q with respect to s_1,

$$\frac{\partial Q\left[s_1, s_2\right]}{\partial s_1} = \left(\Delta Q_{k_2'} - \Delta Q_{k_2}\right) Q\left[s_1, s_2\right]$$

where

$$\Delta Q_{k_2} \equiv \frac{\partial\, \hat{\psi}_1\left[s_1|\, k_2\right] \Big/ \partial s_1}{\hat{\psi}_1\left[s_1|\, k_2\right]}$$

Applying Bayes' Rule

$$\hat{\psi}_1\left[s_1|\, k_2\right] = \int \hat{\psi}_1\left[s_1|\, k_2, \theta\right] \zeta\left[\theta|\, k_2\right] d\theta = \int \hat{\psi}_1\left[s_1|\, k_2, \theta\right] \frac{\psi_2\left[k_2|\, \theta\right] \zeta\left[\theta\right]}{\psi_2\left[k_2\right]} d\theta$$

The component densities are

$$\begin{aligned}\psi_2\left[k_2|\,\theta\right] &= \frac{1}{\sigma_{2|\theta}}\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\\ \hat{\psi}_1\left[s_1|\,k_2,\theta\right] &= \frac{1}{\hat{\sigma}_{1|k_2,\theta}}\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\end{aligned}$$

where

$$\mu_{1|k_2,\theta}\equiv k_2-(t_2-t_1)\,\theta;\ \hat{\sigma}^2_{1|k_2,\theta}\equiv\sigma^2_{\varepsilon_1}+(t_2-t_1)\,\sigma$$

Thus, the denominator of ΔQ_{k_2} is

$$\hat{\psi}_1\left[s_1|\,k_2\right]=\frac{1}{\psi_2\left[k_2\right]}\int\frac{1}{\hat{\sigma}_{1|k_2,\theta}}\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\frac{1}{\sigma_{2|\theta}}\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\frac{1}{\sigma_\theta}\phi\left[\frac{\theta-\mu_\theta}{\sigma_\theta}\right]d\theta$$

and the numerator is

$$\begin{aligned}\frac{\partial}{\partial s_1}\hat{\psi}_1\left[s_1|\,k_2\right] &= \frac{1}{\psi_2\left[k_2\right]}\int\frac{1}{\hat{\sigma}^2_{1|k_2,\theta}}\phi'\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\frac{1}{\sigma_{2|\theta}}\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\frac{1}{\sigma_\Theta}\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta\\ &= -\frac{1}{\psi_2\left[k_2\right]}\int\frac{1}{\hat{\sigma}^2_{1|k_2,\theta}}\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\frac{1}{\sigma_{2|\theta}}\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\frac{1}{\sigma_\Theta}\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta\end{aligned}$$

where the last equality uses $\phi'\left[x\right]=-x\phi\left[x\right]$. Hence, dividing numerator by denominator,

$$\begin{aligned}\Delta Q_{k_2} &= -\frac{\int\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}{\int\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}\\ &= -\frac{\int\left(\frac{s_1-k_2}{\hat{\sigma}_{1|k_2,\theta}}+\frac{(t_2-t_1)\theta}{\hat{\sigma}_{1|k_2,\theta}}\right)\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}{\int\phi\left[\frac{s_1-\mu_{1|k_2,\theta}}{\hat{\sigma}_{1|k_2,\theta}}\right]\phi\left[\frac{k_2-\mu_{2|\theta}}{\sigma_{2|\theta}}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}\\ &= \frac{1}{\frac{1}{t_2-t_1}\hat{\sigma}^2_{1|k_2,\theta}}\left(\mu_A-\frac{\int\theta\phi\left[\frac{\mu_A-\theta}{\sigma_A}\right]\phi\left[\frac{\mu_B-\theta}{\sigma_B}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}{\int\phi\left[\frac{\mu_A-\theta}{\sigma_A}\right]\phi\left[\frac{\mu_B-\theta}{\sigma_B}\right]\phi\left[\frac{\theta-\mu_\Theta}{\sigma_\Theta}\right]d\theta}\right)\end{aligned}$$

where we use the following shorthand:

$$\mu_A\equiv\frac{k_2-s_1}{t_2-t_1},\ \sigma_A\equiv\frac{\hat{\sigma}_{1|k_2,\theta}}{t_2-t_1},\ \mu_B\equiv\frac{k_2-\mu_0+\rho\frac{\sigma_0}{\sigma_\Theta}\mu_\Theta}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_2},\ \sigma_B\equiv\frac{\sigma_{2|\theta}}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_2}$$

Applying Lemma 9,

$$\Delta Q_{k_2} = \frac{1}{\frac{1}{t_2-t_1}\hat{\sigma}_{1|k_2,\theta}} \left(\mu_A - \frac{\frac{\mu_A}{\sigma_A^2} + \frac{\mu_B}{\sigma_B^2} + \frac{\mu_\theta}{\sigma_\theta^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_\theta^2}} \right) ,$$

Hence, $\frac{\partial Q[s_1,s_2]}{\partial s_1}$ takes the sign of

$$\Delta Q_{k_2'} - \Delta Q_{k_2} = \frac{1}{\hat{\sigma}^2_{1|k_2,\theta}} \left(\frac{k_2' - k_2}{t_2 - t_1} - \frac{\frac{k_2'-k_2}{t_2-t_1} + \frac{\frac{k_2'-k_2}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_2}}{\sigma_B^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_\theta^2}} \right) = \frac{k_2' - k_2}{\sigma^2_{s_1|k_2,\theta}} \left(1 - \frac{\frac{1}{\sigma_A^2} + \frac{\frac{t_2-t_1}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_2}}{\sigma_B^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_\theta^2}} \right)$$

Since $k_2' > k_1$, $\frac{\partial Q[s_1,s_2]}{\partial s_1}$ takes the sign of the parenthesied factor and is positive iff

$$1 > \frac{\frac{1}{\sigma_A^2} + \frac{\frac{t_2-t_1}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_2}}{\sigma_B^2}}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma_B^2} + \frac{1}{\sigma_\theta^2}} \iff 1 + \frac{\sigma_B^2}{\sigma_\Theta^2} > \frac{t_2 - t_1}{\rho\frac{\sigma_0}{\sigma_\Theta} + t_2} \iff \rho\frac{\sigma_0}{\sigma_\Theta} + \frac{\sigma_B^2}{\sigma_\Theta^2}\left(\rho\frac{\sigma_0}{\sigma_\Theta} + t_2\right) > -t_1$$

which is always true. □

Lemma 5 implies that for $k_2' > k_2$,

$$\frac{\psi_2\left[k_2'|\, s_1', s_2\right]}{\psi_2\left[k_2|\, s_1', s_2\right]} \overset{(<)}{>} \frac{\psi_2\left[k_2'|\, s_1, s_2\right]}{\psi_2\left[k_2|\, s_1, s_2\right]} \iff s_1' \overset{(<)}{>} s_1$$

Rearranging the LHS of the equivalence yields

$$\frac{\psi_2\left[k_2'|\, s_1', s_2\right]}{\psi_2\left[k_2'|\, s_1', s_2\right]} \overset{(<)}{>} \frac{\psi_2\left[k_2|\, s_1, s_2\right]}{\psi_2\left[k_2|\, s_1, s_2\right]} \iff s_1' \overset{(<)}{>} s_1$$

Hence, MLRP holds for K_2 with respect to s_1. Since MLRP implies FOSD, K_2 can be ranked as claimed.

This completes the proof of Proposition 2.

4.1 Selecting for cumulative ability

This section to be added. Notes in the appendix. These need to be cleaned into proposition-proof form.

5 Empirical Analysis

In this section, we ask whether, conditional on having attended a top graduate business program, the probability of an executive achieving the rank of CEO is correlated (and, if so, in what direction) with the rank of the *undergraduate* institution that he or she attended, unconditionally and conditional on the rank of his or her MBA granting institution, if any.

5.1 Data

Data on individuals' career was retrieved from Compustat's Capital IQ People Intelligence. The profiles include the individuals' title, and we flagged all "Chief Executive Officers" and "Co-Chief Executive Officers" as CEOs. From this data source, we also collected information on the institutions at which individuals earn their undergraduate degrees and, if applicable, graduate degrees. We then matched universities against undergraduate and MBA rankings, respectively.

First, we used the undergraduate university rankings from the 2023 QS World University Rankings, published by Quacquarelli Symonds, which assigns ranks to 400 universities across the world.[3] Second, we used MBA university rankings are the 2023 QS rankings, which assign ranks to the top 250 MBA programs each year. We matched the executive's universities to the QS rankings to create our main dataset. Summary statistics are presented in Table A1.

INSERT TABLE A1 HERE

5.2 Results

Table 2 presents our regression results. Individual columns report coefficient estimates from increasingly restrictive subsamples. The first column includes the entire sample of individual listed in Compustat Capital IQ for which we could match a ranked undergraduate institution; the second column is restricted to individuals who earned an MBA; the third column includes individuals whose MBA is ranked in the top-50; the fourth column includes individuals with a top-10 MBA; and the fifth column includes individuals with a top-5 MBA. In all cases, the dependent variable is

[3]Leading university rankings— for example, the QS World University Rankings, Times Higher Education Rankings, and Shanghai Academic Ranking of World Universities—are highly correlated (Aguillo et al., 2010). Because they cover more universities, we rely on the QS Rankings as a measure of university quality.

an indicator that is equal to 100% if the individual achieves the rank of CEO, and the independent variable of interest is the rank of the individual's undergraduate institution.

In the full sample (column 1) (i.e. unconditional on subsequent MBA achievement), every 10 positions that an individuals' undergraduate *alma mater* is further down the league table is associated with a 0.02% drop in the probability that he/she makes CEO. Of course, this estimate is not the causal effect of university rank on becoming CEO. Indeed, the point of the theoretical model is that human capital and its components—innate ability and privilege, clear determinants of becoming CEO—are unobservable to the econometrician and labor market, especially at the time of earning an MBA. The omitted correlates of undergraduate institution rank that drive becoming CEO are myriad and sure to bias any estimate. Still, the correlation is non-trivial: In our rarefied sample, where 14% of individuals become CEO, a graduate from the number one ranked university is 0.14% more likely reach the position than one from number ten.

Restricting to workers with an MBA eliminates nearly 90% of our observations. Moreover, restricting the sample to those whose MBA signal more precisely measures human capital at that time flips the result: worse ranked undergraduate institutions predict the individual becoming CEO. Furthermore, in support of the theoretical model, as precision of "top-X" MBA gets more precise in columns (3)–(5), the stronger predictor of becoming CEO that low ranked undergraduate *alma mater* becomes, until the magnitude of this opposite signed effect is more than twice that in unconditional sample.

INSERT TABLE 2 HERE

6 Conclusion

We have proposed a model of worker human capital from the perspective of a hiring manager. Workers' human capital (productivity) evolves linearly as a function of stochastic *privilege* (level) and *innate ability* (slope) over time. Employers cannot observe either human capital or its components directly, but instead see only noisy periodic signals of it. Within this framework, we have analyzed the following question: "Facing two candidates with identical contemporaneous human capital signals, when should a manager who seeks the most able employee hire the candidate with weaker earlier signal?" We derived a sharp analytical condition that is a function of (i) the precision

of the late signal, (ii) the dispersion of privilege and (iii) the time between signals. Specifically, holding (ii) and (iii) fixed, above a certain level of late signal precision, the ability of the candidate with the weaker early signal *first order stochastically dominates* the ability of the candidate with the stronger early signal and, below this level, the opposite holds.

Our preliminary empirical tests support the model. We find that although attending a higher ranked undergraduate institution (i.e. the early signal) positively correlates with the probability of rising to CEO (our proxy for high ability), unconditionally, when we condition on having an MBA (i.e. the late signal), the correlation vanishes. When we condition on having a top-50 MBA (i.e. a more precise late signal), attending a higher ranked undergraduate institution negatively correlates with making it to CEO. As we make the second signal more precise–top-10 and then top-5–this negative correlation gets monotonically stronger.

The above tests, while suggestive, are imperfect. First, moving from top-50 to top-10 to top-5 increases not just the precision but also level of the late signal. Second, the model predicts that for sufficiently noisy late signals, the ability of those with better early signals *first order stochastically dominates* the ability of those with weaker ones. Of course, not conditioning on a second signal at all is, in the limit, equivalent to conditioning on a "sufficiently noisy second signal," and our estimated coefficients have the correct sign, we would prefer to get the result without such extreme late signal noise. Unfortunately, "any MBA" was not a sufficiently noisy second signal to induce a positive conditional correlation between better undergraduate institution and the probability of becoming CEO. Finally, our theoretical results imply other empirical hypotheses that our data does not allow us to test. We are gathering new data to improve on these preliminary results.

References

Aguillo, Isidro, Judit Bar-Ilan, Mark Levene, and José Ortega. 2010. "Comparing university rankings." *Scientometrics*, 85(1): 243–256.

Becker, Gary S. 1964. *Human capital: A theoretical and empirical analysis, with special reference to education.* University of Chicago press.

Cunha, Flavio, and James Heckman. 2007. "The technology of skill formation." *American economic review*, 97(2): 31–47.

Jovanovic, Boyan. 1979. "Job matching and the theory of turnover." *Journal of political economy*, 87(5, Part 1): 972–990.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics*, 87(3): 355–374.

Table 1. Summary Statistics

	All Executives N=595,403	With MBA N=69,425	With top 10 MBA N=31.358
CEO flag (%)	14.2	19.2	21.6
Rank of undergraduate institution	352.4	225.4	171.8
Rank of MBA program	–	21.4	3.29

This table reports summary statistics. More description to come.

Table 2. Regressions with Executives' Education

Dependent Variable: Is the individual a CEO?					
	Not conditional on MBA	Conditional on			
		any MBA	top-50 MBA	top-10 MBA	top-5 MBA
	(1)	(2)	(3)	(4)	(5)
Rank of undergraduate	-0.00200***	0.00004	0.00095*	0.00326***	0.00428***
institution	(0.00012)	(0.00050)	(0.00057)	(0.00090)	(0.00106)
Constant	14.90419***	19.20169***	19.42544***	21.06506***	22.35187***
	(0.06311)	(0.18696)	(0.19980)	(0.27572)	(0.31382)
R^2	0.01	0.01	0.01	0.01	0.01
No. of observations	595,403	69,425	58,993	31,358	24,805

This table reports the executive regression results. Variables are defined as in Table 1; that is, the dependent variable equals 100(%) when the individual is a CEO and 0(%) otherwise. Column 1 includes all individuals with a listed undergraduate degree. Columns 2–5 include individuals with MBAs from any, top-50, top-10 and top-5 institutions, respectively.

A Appendix: Auxiliary Results on Normal Random Variables

The PDF of a Normal random variable X with realization x, mean μ, and variance σ^2 is

$$\frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{e^{U\left[x;\left(\mu,\sigma^2\right)\right]}}{\sigma\sqrt{2\pi}} \ ,$$

where

$$U\left[x;\left(\mu,\sigma^2\right)\right] \equiv -\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \ .$$

Let $\{X,Y\}$ be distributed $N\left[\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right]$. The joint PDF of $\{X,Y\}$ is

$$f\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right] \equiv \frac{e^{V\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \ ,$$

The following technical lemmas are useful in Section 2 above:

Proposition 3. *Let $\hat{X} \equiv X+\varepsilon$, where $\varepsilon \tilde{} N\left[0,\sigma_\varepsilon^2\right]$. Then $\left\{\hat{X},Y\right\}$ is distributed $N\left[\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right]$, where $\hat{\sigma}_X^2 \equiv \sigma_X^2+\sigma_\varepsilon^2$, and $\tilde{\rho} \equiv \rho\frac{\sigma_X}{\hat{\sigma}_X}$.*

Proof. Define

$$Z\left[x\right] \equiv U\left[\hat{x};\left(x,\sigma_\varepsilon^2\right)\right] + V\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right]$$

such that

$$Z'\left[x\right] = \frac{\hat{x}-x}{\sigma_\varepsilon^2} - \frac{\frac{x-\mu_X}{\sigma_X^2}-\rho\frac{y-\mu_Y}{\sigma_X\sigma_Y}}{1-\rho^2} \text{ and } Z''\left[x\right] = -\frac{1}{\sigma_\varepsilon^2} - \frac{1}{\left(1-\rho^2\right)\sigma_X^2}$$

We write $Z'' \equiv Z''\left[x\right]$, because $Z''\left[x\right]$ does not depend on x.The joint PDF of $\left\{\hat{X},Y\right\}$ is

$$\begin{aligned}
h\left[\hat{x},y\right] &= \int_{-\infty}^{+\infty} h\left[\hat{x}\middle|\, x,y\right] f\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right] dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_\varepsilon}\phi\left[\frac{\hat{x}-x}{\sigma_\varepsilon}\right] f\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right] dx \\
&= \int_{-\infty}^{+\infty} \frac{e^{U\left[\hat{x};\left(x,\sigma_\varepsilon^2\right)\right]}}{\sigma_\varepsilon\sqrt{2\pi}} \frac{e^{V\left[x,y;\left(\mu_X,\sigma_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\rho\right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} dx \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{Z\left[x\right]} dx
\end{aligned}$$

Applying Lemmas 6 and 7 below, we find

$$\begin{aligned} h\left[\hat{x},y\right] &= \frac{\sqrt{-Z''}}{2\pi\sqrt{2\pi}\hat{\sigma}_X\sigma_Y\sqrt{1-\tilde{\rho}^2}}\int_{-\infty}^{+\infty} e^{V\left[\hat{x},y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right]-\frac{1}{2}\left(Z'[x]\Big/\sqrt{-Z''[x]}\right)^2}dx \\ &= f\left[\hat{x},y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right]\sqrt{-Z''}\int_{-\infty}^{+\infty}\phi\left[Z'\left[x\right]\Big/\sqrt{-Z''}\right]dx \\ &= f\left[\hat{x},y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right]\sqrt{-Z''}\left.\frac{\frac{1}{2}-\Phi\left[Z'\left[x\right]/\sqrt{-Z''}\right]}{\sqrt{-Z''}}\right|_{x=-\infty}^{x=+\infty} \\ &= f\left[\hat{x},y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right] \end{aligned}$$

where the penultimate equality follows because $\int\phi\left[a+bz\right]dz=\left(\Phi\left[a+bz\right]-\frac{1}{2}\right)\big/b$, and $Z'\left[x\right]/\sqrt{-Z''}=c-x\sqrt{-Z''}$.Here, a, b are arbitrary constants and c is a specifc but immaterial one.

□

Lemma 1. Let $\hat{X}\equiv X+\varepsilon_X$ and $\hat{Y}\equiv Y+\varepsilon_Y$, where $\varepsilon_X\tilde{}N\left[0,\sigma_{\varepsilon_X}^2\right]$ and $\varepsilon_Y\tilde{}N\left[0,\sigma_{\varepsilon_Y}^2\right]$. Then $\left\{\hat{X},\hat{Y}\right\}$ is distributed $N\left[\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\hat{\sigma}_Y^2\right),\hat{\rho}\right]$, where $\hat{\sigma}_X^2\equiv\sigma_X^2+\sigma_{\varepsilon_X}^2$, $\hat{\sigma}_Y^2\equiv\sigma_Y^2+\sigma_{\varepsilon_Y}^2$, and $\hat{\rho}\equiv\rho\frac{\sigma_X\sigma_Y}{\hat{\sigma}_X\hat{\sigma}_Y}$.

Proof. Follows directly from a second application of Proposition 3. □

Lemma 6.

$$\frac{1}{\sigma_\varepsilon\sqrt{2\pi}}\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}=\frac{\sqrt{-Z''}}{2\pi\sqrt{2\pi}\hat{\sigma}_X\sigma_Y\sqrt{1-\tilde{\rho}^2}}$$

Proof. The lemma holds iff

$$\frac{1}{\sigma_\varepsilon}\frac{1}{\sigma_X\sqrt{1-\rho^2}}=\frac{\sqrt{\frac{1}{\sigma_\varepsilon^2}+\frac{1}{(1-\rho^2)\sigma_X^2}}}{\sqrt{\sigma_X^2+\sigma_\varepsilon^2}\sqrt{1-\left(\rho\frac{\sigma_X}{\sqrt{\sigma_X^2+\sigma_\varepsilon^2}}\right)^2}}$$

which, after straightforward algebraic simplication, is true. □

Lemma 7.

$$Z\left[x\right]=V\left[\hat{x},y;\left(\mu_X,\hat{\sigma}_X^2\right),\left(\mu_Y,\sigma_Y^2\right),\tilde{\rho}\right]-\frac{1}{2}\left(Z'\left[x\right]\Big/\sqrt{-Z''}\right)^2$$

Proof.

$$
\begin{aligned}
& V\left[\hat{x}, y; \left(\mu_X, \hat{\sigma}_X^2\right), \left(\mu_Y, \sigma_Y^2\right), \tilde{\rho}\right] - \frac{1}{2}\left(Z'[x] \Big/ \sqrt{-Z''}\right)^2 \\
= & -\frac{1}{2}\left(\frac{\frac{(\hat{x}-\mu_X)^2}{\sigma_X^2+\sigma_\varepsilon^2} - 2\rho\sigma_X \frac{\hat{x}-\mu_X}{\sigma_X^2+\sigma_\varepsilon^2}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{1-\rho^2\frac{\sigma_X^2}{\sigma_X^2+\sigma_\varepsilon^2}} + \frac{\left(\frac{1}{\sigma_\varepsilon}\frac{\hat{x}-x}{\sigma_\varepsilon} - \frac{1}{\sigma_X(1-\rho^2)}\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right)\right)^2}{\frac{1}{\sigma_\varepsilon^2}+\frac{1}{(1-\rho^2)\sigma_X^2}}\right) \\
= & -\frac{1}{2}\left(\begin{array}{c} \frac{\frac{(\hat{x}-\mu_X)^2}{\sigma_X^2+\sigma_\varepsilon^2} - 2\rho\sigma_X\frac{\hat{x}-\mu_X}{\sigma_X^2+\sigma_\varepsilon^2}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{\frac{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2}{\sigma_X^2+\sigma_\varepsilon^2}} \\ + \frac{\frac{1}{\sigma_\varepsilon^2}\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 - 2\frac{1}{\sigma_\varepsilon}\frac{1}{\sigma_X(1-\rho^2)}\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right) + \frac{1}{\sigma_X^2(1-\rho^2)^2}\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right)^2}{\frac{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2}{\sigma_\varepsilon^2(1-\rho^2)\sigma_X^2}} \end{array}\right) \\
= & -\frac{1}{2}\frac{\left(\begin{array}{c} (\hat{x}-\mu_X)^2 - 2\rho\sigma_X(\hat{x}-\mu_X)\frac{y-\mu_Y}{\sigma_Y} + \left(\sigma_X^2+\sigma_\varepsilon^2\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \\ + \left(1-\rho^2\right)\sigma_X^2\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 - 2(\hat{x}-x)\sigma_X\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right) + \sigma_\varepsilon^2\frac{1}{1-\rho^2}\left(\frac{x-\mu_X}{\sigma_X} - \rho\frac{y-\mu_Y}{\sigma_Y}\right)^2 \end{array}\right)}{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2} \\
= & -\frac{1}{2}\frac{\left(\begin{array}{c} \hat{x}^2 - 2\hat{x}\mu_X + \mu_X^2 - 2\rho\sigma_X\hat{x}\frac{y-\mu_Y}{\sigma_Y} + 2\rho\sigma_X\mu_X\frac{y-\mu_Y}{\sigma_Y} + \sigma_X^2\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 + \sigma_\varepsilon^2\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \\ + \left(1-\rho^2\right)\sigma_X^2\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 - 2\hat{x}x + 2\hat{x}\mu_X + 2x^2 - 2x\mu_X + 2\rho\sigma_X\hat{x}\frac{y-\mu_Y}{\sigma_Y} - 2\rho\sigma_X x\frac{y-\mu_Y}{\sigma_Y} \\ +\sigma_\varepsilon^2\frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \rho^2\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\} \end{array}\right)}{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2} \\
= & -\frac{1}{2}\frac{\left(\begin{array}{c} \hat{x}^2 - 2\hat{x}x + x^2 + \left(1-\rho^2\right)\sigma_X^2\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 \\ +x^2 - 2x\mu_X + \mu_X^2 - 2\rho\sigma_X(x-\mu_X)\frac{y-\mu_Y}{\sigma_Y} + \sigma_X^2\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \\ +\sigma_\varepsilon^2\frac{1}{1-\rho^2}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \rho^2\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 + \left(1-\rho^2\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right) \end{array}\right)}{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2} \\
= & -\frac{1}{2}\frac{\left(\begin{array}{c} \left\{\sigma_\varepsilon^2 + \left(1-\rho^2\right)\sigma_X^2\right\}\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 \\ + \left\{\sigma_\varepsilon^2 + \left(1-\rho^2\right)\sigma_X^2\right\}\left(\frac{1}{1-\rho^2}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right) \end{array}\right)}{\sigma_\varepsilon^2+(1-\rho^2)\sigma_X^2} \\
= & -\frac{1}{2}\left(\frac{\hat{x}-x}{\sigma_\varepsilon}\right)^2 - \frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{2\left(1-\rho^2\right)} \\
= & U\left[\hat{x}; \left(x, \sigma_\varepsilon^2\right)\right] + V\left[x, y; \left(\mu_X, \sigma_X^2\right), \left(\mu_Y, \sigma_Y^2\right), \rho\right] = Z[x]
\end{aligned}
$$

□

Lemma 8.

$$\sum_{i=1}^{n} \frac{(\mu_i - x)^2}{\sigma_i^2} = \frac{(x - B/C)^2}{D/C} + \frac{A}{C} - \left(\frac{B}{C}\right)^2$$

where $A \equiv \sum_{i=1}^{n} \left\{\mu_i^2 \prod_{j \neq i} \sigma_j^2\right\}, B \equiv \sum_{i=1}^{n} \left\{\mu_i \prod_{j \neq i} \sigma_j^2\right\}, C \equiv \sum_{i=1}^{n} \left\{\prod_{j \neq i} \sigma_j^2\right\}, D \equiv \prod_{i=1}^{n} \sigma_j^2$.

Proof.

$$\sum_{i=1}^{n} \frac{(\mu_i - x)^2}{\sigma_i^2} = \frac{\sum_{i=1}^{n} \left\{\left(\mu_i^2 - 2\mu_i x + x^2\right) \prod_{j \neq i} \sigma_j^2\right\}}{\prod_{i=1}^{n} \sigma_i^2} = \frac{C}{D}x^2 - 2\frac{B}{D}x + \frac{A}{D}$$

Complete the square

$$= \frac{C}{D}\left(x + \frac{-2B/D}{2C/D}\right)^2 + \frac{A}{D} - \frac{(-2B/D)^2}{4C/D}$$

Simplifying yields the form of the lemma. □

Lemma 9.

$$\frac{\int_{-\infty}^{+\infty} x \prod_{i=1}^{n} \phi\left[\frac{\mu_i - x}{\sigma_i}\right] dx}{\int_{-\infty}^{+\infty} \prod_{i=1}^{n} \phi\left[\frac{\mu_i - x}{\sigma_i}\right] dx} = \frac{\sum_{i=1}^{n} \frac{\mu_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Proof.

$$\frac{\int_{-\infty}^{+\infty} x \prod_{i=1}^{n} \phi\left[\frac{\mu_i - x}{\sigma_i}\right] dx}{\int_{-\infty}^{+\infty} \prod_{i=1}^{n} \phi\left[\frac{\mu_i - x}{\sigma_i}\right] dx} = \frac{\int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\sum_{i=1}^{n} \frac{(\mu_i - x)^2}{\sigma_i^2}} dx}{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\sum_{i=1}^{n} \frac{(\mu_i - x)^2}{\sigma_i^2}} dx} = \frac{\int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{(x - B/C)^2}{D/C} + \frac{A}{C} - \left(\frac{B}{C}\right)^2\right)} dx}{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{(x - B/C)^2}{D/C} + \frac{A}{C} - \left(\frac{B}{C}\right)^2\right)} dx}$$

from Lemma 8. Further simplifying,

$$= \frac{\int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\frac{(x - B/C)^2}{D/C}} dx}{\int_{-\infty}^{+\infty} e^{-\frac{1}{2}\frac{(x - B/C)^2}{D/C}} dx} = \frac{\int_{-\infty}^{+\infty} x g[x] dx}{\int_{-\infty}^{+\infty} g[x] dx} = \frac{B}{C} = \frac{\sum_{i=1}^{n} \left\{\mu_i \prod_{j=1, j \neq i}^{n} \sigma_j^2\right\}}{\sum_{i=1}^{n} \left\{\prod_{j=1, j \neq i}^{n} \sigma_j^2\right\}} = \frac{\sum_{i=1}^{n} \frac{\mu_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

where $g[x] \equiv \frac{1}{\sqrt{2\pi}} \frac{C}{D} e^{-\frac{1}{2}\frac{(x - B/C)^2}{D/C}}$ is the density of a normal random variable with mean B/C and variance D/C. □

A1 Notes for Section 4.1

Consider an employer with discount rate $\delta \in (0, \infty)$ and time horizon $T \in (0, \infty)$. Here, the discount rate is the sum of the separation rate and the interest rate. The expected payoff $\pi_T[s_1, s_2]$

from hiring a candidate with signals s_1, s_2 is

$$\begin{aligned}\pi_T[s_1,s_2] &= \mathbb{E}_{\Theta,K_2}\left[\int_{t_2}^{T+t_2} e^{-\delta(t-t_2)}K_t dt\middle| s_1,s_2\right] = \mathbb{E}_{\Theta,K_2}\left[\int_{t_2}^{T+t_2} e^{-\delta(t-t_2)}\left(K_2+\Theta\left(t-t_2\right)+\sigma W_t\right)dt\middle| s_1,s_2\right] \\ &= \mathbb{E}\left[K_2\middle| s_1,s_2\right]\int_{t_2}^{T+t_2} e^{-\delta(t-t_2)}dt + \mathbb{E}\left[\Theta\middle| s_1,s_2\right]\int_{t_2}^{T+t_2} e^{-\delta(t-t_2)}\left(t-t_2\right)dt + \sigma\int_{t_2}^{T+t_2} e^{-\delta(t-t_2)}\mathbb{E}_{\Theta,K_2}[W \\ &= \mu_{2|s_1,s_2}\frac{1-e^{-\delta T}}{\delta} + \mu_{\Theta|s_1,s_2}\frac{1-e^{-\delta T}(1+\delta T)}{\delta^2} + 0\end{aligned}$$

Hence,

$$\pi_T\left[s_1',s_2\right] - \pi_T\left[s_1,s_2\right] = \left(\mu_{2|s_1',s_2}-\mu_{2|s_1,s_2}\right)\frac{1-e^{-\delta T}}{\delta} + \left(\mu_{\Theta|s_1',s_2}-\mu_{\Theta|s_1,s_2}\right)\frac{1-e^{-\delta T}(1+\delta T)}{\delta^2}$$

Let $s_1' > s_1$. In that case, $\mu_{2|s_1',s_2} > \mu_{2|s_1,s_2}$, while $\mu_{\Theta|s_1',s_2} \overset{(<)}{>} \mu_{\Theta|s_1,s_2}$ if $\sigma_{\varepsilon_2}^2 \overset{(<)}{>} \sigma_{\varepsilon_2}^{*2} \equiv \frac{\sigma_0^2\left(1-\rho^2\right)-\rho\frac{\sigma_0}{\sigma_\Theta}\sigma^2}{\rho\frac{\sigma_0}{\sigma_\Theta}+t_1}\left(t_2-t_1\right)$. Therefore, if $\sigma_{\varepsilon_2}^2 \geq \sigma_{\varepsilon_2}^{*2}$, the employer always hires the candidate with the higher first signal. When $\sigma_{\varepsilon_2}^2 < \sigma_{\varepsilon_2}^{*2}$, he hires the candidate with the higher (lower) first signal iff

$$\frac{\Delta\mu_2}{\delta}\left(1-e^{-\delta T}\right) \overset{(<)}{>} \frac{\Delta\mu_\Theta}{\delta^2}\left(1-e^{-\delta T}(1+\delta T)\right) \iff \delta \overset{(<)}{>} \frac{\Delta\mu_\Theta}{\Delta\mu_2}\frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}}$$

where $\Delta\mu_2 \equiv \mu_{2|s_1',s_2} - \mu_{2|s_1,s_2}$ and $\Delta\mu_\Theta \equiv \mu_{\Theta|s_1,s_2} - \mu_{\Theta|s_1',s_2}$

Single-crossing in δ:

- Let $T \to \infty$. Then this condition becomes $\delta \overset{(<)}{>} \frac{\Delta\mu_\Theta}{\Delta\mu_2}$.

  Hence, if $\sigma_{\varepsilon_2}^2 < \sigma_{\varepsilon_2}^{*2}$, then there exists a unique $\delta^* \in (0,\infty)$ such that the employer is indifferent between the two candidates. For all $\delta > \delta^*$ (impatience and/or high employee turnover), he hires the guy with the higher first signal. For all $\delta < \delta^*$ (patience and/or low employee turnover), he hires the guy with the lower first signal.
- Single-crossing also holds for finite T, although it takes remarkably many steps to prove that...:

  To prove it, it suffices to show that

$$\frac{d}{d\delta}\left[\frac{1}{\delta}\frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}}\right] < 0$$

Observe that

$$\frac{d}{d\delta}\left[\frac{1}{\delta}\frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}}\right]=\frac{e^{\delta T}\left(\delta^2T^2+2\right)-e^{2\delta T}-1}{\delta^2\left(1-e^{\delta T}\right)^2}$$

takes on the sign of the numerator. We can write this numerator as

$$e^x\left(x^2+2\right)-e^{2x}-1$$

where $x\equiv\delta T>0$. This expression is negative iff

$$f\left[x\right]\equiv 2+x^2-e^x-\frac{1}{e^x}<0$$

Now notice that

$$f\left[0\right]=2+0^2-e^0-\frac{1}{e^0}=0$$

while

$$\begin{aligned} f'\left[x\right] &= 2x-e^x-\left(-\frac{1}{\left(e^x\right)^2}e^x\right)=2x-e^x+\frac{1}{e^x} \\ f'\left[0\right] &= 0 \end{aligned}$$

Similarly,

$$f''\left[x\right]=2-e^x-\frac{1}{e^x};\ f''\left[0\right]=0$$

$$f'''\left[x\right]=-e^x+\frac{1}{e^x};\ f'''\left[0\right]=0$$

$$f''''\left[x\right]=-e^x-\frac{1}{e^x}<0 \text{ for all } x>0\ .$$

Reversing direction, $f'''\left[0\right]=0$ and $f''''\left[x\right]<0\ \forall x>0$ jointly imply that $f'''[x]<0\ \forall x>0$. In turn, together with $f''\left[0\right]=0$ $f'''[x]<0$, $\forall x>0$, implies that $f''[x]<0$, $\forall x>0$. Together with $f'\left[0\right]$, $f''[x]<0\ \forall x>0$, implies that $f'\left[x\right]\ \forall x>0$. Finally, together with $f\left[0\right]=0$, $f'\left[x\right]<0\ \forall x>0$, implies that $f\left[x\right]<0\ \forall x>0$, which implies that $\frac{d}{d\delta}\left[\frac{1}{\delta}\frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}}\right]<0$.

(At most) single-crossing in T: It seems that

$$\begin{aligned} \frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}} &\rightarrow 1 \text{ for } T \rightarrow \infty \\ \frac{1-e^{-\delta T}(1+\delta T)}{1-e^{-\delta T}} &\rightarrow 0 \text{ for } T \rightarrow 0 \end{aligned}$$

(I have not yet formally proved this.) Hence, for T sufficiently small, the employer always hires the guy with the higher first signal. Iff $\delta < \delta^*$, this flips for $T \rightarrow \infty$.

Observe that

$$\frac{d}{dx}\left[\frac{1-e^{-x}(1+x)}{1-e^{-x}}\right] = \frac{e^{-x}}{\left(e^{-x}-1\right)^2}\left(x-1+e^{-x}\right)$$

takes the sign of

$$f[x] \equiv x-1+e^{-x}$$

where $f[0]=0$. Since

$$f'[x] = 1-e^{-x} > 0$$

for $x > 0$, it follows that

$$\frac{d}{dx}\left[\frac{1-e^{-x}(1+x)}{1-e^{-x}}\right] > 0 \ .$$

Hence, we have (at most) single-crossing in T.